



[www.mecsjs.com](http://www.mecsjs.com)

*Multi-Knowledge Electronic Comprehensive Journal  
for Education and Science Publications (MECSJ)*

ISSN ONLINE ( **2616-9185** )

**ISSUE ( 2 ) , 2017**



ISSN ONLINE ( **2616-9185** )

**a**



## Table Of Content

[www.mecsj.com](http://www.mecsj.com)

| <b>Title Of Paper</b>  | <b>Page Number</b> |
|--|--------------------|
| <b>A comparison of most recent MapReduce joins algorithms</b>  | <b>1 - 34</b>      |
| <b>Challenges of Electronic Arbitration in Electronic Commerce transactions</b>                      | <b>35 - 43</b>     |
| <b>Investigating the Impact of Walking on humans Health</b>  | <b>44 - 53</b>     |
| <b>Operation strategy at Galanz</b>  | <b>54 - 67</b>     |
| <b>The Impact of Poor Planning and Management on the Duration of Construction Projects: A Review</b> | <b>68 - 85</b>     |

## Aim & Scope

Aims to enhance the level of published articles by improving integrity, morals and keeping papers in high professional standards to spread the good knowledge to all up-coming researches, scholars and scientists. So, the main objective of MECS is to advance the professionalism level of existed research, and to enhance the focus on new emerging trends.



## A comparison of most recent MapReduce joins algorithms

Majeed Bander Al-Rewili

Computer Science Engineer, University of Edinburgh

**E-mail:** Maj\_Rewili@hotmail.com

### **Abstract:**

In this interesting line of research, an attempt has been to overview different parallel processing platforms that implement MapReduce jobs. This survey provides a wide-ranging analysis of work and publications related to MapReduce framework to data, and it also can be used as a basis for further research and examination. The scope of this survey is focused on pre-processing, pre-filtering, partitioning, replication, load balancing, performance, memory space, communication cost, and query processing and optimization aspects in the light of big data analysis in MapReduce. Moreover, a set of efficient optimized and improved approaches in the context of analytical query processing and optimizing using MapReduce. It provides an added value to current research published yearly by introducing a comprehensive classification of recently presented papers in the era of join types using MapReduce. From data-centric perspective, the main topic of this approach is intended to highlight the importance of traditional problems of data management and analysis in the regard of efficient big data processing and analysis approaches.

**Keywords:** MapReduce, Hadoop, join types, multi-way join, theta-join, KNN join, top-k join, graph similarity join, semi join, filter join, bloom join, intersection join.



## I. INTRODUCTION

There are many systems have been developed primarily to adopt big data analysis such as Yahoo's PNUTS, Twitter Storm, LinkedIn's Kafka, and especially Google's MapReduce. MapReduce, because of its simplicity, transformed the receiving of big data and large-scale processing; it becomes the most common framework used for vast datasets analysis based on machine learning techniques. Apache Hadoop is an open source which implements MapReduce framework and it has performed high popularity in both academia and industry due to its widespread usage [7].

MapReduce implementation in DBMS supports a set of functions: storage management, data partitioning, data compression, storage management, query optimization, and indexing. Hadoop DB presents the strategies of partitioning and indexing for parallel DBMSs based on MapReduce framework. Hadoop DB architecture includes three layers: top layer, middle layer, and bottom layer. In top layer, Hive is extended to convert queries to MapReduce jobs. In middle layer, MapReduce infrastructure and HDFS are implemented including fault tolerance, shuffling data between nodes, and caching intermediate files. In bottom layer, there are a set of computing nodes distributed in side layer to run individual instance of PostgreSQL DBMS and to store data [17].

Hadoop is an open source implementation of MapReduce which is the most common framework increasingly used by many companies including huge number of users. Hadoop is mainly compound of two parts: Hadoop Distributed File System (HDFS) and MapReduce to achieve distributed processing. Hadoop contains various servers: Job-Tracker and Task-Tracker to perform MapReduce, and Name-Node, Secondary Name-Node, and Data-Node to manage HDFS. MapReduce supports parallel processing of vast datasets; it includes two functions: Map function and Reduce function. Any job which has to be performed by MapReduce should go through these two phases. Map function is also called mapper which takes input including key-value pairs. It also performs some computational processes on the input to produce intermediary outputs formed also with key-value pairs. While Reduce function, which is also called reducer processes the obtained results from Map function; the data are shuffled to perform reduce phase. Shuffle

phase sometimes takes time, network bandwidth, and other resources more than two main functions, Map and Reduce [7].



Data is stored by default in HDFS which consists of several Data-Nodes to store data. It also consists of Name-Node, a master node, to monitor Data-Nodes and maintain all Meta data. Data in HDFS is separated into multiple chunks that contain different Data-Nodes and equivalent in size. Two system processes are established, Job-Tracker and Task-Tracker, in MapReduce runtime. Job-Tracker is responsible to split a job into two phases: map and reduce that the user defines. It also arranges all tasks among different Task-Trackers. After that, Task-Tracker accepts the job and starts to process tasks assigned to map reduce functions. Task-Tracker will take a data chunk defined by Job-Tracker and apply map task on. Once every map task completes, all intermediate results are grouped into reduce tasks in order to obtain the results [18].

HDFS is a distributed file designed to store big data files in a stream data form with access pattern. It is designed to recognize and respond individual machines failures since it is potential to work on commercial hardware. The main workflow is as follows: data are copied to HDFS to perform MapReduce, and then results are also copied from HDFS. So HDFS is usually not the key storage of data. This typical workflow scenario of using HDFS obeys to an access model called write-once read-many. In this model, random access to file parts is essentially costly in comparison with sequential access since HDFS is optimized for streaming access of large files. Files are possible to be only appended; there is no file update support [7].

#### A. Query optimization

Query plan optimization using many plan generations and selection algorithms can be performed and developed to find optimal plan for relational DBMSs. In addition, MapReduce system can further improve these optimization algorithms. Query optimization algorithms and more elaborate algorithms are needed since MapReduce jobs usually run longer than relational queries.

Additionally, query execution time and query optimization time should be balanced to run fast relational optimization algorithms. To reduce the plan search space and to pipeline data between operators, only left deep plans are typically considered in most relational DBMSs. Query execution is more significant for efficiency so there will be no pipeline between the original MapReduce and the operator [17].



In this paper, we considered a set of papers related to MapReduce published early in main database journals and conferences from 2009 to 2016. We attempt to analyze the limitations of existing surveys' approaches related to MapReduce in order to outline their shortcomings and to make a comparison between them. In addition, we aim to define major encountered problems in terms of MapReduce tasks processing in order to provide categorization of entire work and research comprehensively according to the addressed problems. The main contribution of this paper is to present a powerful citation of current problems and their potential solving techniques and to talk about future work to improve novel systems in terms of MapReduce processing tasks. In our survey, we focus on the improvements of MapReduce framework by reviewing the primary MapReduce framework and its multiple implementations. Different approaches have been implemented using MapReduce framework since it has no real specification of the way of implementing components. Therefore, we compare the design and features of different well known implementations of MapReduce framework.

## II. BACKGROUND/LITERATURE REVIEW

In the following section, an overview is introduced to provide many techniques and methods presented in the literature in terms of MapReduce performance improvement. We organize the categorized approaches of MapReduce improvement in a specified classification based on the introduced improvement.

Many purposes have been realized to improve the usefulness of database operators via MapReduce algorithms especially in intensive applications. In MapReduce framework, Map function is able to easily support simple operators such as select and project, but it cannot achieve theta-join, equi-join, multi-way join, and similarity join [17].

### A. Multi way join

Multi-way join is more complex join implementation than binary join. It can be implemented either using only one MapReduce job which is called replicated join or using multiple MapReduce jobs (one job for every join). Multiple MapReduce jobs are used to execute multi-way join by achieving a series of equi-joins. Every single equi-join is performed by one MapReduce job, and



every result of one MapReduce job passes to next MapReduce job as input. Usually, several join orders can lead to different performance based on different query plans that can be generated [17].

In the paper of [20], a multi-way join was presented to compute a set of matrix multiplications among several relations. The proposed algorithm can reduce the number of binary multiplications by taking the advantage of multi-way join operation. The proposed algorithm was implemented based on MapReduce framework, which provides us an ability to achieve the scalability of large matrix multiplication. The paper took a different perspective differs than several papers have investigated matrix multiplication using MapReduce. In the paper, the concept of parallelism was employed in the expansion of the problem from binary multiplication to n-ary multiplication of the whole equation. The multiplication was translated into a join operation in database systems to facilitate the efficiency of the matrices storage and to easier matrices multiplication of the most common matrices in graph data. Three types of algorithms were implemented: S2, P2, and PM. The experiments were processed on real world graph data in the paper have demonstrated the capacity of the parallel m-way join to enhance the process of matrix multiplication. Because of using the raw key implementation, the parallel two-way join algorithm can balance the intra-operation parallelism and inter-parallelism approaches.

In the paper of [12], three-way joins on MapReduce was studied in order to utilize distributed computation of joins using clusters of many machines for efficient graph algorithms. A state-of-the-art MapReduce multi-way join algorithm was shown in the paper to provide the appropriateness of using it with huge datasets. The aggregation step can be integrated into a cascade of two-way joins if the join result needs to be summarized or aggregated to get more efficiency. In the paper, the focus was on three-way joins for MapReduce specially for social networks analysis. However, the result of the join should be preferably cascaded of two-way joins to reduce the communication cost. Multi-way join algorithms are divided into three sub-types including Replicated join, Star join, and Theta-join as shown as follows:

- Replicated join

Replicated join is performed by as a single MapReduce job to perform multi-way joins. There is a special case of replicated join called star join that perform all join conditions on the same attribute or a set of same attributes [17].

- Star join

Star join can be implemented by one MapReduce job by setting the map output key to be the join attribute and deploying load balanced if needed [17].

- Theta join

Theta-join or  $\theta$ -join is a join operator contains one of the following join conditions: ( $<$ ,  $>$ ,  $=$ ,  $<=$ ,  $>=$ , or  $!=$ ) [17]. In real practices, more specifically in complex relations, multi-way theta-join queries are powerful. The most challenging task is to minimize the total processing time span through the best schedule sequence of execution of MapReduce jobs by mapping multi-way theta-join query [34].

In the paper [21], a proposed algorithm to implement theta-join as a single MapReduce job was presented. The implementation was achieved without changing MapReduce framework by constructing proper functions of Map and Reduce. The goal of the paper is to minimize job completion time. To do this, an appropriate join matrix-to-reducer mappings was used to define a great diversity of join implementations. An algorithm was proposed called 1-Bucket-Theta that uses matrix-to-reducer mappings on any join which has output significantly fraction of cross product and on cross product especially. Moreover, even though the proposed algorithm, 1-Bucket-Theta, is slower than other faster algorithms, other algorithms cannot be identified as usable unless knowing the join result in advance or performing expensive an analysis. The proposed algorithm consists of M-Bucket class of algorithms that can exclude large regions of join matrix and reduce input-related costs to improve running time for any theta-join. The proposed approach does not need to change MapReduce model; it supports any theta-join in a single MapReduce job. Indeed, the proposed algorithm can be integrated with high level programming languages on top of MapReduce. There are better algorithms than 1-Bucket-Theta for selective join conditions. On the other hand, these





algorithms include an essential fraction of the join matrix cells that are unassigned to any reducer. In practice, finding enough of such matrix cells can be impossible or computationally very expensive due to complex user-defined join conditions and insufficient input statistics. Due to the lack of proof that better matrix-to-reducer mapping does not miss any output tuple, we cannot use it.

Extending current solutions from traditional distributed and parallel databases for multi-way theta-join queries is relatively difficult to fit huge data volumes. In the paper of [34], a study was conducted from cost effective perspective of the problem of efficient processing of multi-way theta-join queries based on MapReduce identification and scheduling. Efficient processing of multi-way theta-join has not never been fully explores although of many works have been done using key-value pair-based programming model to support join operations. The most challenging task is to minimize the total processing time span through the best schedule sequence of execution of MapReduce jobs by mapping multi-way theta-join query. The main solution provided in the paper includes two parts: using only single MapReduce job for efficient execution of chain-typed theta-join, and how to execute single MapReduce job or a set of MapReduce jobs in a certain order and the corresponding cost metrics. The method can achieve substantial improvement of the join processing efficiency compared to other widely adopted solutions. In fact, the work introduced for the first time the exploration and evaluation of multi-way theta-joins using MapReduce. In the work, a set of rules were established to decompose a multi-way join query, in order to evaluate the cost model to execute multi-way join query for both single MapReduce job and multiple MapReduce jobs. Thus, extensive experiments were conducted to validate the proposed cost model and the solution of multi-way theta-join queries, and to compare with the state-of-art solutions in terms of query evaluation efficiency. A Hilbert curve based space partition method was proposed in the paper to reduce the volume of copying data over network and to adjust the reduce tasks workload. Certainly, the proposed schema in resource restricted scenarios for scheduling can help to achieve the evaluation of complex join queries a near optimal time efficiency.

In the paper [13], a binary theta-join and pre-processing clustering algorithm were introduced in MapReduce framework. The optimal trade-off between the communication cost and the size of the input can be reached using the best-known algorithm which has high join selectivity. Thus, the improvements of the state-of-the-art have been presented when the join selectivity is low. In



addition, load imbalance was considered across reducers to decrease the communication cost and the maximum load of a reducer. The proposed algorithm in the paper is based on 1-Bucket-Theta and M-Bucket. 1-Bucket-Theta requires minimal statistical information and examines all tuples pairs, making it the most generic algorithm. M-Bucket-I is better than 1-Bucket-Theta in cases that the join selectivity is small. The worst-case behavior of 1-Bucket Theta matches the lower bounds for the binary theta-join problem, so an analysis was performed. Clustering histogram buckets were performed to improve these algorithms by achieving more efficient partitioning of histogram buckets to reducers. The imbalance across reducers, the maximum reducer input, and the replication rate are the main factors of the efficiency. In the paper, the results have revealed that load imbalance is not significantly affected by improving the replication rate and maximizing reducer input. The main difference between M-Bucket-O and M-Bucket-I is that the earlier aims to minimize the maximum reducer output, whereas the last aims to minimize the maximum reducer input. Join Matrix (JM) is used to operate M-Bucket-I partitioner. JM includes each cell corresponds to pair of histogram buckets; trying to create a region for each single reducer and fit cells in these regions. The main goal is to improve the quality of the partitioner phase by reducing rows and columns of JM. The results confirmed that the proposed partitioning algorithm provides up to 59% better time performance. Once the selectivity becomes lower and the number of the band condition increase, the improvements become more significant. However, the approach is not intrusive; it can be integrated with the existing state-of-the-art.

In the paper [31], a proposed algorithm called Strict Even Join (SEJ) was designed to partition multi-way theta joins into smaller groups and selects the best one based on one MapReduce job. Therefore, by calling SEJ algorithm, a dynamic algorithm is elaborated to optimize the multi-way theta joins. The experiments have proved the feasibility and efficiency of the proposed randomized algorithm. A method called largrangian was used to minimize the communication cost between map and reduce functions and to compute the estimated results per relation. The experiments in the paper have shown the efficiency and the stability of the proposed algorithm in terms of multi-way joins using one MapReduce job rather than cascades of two-way joins.

## B. Equi join

Equi-join is a special case of theta-join where join condition can be only "=" . MapReduce implementation follows strategies of earlier parallel database implementation on equi-join operator [17]. Equi-join exploits MapReduce key-equality which requires more complex join based data flow management. MapReduce provides balancing between mapper nodes easily due to its simplicity nature. However, in some cases, standard equi-join algorithm could delay job completion whether a reducer receives a larger shared work. For this reason, balance load between reducers can resolve the issue by minimizing the greater amount of work allocated to a reducer and then minimizing job completion time [21]. Equi-join implementation has four variant types: repartition join, map-only join, reduce-only join and semi join [17].

- Repartition join

The default join algorithm and the most basic equi-join implementation for MapReduce in Hadoop is repartition join which is the most general join method that can be implemented as one MapReduce job. In repartition join, map phase repartitions two tables and then tuples are shuffled with the same key. After that, the result of map phase is assigned to the same reducer which joins the generated tuples [17].

- Map-only join

Map-only join consists of only map phase; it partitions input data based on the join key and then shuffles it to the reducers. Map-only join can be implemented on co-partitioned relations based on the join key [17]. Map-side join is an algorithm without Reduce phase [24]. The data sets in addition to their partition are sorted by the same ordering. The two sets of data pre-partitioned into the same number of splits by the same partitioner. This algorithm buffers all records with the same keys in memory, as is the case with skew data may fail due to lack of enough memory [24].

- Replication side join

and direct Reduce-side join is an algorithm which performs data pre-processing in Map phase, during the Reduce phase [24]. The preprocessing is sorting for the keys. Semi-joins join is done filtering is used to filter the original data. The partitioner must split the nodes by the key. The



reducer should have enough memory for all records with a same key. It is the most time-consuming, because it contains an additional phase and transmits data over the network from one phase to another. The algorithm has to pass information about source of data through the network [24].

- Semi join

Semi-join can be implemented on MapReduce even it has been well studied in parallel database systems. It is efficient when the result of semi join is relatively small since it requires several MapReduce jobs and the result of semi join must be implemented first [17].

In the paper [3], a study of the properties hash-based and sort-based equi-join algorithms was focused in case of fully joining datasets loaded into the main memory. In large high performance distributed data processing system, building block of a single node setting is very important factor. When running analytical data processing services on hardware shared among parallel services, memory footprint is an important deployment consideration. The critical contributions of the work are: studying the impact of memory footprint for each join algorithm on the number of parallel queries can be achieved, in addition to improving query response time through allowing system implementers and query optimizers to use the optimal join algorithm. In addition, the impact of two physical characteristics of join algorithms regarding their input and output (data being hash partitioned on the join key and data being pre-sorted on the join key) was considered in the paper to measure the performance. To optimize complex query pipelines with multiple joins. In general, equi-join is expensive process and the improving the overall performance of main memory data processing is relatively a challenging task. The results showed that hash-based join algorithm performs faster than sort-based join algorithms in most cases. Thus, the hash-based algorithm consumes smaller memory footprint compared to sort-based algorithms. When join inputs is already sorted, sort-based algorithms become competitive. The main conclusion of the paper is that considering the physical characteristics of the input and output is required to achieve the best response time and consolidation for main memory equi-join processing.

### C. Similarity join

Similarity join is one of many applications of join conditions where the results are similar to the join condition value but not equal to exact value. Therefore, there have been many proposed algorithms to find top-k most similar pairs, k-nearest returned tuples from two relations, and KNN join which finds the similarity between tuples based on their distances [17]. String similarity joins have received considerable interest. String similarity join is widely applied that aims to find all string pairs based on user defined threshold and a given similarity function [25].

In the paper [27], an efficient set-similarity join algorithm was proposed based on MapReduce to achieve joins in parallelism. For end-to-end set-similarity joins, a three-stage approach was proposed that takes a set of records as input and provides a set of joined records according to the set-similarity condition. In order to minimize the need for replication and to balance the workload, an efficient data-nodes partitioning technique was proposed. Both self-join and R-S joins were used to control the amount of data-nodes in main memory. The data still does not fit into main memory of a node even of the use of the most fine-grained partitioning. Extensive experiments were conducted to get results along with the increasing size of real data sets in order to estimate the scaling up and the speed up of the proposed algorithm and their properties. By exploiting the properties of the MapReduce framework, a discussion of different ways efficiently applied was introduced in terms of multiple inputs, replication of join, and partitioning.

String similarity joins have received considerable interest to design new algorithms called MGjoin with the assistant of an inverted index. String similarity join is widely applied that aims to find all string pairs based on user defined threshold and a given similarity function. In the paper of [25], two-step-filter-and-refine was adopted by the proposed algorithm to identify similar string pairs adopted approach. The proposed algorithm can generate candidate pairs based on inverted index and verify the candidate pairs based on similarity join. On the other hand, the proposed algorithm could result in high verification cost caused by poor filtering power or greater power of filtering computational cost. The proposed approach was the first work to explore multiple prefix filtering method was performed based on different orders and a parallel extension of the algorithm.



Extensive experiments were conducted and have shown that the proposed approach outperforms other approaches mainly state-of-the-art methods in terms of scalability and efficiency.

In the paper of [6], a scalable string similarity join called MASSJOIN was presented based on MapReduce. The proposed approach supports both character based similarity functions and set based similarity functions. Existing partition based signature scheme was extended to perform set based similarity functions, which generates key-value pairs by utilizing the signatures. Using the proposed approach, key-value pairs were merged in order to reduce transmission cost and the number of key-value pairs. Therefore, light-weight filter units were incorporated into key-value pairs in order to improve the performance and omit the factors of increasing transmission cost. The significance of the proposed method was shown by conducting extensive experiments; the results proved that the performance of the proposed method is better than the state-of-the-art approaches.

#### *D. kNN join*

KNN join is useful tool mostly used in data mining applications and spatial multimedia databases. It can produce K Nearest Neighbors (KNN) from one relation for every point in another relation. Performing KNN joins efficiently is a challenging task since it involves both the join and NN search. Hence, the applications continue to expand with the amount of data need to process. KNN execution on large data stored in MapReduce is the main challenging and interesting task since it frequently needed in practice [30]. KNN join is costly operation since NN search and join are expensive especially when datasets are in large multi-dimensions. There has been little research on parallel KNN joins in large data since it incrementally increases being exponential rate of datasets and a challenging task. On the other hand, there have been many parallel algorithms in MapReduce for equi-joins, similarity joins, theta-joins, and spatial range joins. Hence, many challenging and interesting problems were encountered regarding implementing KNN joins in MapReduce [30]. K Nearest Neighbour KNN join is a primitive operation commonly implemented by various applications of data mining. KNN join is designed to find k nearest neighbours from one dataset for every object in another dataset. However, KNN is an expensive operation since it combines k nearest neighbour query and join operation. Moreover, performing KNN join on centralized



machine is difficult with the increasing volume of data [18]. In many application domains, K Nearest Neighbours is one of the popular methods used to achieve query point or a set of query points namely KNN-join. Many problems have received much effort to resolve and to adopt changes to the database specially in stand-alone systems and spatial databases. These problems may limit the efficiency of relational database management system large data applications [28]. Typically, KNN join operation correlates a data object located in one dataset with the corresponding k nearest neighbor located in the same or different dataset [29].

In the paper [30], a novel algorithm was proposed in order to implement parallel KNN joins on large data using MapReduce demonstrated by Hadoop. The extensive experiments in the paper have demonstrated the scalability, efficiency, and effectiveness of the proposed methods in large and synthetic datasets. KNN join is costly operation since NN search and join are expensive especially when datasets are in large multi-dimensions. In the work based on previous observation, a motivation pays an attention to explore the problems associated with KNN joins execution on large data in MapReduce. First, Block Nested Loop Join (BNLJ) was the basic approach was proposed and then it was improved using R-tree indices. The basic approach does not scale well for large and multidimensional data due to the quadratic number of partitions produced (number of dataset input blocks and reducers). MapReduce friendly was introduced to handle this limitation.

MapReduce friendly is an approximate algorithm dependent to multi-dimensional datasets mapping into single dimension. For example, transforming KNN joins and space-filling curves are converted to a set of single dimension range searches. The proposed algorithms presented in the work were applied in MapReduce framework and the above issues raised in Hadoop were handled. The extensive experiments conducted in the work have been implemented over large real datasets, and the results confirmed good approximation quality which constantly outperforms the basic approach. Parallel KNN join in MapReduce was studied in the work including proposing exact H-BRJ and H-zKNNj approximate algorithms.

In the paper [18], an investigation to perform KNN join using MapReduce was presented. In map phase, cluster objects are divided into groups, and then KNN join is performed on each group of objects independently in reduce phase. Hence, the proposed mapping mechanism is designed to exploit distance-filtering rules using Voronoi-diagram-based partitioning method in order to



minimize computational and shuffling costs. Two approximate algorithms were proposed to reduce number of replicas and then reduced shuffling cost. The primary contributions of the paper are: presenting an implementation of KNN joins for multi-dimensional and large volume datasets using MapReduce framework without any modification. Additionally, in order to perform KNN join, an efficient mapping method is designed to divide objects into groups; every group is processed by a reducer. The distances between data partitions are more closely between groups and reduce number of replicas. Moreover, a cost model was developed to compute the number of replicas resulted from shuffling process. The extensive experiments have been conducted demonstrate the efficiency, robustness, and scalability of proposed methods.

In the paper of [28], a new method to achieve both KNN and KNN join in relational database was integrated with further query conditions. The main purpose was to design an algorithm that has the least impact and trivial changes to relational algorithms of database engine. The proposed algorithm uses SQL operators that generate the best plan to be used by query optimizer without radical changes to the database. The proposed approach is guaranteed to find the best-estimated KNN exactly in logarithmic cost in terms of number of block accesses required using only a small number of random shifts for databases in any fixed dimension. The extensive experiments have been conducted have demonstrated the efficiency and practicality of the proposed approach mainly on large, real, and synthetic datasets.

In general, KNN-join is designed to handle static datasets but not frequently updated datasets, whereas KNN-join is an expensive operation since it applied on high dimensional data. In the paper of [29], a novel KNN join method was proposed namely KNN-join+ to provide an effective KNN results with no significant changes to high dimensional datasets. Additionally, the proposed method guarantees to answer KNN queries of the advanced applications with the least workload. The results have revealed the effectiveness of the KNN-join+ to fast process high dimensional KNN join queries in both static and dynamic datasets. The proposed approach outperforms the existing indexing techniques by providing the excellent and scalable choice to handle dynamic dimensional data when dimensions are high especially in terms of sequential scan.



### E. Top-k join

Many applications have used top-k similarity join to calculate the most top-k similar pairs among different data records in a dataset. Typically, the time performance in top-k join is a challenging issue with the increased applications that require processing vast datasets. However, traditional methods cannot easily find the top-k pairs in such massive amounts of data [4].

In the paper [14], a new class of queries called top-k multiple-type integrated query (top-k MULTI) was defined. The main role of top-k MULTI query is to treat several data types to find the relevance between the object and the query. It can deal with many data types such as relational, spatial, and textual data types. The main discrimination between traditional top-k query and top-k MULTI query is that the dependency of component scores on the top-k MULTI query to find final scores. Hence, traditional top-k spatial keyword query can be considered as an instance of top-k MULTI query. In the paper, an integration of the relational data type into the traditional top-k spatial keyword query to create top-k spatial keyword-relational (SKR) query to show the importance of top-k MULTI query. Additionally, an investigation of several approaches to process top-k MULTI query (hybrid index and separate index approaches) and top-k SKR query was presented. The key issue for top-k MULTI query processing is the Scalability due to the multiple data types integrated in a query. In hybrid index approach, all indices for top-k MULTI query are built in an integrated form creating multi-level indices. On the other hand, all individual indices are maintained independently in separate index approach. A new query processing method was proposed for the top-k SKR query called Separate SKR based on separate index approach. Therefore, two methods were presented based on hybrid index approach to the top-k SKR query through expanding characteristic methods for the top-k spatial keyword query. Finally, a comparison of the results of extensive experiments on top-k SKR query using real datasets was performed to measure the efficiency and scalability of the methods from storage and query performance perspectives. The results showed that Separate SKR was more efficient and scalable up to 13 times than extended hybrid index methods. Further, Separate SKR consumes storage space up to 3 times less than extended hybrid index methods. In conclusion, separate index method can be easily extended to encourage a new data type for top-k MULTI query.



The proposed algorithm in [4] namely RDD-based can perform top-k similarity join over large clusters on high dimensional data sets. In general, RDD-based algorithm involves four stages that load multiple high-dimensional records into HDFS to find the top-k closest pairs ordered based on Hamming distances to perform global top-k pairs. An efficient distance function based on Locality Sensitive Hashing (LSH) was developed to increase the process of top-k similarity join and comparisons. All pairs of LSH signatures are split into partitions to minimize the amount of data during the RDD running time. Therefore, the proposed algorithm is capable to calculate top-k closest pairs in parallelism by exploiting a serial computation strategy. The results of experiments have revealed the scalability and effectiveness of RDD-based proposed algorithm.

#### *F. Graph similarity join*

One of the advanced operations used in a wide range of academic, theoretical, real, and practical applications is to identify clusters or close-knit communities in graphs. Practical algorithmic heuristics are required to efficiently embrace the problem either the theoretical algorithm is computationally or inflexible expensive. A set of significant challenges remain in implementing these heuristics to work for large real world graphs such as irregular data access patterns, compound factors, scale of data, intensive operation computation, and better approximation restriction [26].

In a distributed framework like MapReduce, performing graph-analysis is a challenging task. Many approaches have been proposed for graph-analysis of algorithms, but they perform shuffling and storing phases which increase the cost of high communication in MapReduce [8]. Graph similarity joins is very important with the advent of massive graph-modelled data, since it is widely applied for many objectives such as data cleaning [5].

In the paper [8], a new design pattern for a family of iterative graph algorithms for MapReduce framework. The proposed method separates graph topology from the graph-analysis results. In each iteration step, each MapReduce node existing in the graph participates in graph-analysis task and reads the same partition of the graph locally. Each node also reads all the current analysis results from the distributed file system. Using merge-join, the results of iterations are correlated to each graph partition locally, in addition to generate and dump the new improved analysis results in

the graph partition to HDFS. The algorithm requires only one MapReduce job to perform pre-processing graph, and the actual analysis using repartition requires one map-based MapReduce job. All partial results are contained in HDFS which stores the result of map stage to perform merge-join between a partition of the graph and a global file. In detail, the method to perform graph-analysis used parallel merge-join between the partition of graph and a global table containing all partial results of each node. The map-based approach proposed in the paper outperforms the basic approach since it can improve the performance of graph-analysis. At end, the approach can reduce the communication cost and improve the performance by separating the graph topology from the graph-analysis.

A novel MapReduce-based algorithm called pClust-mr was proposed in the paper of [26] for a popular serial graph clustering. Thus, a novel application of the proposed method was developed to cluster biological graphs more specifically to identify dense sub graphs from bipartite graphs. The proposed algorithm uses pipelined MapReduce stages to implement a mixture of shuffling and sorting operations in order to process the edges of the graph as an input. The results have revealed the linear scaling of the time performance on small real world graphs.

In the paper of [5], graph similarity joins are considered under edit distance limitations in order to find the pair of closest to each other lower than a specified threshold. With the use of MapReduce programming model, a scalable algorithm was proposed namely MGSJoin, which applies filtering verification framework to perform the most efficient graph similarity join. The main idea of the algorithm is to count the overlapping graph signatures with filtered candidates. Spectral Bloom filters are introduced to minimize the number of key-value pairs with the potential issue of too many key-value pairs in filtering phase. In addition, multi-way join strategy was integrated to increase the efficiency of GED calculation for verification based on MapReduce. The proposed algorithm is efficient and scalable with prove of extensive empirical experiments demonstration. In the paper, the main focus is on graph similarity join mainly for graph processing data. For example, suppose there are two graph object sets with distance threshold, and we have to return graph similarity join including all pairs of graph objects contained in these two graphs in terms of the lowest distances between them. In pre-processing of graph data mining, graph similarity join has a wide range of applications. The main contribution of the work is to present MapReduce based graph



similarity join algorithm to redesign the current in-memory graph similarity join algorithm. Moreover, large-scale graph datasets can be processed as a resulting baseline of the algorithm. More specifically, Bloom filter capacity was proposed to minimize intermediate key-value pairs. Therefore, optimized verification strategy was presented by multi-way join algorithm that can reduce number of rounds of MapReduce. The results have demonstrated the efficiency and scalability of the proposed algorithm against current solutions with the implementation in real publicly available datasets conducting in wide range of applications.

### G. Bloom join

A Bloom filter is a space-efficient [probabilistic data structure](#); that is used to test whether an [element](#) is a member of a [set](#). A query returns either "possibly in set" or "not in set". Elements can be added to the set, but not removed. An empty Bloom filter is a [bit array](#) of  $m$  bits, all set to 0. There must also be  $k$  different [hash functions](#) defined, each of which [maps](#) or hashes some set element to one of the  $m$  array positions with a uniform random distribution.  $K$  is a constant, much smaller than  $m$ , which is proportional to the number of elements to be added. To add an element, feed it to each of the  $k$  hash functions to get  $k$  array positions. Set the bits at all these positions to one [19].

It reduces transmission cost. Bloom join with open source map-reduce framework of Hadoop improves the performance of query optimization. The reduce side join applied bloom filters which is inexpensive than map-side join [19]. There are two kinds of cases needing to be considered: two-way joins; that occurs between two data sets, and multi-way joins; that occurs between more than two data sets, and it is implemented by a sequence of two two-way joins [24].

Bloom filter: a type of the map-reduce join, the relation queue is used to decide which relations must be further processed. The memory space needed to store a bloom filter is small compared to the amount of data belonging to the set being tested. For improving the performance of query execution the reduce side join is used with filtering on the map side which generates less I/O cost, but there remain many non-joining tuples after filtering [23]. The individual input records can be processed in parallel. Map function does not only tag the input records but also filters them allowing only some of them to be part of the final map output, there is no replication for the



elements. The input to the map function is file split. The hash functions and reduces the total processing cost. It reduces transmission cost, it reduces the amount of data transferred compared to semi-join by utilizing the concept of bloom filters [19]. Two-way join needs less memory space than multi-way join [32].

#### H. MRFA join

This algorithm, used to manage huge amount of data on large scale systems even for highly skewed data. It is Map/Reduce Frequency Adaptive Join algorithm based on distributed histograms -and randomized redistribution approach. The support for fault tolerance and load balancing in Map Reduce and Distributed File System are preserved if possible: the inherent load imbalance due to repeated values must be handled efficiently by the join algorithm and not by the Map/Reduce framework. Join computation in MRFA-Join proceeds in two map-reduce jobs. First, one phase to compute distributed histograms and to create randomized communication templates to redistribute only relevant data while avoiding the effect of data skew, Map phase to generate a tagged "local histogram" for input relations then Reduce phase to create join result global histogram index and randomized communication templates for relevant data [10].

Second, another phase is used to generate join output result by using communications templates carried out in the previous step; Map phase to create a local hash table and to redistribute relevant data using randomized communication templates, then Reduce phase to compute join result. The detailed information provided by distributed histograms, allows to reduce communications costs to only relevant data while guaranteeing perfect balancing processing due to the fact that, all the generated join tasks and buffered data do not never exceed a user defined size using threshold frequencies. This makes the algorithm scalable and outperforming existing map-reduce join algorithms which fail to handle skewed data whenever join tasks cannot fit in the available node's memory. MRFA-Join can also benefit from Map/Reduce underlying load balancing framework in heterogeneous or a multi-user environment since MRFA-Join is implemented without any change in Map/Reduce framework. The overhead related to distributed histograms processing remains very small compared to the gain in performance and communication costs since only relevant data is processed or redistributed across the network [10].



MRFA-Join: a general join framework with filtering techniques in Map-Reduce. To avoid the effect of repeated keys, Map user-defined function should generate distinct output keys even for records having the same join attribute value. The load balance is perfect. To compute the join of two datasets, the input relations are divided into blocks (splits) of data in distributed histograms and to buckets in a randomized key redistribution approach. These splits are also replicated on several nodes for reliability. The detailed information provided by these histograms, communication costs is reduced to only relevant data processing due to the fact that all the generated join tasks and buffered data never exceed a user defined size. The overhead related to distributed histograms processing remains very small compared to the gain in performance and communication costs since only relevant data is processed or across the network. The global redistribution cost is reduced to a minimum [10].

#### 1. Intersection filter

The intersection filters can filter out disjoint elements between two datasets, there are three approaches used to build the intersection filter. First approach is a pair of Bloom filters; specifying the set of intersection by eliminating all disjoint elements between the input datasets. Then filter out the disjoint elements in the input datasets by applying pair of bloom filters on the input datasets. Each tuple in one input dataset is queried into the Bloom filter of the other input dataset by  $k$  hash functions. If its join key is a member of the filter, the tuple containing this key will be returned because the key is a common member of the two input datasets. Otherwise, the tuple will be removed from its dataset because its join key is a disjoint member and this tuple is a non-joining tuple. This approach does not require the filters to have the same size  $m$  and  $k$  hash functions. The second approach is Intersection of un-partitioned Bloom filters; this approach is based on the idea that intersecting Bloom filters will produce a result filter called the intersection filter. There is unfortunately little difference between the intersection filter and the intersection of Bloom filters, it is used un-partitioned Bloom filter, only one intersection Bloom filter is used to remove most non-joining tuples from the input datasets instead of using two filters as the first approach. It should use the un-partitioned Bloom filters with the same size  $m$  and  $k$  hash functions [23].



The last approach is Intersection of partitioned Bloom filters; in this approach the partitioned Bloom filters are used to create the intersection filter. The size of partitioned Bloom filters can be changed after they are created. The filters may have different sizes but their partitions should have the same size. The intersection filter is generated by intersecting pairs of partitions of two partitioned filters. The intersection filter is generated by intersecting pairs of partitions of two partitioned filters. Two filters with 3 partitions are pair wise intersected with the bit-wise and to produce the result filter including three 4-bit partitions. This intersection filter represents the approximate intersection of the two datasets. If there exists at least one partition of the result filter containing all  $m/k$  bits equal to 0, the two input datasets are disjoint. So the join processing can be finished without doing anything. The pre-processing step is written as a standard map include two jobs running in parallel to process the input datasets (R and S) to build the intersection filter [23].

Intersection filter: Three approaches proposed to compute the intersection filter; intersection to Bloom filters, un-partitioned and Partitioned Bloom Filters. It filters out disjoint elements or non-joining tuples from both datasets, not only on one input dataset [23]. The intersection filter uses hash functions to portion the entire data sets. The memory space for first approach is small [11], but un-partition needs less memory space than partition [23]. The first approach needs to maintain two filters while others require one filter on nodes. A pre-processing enables a dramatic reduction in I/O and computational overhead. It produces much less intermediate data. Join processing in intersection filter can minimize disk I/O and communication costs. It is more effective through a cost-based comparison of join using different approaches. The preprocessing increase total cost, but it is small compared with other algorithms [22].

#### J. Parallel join: semi join

There are three ways to implement the semi-join operation; a semi-join using bloom-filter, semi-join using selection, an adaptive semi-join. The preprocessing in adaptive and selection semi-join is unique finding keys which are present in two datasets, and the relation queue is used to decide which relations must be further processed in bloom filter [10]. Delete the tuples that will not be used in join by using the filter will reduce the amount of data transferred over the network and the size of the dataset for the join. These filtering techniques introduce some cost, the semi-join can

improve the performance but the larger data sets will decrease the performance. There is no replication on the data sets. The additional information about the source of data will increase the data transferred. Memory space can be large depend on the size of the input data sets, but it can improve performance and reduce the possibility of memory overflow [24].

When a large portion of the data set does not take part in the join, deleting of tuples that will not be used in join significantly it will reduces the amount of data transferred over the network and the size of the dataset for the join. These filtering techniques introduce some cost, so the semi-join can improve the performance of the system only if the join key has low selectivity. There are three ways to implement the semi-join operation [24]. Parallel join is one of the most expensive operations in terms both I/O and CPU costs.

- A semi join using bloom filter

There are two jobs to perform the semi join. The Map phase and the Reduce phase. In the Map phase, the keys from one set are selected and added to the Bloom-filter. In the Reduce phase combines the output from Map phase into one. The second job filters only the output of the Map phase, increasing the size of the bitmap will increase the accuracy on this approach, but will increase the amounts of memory space needed. The advantage of this method is it's the compactness. The performance of the semi-join using Bloom-filter highly depends on the balance between the Bloom-filter sizes, which increases the time needed for its reconstruction of the filter in the second job, the large size of the data set can decrease the performance of the join [24].

- A semi join using selection

Semi-join with selection extracts unique keys and constructs a hash table. The hash table created in the first step filters the second set. In the context of Map-Reduce, the semi-join is performed in two jobs. Unique keys are selected during the Map phase of the first job and then they are combined into one file during the Map phase. The second job consists of only the Map phase, which filters out the second set. The semi-join using selection has some limitations. Hash table in memory, based on records of unique keys, can be very large, and depends on the key size and the number of different keys [24].



- The adaptive semi join

The Adaptive semi join is performed in one job, but filters the original data on the flight during the join. Similar to the Reduce-side join at the Map phase the keys from two data sets are read and values are set equal to tags which identify the source of the keys. At the Reduce phase keys with different tags are selected. The disadvantage of this approach is that additional information about the source of data is transmitted over the network [24].

#### K. Filter join

This type of join focused on reducing the number of map output records that are not joined. The map output records are replicated multiple times, so filtering out redundant records removes multiple copies of the record in multi-way joins. To join number of datasets simultaneously, some datasets need to be replicated. Replication may degrade the join performance, so it is important to reduce the number of redundant records. There are some filtering techniques to multi-way joins. Multi-way joins can be classified into two types: common attribute joins and distinct attribute joins. A common attribute join combines datasets based on one or more shared attributes, whereas some relations do not have join attributes in a distinct attribute join [16].

- Common attribute joins

The entire input datasets share joins attributes. The input records do not need to be replicated and they can be processed in a similar manner to two-way joins. A set of filters is created and probed in turn depending on the processing order of the input datasets [16]. There is no need to duplicate the entire datasets, the first data set used to make a set of filters to the next dataset. The cost depends on the number of input records, the ratio of the joined records, and the false positive rate of the filters; it is efficient when small portions of records participate in joins. There is no partitioning for input dataset. It needs a small memory space [16].

- Distinct attribute joins

Distinct attribute joins required the replication of some input datasets. The star-dim pattern delivered the best performance, chain join has the least performance between them. There are some equations can be used to select the processing order of the input data sets, and to estimate

the join cost, because the processing order must be selected carefully because it affects the join cost, but there may be a large search space if the numbers of reducers and the input datasets are large, cost for star-fact is less than chain join but more than star-dim. The replication of input records for their corresponding reducers can be implemented in a similar manner to the data partitioning. The input datasets may not have some join attributes. Thus, some of the datasets with missing attributes need to be replicated because their records may be joined to the input records of other datasets with any values of the missing attributes. The filters can be applied in three patterns: chain, star-fact, and star-dim [16].

- Chain

The chain pattern creates filters similar to common attribute joins, except that each set of filters is created for a different join attribute [16].

- Star fact

The star-fact pattern creates filters using the dataset with both join attributes and uses the filters to process the other datasets [16].

- Star-Dim

The star-dim pattern creates filters using the datasets with missing join attributes and uses the filters to process the other dataset [16].

#### *L. SJMR: parallelizing spatial join*

Spatial join merges two spatial data sets with a spatial relationship between the objects. Spatial join is commonly used in applications such as spatial robotics, DBMS, and game programming. Given two sets of multidimensional objects in Euclidean space, a spatial join query can discover all pairs of objects satisfying a given spatial relation, such as intersection. The input dataset is preprocessed to extract some key attributes. The input datasets are replicated to several partitions at Map stage. The filter is used to remove the tuples that cannot be parts of the result. This algorithm depends on good load balancing strategies. The grid partitioning method is used to divide the random data among  $n$  processors, the performance of SJMR improves by the increasing of node number, it's performance is high compared with PPBSM algorithm. Reduce task number and

memory size of each node is increased to make the memory size large enough to filter and refine all in memory without writing operations so it needs large memory space. This method could only proceed when Reduce stage has finished completely, so its cost is high [33].

#### *M. Massively parallel sort merge (MPSM) joins*

MPSM join is a new sort-based parallel join method scaling almost linearly with the number of clusters. Therefore, this sort-based join outperforms hash-based parallel join algorithms on modern multi-core servers. Sort-based algorithms formed the basis for multi-core optimization in recent proposed approaches [1]. There are three type of this algorithm:

- B-MPSM algorithm

It is the basic form of MPSM algorithm, which is unaffected by any kind of skew. It allows some similarity to fragment and replicate distributed join algorithms. It only replicates merge join scans of the threads/cores but does not duplicate any data [1].

- P-MPSM algorithm

It is an improved MPSM version based on range partitioning of the input by join keys [1].

- D-MPSM algorithm

The MPSM can be effectively modified to non-main memory scenarios, in which intermediate data must be written to disk [1].

B-MPSM starts by generating sorted runs in parallel. There is no data duplication but instead data is partitioned into equal sizes. B-MPSM performs a number of worker threads of sort-merge joins in parallel. At result, it a large memory space is needed. P-MPSM uses join keys to split the input data sets. D-MPSM uses a small part of memory to store the data during join processing so it is a RAM-constrained version that spools runs to disk. The performance depends on the number of threads running in parallel. MPSM adds only very little overhead to the overall join processing. The algorithms take advantage of massive thread parallelism, fast inter-processor communication through local memory. Memory space needed for B-MPSM and P-MPSM are large but it is small for D-MPSM [1].



After deeply reviewed papers and works in the field of MapReduce join algorithms, we are motivated to make a comparison between these investigated papers. The comparison will be based on the following criteria (technical aspects of join algorithms): pre-processing, filtering, partitioning, replication, load balancing, performance, memory space, communication cost, query processing, and query optimization. Compared information is displayed in Table 1. It shows different MapReduce join algorithms from literature analyzed from the above-mentioned perspectives.

We classify MapReduce join algorithms into several categories including Multi way join, Equi-join, Similarity join, and Bloom join filter. Further, Multi way join algorithms are also divided into: two-way or three-way; replicated join, and star join, or theta-join. In Equi-join, the algorithms are also subdivided into: repartition join, map-only join, replication side, and semi join. Moreover, similarity join algorithms are classified into: top-k, KNN, string similarity, and graph similarity. Lastly but not least, Bloom join filter is either intersection filter or parallel join

## RESULTS AND DISCUSSION

In terms of theta-join, we investigated four of the most recent papers in MapReduce framework: M-Bucket-Theta [13], M-Bucket [21], MRJ [34], and Random algorithm SEJ [31]. M-Bucket-Theta in [13] a binary theta-join and pre-processing clustering algorithm were introduced. In addition, load imbalance was considered to decrease the communication cost and the maximum load of a reducer. The proposed algorithm in the paper is based on 1-Bucket-Theta that requires minimal statistical information and examines all tuples pairs. The results confirmed that the proposed partitioning algorithm provides up to 59% better time performance. However, [31] still perform join performance less than M-Bucket [21] that implements theta-join as a single MapReduce job without changing MapReduce framework. It implements memory-aware approach, minimizes total cost, and has better performance. Another examined paper was MRJ [34] that provides a solution using only single MapReduce job for efficient execution of chain-typed theta-join near optimal time efficiency. The method can achieve substantial improvement of the join processing efficiency compared to other adopted solutions. A Hilbert curve based space partition method was proposed in the paper to reduce the volume of copying data over network and to adjust the reduce tasks workload. The last approach is Random algorithm SEJ [31] that performs efficient multi-way joins using one

MapReduce job rather than cascades of two-way joins due to the use of largrangian method to reduce communication cost and to increase performance.

In terms of KNN join, we investigated four of the most recent papers in MapReduce framework: PGBJ [18], kNNJoin+ self-join [29], H-zkNNJ [30], and  $z\chi$ -kNN [28]. The proposed method in [18] was designed based on mapping mechanism that exploits distance-filtering rules using Voronoi-diagram-based partitioning method in order to minimize computational and shuffling costs. PGBJ achieved poor performance due to the large memory consumption. However, kNNJoin+ self-join proposed in [29] is more effective that results with no significant changes to high dimensional datasets and with the least workload. , kNNJoin+ outperforms other indexing techniques by providing the excellent and scalable choice to handle dynamic dimensional data when dimensions are high. Another examined approach is H-zkNNJ in [30] that proposed Block Nested Loop Join (BNLJ) for scalability requiring only linear number of blocks, dataset input and reducers. The problems and issued addressed in the study are: the issue of reducing the amount of communication occurred between Map and Reduce phases, the issue of performing random shifts in MapReduce, and the issue of designing a good partition over single dimensional z-values for joins purpose. In [28],  $z\chi$ -kNN uses SQL operators that generate the best plan to query optimizer without radical changes to the database.  $z\chi$ -kNN approach is guaranteed to find the best-estimated KNN exactly in logarithmic cost in terms of number of block accesses.

In terms of top-k join, we investigated two of the most recent papers in MapReduce framework: Top-k MULTI query [14] and RDD-based algorithm [4]. First, Top-k MULTI query in [14] performed efficient and scalable storage and query performance using Separate SKR. Further, Separate SKR consumes storage space up to 3 times less than extended hybrid index methods. However, RDD-based algorithm proposed in [4] is based on Hamming distances and distance function based on Locality Sensitive Hashing. RDD-based algorithm can perform top-k similarity join over large clusters on high dimensional data sets.

In terms of Graph similarity join, we investigated three of the most recent papers in MapReduce framework: MGSJoin [5], pClust-mr [26], and Map-based graph analysis [8]. MGS Join in [5] applies filtering verification framework in graph similarity join mainly for graph processing data. MGSJ join



was presented to redesign the current in-memory graph similarity join algorithm and to use the capacity of Bloom filter capacity to minimize intermediate key-value pairs. Therefore, it used an optimized verification strategy by multi-way join algorithm to reduce number of rounds of MapReduce which results in more efficient and scalable algorithm against other solutions. PClust-mr in [26] was proposed for serial graph clustering using pipelined MapReduce stages to implement a mixture of shuffling and sorting operations. PClust-mr results in linear scaling of the time performance on small real world graphs but it still needs to be improved for large graphs. Map-based graph analysis in [8] requires only one MapReduce job to perform pre-processing graph. It uses parallel merge-join to generate and dump the new improved analysis results in the graph partition to HDFS. It outperforms other approaches due to the improved performance of graph-analysis and the reduced communication cost by separating the graph topology from the graph-analysis.

In terms of String similarity join, we investigated three of the most recent papers in MapReduce framework: MASSJOIN [6], PPJoin+ [27], and MGJoin [25]. MASSJOIN in [6] supports both character based similarity functions and set based similarity functions which generates key-value pairs by utilizing the signatures. MASSJOIN can reduce transmission cost and the number of key-value pairs using lightweight filter units to improve the performance better than others and to omit the factors of increasing transmission cost. PPJoin+ in [27] is a three-stage approach and an efficient data-node partitioning technique proposed to minimize the need for replication and to balance the workload for end-to-end set-similarity joins. However, PPJoin+ still does not fit into memory with the increasing size of real data sets even with the use of self-join and R-S joins to control the amount of data-nodes in main memory. MGJoin in [25] results in high verification cost caused by poor filtering power or greater power of filtering computational cost. MGJoin adopted two-step-filter-and-refine, which was the first work to explore multiple prefix filtering method based on different orders and a parallel extension of the algorithm.

In terms of Multi-way join, we investigated two of the most recent papers in MapReduce framework: m-way: S2, P2, and PM [20] and three-way join [12]. M-way in [20] uses three types of algorithms were implemented: S2, P2, and PM. M-way join can reduce the number of binary multiplications by taking the advantage of multi-way join operation. It has demonstrated the



capacity of the parallel m-way join to enhance the process of matrix multiplication differs than the rest of papers. It also can balance the intra-operation parallelism and inter-parallelism approaches because of using the raw key implementation and parallel two-way join algorithm. Three-way join in [12] utilizes distributed computation of joins using clusters of many machines for efficient graph algorithms. It uses a cascade of two-way joins if the join result needs to be summarized or aggregated to get more efficiency. However, the result of the join should be preferably cascaded of two-way joins to reduce the communication cost.

In terms of Equi-join, we investigated only one of the most recent papers in MapReduce framework: STRSM [3]. It studied the impact of memory footprint for each join algorithm on the number of parallel queries to improve query response time. It allows system implementers and query optimizer to use the optimal join algorithm and to optimize complex query pipelines with multiple joins. However, hash-based join algorithm performs faster and consumes smaller memory footprint compared to sort-based algorithms in most cases.

Two-way join is less efficient than the improved repartition join especially if the size of the relation is small. Therefore, two-way join needs additional Map-Reduce rounds to build the bloom-filters. As a result, it is more efficient than improved repartition join. However, when the size of the relation grows to over 50 million records, the bloom-filters can filter a lot of useless data to save network overhead and processing overhead. The bloom-filter can be used to filter useless data and eventually improve the efficiency of the two-way join and multi-way joins [24].

Bloom Filter, which works better than Semi-Join, reduces amount of data transfer between different sites and performs efficient query processing. Bloom join with open source map-reduce framework of Hadoop improves the performance of query optimization [19].

MRFA-Join algorithm ensures of perfect balancing properties during all stages of join computation [10]. The intersection filter has an extra cost for the preprocessing step, but its efficiency in space-saving and filtering often outweighs these shortcomings [23]. However, its performance is least compared with other join algorithms like bloom join and Reduce-Side-Join [22].

The un-partitioned intersection filters seem more efficient than the joins using the partitioned intersection filter because of their filtering performance. However, the partitioned intersection filter can easily discover disjoint datasets on a join key column and stop the join processing [22].

Common attribute filter joins and distinct attribute filter joins significantly outperformed the repartition join on the other hand, MFR-Join outperforms them and the semi-join with bloom filters [16]. Moreover, common attribute filter joins and distinct attribute filter joins improve the execution time significantly by reducing the amount of intermediate results when small portions of input datasets are joined [16].

The performance of SJMR algorithm was compared with the performance of PPBSM algorithm; the performance of PPBSM is less than the performance of SJMR algorithm. On the other hand, SJMR algorithm uses a technique called reference tile method, which is an improvement to reference point method. Instead of using a duplication and elimination operator at the end of SJMR, it is better to avoid producing duplicates online per reduce job. As a result, the filter step in [33] was modified through a simple test applied during the intersection's checking of rectangles. The results are illustrated in Table 1.

## I. CONCLUSION AND FUTURE WORK

We have produced big effort to create this survey that mainly specialized to MapReduce programming model and software framework. MapReduce is intended to facilitate and simplify the processing of massive amount of data through large clusters of commodity hardware in parallelism, reliable, and fault-tolerant manner. We summarized a number of join algorithms in introduced under Map-Reduce framework, and we compared between them based on a set of [criteria](#). Per join type, some of the investigated approaches showed better performance than other approaches. Several join algorithms have presented disparate results in terms of pre-processing, filtering, partitioning, replication, load balancing, performance, memory space, and total cost. The performance of each algorithm depends on the size and duplicates of the input data sets. We can say that the preprocessing step can improve the performance but it requires additional cost such as that there is a specific algorithm incorporating filtering techniques. However, we cannot conclude has the best and perfect performance of the solution.





We will extend this research again by incorporating other comparison criteria and involving other recently published papers. We also can consider other join types that have not explored yet such as block-nested loop join, hash join, symmetric hash join, natural-join, and self-Join. In addition, we have an opportunity to follow-up enhancements shown by the previous algorithms. This survey is ongoing; it has a chance to be continuously up to date according to the improvements and developments to MapReduce framework as well as the enhancement of distributed and parallel computing based on MapReduce paradigm. One important opportunity for future work is to benefit every researcher interested in this area of research in order to resolve some encountered problems and shortcomings outlined in this paper. Conversely, this research acts as an initiation point to enhance being proposed algorithms in the field of big data analysis techniques based on MapReduce.

## ACKNOWLEDGMENT

We specially thank Dr. Al-badarneh, Amer for his assistance and comments that greatly improved this work, and our parents for their support.

## REFERENCES

- [1] Albutiu, M. C., Kemper, A., & Neumann, T. (2012). Massively parallel sort-merge joins in main memory multi-core database systems. *Proceedings of the VLDB Endowment*, 5(10), 1064-1075.
- [2] Andreas, C. (2010). *Designing a Parallel Query Engine over Map/Reduce* (Doctoral dissertation, Master's thesis, Informatics MSc, School of Informatics, University of Edinburgh).
- [3] Blanas, S., & Patel, J. M. (2013, October). Memory footprint matters: efficient equi-join algorithms for main memory data processing. In *Proceedings of the 4th annual Symposium on Cloud Computing* (p. 19). ACM.
- [4] Chen, D., Shen, C., Feng, J., & Le, J. (2015). An Efficient Parallel Top-k Similarity Join for Massive Multidimensional Data Using Spark. *International Journal of Database Theory and Application*, 8(3), 57-68.
- [5] Chen, Y., Zhao, X., Xiao, C., Zhang, W., & Tang, J. (2014). Efficient and Scalable Graph Similarity Joins in MapReduce. *The Scientific World Journal*, 2014.



- [6] Deng, D., Li, G., Hao, S., Wang, J., & Feng, J. (2014, March). Massjoin: A mapreduce-based method for scalable string similarity joins. In Data Engineering (ICDE), 2014 IEEE 30th International Conference on (pp. 340-351). IEEE.
- [7] Doukeridis, C., & Nørnvåg, K. (2014). A survey of large-scale analytical query processing in MapReduce. *The VLDB Journal*, 23(3), 355-380.
- [8] Elmasri, R., & Navathe, S. (2009). *Fundamentals of database systems*. 人民邮电出版社.
- [9] Gupta, U., & Fegaras, L. (2013, October). Map-based graph analysis on MapReduce. In *Big Data, 2013 IEEE International Conference on* (pp. 24-30). IEEE.
- [10] Hassan, M. A. H., Bamha, M., & Loulergue, F. (2014). Handling data-skew effects in join operations using mapreduce. *Procedia Computer Science*, 29, 145-158.
- [11] Jeffrey, M. C., & Steffan, J. G. (2011, June). Understanding Bloom filter intersection for lazy address-set disambiguation. In *Proceedings of the twenty-third annual ACM symposium on Parallelism in algorithms and architectures* (pp. 345-354). ACM.
- [12] Kimmatt, B., Thomo, A., & Venkatesh, S. (2014, July). Three-way joins on mapreduce: An experimental study. In *Information, Intelligence, Systems and Applications, IISA 2014, The 5th International Conference on* (pp. 227-232). IEEE.
- [13] Koumarelas, I. K., Naskos, A., & Gounaris, A. (2014, March). Binary Theta-Joins using MapReduce: Efficiency Analysis and Improvements. In *EDBT/ICDT Workshops* (pp. 6-9).
- [14] Kwon, H. Y., & Whang, K. Y. (2015). Scalable and efficient processing of top-k multiple-type integrated queries. *World Wide Web*, 1-25.
- [15] Lee, K. H., Lee, Y. J., Choi, H., Chung, Y. D., & Moon, B. (2012). Parallel data processing with MapReduce: a survey. *AcM SIGMOD Record*, 40(4), 11-20.
- [15] Lee, T., Im, D. H., Kim, H., & Kim, H. J. (2014). Application of filters to multiway joins in mapreduce. *Mathematical Problems in Engineering*, 2014.
- [16] Lee, T., Im, D. H., Kim, H., & Kim, H. J. (2014). Application of filters to multiway joins in mapreduce. *Mathematical Problems in Engineering*, 2014.
- [17] Li, F., Ooi, B. C., Özsu, M. T., & Wu, S. (2014). Distributed data management using MapReduce. *ACM Computing Surveys (CSUR)*, 46(3), 31.



- [18] Lu, W., Shen, Y., Chen, S., & Ooi, B. C. (2012). Efficient processing of k nearest neighbor joins using mapreduce. *Proceedings of the VLDB Endowment*, 5(10), 1016-1027.
- [19] Mahajan, S. M., & Jadhav, M. V. P. BLOOM JOIN FINE-TUNES DISTRIBUTED QUERY IN HADOOP ENVIRONMENT.
- [20] Myung, J., & Lee, S. G. (2012, February). Matrix chain multiplication via multi-way join algorithms in MapReduce. In *Proceedings of the 6th International Conference on Ubiquitous Information Management and Communication* (p. 53). ACM.
- [21] Okcan, A., & Riedewald, M. (2011, June). Processing theta-joins using MapReduce. In *Proceedings of the 2011 ACM SIGMOD International Conference on Management of data* (pp. 949-960). ACM.
- [22] Phan, T. C. (2014). Optimization for big joins and recursive query evaluation using intersection and difference filters in MapReduce (Doctoral dissertation, Université Blaise Pascal-Clermont-Ferrand II).
- [23] Phan, T. C., d'Orazio, L., & Rigaux, P. (2013, August). Toward intersection filter-based optimization for joins in mapreduce. In *Proceedings of the 2Nd International Workshop on Cloud Intelligence* (p. 2). ACM.
- [24] Pigul, A. (2012). Comparative Study Parallel Join Algorithms for MapReduce environment. *Труды Института системного программирования РАН*, 23.
- [25] Rong, C., Lu, W., Wang, X., Du, X., Chen, Y., & Tung, A. K. (2013). Efficient and scalable processing of string similarity join. *Knowledge and Data Engineering, IEEE Transactions on*, 25(10), 2217-2230.
- [26] Rytysareva, I., & Kalyanaraman, A. (2012). An efficient MapReduce algorithm for parallelizing large-scale graph clustering.
- [27] Vernica, R., Carey, M. J., & Li, C. (2010, June). Efficient parallel set-similarity joins using MapReduce. In *Proceedings of the 2010 ACM SIGMOD International Conference on Management of data* (pp. 495-506). ACM.
- [28] Yao, B., Li, F., & Kumar, P. (2010, March). K nearest neighbor queries and knn-joins in large relational databases (almost) for free. In *Data engineering (ICDE), 2010 IEEE 26th international conference on* (pp. 4-15). IEEE.



www.mecsjs.com

- [29] Yu, C., Zhang, R., Huang, Y., & Xiong, H. (2010). High-dimensional knn joins with incremental updates. *Geoinformatica*, 14(1), 55-82.
- [30] Zhang, C., Li, F., & Jestes, J. (2012, March). Efficient parallel kNN joins for large data in MapReduce. In *Proceedings of the 15th International Conference on Extending Database Technology* (pp. 38-49). ACM.
- [31] Zhang, C., Li, J., & Wu, L. (2013). Optimizing Theta-Joins in a MapReduce Environment. *International Journal of Database Theory and Application*, 6(4), 91-107.
- [32] Zhang, C., Wu, L., & Li, J. (2013). Efficient processing distributed joins with bloomfilter using mapreduce. *Int J Grid Distrib Comput*, 6(3), 43-58.
- [33] Zhang, S., Han, J., Liu, Z., Wang, K., & Xu, Z. (2009, August). Sjmr: Parallelizing spatial join with mapreduce on clusters. In *Cluster Computing and Workshops, 2009. CLUSTER'09. IEEE International Conference on* (pp. 1-8). IEEE.
- [34] Zhang, X., Chen, L., & Wang, M. (2012). Efficient multi-way theta-join processing using mapreduce. *Proceedings of the VLDB Endowment*, 5(11), 1184-1195.



www.mecsjs.com

# Challenges of Electronic Arbitration in Electronic Commerce transactions

Ahmed, Adnan

Arbitration & Business Law Department, University of Gloucestershire

Email: [Ahmad9012@gmail.com](mailto:Ahmad9012@gmail.com)

## Abstract

Electronic arbitration is inseparable from the advancement of electronic commerce besides cross-border trade transactions. Merchants in such transactions require an efficient, cost effective, and swift dispute resolution mechanism. Therefore, online arbitration is gaining prominence as it is deemed to be a private and a faster means of solving e-commerce disputes. Research indicates that an increasing number of both domestic and international regulations are adhering to the rules governing electronic arbitration. However, just like e-commerce, e-arbitration is also encountering certain challenges that may hinder its efficient operation. This paper will discuss exhaustively the challenges of electronic arbitration e-commerce.

**Keywords:** Electronic arbitration, transactions, challenges, disputes, e-commerce, resolution.



## **Introduction**

Electronic Commerce (e-commerce) is the buying and selling of goods and services over the internet. Berti &Ponti (2007) emphasize that advancements in Information Communication and Technology has led to the possibility of business transactions being carried out virtually. There are three main areas of electronic commerce. These include online retailing, virtual market, and online auctions. Millions of online business transactions are carried out every year. It is no longer a necessity for customers to visit the brick and mortar shops. Similarly, traders no longer have to sign solid copies of contracts. Lew & Kröll (2003) claim that electronic commerce has led to new challenges in various aspects. These challenges require fast and efficient response, and have necessitated the establishment of Electronic arbitration (E-arbitration). Arbitration is a process where an arbitrator, who is a neutral third party, delivers a decision that is final and binding on both parties in a dispute. Electronic arbitration is a method that is used to settle disputes by use of online platforms that provide arbitration services. This is e recent mechanism of dispute resolution that helps the disputing parties to solve their wrangles online.

## **Electronic Arbitration Institutions**

Electronic arbitration has been adopted mostly in Business-2-Business relations. Today, there are various services that provide e-arbitration in different forms. Kaufmann & Schultz (2004) asserts that the key aspects of carrying out electronic arbitration are to improve the speed of solving disputes, reduce costs, and the notion that disputing parties do not have to be available at the same place. Discussed below are some of the providers of electronic arbitration. Founded in 1985, Hong Kong International Arbitration Centre is an institution that provides dispute resolution services. It has established the Electronic Transaction Arbitration Rules that resolves consumer disputes. It has options that conduct hearings in person, by telephone, by video link, by mail, or by any other electronic communication. Since the rules of operation are from 2002, they are mostly dated in shape and language.



www.mecsj.com

Chinese International Economic and Trade Arbitration Commission is a renowned institution that has developed electronic arbitration rules. The institution aims at resolving electronic commerce disputes. Its rules were effective from 2015, and have a more visibly online-based focus. The institution also has an online dispute resolution center where cases can be submitted and are effectively resolved. Raghebi & Omid (2016) affirms that the claimant has the authority to decide on the way of communication, and can also opt to use regular mail for various communications. Research indicates that there is a clear preference for online communication in regard to submission of evidence, communication, as well as witness testimonies (Ramokanate, 2014). Also available are different procedures that have shorter time limits in regard to the value of the claim.

Headquartered in New York, the American Arbitration Association (AAA) is a non-profit organization that resolves disputes between suppliers and manufacturers. It does this at a fast and fairly inexpensive way. The whole process is electronic via the AAA web file platform. Razmi & Afras (2014) emphasize that the process has got two stages that include negotiation and arbitration. The whole process is supposed to be concluded in 66 days. In the arbitration stage, there is no evidentiary hearing other than through the AAA web-file platform. Moreover, only documents are used in the whole process. The arbitration award is to be given on the AAA web-file platform in 30 days.

Swiftcourt is a Swedish digital court that has been instrumental in providing binding electronic arbitration. The court offer services both on a standalone basis or when incorporated into a contract. Their main focus is to provide quick and cost efficient services, as it promises a resolution within six weeks from the date of application. Swiftcourt claims that their services are efficient and cheaper as compared to the offline courts.

### **Challenges facing Electronic Arbitration in electronic commerce transactions**

#### 1. Technology infrastructure

Electronic arbitration depends massively on the availability of technology. Technology plays an important part in order for electronic arbitration to succeed. Without easy access to internet connections and computers, the ability of parties to utilize electronic arbitration tools is extremely



limited. Electronic arbitration would inevitably appeal more to individuals who are essentially experienced in the online environment.

According to Tao (2004), there are a lot of individuals in various countries across the worlds who utilize eBay as a market place where they are able to buy and sell goods at a distance. e-Bay users are some of the people who have taken advantage of online arbitration. Their use of e-arbitration has been promoted by their ability to have computers as well as their ability to access internet. The merchants in eBay already have online arbitration options available to them for settlement of disputes.

Increase in the number of transactions has led to the increase of online transactions. In several countries, electronic arbitration has been promoted through government initiatives that accelerate individuals' access to the internet. Such initiatives include electronic government projects. However, it is important to note that certain electronic applications and platforms require to be developed to fit specific contexts in different countries.

## 2. Legal challenges.

There are various issues that have been raised at the platforms where parties conclude electronic arbitrations. One of the issues has been the integrity of the websites. All of the legal description and details of electronic arbitration are communicated online. They are conducted between two parties that are at a distance from one another. The parties do not have an opportunity to meet physically and mostly never get to see one another. This has always raised a lot of concerns. Torres & Arias (2009) affirms that one of the parties may not be serious enough in the arbitration process.

As a result, the admission to the online arbitration site may be risky on one of the parties to the deal, if the website is merely a fraudulent platform that aims to cheat. In order to curb such problems, recent legislations in various countries have resorted to the introduction of electronic authentication and ratification bodies. These bodies have been instrumental in tracking and revealing the identification of various websites. Moreover, they record, store, and carry out approval of electronic transactions that have been concluded between the parties. If a particular site is found to be unsafe, the parties are warned and notified of the credibility of the site. These bodies have been entrusted to issue certificates of services.

## 3. Enforceability of electronic arbitration.





The subject of enforceability in the context of electronic arbitration encompasses various elements. One of the main concerns of e-arbitration has always been the enforceability of resolutions that are obtained through online arbitration (Tao, 2004). In most instances, parties X and Y do enter into contractual agreements that bind them to comply with the resolutions that are reached through online arbitration. The ability of both parties in dispute to ensure that the other party would definitely comply with the resolution depends on the effectiveness of online arbitration.

According to Ruhl (2011), issues of submission to online arbitration as well as the enforceability of electronic arbitral awards have attracted the operation of the New York Convention. There has been certain doubts being raised as to whether agreements that are submitted to electronic arbitration, to be considered as being in writing. However, article 2(1) of the New York Convention requires that any agreement being submitted to arbitration must be in writing. This implies that an agreement in writing consist of an agreement contained in an exchange of telegrams or letters (Zacks, 2015). Modern scholars claim that online arbitration is definitely influenced by the operation of the New York convention, as its submissions fall within the scope of the phrase “exchange of telegrams”. The New York convention provides increased confidence to parties in regard to the enforceability of agreements to engage in electronic arbitration, as well as the enforceability of arbitral awards that are obtained from electronic arbitration.

However, there are certain inconveniences that are linked to initiating court action in order to implement the provisions of the New York convention. Such inconveniences would arise when a party to an online dispute deliberately refuses to comply with an arbitral ward. The creation of cyber-courts is one of the solutions that have been proposed to curb this dilemma. The cyber-courts would enable a party to obtain an online court order, so as to enforce a binding resolution that is obtained from online arbitration.

#### 4. Security in electronic arbitration.

Parties to a dispute often exchange a lot of information during the course of online arbitration process. Keeping the contents of proceedings as confidential as possible is usually a relatively simple matter, in case of offline disputes. However, a lot of protocols that are used to communicate online have not been inherently formed with security concerns in mind. According to Brunet (2006), the need to ensure that deliberations are kept confidential is vital in developing public confidence in



online arbitration as a legitimate way of resolving disputes that arise from online transactions. For such reasons, it is important that appropriate security measures are implemented during online arbitration processes.

Schlossberger (2015) asserts that there are various elements of security in the context of electronic arbitration. For instance, parties to an electronic dispute need to be certain that communications done during the process is confidential and can never be accessed by third parties without proper authorization. This can be attained by providing the participants involved with authentication credentials, as well as encrypting data by using public key cryptography. Second, the integrity of all transmitted data should be guaranteed. This is somehow achieved by the use of encryption techniques. Digital signatures by individuals involved in an online dispute can further help in verifying the integrity of all communications.

#### 5. Language barriers and culture.

Culture and language are deemed as being amongst the crucial challenges that hamper the interaction between individuals and various electronic arbitration websites. As noted by Ramokanate (2014), there is an urgent necessity to develop software's that will make a significant progress in the translation of all texts into languages that can be understood by everyone. There is also a need to consider cultural barriers, traditions, values, and customs, so as to avoid being a hindrance towards the use of commercial sites.

#### 6. Seat of arbitration

The seat of arbitration is crucial in regard to various aspects. The seat of arbitration is in charge of determining the applicable law in arbitrations. Moreover, based on the New York Convention, enforcement and recognition of the award may be rejected if the award had been set aside by an incompetent authority of the state in which the award was made. In electronic arbitration, the issue of multiple locations has been an obstacle in determining the place of arbitration. In this kind of arbitration, the disputing parties are not the only ones who may be located in different states. The arbitrators in attendance during a deliberation may also be from different countries. In comparative law there is often a tendency of not using the arbitrator's electronic presence or ones technical equipment to determine location. Therefore, the determination of the seat of arbitration must rest on legal criteria. According to Whitley & Kjaergaard (2014), this issue has led certain scholars to conclude that electronic arbitration has no identifiable seats.



However, current views about determination of the seat of online arbitration have significantly solved this concern. Today, the seat of arbitration is determined based on the new codes and not on geographical notion. The choice of the seat of arbitration is initially determined by the parties, either directly or by referring to the arbitration rules. If this fails, then the seat is determined by the arbitrators.

The idea that hearings and procedural acts are conducted elsewhere is irrelevant. Lynch (2003) argues that it is the task of the disputing parties in electronic arbitration deliberations to determine the seat of arbitration. The admissibility of determining the seat of arbitration by the parties as well as the arbitrators has resulted in the conclusion that lack of a physical place is irrelevant.

#### 7. Lack of physical/personal appearance of the parties.

A notable challenge of electronic arbitration has always been the lack of personal appearance of the disputing parties, witness, as well as experts in front of the arbitrators. According to Howell (2008), physical appearance has got crucial importance for the arbitration, in regard to its adjudicatory nature. For instance, sometimes arbitrators often need to understand the emotional situation that the witness is testifying in. This would help the arbitrator in making an assessment according to the credibility of the testimony.

#### 8. Determination of the place of arbitration.

Electronic arbitration is deemed as being placeless, implying that it lacks a fixed location. It is claimed that the concept of place, in arbitration, has become more abstract. The concern of determination of the place of arbitration often become important in instances when the arbitration agreement fail to cover this matter, and the disputing parties fail to agree on this issue after the dispute arises (Wissam, 2013). Since electronic arbitration is carried out in cyberspace, this can undermine the more traditional territorial approach the links that place of arbitration to the place where the involved parties have a domicile.

#### 9. Issues concerning the law to be applied.

Despite the disputes being handled out of court, the issue of applicable law usually emerges in electronic arbitration. In a global perspective, the law that should be applied has always been a great concern. There are still various differences relating to the level of consumer protection. Existing



consumer protective legislation often provide for the consumers' residence to determine the applicable law.

Brunet (2006) claims that it is not an easy task for businesses to know all applicable consumer protective legislation in the world. In order to avoid issues for businesses in their compliance with the consumer protective legislation across the world, it is necessary to create an equal level of applicable consumer protection. The UNCITRALs already has a vision of adopting a global framework on applicable legal principles.

### **Conclusion**

Electronic commerce is perceived to be the modern means that involves the trade of goods and services by use of electronic mechanisms. Since its inception, electronic commerce transactions have seen the emergence of various disputes. Such wrangles have led to the establishment of electronic arbitration, which has been useful in trying to solve the disputes that arise between the traders. Online arbitration is gaining prominence as it is deemed to be a private and a faster means of solving e-commerce disputes. However, electronic arbitration is also encountering certain challenges. Some of the notable challenges include technology infrastructure, legal challenges, enforceability issues, security concerns, language barriers and culture, lack of physical/personal appearance of the parties, determination of the place of arbitration, as well as issues concerning the law to be applied. If these challenges are resolved, electronic arbitration will be best suited mechanism to resolve e-commerce disputes.

### **References**

- Angel, W. (2015). Contemporary issues in international arbitration and mediation: The Fordham papers 2014.
- Berti, S. & Ponti, A. (2007). Comparative law of international arbitration. London: Sweet & Maxwell.
- Brunet, E. (2006). Arbitration law in America: A critical assessment. New York: Cambridge University Press. Commons, 93(3), 726-735.
- Howell, J. (2008). Electronic disclosure in international arbitration. Huntington, N.Y: Juris Net.



www.mecsj.com

- Kaufmann G. & Schultz, T. (2004). Online dispute resolution: Challenges for contemporary justice. The Hague: Kluwer Law Intern.
- Lew, M. & Kröll, S. (2003). Comparative international commercial arbitration. The Hague: Kluwer Law International.
- Lynch, K. (2003). The forces of economic globalization: Challenges to the regime of international commercial arbitration. The Hague [u.a.: Kluwer Law Internat.
- Raghebi, A. & Omidi, R. (2016). Legal principles on applying authority. Parliament on electronic contracts. Journal of Current Research in Science, (2), 178-177.
- Ramokanate, L. (2014). A legal framework for the treatment of input errors in electronic contracts (Doctoral dissertation). North-West University.
- Razmi, M. & Afras, M. (2014). Poition of Electronic Evidence in Iran Law and International Documents. Indian J. Sci. Res, 5(1), 425-431.
- Ruhl, G. (2011). Consumer Protection in Choice of Law. Cornell Int'l LJ, 44, 568-569.
- Schlossberger, O. (2015). Anti-Money Laundering. EU Press.
- Tao, J. (2004). Arbitration law and practice in China. The Hague: Kluwer Law Internat.
- Torres, T. & Arias, M. (2009). Encyclopedia of human resources information systems: Challenges in e-HRM. Hershey, Pa: IGI Global.
- Whitley, E. & Kjaergaard, A. (2014). Who do you think you are & quest; a review of the complex interplay between information systems, identification and identity.
- Wissam, A. (2013). The principles of private international law. Rafidain of Law Journal, 16(57).
- Zacks, E. (2015). Shame, Regret, and Contract Design. Marquette Law Scholarly European Journal of Information Systems, 23(1), 17-35.



www.mecsjs.com

# Investigating the Impact of Walking on humans Health

Farah Mohammed

Public Health, University of Jordan

*Email:* fofo181198@gmil.com

## Abstract

Recent studies have demonstrated that brisk walking help to reduce anxiety and tension and assist in weight loss. Regular walking helps improve cholesterol profile, help control hypertension, and slow the process of osteoporosis.

The purpose of this paper is to summarise the benefits of walking on physical and mental health to encourage people to participate in walking as a regular and sustainable exercise, and to determine of the limitation of regular walking. The researcher used the descriptive approach for achieve the research objectives.

Research finding indicated that walking can relieve the human body fat, improve their heart health and decrease the possible risk of heart attack and stroke. Additionally walking improve their ability in lowering high blood pressure, reduce risk of diabetes, and Decrease stress and prevent depression.

Keywords: Walking, humans health, Physical and mental health.



## 1. Introduction

Health is a dynamic process, and it is always changing. All humans may have times of good health, times of sickness, and maybe even times of serious illness. This study came to investigate the impact of Walking on reducing the sickness time through regular walking in daily manner.

When most people are asked what it means to be healthy, they normally respond with the four components of fitness mentioned earlier (cardio respiratory ability, body composition, flexibility, and muscular ability). (canfitpro, 2016)

Those of who participate in regular physical activity to improve the current and future level of their health. Those strive toward an optimal state of well-being. As their lifestyle improves, their health also improves and has less sickness and disease. (canfitpro, 2016)

Walking is one of the less expensive and most broadly attainable exercises of physical activity. It is rarely linked with physical injury and can easily be adopted by people of different ages, including those who have never participated in physical activity.

Walking is currently considered as the most popular action with regard to common realised physical activity all over the world. According to some published studies in the United Kingdom and United States that directly demonstrated that the prevalence of walking is two to three times higher than those of the next most frequently reported activities.

Walking is an automatic, intrinsic human function, harmonious, dynamic, aerobic activity of large skeletal muscles that confers the multifarious benefits with minimal adverse effects. brisk Walking and regularly in sufficient quantity increase 70% of maximal heart rate, develops and sustains physical fitness and the cardiovascular capacity and stamina for bodily work and movement in everyday life. (Notthoff, 2014)

Walking is considered as a convenient activity and may be accommodated in occupational and domestic routines. It is self-regulated in intensity, duration and frequency, is inherently safe, duration



and frequency, it is a year-round, readily repeatable, self-reinforcing, habit-forming activity and the main option for increasing physical activity in populations. (Kohl & Cook, 2013)

Studies have shown that walking has higher levels of involvement than other exercises of physical activity, possibly because it is convenient and overcomes many of the commonly perceived barriers to physical activity: lack of time, lack of fitness or lack of skill. (Hovell, 2017). This study came to investigate the relation between walking activity and public health especially physical and mental health.

### **1.1 Research problem**

Walking rates have declined steadily all over the world during the last decades, because the evolution of transportation and manufacturing of elevators and electrical stairs. Therefore, this paper has been prepared to investigate solutions and methods to inducement the people on regularly walking.

### **1.2 Research Questions**

This paper came to answer the following questions:

- 1- What is the impact of walking on humans' health?
- 2- Why walking is so accessible?
- 3- What is the limitation of regular walking?

### **1.3 Objective**

The main objectives that this paper aims to answer the following:

- 1- The researcher seeks to determine the benefits of regularly walking.
- 2- The researcher seeks to determine the impact of regularly walking on humans' health.
- 3- To encourage people to participate in walking as a regular and sustainable exercise.





## **Methodology**

The searcher searched, and compiled of information concerned with the effect of walking on the health. So a descriptive approach was used to achieve the research objectives through reviewing several literatures and past studies on the study topic.

## **2. Health**

### **2.1 The concept of health**

The world health organization defined health is “A state of complete physical, mental and social well-being and not merely the absence of disease or infirmity”. (Hofgastein, 2011)

This means that health is a healthful lifestyle is represented in mental and physical health to support an individual's function in society.

### **2.2 Type of health**

We have five types of health mental, physical, social, emotional, and spiritual health:

1) - *physical health*

2) *Physical health involves pursuing a healthful lifestyle to Maintaining physical fitness, decrease the risk of noncommunicable disease. (Nordqvist, 2017)*

3) - *Mental health*



www.mecsj.com

Mental health is the ability of person to grow and learn intellectually. and refers to a person's emotional, individual's perception, psychological wellbeing, and social. (Nordqvist, 2017)

#### - **Social health**

The ability of person to interact well with people and the environment around him and to have satisfying personal relationships.

#### - **Emotional health**

The ability of person to control in him emotions so that him feel comfortable expressing them and can express them appropriately.

#### - **Spiritual health**

It varies from person to person but has the concept of faith at its core. A belief in some unifying force. (canfitpro, 2016)

### **3. Walking**

#### **3.1 The concept of walking**

Walking is an important locomotor skill and used in everyday activities like walking to market or job and used in sports, used in play and dance activities. and the Walking is the transfer of weight from one foot to the other while moving forwards or backwards. the arms swing freely in opposite directions, and One foot is always in contact with the ground. (zealand, 2012)

This means that walking is the continuous move process using in everyday activities like walking to market or school.

#### **3.2 In the following the reasons why walking is so accessible: (Fenton, 2011)**



- walking is free this mean you do not to need expensive gym membership or special shoes to take part
- you can wear your daily clothes that you like so you do not feel uncomfortable.
- walking is safety, so you are unlikely to get injured.
- it's fun to go with your friends for a walk.
- the walking is a healthy way to go where you need be.
- you can be walking almost anywhere at any time.
- Walking is a great 'gateway' to the world of exercise because it often inspires people to try out other healthy activities.

### **3.3 The limitation of regular walking:**

- people may not realize how long it takes to walk short distances.
- lack of safe and attractive places to walk.
- adverse weather (either too hot or too cold).
- In addition to suffer the people with disabilities, elderly people or parents with kids. (Torner & Neogi, 2014)

## **4. Impact of walking on health:**

### **4.1 There are many health benefits of walking. And exercise walking is so very easy:**

1. **Reach and preserve a healthy weight:** brisk Walking reduces your body fat, increases lean muscle tissue, burns calories and improves metabolism. walking can help you achieve long-lasting weight control. Combined with a healthy lifestyle and nutritious eating.



4) 2. *Regulates Blood Pressure and cholesterol: regular walking can also help lowering high blood pressure, decreasing low density lipoprotein (LDL) cholesterol and increasing high density lipoprotein (HDL) instead of taking risky medications with unpleasant side effects. (Greene, 2014)*

5) 3. *manage and reduce risk of diabetes: daily Walking progresses your body's natural ability to process sugar (glucose tolerance) and to maintain a healthy weight. This help you prevent or even reverse the effect of type 2 diabetes. (Biswas, 2017)*

6) 4. *Lower your risk of heart attack and stroke: Walking helps to improve your heart health and decrease your risk of heart attack and stroke, just 30 minutes of walking daily greatly improves your circulation and helps keep your heart and blood vessels healthy - Based on large studies of both women and men. (Biswas, 2017)*

7) 5. *Decrease stress and prevent depression: Regular exercise walking is a great way to improve your mood, breathing, to reduce stress, anxiety and depression by improving circulation and stimulates the nervous system receptors and decreases the production of the stress hormones. (Greene, 2014)*

8) 6. *Stay strong, active and healthy: As you age, exercise walking can improve your stamina and keep you fit. Walking exercise strengthens bones, muscles and joints, helps prevent falls and hip fractures, improves your immune system and extends your life expectancy. (Biswas, 2017)*

9) 7. *Strengthens Bones: Studies show that regular walking can relieve arthritis and back pain and reducing the risk of osteoporosis, fracture. (Greene, 2014)*

10) 8. *Improves Digestion: Walking after meals helps you reduce weight and supports your digestive system, and reduce the risk of colon cancer, diarrhea, constipation, bloating. (Biswas, 2017)*

#### **4.2 The impact of walking on mental health**

- relieve physical symptoms of anxiety associated with minor stress.
- increase self-reported energy levels when older adults set their own pace.
- advance sleep quality.



[www.mecsj.com](http://www.mecsj.com)

- raise affective response, resulting in increased psychological well-being for individuals with type 2 diabetes.
  - be related with better cognitive performance at school.
  - progress the cognitive functioning of older adults (compared to stretching and toning).
  - progress cognitive performance and reduce cognitive decline among older people.
  - growth the size of the hippocampus and prefrontal cortex, potentially beneficial for memory.
- (Hancock, 2012)



## 5. Conclusion

Walking is among the most cost-effective and accessible means of exercise. walking help to maintain physical and mental health in the all age by preventing a variety of health problems. In this paper we have defined the health is represented in mental and physical health to support an individual's function in society ; the concept of walking is the continuous move process using in everyday activities ; the identification of the types of health (physical and mental) , obstacles to the practice of walking such as adverse weather, lack of safe and attractive places to walk, and people may not realize how long it takes to walk short distances; and the impact of walking on physical and mental health such as Reach and preserve a healthy weight, Regulates Blood Pressure and cholesterol, manage and reduce risk of diabetes, Decrease stress and prevent depression, Strengthens Bones, Stay strong, active and healthy, Lower your risk of heart attack and stroke, and Improves Digestion.

Research finding currently indicate that walking can relieve symptoms of depression and anxiety, improve sleep quality, reduces your body fat, lowering high blood pressure, reduce risk of diabetes, to improve your heart health and decrease your risk of heart attack and stroke resulting in improvements in individual quality of life and reductions in the medical costs associated with treating these disorders, and improve cognitive performance (performance in mental processes such as thinking, understanding and remembering).

Although there is a body of studies on the benefits of walking for mental and physical health, few of people have a regular walking. therefore, we need to more studies into the relationship between the walking and its effect on mental and physical health.



## References

- Biswas, C. (2017, 11 17). Top 20 Health Benefits of walking Daily. stylecraze. Retrieved from <http://www.stylecraze.com/articles/best-benefits-of-walking/#gref>
- canfitpro. (2016). The importance of health, fitness, and wellness. In Foundations of Professional Personal Training 2nd Edition With Web Resource. Human Kinetics.
- Fenton, K. (2011). Walking works. It makes people happy, keeps everyone healthy and can even save your life. walking works. Retrieved from [https://www.walkingforhealth.org.uk/sites/default/files/Walking%20works\\_summary\\_AW\\_Web.pdf](https://www.walkingforhealth.org.uk/sites/default/files/Walking%20works_summary_AW_Web.pdf)
- Greene, M. (2014). The Health Benefits of Walking. Retrieved from washburn: <http://www.washburn.edu/campus-life/recreation-wellness/employee-wellness/documents/EW-PH-Benefits-of-Walking.pdf>
- Hancock, C. (2012). The benefits of regular walking for health, . united kingdom: C3 Collaborating for Health.
- Hofgastein, B. (2011). Designing the road to better health and well-being in Europe” at the 14th European Health Forum Gastein. Austria: WHO.
- Hovell, S. &. (2017). Overcoming Barriers to Physical Activity. Center for disease control and prevention .
- Kohl, H. W., & Cook, &. H. (2013). Educating the Student Body: Taking Physical Activity and Physical Education to School. US: Committee on Physical Activity and Physical Education in the School Environment.
- Nordqvist, C. (2017). Health: What does good health really mean? united states: Mwdical News Today.
- Notthoff, N. (2014). Positive messaging promotes walking in older adults. 29(2), 329-341.
- Torner, J., & Neogi, &. T. (2014). Daily walking and the risk of incident functional limitation in knee osteoarthritis: an observational study. bosten: bosten university.



www.mecsjs.com

## Operation strategy at Galanz

Li, Yang Won

Department of Management, City University of Hong Kong, Kowloon, Hong Kong

**Email:** Won.li@adelaide.edu.au

### III. ABSTRACT

In this research, we investigate the growth of Galanz as a simple industry to become one of the mostly common manufacturers in the world. In this research, we aim to introduce students to the concept of the order of the winner to classifiers, operations priorities / goals. Another goal is to show how the operation priorities should reflect the requirements of the customers and influence on the company wants to compete. Further, we demonstrated how a company can gain a competitive advantage through lower costs and a strategy on how to support the business strategy with the strategy and operations capabilities. The research also showed the students how a company can build several opportunities over time, and how the strategy and operational capabilities of the company may change over time. We also provided students with the opportunity to analyze the trade-offs in the strategic, operational and marketing decisions as the business expanded from domestic to international market and from OEM to ODM and OBM. The challenges of students to develop coherent plans of action that relate to future growth objectives are also involved. Finally, we help student to understand the great opportunities and challenges, and the management of the activities of the supply chain in China.

**Keywords:** Galanz, operation priorities, OEM, ODM, OBM.





#### IV. INTRODUCTION

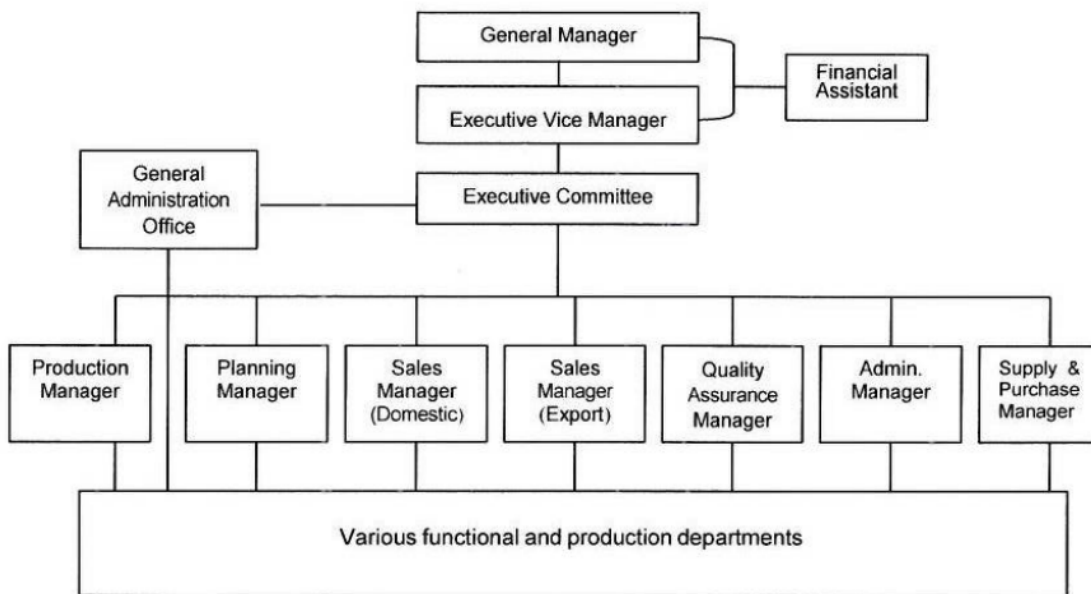
Galanz has rapidly grown that transformed from simple manufacturer to become one of the world class manufacturers of microwave ovens. It is a member of Shunde Galanz Enterprises Group Co. Ltd. Galanz is the producer of fifty percent of microwave ovens in the world in 2003.

Galanz started its business with a rich competitive strategy developed based on cost control, economies of scale, and full utilization of operating resources and capacities. The case of Galanz introduces new concepts of winning and shows how to reflect the needs of customers based on priorities of operations that influence the competitive advantage of the company[1].

Galanz has achieved economy of scale through combining OEM, ODM, and OBM. It also has achieved 60%-70% of the local production in 2002, and 50% of the world-wide market in 2007. The increasing production lines refer to many reasons: first, free production line transfer; second, the obtained right of using additional capacities for production; third, the encouraged component suppliers of Galanz' manufacturing facilities.

In this case, the operational and competitive aspects of Galanz require us to trigger different kinds of growing strategies. Many reasons stand beside the withdraw of many industry players from the market such as the adoption of penetration pricing leveraging economies of scale, the expansion of production capability to exceed market demand, and the aggressive pricing strategy.

### Company structure of Galanz group in 2002



### V. OBJECTIVES

1. Introduce students to the concept of the order of the winner to classifiers, operations priorities / goals.
2. Show how the operation priorities should reflect the requirements of the customers and influence on the company wants to compete.
3. Demonstrate how a company can gain a competitive advantage through lower costs and a strategy on how to support the business strategy with the strategy and operations capabilities.
4. Show the students how a company can build several opportunities over time, and how the strategy and operational capabilities of the company may change over time.



www.mecsj.com

5. Provide students with the opportunity to analyze the trade-offs in the strategic, operational and marketing decisions as the business expanded from domestic to international market and from OEM to ODM and OBM.
6. Challenge students to develop coherent plans of action that relate to future growth objectives.
7. Help you understand the great opportunities and challenges, and the management of the activities of the supply chain in China.

## **VI. WHY MICROWAVE OVENS?**

In 1980, microwave ovens market was in its infancy in china since the competition and demand size was small from the perspective of the competitive environment. The microwave ovens had become common in modern cities in China in 1991, introducing an opportunity for an excellent business. Beside the development of manufacturing microwave ovens, the price of the microwave ovens was high and unaffordable to the most Chinese[2].

Galanz has transformed its manufacturing from world factory to world brand after it entered as a dominant player in the global home appliance market of microwave ovens. In Galanz management, a highly-centralized decision making and execution system has been adopted. It included three layers: senior executives, general management, and operational staff. From communication and information flow perspective, the strategies were hard to implement since they are unclear and slow.

## **VII. THE START-UP**

Many Chinese businesspersons utilized the opportunity in manufacturing microwave ovens. Whoever, starting a microwave oven production business was not completely emphasized without any challenge, despite the technical ease of producing microwave ovens. At that time, the only person that had the determination of working to outperform business challenges was Liang Senior that first import equipment and technology from overseas during the initial stages of Galanz. In early



stages, Galanz has searched of engineering professionals knowledgeable of microwave oven technologies to help to set up factor[3].

In rapid way, Galanz used the Japanese company as outsourcing of the production technology to produce oven for domestic market under Galanz brand name. This was to solve the problem of outsourcing part of magnetron production to other companies. Although it has used outsourcing for production, it produced ovens with low cost based on enhanced ability of competing against leading, successful players such as LG, Panasonic, and Toshiba

## VIII. THE EARLY SUCCESS

The role of technology has played a critical role in the success of Galanz. This is caused by the enhanced technologies that make the management of large set of information to become much easier. In special case, the manufacturing of Magnetron used as equipment in producing microwave ovens was the main proficiency and expertise of Galanz Company as the time passed[4].

The microwave oven business had many winners' qualifiers during the earlier stages of development. The characteristics or traits of business refer to the orders winners that perform a critical activity in competitive advantage of the Galanz Company. Competitive advantage helps the company to differentiate itself from the others. However, order qualifiers act as characteristics of compulsorily exhibiting the competitive advantage of the company to differentiate itself in the market.

The case of Galanz indicated that it was owned by the foreign trade department of Guangdong province and it has successfully gotten its desired exports of products larger than 23 million in 1992. At that time, it has converted its business from manufacturing feather products into manufacturing microwave ovens[5].



## IX. OEM PRODUCTION LINE

The first method that Galanz has implemented to achieve low cost of production and to intensify its production capability was through a free production line transfer. After a few years, Galanz continued to quickly enlarge its production scale with quality of products as a result of providing staff training, production site enhancements, and parts customization. Furthermore, the increased production capacity of Galanz has been improved by utilizing the labour resources and production facilities. The operations were divided into three shifts per day (24 hours a day), seven days a week, and 365 days a year. In sum, the company through reduced cost achieved an increasing in its production scale[4].

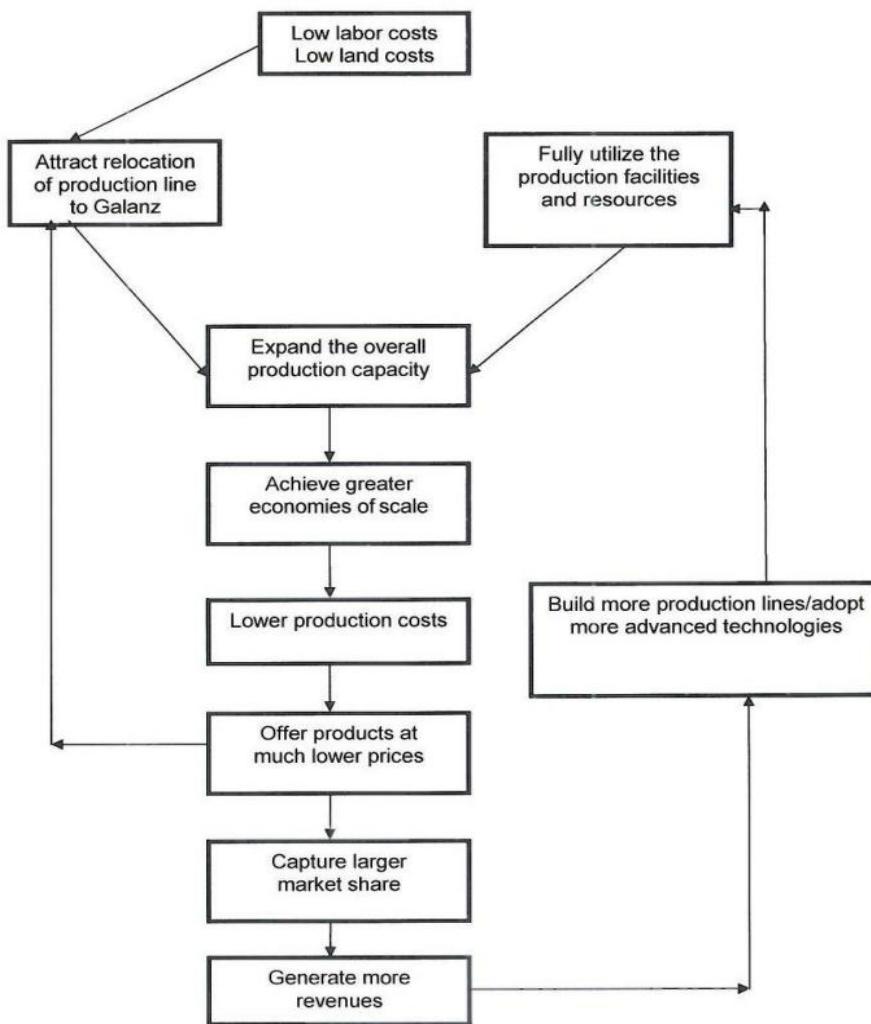
One of the main causes of fierce competition of branded products is globalization. There is a conflict in the interest in OEM and OBM. In OEM, service business and setup sales depend on the channel partner that lead to have a great competition to OEM customers.

## X. PRICE WAR

Price could fully dominate the internal microwave oven market, which stimulates Galanz to sustain the cost leadership. Consequently, Galanz decided to reduce its prices to obtain more market share than other competitor and to lead a position in global markets. This price-cutting method was relied on by Galanz to set the average unit cost of production required for growth. The company pushed its sales team to work harder since it enjoyed tremendous economies of scale for low cost production and additional inventory[6].

The main objective of price war was to damage the confidence of other competitors in market with little investment value. Some backwards that were utilized for improving the efficiency of the business targets for product pricing and profit margins.

### **Galanz Price cutting cycle**



## XI. TRANSFORMATION FROM OEM TO ODM

One of the main success factors of Galanz is its entrance to the global market using OEM business that enabled the company to utilize resources such as manufacturing equipment. Thus, it has exceeded other Chinese companies that export microwave ovens with no brand recognition to users. In short, the transformation of Galanz from OEM to ODM after magnetrons' production has increased the demand of branded products. These products are with good quality and low cost.

In many cases, the transformation from OEM to ODM has emerged because of the lack of magnetron and retrenchment of the suppliers. Thus, Galanz has designed and developed its own magnetron started mass production. However, foreign market was nor extremely familiar with the brand Galanz[7].



## **XII. DEVELOPMENT OF PRODUCTION OF OWN MAGNETRON**

Galanz has converted its business directions from “made in china” to “created in china”. The import of advanced technologies from overseas partners inspired Galanz to establish technical capabilities for equipment to design and develop its own magnetron [8].

The most important objective to Galanz was cost, but it has a plentiful labour supply. Further, it can deliver quality products better than the rest of leading companies in the market. One of the big challenges faced by Galanz Company was the refuse of supplying of the most important components magnetron used in production, stimulating it to design and develop its own innovation. Subsequently, Galanz has initiated major invest in magnetron in 1997, as well as it was able to design and produce magnetron by 2000. After that, the demand on magnetron was 25 million in 2003, while its capacity was only 16 million[9].

## **XIII. ENHANCEMENT OF R&D CAPABILITY AND PRODUCTION INNOVATION**

Galanz started to raise up the investment in R&D activities after touching many benefits from gained from the design and production of magnetron based on R&D activities. It aimed to enhance the internal R&D structure and to facilitate the design and development

of new product. Further, the investment in R&D has given more than 3% of annual revenue. The new technological invention of light wave ovens has improved the technical capability of Galanz through investment in R&D and import of technologies from overseas partners. This led the company to provide more ODM service to its large OEM clients and to receive more orders for Galanz brand products[10].

In addition, Galanz imported new technologies and investments in R&D to achieve differentiation and reduced cost. Moreover, the enhanced R&D capacity and production that have been setup with focus on new features and technologies have more than 3% of annual revenue.



#### XIV. THE DRIVER OF SELF SUFFICIENCY

Galanz has begun to integrate its supply chain through adding more manufacturers to the original components to improve quality and reduce costs at the same time. This has resulted Galanz to produce 90% of the total components of microwave ovens[11].

Low price strategy that has been previously implemented by Galanz to deal with win orders was based on selling products in lower costs than the competitors. Additionally, the company has achieved economies of scale and increase in demand due to the successful achievement of cost reduction and with the aid of low price strategy. Consequently, the company was able to sell products at lower prices than competitors and it has always been up for the price wars to achieve economies of scale. In short term, Galanz company has successfully made full utilization in terms of production capability among foreign manufacturers[12,10].

The flexibility, cost, innovation, and delivery service objectives are the main important operations goals of Galanz Company. Initially, the company did not face any competition in the industry during the initial years. After a few years, the company was engaged in outsourcing the equipment used in manufacturing microwave ovens, and it continued to do this process to sell products to many countries around the world. Subsequently, it also was engaged in both producing and selling the equipment used to produce microwave ovens and to sell them to the manufacturer at worldwide level[13].

The company gained the competitive advantage by cheap access to land and labour. Further, the company has utilized raw materials with added value for manufacturing products to be sold in high prices. As a result, the company has followed low price strategy as a main objective to gain market share and make money. The most important operations objectives to the company during its early stages of development were cost, quality, and timely delivery. Recently, the whole country has transformed to modernize strategies. Therefore, the priorities of these important operational objectives have been evolved and changed over time by focusing on both low cost and quality of innovation.





## **XV. OEM/ODM VS. OBM IN THE OVERSEAS MARKET**

The quality of products of Galanz earned approval between the global industry players of microwave ovens. However, Galanz stayed on top among microwave oven players in the domestic market. The primary business of Galanz in 2003 was OEM microwave ovens but without any brand known to the users. Therefore, the experience of Galanz in changing the strategic relationship between international companies with the causes of globalization in terms of competition [14].

The percentage of Galanz OBM and OEM microwave ovens has raised noticeably from 1997 to 2003. On the other hand, Galanz was careful when it puts its effort to expand OBM sales without harming the interests in OEM customers to maintain OEM business [15].

The company has faced many challenges especially when it ordered the priorities to decide which production order among OEM, OBM, and ODM. The company overlooked the significance of compliance with principles of practice to manage the efficiency of production with respect to the training of employees. The management of Galanz has recognized the importance of technical and management skills, work efficiency and attitudes, as well as the low cost labour [16].

Through OBM business, users can explore the brand name of Galanz, at the same time price reduction is affected by cost leadership. The combination of small and inefficient industrial players in price war had given a great chance to discourage new entrants to the market. Therefore, an excessive entry is encouraged by low risk, low cost, and united acquisition to gain high profit margin in the industry.

## **XVI. PRODUCTION PLANNING**

Galanz has adopted large scale production with high efficiency and low cost compared to past products that were low quality in the domestic OBM market. At that time, the company has put little effort on strengthening the capacity of forecasting and planning for production scale. The company wanted to accurately forecast the demand to produce the requested products with required quantity and quality. However, the company might suffer from the shortage of products wanted by customers or unsold inventory [17].

In the mode of OEM, the customer can place his/her order and pay to the company only on delivery. However, this mode resulted in poor planning and forecasting. Whereas the OBM business allowed Galanz to get more accuracy in forecast that plays a core role in the success of the company. Conversely, the customers can request different product configurations at any time due to the increased product variety and the shortened product lifecycle. OBM business in the overseas market enabled Galanz to adapt mass production systems in meeting high variety products and the demand of customers, as well as the low volumes [18].

High quality, fast delivery, reliable delivery, and flexibility are the main priorities to achieve competitive advantage. These priorities formulate the success factors to lead company to greater success and execute a combined scheme in an international expansion. Hence, many advantages have been occupied with forming a joint venture in Galanz. Moreover, the combination of OEM, OBM, and ODM businesses draws the future of competitive strategy for Galanz. The competitive strategy involves an effective sharing of value chain activities, effective resource allocation for competitive advantage, and vertical relationship for product management adoption.

## **XVII. THE FUTURE OF THE COMPANY**

Galanz has faced many challenges concerned in low cost competitive strategy and due to many critical changes despite its spectacular growth. The success and the strategic company direction is influenced by many issues such as the competitive strategy [19].

Galanz company dumps its products based on an efficient production planning to produce proper number of products compatible with the demands. Therefore, it follows production planning capabilities and sales forecasting.

The abundant supply of cheap labour and land has permitted an increased market share with cost leadership strategy. This also has allowed Galanz to fully utilize resources toward product-oriented process and increased in production sale, as well as decreases

production cost. Galanz has focused on enhancing the distribution of product and improving the design and development of products. It also has strategic partnerships with different companies in international market mainly in OEM business. The employment of OEM method was on of the success factors of Galanz Company.

**STAGES OF GALANZ PRICE WAR IN CHINA (1996-2002)**

| Month/Year | Product   | Price Reduction | Result  |
|------------|---|-----------------|---|
| Aug. 1996  | Full range                                      | 40% discount    | <ul style="list-style-type: none"> <li>• Sales rose to 650,000</li> <li>• Domestic market share more than 35%</li> </ul>                            |
| Oct. 1997  | Full range                                      | 29-40% discount | <ul style="list-style-type: none"> <li>• Sales rose to 1,980,000</li> <li>• Domestic market share more than 47.6%</li> </ul>                        |
| May 1998   | Full range                                      | 30% discount    | <ul style="list-style-type: none"> <li>• Sales rose to 4,500,000</li> <li>• Domestic market share more than 60%</li> </ul>                          |
| June 2000  | Fine Golden Flower Series (mid-ranged products) | 40% discount    | <ul style="list-style-type: none"> <li>• Sales rose to 1,000,000 in the two product lines</li> <li>• Domestic market share more than 76%</li> </ul> |
| Oct. 2000  | Black Edition Series (high-end products)        | 40% discount    | <ul style="list-style-type: none"> <li>• International market share more than 30%</li> </ul>  |
| Apr. 2001  | Products below RMB 300                          | 30% discount    | <ul style="list-style-type: none"> <li>• Products were popular in the low season</li> </ul>   |
| Jan. 2002  | Digital Temperature Control Series              | 30% discount    | <ul style="list-style-type: none"> <li>• Galanz dominated the microwave oven market</li> </ul>  |

Galanz was formed by Liang Zhaoxian’s father, Liang Qingde in 1978. It’s headquarter is located in Shunde in the province of Guangdong. The company has initially created for manufacturing down feather products as a primary business. The name of Galanz company at that time was Guizhou down product factory [20].

Galanz hold order qualifiers in terms of the delivery of quality of service of products delivery. Conversely, order winners of the company have been shown through the low coat and low price of products. During the early stage of development, the classification of order qualifiers and winner qualifiers were on the basis of development.



## XVIII. CONCLUSION

In this research, we can conclude that Galanz has transformed its manufacturing from world factory to world brand after it entered as a dominant player in the global home appliance market of microwave ovens. In rapid way, Galanz used the Japanese company as outsourcing of the production technology to produce oven for domestic market under Galanz brand name. At that time, it has converted its business from manufacturing feather products into manufacturing microwave ovens. The most important objective to Galanz was cost, but it has a plentiful labour supply. In short term, Galanz company has successfully made full utilization in terms of production capability among foreign manufacturers.

## XIX. BIBLIOGRAPHY

- [1] Ansoff, H. I. (1965). *Corporate strategy: An analytic approach to business policy for growth and expansion*. McGraw-Hill Companies.
- [2] Au, K., Peng, M. W., & Wang, D. (2000). Interlocking directorates, firm strategies, and performance in Hong Kong: Towards a research agenda. *Asia Pacific Journal of Management*, 17(1), 29-47.
- [3] Chen, J., Zhu, Z., & Anquan, W. (2005). A system model for corporate entrepreneurship. *International Journal of Manpower*, 26(6), 529-543.
- [4] Child, J., & Rodrigues, S. B. (2005). The internationalization of Chinese firms: A case for theoretical extension?[1]. *Management and organization review*, 1(3), 381-410.
- [5] Datta, D. K., Rajagopalan, N., & Rasheed, A. (1991). Diversification and performance: Critical review and future directions. *Journal of Management Studies*, 28(5), 529-558.
- [6] Deng, P. (2003). Foreign investment by multinationals from emerging countries: The case of China. *Journal of Leadership & Organizational Studies*, 10(2), 113-124.
- [7] Deng, P. (2004). Outward investment by Chinese MNCs: Motivations and implications. *Business horizons*, 47(3), 8-16.
- [8] Dunning, J. H. (2006). Comment on Dragon multinationals: New players in 21 st century globalization. *Asia Pacific Journal of Management*, 23(2), 139-141.



www.mecsjs.com

- [9] Gereffi, G., Humphrey, J., & Sturgeon, T. (2005). The governance of global value chains. *Review of international political economy*, 12(1), 78-104.
- [10] Gilboy, G. J. (2004). The myth behind China's miracle. *Foreign affairs*, 33-48.
- [11] Hoskisson, R. E., & Hitt, M. A. (1990). Antecedents and performance outcomes of diversification: A review and critique of theoretical perspectives. *Journal of management*, 16(2), 461-509.
- [12] Kotabe, M., & Murray, J. Y. (2004). Global sourcing strategy and sustainable competitive advantage. *Industrial Marketing Management*, 33(1), 7-14.
- [13] Mathews, J. A. (2006). Dragon multinationals: New players in 21st century globalization. *Asia Pacific journal of management*, 23(1), 5-27.
- [14] Nolan, P. (2002). China and the global business revolution. *Cambridge Journal of Economics*, 26(1), 119-137.
- [15] Peng, M. W. (2001). The resource-based view and international business. *Journal of management*, 27(6), 803-829.
- [16] Plambeck, E. L., & Taylor, T. A. (2005). Sell the plant? The impact of contract manufacturing on innovation, capacity, and profitability. *Management science*, 51(1), 133-150.
- [17] Justin Tan, J., & Litsschert, R. J. (1994). Environment- strategy relationship and its performance implications: An empirical study of the Chinese electronics industry. *Strategic management journal*, 15(1), 1-20.
- [18] Williamson, P., & Zeng, M. (2004). Strategies for competing in a changed China. *MIT Sloan Management Review*, 45(4), 85.
- [19] Wright, M., Filatotchev, I., Hoskisson, R. E., & Peng, M. W. (2005). Strategy research in emerging economies: Challenging the conventional wisdom. *Journal of management studies*, 42(1), 1-33.
- [20] Yin, R. K. (2013). *Case study research: Design and methods*. Sage publications.



# **The Impact of Poor Planning and Management on the Duration of Construction Projects: A Review**

Farah Jawdat Ibrahim Khalid

Civil Engineering, Hashemite University

Email:faroo7khalid@gmail.com

## **Abstract**

The construction business is a noteworthy player in the economic sectors, creating both, wealth and employment. However numerous ventures encounter broad delays and time overruns leading to as surpass in starting time and cost estimations. Despite the fact that project management has demonstrated its accomplishment in construction as a procurement strategy, there are issues related with the way in which the projects have been planned or managed that have prompted venture postponements, cost overruns and delays with low client fulfillment. The aim of this paper is to review the impact of poor planning and management on the duration of construction projects. This research is based on a descriptive methodology, in which the researcher tends to review previous studies and literature that will help in identifying the relationship between the poor planning and management of project and the occurrence of delays in the construction projects. Literatures analysis discloses that poor project planning and management is cited by several researchers as a delay factor in the construction projects. The researcher concludes that poor planning and management of the construction projects may lead to several negative effects on the duration and completion of projects. Construction delays and duration issues are frequently responsible of transforming productive ventures into loosing projects. These delays can be reduced or prevented by an increased pre- project planning and successful project management as they are one of the most critical success factors of the construction project accomplishment.

**Keywords:** Poor planning, project management, construction.

## **1. Introduction:**

The construction sector consists of a range of activities related to buildings and engineering constructions of all types, as well as maintenance, planning and management. This sector is closely related to various other economic sectors, making it an important and reliable indicator of the

movement and trends of the national economy (Sweis, Sweis, Hammad, & Shboul, 2008). However, there are many risks that are encountering construction projects and threatening its operations. These risks are caused mainly by poor planning and management that affect the progress of the project and result in delayed delivery or increased cost, and sometimes poor construction quality (Gajewska, & Ropel, 2011).

A construction project is normally recognized to be successful when it is finished within its budget, on its planning time and according to its standards and specifications Olawale and Sun (2012); (Frimpong, et al. (2003); Majid (2006). In the development and construction business, contractual workers and engineers have a tendency to limit the duration of their ventures and amplify their profit to increase their construction market share and to provide this sector with successful progression and development. To accomplish this tendency, it is vital for project managers to deliberately recognize the schedules and plans of a project and measure their effects before the implementation stage (Gunduz, Nielsen, & Ozdemir, 2013).

Construction projects are of a special nature and characterized by their long period of time, which may lead to variations in conditions and possibilities. The long duration of the projects and their multi-stage processes starting from the preparing and beginning of the project to implementation and final delivery of it; lead to several conditions, possibilities, uncertainties and the possibility of falling into the risk of extending the duration of the project or incurred financial or other losses, which adversely affect the operation of the project and the economics of construction (Gajewska, & Ropel, 2011). In spite of the similarity of construction projects in terms of nature and purpose, the variation in their operating place or any other variable can change the preparation process of project and the way to manage it. So, this requires a good management and planning of the project in a way that avoids falling into constructing and planning risks and guarantees success and achievement of its objectives (Falqi, 2004).

Those projects may vary in duration, size, environment, complexity, objectives, conditions, organization structures, deadlines, financial intensity, uncertainty, and other different dimensions Keung and Shen (2012); Zou et al. (2007). However, in a construction project where time really worth money, time planning and management is vital (Duran, 2006). The delay in construction projects, which means the non-completion and achievement of the project inside the predefined



construction contract's duration, is considered to be a standout amongst the most repeating issues in the construction business (Mahamid et al. 2012). As indicated by Duran (2006), the industry and business of construction has an awful notoriety for familiarizing with delays and postpones. Postpone examination is either disregarded or done subjectively by only including a contingency. Therefore, numerous construction projects fail to meet the planning timetable's due dates (Duran, 2006). One of the interested examples among many other countless examples is Indonesia, in which Trigunarsyah (2004) recognized that 47% of the construction ventures were finished within its timetable, only 15% of construction project were in front of the timetable, and 38% were behind it which is a large percentage that requires to be taken into account.

Indonesia is not the only country facing this delay and the postponement of the construction project. On the contrary, most of the world's countries, specifically developing countries face this specified problem in their construction sectors and it becomes a common issue for the construction industry, which are primarily caused by poor pre-construction planning and mismanagement of the project at its various stages (Duran, 2006). Therefore, this research paper represents an investigation attempt to shed light on and review some of the prominent researches and literature that studied the relationship between poor management and planning of the project and the delays that result from it, as the construction delays has become a common bad phenomenon associated with this sector.

### **1.1 Problem Statement and Study objectives:**

Delays in construction ventures are common in the developing countries as well as developed countries' industry (Falqi, 2004). Construction delays and duration issues are frequently responsible of transforming productive ventures into losing projects. These delays can be reduced or prevented by an increased pre- project planning and successful project management as they are one of the most critical success factors of the construction project accomplishment (Yang et al., 2012). Construction projects' managers ought to deliberately evaluate the likelihood of postponements and delays to secure project achievement. Foreseeing and predicting the likelihood of postponements and project preplanning play a crucial role towards the success and achievement of project within its planned timeframe and budget (Kim et al., 2009). Hence, this paper tends to explore and review "**the impact of poor planning and management on the duration of construction projects**", which is achieved through the following subdivided objectives:



- 1) To identify project planning and management process definitions.
- 2) To clarify construction projects situation in developing countries.
- 3) To identify delays, its types and responsibilities.
- 4) To review the relation between the project planning and management and the duration of the projects.

## **1.2 Methodology**

Research methodology can be defined as a subject that deals with how research or study is carried out in a scientific way. The importance of research methodologies lies in its ability to highlight and give essential training in the arrangement and collection of material in a way that can be recognized in an easy way (Saunders, 2011). This research is based on a descriptive methodology, in which the researcher tends to review previous studies and literature that will help in identifying the relationship between the poor planning and management of project and the occurrence of delays in the construction projects. Primary data will be collected from the literature, related studies, cases and investigations which will help in identifying the impact of poor planning and management on the duration of construction projects.

## **2. Literature Review:**

Project Management Institute defines a project as a temporary activity that is initiated to create a unique product, service or outcome that has a specific beginning and end, and is reached when the project objectives are achieved or when the project is terminated because its objectives cannot be achieved or when the need arises for this project no longer exists. The term "temporary" does not necessarily mean shortening the time period, nor does it generally apply to the service or product produced by the project, since most projects have a result that is expected to continue and last, for example: the project which includes the construction of a monument, will produce a result that is expected to last for centuries. Also, the projects may have social, economic or environmental impacts that are greater than the projects themselves in continuity (Snyder, 2014).

The construction project, as defined by Gould and Joyce (2009), is an idea that is defined by the owner and then developed by the designers and then produced by the contractor, where the contractor



returns it to the owner after implementation have accomplished as required. This project requires many processes in order to be successful such as, managing, planning, controlling, monitoring and many other processes. The project management and planning process is an important process and the negligence of it leads to many problems in the construction project, including economic or time duration problems (Lester, 2006).

A literature investigation uncovers that poor venture planning and management is cited by numerous researchers as a delay factor in the construction ventures (Hoseini, 2015). Project planning and management have a wide range of aspects and this review just addresses time aspect of this process; hence it would be important to clarify the project planning and management process precisely.

#### – **Concept of Project Planning and Management**

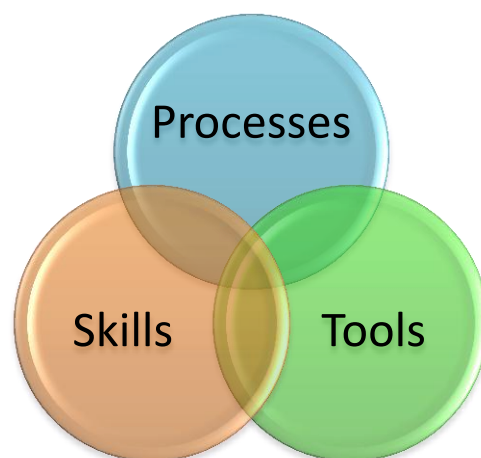
Project is characterized as succession of tasks and activities that have certain targets with specific standards, recognized beginning and complete dates, budgetary constraints, requires nonhuman and human assets and include different functions (Kerzner, 2009). The essential challenge of venture management and administration is to accomplish venture targets considering its constraints and limitations (Zidane, 2012). This requires applying skills, techniques, tools, knowledge to project tasks in order to achieve project prerequisites (PMI, 2013, Westland, 2007). As indicated by (PMI, 2013), project management includes:

- 1) Identifying of the Project requirements.
- 2) Addressing project stakeholder desires.
- 3) Correspondence administration between the partners.
- 4) Adjusting venture imperatives.

Project management must combine the following three components in order to achieve those previous mentioned requirements (Westland, 2007):

- 1) Skills: experience and skills in order to minimize the risks within the venture and in the same manner increase the probability of its success.

- 2) Processes: several techniques and process including ; cost management, time management , risk management , quality management and other processes that are required in order to control and monitor cost , time ,scope and quality of projects.
- 3) Tools: such as; financial software, planning software, review and audit forms that are used by mangers in order to increase project success probability.



**Figure (1): Project management components (Westland, 2007)**

According to (PMI, 2013), "project time management is the processes required to manage timely completion of the project". Indeed, time management is a process that records and controls time spent to finish each activity (PMI, 2013).

It is impossible to satisfy the venture management necessities and apply its parts without planning for the project (Zwikael et al., 2014). Indeed, project planning is a proposition of how to do all the venture administration activities and tasks to accomplish the project objectives. The essential function of time project management is to serve the venture manager as a guide to demonstrate the path from venture begins date to its complete date (Mantel, 2001). Without having the guide, venture manager cannot achieve the project tasks which result in failure in achieving objectives of venture (Mantel, 2001).

Several investigators identify list of the things and steps that must be followed in order to have a good plan for the project. However, PMI introduces project planning process as following steps: create project management plan, Gather requirements, Describe scope, Make Work

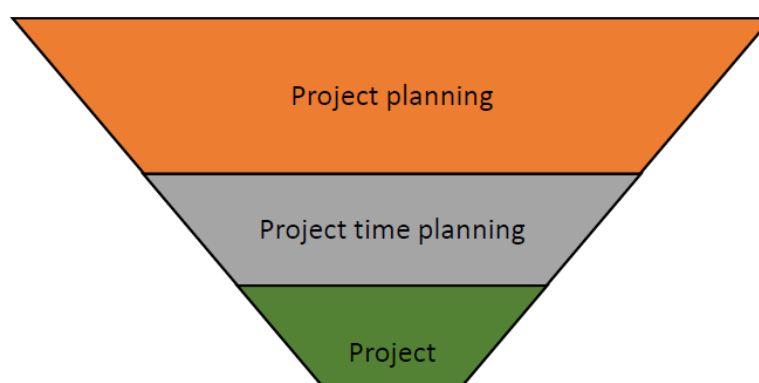


Breakdown Structure (WBS), Express activities, order activities, Evaluate activity resources, Evaluate activity duration, create schedule, evaluate costs, Decide budget, Plan quality, create human resources plan, communication plan, Risk management Plan, clarify risks, Perform quantitative and qualitative risk analysis, Plan risk Responses and finally Plan procurement (PMI, 2013).

Project planning activities continue for the whole duration of the construction phases. Those planning periods are described as in the following (Hoseini, 2015):

- 1) Initiation planning: The initiation planning process frequently incorporates initial scope description and budget preparation from the proprietor association's point of view (Lines et al., 2015)
- 2) Preconstruction planning: After the project budgetary and scope clarification, the project move into the detailed planning phase. This preconstruction period may include the design phase, the planning phase, and the award phase (Lines et al., 2015).
- 3) Construction execution planning: planning may happen amid the execution stage, which comprises of all exercises past activation and completion of preconstruction planning (Lines et al., 2015).

Project schedule is the output of venture time management and planning process (PMI, 2013, Lines et al., 2015). This schedule enables the venture administrator to control the measure of time spent by every action inside the project (Westland, 2007). As specified previously, venture planning is a wide-ranging term and comprises diverse perspectives. The Following figure demonstrates venture planning is limited to project time management and the result of it is project schedule.





**Figure (2): project planning and management concept (Lines et al., 2015)**

– **Construction Projects Situation in Developing Countries**

Long et al (2004) state that performance of construction projects in developing countries is influenced by many factors that results in project delays, poor quality, safety issues and cost over runs. Problems in construction projects are evident across the globe. However, there is a need to focus on factors based on geographical, regional and country. Various factors that influence project performance are same in different countries, but few are different in developing countries including incompetent contractors or designers, poor quality of change management, social constraints, challenges due to technological advancements, site specific issues, lack of advanced tools and techniques, poor planning and estimation of project. All such factors influence the performance of construction projects in developing countries and results in delays. Frimpong, Oluwoye and Crawford (2003) state that cost over runs and delays are very common in building projects. Developing countries have no exception rather they have to face even greater number of problems causing delays and cost overrun than developed countries. Different factors cause delays in construction such as poor construction management skills of contractors, lack of technical performance, increased material prices, and procurement of material and disbursement difficulties from agencies. All these factors need to be identified properly in projects and then proper planning and management is required to cope with the challenges. Developing countries can overcome these problems by improving their project planning, improving monitoring measures, establishing controls and proper knowledge about the construction project management techniques and their implementation. If managed properly, cost and delay problems can be controlled.

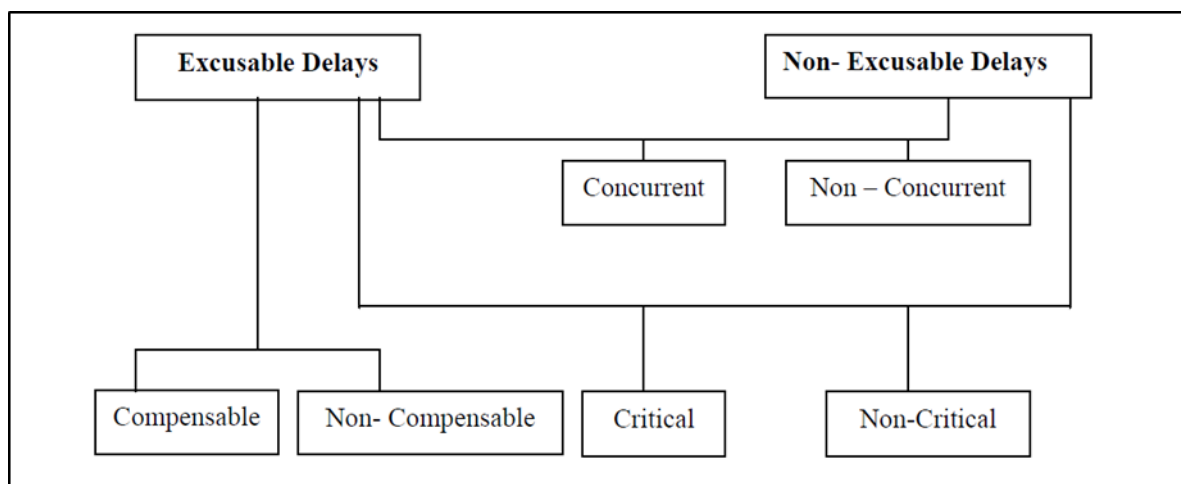
According to Enshassi, Mohamed and Abushaban (2009), construction projects have to face many complex issues and problems due to the impact of certain factors. The most important factors influencing the performance of construction projects in developing countries are lack of resources required for construction, hindrances at borders or due to roads closures thus resulting in material shortage, lack of leadership skills, appreciation of material prices, lack of qualified and experienced personnel and poor quality of construction material and equipment. All these factors influence the performance of project resulting in delays, claims and disputes.

Construction project managers can control delays and cost overruns by focusing on these factors. Sambasivan and Soon (2007) stated that delay of construction projects is a global phenomenon and the construction inductor in developing countries is no exception. Delays in construction projects are caused by improper planning of contractors, inadequate experience, poor site management, inadequate finance, delayed payments, conflicts with subcontractors, lack of material, lack of equipment, and communication gap between parties and errors in construction methods. Such factors results in disputes, cost overrun, delays, litigation, arbitration and total abandonment.

#### – **Delays, Its Types and Responsibilities**

Delay in construction projects is defined as the time difference between the date of termination of the project listed or specified in the contract and the actual date of completion of the project (Falqi, 2004), or the time difference between the planned time and the actual time of project activities (González et al., 2008). The delay in the construction projects is also defined as the time beyond the completion date specified in the contract or beyond the date agreed upon by the parties to hand over the project whether or not the owner is allowed to extend the time or is subjected to a fine or penalty or not (Al-Ghafly, 1995), and the delay was defined as something unexpected and non-existent Planned occurred that cause a delay in the project schedule (Trauner, 2009). In another definition of Stumpf (2000), the delay is an act or event that results in an extension of the time required to perform a specific task required by contract or contract frame, where the delay appears in the form of additional days or late onset of subsequent activity, which may or may not include changes in the scope of the contract.

Several studies and researchers like (Alaghabri et al., 2007); (Vidalis and Najafi, 2002) have classified the delays faced by construction projects into several major types, which are: Concurrent delay, Critical delay, Compensable delay, and Excusable delay. Figure 3 represents the sequential relationships between those different groups of delays (Vidalis and Najafi, 2002).



**Figure (3): the sequential relationships between different groups of delays (Vidalis and Najafi, 2002).**

The project process is affected internally or externally by those previous types of delays. The external causes of delay are due to outside construction project's factors such as utilities, subcontractors, governments, labor unions, suppliers, nature, etc. however, the internal causes of delays are resulted from the contractors , designers , owner and consultants (Vidalis and Najafi, 2002) .

The responsibilities of delay categorized by (Vidalis and Najafi, 2002) as following:

- Neither party responsible: The cause of this delay is neither of the project's parties, and occurs for reasons beyond the control of the parties to the contract, or the so-called "acts of God". In this case, the contractor obtains an extension of the period of execution as the owner deems appropriate and does not receive compensation for the collateral damages. Also he shall not be liable to pay fines or penal conditions.
- Owner responsible: this delay caused by the owner or one of his agents and representatives, and here the contractor is entitled to claim compensation for material damage suffered as well as the extension of the implementation time.

- Contractor or subcontractor responsible: it is considered as a breach of the contract by the Contractor. In this case, the Contractor shall not be entitled to extend the period of implementation of the Project nor to receive compensation. Also, the Owner shall also be entitled to impose delay penalties on the contractor and deduct it from his dues.
- Both parties are responsible: in this case the contractor gets additional time to complete the work, but does not receive compensation and does not pay any penalty or fines.

Furthermore, (Doloi et al., 2012) examined the critical key factors that cause the delays in construction project, where they mentioned that the most critical ones are:

1. Absence of commitment.
2. Inadequate site management.
3. Weak and inadequate site management.
4. Unsuitable planning.
5. Absence of clearness in project scope.
6. Lack of communication.

And all of those previous causes are related to the improper and inappropriate management and planning for the project, which its relation and impact will be identified clearly in the following.

– **The Impact of Poor Planning and Management on the Duration of Construction Projects**

Several researchers and investigators identified causes of construction delays based on the country in which they develop their study. One of the most common delay factors that were mentioned by several researchers was "poor project planning and management". This factor was cited by several authors indicated in the following table (Hoseini, 2015):

**Table (1): poor project management and planning as common delay factor in literature**

| Common delay and time overrun factor | Researcher |
|--------------------------------------|------------|
|--------------------------------------|------------|



|   |   |
|---|---|
| <p>Poor project management and planning</p> | <p>(Khoshgoftar et al., 2010), (Al-Kharashi and Skitmore, 2009), (Faridi and El-Sayegh, 2006), (Assaf and Al-Hejji, 2006), (Sweis et al., 2008), (Pourrostan and Ismail, 2011), (Akogbe et al., 2013), (Marzouk and El-Rasas, 2014), (Muya et al., 2013) , (Toor and Ogunlana, 2008), (Sambasivan and Soon, 2007)</p> |
|---|---|

Considering those previous mentioned researches , poor planning and management was mentioned in their works as general term of delay and time overrun , including : resource planning ,time planning, financial planning, equipment and site management , etc. (Hoseini, 2015).

The advantages of pre-project planning incorporate expanded benefit, higher quality and decreased hazard (González et al. 2008). The efforts used up in this early venture stage influence the level of progress amid start-up and in the detailed outline and development stages (Yang and Wei 2010). This finding is upheld by Thomas and Ellis (2007), who utilized straightforward pre-project planning and management techniques to diminish introductory construction duration by 30%. Hanna and Skiffington (2010) contended that expanded construction management and arranging enables the contractual worker to be more proactive than responsive with respect to basic factors that influence a venture. As per Gibson et al. (2006), the beneficial outcomes of reasonable planning for expense and time before configuration, and also amid the construction stage, are enhanced project result, more noteworthy client fulfillment, and decreased venture duration and cost. Consequently, distinguishing and testing for huge contrasts in the effects of the vital success factors on cost, quality, and time are essential in managing these elements in pre-project arranging as an approach to secure the best outcome of the venture.

As indicated by Dvir et al. (2003), there is a solid relationship between fruitful project planning and management and the achievement of a venture from the point of view of venture partners. These researchers likewise demonstrated that reasonable definitions of useful and specialized details in venture planning can prompt more viable execution of tasks. They additionally

found a solid relationship between effective usage of planning systems and advantages to construction project partners. Such discoveries are affirmed in a later report which demonstrated that venture success can be estimated in perspective of the quality of the project planning and management; though poor management and planning implies uncontrolled changes in the arranging factors of quality, time and cost (Dvir and Lechler, 2004). Zwikael (2009) contended that numerous construction ventures will probably be liable to the danger of poor project arranging when contrasted with ventures in non-construction areas. Zwikael evaluated the significance and impact of project planning in construction projects and found that the degree of utilization of proper venture planning by venture managers and other project partners was not at the ideal level of project prerequisites. He promote contended that a solid attention ought to be set on characterizing the venture scope, venture exercises and expenses. With respect to planning, the improvement of a good venture schedule is imperative to a comprehension of project execution and control. Good scheduling demonstrates a guide and map for venture chiefs, schedulers and planners in observing and following basic milestones and activities amid the progress of venture (Baldwin and Bordoli, 2014).

### 3. Conclusion:

The reason for construction project planning and management is to organize, plan, coordinate, monitor and control the utilization of project goals in the best way as indicated by project partners' needs. It includes numerous processes and sub-processes and incorporates the identifying of venture scope, duties and responsibilities of the project workers, cost evaluating, venture stakeholder management, and additionally the use of arranging and control strategies and instruments. These require learning of the essentials of project management keeping in mind the end goal to create fruitful project schedules and plans, which are vital for the conveyance of the venture to time, quality and cost destinations. Where there is an absence of knowledge, the application of venture planning and management ideas will bring about fragmented and incomplete venture plans or poor planning and, henceforth, loss of project's execution efficiency and performance.

Due to this review of several researches and literature, the researcher can conclude that poor planning and management of the construction projects may lead to several negative effects on the duration and completion of projects. Construction delays and duration issues are frequently responsible of transforming productive ventures into losing projects. These delays can be reduced or

prevented by an increased pre- project planning and successful project management as they are one of the most critical success factors of the construction project accomplishment

### Reference:

- 1) Akogbe, R. K. T., Feng, X., & Zhou, J. (2013). Importance and ranking evaluation of delay factors for development construction projects in Benin. *KSCE Journal of Civil Engineering*, 17(6), 1213-1222.
- 2) Alaghbari, W. E., Razali A. Kadir, M., Salim, A., & Ernawati. (2007). The significant factors causing delay of building construction projects in Malaysia. *Engineering, Construction and Architectural Management*, 14(2), 192-206.
- 3) Al-Ghafly, M. A. (1995). *Delay in the construction of public utility projects in Saudi Arabia* (Doctoral dissertation, King Fahd University of Petroleum and Minerals).
- 4) Al-Kharashi, A., & Skitmore, M. (2009). Causes of delays in Saudi Arabian public sector construction projects. *Construction Management and Economics*, 27(1), 3-23.
- 5) Assaf, S. A., & Al-Hejji, S. (2006). Causes of delay in large construction projects. *International journal of project management*, 24(4), 349-357.
- 6) Baldwin, A., & Bordoli, D. (2014). *Handbook for construction planning and scheduling*. John Wiley & Sons.
- 7) Doloi, H., Sawhney, A., Iyer, K. C., & Rentala, S. (2012). Analysing factors affecting delays in Indian construction projects. *International Journal of Project Management*, 30(4), 479-489.
- 8) Duran, O. (2006). Current risk management applications in Turkish construction industry, an unpublished Master thesis. *Gaziantep University, Gaziantep*.
- 9) Dvir, D., & Lechler, T. (2004). Plans are nothing, changing plans is everything: the impact of changes on project success. *Research policy*, 33(1), 1-15.
- 10) Dvir, D., Raz, T., & Shenhar, A. J. (2003). An empirical analysis of the relationship between project planning and project success. *International journal of project management*, 21(2), 89-95.
- 11) Enshassi, A., Mohamed, S., & Abushaban, S. (2009). Factors affecting the performance of construction projects in the Gaza strip. *Journal of Civil engineering and Management*, 15(3), 269-280.



- 12) Falqi, I. (2004). Delays in project completion: a comparative study of construction delay factors in Saudi Arabia and the United Kingdom. *Unpublished MSc. Thesis, School of the Built Environment, Heriot-Watt University.*
- 13) Faridi, A. S., & El-Sayegh, S. M. (2006). Significant factors causing delay in the UAE construction industry. *Construction Management and Economics*, 24(11), 1167-1176.
- 14) Frimpong, Y., Oluwoye, J., & Crawford, L. (2003). Causes of delay and cost overruns in construction of groundwater projects in a developing countries; Ghana as a case study. *International Journal of project management*, 21(5), 321-326.
- 15) Gajewska, E., & Ropel, M. (2011). Risk Management Practices in a Construction Project—a case study. *Swedia, Chalmers University Of Technology.*
- 16) Gibson Jr, G. E., Wang, Y. R., Cho, C. S., & Pappas, M. P. (2006). What is preproject planning, anyway?. *Journal of Management in Engineering*, 22(1), 35-42.
- 17) González, V., Alarcón, L. F., & Mundaca, F. (2008). Investigating the relationship between planning reliability and project performance. *Production Planning and Control*, 19(5), 461-474.
- 18) Gould, F. E., & Joyce, N. (2009). *Construction project management*. Upper Saddle River, N.J.: Pearson Prentice Hall.
- 19) Gunduz, M., Nielsen, Y., & Ozdemir, M. (2013). Fuzzy assessment model to estimate the probability of delay in Turkish construction projects. *Journal of Management in Engineering*, 31(4), 04014055.
- 20) Hanna, A. S., & Skiffington, M. A. (2010). Effect of preconstruction planning effort on sheet metal project performance. *Journal of Construction Engineering and Management*, 136(2), 235-241.
- 21) Hoseini, E. (2015). *Project Time Planning in Norwegian Construction Industry-An Empirical Study* (Master's thesis, NTNU).
- 22) Kerzner, H., & Kerzner, H. R. (2017). *Project management: a systems approach to planning, scheduling, and controlling*. John Wiley & Sons.
- 23) Keung, C. C., & Shen, L. Y. (2012). Measuring the networking performance for contractors in practicing construction management. *Journal of Management in Engineering*, 29(4), 400-406.



- 24) Khoshgoftar, M., Bakar, A. H. A., & Osman, O. (2010). Causes of delays in Iranian construction projects. *International Journal of Construction Management*, 10(2), 53-69.
- 25) Kim, S. Y., Van Tuan, N., & Ogunlana, S. O. (2009). Quantifying schedule risk in construction projects using Bayesian belief networks. *International Journal of Project Management*, 27(1), 39-50.
- 26) Lester, A. (2006). *Project management, planning and control: managing engineering, construction and manufacturing projects to PMI, APM and BSI standards*. Elsevier.
- 27) Lines, B. C., Sullivan, K. T., Hurtado, K. C., & Savicky, J. (2015). Planning in construction: longitudinal study of pre-contract planning model demonstrates reduction in project cost and schedule growth. *International Journal of Construction Education and Research*, 11(1), 21-39.
- 28) Long, N. D., Ogunlana, S., Quang, T., & Lam, K. C. (2004). Large construction projects in developing countries: a case study from Vietnam. *International Journal of project management*, 22(7), 553-561.
- 29) Mahamid, I., Bruland, A., and Dmaid, N. (2012). Causes of Delay in Road Construction Projects. *Journal of Management in Engineering*, 28(3), 300-31.
- 30) Majid, I. A. (2006). *Causes and Effects of delays in ACEH Construction Industry* (Doctoral dissertation, Universiti Teknologi Malaysia).
- 31) Mantel, S. J. 2001. *Project management in practice*, New York, Wiley.
- 32) Marzouk, M. M., & El-Rasas, T. I. (2014). Analyzing delay causes in Egyptian construction projects. *Journal of advanced research*, 5(1), 49-55.
- 33) Muya, M., Kaliba, C., Sichombo, B., & Shakantu, W. (2013). Cost escalation, schedule overruns and quality shortfalls on construction projects: The case of Zambia. *International Journal of Construction Management*, 13(1), 53-68.
- 34) Olawale, Y., & Sun, M. (2012). PCIM: Project control and inhibiting-factors management model. *Journal of management in engineering*, 29(1), 60-70.
- 35) PMI (2013). *A Guide to the project management body of knowledge: (PMBOK guide)*, Atlanta, Project Management Institute.
- 36) Pourrostan, T., & Ismail, A. (2011). Significant factors causing and effects of delay in Iranian construction projects. *Australian Journal of Basic and Applied Sciences*, 5(7), 450-456.



- 37) Sambasivan, Murali, and Yau Wen Soon. "Causes and effects of delays in Malaysian construction industry." *International Journal of project management* 25, no. 5 (2007): 517-526.
- 38) Snyder, C. S. (2014). A guide to the project management body of knowledge: PMBOK (®) guide. *Project Management Institute: Newtown Square, PA, USA*.
- 39) Stumpf, G. R. (2000). Schedule delay analysis. *COST ENGINEERING-ANN ARBOR THEN MORGANTOWN-*, 42(7), 32-32.
- 40) Sweis, G., Sweis, R., Hammad, A. A., & Shboul, A. (2008). Delays in construction projects: The case of Jordan. *International Journal of Project Management*, 26(6), 665-674.
- 41) Thomas, H. R., & Ellis Jr, R. D. (2007). Contractor prebid planning principles. *Journal of Construction Engineering and Management*, 133(8), 542-552.
- 42) Toor, S. U. R., & Ogunlana, S. O. (2008). Problems causing delays in major construction projects in Thailand. *Construction management and economics*, 26(4), 395-408.
- 43) Trauner, T. J. (2009). *Construction delays: Understanding them clearly, analyzing them correctly*. Butterworth-Heinemann.
- 44) Trigunarysyah, B. (2004). Constructability practices among construction contractors in Indonesia. *Journal of construction engineering and management*, 130(5), 656-669.
- 45) Vidalis, S. M., & Najafi, F. T. (2002, June). Cost and time overruns in highway construction. In *4th transportation specialty conference of the Canadian society for civil engineering, Canada* (Vol. 5, No. 8).
- 46) Westland, J. (2007). *The Project Management Life Cycle: A Complete Step-By-Step Methodology for Initiating, Planning, Executing & Closing a Project Successf*. Kogan Page Publishers.
- 47) Yang, J. B., & Wei, P. R. (2010). Causes of delay in the planning and design phases for construction projects. *Journal of Architectural Engineering*, 16(2), 80-83.
- 48) Yang, L. R., Chen, J. H., & Huang, C. F. (2012). Requirements definition and management practice to improve project outcomes. *Journal of Civil Engineering and Management*, 18(1), 114-124.
- 49) ZIDANE, Y. (2012). Project Change in Large Scale Engineering Projects.
- 50) Zou, P. X., Zhang, G., & Wang, J. (2007). Understanding the key risks in construction projects in China. *International Journal of Project Management*, 25(6), 601-614.



[www.mecsjs.com](http://www.mecsjs.com)

51) Zwikael, O. (2009). Critical planning processes in construction projects. *Construction Innovation*, 9(4), 372-387.

52) Zwikael, O., Pathak, R. D., Singh, G., & Ahmed, S. (2014). The moderating effect of risk on the relationship between planning and success. *International Journal of Project Management*, 32(3), 435-441.

53) Saunders, M. N. (2011). *Research methods for business students*, 5/e. Pearson Education India.