

NoSQL: Future of BigData Analytics Characteristics and Comparison with RDBMS



Muhammad Arshad, M. Nawaz Brohi, Tariq Rahim Soomro,
Taher M. Ghazal , Haitham M. Alzoubi , and Muhammad Alshurideh 

Abstract The growth of digital world has become so fast and the data volume turned to more composite in relations to size from terabyte and higher, the variation of data including structured, unstructured and hybrid, pace has gone so high in growth of data. This is called 'Big Data' as a worldwide phenomenon. Typically, this is measured as the collection of data from multiple sources has become so huge that it can't be well exploited or maintained by using regular methods used for data management: e.g., RDBMS the conventional relational database management systems or search engines used conventionally. The experts have been working to solve handle these problems, the traditional relational database management systems are modified by precisely

M. Arshad
Department of Computer Science, SZABIST Dubai, Dubai, UAE

M. N. Brohi
Bath Spa University RAK Campus, RAK, UAE
e-mail: mnbrohi@bathspa.ac

T. R. Soomro
Institute of Business Management, CCSIS, Karachi, Sind, Pakistan
e-mail: tariq.soomro@iobm.edu.pk

T. M. Ghazal
School of Information Technology, Skyline University College, Sharjah, UAE
e-mail: taher.ghazal@skylineuniversity.ac.ae

Center for Cyber Security, Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia (UKM), Selangor, Malaysia

H. M. Alzoubi (✉)
School of Business, Skyline University College, Sharjah, UAE
e-mail: haitham.alzubi@skylineuniversity.ac.ae

M. Alshurideh
Department of Marketing, School of Business, University of Jordan, Amman, Jordan
e-mail: malshurideh@sharjah.ac.ae

Department of Management, College of Business Administration, University of Sharjah, Sharjah, UAE

designed a set of substitutes DBMS's; which are search based systems, New SQL and NoSQL. This goal of this study is to provide characteristics, classification and evaluation of database management system called NoSQL in Big Data Analytics. The study is planned to help people, specifically the companies to gain. The current computation power also falls less against the massive extended storage volume. The fast growth of unstructured data requires a complete model change for the new age to meet the incoming challenges and observe the progress of new proficient techniques for data engineering. The explanation of BigData, RDBMS and NoSQL have also been described in this study to put light on difference of data and requirement of new technology. This study will highlight the differences, capabilities of both RDBMS and NoSQL database management systems in accordance with BigData. With the help of qualitative and quantitative approach the study has emphasized on the limitation of RDBMS and requirement of new technology NoSQL used for unstructured BigData used for analytics.

1 Introduction

NoSQL, for "Not Only SQL", attributes to a diverse and progressively more recognizable collection of non-relational data management systems; where SQL is not being used for manipulation of data as well as databases are not built primarily on tables (Akhtar et al., 2021; Al Ali, 2021; Alhamad et al., 2021, 2022; Ali et al., 2021). While working with massive data whose architecture does not necessitate a relational model NoSQL database management system are preferable. These structures are disseminated, non-relational databases planned for all-encompassing wide range data storage as well as for densely-parallel data operations and processing over a great amount of commodity servers. Non- SQL languages and systems are also employed by them to interrelate with data and information (although several latest APIs that renovate SQL inquiries to the system's indigenous query language or tool). NoSQL database systems came up next to chief Internet corporations, like Facebook, Amazon and google which experienced difficulties while working with massive amounts of data and information with usual RDBMS systems were not able to deal with. Multiple processes can be carried out by them like, investigative and prophetic analytics, ETL-style the transformation of data, and OLTP non-mission-critical (like handling transactions between organizations and long duration transactions). Unlike the conventional DBMSs and data warehouses, these systems as inspired by Web 2.0 applications are planned to reach out to the maximum number of those users who are doing updates and the reads as well (Ali et al., 2022; Alnazer et al., 2017; Alnuaimi et al., 2021; Alsharari, 2021; Alshurideh et al., 2022).

A relational database system which is planned and intended to offer ACID (Atomicity, Consistency, Isolation, Durability) properties, traditional SQL-based OLAP in Big Data domain and real-time OLTP (Online Transaction) are known as NewSQL system. Utilizing NoSQL-approach features like column-based information storage and scattered styles this system has covered up all the limitations encountered by

conformist RDBMS system. Other novel features introduced by this system are in-memory processing, symmetric multiprocessing (SMP) or Massively Parallel Processing.

The data that is too much comprehensive, ambiguous, readily changing is hard to be tackled using the conventional methods as per the observation of analytics. These days every research institution, businesses, and governments are generating extraordinary amounts of data which is too complex as well. Hunting of required information from such enormous data is very crucial for the organizations. It's a great challenge to extract the meaningful insight out of the bulk of data swiftly. That is why analytics has become inextricably important to understand the significance of Big Data to perk up their business performance and boost their market share. In the past few years, the ways to deal with the variety, velocity, and volume of big data has improved to a great extent. The immense rise in data size necessitates rapid analytics with each new inquiry by the application user. This situation has escorted the technologists to introduce a new DBMS system that can triumph over this processing holdup at Database side. As the structural design of RDBMS has restrictions to handle such huge data and carry out analytics. NoSQL architecture is specifically designed to deal with such speed breakers. Thanks to the flexible and adaptable architecture of NoSQL massive volumes of information can be processed rapidly.

In this paper, we will discuss the conventional RDBMS features as well as its limits to manage the huge data. Furthermore, NoSQL Databases will be discussed along with their types and distinctive characteristics to deal with Big Data will be explained. The application areas where NoSQL databases can incorporate will also be discussed. By means of the industry's experience with NoSQL, we will also attempt to elucidate the problems that can be encountered using later mentioned systems for Big Data. Finally, there is going to be a comparison between the two systems about how they contend with normal data to Big Data.

2 Literature Review

From the several data-models, the model that has been surpassing all others is a relational model from the start of early 80 s, with achievements like Oracle databases, MySQL and Microsoft SQL Server also known as Relational Database Management System (RDBMS). The models mentioned before are all designed on the relational model. The main reason to build RDBMS was to provide data processing to businesses and from that time till now RDBMS is proving as the best tool for information storage whether the information is personal data, financial statements, transaction processing and so on (Aziz & Aftab, 2021; Cruz, 2021; Eli, 2021; Farouk, 2021; Ghazal et al., 2021a, 2021b, 2021c, 2021d).

2.1 *Big Data*

As time went by, the provisions for data kept on growing. This data has seen a revolution from structured to unstructured in form and from Megabytes, Gigabytes, Terabytes to Petabytes in size this ever-changing data has caused people to consider about a different solution to supervise this big amount of data. As the data became big, comprehensive, multifaceted, structured or amorphous, and diverse, it needed noteworthy consideration and concentration. Vast amounts of data are being produced at a very swift rate from a variety of distinct prospective areas, systematic tools, and the internet, especially the world known social media, mentioning a few of them. This type of data was termed as Big data. Big data points to those datasets the size of which are past the ability of typical database software techniques to capture, process and manage, store, and examine (Al-Khayyal et al., 2021; Al Batayneh et al., 2021; Alshurideh, 2022; Alshurideh et al., 2022). This characterization is deliberately subjective because it is evident that as technology progresses over the passage of time, the volume of datasets that become licensed as big data will also enlarge (Al Guergov & Radwan, 2021; Hamadne et al., 2021; Hanaysha et al., 2021a, 2021b; Joghee et al., 2020; Naqvi et al., 2021; Shebli et al., 2021).

4v's can explain the nature of Big data easily. 4v's include Volume, Variety, Velocity, Variability.

2.1.1 Volume

The magnitude of produced and stored data, its importance, value and potential insight is determined by its size. Means whether the data is worthy enough to be termed as Big data or not.

2.1.2 Variety

Variety is the kind, category, and the architecture of the data. This assists experts who scrutinize it to efficiently employ the consequential insight. Big data it the collection of unstructured type data including videos, audios, images and text; also, it uses data fusion to fill in the missing pieces.

2.1.3 Velocity

It can be described as, the pace of data production and processing in the system to comply with the business requirements, tests, and obstacles that stretch out in the technique of escalation, improvement, and lastly the growth. Big data is usually obtainable in up to date format.

2.1.4 Variability

Irregularity and unpredictability of the data set can obstruct the course of action to hold and handle it (Fig. 1).

While considering big data, one problem is its growth, but the other bigger issue is the dire need to manage and store not just the structured data but also the unstructured data as well as the pictures, videos and files. An analytical case in point is that the relational model cannot deal with the data traffic that the social media sites like Facebook and Twitter produce, also it is not the type of data they want to store. Now, for the conventional data processing systems, this significant velocity of the increase in the volume of the data poses a solemn challenge (Alshurideh et al., 2020; Alzoubi, 2021a, 2021b; Alzoubi et al., 2022a; Alzoubi et al., 2020b).

Newly, though, in many cases, the utilization of relational databases escorted to troubles both due to discrepancy and glitches in the design of data and restraints of parallel scalability among various servers and massive size of data. The two main tendencies that brought the mentioned issues into the consideration of international software community are.

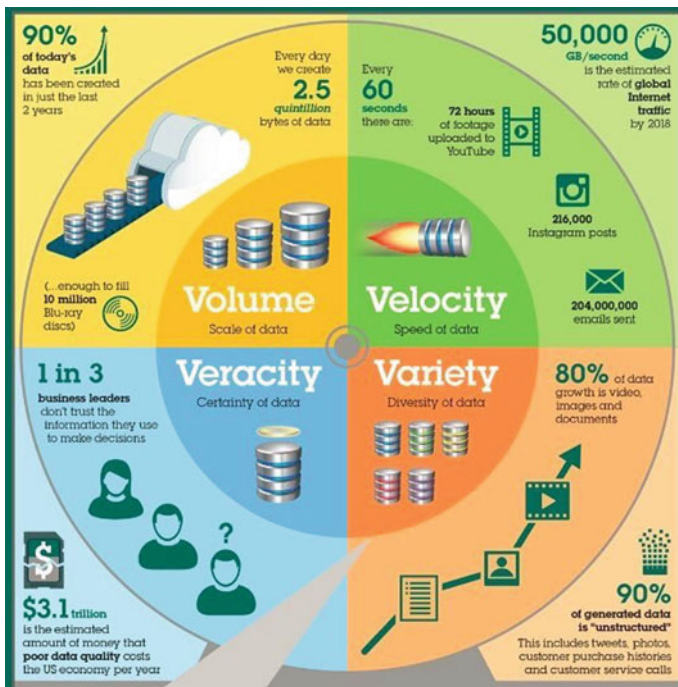


Fig. 1 BigData 4 V explanation (“Big Data,” IBM, [Online]. Available: <http://www.ibmbigdatahub.com/infographic/extracting-business-value-4-vs-big-data>)

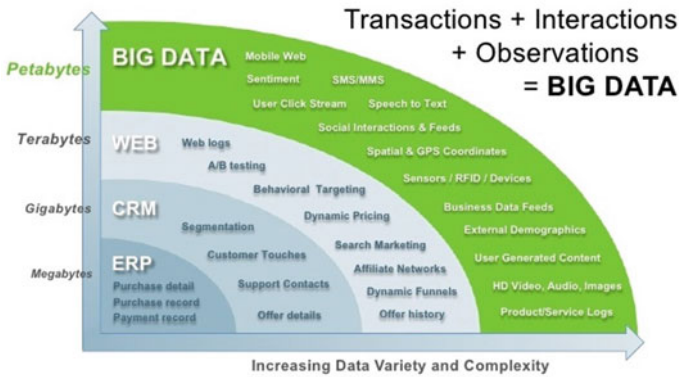


Fig. 2 BigData Transactions with Interaction and Observations (Slide Share, [Online]. Available: <https://www.slideshare.net/cloudstack/vbacd-july-2012-apache-hadoopnow-and-beyond>. [Accessed 12 February 2018].)

1. The massive increase of the quantity of data produced by sensors, systems and users, additionally fast paced by the attention of a huge module of this volume on big disseminated systems like Google, Amazon and other cloud services.
2. The escalating interconnection and intricacy of data fast paced by Web 2.0, Internet, social networks and exposed and uniform access to sources of data consists of different systems at a very large number (Fig. 2).

Big Data = Transaction + Observation + Interactions.

Due to this very reason, a lot of emerging companies took up different kinds of non-relational databases, which are also known as NoSQL databases and the application run arise e.g. Yahoo which used PNUTTS to fulfill enormously parallel and physically worldwide dispersed database system to run their web based applications.

Since their release NoSQL (“Not Only SQL”) systems are extensively accepted in several realms. The main idea behind NoSQL systems is to hold up applications not properly served by relational systems, specifically those which involved in managing and processing BigData. NoSQL system can be classified as graph databases, document stores and key-value stores. It is important to mention here that, there is not one specific query language like standard query language used in RDBMS or a typical APIs used to communicate with various NoSQL systems. Normally, customers are required to use custom build APIs at programming level to communicate (Alzoubi & Ahmed, 2019; Alzoubi & Aziz, 2021; Alzoubi & Yanamandra, 2020; Alzoubi et al., 2020, 2021). This makes portability to be reduced and necessitates code at system-specific level.

2.2 Characteristics of RDBMS

The data organized in relational databases is in the form of tables, which are made up of columns and rows. In order to remove ambiguities during queries, these tables cannot have duplicate rows and every table has been assigned a primary key to a column that distinctively recognize each and every row that is known as record. For example, Fig. 3 demonstrates that Product_ID in the product table is the primary key. Author_ID Column used as Foreign Key in the Product_Book table, which is a child table, Author_ID column is used to reference Author table which is a parent table. The keys used like Foreign and performance, that scan be described as supplementary table relationships may possibly be mandatory while retrieval of data.

Multiple tale inheritances are utilized by libraries' database to hoard familiar traits in a common table knows as Product table (please refer to Fig. 3), every single one of unique attributes is kept in specific type product tables. This said method is far additional well-organized in contrast to concrete table inheritance where for every product category a new table is fashioned, and the queries used are custom-made for specific products. But, as explained to get all the significant characteristics of an established solution multiple table inheritance necessitate many joined operations.

The RDBMS make it certain to have Database ACID properties as the primary prerequisite. For databases, ACID properties act as the vital concept. The abbreviation globally knows as Atomicity, Consistency, Isolation, and Durability.

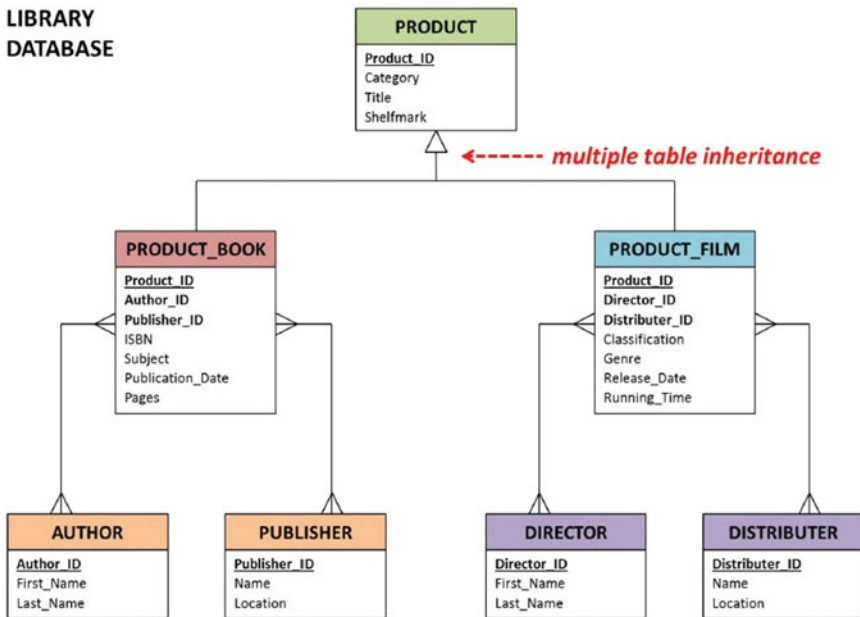


Fig. 3 Relational Database Schema [17]

The ACID makes certain arrangements to business to keep these purchase of sweater dealings from overlaying one another thus the merchant is kept safe from the flawed register and account balances.

2.2.1 Atomicity

Atomicity is the first ACID property and it is best explained by the phrase “all or nothing” to understand this let’s consider an example; When a database gets an update, either all of it is available or none of the updates happen to be accessible to anybody past the application or user executing the update. The above-mentioned action performed on database is called a transaction and it is either assigned or canceled. In other words, only a part of an update cannot be put into the database, you get whole of it or none.

2.2.2 Consistency

ACID property of consistency makes sure that if there is a change to values in an instance then there will be a consistent change to other all values in that specific instance. The constraint of consistency is a base on data and it assists in the system as precondition, the condition after the execution knows as post-condition and at the end the change or transformation condition to be ensured on every transaction (Kashif et al., 2021; Khan, 2021; Lee & Ahmed, 2021; Lee et al., 2022a, 2022b).

2.2.3 Isolation

The third property known as isolation section of the ACID is needed when there are many transactions running in parallel known as concurrent. The executions of transaction that take place in parallel are known as concurrent transactions, for example shared multiple users accessing shared objects in the figure the scenario is explained at the top as actions happening over time. Basically, the safety precautions employed by DBMS to thwart clashes between concurrent transactions is known as isolation.

Let’s understand this with the help of an example, if there are two parties updating the similar catalog article, the changes made by the first party should not depend on or be affected by the changes made by the other party to be able to work in isolation means that both parties act as they are the solitary user. Every change must be kept isolated from other users of the same catalog (Fig. 4).

Serializability is another important concept which should be understood while debating separation over transactions. The execution of transactions could be serialized once the consequence on the database is unchanged whether the transactions are executed in an interleaved manner or in serial order. Concurrent transaction i.e. Transactions 1 through Transaction 3 are being executed at the same time as it can be seen in the figure. An important point to keep in mind here is that in serialized

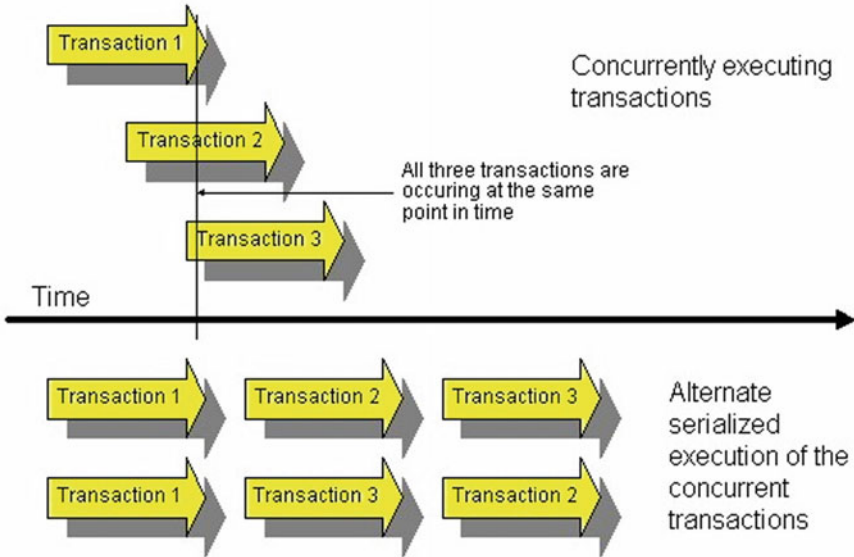


Fig. 4 Isolation Relational Database Property (“Database ACID Properties,” [Online]. Available: https://www.servicearchitecture.com/articles/database/acid_properties.html.)

execution it is not compulsory that transactions started first will be the ones automatically completed before the finishing the other transactions in the sequential execution (Mehmood, 2021; Mehmood et al., 2019; Miller, 2021; Mondol, 2021; Obaid, 2021).

2.2.4 Durability

Durability is that ACID property which attends to the need of keeping a record of committed transactions. These updates are not supposed to be missed in any case as it’s a very critical thing. It’s the systems capability to recover the completed transactions in system or storage media failure. The durability features are as follows:

- The recovery of recently committed transactions in case of database failure
- The recovery of recently committed transactions in case of application failure
- The recovery of recently committed transactions in case of CPU failure
- The recovery of recently committed transactions in case of storage.

The restrictions of RDBMS’s are to deal with amorphous, diverse, heterogeneous, massive amounts of data. For RDBMS retailer it is a huge challenge because of its architecture. This test to manage the big data has compelled them to devise a new technology that can handle the said amounts of data and information.

SQL-Like centralized databases have been pushed towards their perimeters by computational processing and storage requirements of applications like Big Data used for Analytics, Social Networking and Business Intelligence having large than

petabyte datasets. So, these limitations paved a way that directed to the growth of horizontally ascendable, dispersed non-relational type of data stores, which are called as No-SQL databases, like Google's has built open source HBase implementation and Bigtable and the well-known Facebook has Cassandra. The effectiveness, competence, and cost-effectiveness of these approaches are gained by the embodiment of distributed architecture-based key-value stores, for example Voldemort and Cassandra. In regard to Data warehousing, Web 2.0, Grid and applications used in Cloud were very hard with RDBMS and it was a major drawback of this system. Pokorny (2013), emphasizes mainly to NoSQL databases from the perspective of cloud environment, chiefly concurrency model and scalability in horizontal fashion. The Relational Database Management Systems (RDBMS) and NoSQL databases are differing from each other, but NoSQL databases did not assure ACID properties (Radwan & Farouk, 2021; Shamout et al., 2022).

2.3 Characteristics of NoSQL Databases

The conventional database systems are designed on the basic idea of execution of transactions in the manner to keep the data veracity and reliability. This keeps the data consistent while managing it. The features of transactions are also known as ACID (Atomicity, Consistency, Isolation, and Durability) as we have already discussed. Though, developing a system compliant with ACID has made known to be a trouble. CAP-theorem has been observed, i.e. clashes arose among distributed systems diverse sides of high availability that are not completely resolvable.

2.3.1 Strong Consistency

On updates to the data set the version of data seen by the clients is totally same e. g. through the method of two-phase commit protocol (XA transactions), and ACID.

2.3.2 High Availability

In the case, if a few of the machines in a cluster are not working, all the clients still can always find a copy of the requested data. Down machines do not create a problem in this matter.

2.3.3 Partition-Tolerance

The goal of the entire system always is to maintains its characteristics and features even while being positioned on various servers at the same time being transparent to

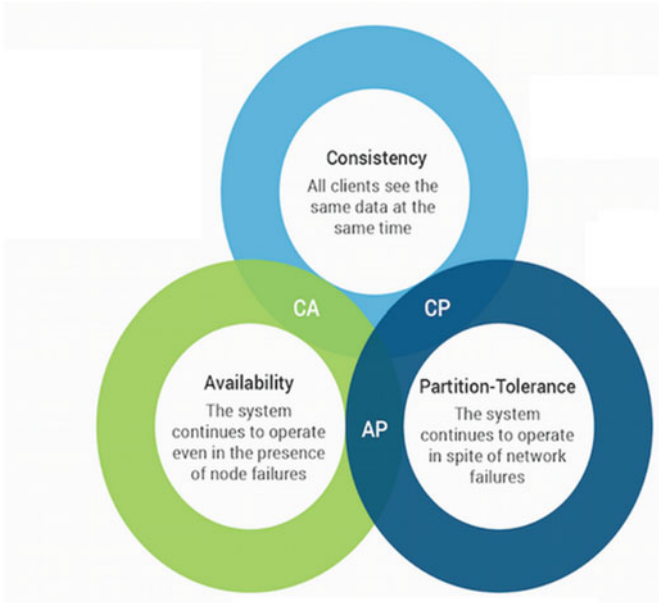


Fig. 5 Characteristics of NoSQL Database

the client. According to the CAP-Theorem, at the same time, out of three only two dissimilar aspects of scaling out can be attained entirely (see Fig. 5).

To attain improved Availability and Partitioning a lot of the NoSQL databases mentioned above have lessened the needs for Consistency. This step laid the way to develop systems globally called as a BASE (Basically Available, Soft-state, Eventually consistent). NoSQL databases have been classified according to the CAP theorem by Han, J. They had compared different NoSQL databases by executing multiple different criteria (Al Ali et al., 2021; Alzoubi et al., 2021; Batayneh et al., 2021).

Main Usages of NoSQL Database can be categorized as (1) Huge-scale and wide data calculation and processing (processing in parallel in the distributed systems); (2) Embedded IR (general machine-to-machine data search and reclamation); (3) Investigative analytics done on unstructured and structured data (knows as expert level); (4) Huge size data storage (unstructured, semi-structured, small-packet structured) (Afifi et al., 2020).

They prove valuable as well for machine-to-machine communication for information and data retrieval, recovery and exchange, for dispensing large number of executions, to the extinct ACID restrictions can be made soft, or the way is to apply them on application side not on DBMS side. In conclusion, when we are to deal with semi-structured or hybrid data these systems act as very good probing analytics, nonetheless to get to the lowermost of intellect, the researcher should be a skillful mathematician working in accordance with an expert programmer (Ghazal, 2021; Ghazal et al., 2021a, 2021b).

Table 1 Key-Value store NoSQL Database

CAR	
Key	Value
1	Make: Infiniti Model: Q50 Color: White Year: 2016
2	Make: Volvo Model: XC90 Color: Beige Year: 2009 Transmission: Auto Type: 4*4

“Characteristics of NoSQL,” [Online]. Available: <https://www.forbes.com/sites/forbestechcouncil/2017/09/18/renaissance-in-cloud-datamanagement/#4509108030cf>

2.4 Classification of NoSQL Databases

NoSQL databases have been classified by Leavitt (2010) in three types: Key-value stores e.g. SimpleDB column-oriented databases—e.g. Cassandra, HBase, Big Table and document-based stores—e.g. CouchDB, MongoDB. In this segment, according to the suitability of different kinds of tasks we categorize NoSQL Databases into four basic categories,

- (1) Key-Value stores.
- (2) Document databases (or stores).
- (3) Wide-Column stores.
- (4) Graph databases.

2.4.1 Key-Value Stores

Classically, in these DBMS the data objects are stored as alpha-numeric identifiers (keys) and related values in plain, standalone tables (also known as —hash tables||). The values could possibly be as simple as text strings or could be more complex like lists and sets. Data searches can usually one can perform data searches only with the use of keys, not values, and they are restricted to precise matches. See Table 1.

2.4.2 Document Databases

As the name indicates, document databases and the idea were derived from Lotus Notes, these mainly are designed and intended to store and manage the different kind of documents. Customary data exchange systems like JSON (Javascript Option Notation), XML, or BSON (documents are encoded by Binary JSON). Contrasting to

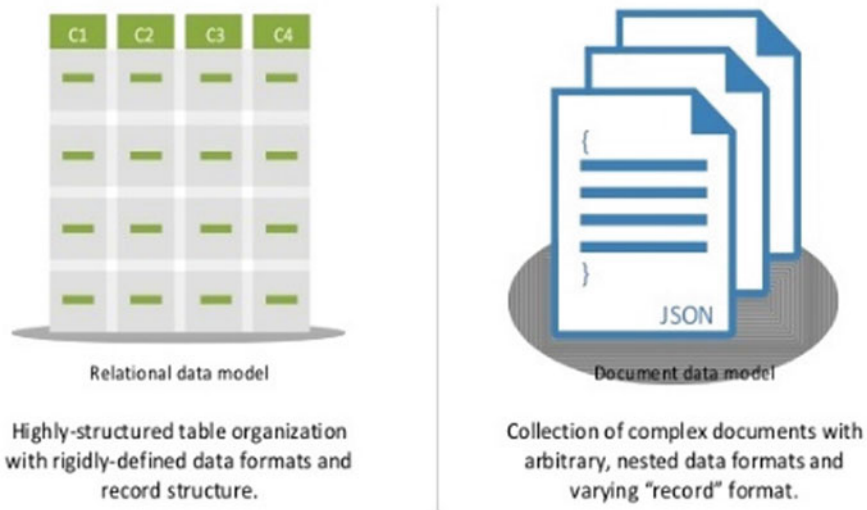


Fig. 6 Document Type Database vs Relational Database (Slide Share, [Online]. Available: <https://www.slideshare.net/cloudstack/vbacd-july-2012-apache-hadoopnow-and-beyond>. [Accessed 12 February 2018].)

the uncomplicated key-value stores illustrated above, in the value column of document databases structured and unstructured data is present—particularly attribute name/value pairs. Hundreds of these attributes can dwell into a single column, also from row to row the type and number of attributes recorded can differ. In document databases, the values and keys are totally searchable which is in contrast with simple key-value stores (Ghazal et al., 2013, 2021c; Kalra et al., 2020) (Fig. 6).

2.4.3 Wide-Column (or Column-Family) Stores (BigTable-Implementations)

Wide-Column (or Column-Family) stores (after this WC/CF) are just like document databases. To house multiple attributes for each key they utilize a column based distributed data structure. Whilst several WC/CF stores comprise a Key-Value DNA (for example the Cassandra Dynamo-inspired), the majority are designed like Google’s Bigtable, that is petabyte scale internal system based on distributed storage for data. This system is developed by Google for its famous search engine and additional products like Finance by Google and Google Earth. In general, the capability is not only to reproduce Google’s storage structure BigTable, but Google’s file system which is distributed in architecture (GFS) and its processing framework which is parallel MapReduce too. Similar scenario is with Hadoop, which comprises the file system called the Hadoop File System (HDFS, based on GFS) + Hbase (a



Fig. 7 NoSQL Wide Column Database (“Graph Databases: NOSQL and Neo4j,” infoQ, [Online]. Available: <http://www.infoq.com/articles/graphnosql-neo4j> [Accessed 26 March 2018].)

Bigtable-style storage system) + MapReduce (Khan et al., 2021; Lee et al., 2021) (Fig. 7).

2.4.4 Graph Databases

Relational databases have been replaced by graph databases with more organized relational graphs having key value pairing interconnected to each other. These are like object-oriented databases because of the graph is illustrated as object-oriented network of graph nodes, (objects in concept), relationship of node knows as edges and properties (the object characteristics stated as key-value pairs). The four NoSQL forms conferred here are those that are related with relations. Among other NoSQL DMS, these types are considered more human-friendly because they focus on the visual depiction of data and information (Fig. 8).

Many big organizations that deal with big data have now adopted NoSQL. Following is the table that shows a few of these big businesses (Table 2).

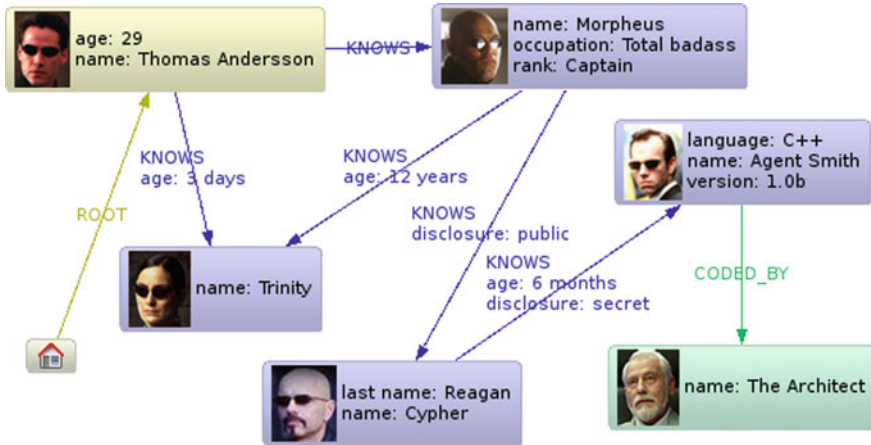


Fig. 8 NoSQL Graph Database Store (“Graph Databases: NOSQL and Neo4j;” infoQ, [Online]. Available: <http://www.infoq.com/articles/graphnosql-neo4j>. [Accessed 26 March 2018].)

Table 2 NoSQL type used by Companies

Company name	NoSQL name	NoSQL storage type
eBay	Cassandra MongoDB	Column Document
Facebook	Cassandra	Column
Lots of Words	CouchDB	Document
MongoHQ	MongoDB	Document
Mozilla	Hbase	Column
Netflix	Hbase Cassandra	Column Column
Twitter	Cassandra	Column

Big businesses because of their high data storage demands have converted to NoSQL and its experts are also in the favorable light now.

2.5 Comparison Between RDBMS and NoSQL

In this study as the characteristics of both RDBMS and NoSQL has been described, the comparison between RDBMS and NoSQL has been analyzed in detail. Which shows the major differences and capabilities of both systems according to customer needs (Table 3).

Table 3 Comparison between RDBMS and NoSQL (Author Created)

S. No	Description	RDBMS	NoSQL
1	Data size	Gigabyte, Terabyte	Petabytes and greater
2	Schema	Static Schema	Dynamic Schema
3	Type	Relational	Non-Relational
4	Data	Structured	Unstructured
5	Scalability	Vertical	Horizontal
6	Language	Standard Query Language	Un-structured Query language
7	Joins	Helpful to design complex queries	No Joins, don't have interface to design complex queries
7	OLTP	Recommended and best used for OLTP Systems	Less likely to be considered as OLTP systems
8	Flexible	Rigid Schema, relationship bounded	Very flexible and no rigid Schemas
9	Auto elasticity	Require down time in most cases	Automatic, no outage required
10	Consistency	Strong consistency supported	Consistency varies per solution; some solutions have tunable consistency
11	Scale	Scale well vertically	Scale well horizontally
12	Transaction	ACID	CAP Theorem

“Characteristics of NoSQL,” [Online]. Available: <https://www.forbes.com/sites/forbestechcouncil/2017/09/18/renaissance-in-cloud-datamanagement/#4509108030cf>

3 Material and Methods

In 1980's the primary cohort of commercial systems come into sight by Teradata Corporation and in the same time, the necessity surfaced for well-defined systems to determine the ability of DBMS dealing with very big quantities of data. Motivated by vendor's desires to weigh the commercial systems, in the starting of 90's the Transaction Processing Performance Council designed a series of data warehouse end-to-end benchmarks. Likely systems have been developed by TPC-H and PC-R at the beginning of 2000 (the details are all accessible from the TPC website²). With some update on a venture data warehouse, these benchmarks are limited to a data size of a terabyte, highlighting single and multi-user performance of complex SQL query processing abilities. Even before this, academia had started developing micro-benchmarks like EXRT and XMark benchmarks for XML-related DBMS technologies and the OO7, the Wisconsin benchmark, and BUCKY benchmarks for object-oriented DBMSs, (Matloob et al., 2021; Naqvi et al., 2021).

With the passage of time, the volume of data kept on growing from megabytes to petabytes in size and from simple data models (a few tables with a small number of relationships) to complex ones (big tables with many complex relationships). This change in the demand for data needs has led TPC to act in response. In the dawn

of 2000's TPC-DS developed its next generation decision support benchmark. Its foundation is based on the SQL programming language, but it consists of several big data elements, like exceedingly large system sizes and data. Even though the existing limit is 100 terabytes; the schema and data generator can be expanded to petabytes. Quite composite analytical data queries are also contained by it using sophisticated and complicated SQL structures and a synchronized update model.

3.1 Adaptation of NoSQL

The term NoSQL was invented in 1998. Lots of people assume NoSQL is a deprecating term fashioned to jab at SQL but in actual, the term NoSQL stands for Not Only SQL. Putting forth the idea that both these technologies SQL and NoSQL can exist together in their own specific place. For the previous few years, NoSQL technology has been heard and seen in the news most likely because of the reason that as many of the Web 2.0 leaders have taken the NoSQL technology. Facebook, Twitter, Digg, Amazon, LinkedIn and Google all these companies use NoSQL in one way or another.

The main factors behind the adaptation of NoSQL includes flexibility of data, no rigid schema and scalability.

3.2 Questionnaire Development

As discussed earlier technology is changing extensively, and data is becoming more crucial to any organization. The accessing and manipulation of data is much more necessary than saving it to storage. Accessing and storing of data to storage obviously is time consuming. The RDBMS's due to their architecture must consider the data types, relations and other hidden processes involved in execution and storing the data to disk and the same when accessing the data from storage. This adds up the time which is very crucial for the applications to respond. The massive data growth requires additional storage on the fly. This is difficult to manage in RDBMS environment (Rehman et al., 2021; Suleman et al., 2021).

On the other hand, accessing and storing of heterogenous type of data in NoSQL environment is very fast. The flexibility of multitype data such as text, images, videos and documents are managed very efficiently. The schema free and no predefined architecture gives a strong advantage over RDBMS. Dealing with enormous growth of data is very easy in NoSQL environment as it gives flexible scalable architecture by adding more and more additional servers called shards in running environment.

Qualitative Study has been done on previous researches for the same topic. This has helped enormously to get the basic idea about the database vendors and organizations requirements and improvements which have been done to accomplish the

day-to-day challenges. This also has shed light on the technology enhancements in this specific field within last 2 decades.

3.3 Data Collection

A questionnaire survey has been conducted as part of quantitative method. To get the current situation in dealing with bigdata and analytics a questionnaire has been developed and spread in IT field. The targeted people were IT company CEO's, CTO's, Database Administrators which are dealing with day to day management challenges and the network professionals which are designing networks to deal with bigdata. The results are then analyzed by the respondents and have helped in getting the result and findings for this study. As the restriction of direct access to all the respondents the study focused on various methods to collect data which include professional groups, emails, printed copies to professionals in contact. The results are then analyzed and discussed in Chap. 4.

4 Results and Findings

As the world already has adopted the NoSQL Database, to analyze the levels and reasons of moving the software developing companies towards it a survey has been conducted. The survey has been conducted to analyze and judge the necessity of NoSQL with respect to business needs, Data management problems, flexibility or low-cost DBMS adaptation.

The survey has been conducted among database administrators, IT company CEO's, technology decision makers, software developers and IT experts. The survey has been conducted considering those people doing business in RDBMS and Big Data.

The main points of the survey are given below.

- About half of the more than 250 respondents pointed to the fact that they have worked on NoSQL projects in past couple of years. The companies which have secured large software projects more than 50% of their projects are to deal with Big Data for their clients. The reasons of adaption of NoSQL is described by them as:
- 49% referred to inflexible schemas to be the major reason for their migration to NoSQL technology from the relational database system. A prime reason for switching to NoSQL is the deficiency of scalability and high latency/low performance when dealing with Big Data.
- Overall, 40% were of the viewpoint that NoSQL is very essential and significant to their daily operations, and it is continuing to become more important.

- The management of Big Data in NoSQL is much easier than in RDBMS due to the pre-defined architecture and limitation. As the capturing of multi structural data is greeted in NoSQL. The type of data is high rank in considering the NoSQL database to be used (Figs. 9, 10).

As per the above survey results explains according to software professionals the major factors which take part in deciding about NoSQL and RDBMS are BigData, unstructured data and the management of the same. This means when the experts have to deal with massive amount of unstructured data, they are more likely to adopt NoSQL (Fig. 11).

- Another question regarding the performance of NoSQL using BigData shows the experts agreed on having better performance than RDBMS.

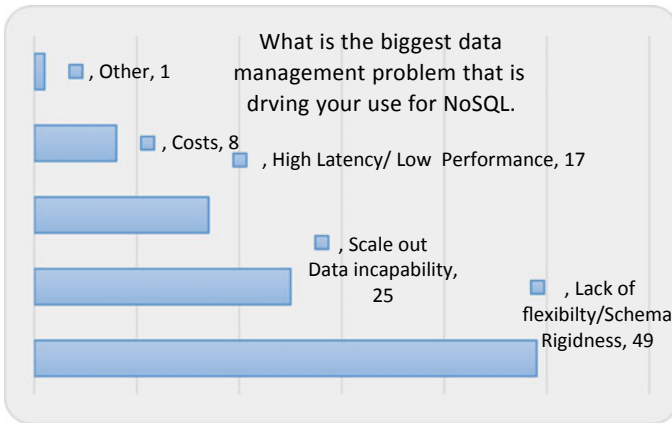


Fig. 9 Problems to driving towards NoSQL (Created by authors)

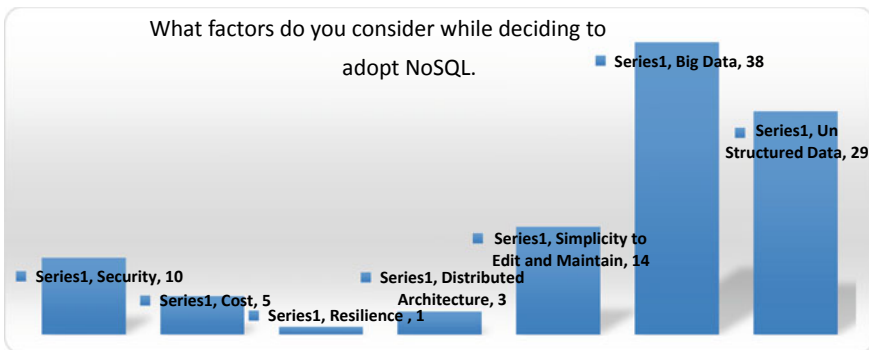


Fig. 10 Factor to decide NoSQL (Created by authors)

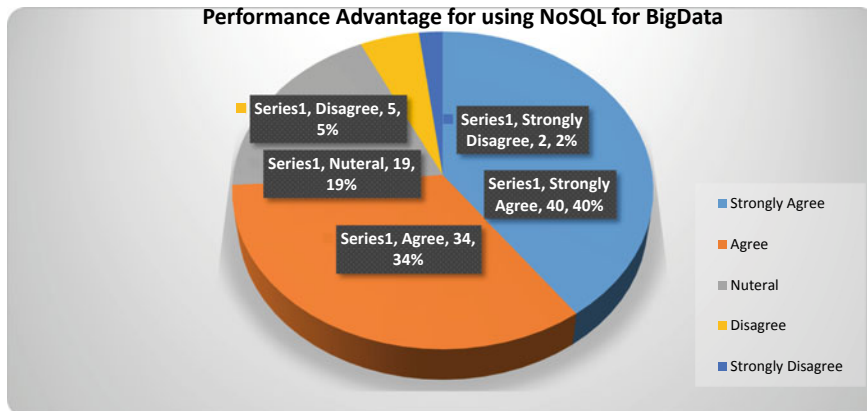


Fig. 11 NoSQL Performance with BigData (Author Created)

The architecture of DBMS and data has the main effect on performance. Performance is the major requirement of all bigdata based companies. At the same time fast data store and access is the major challenge for DBMS vendors and they keep doing research on this to get the customer’s confidence. In the survey 40% of the respondents strongly agree that NoSQL perform well when dealing with BigData along with 34% agree for the same.

Those organizations that show considerably high needs of storing data are considering NoSQL seriously and the demand of NoSQL database experts have also risen to a higher level in these developing organizations.

Overall, the results of the survey can be concluded as more than 65% of the respondent are agree that when it comes to deal with BigData they have used, or their first choice is to go with NoSQL. The reasons that have emerged from the survey answers are particularly highlighted as pre-defined schema in RDBMS. This has very strong characteristic and work very well in small to large databases which deals with structured and well-organized data. This kind of architecture is very popular and successful in OLTP environments like Banks, online retail shops and university library systems. The ERP systems also have pre-defined database architecture and are doing well with RDBMS. However, when it comes to very large to huge databases with the data volume from petabyte and greater, unstructured type of data including text, files, images and videos and all incoming very massively, it cannot be handled by RDBMS with high performance. The massive increase in data also require scalability at hardware level and the DBMS system should accept it on the fly. This has been the major characteristic of NoSQL DBMS that it dynamic schema and accept any kind of data. As the previous studies has shown social media is growing very widely in terms of data and dealing with all kinds of data, all the social media systems are using NoSQL in one way or another. So, the other top factors highlighted as scalability, performance when dealing with BigData, simple in maintaining the system.

5 Conclusion

Storage and processing requirements of some applications like Analytics for Big Data, Business Intelligence and social networking which is growing rapidly over peta-byte datasets have forced RDBMS to their limits. This has directed to the development of technology that is horizontally scalable, dispersed non-relational database named as No-SQL. The study speculate about the primarily usages of NoSQL Databases: The larger scale data processing system (parallel processing over distributed systems), (machine-to-machine data look-up & recovery); Analytics on semi-structured data (professional level); Huge capacity data storage (structured, semi-structured, unstructured) NoSQL is a huge and growing field, for the purposes of this study, characteristics (benefits and features of NoSQL DBMS); classification (the four categories with their features); and the comparison and assessment (with a table on basis of few characteristics- strategy, integrity, attributes, distribution) of different kinds of NoSQL databases. The study has also shown the difference between DBMS and NoSQL with present state and reason of acceptance of NoSQL databases. This study with motivation has provided an autonomous understating about the weakness and strengths of NoSQL databases which are supporting the applications that are dealing with large volume of data. The study has concluded the applications dealing with BigData performs well in NoSQL environment. Still the requirement varies from solution to solution. NoSQL will be emerging as the solution for BigData analytics in future.

5.1 Future Work

As the technology has been changing rapidly, the business is more relying on analytics. The analytics is based on data and the data is growing massively. The fast processing of this data is the major need of every organization. This is forcing technology leaders to put more efforts in making the DBMS more efficient in dealing with BigData. NoSQL has fulfilled the requirement to some level. However, there are still challenges that need more efforts from DBMS vendors to get confidence of customers. The main challenges are like data security and more compatible environments to most available development systems. Hope this study will help the users to understand the different DBMS architectures, BigData and its requirements and how to deal with this by understanding the nature. And will help users to guide to make decision while choosing DBMS according to their need and DBMS capabilities.

References

- Afifi, M. A. M., Kalra, D., Ghazal, T. M., & Mago, B. (2020). Information Technology Ethics and Professional Responsibilities. *International Journal of Advanced Science and Technology*, 29, 11336–11343.
- Akhtar, A., Akhtar, S., Bakhtawar, B., Kashif, A.A., Aziz, N., Javeid, M.S., (2021). COVID-19 Detection from CBC using machine learning techniques. *International Journal of Innovation and Technology Management* 1, 65–78. <https://doi.org/10.54489/ijtim.v1i2.22>
- Al-Khayyal, A., Alshurideh, M., Al Kurdi, B., Salloum, S.A., (2021). Factors influencing electronic service quality on electronic loyalty in online shopping context: data analysis approach, in: *Enabling AI Applications in Data Science*. Springer, pp. 367–378.
- Al Ali, A., (2021). The Impact of Information Sharing and Quality Assurance on Customer Service at UAE Banking Sector. *International Journal of Innovation and Technology Management* 1, 01–17. <https://doi.org/10.54489/ijtim.v1i1.10>
- Al Batayneh, R.M., Taleb, N., Said, R.A., Alshurideh, M.T., Ghazal, T.M., Alzoubi, H.M., (2021). IT Governance Framework and Smart Services Integration for Future Development of Dubai Infrastructure Utilizing AI and Big Data, Its Reflection on the Citizens Standard of Living, In: *The International Conference on Artificial Intelligence and Computer Vision*. pp. 235–247. Springer.
- Al Shebli, K., Said, R.A., Taleb, N., Ghazal, T.M., Alshurideh, M.T., Alzoubi, H.M., (2021). RTA's employees' perceptions toward the efficiency of artificial intelligence and big data utilization in providing smart services to the residents of Dubai, In: *The International Conference On Artificial Intelligence And Computer Vision*. pp. 573–585. Springer.
- AlHamad, A., Alshurideh, M., Alomari, K., Kurdi, B., Alzoubi, H., Hamouche, S., & Al-Hawary, S. (2022). The effect of electronic human resources management on organizational health of telecommuni-cations companies in Jordan. *International Journal of Data and Network Science*, 6, 429–438.
- Alhamad, A.Q.M., Akour, I., Alshurideh, M., Al-Hamad, A.Q., Kurdi, B.A., Alzoubi, H., (2021). Predicting the intention to use google glass: A comparative approach using machine learning models and PLS-SEM. *International Journal of Data and Network Science* 5. <https://doi.org/10.5267/j.ijdns.2021.6.002>
- Ali, N., Ahmed, A., Anum, L., Ghazal, T.M., Abbas, S., Khan, M.A., Alzoubi, H.M., Ahmad, M., (2021). Modelling supply chain information collaboration empowered with machine learning technique. *Intelligent Automation and Soft Computing* 30, 243–257. <https://doi.org/10.32604/iasc.2021.018983>
- Ali, N., M. Ghazal, T., Ahmed, A., Abbas, S., A. Khan, M., Alzoubi, H., Farooq, U., Ahmad, M., Adnan Khan, M., (2022). Fusion-Based supply chain collaboration using machine learning techniques. *Intelligent Automation and Soft Computing* 31, 1671–1687. <https://doi.org/10.32604/iasc.2022.019892>
- Alnazer, N. N., Alnuaimi, M. A., & Alzoubi, H. M. (2017). Analysing the appropriate cognitive styles and its effect on strategic innovation in Jordanian universities. *International Journal of Business Excellence*, 13, 127–140. <https://doi.org/10.1504/IJBEX.2017.085799>
- Alnuaimi, M., Alzoubi, H. M., Ajelat, D., & Alzoubi, A. A. (2021). Towards intelligent organisations: An empirical investigation of learning orientation's role in technical innovation. *International Journal of Innovation and Learning*, 29, 207–221. <https://doi.org/10.1504/IJIL.2021.112996>
- Alsharari, N. (2021). Integrating blockchain technology with internet of things to efficiency. *International Journal of Innovation and Technology Management*, 1, 1–13.
- Alshurideh, M. (2022). Does electronic customer relationship management (E-CRM) affect service quality at private hospitals in Jordan? *Uncertain Supply Chain Manag.*, 10, 1–8.
- Alshurideh, M., Al Kurdi, B., Alzoubi, H., Ghazal, T., Said, R., AlHamad, A., Hamadneh, S., Sahawneh, N., Al-kassem, A., (2022a). Fuzzy assisted human resource management for supply chain management issues. *Annals of Operations Research* 1–19.

- Alshurideh, M., Gasaymeh, A., Ahmed, G., Alzoubi, H., Kurd, B.A., (2020). Loyalty program effectiveness: Theoretical reviews and practical proofs. *Uncertain Supply Chain Management* 8. <https://doi.org/10.5267/j.uscm.2020.2.003>
- Alshurideh, M.T., Al Kurdi, B., Alzoubi, H.M., Ghazal, T.M., Said, R.A., AlHamad, A.Q., Hamadneh, S., Sahawneh, N., Al-kassem, A.H., (2022b). Fuzzy assisted human resource management for supply chain management issues. *Annals of Operations Research* 1–19.
- Alzoubi, Ali, 2021a. The Impact of Process Quality and Quality Control on Organizational Competitiveness at 5-star hotels in Dubai. *International Journal of Innovation and Technology Management* 1, 54–68. <https://doi.org/10.54489/ijtim.v1i1.14>
- Alzoubi, Asem, 2021b. Renewable Green hydrogen energy impact on sustainability performance. *International Journal of Computer Integrated Manufacturing* 1, 94–110. <https://doi.org/10.54489/ijcim.v1i1.46>
- Alzoubi, H., & Ahmed, G. (2019). Do TQM practices improve organisational success? A case study of electronics industry in the UAE. *International Journal of Economics and Business Research*, 17, 459–472. <https://doi.org/10.1504/IJEBR.2019.099975>
- Alzoubi, H., Ahmed, G., Al-Gasaymeh, A., & Kurdi, B. (2020a). Empirical study on sustainable supply chain strategies and its impact on competitive priorities: The mediating role of supply chain collaboration. *Management Science Letters*, 10, 703–708.
- Alzoubi, H., Alshurideh, M., Kurdi, B. A., & Inairat, M. (2020b). Do perceived service value, quality, price fairness and service recovery shape customer satisfaction and delight? A practical study in the service telecommunication context. *Uncertain Supply Chain Management*, 8, 579–588. <https://doi.org/10.5267/j.uscm.2020.2.005>
- Alzoubi, H., Alshurideh, M., Kurdi, B., Akour, I., & Aziz, R. (2022). Does BLE technology contribute towards improving marketing strategies, customers' satisfaction and loyalty? The role of open innovation. *International Journal of Data and Network Science*, 6, 449–460.
- Alzoubi, Haitham M. Alshurideh, M., Ghazal, T.M., (2021a). Integrating BLE beacon technology with intelligent information systems iis for operations' performance: a managerial perspective, In: *The International Conference on Artificial Intelligence and Computer Vision*. pp. 527–538.
- Alzoubi, H. M., & Aziz, R. (2021). Does emotional intelligence contribute to quality of strategic decisions? *The Mediating Role of Open Innovation*. <https://doi.org/10.3390/joitmc7020130>
- Alzoubi, Haitham M., Vij, M., Vij, A., Hanaysha, J.R., (2021b). What leads guests to satisfaction and loyalty in UAE five-star hotels? AHP analysis to service quality dimensions. *Enlightening Tour*. 11, 102–135. <https://doi.org/10.33776/et.v11i1.5056>
- Alzoubi, H. M., & Yanamandra, R. (2020). Investigating the mediating role of information sharing strategy on agile supply chain. *Uncertain Supply Chain Manag.*, 8, 273–284. <https://doi.org/10.5267/j.uscm.2019.12.004>
- Aziz, N., & Aftab, S. (2021). Data mining framework for nutrition ranking: methodology: SPSS modeller. *International Journal of Innovation and Technology Management*, 1, 85–95.
- Cruz, A. (2021). Convergence between blockchain and the internet of things. *International Journal of Innovation and Technology Management*, 1, 35–56.
- Eli, T. (2021). Students Perspectives on the Use of Innovative and Interactive Teaching Methods at the University of Nouakchott Al Aasriya, Mauritania: English Department as a Case Study. *International Journal of Innovation and Technology Management*, 1, 90–104.
- Farouk, M., 2021. The Universal Artificial Intelligence Efforts to Face Coronavirus COVID-19. *International Journal of Computer Integrated Manufacturing* 1, 77–93. <https://doi.org/10.54489/ijcim.v1i1.47>
- Ghazal, T., Soomro, T. R., & Shaalan, K. (2013). Integration of project management maturity (PMM) based on capability maturity model integration (CMMI). *European Journal of Scientific Research*, 99, 418–428.
- Ghazal, T.M., (2021). Positioning of UAV base stations using 5G and beyond networks for IoMT applications. *Arabian Journal for Science and Engineering* 1–12.

- Ghazal, T. M., Anam, M., Hasan, M. K., Hussain, M., Farooq, M. S., Ali, H. M., Ahmad, M., & Soomro, T. R. (2021a). Hep-Pred: Hepatitis C staging prediction using fine gaussian SVM. *Computers, Materials and Continua*, *69*, 191–203.
- Ghazal, T. M., Hasan, M. K., Alshurideh, M. T., Alzoubi, H. M., Ahmad, M., Akbar, S. S., Al Kurdi, B., & Akour, I. A. (2021b). IoT for smart cities: machine learning approaches in smart Healthcare—a review. *Futur. Internet*, *13*, 218. <https://doi.org/10.3390/fi13080218>
- Ghazal, Taher M, Hussain, M.Z., Said, R.A., Nadeem, A., Hasan, M.K., Ahmad, M., Khan, M.A., Naseem, M.T., (2021b). Performances of K-means clustering algorithm with different distance metrics.
- Ghazal, Taher M, Said, R.A., Taleb, N., (2021c). Internet of vehicles and autonomous systems with AI for medical things. *Soft Computing* 1–13.
- Guergov, S., Radwan, N., (2021). Blockchain convergence: analysis of issues affecting IoT, AI and Blockchain. *International Journal of Computer Integrated Manufacturing* 1, 1–17. <https://doi.org/10.54489/ijcim.v1i1.48>
- Hamadneh, S., Pedersen, O., & Al Kurdi, B. (2021). An investigation of the role of supply chain visibility into the scottish food supply chain. *Journal of Legal, Ethical and Regulatory Issues*, *24*, 1–12.
- Hanaysha, J.R., Al-Shaikh, M.E., Joghee, S., Alzoubi, H., (2021a). Impact of innovation capabilities on business sustainability in small and medium enterprises. *FIIB Business Review* 1–12. <https://doi.org/10.1177/231971452111042232>
- Hanaysha, J. R., Al Shaikh, M. E., & Alzoubi, H. M. (2021b). Importance of marketing mix elements in determining consumer purchase decision in the retail market. *International Journal Services Science Management Engineering Technology*, *12*, 56–72.
- Joghee, S., Alzoubi, H. M., & Dubey, A. R. (2020). Decisions effectiveness of FDI investment biases at real estate industry: Empirical evidence from Dubai smart city projects. *International Journal of Scientific & Technology Research*, *9*, 3499–3503.
- Kalra, D., Ghazal, T. M., & Afifi, M. A. M. (2020). Integration of collaboration systems in hospitality management as a comprehensive solution. *International Journal of Advanced Science and Technology*, *29*, 3155–3173.
- Kashif, A.A., Bakhtawar, B., Akhtar, A., Akhtar, S., Aziz, N., Javeid, M.S., (2021). Treatment response prediction in hepatitis c patients using machine learning techniques. *International Journal of Innovation and Technology Management*. 1, 79–89. <https://doi.org/10.54489/ijtim.v1i2.24>
- Khan, M.A., (2021). Challenges facing the application of IoT in medicine and healthcare. *International Journal of Computer Integrated Manufacturing*. 1, 39–55. <https://doi.org/10.54489/ijcim.v1i1.32>
- Khan, M.F., Ghazal, T.M., Said, R.A., Fatima, A., Abbas, S., Khan, M A, Issa, G.F., Ahmad, M., Khan, Muhammad Adnan, (2021a). An IoMT-Enabled smart healthcare model to monitor elderly people using machine learning technique. *Computational Intelligence and Neuroscience*.
- Khan, Q.-T.-A., Ghazal, T.M., Abbas, S., Khan, W.A., Khan, M.A., Said, R.A., Ahmad, M., Asif, M., (2021b). Modeling habit patterns using conditional reflexes in agency.
- Leavitt, N., (2010). Will NoSQL databases live up to their promise? *Computer* (Long. Beach. Calif). *43*, 12–14.
- Lee, C., Ahmed, G., (2021). Improving IoT privacy, data protection and security concerns. *International Journal of Innovation and Technology Management*. 1, 18–33. <https://doi.org/10.54489/ijtim.v1i1.12>
- Lee, K., Azmi, N., Hanaysha, J., Alzoubi, H., & Alshurideh, M. (2022a). The effect of digital supply chain on organizational performance: An empirical study in Malaysia manufacturing industry. *Uncertain Supply Chain Manag.*, *10*, 495–510.
- Lee, K., Romzi, P., Hanaysha, J., Alzoubi, H., & Alshurideh, M. (2022b). Investigating the impact of benefits and challenges of IOT adoption on supply chain performance and organizational performance: An empirical study in Malaysia. *Uncertain Supply Chain Management*, *10*, 537–550.

- Lee, S.-W., Hussain, S., Issa, G. F., Abbas, S., Ghazal, T. M., Sohail, T., Ahmad, M., & Khan, M. A. (2021). Multi-Dimensional trust quantification by artificial agents through evidential fuzzy multi-criteria decision making. *IEEE Access*, 9, 159399–159412.
- Matloob, F., Ghazal, T.M., Taleb, N., Aftab, S., Ahmad, M., Khan, M.A., Abbas, S., Soomro, T.R., (2021). Software defect prediction using ensemble learning: A systematic literature review. *IEEE Access*.
- Mehmood, T. (2021). Does information technology competencies and fleet management practices lead to effective service delivery? empirical evidence from e-commerce industry. *International Journal of Innovation and Technology Management*, 1, 14–41.
- Mehmood, T., Alzoubi, H.M., Ahmed, G., (2019). Schumpeterian entrepreneurship theory: evolution and relevance. *Academy of Entrepreneurship Journal*. 25.
- Miller, D., (2021). The best practice of teach computer science students to use paper prototyping. *International Journal of Innovation and Technology Management*. 1, 42–63. <https://doi.org/10.54489/ijtim.v1i2.17>
- Mondol, E.P., (2021). The impact of block chain and smart inventory system on supply chain performance at retail industry. *International journal of computer integrated manufacturing* 1, 56–76. <https://doi.org/10.54489/ijcim.v1i1.30>
- Naqvi, R., Soomro, T.R., Alzoubi, H.M., Ghazal, T.M., Alshurideh, M.T., (2021). The nexus between big data and decision-making: a study of big data techniques and technologies, In: *The International Conference on Artificial Intelligence and Computer Vision*. pp. 838–853.
- Obaid, A.J., (2021). Assessment of smart home assistants as an IoT. *International Journal of Computer Integrated Manufacturing*. 1, 18–36. <https://doi.org/10.54489/ijcim.v1i1.34>
- Pokorny, J. (2013). NoSQL databases: A step to database scalability in web environment. *International Journal of Web Information Systems.*, 9, 69–82.
- Radwan, N., Farouk, M., (2021). The growth of internet of things (IoT) in the management of healthcare issues and healthcare policy development. *International Journal of Innovation and Technology Management*. 1, 69–84. <https://doi.org/10.54489/ijtim.v1i1.8>
- Rehman, E., Khan, M. A., Soomro, T. R., Taleb, N., Afifi, M. A., & Ghazal, T. M. (2021). Using blockchain to ensure trust between donor agencies and ngos in under-developed countries. *Computers*, 10, 98.
- Shamout, M., Ben-Abdallah, B., Alshurideh, M., Alzoubi, H., Al Kurdi, B., & Hamadneh, S. (2022). A conceptual model for the adoption of autonomous robots in supply chain and logistics industry. *Uncertain Supply Chain Management.*, 10, 1–16.
- Suleman, M., Soomro, T.R., Ghazal, T.M., Alshurideh, M., (2021). Combating against potentially harmful mobile apps, In: the international conference on artificial intelligence and computer vision, pp. 154–173 Springer.