International Journal of Sciences: Basic and Applied Research (IJSBAR)

ISSN 2307-4531
(Print & Online)

https://gssrr.org/index.php/JournalOfBasicAndApplied/index

Security Implications of Adopting a New Data Storage and Access Model in Big Data and Cloud Computing

Mathias Philippe Totto Ndong [a*], Agnes Enangue [b], Essomba Mbondjo Richard[c], Samuel Batambock[d], Jean Nyatte Nyatte[e], Kikmo W C[f]

[a,b,c,d,e,f]*Energy, Materials, Modeling and Methods research laboratory (E3M)*
[a,b,c,d,e,f]*National Higher Polytechnic School of Douala, P.O.Box 2701 Douala-Cameroon*
[a]*Email: totto.philippe@gmail.com*

Abstract

This article examines the security implications of using cloud computing and Big Data. It employs a mixed methodology of qualitative and quantitative research and takes a critical realist epistemological approach. The objective is to identify the components of a theory for predicting and explaining [1, 4] the security implications associated with adopting the services provided by cloud computing and Big Data. The integration of various information sources and the widespread use of computing across diverse fields have resulted in a significant increase in data volume, scale, quantity, and diversity. Consequently, data management, storage, retrieval, and access have undergone significant changes. The latest developments in IT have brought forth novel technologies such as Cloud Computing and Big Data. Big Data comprises of technologies that rely on NoSQL (Not only SQL) databases, which enable the growth of data volumes, numbers, and types on a large scale. The new NoSQL systems are seen as solutions for meeting scalability requirements of large IT firms. Multiple open-source and pay-as-you-go NoSQL models are available for purchase.

Keywords: Cloud Computing; Big Data; NoSQL.

Received: 8/30/2023
Accepted: 10/5/2023
Published: 10/15/2023

----

* Corresponding author.

1. The context

The abundance of sources for generating digital data (Internet, computers, mobile phones, digital cameras, household appliances, sensors, and so on). The widespread use of information technology (IT) across various sectors and fields (including E-Commerce, E-Administration, E-Government, Astrology, Meteorology, MultiMedia, Digital Cinema and Television, Digital Archives, Digital Libraries, Social Networks, etc.) has caused a surge in data volume. This elevated level reflects the change in scale in terms of three dimensions: data volumes, data types, and data numbers. Several terms have been used in literature to describe this phenomenon, such as "data deluge", "data tsunami", "data explosion", and "data boom". It is challenging to estimate how much digital data businesses, public entities, and individuals produce daily, whether it is in the form of photographs, video clips, texts, tweets or emails. Since the 1990s, IT designs have depended upon data warehouses, which are typically centralized in servers connected to storage bays. These architectures have presented scalability challenges when it comes to adding power on demand. Therefore, traditional DBMSs and even data warehouses have struggled to adapt to the increasing volume of data and varying multi-structured data formats, such as video, images, text, and web data, and the frequency at which they are generated. Since the inception of computing, the location of data in relation to the computer processing it and the access methods have continuously evolved. The magnitude of the exact number is causing issues for storing and accessing data.

General issues

The definition of the information and communication technology needs of business units, dictated by CIOs, has lost ground in the face of increasing consumerism in IT. The de-intermediation of CIOs [2, 3]is materialising through their elimination for access to technology, especially hosted cloud tools. As a result, fragmented solutions proliferate in the different departments of the company, highlighting a problem of IS balkanisation that has already been studied with the large-scale adoption of ERP [5, 8].

The deployment of hosted services in big data and/or cloud computing mode at the business unit level has been motivated by the quest for the agility and flexibility required in a competitive environment. The potential difficulty in implementing clouds is the geographical dispersion, as data is outsourced to another country or even continent, resulting in data and metadata being transported across distant networks. While the difficulty of locating data is taken into account in the concept of clouds, the location of metadata is more difficult to determine. Metadata management therefore remains the main technical constraint on the scalability of clouds.

It is quite possible that in a few years' time, a global IT strategy that ignores cloud computing will be doomed to failure. Since 2013, however, a number of questions have been raised about the security of data hosted in the cloud. A former NSA consultant revealed the existence of an electronic surveillance programme (PRISM), including the collection of data hosted by US cloud service providers. Under the USA PATRIOT Act, the US security services have the right to access digital data held by individuals and companies without prior authorisation and without informing users.

These revelations were also in line with Richard Stallman's assessment of cloud computing. The founding father of the free software movement spoke out against cloud computing in 2008, stating that "It's as bad a choice as using a proprietary program... If you use a proprietary program or a third-party server, your data is no longer secure. Data confidentiality is at the centre of current debates. Considered to be the driving force behind the digital economy, data is a company's intangible capital. The loss of control over data can create risks with a high impact on the business. Research questions include How does the adoption of cloud computing affect the security of corporate data? How does cloud computing expose organisations to new risks?

How can IT governance be controlled when the CIO has a more managerial profile? In such circumstances, CIOs are challenged to rethink their technological and relational approaches in order to maintain control of the information system. According to a study published by Gartner in 2016, by 2020 all companies will have integrated cloud computing, in part or in full, into their technology strategies. IT departments must therefore implement an appropriate security policy to minimise the risks without limiting the flexibility of the service.

Our contribution

Our contribution is to provide a set of criteria and indicators that interested parties can use to make decisions about appropriate solutions for their businesses, by developing a comparative study of a range of NoSQL solutions widely available on the market. In a first approach, we will explore the world of Big Data and Cloud Computing and the new requirements imposed by the transition to new dimensions that have led to the concepts of Big Data and Cloud Computing, as well as the limitations of traditional systems. We will then present the new data management systems designed to meet the new requirements imposed by the transition to new dimensions, namely NoSQL and NewSQL.

1.1. Big data

The evolution of the architectures of the various data management systems that have led to big data, as well as the basic concepts of this latest generation, will be discussed in the following sections.

a. The evolution of data management systems

Every innovation in data management is seen as a new departure from what has gone before. However, whether revolutionary or incremental, most new stages and waves of data management evolution build on their predecessors, so that the tools, technologies and practices used to solve new and different problems are not overturned. The evolution of data management systems must include technological developments in hardware, storage, networking and new computing models such as virtualisation and cloud computing. The combination of new technologies and the reduction of various costs, from storage to processing, has revolutionised the way data is managed and used, opening up new possibilities. Each birth of a new data management model is justified by the need to solve a specific type of data management problem. Each of these generations or phases has evolved as a result of cause and effect. When a new technological solution is introduced to the market, new approaches are discovered. When databases When relational data systems hit the market, we had to develop a set of tools that allowed managers to explore the associations and relationships between data elements. As companies began to store unstructured data, it became apparent that the various servers were being overwhelmed by the immense amounts of data being hosted, so to take advantage of this data, new tools and innovative approaches were required, as well as the need to develop new functionalities and analysis tools to obtain information that would be useful to the business. The different generations of data management technologies over the last few decades have ushered in a new episode: the era of big data. This latest trend is not an independent technology, but rather an accumulation of the last 50 years of technological evolution.

b. Big Data Management

Is big data really new, or is it an evolution of data management systems? The answer is yes - in fact, it's both. Like other generations, Big Data is the result of the evolution of data management practices over the last five decades. What is new is that, for the first time, the cost of processing and storage has reached a level of reduction that is considered a tipping point. It's important to remember that just a few years ago, organisations typically compromised by storing snapshots or subsets of important information because storage costs and processing limitations didn't allow them to store everything they wanted to analyse. It is now possible to virtualise data so that it can be stored efficiently and to use more cost-effective cloud storage. In addition, improvements in network speed and reliability have removed other physical limitations to managing massive amounts of data at an acceptable speed, and there is also the impact of economic and technological changes in storage.

b.1. Big Data definition

Big Data literally means massive data or mega data. It is a set of heterogeneous data entities that are permanently scalable and cannot be handled by traditional data management systems. Big Data is also a distributed and scalable architecture for processing and storing large volumes of data. In fact, approximately 2.5 billion gigabytes of data are generated every day from the various domains created by different digital tools: videos posted, messages sent, GPS signals, transaction records from online purchases, and many more. These massive amounts of data are known as big data. Web giants such as Yahoo, Facebook, Amazon and Google were the first to use this type of technology to give everyone real-time access to their huge databases. The emergence of big data is seen as a new industrial revolution, similar to the discovery of steam, electricity, the telephone and computers. Others describe this phenomenon as the latest episode of the third industrial revolution, known as the 'information revolution'. However, there is no universal or precise definition of big data. As a polymorphous and complex concept, its interpretation varies according to the communities interested in it as service providers or users. Big Data is also defined in terms of the way in which large volumes of data can be optimally processed and exploited. The emergence of big data is seen as a new industrial revolution, similar to the discovery of steam, electricity, the telephone and computers. Others describe it as the latest episode of the third industrial revolution, known as the 'information revolution'.

b.2. Data Collection and Loading Processes

A layer responsible for data loading in a big data architecture is capable of managing large volumes of heterogeneous data, where the data needs to be validated, cleansed, transformed, reduced (compressed) and integrated into the big data stack for processing. It collects the relevant information and loads it into the big data storage layer [11].

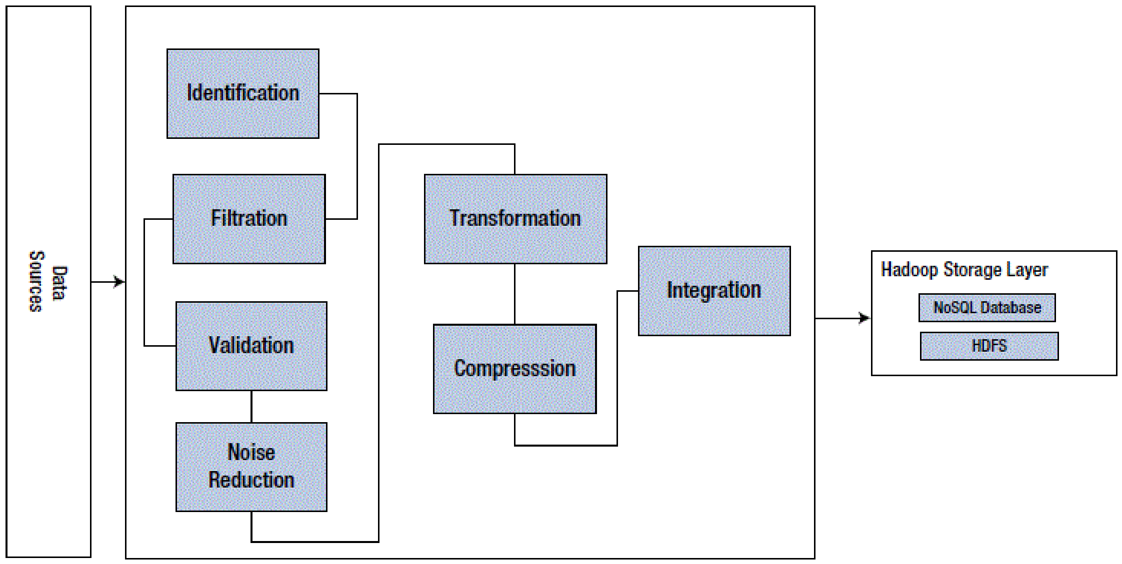

Figure 1: Big data loading process [11].

This layer must include the following components

- Identify the various data formats.

- Filtering and selection of relevant incoming information.

- Ongoing validation and analysis of the data.

- Noise reduction, which involves cleaning up the data.

- Transformation may involve splitting, converging, standardising or synthesising data.

- Compression involves reducing the size of the data without loss of relevance.

- Integration involves integrating all the collected data into the big data storage layer (usually a NoSQL database).

b.3. Key players

The players are the organisations that have designed, developed, deployed and used big data. They are mainly the founders of this new concept and the "creators" of big data technologies. The most important innovations in this field have come from the leaders of the Web [10]: Google (MapReduce and BigTable), Amazon (Dynamo, S3), Yahoo! (Hadoop, S4), Facebook (Cassandra, Hive), Twitter (Storm, FlockDB), LinkedIn (Kafka, SenseiDB, Voldemort), LiveJournal (Memcached) and others. The Apache Foundation has been particularly active in this area, launching or supporting more than a dozen projects, both mature and in incubation: Hadoop, Lucene/Solr, Hbase, Hive, Pig, Cassandra, Mahout, Zookeeper, S4, Storm, Kafka, Flume, Hama, Giraph, etc. In addition to web companies, the scientific sector and, more recently, the promoters of Open Data and its variant, Open Linked Data, derived from the Semantic Web, have also made significant contributions to the field of Big Data [10].

c. Data Storage and Virtualisation

The difficulties associated with managing and storing data in data warehouses, due to its volume, type and form, hinder its proper use. The new architectures developed to respond favourably to these new needs have emerged as NoSQL databases. NoSQL databases aim to scale horizontally by relaxing some or all of the strict transactional conditions (ACID) considered fundamental to traditional systems, and by abandoning the relational model. Virtualisation is used to better control the use and performance of IT resources, particularly storage. This is a fundamental technology used in both big data and cloud computing. Data virtualisation can be used to create a platform for dynamically linked data services. This allows data to be easily searched and linked through a unified reference source. As a result, data virtualisation provides an abstract service that presents data in a consistent form regardless of the underlying physical database. In addition, data virtualisation exposes cached data to all applications to improve performance. Storage virtualisation combines physical storage resources to share them more efficiently, reducing storage costs and making it easier to manage the data required for analysis. Data and storage virtualisation plays an important role in facilitating and reducing the cost of storing, searching and analysing data, both structured and especially unstructured, hosted in big data environments. With virtualisation, the database can be stored as a virtual image and accessed when needed without consuming more data centre resources [8].

1.2 Cloud Computing

The impact of cloud computing on the process of outsourcing information systems (IS) is raising increasingly complex questions for both practitioners and theorists of the digital economy. Developed from the hybridisation of several massive information processing technologies, cloud computing has become a paradigm for organisational transformation. The cloud has been in widespread use since the beginning of the new millennium, and although it is a relatively recent development, its history is directly linked to the development of the Internet. This technology allows users to reduce their software and hardware operating costs directly online, and to gain in performance and scalability. Cloud computing is currently developing in a very favourable context in the business world, with the majority of large companies using or planning to use cloud services in the next few years.

a. Cloud Definition

The National Institute of Standards and Technology (NIST) defines Cloud Computing as a model for delivering a shared set of computing resources, including servers, storage, applications, processing, and delivery platforms, on-demand and over a network that can be rapidly provisioned with minimal management and interaction with the service provider. Cloud computing makes a hardware and software infrastructure dynamic and flexible by exposing the capabilities of data centres as a "network of virtual services". In this infrastructure, users can access and deploy applications from the Internet according to their needs and the quality of service they require [12]. The cloud is the future of software and hardware infrastructures.

This trend is justified by the increasingly low costs involved (measurable and billable service: a computer rented in the cloud costs less than €5/month), Attractive offers and improved connection speeds. A very large number of organisations have already migrated some of their services to the cloud or plan to do so in the near future. The range of cloud services on offer continues to grow at all levels, from specific applications, embedded applications, CRM, ERP, web services and libraries to complete infrastructures. The cloud computing market is expected to exceed $241 billion by 2020. Cloud computing has three service models and four deployment models, which are briefly described below.

b. Service models

The cloud comes in the form of 3 offerings: SaaS, PaaS and IaaS (Figure 2) [12].

Figure 2: The 3 cloud computing models.

b.1. Deployment models

We distinguish between two main forms of cloud computing deployment, depending on the network in which the services are available: Public Cloud and Private Cloud, as well as two other derived forms, namely Community Cloud and Hybrid Cloud [13].

b.1.1 Public Cloud

The public cloud is a collection of hardware, networking, storage, services, applications and interfaces owned and operated by other companies and individuals. These external providers have highly scalable data centres that hide the details of the infrastructure from the customer. The services provided are available over the Internet to anyone who pays for them. The principle is to host applications, usually web applications, in a shared environment with an unlimited number of users. The services offered by public clouds are not all the same. Some are highly scalable, with high levels of security and better quality of service. Others are less robust and less secure, but much cheaper to use. Your choice will depend on the nature and importance of your data and the level of risk you are prepared to take. The best-known public cloud providers are Google, Microsoft and Amazon.

b.1.2. Private Cloud

The private cloud is a collection of hardware, networking, storage, services, applications and interfaces owned and operated by an organisation for use by its employees, partners and customers. It can be built and managed by a third party for an organisation's exclusive use. It is an environment deployed within a company or organisation that owns it, using its internal infrastructure. The resources are owned and controlled by the company's IT department, or the services are accessed via the private network. The private cloud is a highly controlled environment behind a firewall that is not available to the public. If an organisation is managing a big data project that requires the processing of massive amounts of data, the private cloud may be the best choice in terms of latency and security. Today, this model corresponds to the evolution of virtualised data centres and the emergence of IT as a service: the information system and IT equipment are transformed into a service centre for the rest of the business.

b.1.3. Community Cloud

The Community Cloud is an extension of the Private Cloud. In this model, resources, services and ownership are shared at the level of a community (e.g. state, city, academy, etc.). Services are accessed through the interconnection of networks belonging to the organisations in that community.

b.1.4. Hybrid Cloud

A hybrid cloud is a combination of a private cloud and the use of public cloud services. Some services are accessible via a private network and others via the Internet. The future should confirm the emergence of the hybrid cloud, which consists of the co-existence and communication between a private cloud and a public cloud in an organisation, sharing data and applications.

c. Characteristics and benefits

A cloud can be characterised by a number of features, the most important of which are listed below [12]:

- Universal network access over the network

A cloud environment is necessarily based on the Internet and is accessible via this network, regardless of the device being used (PC, Mac, TV, tablet and smartphone).

- Pooling of resources

In a cloud-like environment, we don't think in terms of the number of servers, the size of disks or the number of processors, but rather in terms of computing power, storage capacity, etc.

- Elasticity

In a cloud environment, resources can be scaled up very quickly to meet high demand (e.g. to meet the large number of customer purchase requests on an e-commerce web platform during the festive season) and available bandwidth.

- Self-service

In a cloud, a user can consume services or resources without having to make a request to the provider (IT team).

- Measurable and chargeable service (Pay As You Go)

In a cloud environment, the solution provider is able to accurately measure the consumption of various resources (CPU, storage, bandwidth, etc.). This measurement is then used to bill the customer according to usage. Pay As You Go is a typical billing option for a cloud computing provider, meaning that the customer is billed according to the resources used.

- Fault Tolerance

Cloud architectures must be fault tolerant, providing continuous service despite the failure of one or more system components.

- Dynamic workplace

The cloud supports the organisation's new vision of the employee's workplace and their physical presence on company premises. So you don't have to work on company premises, you can work anywhere? The key is to be cost effective and efficient. For an organisation, using the cloud, regardless of the service or delivery model, offers a range of benefits. However, there can be disadvantages, particularly with non-private clouds, the most worrying of which are:

- The problem of bandwidth reliability to the provider.

- The issue of security and confidentiality.

2. Cloud computing and big data

The power of the cloud comes from the fact that users can access the IT resources they need with a minimum of IT support, rather than having to invest in hardware or software. One of its key features is horizontal elasticity or scalability: users can add or remove resources in near real time as their needs change. The cloud plays an important role in the vast world of big data. Spectacular improvements occur when these infrastructures are combined with developments in data management. An infrastructure optimised for horizontal scalability encourages and supports the practical implementation of big data. Two popular examples of the benefits of cloud hosting for big data are Google and Amazon.com. Both companies rely on their ability to manage massive amounts of data to drive their businesses. They have developed infrastructures and technologies to support applications at scale. Think of Gmail or the millions of messages that Google processes every day as part of this service. Google has been able to optimise the Linux operating system and its software environment to support email in the most efficient way; as a result, it can easily support hundreds of millions of users. More importantly, Google is able to collect and use large amounts of data about its email and search engine users to run the business and improve its services. Similarly, Amazon.com optimises its IaaS data centres to support its workloads so that it can continue to offer new services and support a growing number of customers. To develop its retail business, Amazon needs to be able to manage data about its products, its customers and its network of partner merchants. Targeted advertising based on customers' buying habits is essential to the company's success. These companies now offer a range of cloud services for their big data [8].

2.1 Databases in Cloud Computing

All the different cloud computing offerings are playing a major role in the strong demand from companies to migrate their various services to the cloud. Database services are not excluded from this trend. A cloud database is a type of database service that is developed, deployed and delivered via a cloud platform. This enables organisations, end users and their applications to store, manage and retrieve data from the typically private cloud [13]. There are currently several offerings and deployment methods for databases in the cloud, although there are certain limitations that can slow the transition of databases to the cloud. Table 1 below gives some examples of relational databases being used in the cloud, deployed in virtual machines or as a service.

Table 1: Examples of relational databases delivered as a service in the cloud.

Deployment on virtual machine	Database deployment as a service
Oracle Database	Amazon Relational Database service
IBM DB2	Bitcan (MYSQL, MongoDB)
Ingres (Database)	Microsoft SQL Azure (MS, SQL)
PostgreSQL	Heroku PostgreSQL as a service(shared and Dedicated Database Options)
MYSQL	Clustrix Database as a service
NuoDB	Enterprise DB Postgres Plus Cloud Database

2.1.1 Pros and cons of databases in the cloud

There is no doubt that deploying databases in cloud computing environments can bring greater benefits to users, but there are a number of constraints and risks that need to be considered before deciding on such a migration:

- Benefits

- Cost savings

Trading the high costs of installing, maintaining and upgrading an on-premises IT infrastructure for the operating costs of a cloud database subscription makes good business sense, especially in the short to medium term. However, it is important to be aware of the potential hidden costs associated with this strategy.

- Elasticity

As a business grows and the number of users increases, rather than investing in software licences and additional in-house server capacity, some service providers allow the monthly subscription to be scaled to the needs of the business.

- Accessibility

In general, all you need to access a cloud database is a browser and a good internet connection. These solutions are available on a wide range of desktops and mobile devices.

- Security

This is an advantage in the sense that security is better handled in cloud platforms for companies that do not have the resources and expertise to do so.

- Resilience

Because the IT infrastructure and data reside in the cloud service provider's data centre, if the company's premises suffer a disaster of any kind, it is relatively easy to get the company's services up and running again from another location equipped with computers connected to the Internet.

- Disadvantages

- Security

This is the number one concern for businesses. The downside of security lies in the nature of hosting. The cloud platform, if it is external (not installed on the internal network or with an opening to the outside world), must be sufficiently secure to avoid the risk of intrusion or theft of data through piracy. More generally, a confidentiality clause and trust in its staff are essential to ensure that data is not deliberately leaked.

- The connection

This is the bottleneck problem that can reduce performance. If the Internet connection is poor (low speed, frequent interruptions), the user will not be able to access his work platform. In this case, the idea is to continue working on a local application that will synchronise with the server as soon as the user has access to the network again.

- Service interruption

One of the goals of cloud providers is to plan well for the inevitable service interruptions, whether caused by natural disaster, human error or the many causes in between. Downtime is always annoying, but a prolonged outage can be disastrous if it affects a critical application. Customers should carefully examine their service provider's SLAs and historical performance before outsourcing business-critical applications to a public cloud platform. Tools such as Compuware's Outage Analyzer and Is It Down Right Now provide continuous monitoring of service interruptions in the cloud.

- Cost

Without a cost-benefit analysis, cost can be a long-term disadvantage.

- Regulatory compliance

If company data is stored in a service provider's data centre, it can be difficult to ensure compliance with relevant government data protection regulations.

- Performance issues

An application hosted in a remote data centre and accessed via an Internet connection can raise performance concerns compared to software running on a local machine or a viable corporate LAN. Of course, some tasks are better suited to the SaaS model than others. Application performance management tools can help customers and suppliers monitor the performance of their applications.

- Data mobility

The SaaS market is full of start-ups, some of which will inevitably fail. What happens to the data if the service provider goes out of business, or if the customer needs to switch SaaS providers for one reason or another? When a customer chooses a SaaS provider, they need to ensure that they avoid any pitfalls by preparing an exit strategy.

- Software Integration

Enterprises adopting multiple SaaS applications, or wanting to connect hosted software to existing on-premises applications, face the problem of software integration. If it is not possible to manage the relevant libraries, APIs and data structures in-house, there is a relatively new class of integration-as-a-service products, including Boomi (a Dell company), CloudSwitch and Informatica.

2.2 Different NoSQL models

There are a variety of technologies and solutions, all labelled as NoSQL, that differ in the way they represent data. These different NoSQL systems use different approaches that fall into four categories [3, 4]:

-Key-value oriented

-Columnar

-Document-oriented

-Graph-oriented

Several NoSQL architectures have been implemented by the major players in the field, such as Google, Yahoo, Facebook, Twitter, Amazon, etc., to host these large volumes of data on their servers. There are currently more than 305 NoSQL solutions on the IT market, ranging from :

-Key/value databases such as: Redis, Memcached, Voldemort, Riak, Hazelcast, Ehcache, Oracle NoSQL;

-Document databases such as: MongoDB, Amazon DynamoDB, Couchbase, CouchDB, Microsoft Azure DocumentDB, Cloudant;

-Columnar databases such as: Cassandra, HBase, Bigtable, Accumulo, Hypertable;

-Graph-oriented databases such as: Neo4j, Titan, Giraph, InfiniteGraph;

-Multi-model databases such as OrientDB, MarkLogic, ArangoDB, Virtuoso;

Recent studies show that the document-oriented NoSQL type is still the most popular model, as shown in Figure3[16].This is probably due to its ease of use(JSON type).On the other hand, column-oriented systems are neck-and-neck with key-value systems.Graph-oriented systems come last,as they are mainly reserved for very specific applications.

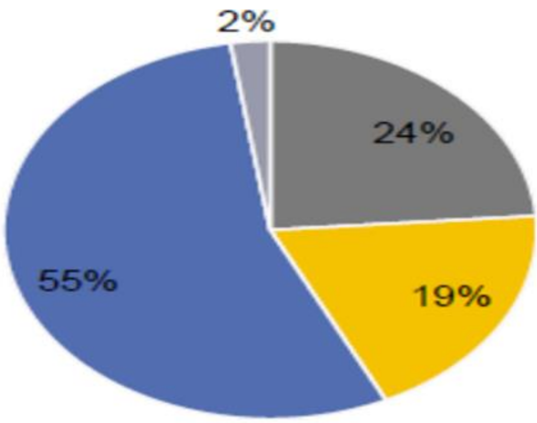

Figure 3: Most commonly used NoSQL types [16].

- (MongoDB, CouchDB…)
- Wide Column Store

  (Cassandra, HBase…)

- Key-Value Store

  (Redis, Memcached…)

- Graph stores

  (Neo4j, Titan…)

For the month of May 2022, the top 19 rankings are shown in Table 3 below:

DB-Engines Ranking

The DB-Engines Ranking ranks database management systems according to their popularity. The ranking is updated monthly.

Read more about the method of calculating the scores.

394 systems in ranking, May 2022

Rank May 2022	Rank Apr 2022	Rank May 2021	DBMS	Database Model	Score May 2022	Score Apr 2022	Score May 2021
1.	1.	1.	Oracle	Relational, Multi-model	1262.82	+8.00	-7.12
2.	2.	2.	MySQL	Relational, Multi-model	1202.10	-2.06	-34.28
3.	3.	3.	Microsoft SQL Server	Relational, Multi-model	941.20	+2.74	-51.46
4.	4.	4.	PostgreSQL	Relational, Multi-model	615.29	+0.83	+56.04
5.	5.	5.	MongoDB	Document, Multi-model	478.24	-5.14	-2.78
6.	6.	↑ 7.	Redis	Key-value, Multi-model	179.02	+1.41	+16.85
7.	↑ 8.	↓ 6.	IBM Db2	Relational, Multi-model	160.32	-0.13	-6.34
8.	↓ 7.	8.	Elasticsearch	Search engine, Multi-model	157.69	-3.14	+2.34
9.	9.	↑ 10.	Microsoft Access	Relational	143.44	+0.66	+28.04
10.	10.	↓ 9.	SQLite	Relational	134.73	+1.94	+8.04
11.	11.	11.	Cassandra	Wide column	118.01	-3.98	+7.08
12.	12.	12.	MariaDB	Relational, Multi-model	111.13	+0.81	+14.44
13.	13.	13.	Splunk	Search engine	96.35	+1.11	+4.24
14.	14.	↑ 27.	Snowflake	Relational	93.51	+4.06	+63.46
15.	15.	15.	Microsoft Azure SQL Database	Relational, Multi-model	85.33	-0.45	+14.88
16.	16.	16.	Amazon DynamoDB	Multi-model	84.46	+1.55	+14.39
17.	17.	↓ 14.	Hive	Relational	81.61	+0.18	+5.42
18.	18.	↓ 17.	Teradata	Relational, Multi-model	68.39	+0.82	-1.59
19.	19.	19.	Neo4j	Graph	60.14	+0.62	+7.91

Table 3: Top 19 DBMS [18].

In the next section, we will look at the actual description of NoSQL databases, focusing on very popular solutions according to a ranking by Solid IT [12]. This DB Engines ranking is an up-to-date list of 343 database management systems, ranked according to their popularity. They measure the popularity of a system using the following parameters:

1. Number of citations of the system on web sites;

2. General interest in the system;

3. Frequency of technical discussions about the system;

4. Number of job advertisements mentioning the system;

5. Number of profiles in professional networks mentioning the system;

6. Relevance in social networks.

3. Conclusion

With the dominance of the internet across most sectors, including industry, media, communications and even homes, the technologies of Big Data and Cloud Computing are experiencing growth. Over the next few years, we anticipate that these technologies will be increasingly employed to confront new data management issues. It is reasonable to expect NoSQL systems to respond positively to scale changes and to be perceived as the most fitting solutions to manage the vast quantities of data that characterise Big Data. Many users of conventional relational database management systems, commonly referred to as "SQL", aim to transition to these contemporary "NoSQL" alternatives in preparation for the anticipated surge in their data size in the years to come. Several NoSQL models with distinct architectures are being designed, developed and implemented across diverse industries. The absence of standardization is a critical aspect of the NoSQL movement. The extensive variety of solutions available in the market presents a dilemma to decision-makers when selecting the ideal model for their operating environment. The absence of standardisation is a significant feature of the NoSQL trend, and with the abundance of existing options in the IT industry, decision-makers who have opted for NoSQL are confronted with substantial issues in selecting the appropriate model for their operational surroundings. Consequently, CIOs are required to reconsider their technological and relationship approaches to exercise authority over the information system. IT departments must therefore implement an appropriate security policy that minimises risk without limiting service flexibility. A security policy is typically based on an information systems governance framework. Information systems governance, as defined by CobiT, "includes the command and operational structures and processes that drive enterprise IT to support and extend the strategies and objectives of the business" [9, 12]. IS governance aims to reduce IT-related risks and offers a comprehensive approach to information security management that goes beyond operational matters [3, 7]. While IS governance is especially relevant in the context of cloud computing, the CIO - a critical figure in implementing IS governance - is increasingly marginalized. Their role as IT solution integrators is being questioned, and their position in the company is being weakened. Several researchers have observed significant changes in their abilities [6, 9, 11]. The matter of security in cloud computing extends beyond the operational stage of the service. It is also relevant during the termination of the contract with the cloud provider when the reversibility of the service is in question. Although the topic currently garners limited attention from researchers, the reversibility of cloud computing is a significant security concern for enterprises. This prompts the question: how can the reversibility of a cloud computing service be defined?

References

[1] Adler P. S., Ferdows K. (1990), “The chief technology officer”, California Management Review, vol. 32, n°3, p. 55-62.

[2] Armbrust M., Stoica I., Zaharia M., Fox A., Griffith R., Joseph A. D., Katz R., Konwinski A., Lee G., Patterson D., Rabkin A. (2010), “A view of cloud computing”, Communications of the ACM, vol. 53, p. 50-58. doi: 10.1145/1721654.1721672.

[3] Becker J., Bailey, E. (2014), “A Comparison of IT Governance & Control Frameworks in Cloud Computing”, In: Proceedings of Americas Conference on Information Systems, Savannah, GA.

[4] ernard J.-G., Rivard S., Aubert B. A. (2004), “L’exposition au risque d’implantation d’ERP : éléments de mesure et d’atténuation”, Systèmes d’Information et Management, vol. 9, p. 25-50. doi : 10.9876/sim.v9i2.155. [5] D. Monthary and M. Platzer , building technology: All trades. Editions le Moniteur, ISBN 978-2-281-1144 6-1, (2006) 105 P.

[6] Bouaynaya, W., Bidan, M. (2017), « Une exploration qualitative du rôle des opérateurs du Cloud Computing dans l’acheminement des données des PME », Management & Avenir, vol. 93, p. 65-83.

[7] Burgelman, R. A. (1996), « A process model of strategic business exit: Implications for an evolutionary perspective on strategy », Strategic management journal, vol. 17, n°1, 193-214.

[8] Buyya R., Ranjan R., Calheiros R. (2010), « Intercloud: Utility-oriented federation of cloud computing environments for scaling of application services », Algorithms and architectures for parallel processing, p. 13-31.

[9] Objelean, A (2015). Introduction to Couchbase-NoSQL Document Database. http://alexo.github.io/articles/Introduction_to_couchbase_no_sql_document_database/. (Consulté en Mai 2023). [10] iang, Y (2012). HBase Administration Cookbook. Packt Publishing Open source community experience distilled. Isbn : 1849517142 9781849517140.

[11] Journal du Net (2013). Comparatif des bases de données NoSQL. https://www.journaldunet.com/developpeur/outils/comparatif-des-bases-nosql/. (Consulté en juin 2023)

[12] Hashem, H (2016). Modélisation intégratrice du traitement BigData. Thèse de Doctorat (Université Paris Saclay).

[13] Hadrien, F (2015). SQL, NoSQL, NewSQL stratégie de choix, Bachelor HES (Haute École de Gestion de Genève).

[14] Hecht, R, Jablonski, S (2011). NoSQL evaluation: A use case-oriented survey. In Proceedings of the International Conference on Cloud and Service Computing, IEEE. November 22-24 2012, Huashan

Road, Shanghai, China pp. 336-341.

[15] hitehorn, M (2015). Quand faut-il envisager d'utiliser une base de données NoSQL (plutôt qu'une base relationnelle) ? Le Grand Guide des Bases de Données (part. 2) : au-delà du relationnel (University de Dundee). [16] Fimbel, E. (2003), « Les facteurs décisionnels de l'externalisation des systèmes d'information: référentiels théoriques, éléments empiriques et proposition typologique », Systèmes d'information et management, vol. 8, n°4, Pp 31.

[17] Fréry F., & Law-kheng F. (2007), « La réinternalisation, chaînon manquant des théories de la firme », Revue française de gestion, vol. 8, p. 163-179.

[18] Oracle Real Application Clusters (RAC), (2013). An Oracle White Paper.

[19] Kouedi, E (2012). Approche de migration d'une base de données relationnelle vers une base de données NoSQL orientée colonne, Mémoire de Master (Université de YaoundeI).

[20] Lasalle, MF (2006). Cours du Relationnel à l'objet : Limites du Relationnel, http://circe.univ-fcomte.fr/Marie-France-Lasalle/bda/COURSHTM/cours/chap01/lec02/ page01.htm.

[21] Maletras, X (2012). Le NoSQL - Cassandra, Thèse Professionnelle (Université Paris 13).