# (Allaymoun et al., 2022)

Every company’s success is built on the basis of data. Consider the amount of knowledge and data that is now available as a result of technological advancements and the internet. As storage capacity and data collection technology have grown, massive amounts of data have been readily available. Every second, new data is generated, which must be saved and processed in order to extract value. Furthermore, since data storage costs have decreased, businesses must extract as much value as possible from vast amounts of data. A new type of big data analytics, as well as unique storage and analysis approaches, are necessary due to the quantity, diversity, and rapid change of such data. Massive amounts of large data must be scrutinized carefully, and essential information recovered.

Big data is a term that refers to large data sets with complex structures that are difficult to store, analyze, and visualize for subsequent processes or consequences. The analysis of enormous amounts of data to identify hidden patterns and links is known as big data analytics. These critical pieces of information for businesses and organizations can help them gain richer, deeper insights and gain a competitive advantage. As a result, large data installations must be carefully considered and implemented. Big data research is at the Centre of today’s research and industry. These data are derived from online transactions, emails, videos, audio, images, click streams, logs, postings, search queries, health records, social networking interactions, science data, sensors, and mobile phones and associated apps.

Big Data Analysis Has a Lot of Benefits Two important advantages of high-dimensional data analysis are the establishment of efficient techniques that may effectively estimate future observations as well as acquiring insight into the link between characteristics and reactions for scientific aims. Because of the large sample size, big data also leads to two additional goals: understanding heterogeneity and commonality across distinct subpopulations. To put it another way, big data promises to (1) explore the hidden structures of each data subpopulation, which has previously been impossible to do and can even be treated as “outliers” when the sample size is small; and (ii) extract important common traits across many subpopulations, despite significant individual differences.

For both academia and industry, big data analytics has become a very broad topic of research. As it aims to extract knowledge, information, and insight from vast amounts of data, big data analytics has recently captivated the interest of academics and industry alike. Two of the most significant advances characterizing the new analytical tools that are developing are big data and cloud computing. Many sectors may find it easier to adopt big data analytical skills offered via cloud delivery models, and, more crucially, it may simplify critical insights that might give them a competitive advantage. Many companies offer online Big Data analysis tools, the most well-known of which is Amazon’s Big Data Analytics Platform. SAP Big Data Analytics, IBM InfoSphere Big Insights, TERADATA Big Data Analytics, Big Data Platform, Cloudera Big Data Solution, and other big data analytics solutions are available. These companies use a range of technologies to analyze massive volumes of data and provide a simple or easy user experience.

There has been a lot of study done on big data analysis to produce graphical results and reports that can assist decision-makers. A big data service is any data-based resource that is made available through the Internet. The performance of a big data service is determined by the data collected by data collectors. On the other hand, the problem of efficient pricing and data distribution in big data services has gotten less attention. We propose an auction-based big data market model to address this. The impact of data size on big data analytics performance, such as machine learning algorithms, is used to calculate data cost and utility. Big data services are categorized as digital products, and they differ from traditional goods in that they have an “infinite supply,” whereas traditional things have a finite supply. As a result, provide a Bayesian profit maximization auction that is honest, rational, and computationally efficient. The best service price and data size are obtained by solving the profit maximization auction. Finally, results from trials utilizing a real-world taxi trip dataset show that our big data market model and auction mechanism effectively address the profit maximization challenge for service providers.

Make use of simple graphics or statistics. When large amounts of disparate data are present, fundamental analysis is typically used, according to the text. Advanced analytics is a set of techniques that may be applied to both structured and unstructured data to perform complex data analysis. The techniques discussed include sophisticated statistical models, machine learning, neural networks, text analytics, and other advanced data analysis approaches. Among other things, advanced analytics can be used to find patterns in data, and predict, and comprehend complex scenarios. He argues that he uses analytics as part of a business process when he does so. For example, statisticians in an insurance firm may develop a model that predicts the likelihood of a false claim. The form, together with some decision rules, can be entered into the company’s claims processing system to report claims with a high potential for fraud. These allegations will be passed on to the investigation team, who will look into them further. The form itself may not be evident to the end user in other instances.

The presentation of data in a pictorial or graphical style is known as data visualization, and the software that makes that presentation is known as a data visualization tool. Data visualization gives users an intuitive way to explore and evaluate data interactively, allowing them to spot interesting patterns, infer correlations and causality, and support meaning-making activities.

Data visualization is not a new concept; it has existed for quite some time. The use of data visualization to convey messages and represent complex things is simple and quick. There are a variety of strategies that can be used to detect patterns in the data it pulls. Because data is growing at such a rapid rate, traditional data presentation approaches are becoming obsolete. The objective of big data visualization is to find intriguing patterns and relationships. We must carefully select the dimensions of the data to be visualized; if we reduce the dimensions to reduce the size of our visualization, we risk losing interesting patterns; on the other hand, if we use all of the dimensions, we risk creating visualizations that are too dense to be useful to users.

Data visualization is considered the most efficient solution in IoT. It visualizes the endless stream of information collected by the respective systems to exponentially increase its value due to the meaningful insights it provides [20]. Data visualization is practically a state-of-the-art technology that enhances individuals' understanding of data by providing the visual presentation. Visualization is primarily the graphical representation of information to give the viewer a qualitative understanding of the information contained within the data. Through its respective techniques, the user gets to have an entirely new meaning of the data in connection with identifying the hidden trends and patterns. IoT Data visualization is the approach where the raw data within the different data streams is presented in a more insightful manner [8]. The data is analyzed by considering particular patterns and behaviors such that the insights become valuable in decision- making.

# (Il-Agure & Dempere, 2022)

## BIG DATA in IoT

According to Kundhavai & Sridevi [13], IoT and Big Data are two evolving technologies that go hand in hand. While the primary idea behind IoT is having every object get an IP address and be linked to other things, Big Data focuses on managing the enormous data amounts generated from the use of IoT. With IoT technologies allowing the connection of trillions of devices that will produce massive volumes of data, it can be challenging to attain efficiency during data collections. The real-time or close-to real-time communication allowed by the IoT technologies potentially mean that the users will be dealing with raw data [13]. To this effect, Big Data analytics are required to examine the raw data to come to meaningful conclusions that will be used to make optimized decisions within the respective fields.

According to Aziz et al. [4], IoT data generally satisfies the criteria established to determine if particular information falls under the Big Data category. The author cited various authors, such as [6], who asserted that many devices in IoT networks would generate massive amounts of data. Specifically, the criteria are five features, volume, variety, velocity, veracity, and value. The volume refers to the fact that there is much more data than ever before, and it continues to grow very fast. IoT devices will generate enormous data amounts such that the right tools are required for the processing, encompassing their classification, processing, and analysis. The variety feature refers to the data coming in diverse forms that are typically incompatible with each other. IoT-generated data usually is variable in terms of its structure, the time it is attractive and could be of uncertain provenance.

Regarding the velocity feature, the data is considered streaming or instead arriving continually in real-time, and the users are interested in obtaining meaningful information from them instantly [4]. loT devices are constantly streaming in raw data in real-time, for instance, in the health sector, where raw streams of sensor values are continually streaming in with a subsequent need for them to be converted to semantically meaningful activities. The veracity feature refers to the data's veracity. IoT sensors do not have margins of error in their measurement. It is essential that information is appropriately stored, accurate and complete. The “truthiness” of data forms the basis of many business decisions [12]. It is necessary to differentiate between reliable and unreliable data. The value feature refers to the data providing practical business value to their users for competitive advantage, which is evident in IoT data.

Figure 1 depicts large amounts of unstructured data are generated by IoT devices which are collected in the big data system. This IoT generated big data which extensively relies on the 3V factors that are volume, velocity, and variety. In the big data system which is basically a shared distributed database, a huge amount of data is stored in big data files. The analysis of the stored IoT big data using analytic tools like Hadoop MapReduce ensues to finally generate the reports of analyzed data. The interactions between Big data and IoT are mutually beneficial. Therefore, as Big Data focuses on allowing real-time analysis of the data generated by IoT technologies to optimize its use, IoT will also bring considerable benefits to Big Data. The spiking growth in the use of IoT technologies will generally prompt developers to demand greater Big Data capacities and further development [14]. Therefore, it becomes vital that the data storage technologies, which are also part of the IoT networks, have better systems capable of processing more data. The interactions will hence allow technological growth in the two fields concurrently.

## Data Visualization in Iot

According to Endert et al. [9], IoT and data have a natural connection. Notably, the retrieval and visualization of essential insights from the data in IoT devices depend on the kind of data analysis this data has undergone. Data analytics is characterized as a procedure that consists of analyzing data sets in terms of shapes sizes and changing data properties to separate the most fundamental conclusions and important insights [9]. The data analytics will provide the basic undertaking required to advance and accomplish IoT applications and ventures. The data analytic tools will empower the strong units to make effective use of the datasets, as explained in the findings.

McCarthy [16], indicated that the transformative experience expected from the significant insights obtained from IoT data would require robust reporting and visualization tools. Therefore, to ensure that accurate reports and visualizations are present from the data visualization process in IoT networks, there are five essential factors to be considered. The first factor is pinpointing the actual thing that will be communicated. Despite the IoT deployments collecting a lot of data, the decision-makers need to concentrate on the data that will help attain specific objectives [16]. This is understood by asking stakeholders what they want to learn from the data and what decisions are made from the information. This will ensure that the process incorporates only the datasets aligned with the intended objectives [16]. The second discussed factor was selecting the right reporting style that will communicate with the audience. The representation of data mustn't become too complex to get confusing. The third factor is ensuring that the reports are simplified as they are critical to effective data visualization [16]. The fourth factor is considering the enterprise data integrations to create new data views vital in the decision-making. Finally, best practices must be set to streamline the reporting process and ensure cohesion.

According to Allam [3], the IoT visualization systems will have a custom dashboard interface that aids the users in analyzing ridge available raw metrics and obtaining information on the models' functionality. The process considers the observations collected from the numerous geographically dispersed IoT sensors and the several AI models added to the findings. The primary objective of the programs is to increase the operators' confidence in the models. The modular visualization framework will need to retain vital features, including the ability to have the features updated in real-time, the interactive elements, openness, and clarity. Considering the IoT metrics are extensively complex, and new measurements are considered in full detail, the visualizations are required to ensure the new measurements can be accessed in real-time [3]. The dashboard could have an immersive user experience that allows accessibility and communication among the operators with the data. The dashboard could also include ways to view and simulate the implemented AI models to improve both the models' clarity and interpretability. In [3], the authors discussed visualization models including Tableau, Thingsboard, IBM Watson, Grafana, and Kibana Platform. The research would ultimately indicate that the integration of human judgment into the methods for data processing allows visual analysis to uncover information and obtain practical perspectives from this data collection.

Khalid et al. [11], investigated the most prevalent issues in IoT data visualization, indicating that many studies have focused on the respective techniques. Considering the domain, different visualization techniques must be selected. Each of them will have specific issues that are not in other techniques. It will be preferred since it demonstrates what the domain experts want to see. Considering clustering as an example, the study stated that it could be overwhelming to illustrate the points in the Cartesian graph. Large data volumes could bring issues such as data overloading and over plotting. The approaches to handle the techniques include reducing dimensionality, caching, and prefetching together with incremental and adaptive processing in visual analytics [11] While the modern approaches address the challenges in data visualization, there is also the matter of human errors and oversimplification of the data. Therefore, machine learning has been considered to reduce the errors that occur in visualization. Notably, [11] asserted that the key to marketing in the future is in the visualization of IoT data. However, there is still a greater chance of a declining number of data scientists' due to automation. Therefore, while this may not predict a decrease in the number of data scientists, it acknowledges a lower quality of produced data considering that the job will be averaged. The absence of quality data scientists and increasing fresh data scientists will prompt better visualization guidance. The existence of too much data to be explored but with minimal human resources, the latter will be degraded to a particular extent to accommodate the demand as stipulated in the law of economics.

# (Mukherjee et al., 2022)

A diagram of a big data flow

Description automatically generated

# (Rana et al., 2023)

## TYPES OF BIG DATA

Structured Data

* The data which can be represented in a structured way.
* Mostly in the form of rows and columns.
* Such tables can be joined in a data model to form relationships.
* Updating, Deleting, or modifying data becomes easy in structured big data.
* Data mining from structured data can be easily done.
* Ensuring & enabling security features are easy [5].

Unstructured Data

* Data which can’t be represented in a definite structure.
* Data which can’t be confronted with a data model.
* Data which can’t be stored in relational databases.
* Such data is portable.
* It can be easily scaled.
* But it is difficult to store such data due to lack of a proper schema [5].

Semi-Structured Data

* This data has some structure.
* But it cannot be confronted with a data model.
* It contains Metadata which can be used to group and describe data.
* Entities kept in the same group may or may not have similar properties.
* Their size may also differ even if they ’re in the same group.
* It is flexible and portable.
* Queries of Structured Big Data are more efficient than Semi-Structured [5],

A diagram of a big data

Description automatically generated

## POWER BI

There are a number of reasons why we prefer a BI tool like Microsoft Power BI [6] over traditional tools used for gaining insight from data such as Excel in such scenarios where we have a huge amount of data.

The processing speed is faster in Power BI than in Excel.

The visualizations created in Power BI are more appealing.

The Power Queries help to manipulate such a huge amount of data very easily.

Excel can only hold up to 1.4M rows & 16.38k columns. So, in the case of big data, it fails [7].

This paper consists of the latest craze of Microsoft power BI as a hustler apparatus with the ability to frame and draw insights from data that can be accessed all over the corporation. Big data is renovating industries of all domains [8]. Main components of Power BI are shown in

## Overview of Big Data

Big data is a huddle of huge quantities of information which keeps on growing exponentially with time. The amount of data is so huge that traditional data analytics tools fail to analyse such a large amount of data. There are various examples of Big Data in daily life like the data of the Stock Exchange of New York, which generates 1 TB+ data daily for new trade. Platforms of social media like Facebook generate 400 TB+ data daily

# (Kahveci et al., 2022)

## Data acquisition

Data acquisition in Industry 4.0 big data analytics systems enables the collection of data from field devices for purposes such as data storage, visualisation, and analytics. OPC Unified Architecture (OPC-UA) and Modbus are two common data communication protocol examples used in the acquisition of data in smart manufacturing settings [17]. These protocols can collect production data in real-time or in batches (or in some instances both). Data communication protocols such as OPC-UA, Modbus, and MTConnect, for example, can allow big data analytics systems to gather production data in real-time, whilst software utilities such as Apache Hadoop can collect data in both real-time and batch-oriented modes [18]. It should be noted that IoT gateways are frequently used to realise data gathering and integration in IoT-enabled big data analytics systems. These gateways provide a variety of important services, including protocol translation, encryption, data processing, management, and filtering, as well as wireless networking of legacy and distant industrial equipment [19]. Furthermore, IoT gateways may be configured to connect with end sensor nodes or I/O devices using protocols such as MQTT, Constrained Application Protocol (CoAP), Hypertext Transfer Protocol (HTTP), and many others. Currently, the majority of shop floor communication protocols are proprietary, which poses a challenge for information technology (IT) / operational technology (OT) integration required to promote data inseparability and connectivity [20]. Therefore, there is a need to develop methods to make use of open communication standards and promote seamless integration between devices and systems.

## IoT-enabled cloud platforms

Sensors are now widely accessible and reasonably priced, allowing industrial equipment, machinery, and other devices to produce massive amounts of data [21]. To make use of these data sources, industrial activities must first be linked to the digital world. In other words, to analyse raw production data, it must be acquired, connected to a network, and stored digitally. In today’s manufacturing, GE’s Predix, ABB’s Ability, Siemens’s MindSphere, Schneider Electric’s EcoStruxure Platform and Honeywell’s Forge are a few examples of commercially available industrial IoT solutions which can help industries to utilise the big data. Furthermore, cloud service providers like Amazon Web Services (AWS), Microsoft Azure, Google Cloud, Oracle, and IBM’s IoT platforms offer capabilities that may be utilised for a variety of industrial applications. The majority of these cloud platforms include distributed computing, big data analytics solutions, tools for data and device management, machine-to-machine (M2M) communication capabilities, and a wide variety of supporting services. In smart manufacturing, field devices can be connected to IoT cloud platforms using several methods, including "plug and play", open communications standards for industrial automation, such as the OPC UA, and publish-subscribe network protocols, for example MQTT [22]. Naturally, cloud-based solutions offer scalable compute power, data storage, and a wide range of services to meet individual requirements. Also, industrial-grade IoT solutions prioritise device safety as well as cyber-security which is a critical component of the industrial cloud computing infrastructures [23]. On the other hand, in most cases, they introduce dependencies on external connectivity, proprietary technologies, limited support for industrial protocols, and custom implementation.

## Data visualisation

There are a large number of data visualisation and analytics software and platforms available both on-premises and on the cloud. Each of them has its advantages and disadvantages. It is inaccurate to assert that one tool is superior than another without considering the specific needs of a certain use case. Increasingly, cloud-native data visualisation and analytics solutions take the advantage of serverless computing and can automatically scale without any infrastructure to manage. Amazon QuickSight by AWS and Power BI on Microsoft Azure are two good examples to create insights from big datasets. Alternatively, Chronograf, Grafana, Kibana, and Splunk can be utilised on-premises, on-demand, or in hybrid environments. For custom dashboards which requires hands-on development, a variety of free and commercial libraries are also available such as D3.js, CanvasJS, and Chart.js [31], [32], [33]. It is crucial to make better use of the overwhelming amount of OT level data by improving the accuracy and speed of decision making process. Correct decisions are expected if data presentation and visualisation delivered in a way that capitalises human perception. Moreover in case of poorly designed dashboards, end users could show less interest in the solution despite of its reliability and effectiveness. Therefore, the graphics must be designed carefully in a smart way that attracts the audience. This can be achieved by providing visual functions to end users enabling them displaying the right information in an appropriate format.

# References

Allaymoun, M.H. *et al.* (2022) ‘Data Visualization and statistical graphics in big data analysis by Google Data studio – sales case study’, *2022 IEEE Technology and Engineering Management Conference (TEMSCON EUROPE)* [Preprint]. doi:10.1109/temsconeurope54743.2022.9801964. Available at: https://library.ittralee.ie/ (Accessed 14 November 2023).

Il-Agure, Z. and Dempere, J. (2022) ‘Review of data visualization techniques in IOT Data’, *2022 8th International Conference on Information Technology Trends (ITT)* [Preprint]. doi:10.1109/itt56123.2022.9863948. Available at: https://library.ittralee.ie/ (Accessed 14 November 2023).

Kahveci, S. *et al.* (2022) ‘An end-to-end big data analytics platform for IOT-enabled Smart Factories: A case study of battery module assembly system for electric vehicles’, *Journal of Manufacturing Systems*, 63, pp. 214–223. doi:10.1016/j.jmsy.2022.03.010.

Mukherjee, S. *et al.* (2022) ‘Leveraging Big Data Analytics in 5g‐enabled IOT and industrial IOT for the development of Sustainable Smart Cities’, *Transactions on Emerging Telecommunications Technologies*, 33(12). doi:10.1002/ett.4618. Available at: https://library.ittralee.ie/ (Accessed 14 November 2023).

Rana, R., Paliwal, N. and Singhal, A. (2023) ‘A study of Business Insight Tool using big data analytics’, *2023 International Conference on Computational Intelligence, Communication Technology and Networking (CICTN)* [Preprint]. doi:10.1109/cictn57981.2023.10140726. Available at: https://library.ittralee.ie/ (Accessed 14 November 2023).