

BDO IN SOUTH AFRICA | FINANCIAL SERVICES

AN INTRODUCTION TO BIG DATA & DATA SCIENCE: ORGANISATIONAL APPLICATIONS

BIG DATA

TABLE OF CONTENTS

AN INTRODUCTION TO BIG DATA & DATA SCIENCE: ORGANISATIONAL APPLICATIONS.....	3
INTRODUCTION.....	3
APPLICATIONS FOR BIG DATA ANALYSIS.....	4
BIG DATA'S INTEGRATION WITH OTHER EMERGING TECHNOLOGIES.....	5
BIG DATA IN FINANCIAL SERVICES.....	6
THE RISKS AND REGULATIONS OF BIG DATA.....	8
CONCLUSION.....	9
REFERENCE LIST.....	10

INTRODUCTION

APPLICATIONS FOR
BIG DATA ANALYSIS

BIG DATA'S
INTEGRATION WITH
OTHER EMERGING
TECHNOLOGIES

BIG DATA IN
FINANCIAL
SERVICES

THE RISKS AND
REGULATIONS OF
BIG DATA

CONCLUSION

REFERENCE
LIST



AN INTRODUCTION TO BIG DATA & DATA SCIENCE: ORGANISATIONAL APPLICATIONS

Malcolm Hodgson & Christopher O'Flaherty,
BDO Financial Services Technology Junior Analysts

INTRODUCTION:

The use of Big Data and Data Science encompasses a number of innovative, game changing, initiatives and possibilities. Gartner defines Big Data as a "high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation."¹ Big Data is the phenomenon wherein data has exponentially grown and keeps growing in numerous day to day activities. Data Science encompasses using this Big Data for analysis and inference.

Big Data comprises of the five V's:

1) Volume

a. Volume refers to the sheer size of Big Data. Volume defines the data infrastructure capability and the planning of current and future storage.

2) Velocity

a. Velocity refers to the speed that the data flows at. Companies may require data to be managed and processed in real time.

3) Variety

a. Variety refers to the structure that data may comprise of. Data can be in one of three formats:

- i. Structured
- ii. Semi-structured
- iii. Unstructured

4) Veracity

a. Veracity refers to the inconsistencies and uncertainty in data. If data isn't managed properly, data can get messy.

5) Value

a. Data in itself is of no use to anyone unless it is extracted into something valuable. Data needs to be converted and thereafter often analysed to add value.²

¹<https://www.gartner.com/en/information-technology/glossary/big-data#:~:text=Big%20data%20is%20high%2Dvolume,decision%20making%2C%20and%20process%20automation.>

²<https://www.geeksforgeeks.org/5-vs-of-big-data/>

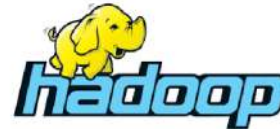
APPLICATIONS FOR BIG DATA ANALYSIS

Data analysis can aid trend analysis, decision making and has numerous other uses. Owing to this, several tools have been developed. This insight can allow companies to make better decisions and thus improve processes. A few of the important tools and applications include Hadoop, R, Power BI and Splunk, however various others exist.

1) Hadoop

Hadoop is an open source software framework for storing massive pools of data. It incorporates large storage capabilities and exceptional processing power to handle concurrent tasks and jobs. Hadoop utilises distributed principles. Hadoop's distributed processing for huge datasets incorporates multiple machines simultaneously. Its storage is decentralised and works on nodes. A single NameNode tracks where the data is housed in the cluster of servers, while the DataNodes store the actual data. Hadoop boasts a variety of advantages, such as:

- The ability to store and process large amounts of data
- Computing power
- Fault tolerance
- Flexibility
- Low Cost
- And Scalability³



2) R

Next up on our list of tools is R. R is a programming language and environment that specialises in statistical computing and graphics. R provides a wide variety of statistical and graphical techniques, including the following:

- Effective data handling and storage
- Large integrated collection of intermediate tools for data analysis
- A well developed, simple and effective programming language which includes conditionals, loops, user-defined recursive functions and input and output facilities.⁴



3) PowerBI

Power BI is an application utilised to turn unrelated data into sources of coherent, interpretable, visually immersive and interactive insights. Power BI can collect your company's data, whether on the cloud or locally, and provide immediate access to this data. Power BI specialises in interactive reporting, which allows users to see all of their data on one screen, dynamically updating as data changes. Furthermore, Power BI also specialises in the following:

- Financial Overview
- User-friendly mobility
- Data Visualisation⁵



4) Splunk

Splunk is an advanced and effective technology that indexes, and searches log files stored in a system. Splunk analyses machine generated data to provide operational intelligence. Splunk does not require a database to store its data, it makes use of its indexes to store its data. Splunk technology has the following advantages:

- It is easy and scalable to implement
- It creates and shares analytical report with interactive charts, graphs and tables
- It finds useful information automatically⁶



INTRODUCTION

APPLICATIONS FOR
BIG DATA ANALYSISBIG DATA'S
INTEGRATION WITH
OTHER EMERGING
TECHNOLOGIESBIG DATA IN
FINANCIAL
SERVICESTHE RISKS AND
REGULATIONS OF
BIG DATA

CONCLUSION

REFERENCE
LIST

³https://www.sas.com/en_us/insights/big-data/hadoop.html

⁴<https://www.r-project.org/about.html>

⁵<https://www.activebs.com/en/power-bi/why-power-bi#:~:text=of%20Power%20BI-,Data%20Analysis,their%20data%20quickly%20and%20easily.>

⁶<https://intellipaat.com/blog/what-is-splunk/>

A BIG DATA'S INTEGRATION WITH OTHER EMERGING

1) IoT

Internet of Things (IoT) refers to a system of interconnected devices via the internet. IoT devices transfer data over the internet without manual intervention. Even though Big Data and IoT have evolved independently, recently they have become interrelated. Studies estimate that by 2020 around 4.4 Trillion GB's of data will have been generated through the internet of things.

The use of IoT devices is exponentially rising. IoT devices generate large pools of data which can be stored and analysed to aid decision making. With organisations gathering data for analytical purposes, IoT acts as a major source of that data. The exponential growth of IoT encompasses numerous uses and capabilities, however, a number of risks are also introduced.⁷

2) AI

Artificial intelligence refers to the simulation of human intelligence. The term is also loosely applied to any machine that can perform human-like tasks, such as learning and problem solving. The more data AI collects, the more accurate it becomes at being able to process a 'human-like' task.

Without data, these AI solutions would not be possible. Through AI and data analysis, numerous solutions are created including weather prediction and music selection. Numerous applications, such as Spotify, use your data and machine learning algorithms to improve a user's experience and uncover certain trends. AI's ability to work so well with Big Data is the primary reason why AI and Big Data are now seemingly inseparable. The primary challenge is, and always will be, the data. If the data fed into AI is unstructured or incorrect, the insights provided will be messy and unstructured as well.⁸

3) Cloud Computing

Cloud computing is the practice of using a network of remote servers hosted on the internet to deliver a variety of services. Instead of keeping one's files on a hard drive / storage device, cloud-based storage makes it possible to store data remotely. When Big Data involves manipulating Petabytes, local storage becomes costly. Cloud computing's scalable environment enables users to save this large amount of data, safely, offsite while incorporating fault tolerant security.⁹

4) Drone Technology

Drones are more commonly being used in numerous sectors, whether it be for stock counts in audits or monitoring crops in agriculture. Drones collect data and store this data which is thereafter analysed to improve decision making processes.

INTRODUCTION

APPLICATIONS FOR
BIG DATA ANALYSISBIG DATA'S
INTEGRATION WITH
OTHER EMERGING
TECHNOLOGIESBIG DATA IN
FINANCIAL
SERVICESTHE RISKS AND
REGULATIONS OF
BIG DATA

CONCLUSION

REFERENCE
LIST

⁷<https://www.whizlabs.com/blog/iot-and-big-data/#:~:text=The%20role%20of%20big%20data%20in%20IoT%20is%20to%20process,in%20the%20big%20data%20system.>

⁸<https://online.maryville.edu/blog/big-data-is-too-big-without-ai/#:~:text=AI's%20ability%20to%20work%20so,rules%20for%20future%20business%20analytics.>

⁹<https://www.thornotech.com/2018/09/big-data-in-the-cloud/>



BIG DATA IN FINANCIAL SERVICES

The combination of emerging technologies results in numerous, innovative, uses in financial services and various other sectors. A few of the uses include:

- Wireless Payments
- Usage Based Insurance
- Personalised customer experience
- Risk analysis and management
- Intelligent asset monitoring
- Algorithmic trading
- Fraud detection
- Process automation

We have observed numerous companies utilising these technologies. A few examples include:

Mastercard increases contactless payments effort in Africa and Middle East

By Divyansh Rajgopal | Apr 7, 2020

The initiative is in line with recommendations from global and regional health authorities and governments to practice social distancing that has led a growing number of merchants to encourage consumers to pay with contactless over cash to avoid human-to-human contact.

*MasterCard increases contactless payments in Africa and the Middle East. Payments have become easier than ever through being able to tap one's bank card to pay or to scan a QR code using one's phone through applications such as Snapscan or Zapper.

These initiatives assist with social distancing during the Covid-19 pandemic, but also increase efficiency when transacting. Moreover, Zapper offers a loyalty program to increase use on the platform and thus allows shoppers to get more value for money.

FNB scoops international Data Anywhere Award in New York

Partner | 4 October 2019

The Data Anywhere category recognised organisations that have modernised their data architecture to unlock value from any data, anywhere – on-premise or across hybrid and multiple public clouds.

*FNB scoops "International data anywhere award" in New York: This has been awarded owing to FNB being able to create a value-add from any data created through their platforms. This initiative allowed users of FNB to make use of real time loans and credit scoring through the banking application. The improvement of financial well-being has also been affected wherein users have improved credit status as well as increased savings amounts. Strategic decisions and analysis has improved all round.

INTRODUCTION

APPLICATIONS FOR
BIG DATA ANALYSIS

BIG DATA'S
INTEGRATION WITH
OTHER EMERGING
TECHNOLOGIES

BIG DATA IN
FINANCIAL
SERVICES

THE RISKS AND
REGULATIONS OF
BIG DATA

CONCLUSION

REFERENCE
LIST

Big Data: The Key to Eradicating Insurance Fraud

For local insurers, one of the most obvious advantages of Big Data lies in fraud reduction. Times of economic downturn tend to be accompanied by an increase in insurance fraud – wreaking havoc with insurers’ business models, and adding new layers of legal and investigative costs.

*Big Data: The key to eradicating insurance fraud. The use of big data allows insurance providers the ability to obtain and validate information earlier, especially when admitting new customers. Predictive analysis can also be completed on an individual level to test each potential candidate. Image recognition and voice biometrics assist with verification and identification. Insurance fraud is far more difficult as trackers are built into the IoT technology of insured devices.

The Discovery Insure Vitality Drive Sensor wins at the MTN Business Internet of Things (IoT) Awards.

NEWS October 2019

Discovery Insure's innovative Vitality Drive Sensor which uses vehicle telematics to accurately track driving behaviour has won the Best Commercial IoT solution in the MTN Business Internet of Things (IoT) Awards.

The device is installed on the car's windscreen and interacts with Discovery's smartphone driving application to measure and reward users for good driving behaviour.

The solution has collected more than 11 billion kilometres of driving data to date.

*Discovery's Vitality package encourages users to meet exercise, driving, and recently spending goals to obtain rewards such as gift vouchers for Netflix, Vida and many other stores. Discovery are utilising IoT devices, such as trackers in your cars or the smart watch on your wrist, creating magnitudes of data. Discovery know where you are, where you've been, and how fast you drove there. They know your weight, your height, your blood pressure, if you smoke and if you have HIV. They know where you shop and what you buy. Annually you also perform questionnaires where they ask about your diet and how many food and alcohol portions you consume per day. Between Vitality drive and bank, it can be known that you are not home. You may have items insured with discovery too, giving them the information of the valuable possessions in your home.

You are rewarded for giving Discovery this information, but you are essentially helping them grow their big data collection. Data has become the most valuable commodity and can be used to make strategic business decisions and can even be sold.

If a malicious user were to get their hands on this information, they could pinpoint the best assets in someone's house and the best time to break in and steal these things.



THE RISKS AND REGULATIONS OF BIG DATA:

Data science encompasses numerous innovative uses. However, with innovation, comes risk. A few of the major risks of Big Data include:

- **Interoperability and integration:** The integration of Big Data with existing systems (including legacy systems) needs to be managed. Interoperability with legacy systems and new tools introduce new risks that may not be accounted for on legacy systems and thus need to be examined.
- **Operational:** A number of operational risks arise with Big Data usage. Firstly, increased costs arise with data storage facilities and scalability needs to be accounted for. Insufficient security and awareness training could further lead to operational risks. The syncing of data may result in risk if not synced correctly resulting in incorrect extractions and analysis.
- **Third party:** Working with third party vendors introduce risks and responsibilities. Whether it be certain tools or applications or data warehouses and cloud storage.
- **Governance and regulatory:** Data, having become one of the most valuable possible commodities, encompasses a number of ethical concerns and requires several regulations. Recently the POPI act has come into effect. The protection of personal information act has been passed to ensure that data is used appropriately and to ensure that if data breaches were to occur, those whose data has been stolen, are timeously made aware of this. By using data unethically, this can be detrimental for a business. Not adhering to the necessary regulations will not only result in large administrative fines and bad public relations, but also may result in imprisonment. Data use and privacy has become highly topical wherein companies have used and abused data as they please. In recent times Cambridge Analytica have used Facebooks data to aid the Trump campaign and Brexit. However, millions of people did not know their data was being used for these purposes.
- **Data management:** The management of large pools of unstructured data should be accounted for in terms of analysis. Data should be cleaned, sorted and verified before used for decision making processes.
- **Identity and access management:** IoT creates massive amounts of data. IoT devices however often have weak security protocols and thus data may not be accurate. Furthermore, breached IoT devices could allow intruders access into systems to manipulate pools of data. Data storage should feature the necessary protocols and these risks should be accounted for when applying IoT technology.
- **Business continuity:** Large volumes of data result in an increased cost to data management. Businesses should ensure that this cost can be scalable. Furthermore, availability of systems is necessary and thus redundancy and fault tolerance is of utmost importance should a breach occur. Big Data aids numerous decision-making processes. If analysis is incompetent, this could detrimentally effect business continuity.
- **Information security:** systems should be monitored. Vulnerability assessments should be performed as well as the implementation of loggings tools to monitor data warehouses. Security should be built in from the ground-up fostering a secure architecture. Poor patch management practices result in breaches and thus continuous updates should be controlled.

INTRODUCTION

APPLICATIONS FOR
BIG DATA ANALYSISBIG DATA'S
INTEGRATION WITH
OTHER EMERGING
TECHNOLOGIESBIG DATA IN
FINANCIAL
SERVICESTHE RISKS AND
REGULATIONS OF
BIG DATA

CONCLUSION

REFERENCE
LIST

```

mirror_mod.use_z = False
    _operation = "MIRROR_Z"
    mirror_mod.use_x = False
    mirror_mod.use_y = False
    mirror_mod.use_z = True

```

```

selection at the end -add
    mirror_ob.select= 1
    modifier_ob.select=1
    context.scene.objects.active
    ("Selected" + str(modifier
    mirror_ob.select = 0
    bpy.context.selected_obj
    data.objects[one.name].sel
    print("please select exactly

```

```

----- OPERATOR CLASSES -----

```

```

types.Operator):
    on X mirror to the selected
    object.mirror_mirror_x"
    mirror X"

```

```

(context):
    context.active_object is not

```


CONCLUSION:

It is evident that Big Data and Data Science encompass a number of innovative and exciting uses. The existing tools are allowing users globally to analyse and sort data, which makes decision making easier. Moreover, there are constantly new and exciting methods for the use of this technology, changing life as we know it on a day to day basis. With innovation, however, does come risk. If managed properly and mitigated timeously, this can result in numerous successful ventures. However, if not accounted for, can be the reason a business fails.

INTRODUCTION

APPLICATIONS FOR BIG DATA ANALYSIS

BIG DATA'S INTEGRATION WITH OTHER EMERGING TECHNOLOGIES

BIG DATA IN FINANCIAL SERVICES

THE RISKS AND REGULATIONS OF BIG DATA

CONCLUSION

REFERENCE LIST



REFERENCE LIST:

- 1) <https://www.forbes.com/sites/gilpress/2013/05/09/a-very-short-history-of-big-data/#379b4c6965a1>
 - 2) <https://www.gartner.com/en/information-technology/glossary/big-data#:~:text=Big%20data%20is%20high%2Dvolume,decision%20making%2C%20and%20process%20automation.>
 - 3) <https://www.geeksforgeeks.org/5-vs-of-big-data/>
 - 4) https://www.sas.com/en_us/insights/big-data/hadoop.html
 - 5) <https://www.r-project.org/about.html>
 - 6) <https://www.activebs.com/en/power-bi/why-power-bi#:~:text=of%20Power%20BI-,Data%20Analysis,their%20data%20quickly%20and%20easily.>
 - 7) <https://www.whizlabs.com/blog/iot-and-big-data/#:~:text=The%20role%20of%20big%20data%20in%20IoT%20is%20to%20process,in%20the%20big%20data%20system.>
 - 8) <https://online.maryville.edu/blog/big-data-is-too-big-without-ai/#:~:text=AI's%20ability%20to%20work%20so,rules%20for%20future%20business%20analytics.>
 - 9) <https://www.thorntech.com/2018/09/big-data-in-the-cloud/>
- *<https://www.iol.co.za/business-report/companies/mastercard-increases-contactless-payments-effort-in-africa-and-middle-east-46380211>
- * <https://businesstech.co.za/news/industry-news/344192/fnb-scoops-international-data-anywhere-award-in-new-york/>
- * <https://www.saia.co.za/info-center/saia-news/2017/07/07/big-data-key/>
- * <https://www.discovery.co.za/corporate/news-vitality-drive-sensor-wins>

INTRODUCTION

APPLICATIONS FOR
BIG DATA ANALYSISBIG DATA'S
INTEGRATION WITH
OTHER EMERGING
TECHNOLOGIESBIG DATA IN
FINANCIAL
SERVICESTHE RISKS AND
REGULATIONS OF
BIG DATA

CONCLUSION

REFERENCE
LIST

WE TAKE IT PERSONALLY.
FOR FURTHER INFORMATION, CONTACT:

CHRISTOPHER O'FLAHERTY

BDO Financial Services Technology Junior Analyst
coflaherty@bdo.co.za

MALCOLM HODGSON

BDO Financial Services Technology Junior Analyst
mhdgson@bdo.co.za

NEVELLAN MOODLEY

Head of BDO Financial Services Technology
nmoodley@bdo.co.za



/BDOSouthAfrica



/bdoafrica



/bdo_sa



/company/bdo-south-africa

www.bdo.co.za