Transforming the Telecoms Business using Big Data and Analytics

Event: ICT Forum for HR Professionals
Venue: Meikles Hotel, Harare, Zimbabwe
Date: 19th – 21st August 2015

Objectives

- Describe the concept of big data
- Importance of big data
- Understand the three dimensions of big data

Big data – A definition

- Big data broadly refers to the new methods and technologies for collecting, managing, and analyzing in real time the vast amount of both structured and unstructured data.

Big Data - Applications

- Organizations collect and analyze vast stores of data for insights that can help identify trends, predict behavior, and empower decision makers. It considers:

  - How much data is generated

  - How this data is identified and managed as an asset to the organization

  - How this data is turned into usable information

  - How organizations use this data to make decisions

Growth of Data

- Within the last decade, the volume of data that was produced in a year is now produced in a week.

- That amounts to over 20 exabytes of data produced a week.

- Data continues to grow exponentially as more of the unconnected become connected.

Memory unit	Size	Binary size
kilobyte (kB/KB)	10^3	2^{10}
megabyte (MB)	10^6	2^{20}
gigabyte (GB)	10^9	2^{30}
terabyte (TB)	10^{12}	2^{40}
petabyte (PB)	10^{15}	2^{50}
exabyte (EB)	10^{18}	2^{60}
zettabyte (ZB)	10^{21}	2^{70}
yottabyte (YB)	10^{24}	2^{80}

Dimensions of Big Data

There are three primary dimensions of Big Data that must be accounted for:

- Volume - Volume describes the amount of data being transported and stored.
- Variety - Variety describes the type of data it is
- Velocity - Velocity describes the rate at which this data is moving – dependent infrastructure (input/output, bandwidth, and latency)
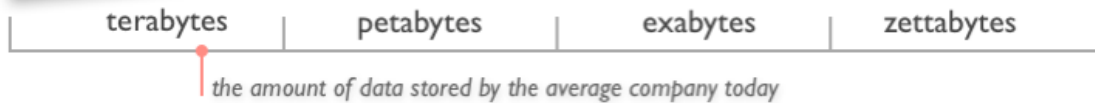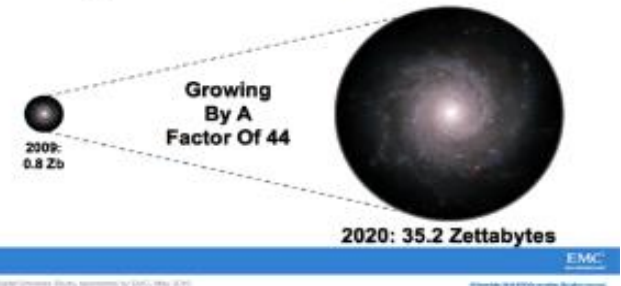
Volume

Data Volume
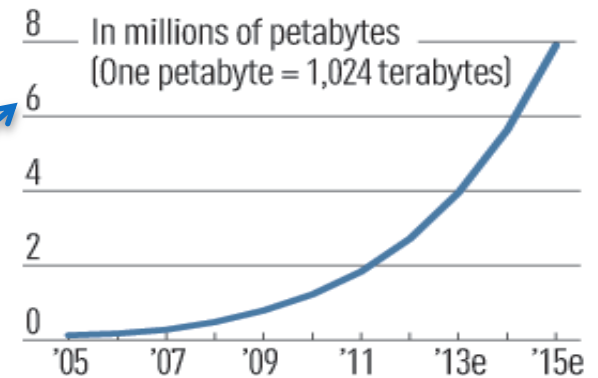
- 44x increase from 2009 2020
- From 0.8 zettabytes to 35zb
- Data volume is increasing exponentially

The Digital Universe 2009-2020
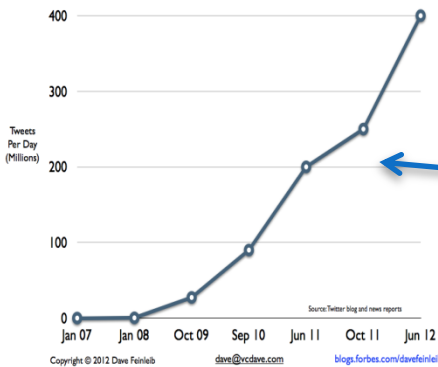
Growing By A Factor Of 44

2009: 0.8 Zb

2020: 35.2 Zettabytes

terabytes	petabytes	exabytes	zettabytes

the amount of data stored by the average company today

Twitter: Tweets Per Day

Source: Twitter blog and news reports

Copyright © 2012 Dave Feinleib   dave@vcdave.com   blogs.forbes.com/davefeinleib

Data storage growth

In millions of petabytes
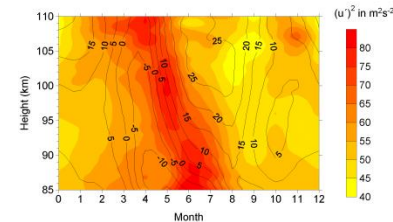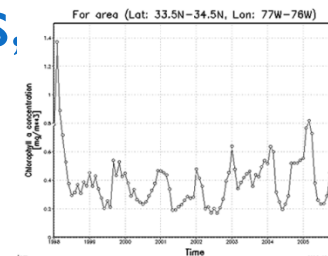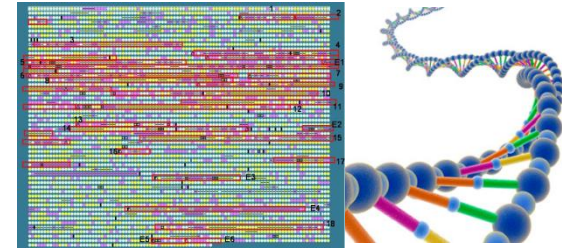(One petabyte = 1,024 terabytes)

Exponential increase in collected/generated data

Velocity

- Various formats, types, and structures
- Text, numerical, images, audio, video, sequences, time series, social media data, multi-dim arrays, etc…
- Static data vs. streaming data
- A single application can be generating/collecting many types of data

To extract knowledge➔ all these types of data need to linked together

Velocity

- Data is begin generated fast and need to be processed fast
- Online Data Analytics
- Late decisions ➜ missing opportunities
- **Examples**
  - **E-Promotions:** Based on your current location, your purchase history, what you like ➜ send promotions right now for store next to you

  - **Healthcare monitoring:** sensors monitoring your activities and body ➜ any abnormal measurements require immediate reaction

Sources of Data

- The vast amount of Data come from many quarters.
  - Social media sites
  - Sensors
  - Digital photos/Videos
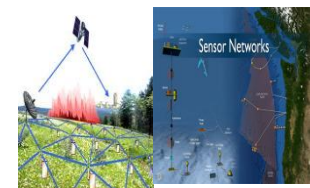 - Business transactions
  - Location-based data

Sources of Data

Social media and networks
(all of us are generating data)

Mobile devices
(tracking all objects all the time)

Scientific instruments
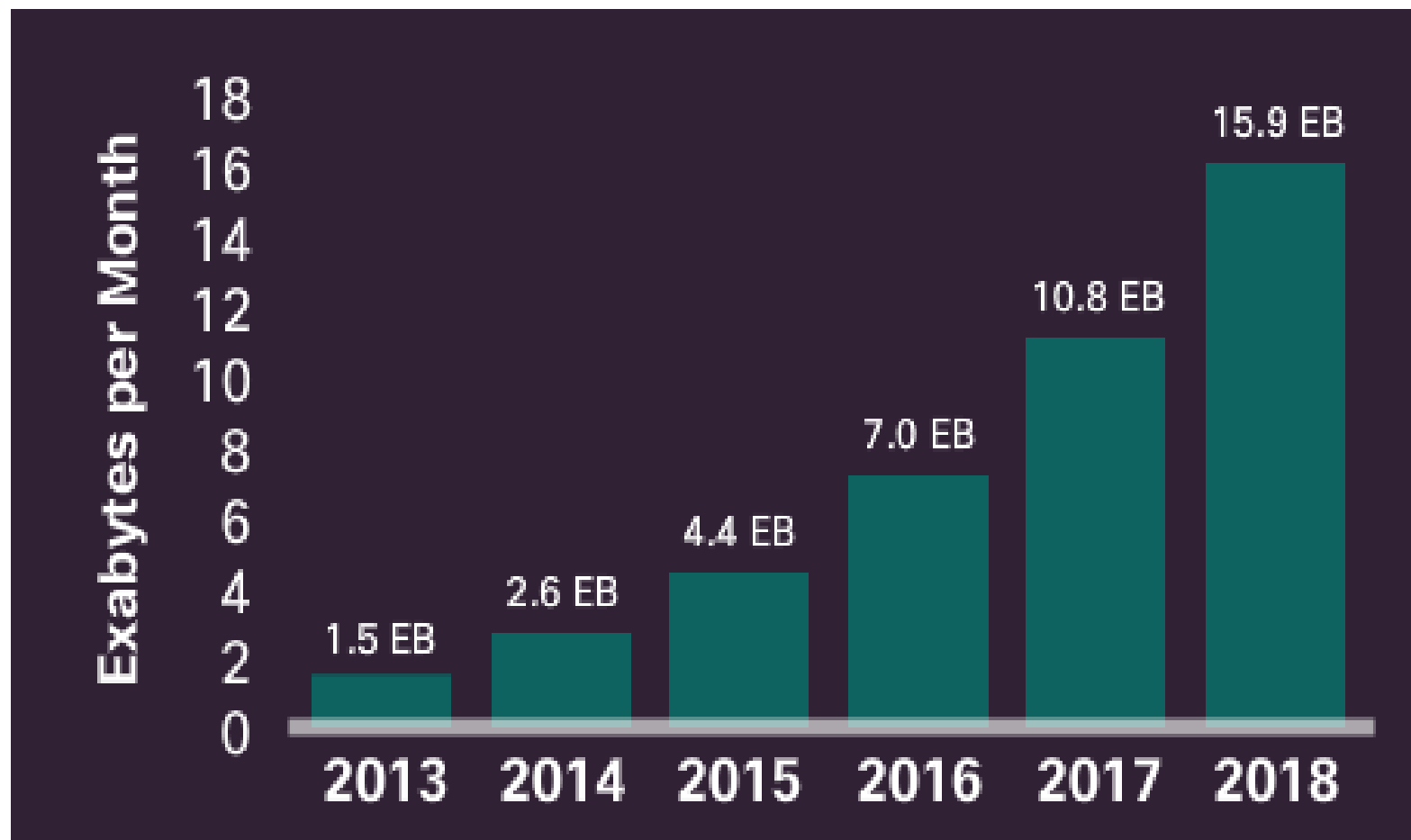(collecting all sorts of data)

Sensor technology and networks
(measuring all kinds of data)

- The progress and innovation is no longer hindered by the ability to collect data

- But, by the ability to manage, analyze, summarize, visualize, and discover knowledge from the collected data in a timely manner and in a scalable fashion
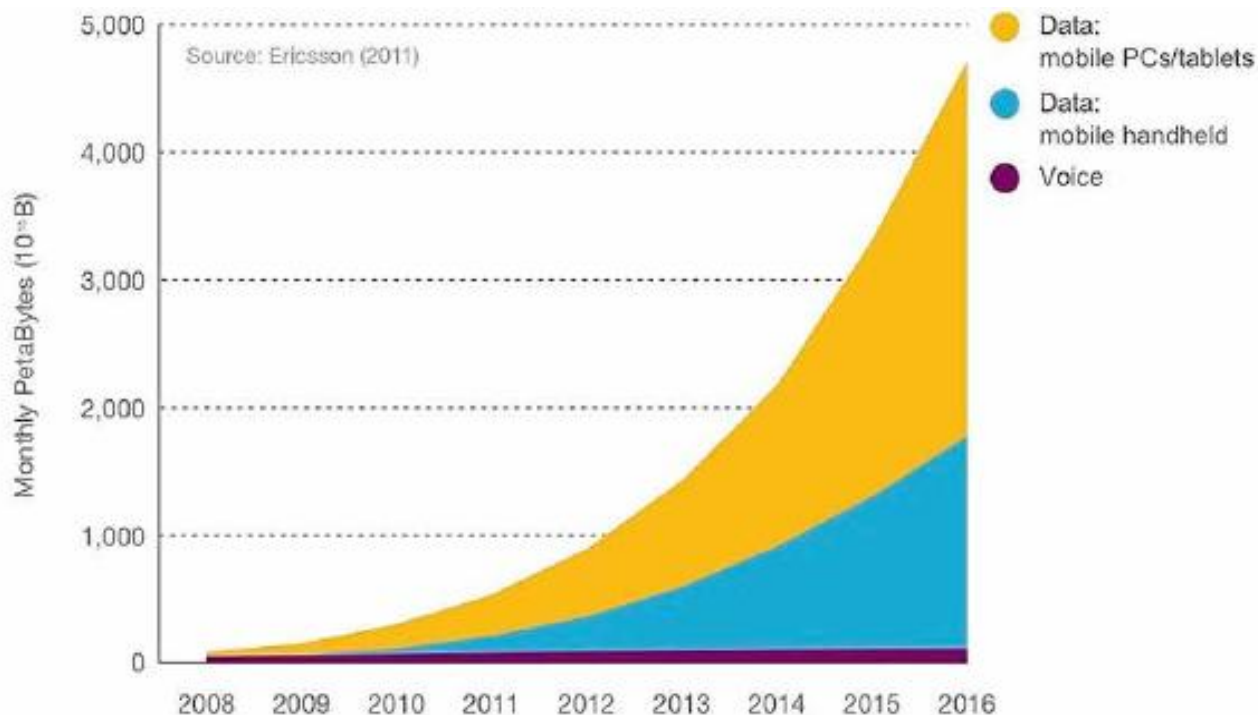
Big Data

Source: Cisco VNI Mobile, 2015

Growth of Data

- Mobile data traffic is expected to grow by approximately 60 percent year over year through 2016, driven mainly by video.

The big data in telecommunication business

- Unstructured data
- Network performance data
- Network traffic data
- Call detail record
- Internet related data

See next slides

What is Data Mining?

- Discovery of useful, possibly unexpected, patterns in data

- Non-trivial extraction of implicit, previously unknown and potentially useful information from data

- Exploration & analysis, by automatic or semi-automatic means, of large quantities of data in order to discover meaningful patterns

Data Mining Tasks

- Classification [Predictive]

- Clustering [Descriptive]

- Association Rule Discovery [Descriptive]

- Sequential Pattern Discovery [Descriptive]

- Regression [Predictive]

- Deviation Detection [Predictive]

- Collaborative Filter [Predictive]

In a Nutshell…

- Big Data is about better analytics!

Big Data and Business Intelligence (BI)

Contributes to:

- Business Alignment Strategy → Business Effectiveness
- Organizational and Behavioral Strategy → Business Efficiency
- Technology Strategy → IT Efficiency

Value Drivers
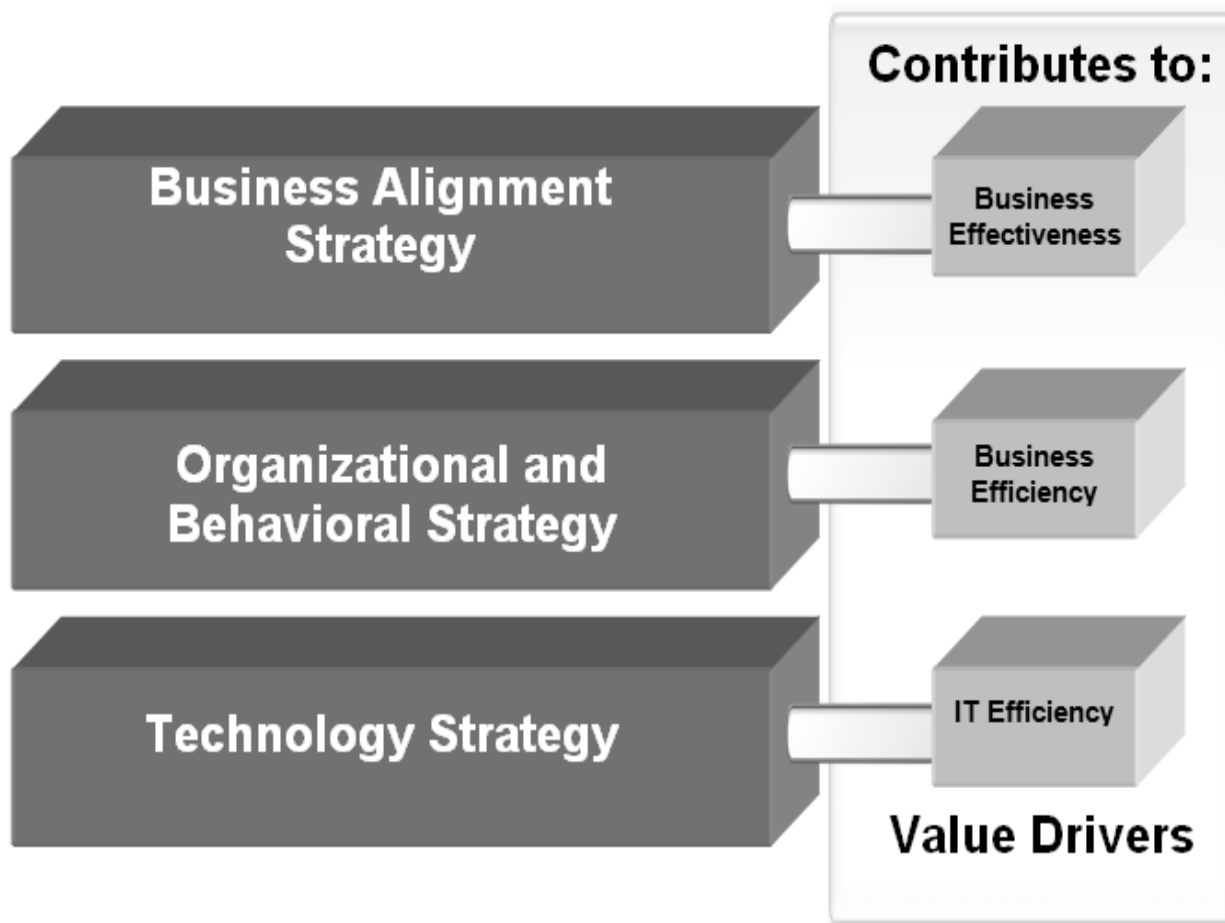
- Business value outcomes tied to the business strategy resulting from business decisions

- Higher productivity, faster time to complete tasks

- Lower total cost of ownership and greater efficiencies in IT

Big Data Technology

- NoSQL (Not Only SQL): Databases that "move beyond" relational data models (i.e., no tables, limited or no use of SQL)
  - Focus on retrieval of data and appending new data (not necessarily tables)
  - Focus on key-value data stores that can be used to locate data objects
  - Focus on supporting storage of large quantities of unstructured data
  - SQL is not used for storage or retrieval of data
  - No ACID (atomicity, consistency, isolation,

NoSQL

- NoSQL focuses on a schema-less architecture (i.e., the data structure is not predefined)

- In contrast, traditional relation DBs require the schema to be defined before the database is built and populated.
  - Data are structured
  - Limited in scope
  - Designed around ACID principles.

Hadoop

- Hadoop is a distributed file system and data processing engine that is designed to handle extremely high volumes of data in any structure.

- Hadoop has two components:
  - The Hadoop distributed file system (HDFS), which supports data in structured relational form, in unstructured form, and in any form in between
  - The MapReduce programing paradigm for managing applications on multiple distributed servers

- The focus is on supporting redundancy, distributed architectures, and parallel processing

Hadoop Related Names

- Apache Avro: designed for communication between Hadoop nodes through data serialization

- Cassandra and Hbase: a non-relational database designed for use with Hadoop

- Hive: a query language similar to SQL (HiveQL) but compatible with Hadoop

- Mahout: an AI tool designed for machine learning; that is, to assist with filtering data for analysis and exploration

- Pig Latin: A data-flow language and execution framework for parallel computation

- ZooKeeper: Keeps all the parts coordinated and working together

What to do with the data
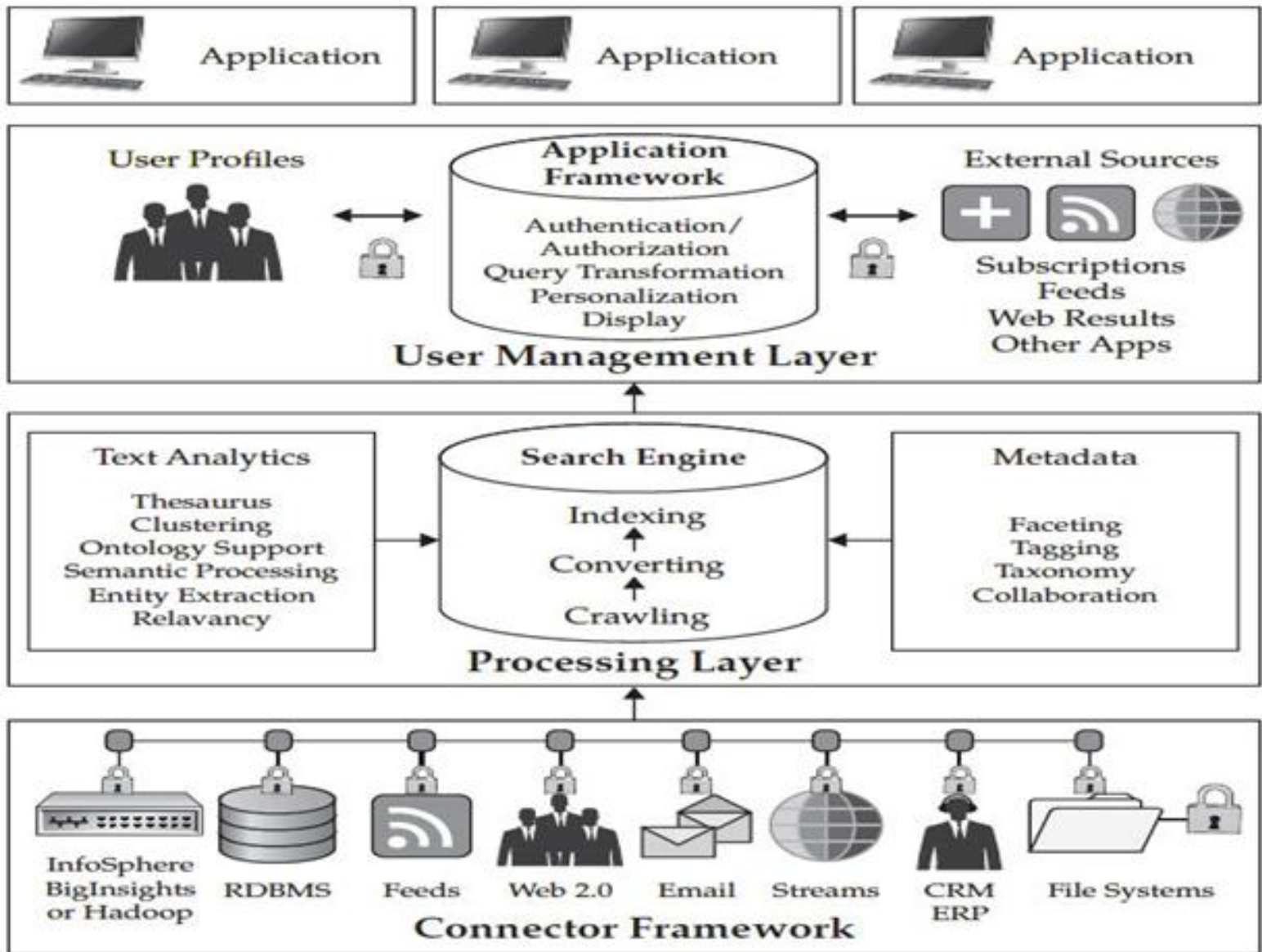
The big data in telecommunication business

- Unstructured data like the emails, text-based documentation, contracts, and social media sources, which contains critical information about customer engagement and customer satisfaction.

- Network performance data including: gigabytes of events logs about services, network, transmission - this provides insights into root cause of alarms and affected services and customers.

The big data in telecommunication business – Cont'd

- Network traffic data – This includes identification of the type of data passing over the network, bandwidth usage and the type of user which is needed to enable better traffic engineering and capacity management and hence informed investment decisions.

- Call detail record which can provide useful call patterns by correlating data about, frequency, date and time, duration, called parties, subscriber location etc.

The big data in telecommunication business – Cont'd

- Internet related data showing IP addresses used, GPS location, frequency, type of data, social networking websites accessed, other sites like new and entertainment, bandwidth requirements etc

Application of Telecoms Big Data

- All these new, vast, and complex type of data is what constitutes the "Big data" and can provide significant, material and actionable insights into customers, products, and operations .

- The new data sources from Web activities, mobile data, user location combined with traditional data enable businesses gain more understanding of the customer requirements.

Big data technology

- Big data tools and software are required to process the extremely large volumes of data that a business has collected to determine which data is relevant and can be analyzed to drive better business decisions in the future.

- Telcos require technology that gathers the vast amounts of data generated by 4G networks, CDRs, clickstreams, IPv6 devices, location sensors, and machine-to-machine monitors in a single format information platform.

Big data technology

- The technology must have the capability to integrate data in near real time, scale cost-effectively and integrate with legacy systems and technologies, and shrink batch windows for high performance.

- Some of the current big data languages capable of efficiently processing massive amounts both structured and unstructured data include the Hadoop and the R language.

Benefits of Big data

- Optimizing routing and quality of service by analyzing network traffic in real time
- Analyzing call data records in real time to identify fraudulent behavior immediately
- Allowing call center reps to flexibly and profitably modify subscriber calling plans immediately
- Tailoring marketing campaigns to individual customers using location-based and social networking technologies
- Using insights into customer behavior and usage to develop new products and services

Benefits of Big data

- Used in finance and HR to provide invaluable strategic asset with access to actionable business insights that improve talent acquisition, retention, development and organizational performance .

- Improved fraud management by correlating internal location, usage, and account data with external sources such as credit reports, operators could significantly increase the detection of fraudulent activity such as looping or call forwarding on hacked PBXs.

- Monetizing the data itself and selling insights about customers to third parties.

Discussion

- How can big data has been used to improve the human relation function in an organization.

- What are the challenges faced by telecommunications business in the implementation of big data

- What are the security considerations should be put in place to protect privacy of client information

- Most telecommunications organizations in Africa are also engaged in other services like mobile money, how can the use of big data impact on improvement of such services

Summary

- Big Data creates new ways of approaching Telecommunications: business: discovery-led rather than theory-led.

- Big data offers telecom business a real opportunity to gain a much more complete picture of their operations and their customers, and to further their innovation efforts.

- Telecommunications organizations that can incorporate these new strategies of learning consumer needs into their organizational processes will gain a more competitive advantage than their counterparts who stick to the traditional methods of learning the market requirements.

Thank you.