Overview of Big Data

Prof Bharati Bhole

Unit I

Introduction to Big Data, Characteristics of Data, and Big Data Evolution of Big Data, Definition of Big Data, Challenges with big data, Why Big data? Data Warehouse environment, Traditional Business Intelligence versus Big Data. State of Practice in Analytics, Key roles for New Big Data Ecosystems, Examples of Big Data Analytics.

Big Data Analytics, Introduction to big data analytics, Classification of Analytics, Challenges of Big Data, Importance of Big Data, Big Data Technologies, Data Science, Responsibilities, Soft state eventual consistency. Data Analytics Life Cycle.

Topics

- Big Data Overview
  - What is Big Data?
  - Data -> Big Data
  - Definition of Big Data
  - Data Structure / Types of Big Data
  - Growth of Data - Tools and Techniques
  - Characteristics of Data
  - Evolution of Big Data
  - Characteristics of Big Data
  - Challenges With Big Data
  - Advantages of Big Data
  - Disadvantages of Big Data
  - Applications of Big Data
  - Why Big Data
  - BI vs Big Data
  - DW Environment
  - Hadoop Environment
  - Coexistence of Big Data & DW
  - Analysts Perspective on Big Data

- State of the Practice in Analytics
  - BI VS DS
  - Current Analytical Architecture
  - Drivers of Big Data
  - Emerging New Big Data Ecosystem
  - Key Roles for the new Big Data Ecosystem
  - Data Scientist Skill Set
  - Examples of Big Data Analytics
  - Big Data Analytics
  - Importance of Big Data Analytics
  - What Big Data Analytics is not?
  - Classification of Big Data Analytics
  - Big Data Technologies

State of the Practice in Analytics

Prof Bharati Bhole

State of Practice in Analytics

TABLE 1-2 *Business Drivers for Advanced Analytics*

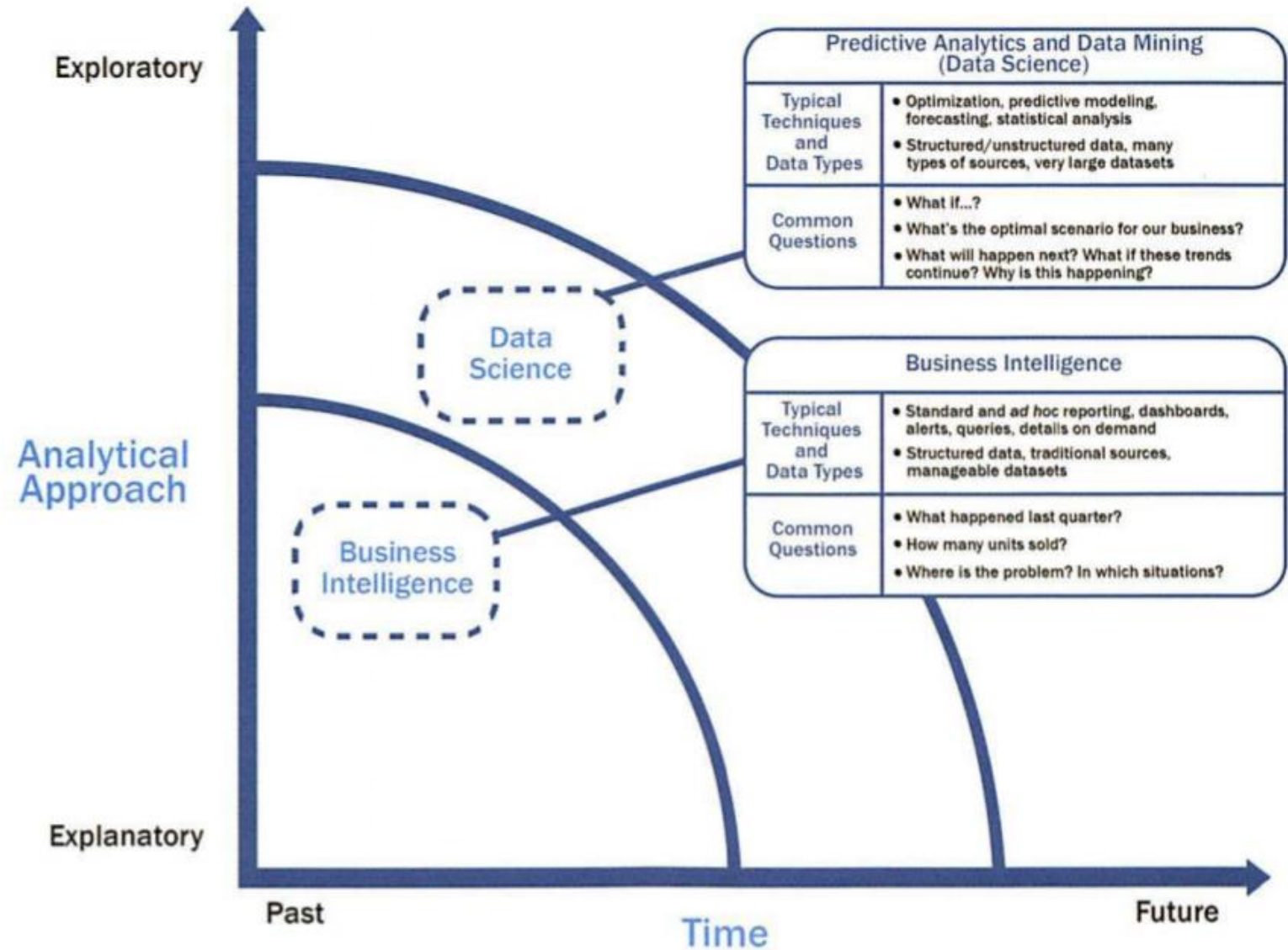| Business Driver	Examples
Optimize business operations	Sales, pricing, profitability, efficiency
Identify business risk	Customer churn, fraud, default
Predict new business opportunities	Upsell, cross-sell, best new customer prospects
Comply with laws or regulatory requirements	Anti-Money Laundering, Fair Lending, Basel II-III, Sarbanes-Oxley (SOX)

BI vs DS

FIGURE 1-8 Comparing BI with Data Science

Prof Bharati Bhole

Current Analytical Architecture

- For data sources to be loaded into the data warehouse, data needs to be well understood, structured, and normalized with the appropriate data type definitions
- Additional local systems may emerge in the form of departmental warehouses and local data marts that business users create to accommodate their need for flexible analysis
- Once in the data warehouse, data is read by additional applications across the enterprise for BI and reporting purposes.
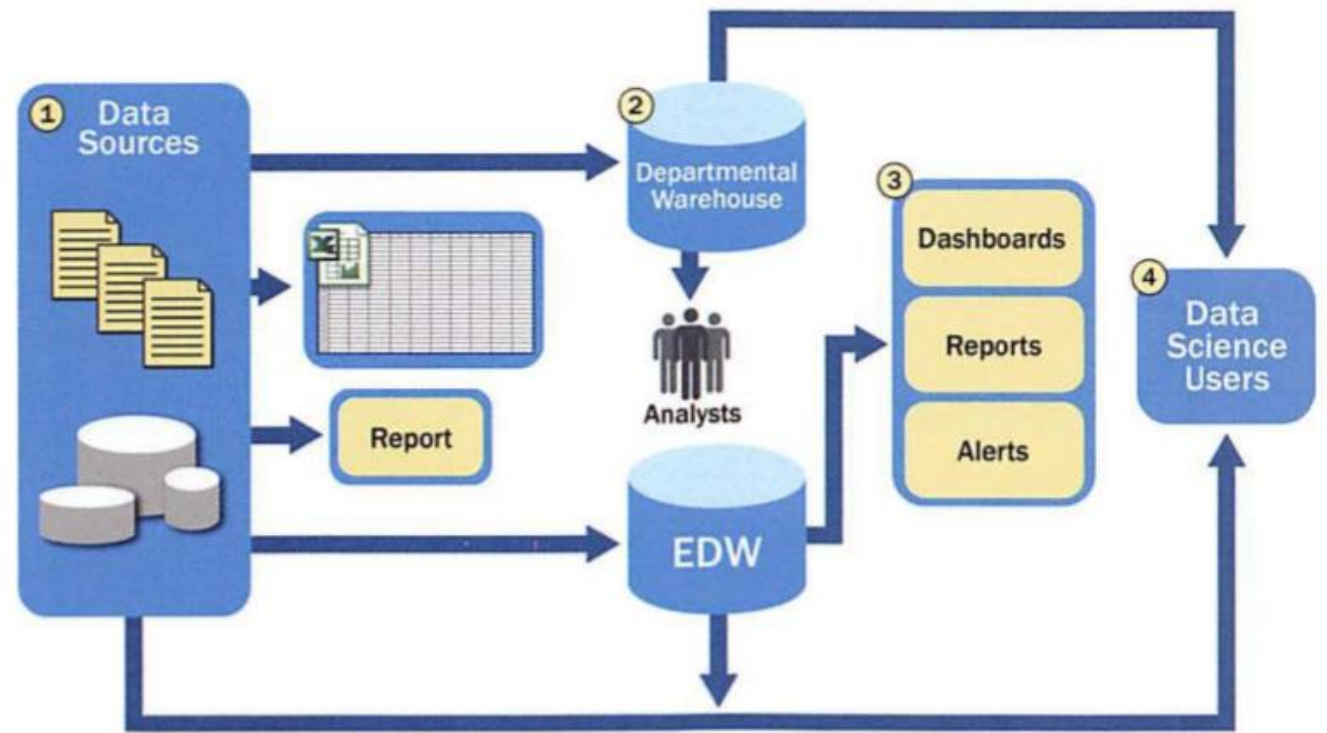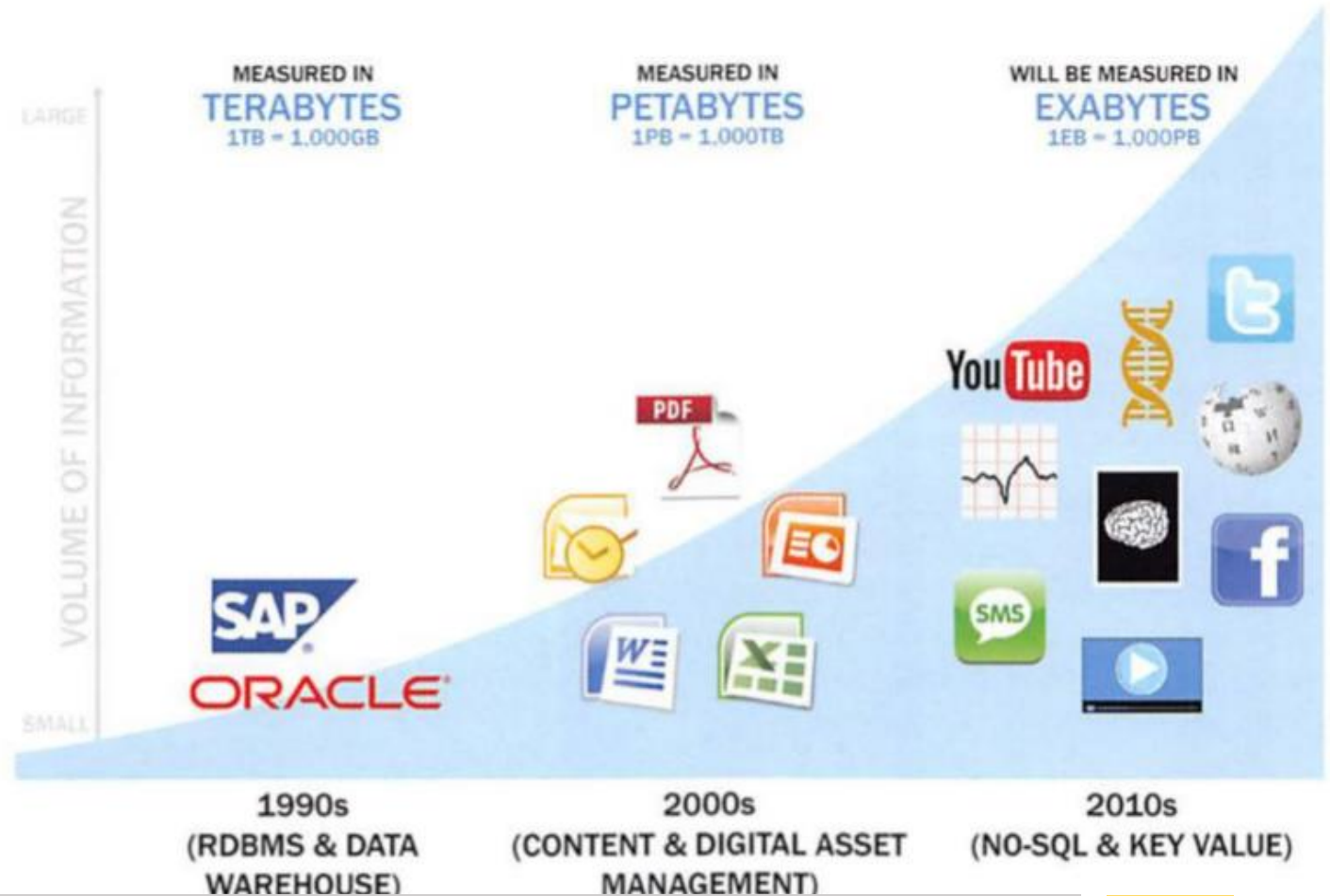- At the end of this workflow, analysts get data provisioned for their downstream analytics.

FIGURE 1-9 Typical analytic architecture

Drivers of Big Data

- Medical information, such as genomic sequencing and diagnostic imaging
- Photos and video footage uploaded to the World Wide Web
- Video surveillance, such as the thousands of video cameras spread across a city
- Mobile devices, which provide geospatial location data of the users, as well as metadata about text messages, phone calls, and application usage on smart phones
- Smart devices, which provide sensor-based collection of information from smart electric grids, smart buildings, and many other public and industry infrastructures
- Nontraditional IT devices, including the use of radio-frequency identification (RFID) readers, GPS navigation systems, and seismic processing

MEASURED IN
TERABYTES
1TB = 1.000GB

MEASURED IN
PETABYTES
1PB = 1.000TB

WILL BE MEASURED IN
EXABYTES
1EB = 1.000PB

VOLUME OF INFORMATION

LARGE

SMALL

1990s
(RDBMS & DATA WAREHOUSE)

2000s
(CONTENT & DIGITAL ASSET MANAGEMENT)

2010s
(NO-SQL & KEY VALUE)

Emerging Big Data Ecosystem & a New Approach to Analytics

FIGURE 1-11 Emerging Big Data ecosystem

Key Roles for New Big Data Ecosystem

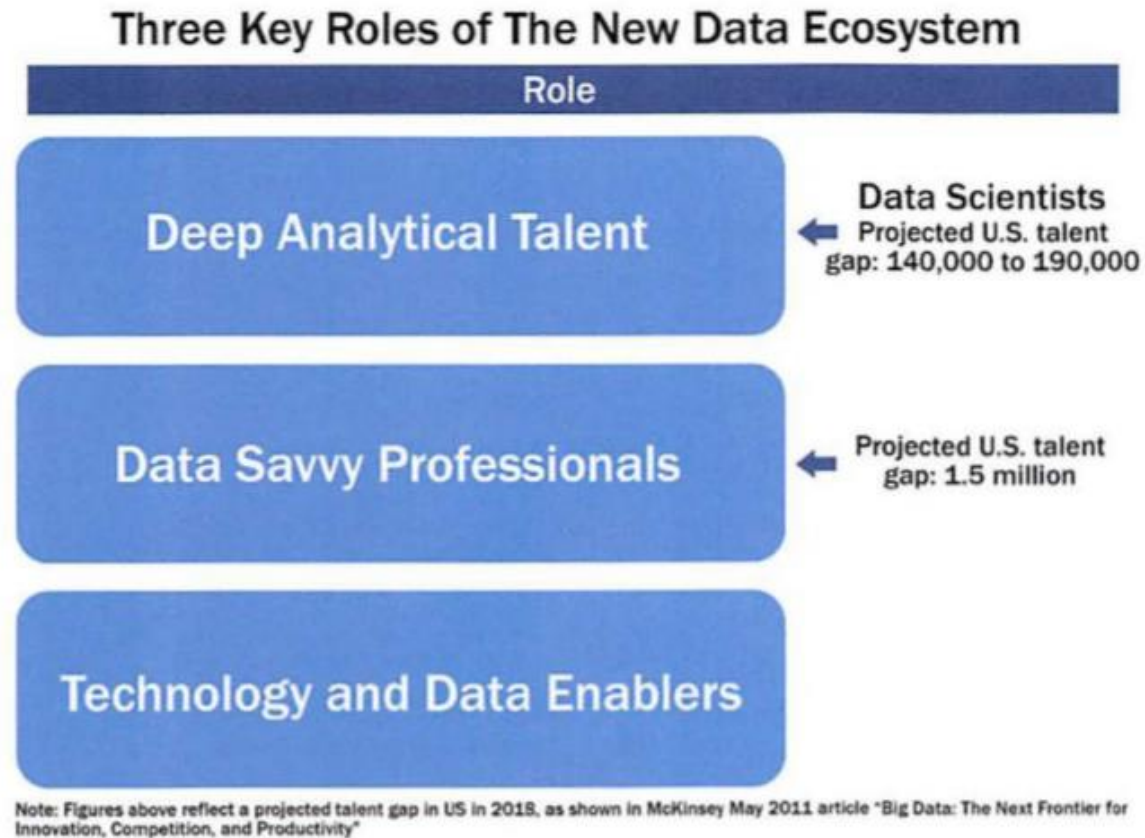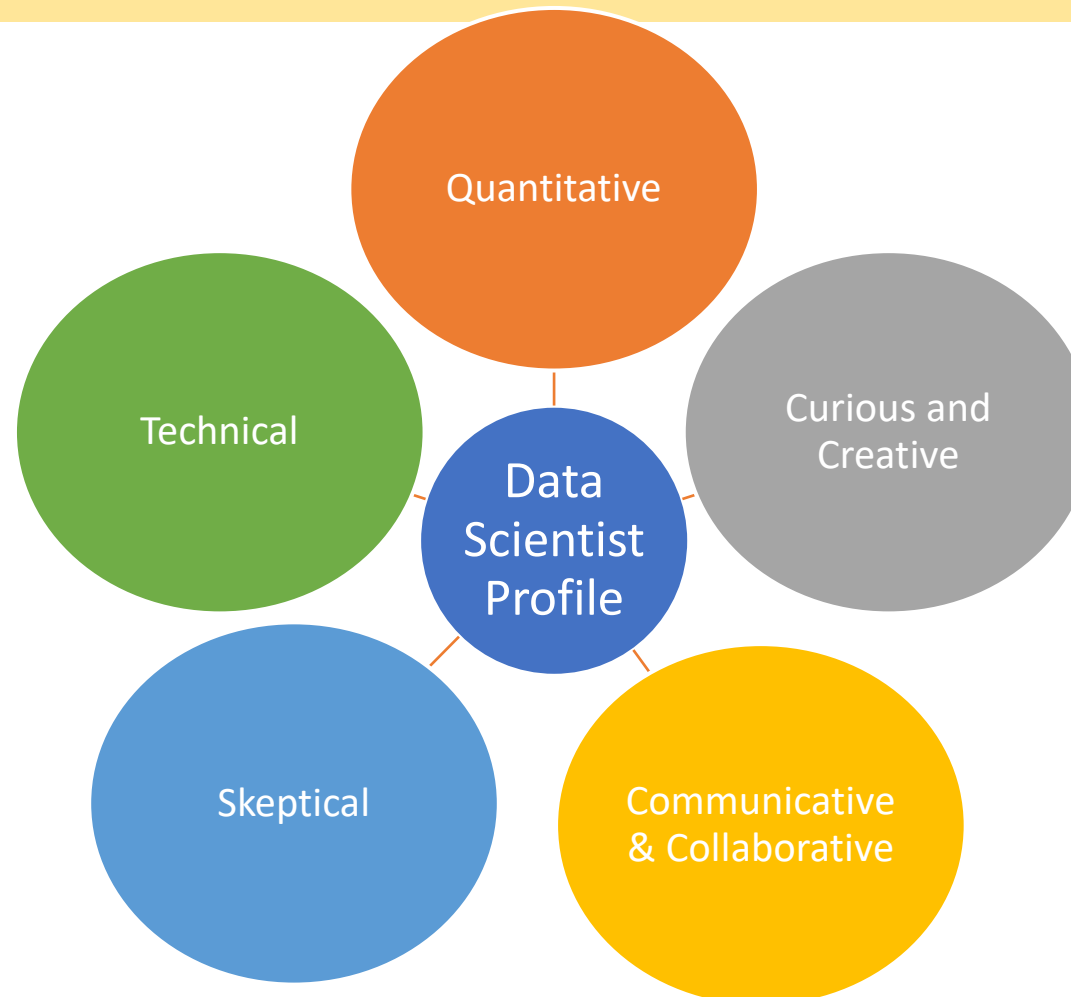# Key Roles for the New Big Data Ecosystem

FIGURE 1-12 *Key roles of the new Big Data ecosystem*

Data Scientist – Skill Set

Prof Bharati Bhole

Examples of Big Data Analytics

- Big Data Analytics in different areas: retail, IT infrastructure, and social media.

- Retail - Big Data presents many opportunities to improve sales and marketing analytics

  Life-Event Situations
  - Marriage, when people tend to buy many new products
  - Divorce, when people buy new products and change their spending habits
  - Pregnancy, when people have many new things to buy and have an urgency to buy them.

- Hadoop represents another example of Big Data innovation on the IT infrastructure.

- Social media represents a tremendous opportunity to leverage social and professional interactions to derive new insights.

Big Data Analytics

Prof Bharati Bhole
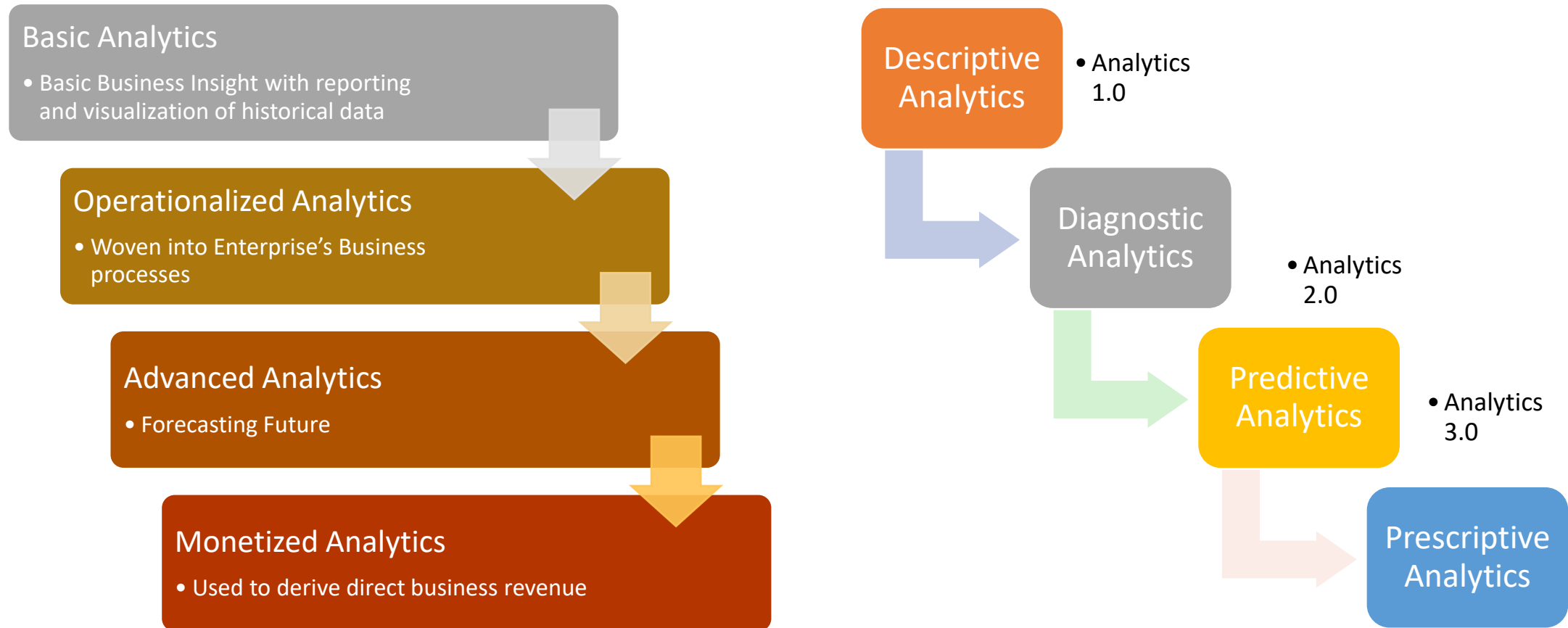
What Big Data Analytics is & is not?

What is Big Data Analytics?

- Technology enabled Analytics
- Competitive advantage
- Reacher / Deeper business insight
- Real-Time Analytics
- Collaboration with data scientist and business users
- Working with huge datasets
- Better and faster decisions in real time
- Moce code to data with greater speed & efficiency

What Big Data Analytics is not?

- Only about Volume
- Just About Technology
- Meant to Replace RDBMS
- Meant to replace DW
- Only used by huge online companies like Google / Amazon
- One-size fill all RDBMS built on shared disk and memory

Classification of Big Data Analytics

Basic Analytics
- Basic Business Insight with reporting and visualization of historical data

Operationalized Analytics
- Woven into Enterprise's Business processes

Advanced Analytics
- Forecasting Future

Monetized Analytics
- Used to derive direct business revenue

Descriptive Analytics
- Analytics 1.0

Diagnostic Analytics
- Analytics 2.0

Predictive Analytics
- Analytics 3.0

Prescriptive Analytics

Classification of Big Data Analytics

Table 3.1   Analytics 1.0, 2.0, and 3.0

| Analytics 1.0	Analytics 2.0	Analytics 3.0
Era: mid 1950s to 2009	2005 to 2012	2012 to present
Descriptive statistics (report on events, occurrences, etc. of the past)	Descriptive statistics + predictive statistics (use data from the past to make predictions for the future)	Descriptive + predictive + prescriptive statistics (use data from the past to make prophecies for the future and at the same time make recommendations to leverage the situation to one's advantage)
Key questions asked: What happened? Why did it happen?	Key questions asked: What will happen? Why will it happen?	Key questions asked: What will happen? When will it happen? Why will it happen? What should be the action taken to take advantage of what will happen
Data from legacy systems, ERP, CRM, and 3rd party applications.	Big data	A blend of big data and data from legacy systems, ERP, CRM, and 3rd party applications.
Small and structured data sources. Data stored in enterprise data warehouses or data marts.	Big data is being taken up seriously. Data is mainly unstructured, arriving at a much higher pace. This fast flow of data entailed that the influx of big volume data had to be stored and processed rapidly, often on massive parallel servers running Hadoop.	A blend of big data and traditional analytics to yield insights and offerings with speed and impact.
Data was internally sourced.	Data was often externally sourced.	Data is both being internally and externally sourced.
Relational databases	Database appliances, Hadoop clusters, SQL to Hadoop environments, etc.	In memory analytics, in database processing, agile analytical method machine learning techniques, etc.

Importance of Big Data Analytics

- Reactive – Business Intelligence
- Reactive – Big Data Analytics
- Proactive – Analytics
- Proactive – Big Data Analytics

Prof Bharati Bhole

Big Data Technologies

- The first requirement is of cheap and abundant storage.

- We need faster processors to help with quicker processing of big data.

- Affordable open source, distributed big data platforms, such as Hadoop.

- Parallel processing, clustering virtualization, large grid environments, high connectivity, and high throughputs rather than low latency.

- Cloud computing and other flexible resource allocation arrangements

Data Science

Business Acumen

Data Scientist

Technology Expertise

Mathematics Expertise
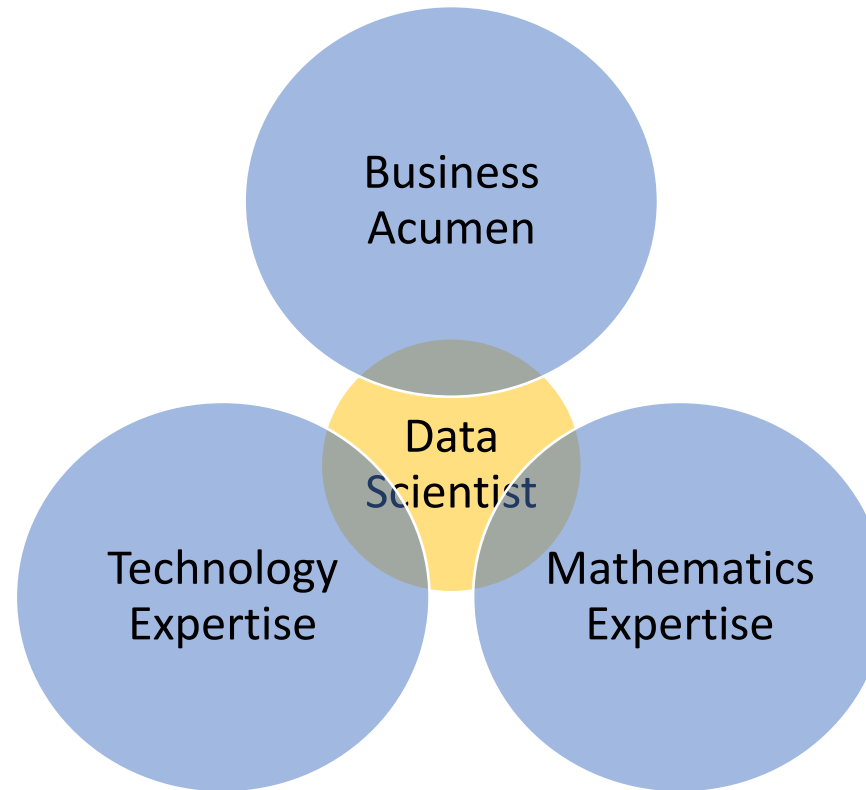
Responsibilities of Data Scientist

- Data Management

- Analytical Techniques

- Business Analysis

Thank You….

Revise the topics from Syllabus References...

Fill Your Attendance Form….!

Prof Bharati Bhole