Innovating Big Data Analytic

Presentation for
IEDE TSINGHUA UNIVERSITY

September 28, 2018

GROUP LEADER KONJOH SELABI ELVIS
MEMBERS:
1. BUKUMBA BIJANU AMANY
2. LEMDJOU YANICK BRICE
3. MEYOF KWEYUI NGONGANG MARK
4. KETCHAYA YANICK BLAISE

BIG DATA & DATA ANALYTICS

SERVICES

OUTLINE

1. Overview
2. New paradigm?
3. Relationship with other fields
4. The future of Big Data

What is Big Data?

- Massive sets of unstructured/semi-structured data from Web traffic, social media, sensors, etc
- Petabytes, exabytes of data
    - Volumes too great for typical DBMS
- Information from multiple internal and external sources:
    - Transactions
    - Social media
    - Enterprise content
    - Sensors
    - Mobile devices

- In the last minute there were …….

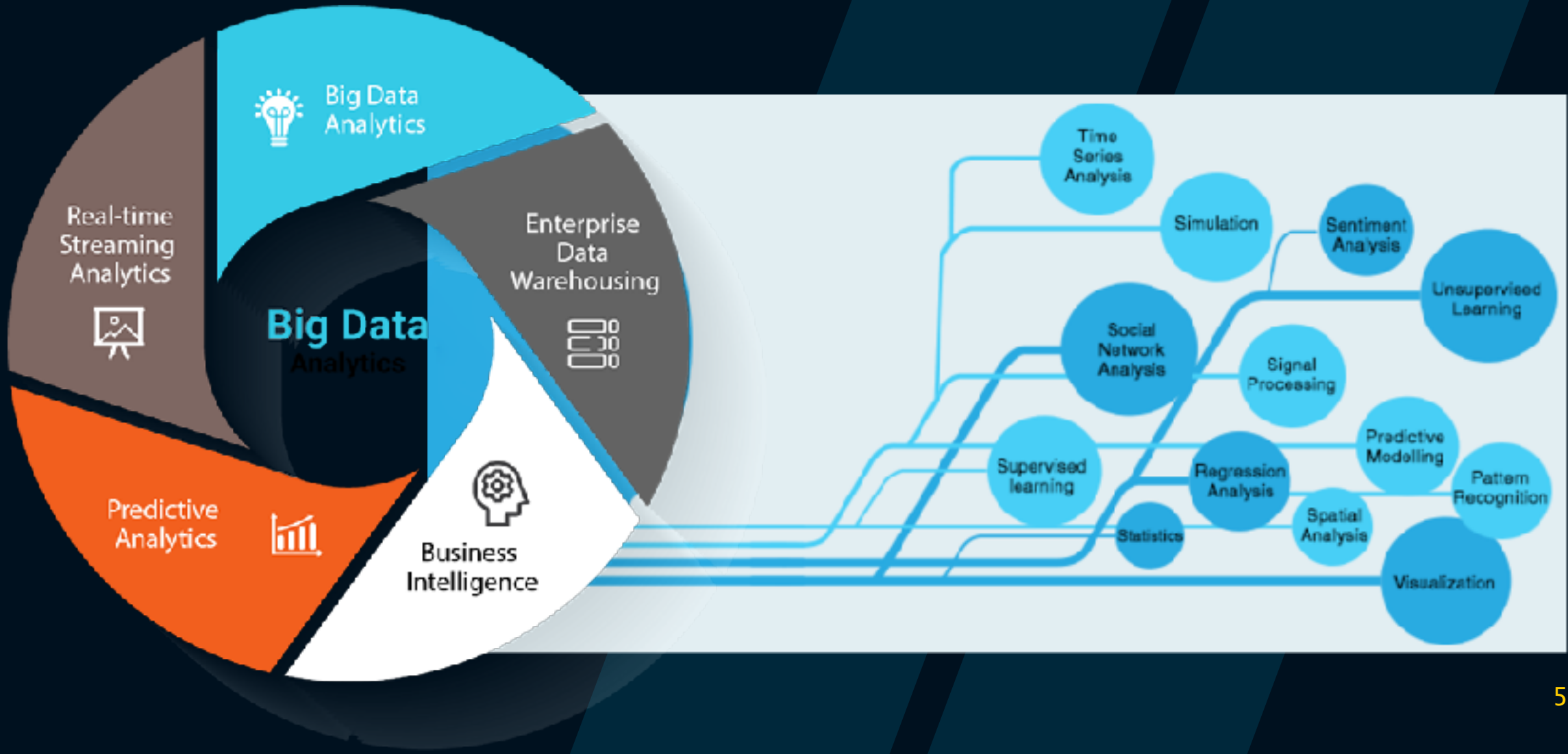
- **204 million emails sent**
- **61,000 hours of music listened to on Pandora**
- **20 million photo views**

- **100,000 tweets**
- **6 million views and 277,000 Facebook Logins**
- **2+ million Google searches**
- **3 million uploads on Flickr**

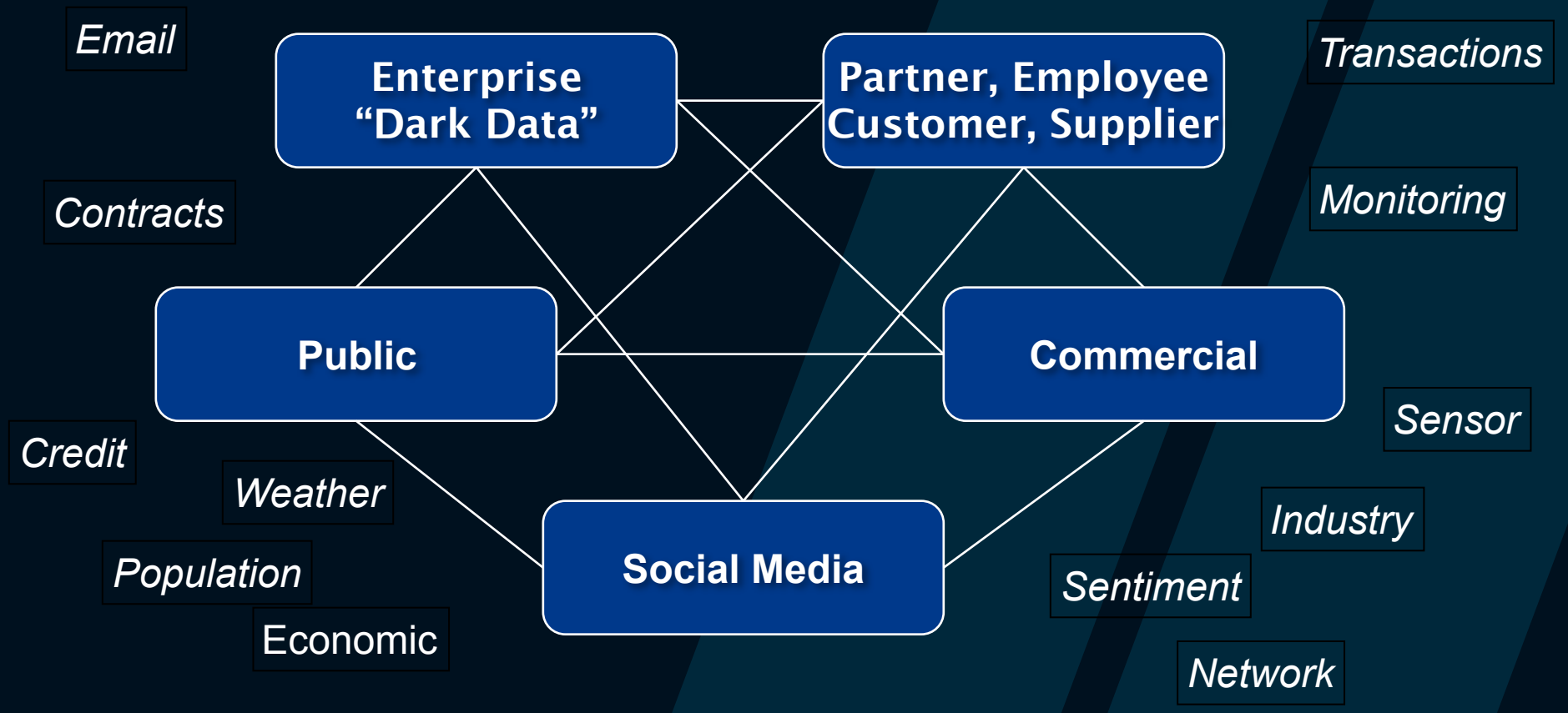
What is Big Data? continued

- Companies leverage data to adapt products and services to:
  - Meet customer needs
  - Optimize operations
  - Optimize infrastructure
  - Find new sources of revenue
  - Can reveal more patterns and anomalies

Overview


Big Data Analytics
Enterprise Data Warehousing
Business Intelligence
Predictive Analytics
Real-time Streaming Analytics
Big Data Analytics
Time Series Analysis
Simulation
Sentiment Analysis
Unsupervised Learning
Social Network Analysis
Signal Processing
Predictive Modelling
Supervised learning
Regression Analysis
Pattern Recognition
Statistics
Spatial Analysis
Visualization

Where does Big Data come from?



Why is Big Data So Important?

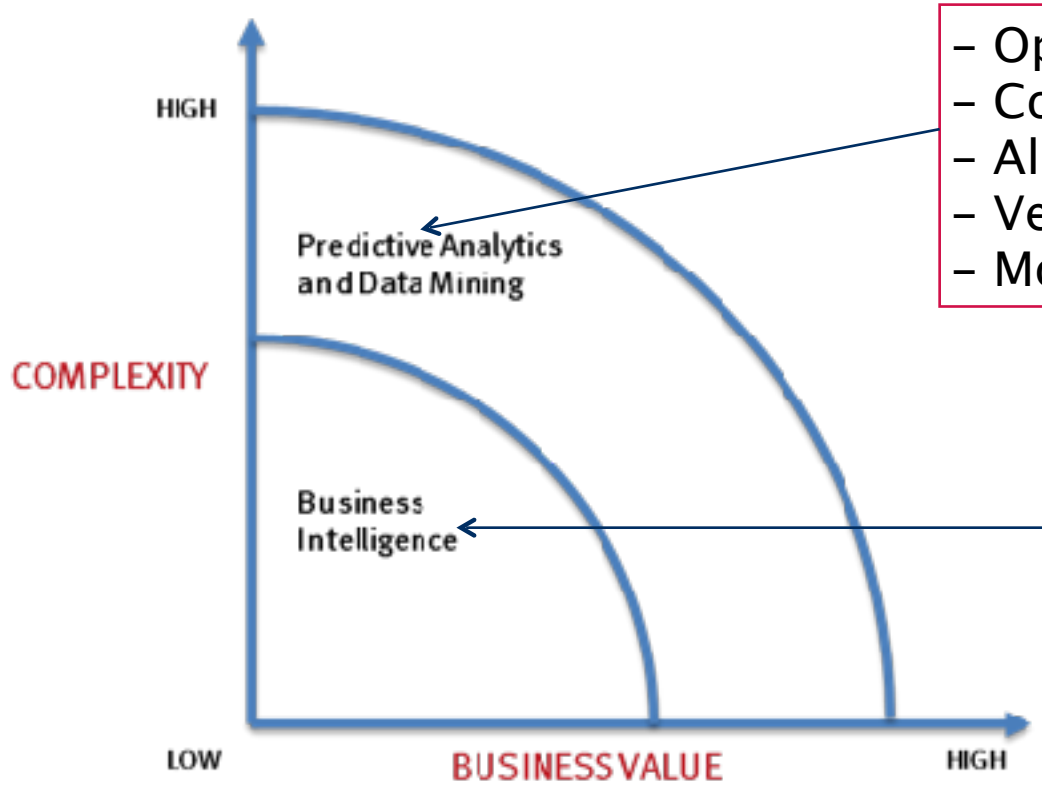
Here are a few of the most important ways big data can transform an organization:

- **Business intelligence:** weapon in the fight for the modern market
- **Innovation** : big data is used to drive new, creative products and tools to market
- **Lowered cost of ownership:** IT professionals measure operations not by the price tags on equipment, but on a variety of factors, including annual contracts, licensing, and personnel overhead.

Types of Data

- When collecting or gathering data we collect data from individuals cases on particular variables.
- A *variable* is a unit of data collection whose value can vary.
- Variables can be defined into *types* according to the level of mathematical scaling that can be carried out on the data.
- There are four types of data or levels of measurement:
  1. **Categorical (Nominal)** [customer's location (America, Europe, Asia)]
  2. **Ordinal** [college football rankings]
  3. **Interval** [Scale data]
  4. **Ratio** [Scale data]

What's driving Big Data

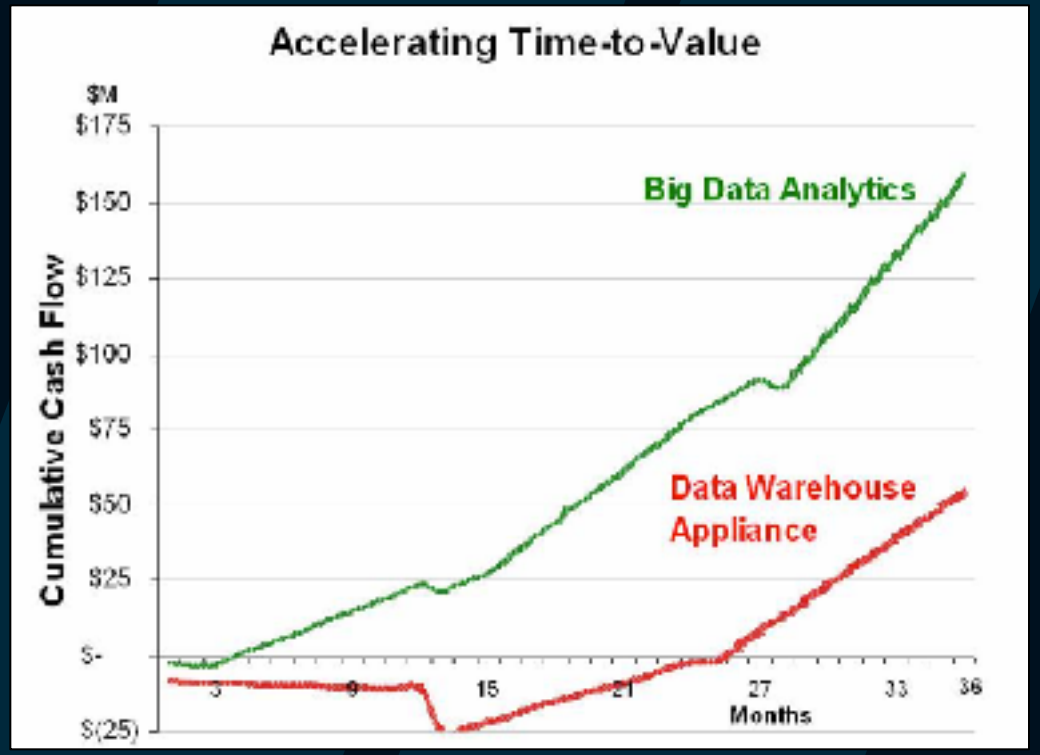


- Optimizations and predictive analytics
- Complex statistical analysis
- All types of data, and many sources
- Very large datasets
- More of a real-time

- Ad-hoc querying and reporting
- Data mining techniques
- Structured data, typical sources
- Small to mid-size datasets

Value of Big Data Analytics

- Big data is more real-time in nature than traditional DW applications
- Traditional DW architectures (e.g. Exadata, Teradata) are not well-suited for big data apps
- Shared nothing, massively parallel processing, scale out architectures are well-suited for big data apps



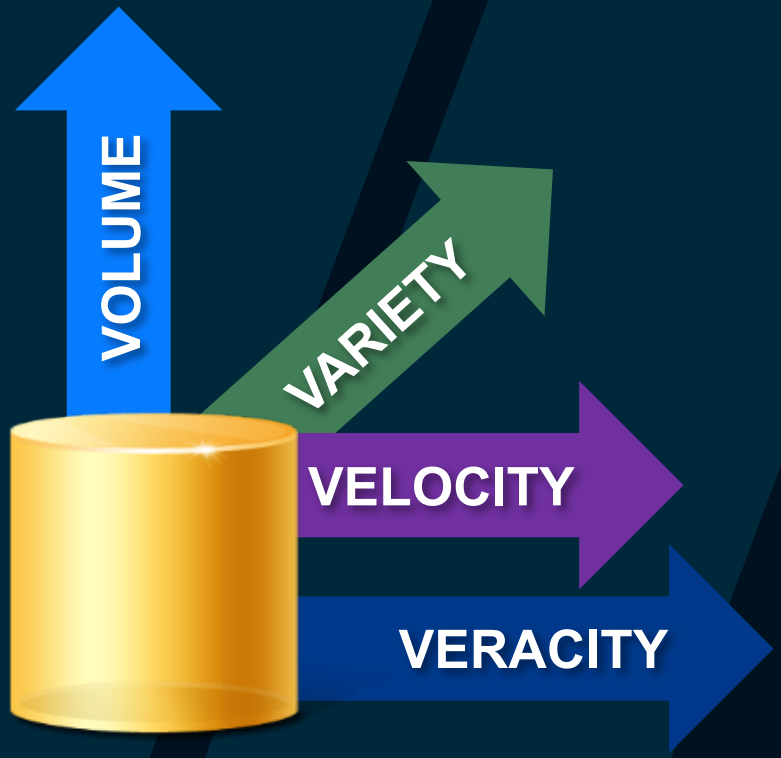
Big Data Characteristics

Growing quantity of data
e.g. social media, behavioral, video

Quickening speed of data
e.g. smart meters, process monitoring

Increase in types of data
e.g. app data, unstructured data

Need of veracity in the use of data
e.g. developing trust



Big Data Characteristics (Count..)

Big data gives you the ability to achieve superior value from analytics on

- data at higher volumes,
- velocities,
- varieties
- veracities.

With higher data volumes, you can take a more holistic view of your subject's past, present and likely future.

The FOUR V's of Big Data

From traffic patterns and music downloads to web history and medical records, data is recorded, stored, and analyzed to enable the technology and services that the world relies on every day. But what exactly is big data, and how can these massive amounts of data be used?

As a leader in the sector, IBM data scientists break big data into four dimensions: **Volume, Velocity, Variety and Veracity**.

Depending on the industry and organization, big data encompasses information from multiple internal and external sources such as transactions, social media, enterprise content, sensors and mobile devices. Companies can leverage data to adapt their products and services to better meet customer needs, optimize operations and infrastructure, and find new sources of revenue.

By 2015
4.4 MILLION IT JOBS
will be created globally to support big data, with 1.9 million in the United States

Volume
SCALE OF DATA

40 ZETTABYTES
[40 TRILLION GIGABYTES]
of data will be created by 2020, an increase of 300 times from 2005

It's estimated that
2.5 QUINTILLION BYTES
[2.5 TRILLION GIGABYTES]
of data are created each day

6 BILLION PEOPLE
have cell phones

WORLD POPULATION 7 BILLION

Most companies in the U.S. have at least
100 TERABYTES
[100,000 GIGABYTES]
of data stored

Variety
DIFFERENT FORMS OF DATA

As of 2013, the global size of data in healthcare was estimated to be
150 EXABYTES
[161 BILLION GIGABYTES]

By 2014, it's anticipated there will be
420 MILLION WEARABLE, WIRELESS HEALTH MONITORS

4 BILLION+ HOURS OF VIDEO
are watched on YouTube each month

30 BILLION PIECES OF CONTENT
are shared on Facebook every month

400 MILLION TWEETS
are sent per day by about 200 million monthly active users

Velocity
ANALYSIS OF STREAMING DATA

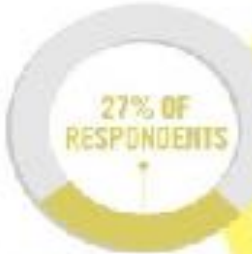
The New York Stock Exchange captures
1 TB OF TRADE INFORMATION
during each trading session

Modern cars have close to
100 SENSORS
that monitor items such as fuel level and tire pressure

By 2016, it is projected there will be
18.9 BILLION NETWORK CONNECTIONS
– almost 2.5 connections per person on earth

Veracity
UNCERTAINTY OF DATA

1 IN 3 BUSINESS LEADERS
don't trust the information they use to make decisions



A recent survey asked executives how much of their data was inaccurate

Poor data quality costs the US economy around
$3.1 TRILLION A YEAR

Sources: McKinsey Global Institute, Twitter, Cisco, Gartner, EMC, SAS, IBM, MEPTEC, QAS

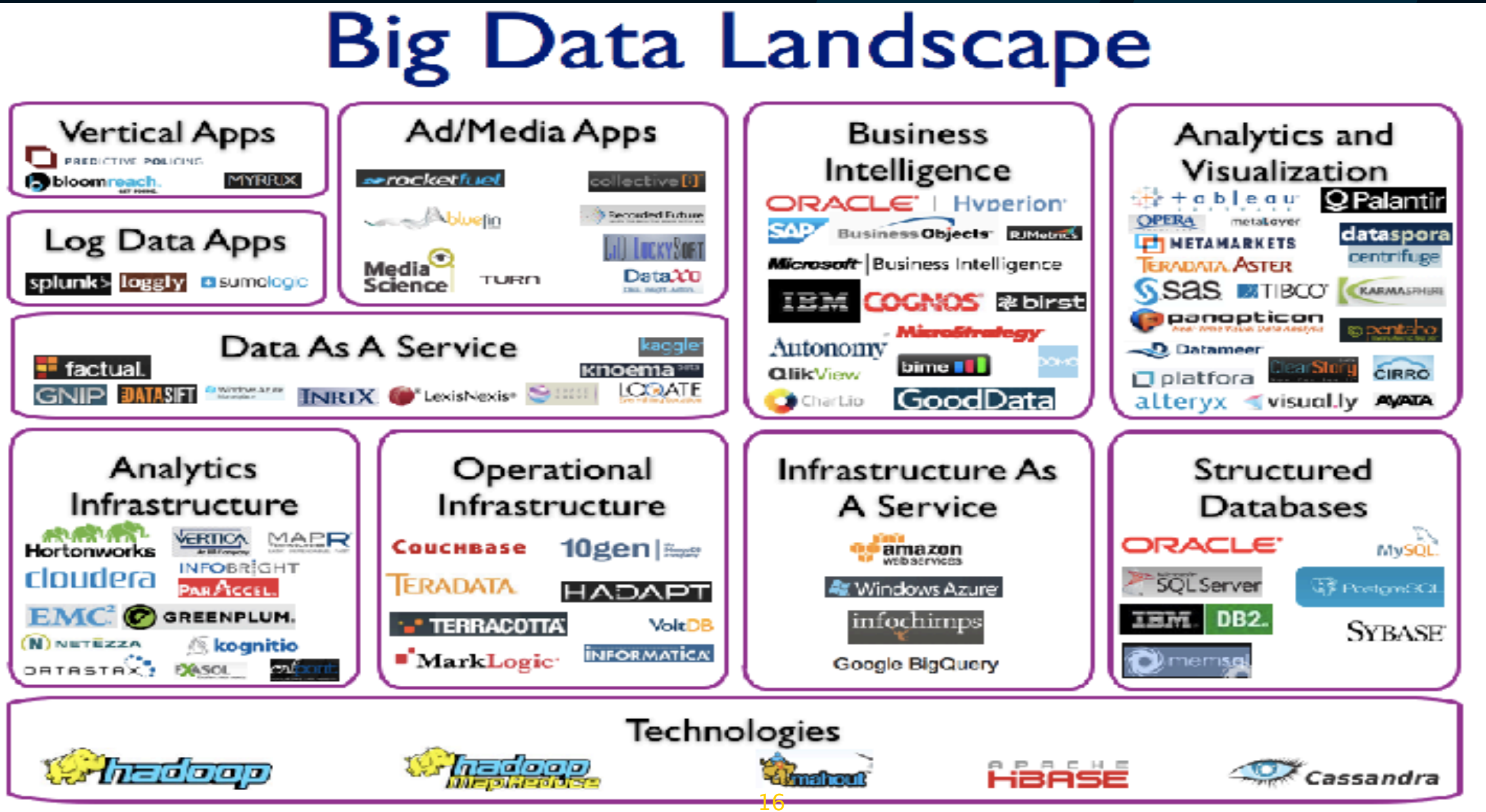
The New Analytic Paradigm

1. You will be expected to do something with information
2. There really is more to know
3. You will have to know more about knowing
4. Brain science and decision science are converging
5. The environment is changing our brain
6. Information management is the essence of leadership
7. A more connected world means much more data is available (and accessible)
8. Math matters (but so does logic and rules)
9. There are significant downsides to not knowing
10. Knowing can change the world

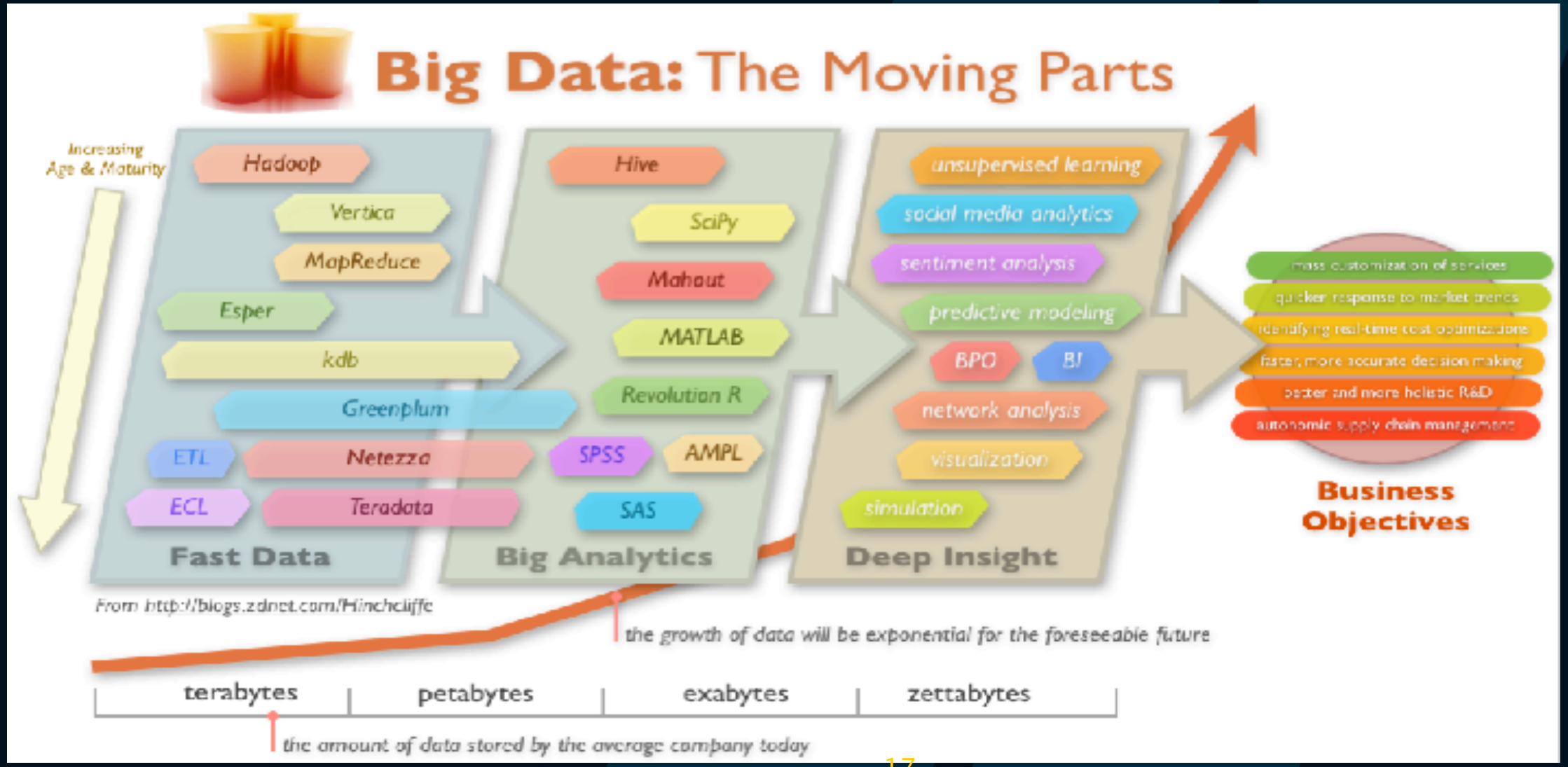
What Technology Do We Have

For Big Data ??

Big Data Landscape

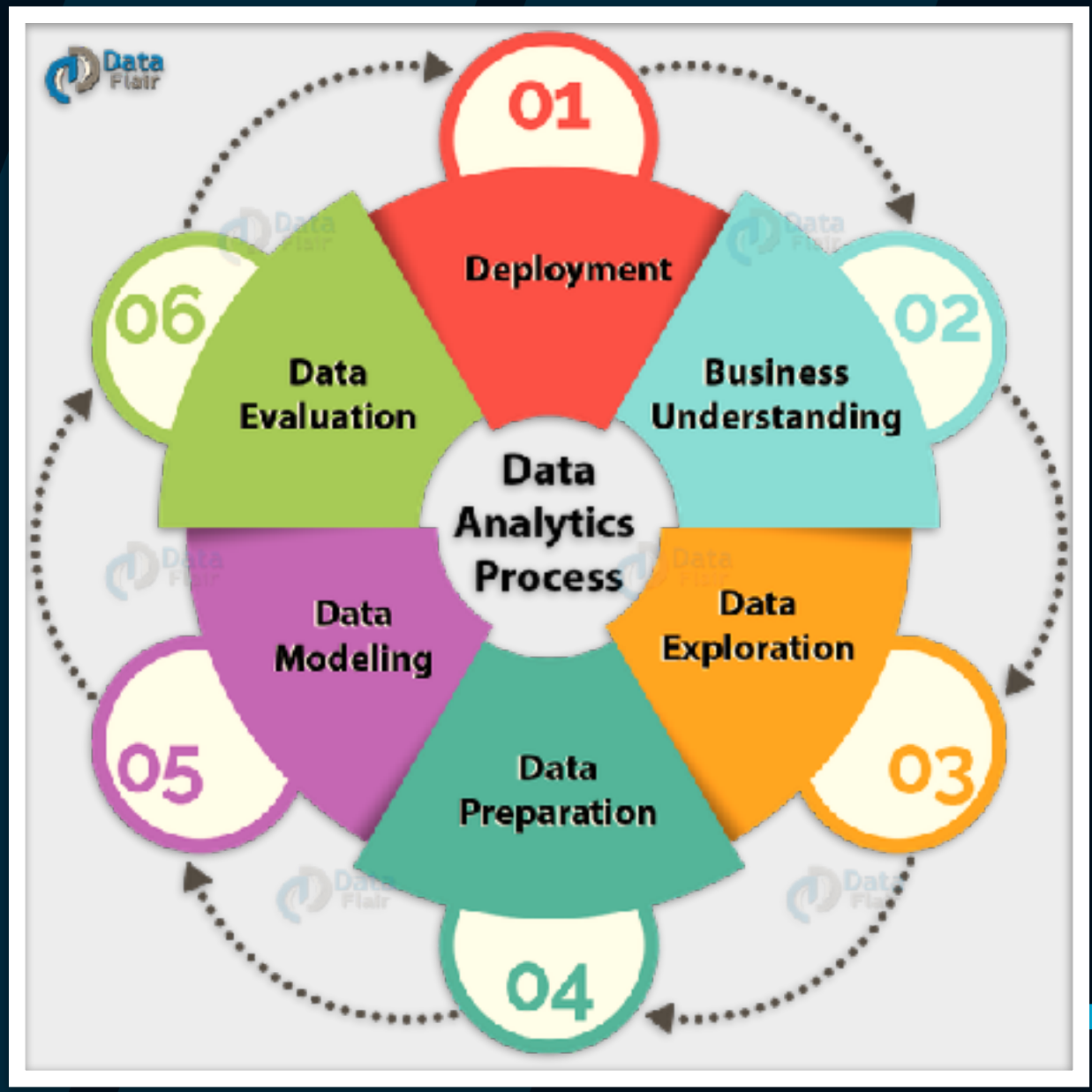


Big Data Technology



Big Data Techniques

- Association Rule Learning
- Classification Tree analysis
- Genetic algorithms
- Machine Learning
- Regression analysis
- Sentiment Analysis
- Social network analysis



Relation between big Data and other fields

HEALTH CARE	ECONOMY	DATA MINING	AGRICULTURE	HADOOP
Traditionally, the health care industry has lagged behind other industries in the use of big data Health care stakeholders now have access to promising new threads of knowledge.	Designed from the ground up to deal intelligently with commodity hardware, Hadoop can help organizations transition to low-cost servers.	Decision trees automatically help users understand what combination of data attributes result in a desired outcome..	Uses sensor data to optimize crop efficiency. The simulations allow it to discover the optimal environmental conditions for specific gene types.	In every vertical there are data tasks with which Hadoop can assist. These tasks have different terms depending on the industry but they all come down to either advanced analytics or data processing.

Big Data Trends

Big Data Trends

1. Big Data and Open Source

- Experts say that in 2017, many enterprises will expand their use of Hadoop and NoSQL technologies, as well as looking for ways to speed up their big data processing.
- Many will be seeking technologies that allow them to access and respond to data in real time.

2-Machine Learning

- As **big data analytics** capabilities have progressed, some enterprises have begun investing in machine learning (ML).

3-Big Data Intelligent Apps

- Another way that enterprises are using machine learning and AI technologies is to create intelligent apps.
- These applications often incorporate big data analytics, analyzing users' previous behaviors in order to provide personalization and better service.

4- IoT

- The Internet of Things is also likely to have a sizable impact on big data.
- According to a **September 2016 report from IDC,** "31.4 percent of organizations surveyed have launched IoT solutions, with an additional 43 percent looking to deploy in the next 12 months."

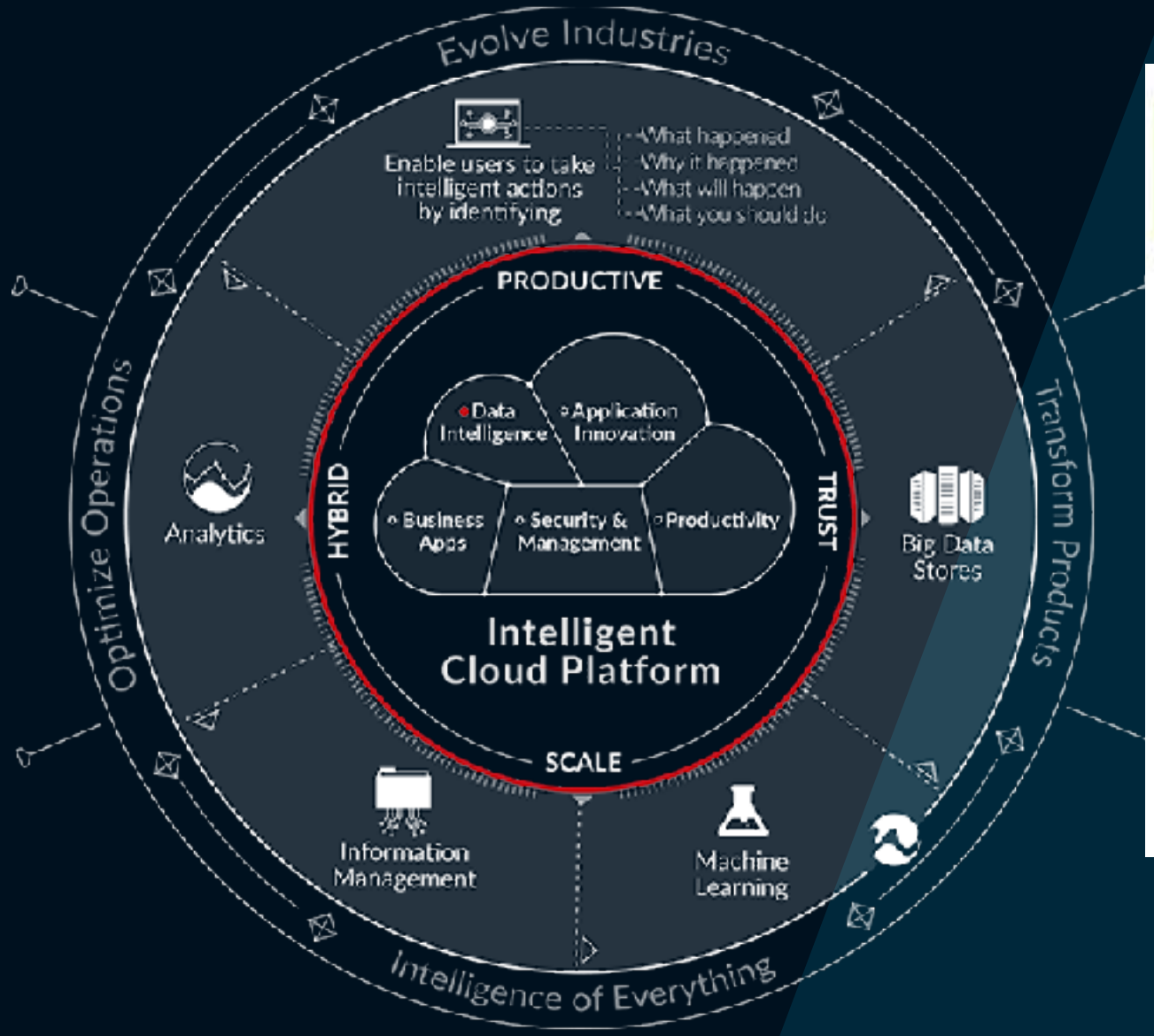
5- Edge Computing

- One new technology that could help companies deal with their IoT big data

6- Self-Service

- As the cost of hiring big experts rises, many organizations are likely to be looking for tools that allow regular business professionals to meet their own big data analytics needs.

FUTURE FOR BIG DATA



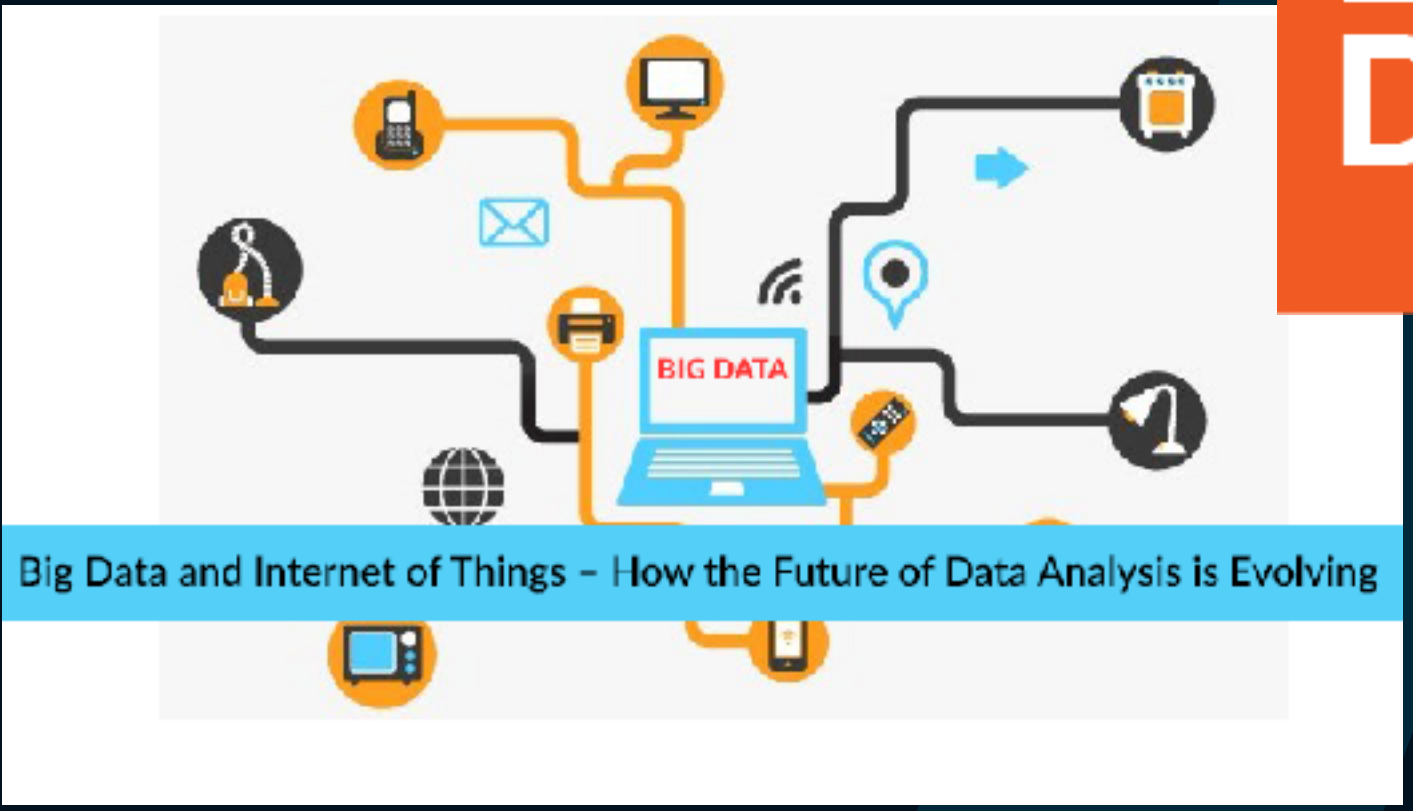
85% of fortune 500 companies are ill prepared to exploit big data.

The digital universe is doubling every 2 years, and will reach 40,000 exabytes (40 trillion gigabytes) by 2020.

By 2015, 4.4 million IT jobs globally will be created to support Big Data, generating 1.9 million IT jobs in the United States.

Fewer than 1 in 5 business report being exactly where they want to be in managing and using data.


BIG
DATA
ANALYTICS
DATA
PAYMENT
CARE
COORDINATION
COVERAGE


BIG DATA
Big Data and Internet of Things – How the Future of Data Analysis is Evolving

Real time analytics

- So Fast data is not about just volume of data like Data Warehouses in which data is measured in GigaBytes, TeraBytes or PetaBytes.
- Instead, we measure volume but concerning its incoming rate like MB per second, GB per hour, TB per day.
- So Volume and Velocity both are considered while talking about Fast Data.

Big data isn’t just an important part of the future, it may be the future itself. The way that business, organizations, and the IT professionals who support them approach their missions will continue to be shaped by evolutions in how we store, move and understand data.

Talent’s CEO


Q & A

THANK YOU!