# 15CSE334- Bigdata Analytiics

## Introduction





15CSE334- Bigdata Analytiics
A.Baskar

# Overview

- What is Big data?
- What makes data "Big"?
- Who's generating Big data ?
- Characteristics of Big data
- What is Big data Analytics?
- Hype cycle for Big data Analytics
- Challenges in handling Big data
- Difference between traditional BI and Big data
- What technology do we have for Big data analytics?
- Case study

# What is Big data?

- **No single standard definition…**

- *Extremely large data sets that may be analyzed computationally to reveal patterns, trends, and associations, especially relating to human behavior and interactions.*

- *Big data is an evolving term that describes any voluminous amount of structured, semi-structured and unstructured data that has the potential to be mined for information.*

- *"**Big Data**" is data whose scale, diversity, and complexity require new architecture, techniques, algorithms, and analytics to manage it and extract value and hidden knowledge from it…*

# What makes Data "Big"

- *The evolution of data sources over the past several years.*
  - The 1990s: Business software's –Capture sales patterns
  - The 2000s: Rapid growth of Internet
  - The 2010s: The term "Internet of Things"

  - **what makes today's data "big"?**

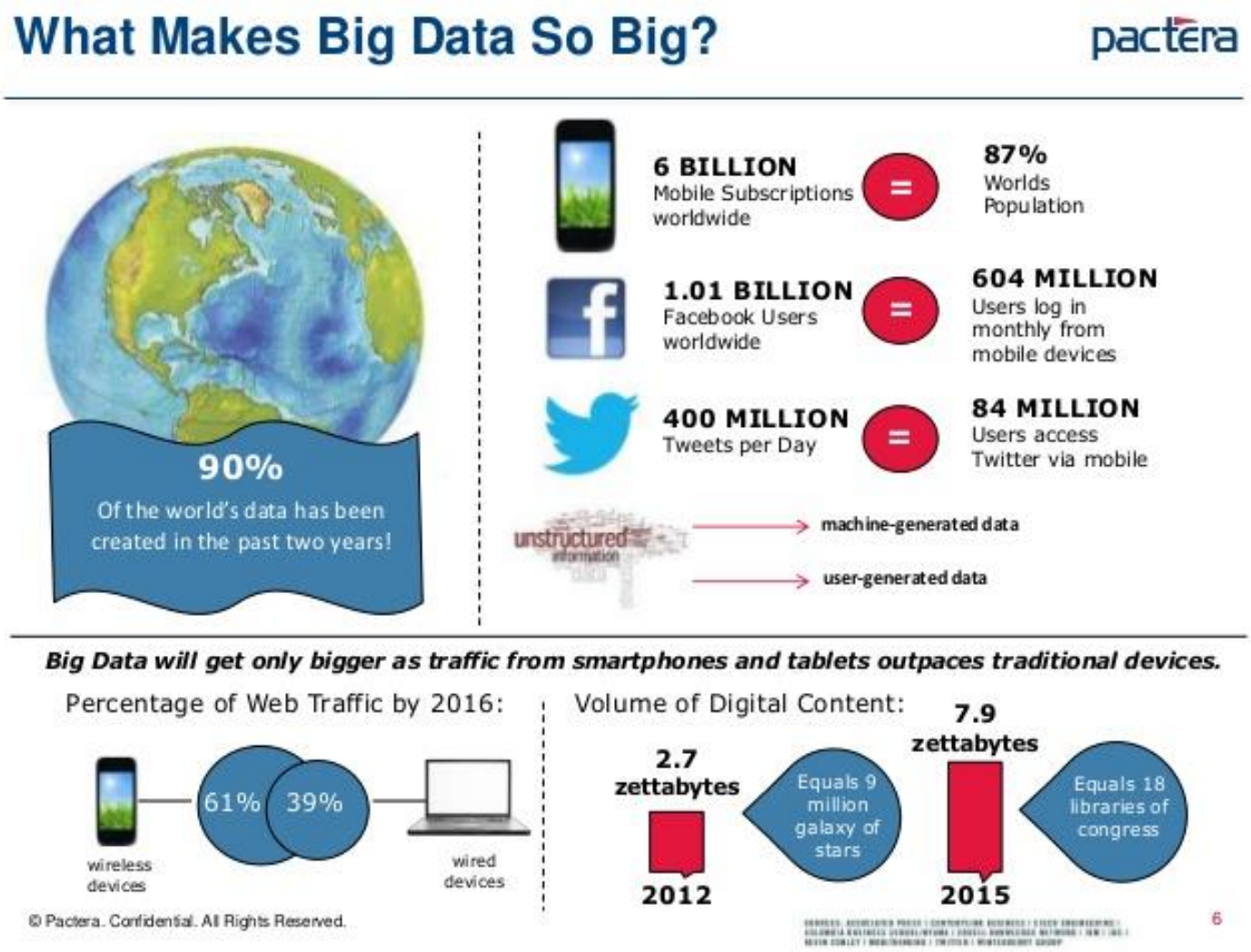
    One important consideration is the ***availability of new data sources today.***

***Examples:***

According to internetworldstats.com, there are almost 5 billion Internet users worldwide

Cisco Systems, a leading networking technology company, estimates there will be 50 billion connected devices by the year 2020, each generating data about their usage.

# What makes Data "Big"




A.Baskar

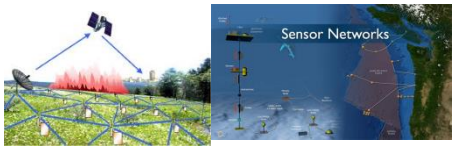
# Who's Generating Big Data


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

**Scientific instruments**
(collecting all sorts of data)


**Mobile devices**
(tracking all objects all the time

A.Baskar




**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

# The Model Has Changed…

- **The Model of Generating/Consuming Data has Changed**

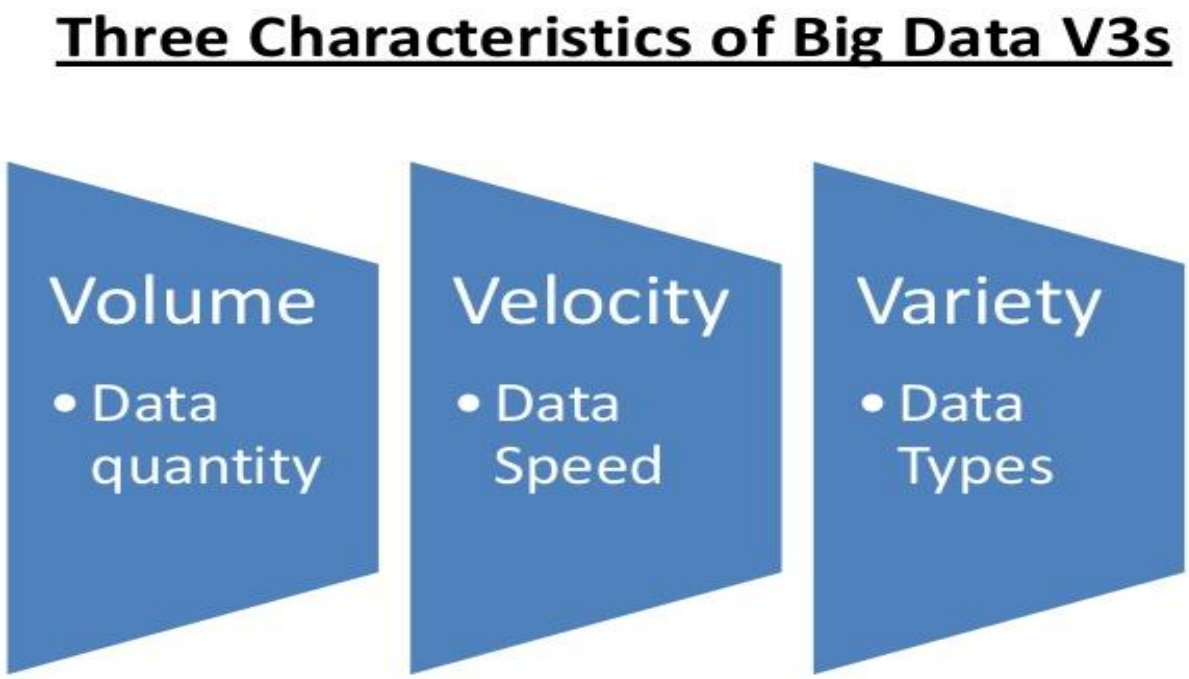
**Old Model:** Few companies are generating data, all others are consuming data




**New Model:** all of us are generating data, and all of us are consuming data

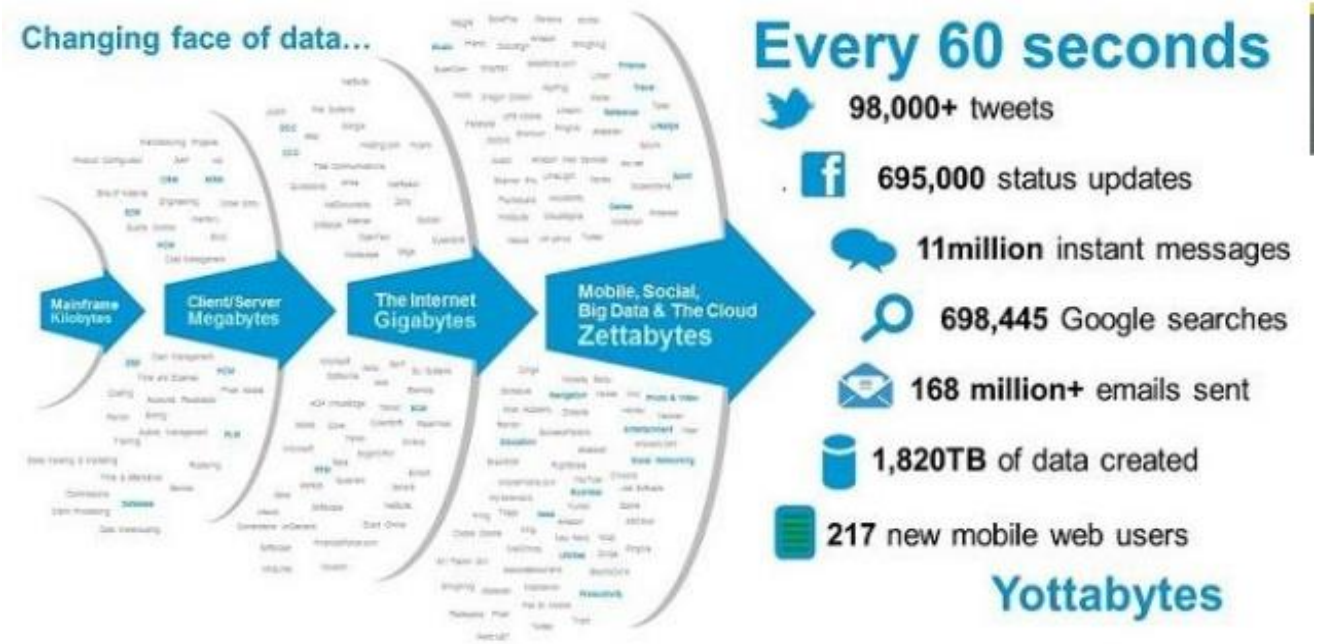


# Characteristics of Big Data



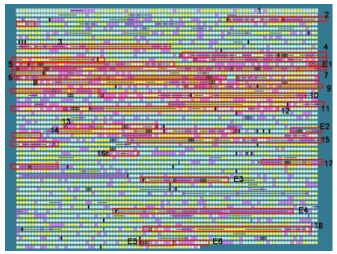
# Characteristics of Big Data

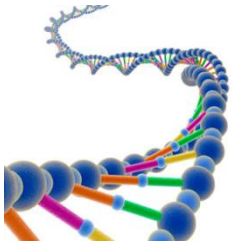
- **Data Volume**
  - 44x increase from 2009 2020
  - From 0.8 zettabytes to 35zb
- Data volume is increasing exponentially

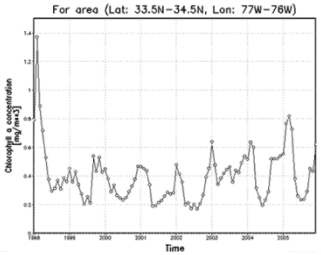


# Characteristics of Big Data: Variety

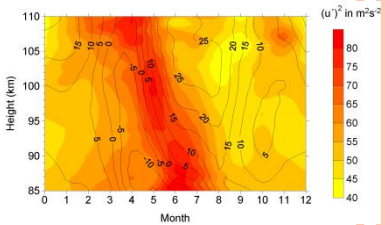
- Various formats, types, and structures
- Text, numerical, images, audio, video, sequences, time series, social media data, multi-dim arrays, etc…
- Static data vs. streaming data
- A single application can be generating/collecting many types of data








A.Baskar


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

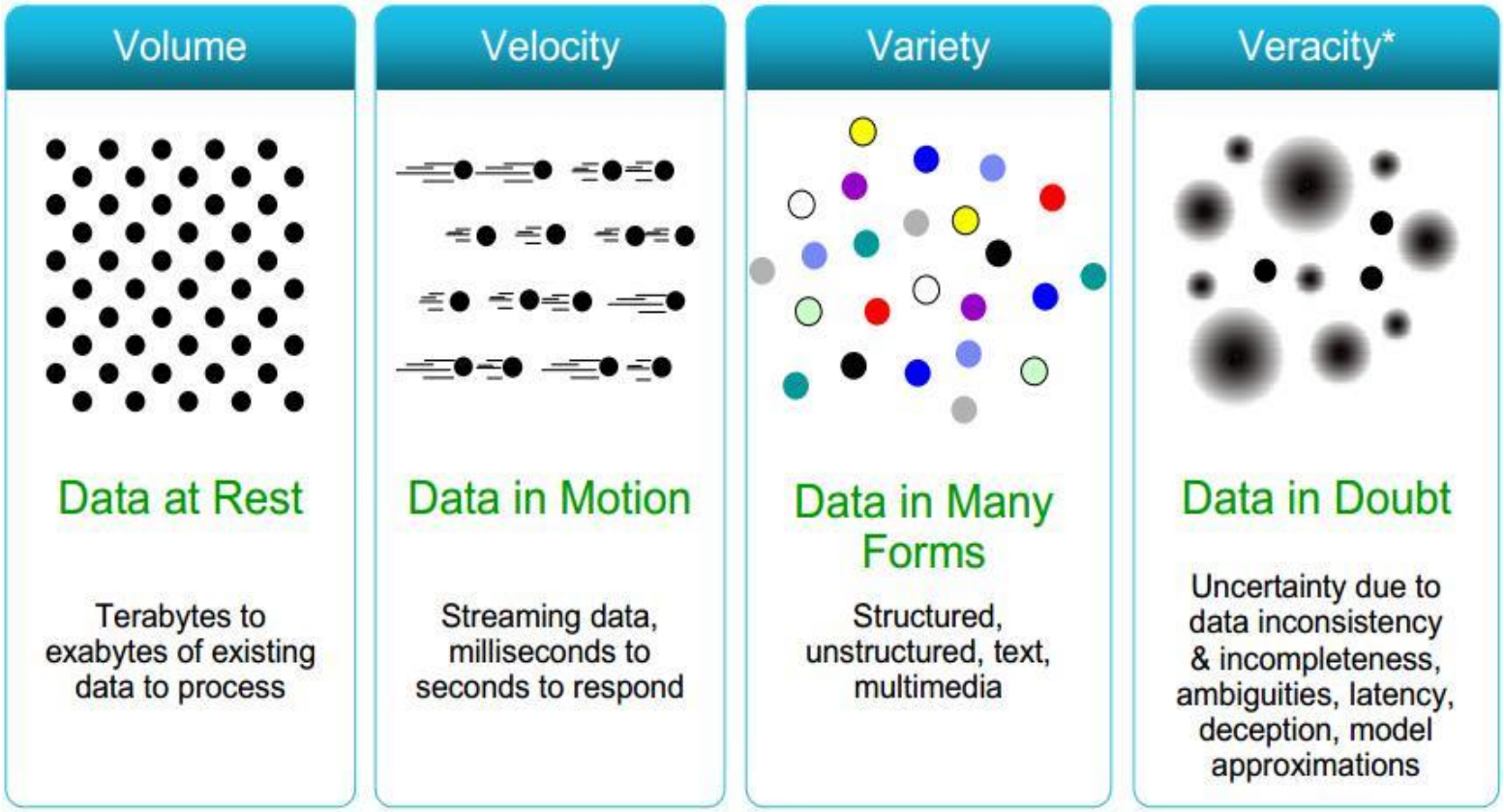
# Characteristics of Big Data: Velocity

- 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

A.Baskar

# Some Make it 4V's




A.Baskar

# VERACITY

## Big Data - Veracity

**Veracity**

1 in 3 Business Leaders don't trust the data they use for decisions

Source: IBM

Further challenge as variety and number of sources grows

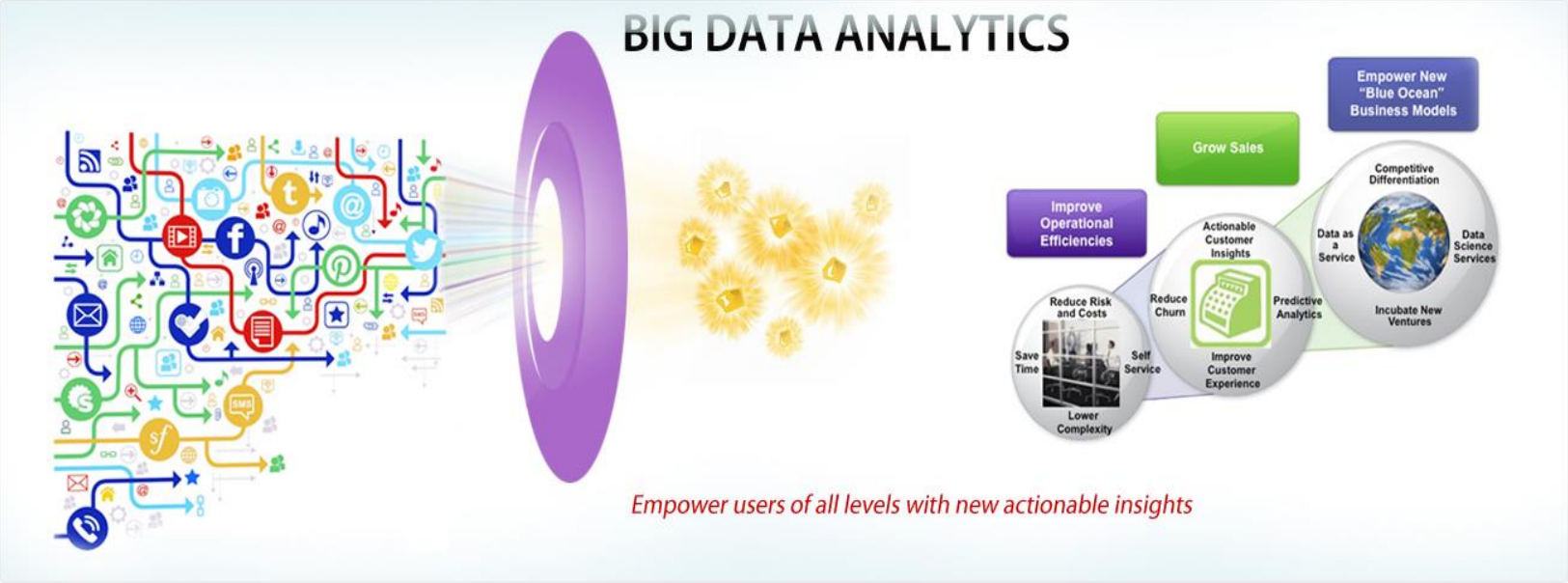

8

monster.ie

A.Baskar

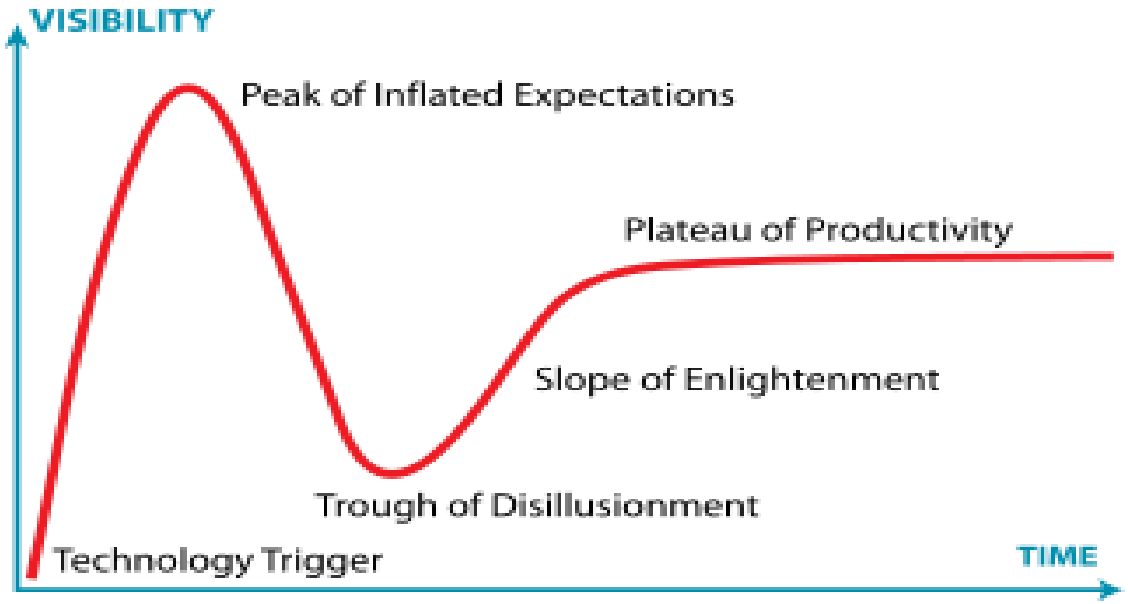
# What is Big Data analytics

- Big data analytics is the process of examining large data sets containing a variety of data types -- i.e., big data -- to uncover hidden patterns, unknown correlations, market trends, customer preferences and other useful business information.



# Gartner Hype cycle

- *The **Hype Cycle** is a branded graphical presentation developed and used by American Information **Technology** (IT) research and advisory firm **Gartner** for representing the maturity, adoption and social application of specific **technologies**.*

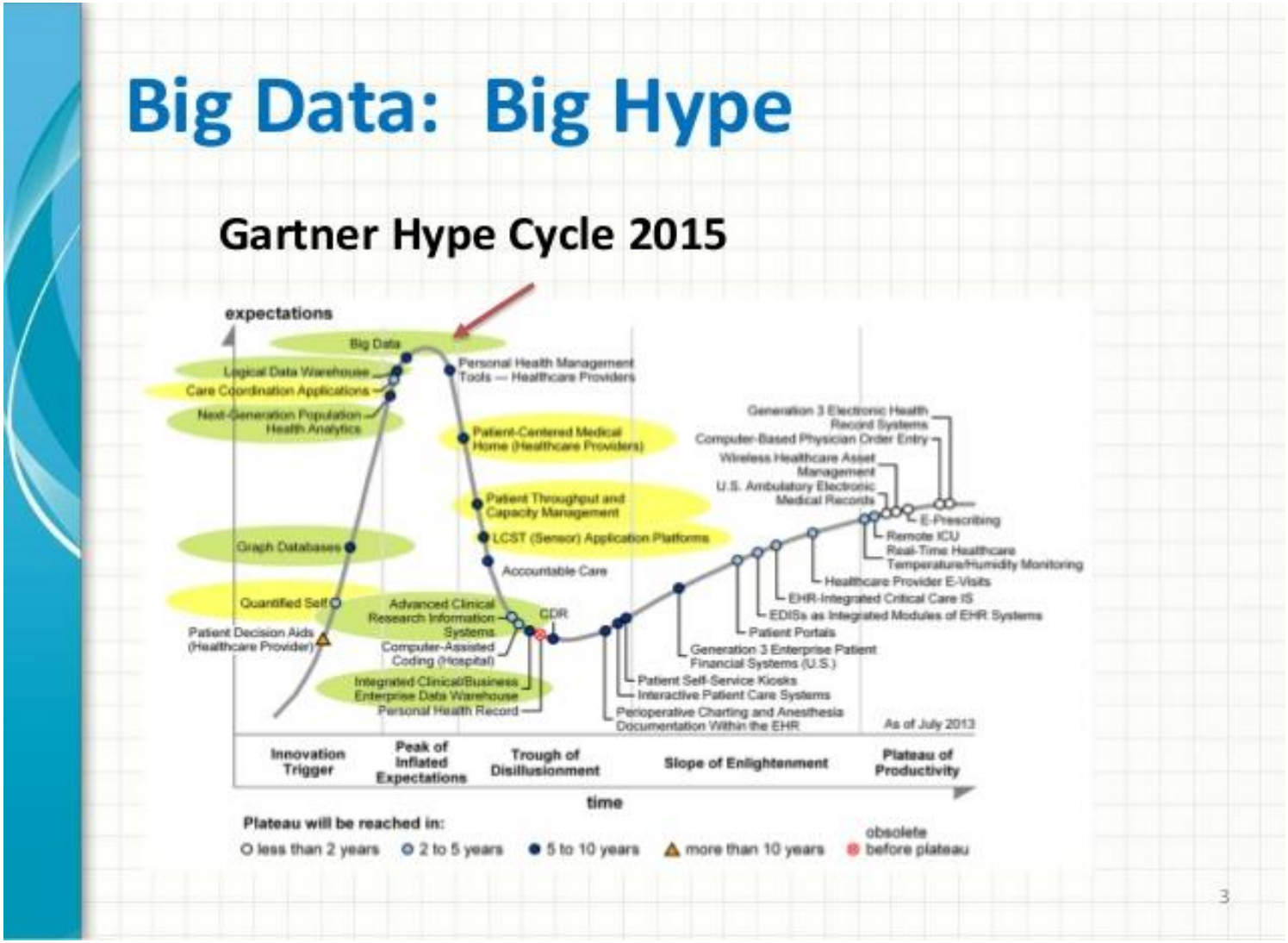



A.Baskar

# Gartner Hype cycle

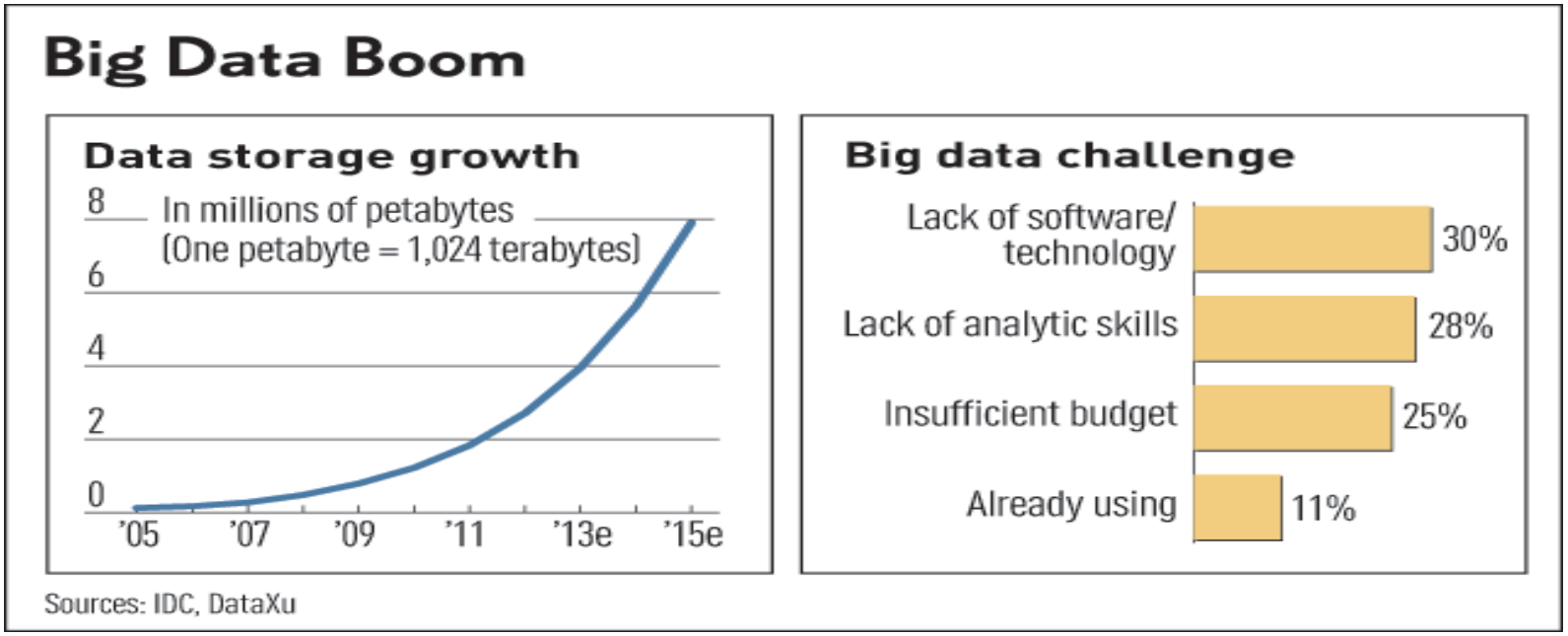
| No. | Phase | Description |
|---|---|---|
| **1** | Technology Trigger | A potential technology breakthrough kicks things off. Early proof-of-concept stories and media interest trigger **significant publicity**. Often no usable products exist and commercial viability is unproven. |
| **2** | Peak of Inflated Expectations | Early publicity produces a **number of success stories**—often accompanied by scores of failures. Some companies take action; most don't. |
| **3** | Trough of Disillusionment | Interest wanes as **experiments and implementations fail to deliver**. Producers of the technology shake out or fail. Investments continue only if the surviving providers improve their products to the satisfaction of early adopters. |
| **4** | Slope of Enlightenment | More instances of how the **technology can benefit the enterprise** start to crystallize and become more widely understood. Second- and third-generation products appear from technology providers. More enterprises fund pilots; conservative companies remain cautious. |
| **5** | Plateau of Productivity | **Mainstream adoption starts to take off**. Criteria for assessing provider viability are more clearly defined. The technology's broad market applicability and relevance are clearly paying |

# Gartner Hype cycle




A.Baskar

# Challenges in Handling Big Data




- **The Bottleneck is in technology**
  - New architecture, algorithms, techniques are needed
- **Also in technical skills**
  - Experts in using the new technology and dealing with big data

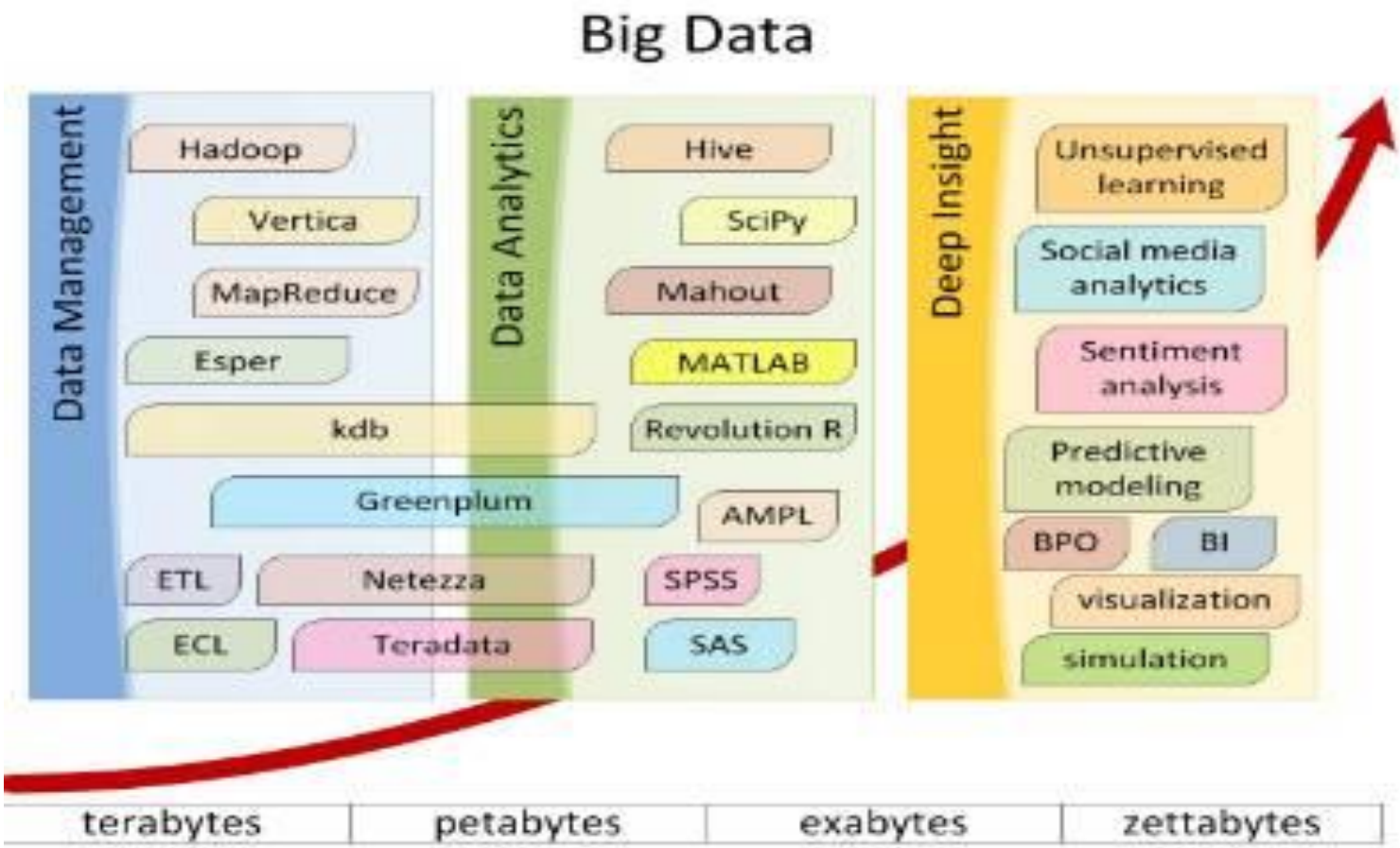
A.Baskar

# Difference between traditional BI and Big data

| | Traditional Analytics (BI) | vs Big Data Analytics |
|---|---|---|
| Focus on | • Descriptive analytics<br>• Diagnosis analytics | • **Predictive analytics**<br>• **Data Science** |
| Data Sets | • Limited data sets<br>• Cleansed data<br>• Simple models | • Large scale data sets<br>• More types of data<br>• Raw data<br>• Complex data models |
| Supports | **Causation:** what happened, and why? | **Correlation**: new insight<br>More accurate answers |

A.Baskar

# What technology do we have for Big data analytics?



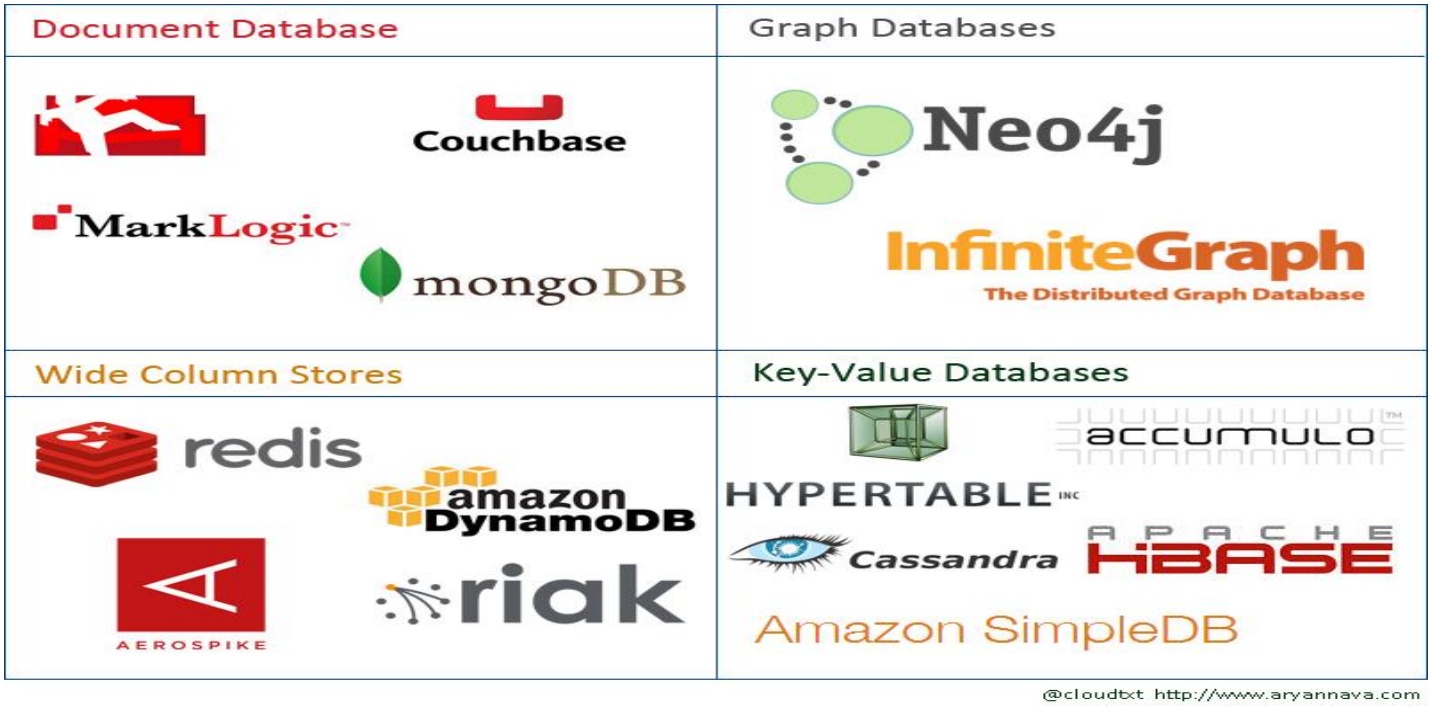

A.Baskar

# TECHNOLOGY FOR BIG DATA ANALYTICS

- **Storing Big data :**

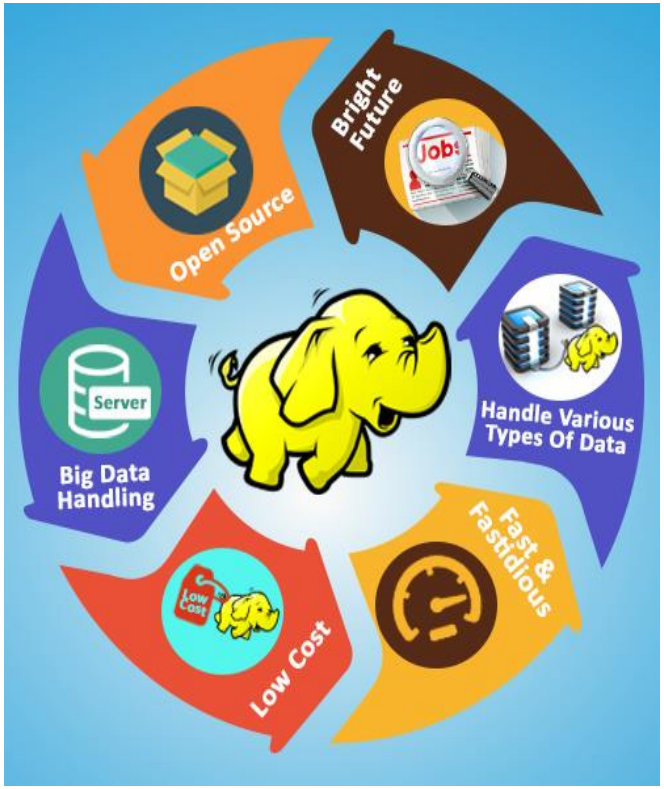
**NoSQL databases:**

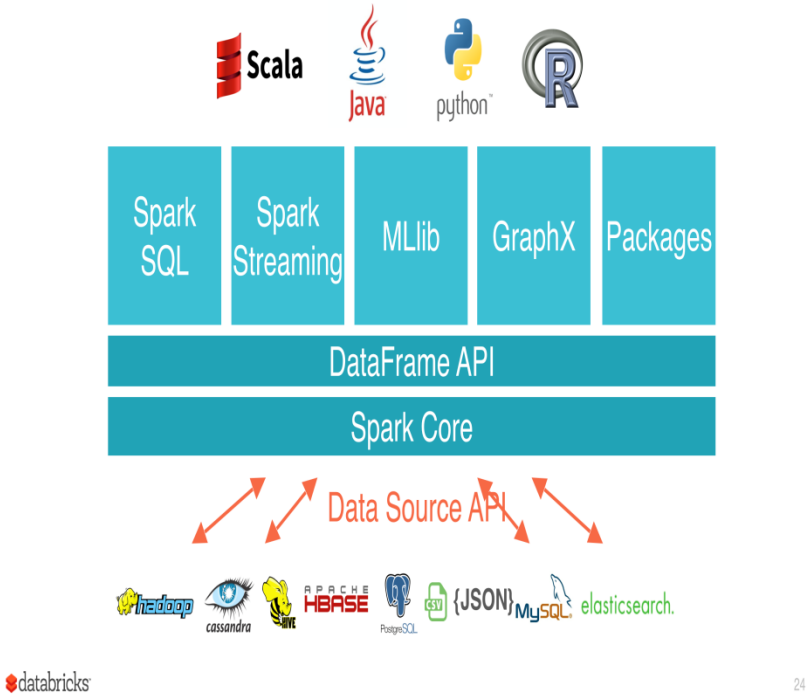
*There are several database types that fit into this category, such as key-value stores and document stores, which focus on the storage and retrieval of large volumes of unstructured, semi-structured, or even structured data.*



# TECHNOLOGY FOR BIG DATA ANALYTICS

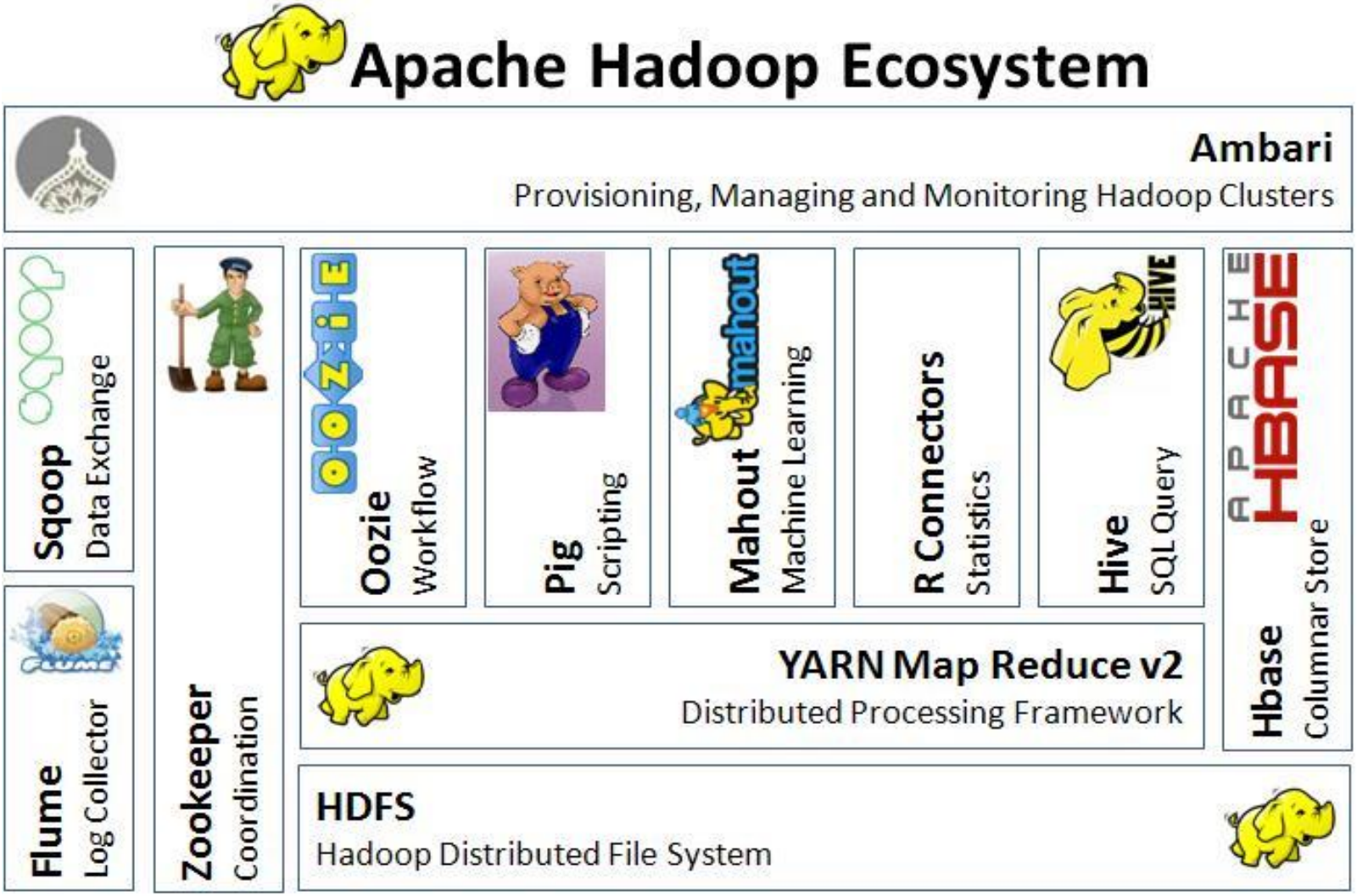
- **Data processing:**



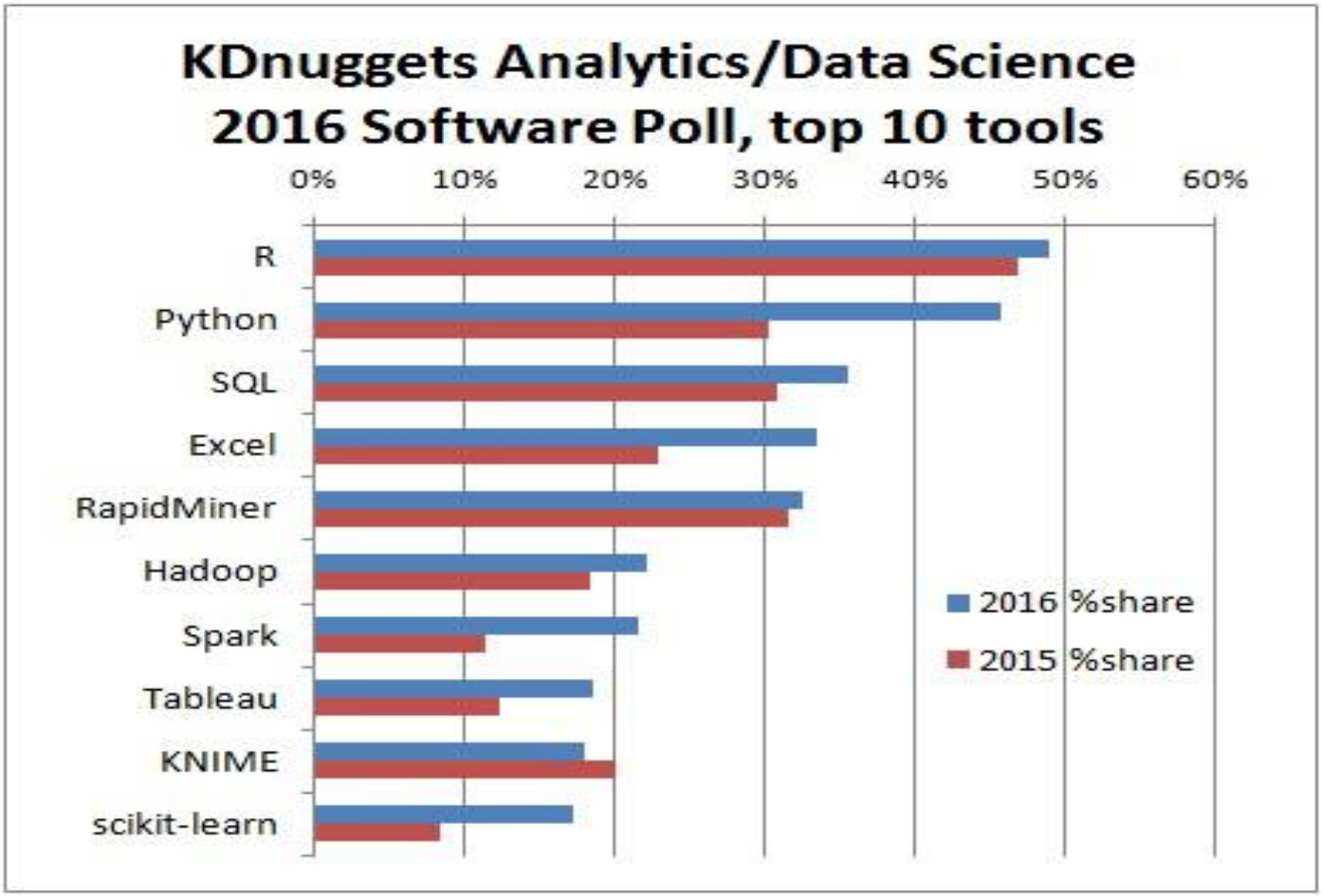

A.Baskar

# Hadoop Eco systems




A.Baskar

# Data analysis Technology for Big data

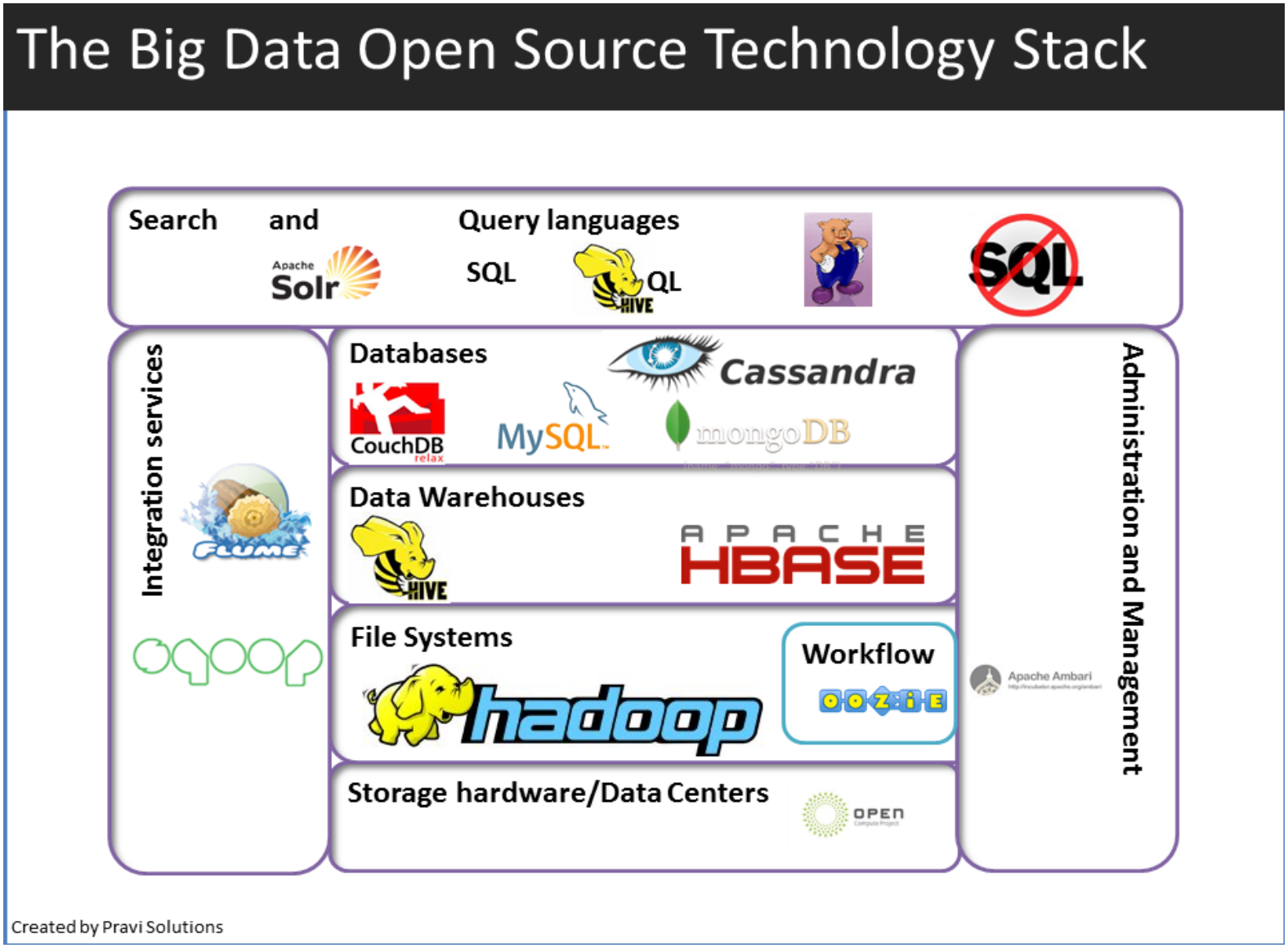



A.Baskar

# Vendors of Big data technologies

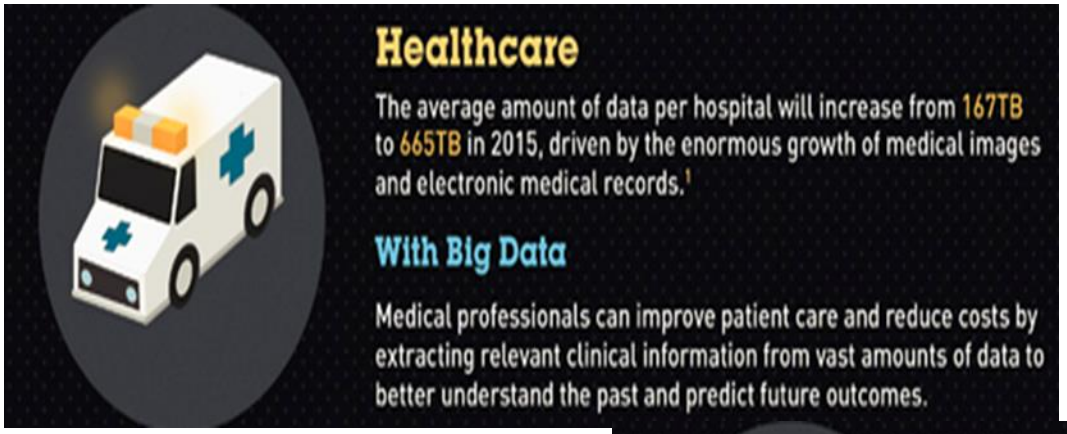

A.Baskar

# Vendors of Big data technologies




A.Baskar

# Case study



## Healthcare

The average amount of data per hospital will increase from 167TB to 665TB in 2015, driven by the enormous growth of medical images and electronic medical records.[1]

### With Big Data

Medical professionals can improve patient care and reduce costs by extracting relevant clinical information from vast amounts of data to better understand the past and predict future outcomes.

## Customer Service

Today, 86% of consumers quit doing business with a company because of a bad customer experience, up from 59% four years ago.[2]

### With Big Data

Service representatives can use data to gain a more holistic view of their customers, understanding their likes and dislikes in real-time in order to resolve a problem or capitalize on happy clients faster.

## Insurance

Insurance companies and government agencies each gather fraud data related to their own individual missions. But the kind, quality and volume of data compiled varies widely.[3]

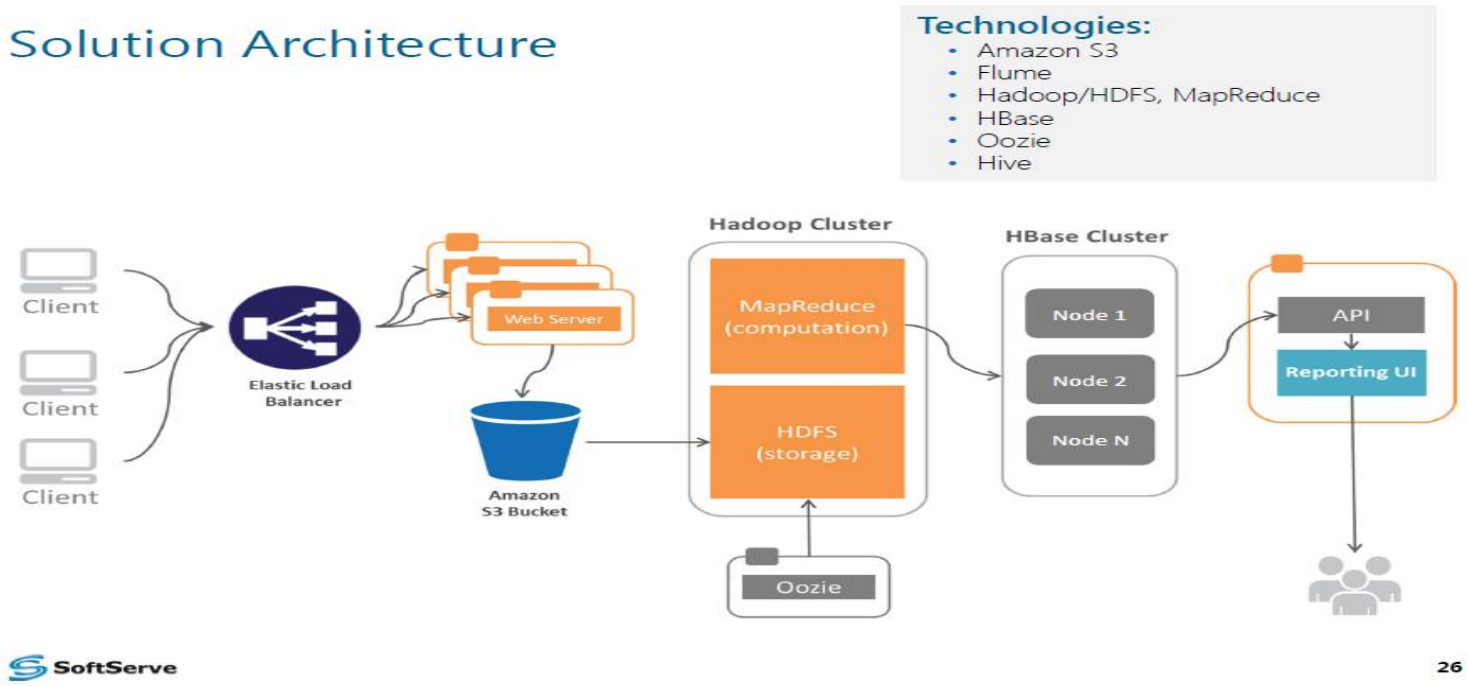
### With Big Data

An insurance or citizen services provider can apply advanced analytics to data and detect fraud quickly, before funds are paid out.

# Click stream analysis

- **Goal** :

*Provide the ability to understand how end-users are interacting with service content, products, and features on sites*




A.Baskar

# To Design Big data solutions

- Understand data users and sources
- Discover architecture drivers
- Select proper reference architecture
- Do trade-off analysis, address cons
- Map reference architecture to technology stack
- Prototype, re-evaluate architecture
- Estimate implementation efforts
- Set up devops practices from the very beginning
- Advance in solution development through "small wins"
- Be ready for changes, big data technologies are evolving rapidly