Big Data and its Importance

Introduction to Big Data

What is Big Data?

What makes data, "Big" Data?

What's Big Data?

The term "**big data**" refers to data sets so large and complex that traditional tools, like relational databases, are unable to process them in an acceptable time frame or within a reasonable cost range.

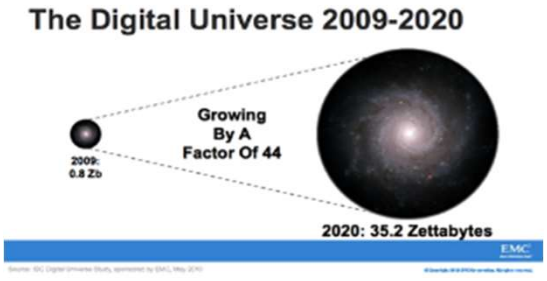
Gartner analyst Doug Laney described big data in terms of what is now known as the 3Vs:

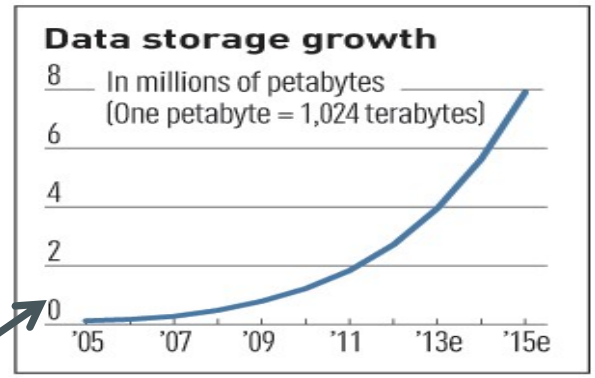
- **Volume:** The overall size of the data set
- **Velocity:** The rate at which the data arrives and also how fast it needs to be processed
- **Variety:** The wide range of data that the data set may contain—that is, web logs, audio, images, sensor or device data, and unstructured text, among many others types

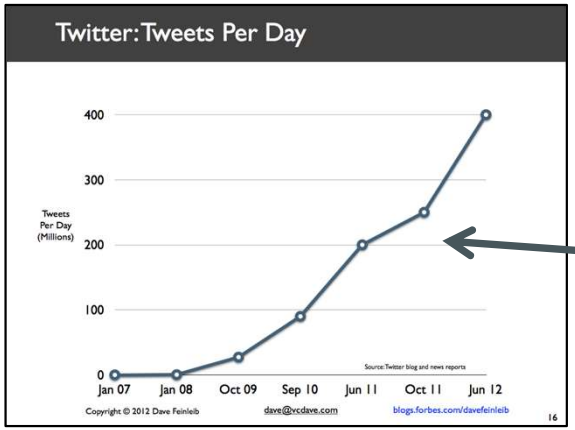
Problems occur in sourcing, moving, storing, searching, and analysing the big data, but with the right tools these problems can be addressed. A rich set of big data processing tools (provided by the Apache Software Foundation, Lucene, and third-party suppliers) is available to help the businesses to fulfil their big data needs.

Characteristics of Big Data: 1-Scale (Volume)

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



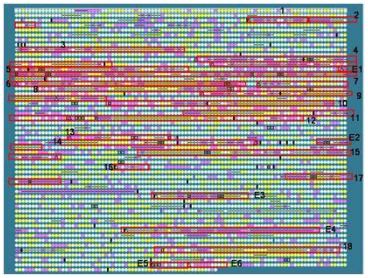


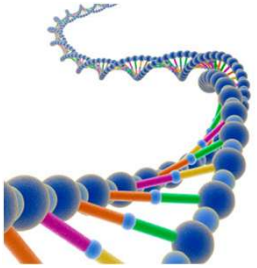


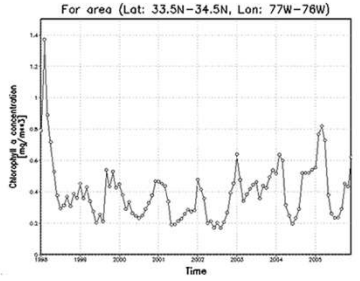
Exponential increase in collected/generated data

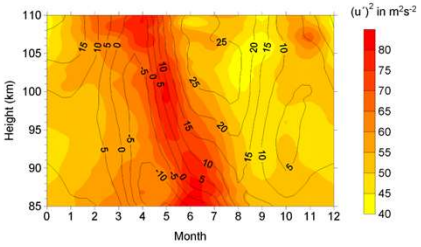
Characteristics of Big Data: 2-Complexity (Varity)

- 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









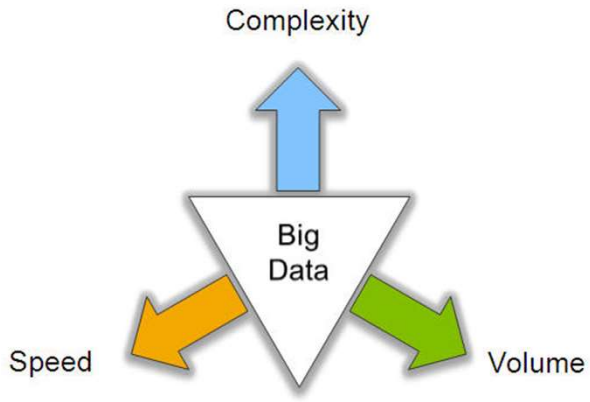
To extract knowledge➔ all these types of data need to linked together

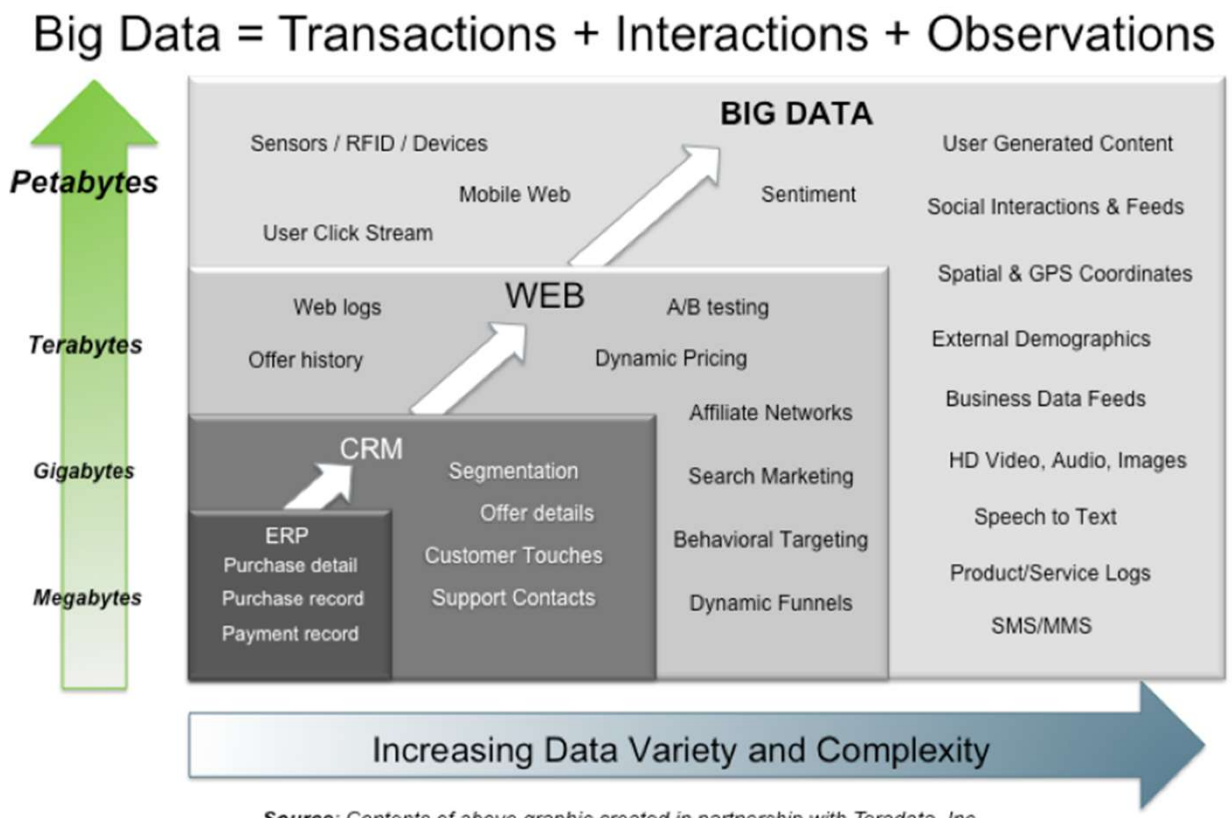
Characteristics of Big Data: 3-Speed (Velocity)

- 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

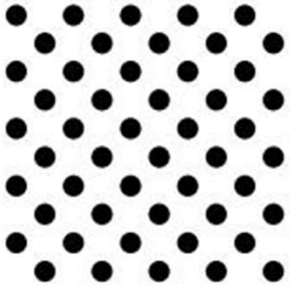
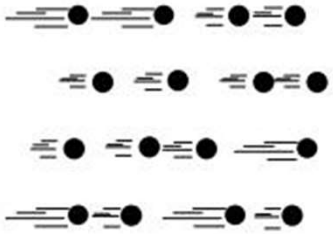
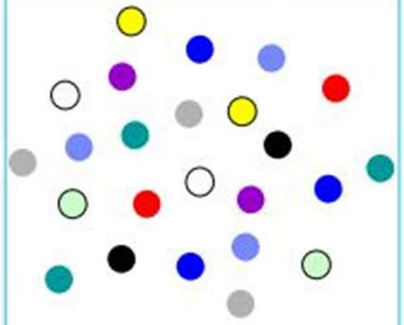
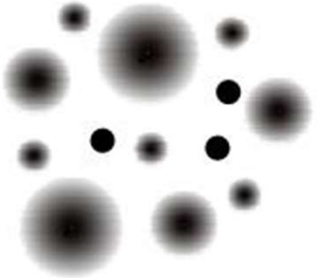
Big Data: 3V's



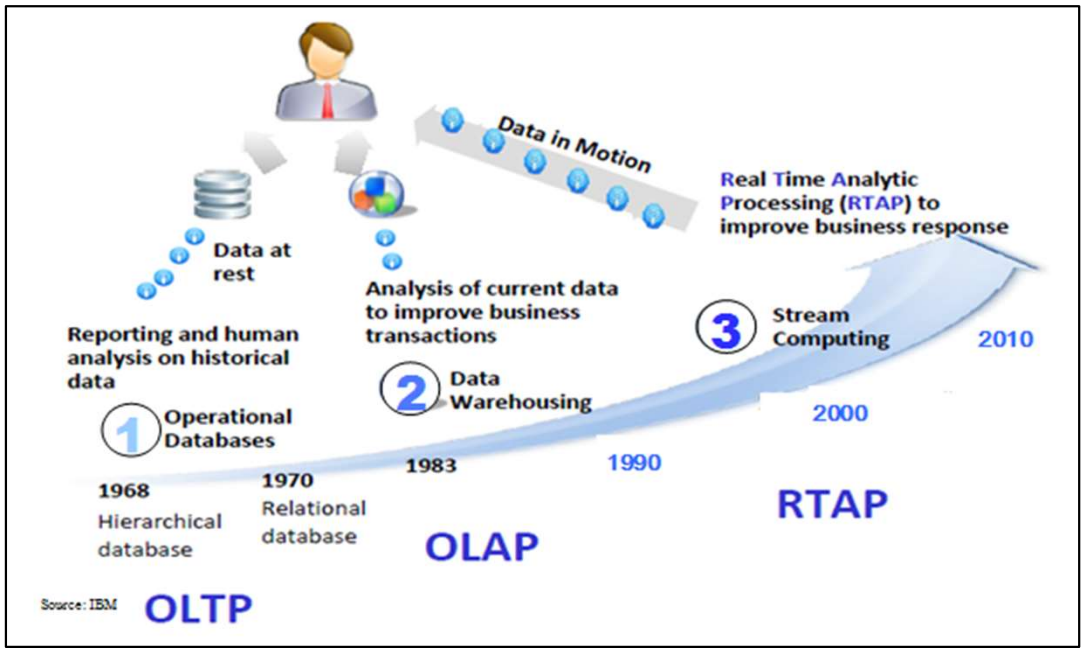


***Source**: Contents of above graphic created in partnership with Teradata, Inc.*

Some Make it 4V's

Volume	Velocity	Variety	Veracity*
			
Data at Rest	**Data in Motion**	**Data in Many Forms**	**Data in Doubt**
Terabytes to exabytes of existing data to process	Streaming data, milliseconds to seconds to respond	Structured, unstructured, text, multimedia	Uncertainty due to data inconsistency & incompleteness, ambiguities, latency, deception, model approximations

Harnessing Big Data



- **OLTP:** Online Transaction Processing (DBMSs)
- **OLAP:** Online Analytical Processing (Data Warehousing)
- **RTAP:** Real-Time Analytics Processing (Big Data Architecture & technology)

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)

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

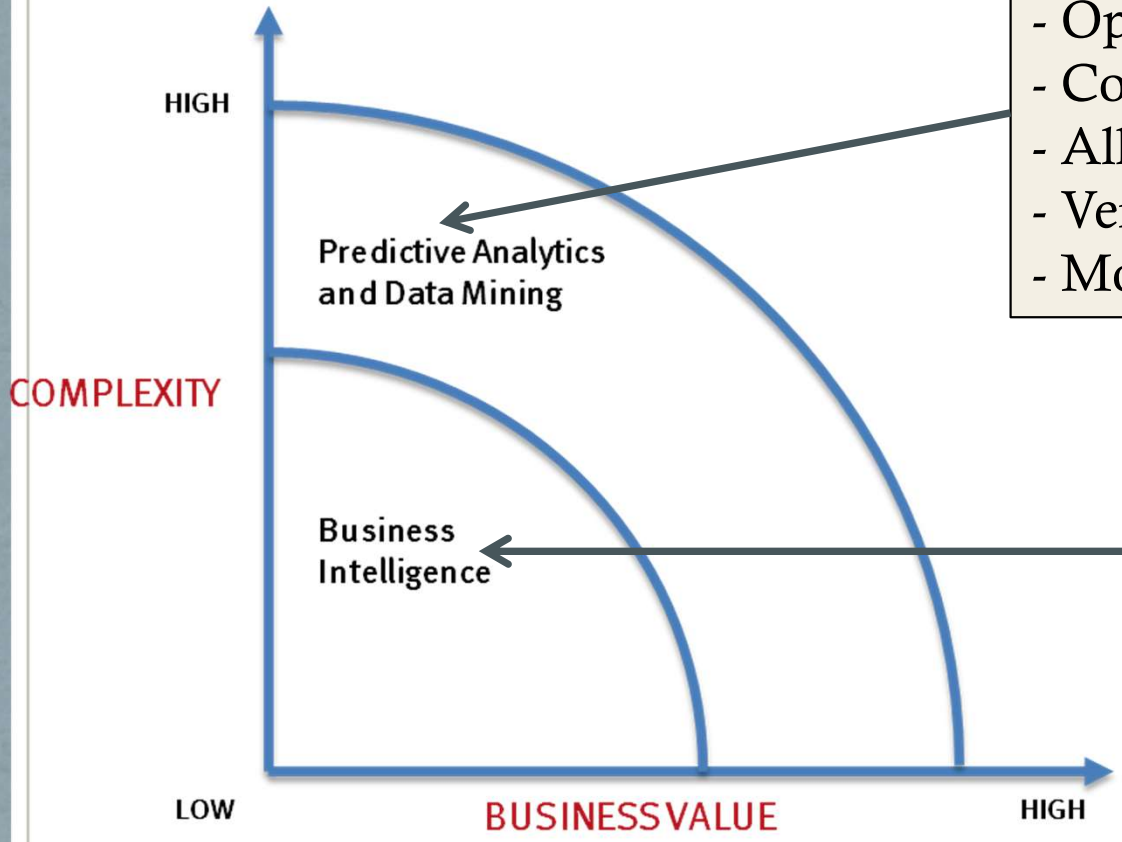
The Model Has Changed…

- **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

What's driving Big Data

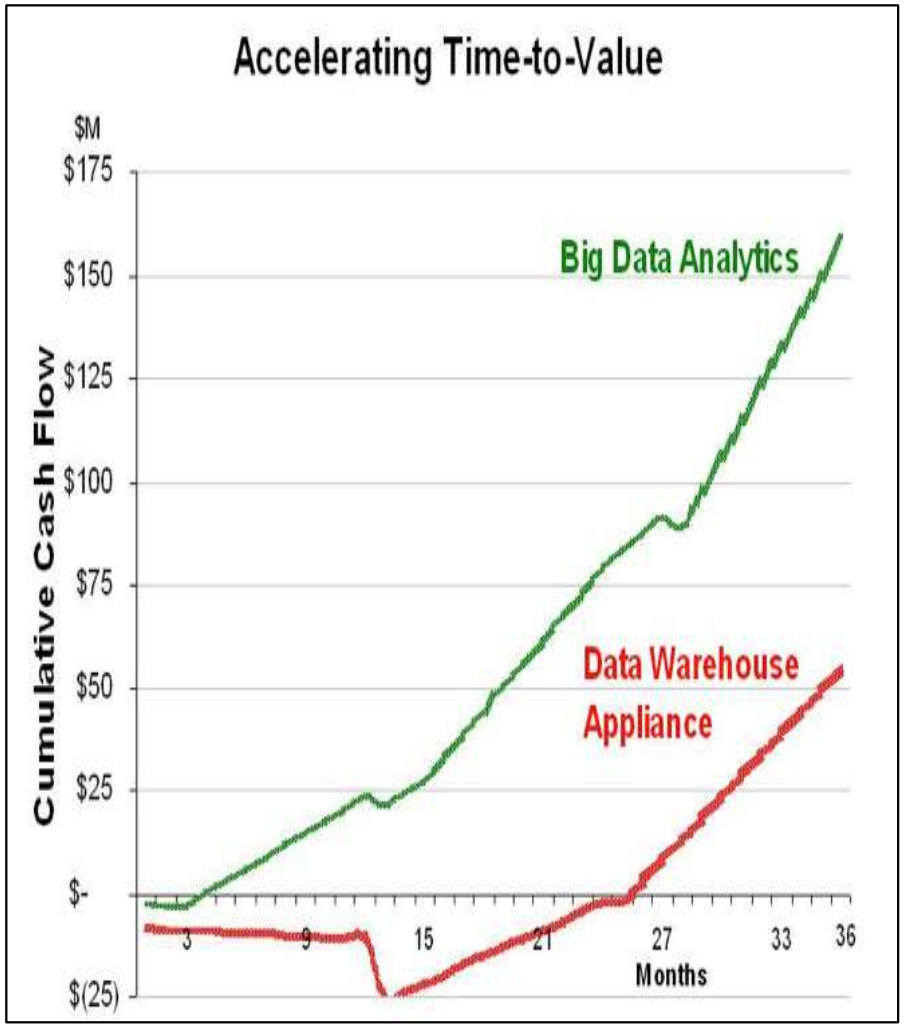


- Optimizations and predictive analytics
- Complex statistical analysis
- All types of data, and many sources
- Very large datasets
- More of a real-time

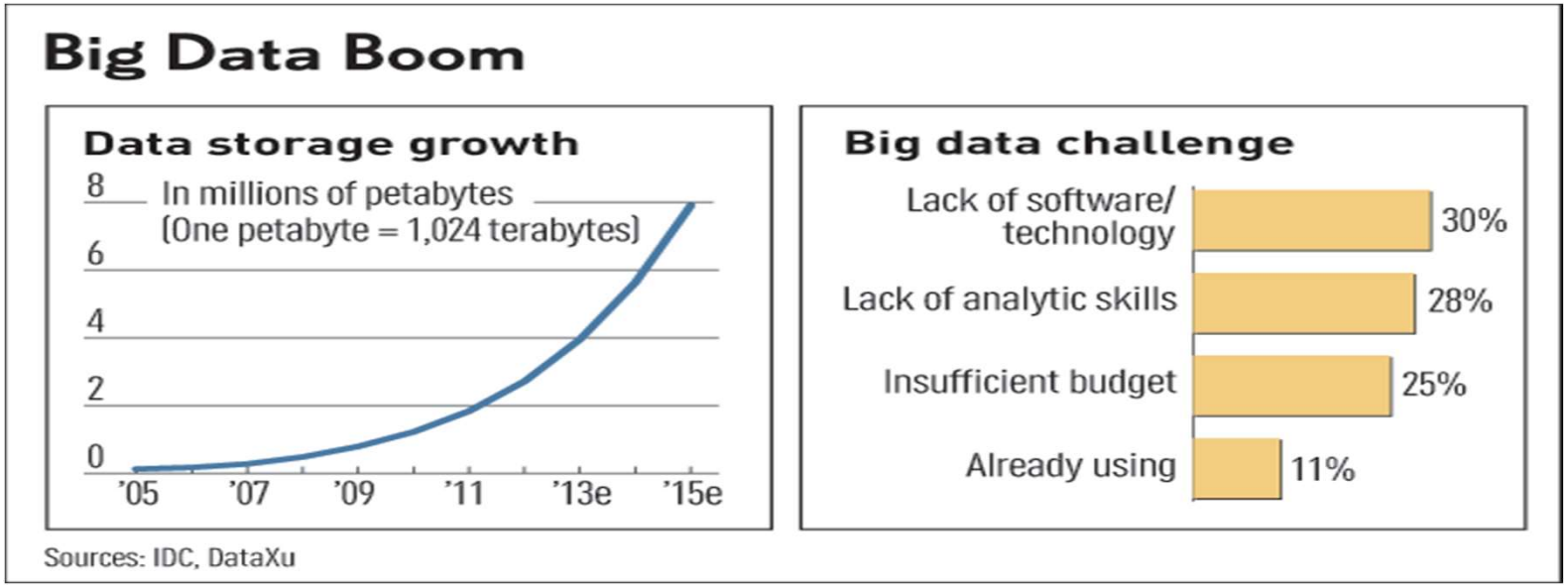
- Ad-hoc querying and reporting
- Data mining techniques
- Structured data, typical sources
- Small to mid-size datasets

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 analytics refers to managing very large amounts of data and extracting value and knowledge from them.



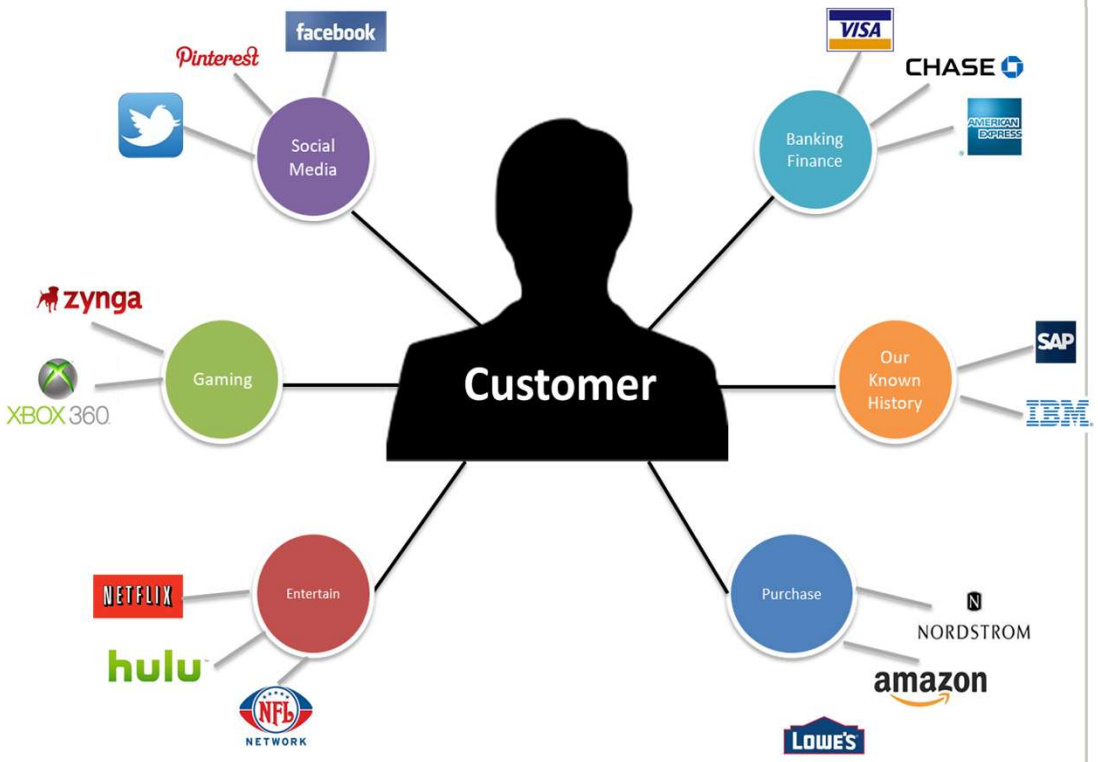
Challenges in Handling Big Data



- **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

Importance of Big Data

- Regardless of the size of the organization or industry, all businesses have access to Big Data as they share and store more information in varying formats, including:
  - E-Mail and Instant Messaging
  - Collaborative Intranets and Extranets
  - Public Websites, wikis, and Blogs
  - Social Media Channels
  - Video and Audio files
  - Data from industrial sensors, wearables and other monitoring devices



- The ability to extract high value from this data to enable innovation and competitive gain is the purpose of **Big Data analytics**. Conducting analytics on large sets of data, business users and executives are able to see patterns and trends in performance, new relationships between data sets and potentially new sources of revenue.

Benefits of Big Data Analytics

- Leveraging a Big Data analytics solution can help the businesses unlock the strategic value of this information by allowing the businesses to:

  - Understand where, when and why their customers buy
  - Protect their client base with improved loyalty programs
  - Seize cross selling and upselling opportunities
  - Provide targeted promotional information to their prospective and existing clients
  - Optimize Workforce planning and operations
  - Address inefficiencies in their supply chain
  - Predict market trends and future needs
  - Become more innovative and competitive
  - Discover new sources of revenue

Benefits of Real-Time Big Data Analytics