Units overview

- UNIT I INTRODUCTION TO BIG DATA
- UNIT II MINING DATA STREAMS
- UNIT III HADOOP ENVIRONMENT
- UNIT IV DATA ANALYSIS SYTEMS AND VISUALIZATION
- UNIT V FRAMEWORKS AND APPLICATIONS

Fields that generates data

All most all fields generate big data. Some major fields where big data plays a major role is

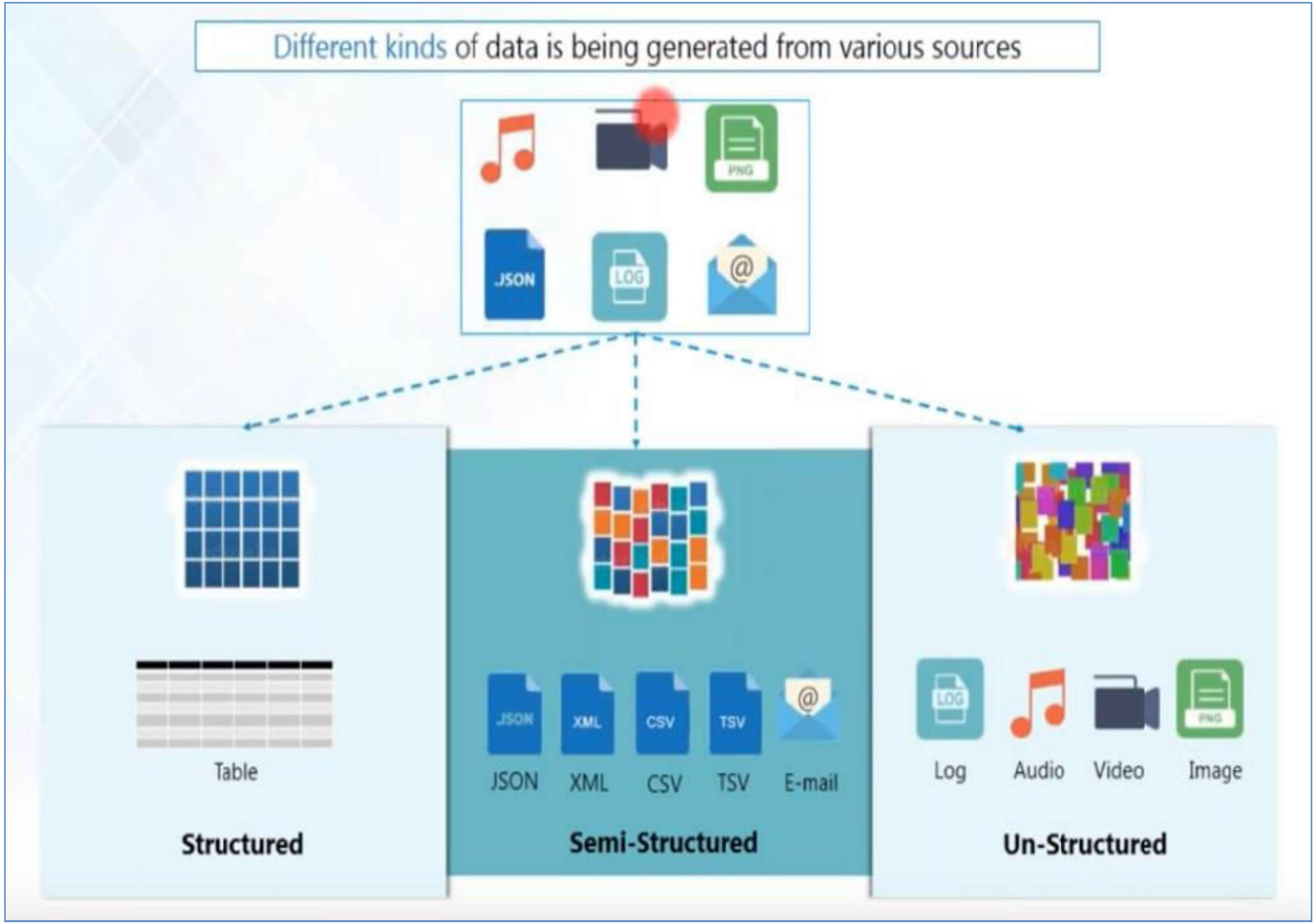
I. **Social networking sites:** social media that carry information, posts, links etc of different peoples from all over world like Facebook twitter etc.

II. **Search engines:** there are lots of data from different databases that retrieve from search engines.

III. **Medical history :** medical history of patients for various health issues from hospitals

IV. **Online shopping:** shopping online help to know the preferences of customers on different products.

V. **Stock exchange:** shares of different companies hold by stock


Different kinds of data is being generated from various sources
PNG
JSON
LOG
@
Table
Structured
JSON
XML
CSV
TSV
E-mail
Semi-Structured
LOG
Log
Audio
Video
PNG
Image
Un-Structured

Traditional Data Vs Big Data

Feature	Traditional Data	Big Data
Data architecture	Centralized database	Distributed database
Types of data	Structured data	Unstructured and semi-structured
Volume	Small amount of data. Range- Gigabyte - terabytes	Large amount of data. Range- petabytes.
Data schema	Fixed schema	Dynamic schema
Data Relationship	Relationship with data is explored easily	Difficulty in relationship between data items.
Scaling	More than one server for computing	Single server for computing
Accuracy	Less accurate results	High accurate results

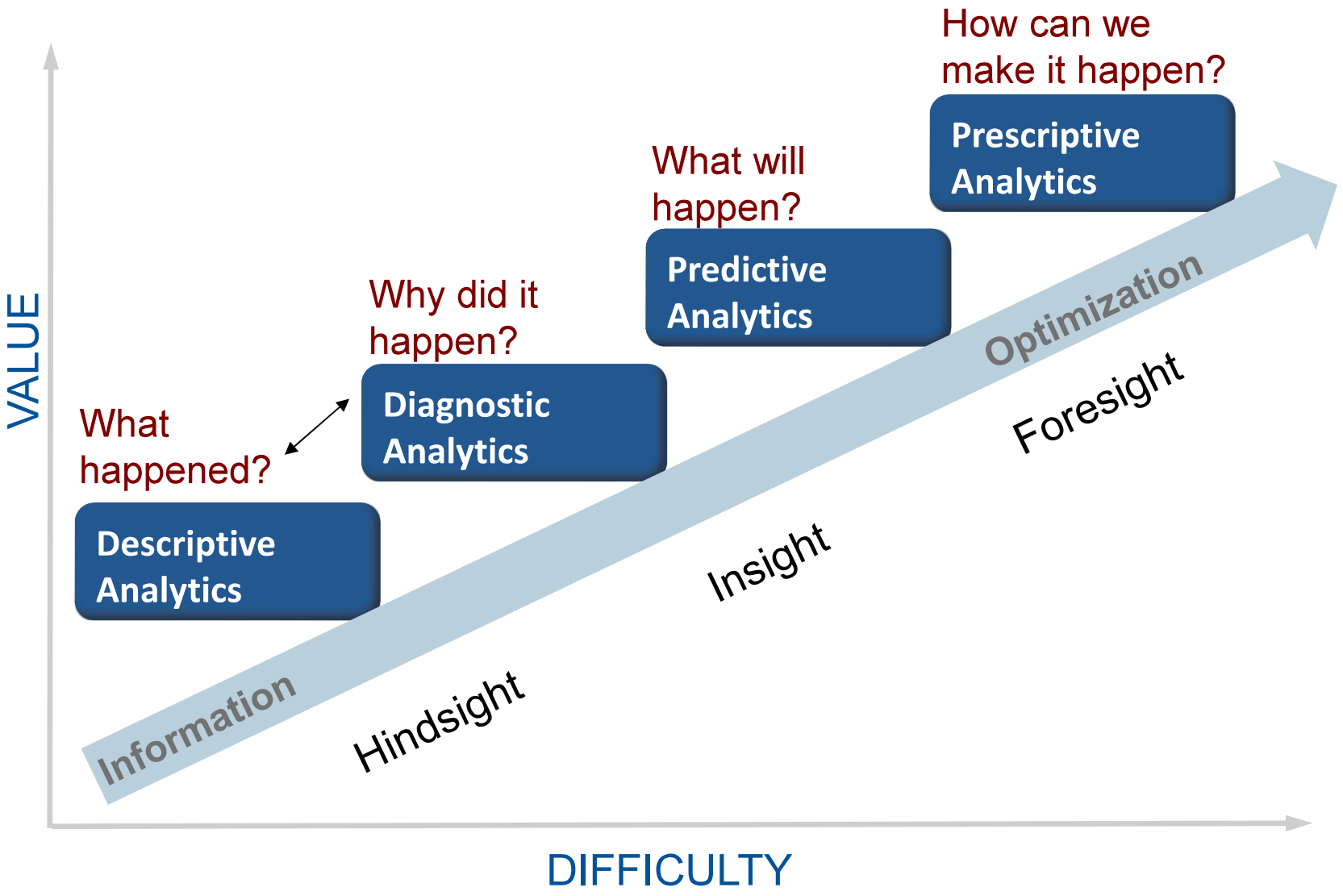
Big Data Analytics

- Big data analytics is a **method to uncover the hidden designs in large data, to extract useful information** that can be divided into two major sub-systems: data management and analysis.
- Big data analytics is a process of inspecting, differentiating and transforming big data with the goal of identifying useful information, suggesting conclusion and helping to take accurate decisions.
- Analytics include both data mining and communication or guide decision making.

The importance of big data analytics

- Big data analytics through specialized systems and software can lead to positive business-related outcomes:
- New revenue opportunities
- More effective marketing
- Better customer service
- Improved operational efficiency
- Competitive advantages over rivals

Analytics Models

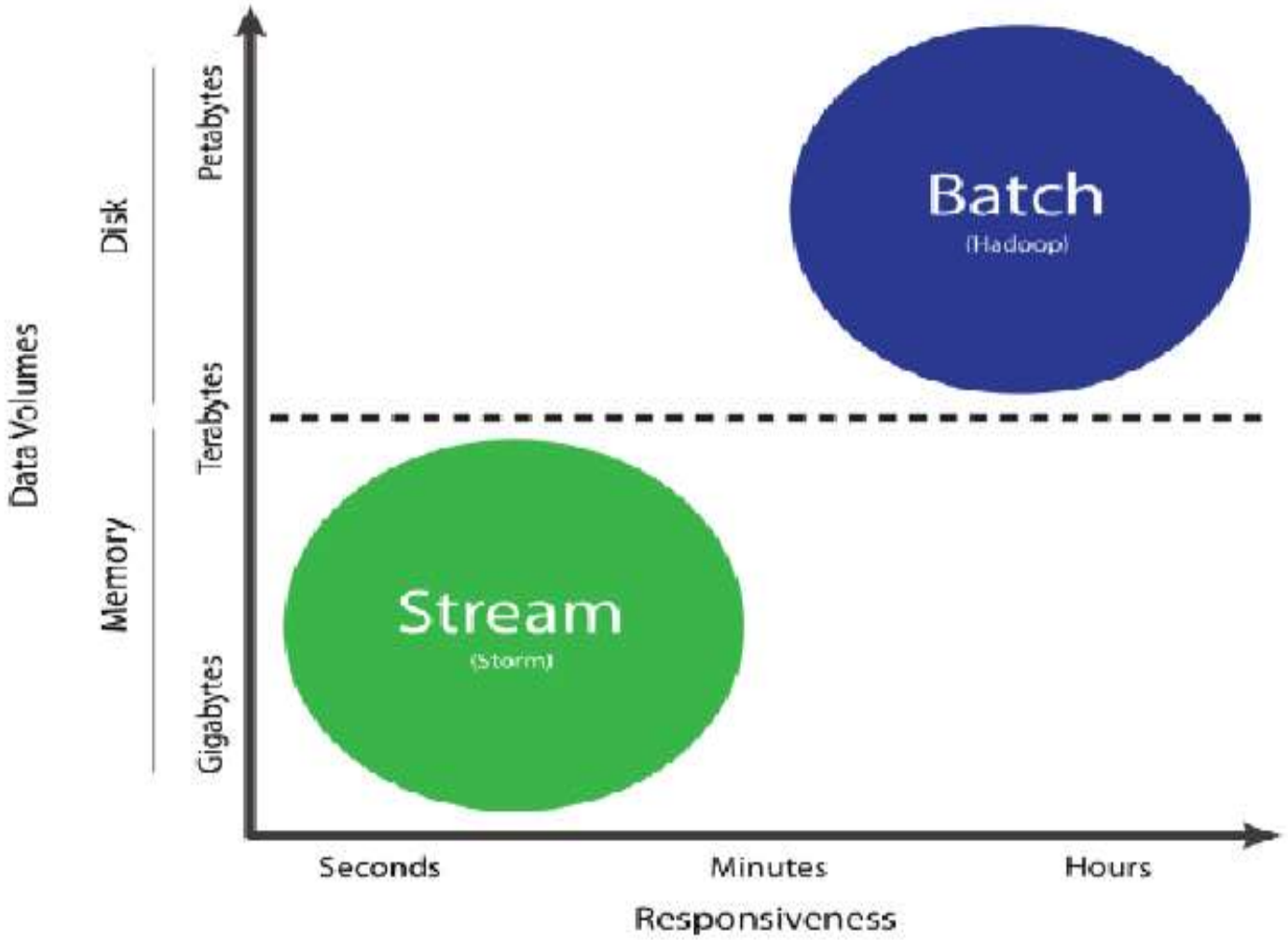


How big data analytics works

Once the data is ready, it can be analyzed with the software commonly used for advanced analytics processes. That includes tools for:

- **data mining**, which sift through data sets in search of patterns and relationships;
- **predictive analytics**, which build models to forecast customer behavior and other future developments;
- **machine learning**, which taps algorithms to analyze large data sets; and
- **deep learning**, a more advanced offshoot of machine learning.

Big Data technologies can be divided into two groups: *batch processing, which are analytics on data at rest, and stream processing, which are analytics on data in motion*



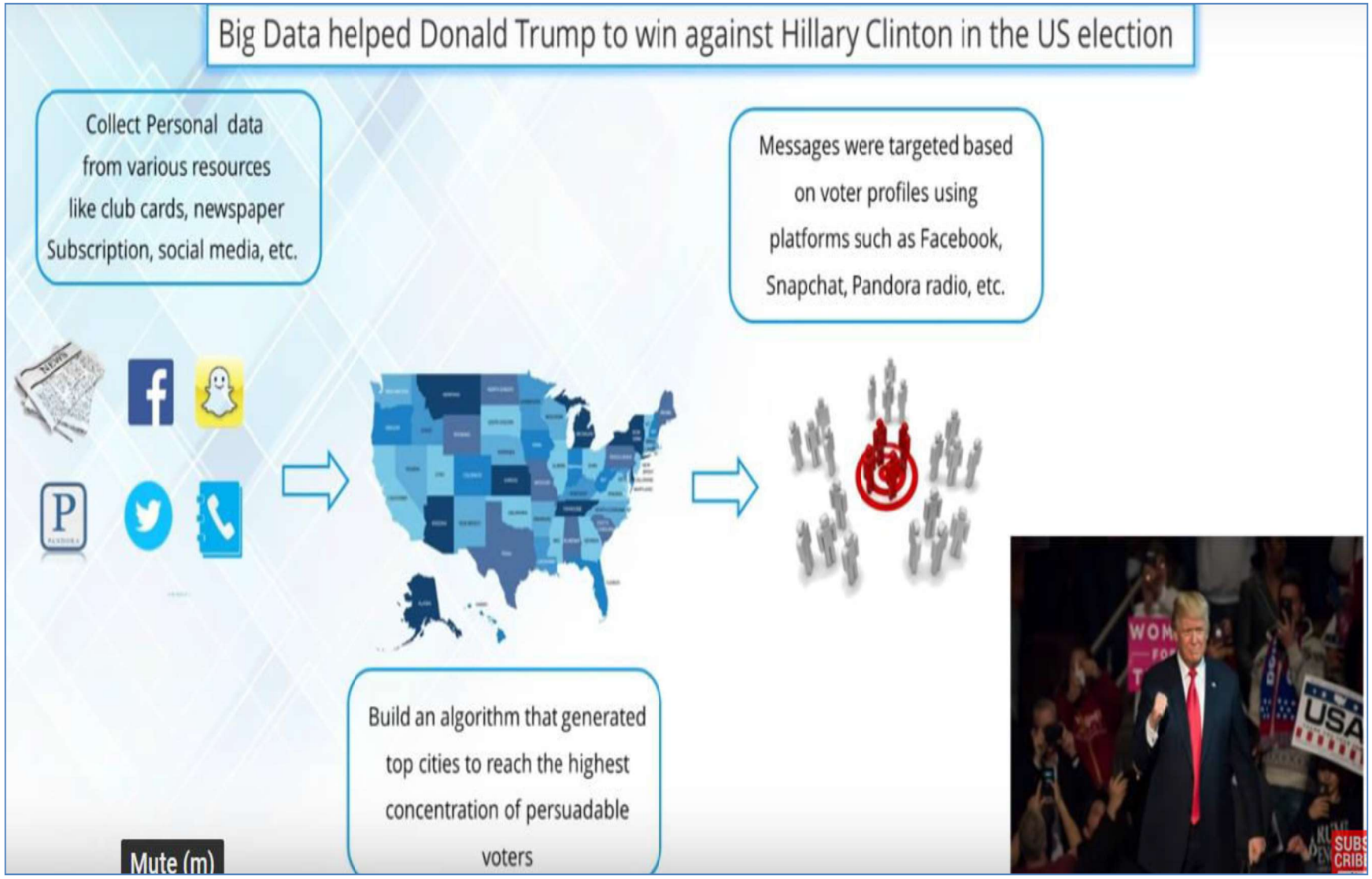
Applications and key data sources for big data and business analytics

S.No	Application	Key Data Sources	Features
1	Healthcare	Electronic health record, patients' information, images, health history data.	Support improved health monitoring, study patients' immune systems, activity recommendation for elderly physical health
2	Financial Industries	Financial reports, stock news, blog post, social media, and annual general meeting information	Provide a mechanism for fraud detection, mitigate against money laundry and decision making
3	Network Optimization	Network signal information, information between network users, weblog, geo-location data, sensor data, video camera, and network log	**Efficient network signaling,** prediction of network variation, network management and to generate cell deployment information

S.No	Application	Key Data Sources	Features
4	Travel estimation	GPS data, location data, satellite imagery, personal information, call data record(CDR)	Provide information for complex route recommendation, location tracking, drone routing for a military operation, emergency situation and infectious disease identification
5	User Behavior Modeling	Log data, social media data, blog post, tweets, and product review	Effective and efficient individual service recommendation.
6	User mobility modeling	Location data, GPS	Maintain global movement pattern to enable disease containment and transportation planning
7	Service Recommendation	Customer product review, product selection, location data, buying behavior data.	Enhanced product buying using customer product review and ascertain weaknesses and strength of products

S.No	Application	Key Data Sources	Features
8	Energy Consumption Analysis	Gas status, consumption pattern data, location data, smart meter reading data, and usage history.	Promote green energy, conservation, and efficiency through energy consumption prediction.
9	Crowdsourcing and sensing	Sensing data such as accelerometer, gyroscopes, magnetometer, electrocardiograph (ECG), pulse rate, electromyography (EMG), online questionnaire and survey.	Approach for large scale data collection project using a smartphone and online platforms.
10	Educational development	Student information, examination information, student enrollment, course allocation, course contents,	Predict student enrollment ratio and dropout rate after particular course or session

Use cases for Big data analytics



Walmart boosted its sales by leveraging the power of Big Data

While forecasting the demand for emergency supplies for approaching Hurricane Sandy, they gain some amazing insights:

Walmart

Extra supplies of Strawberry Pop Tarts were dispatched to stores in Hurricane Sandy's path in 2012, and sold extremely well

Along with flashlights and emergency equipment, they found an upsurge in sales of strawberry Pop Tarts

Apixio uses big data analytics to improve healthcare decision

APIXIO

80% of medical and clinical information about patients is in unstructured format, such as written physician notes

Analysis of medical data using variety of different methodologies & algorithms that are machine learning based and have NLP capabilities

The patient data model generated is aggregated across population to derive larger insights like disease prevalence, treatment patterns, etc.


ANY
QUESTIONS?

Thank You!