Big Data Analytics in the Management of Business

Big Data and Big Data Analytics

Big Data Every Where!

- Lots of data is being collected and warehoused
  - Web data, e-commerce
  - purchases at department/ grocery stores
  - Bank/Credit Card transactions
  - Social Network

Type of Data

- Relational Data (Tables/Transaction/Legacy Data)
- Audio, Video....
- Non-relational data sources
- Text Data (Web)
- Semi-structured Data (XML)
- Graph Data
  - Social Network, Semantic Web (RDF), ...
- Streaming Data
  - You can only scan the data once

What is BIG DATA?

- 'Big Data' is similar to 'small data', but …..

- It requires different approaches: – Techniques, tools and architecture

- It generates value from the storage and processing of very large quantities of digital information that cannot be analyzed with traditional computing techniques.

Characteristics/ Dimensions of Big Data

- V3s Volume Velocity Variety

- Data quantity • Data Speed • Data Types

Big Data Volume

- A typical PC might have had 10 gigabytes of storage in 2000.

- Today, Facebook ingests 500 terabytes of new data every day.

- Boeing 737 will generate 240 terabytes of flight data during a single flight across the US.

- The smart phones, the data they create and consume; sensors embedded into everyday objects will soon result in billions of new, constantly-updated data feeds containing environmental, location, and other information, including video.

- Facebook processes 1 million photographs per second.

- Facebook stores 260 billion photos using storage space of over 20 PB

- Big data volumes are relative and vary by factors, such as time and type of data.

- It will change in future….

- Definition of big data depends on the industry.

- It is impractical to define a specific threshold for big data volumes.

Big Data Velocity

- The rate at which data are generate and the speed at which it should be analysed and acted upon.

- Digital devices – sensors, smartphones….real-time analytics and evidence-based planning.

- E.g. Walmart – > 1 million transactions per hour. Data from mobile devices and flowing through mobile apps produces torrents of information that can be used to generate real-time, personalized offers for everyday customers.

- Customer info: geospatial location, demo, past buying patterns..

- Retailers – Streaming data sources – real-time analytics
- Enables firms to create real-time intelligence from high volumes of "perishable" data.

Big Data Velocity

- Clickstreams and ad impressions capture user behavior at millions of events per second

- high-frequency stock trading algorithms reflect market changes within microseconds

- machine to machine processes exchange data between billions of devices

- infrastructure and sensors generate massive log data in real time

- on-line gaming systems support millions of concurrent users, each producing multiple inputs per second.

- Time sensitive environments – Value proposition of the data degrades – worthless. E.g. health of a patient, health of an investment portfolio.
- Analytics – Data streaming analytics, data in-motion analytics

Big Data Variety

- Big Data isn't just numbers, dates, and strings.

- Big Data is also geospatial data, 3D data, audio and video, and unstructured text, including log files and social media.

- Traditional database systems were designed to address smaller volumes of structured data, fewer updates or a predictable, consistent data structure.

- Big Data analysis includes different types of data

- Data in Video formats – Largest component of big data

- Structural heterogeneity in a dataset.

- Structured data – 5% - tabular, relational databases

- Unstructured – Text, images, audio, video

- Semi-structured – no standards. E.g. XML (textual language for exchanging data on the Web). XML documents contain user-defined data tags which make them machine-readable.

- Unstructured data – Internal sources – e.g. Sensor data, External sources- Social Media

- Innovation in new data management technologies and analytics – Leverage data in their business processes.
- E.g. Facial recognition tech – Intelligence about store traffic – product promotions, placement, and staffing (brick-mortar retailers)
- E.g. Clickstream data – customer behaviour and browsing patterns – Online retailers – advised the timing and sequence of pages viewed by customers
- SMEs – semi-structured data to improve website design and implement effective cross-selling and personalized product recommendation systems

BDA: Text analytics

**Information Extraction – structured data from unstructured text.

• E.g. Drug name, dosage, and frequency from medical prescriptions

**Text summarization – automatically produces a summary of a single or multiple documents

e.g. Scientific and news articles, advt, emails, and blogs

** Question answering tech – answers to questions posed in natural language (NLP)

** Sentiment Analysis

BDA: Audio Analytics, Speech Analytics

- Unstructured audio data

- Human application – Speech Analytics

- Primary users: Call centres, healthcare

e.g. Call centers – analysis of millions of hours of recorded calls. – improve customer experience, evaluate agents performance ……

e.g. Healthcare – support diagnosis and treatment of certain medical conditions that affect the patients communication patterns (e.g. depression, schizophrenia, and cancer), infants cries, ….

BDA: Video Analytics

- To monitor, analyse and extract meaningful info from video streams
- Real-time and pre-recorded videos
- E.g. CCTV cameras, video sharing websites... leading contributors.
- Challenge – size of video data..
- Primary application – automated security and surveillance systems
- Retail sector – Study of buying behaviour of groups. Among family members who shop together, only one interacts with the store at the cash register, causing the traditional systems to miss data on buying patterns of other members.

Big Data Technology

- MapReduce

- Hadoop

- NoSQL

Location Analytics

- What is it?
  - Augmenting mission-critical, enterprise business systems with complementary content, mapping, and geographic capabilities
  - Mapping & Visualization: use maps as the media to visualize data
  - Spatial analytics: merging GIS w/ other types of analytics
  - Find spatio-temporal patterns indicative of physical activities or social behavior
  - Data/information enrichment: add maps, imagery, demographics, consumer and lifestyle data, environment and weather, social media, etc.
- Ubiquity of GPS on cellphones, cars, wristwatches, laptops, tablets, etc.

Ref: http://www.esri.com/software/location-analytics
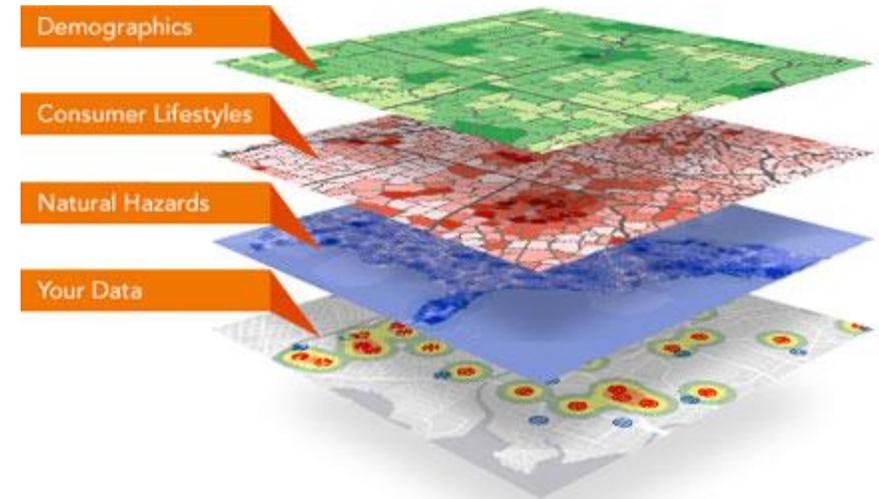
Web Analytics

- What is it?
  - Now: The study of the behavior of web users
  - Future: The study of one mechanism for how society makes decisions
  - Example: Behavior of Web Users
    - How many people clicked on Ebola (or related terms in the past 2 months)
    - Their location, their dwell time, the number of sites they examined, the difficulty or complexity of the material on the web site
    - What can this tell us about popular concern about Ebola?
    - Can it help decision makers to better present information and decisions
  - Commercially, it is the collection and analysis of data from a web site to determine which aspects of the website achieve the business objectives

The Model Has Changed...

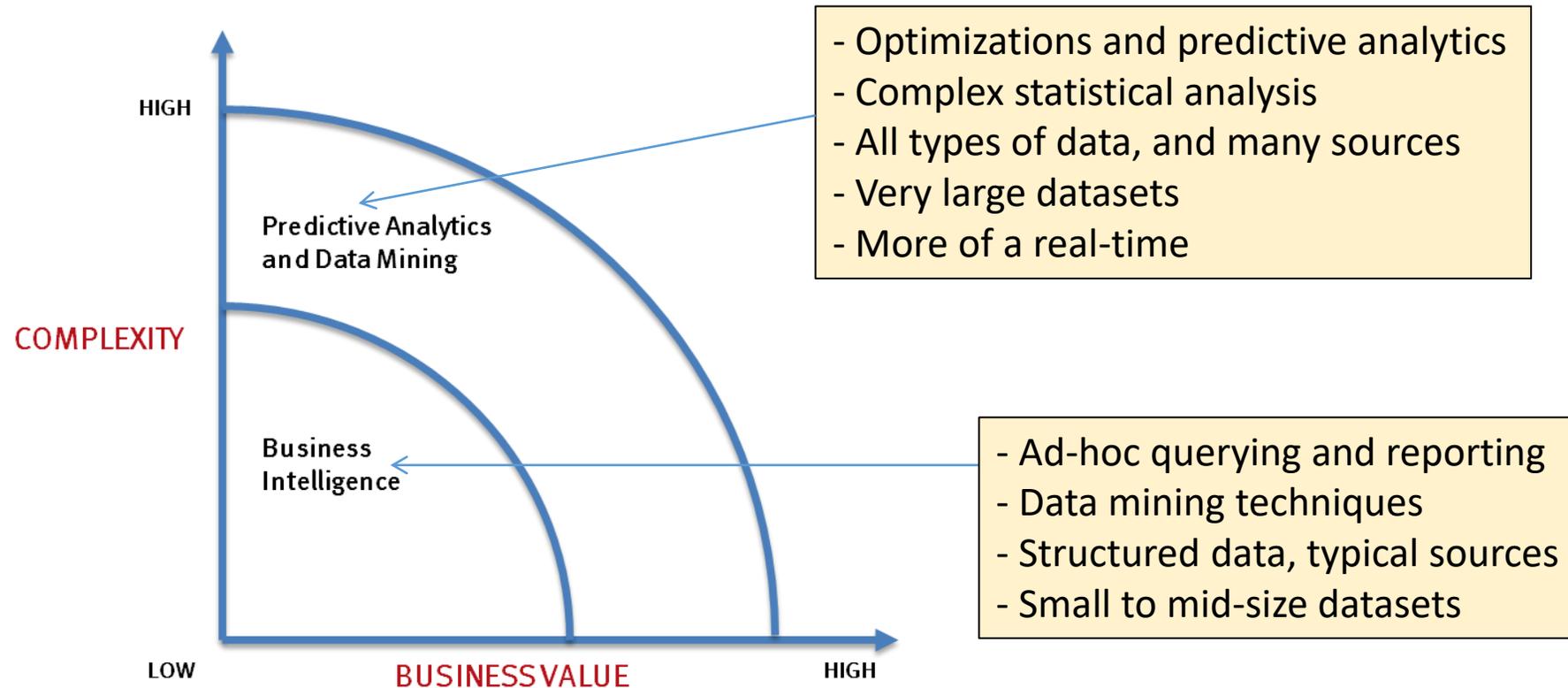- **The Model of Generating/Consuming Data has Changed**

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

HIGH

COMPLEXITY

Predictive Analytics
and Data Mining

Business
Intelligence

LOW   BUSINESS VALUE   HIGH

Big data is a business priority

– inspiring new models and processes for organizations, and even entire industries

Government achieves significant cost savings and ability to react to potential threats quickly

Government cuts acoustic analysis from hours to **70 Milliseconds**

Utility provider improves prediction of power outages

Utility avoids power failures by analyzing **10 PB** of data in minutes

Hospital detects and intervenes in potentially life-threatening conditions

Hospital analyzes streaming vitals to intervene **24 hours earlier**

Retailer optimizes inventory levels and product mix

Retailer reduces time to run queries by **80%**

Stock exchange reduces time to insights to achieve optimal buying / selling strategies

Stock Exchange cuts queries from 26 hours to **2 minutes** on **2 PB**

Telco provider improves ability to quickly address network issues / opportunities
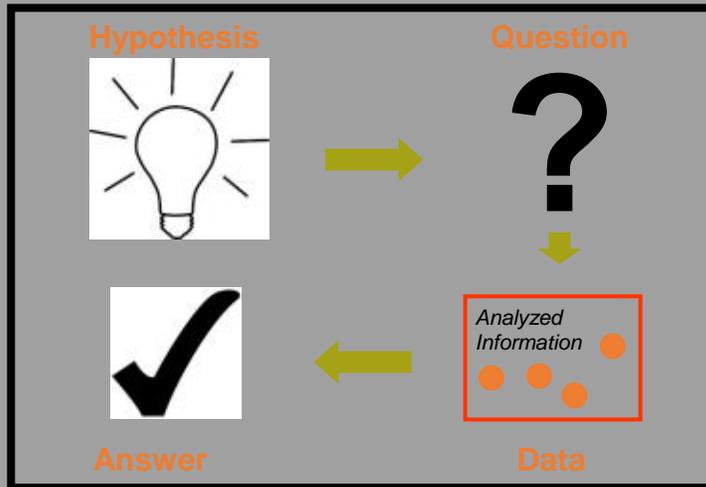
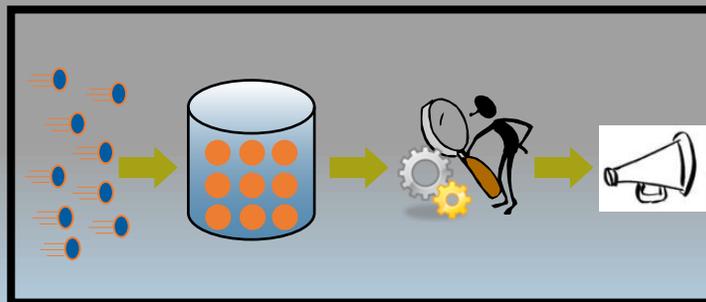Telco analyses streaming network data to reduce hardware costs by **90%**

The Big Data Approach to Analytics is Different

Traditional Analytics
Structured & Repeatable
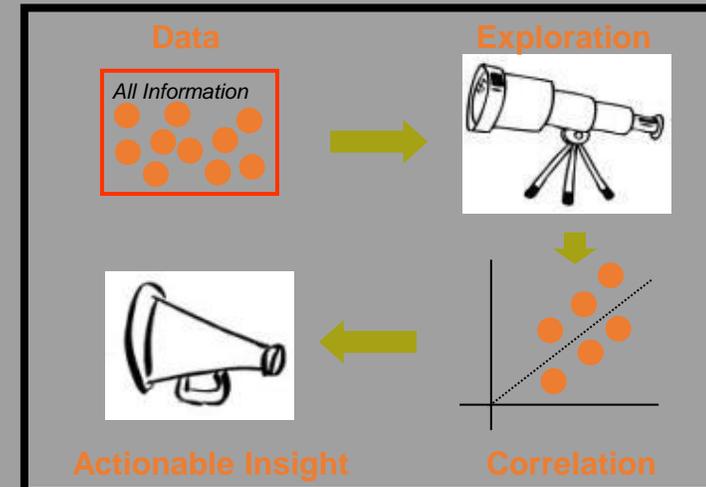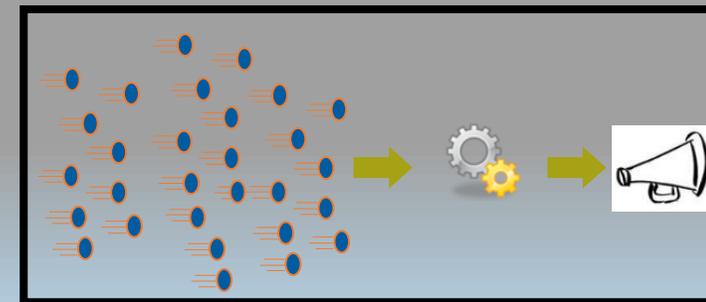**Structure built to store data**

Hypothesis → Question

Answer ← Data (Analyzed Information)

Start with hypothesis
Test against selected data

Analyze after landing…

Big Data Analytics
Iterative & Exploratory
Data is the structure

Data (All Information) → Exploration

Actionable Insight ← Correlation

Data leads the way
Explore *all* data, identify correlations

Analyze in motion…