SNS COLLEGE OF ENGINEERING

Kurumbapalayam(Po), Coimbatore – 641 107
Accredited by NAAC-UGC with 'A' Grade
Approved by AICTE, Recognized by UGC & Affiliated to Anna University, Chennai

Department of Artificial Intelligence and Data Science

Course Name – Foundations of Data Science

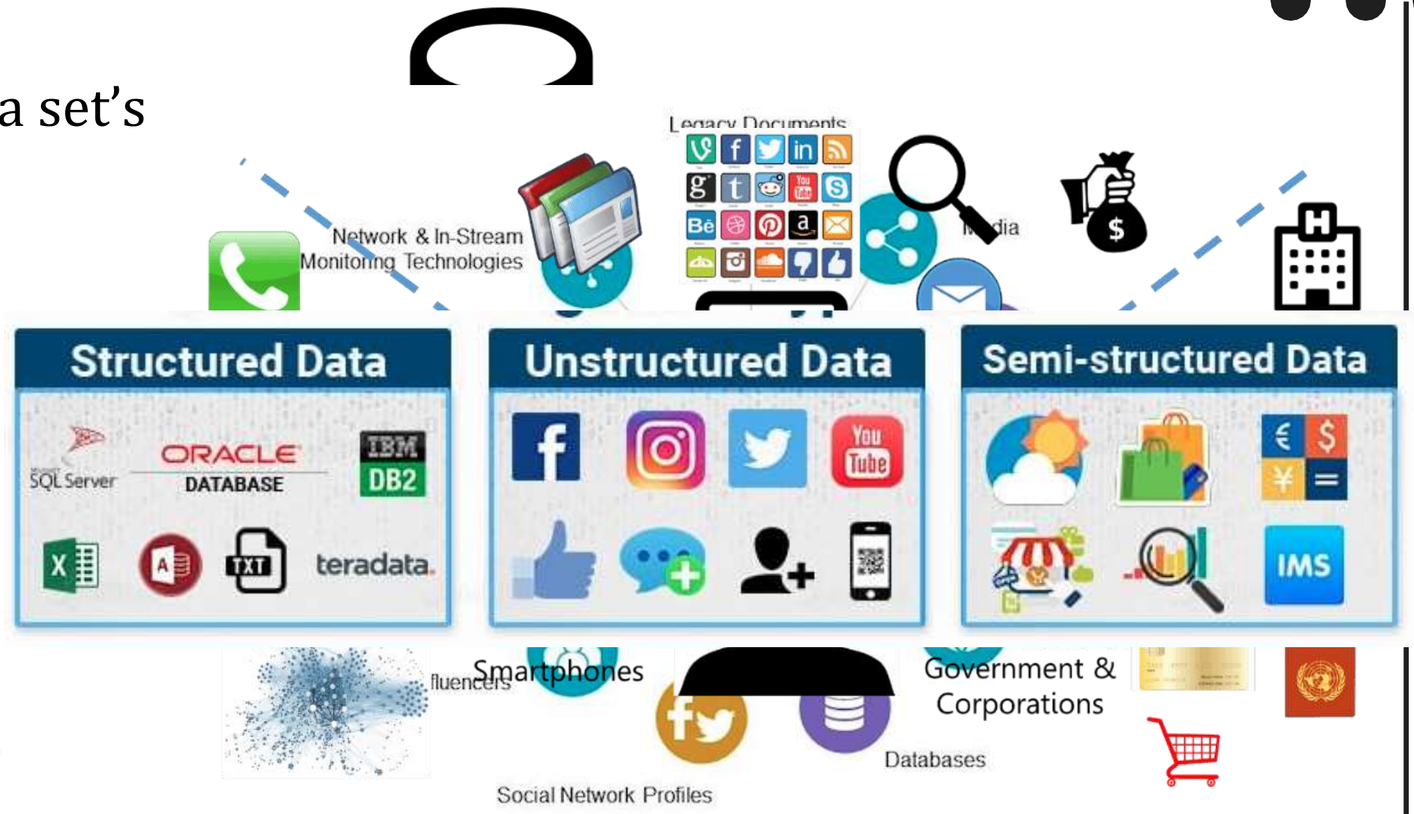
III Year / V Semester

Unit 1 – Introduction

Topic 6: Evolution of Big Data

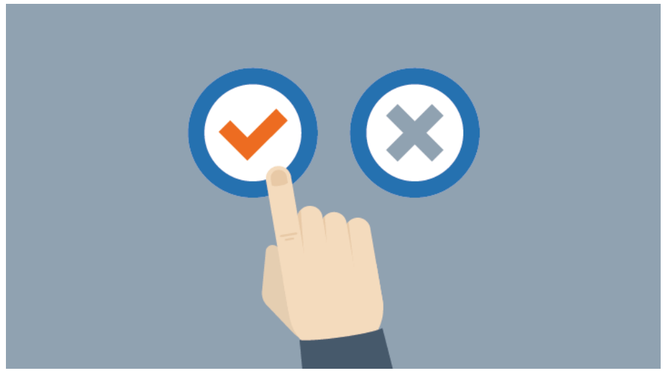
What is Big Data?

- Larger or Voluminous, Complex data set's
- From different sources
- Different Types
- Traditional Database cant handle it.



What is Big Data Analytics?

- Gathering Data
- Storing
- Analyzing or Processing
- Get Useful Business Intelligence
- To make better decisions for business growth.



Evolution of Big Data

1980 – 2000

- WWW
- DBMS
- OLAP
- Dashboard & Score cards
- Data Mining & statistical analysis

Evolution of Big Data

1980 – 2000

- Big Data Coined by John Mashey in 1998
- *Big Data… and the Next Wave of Infrastress*

- In 2000, Francis Diebold presented a paper titled “’ Big Data’ Dynamic Factor Models for Macroeconomic Measurement and Forecasting” to the Eighth World Congress of the Econometric Society.

Diebold, F.X. (2003),
"Big Data' Dynamic Factor Models for Macroeconomic Measurement and Forecasting"
(Discussion of Reichlin and Watson papers), in M. Dewatripont, L.P. Hansen and S.Turnovsky (Eds.),
Advances in Economics and Econometrics, Eighth World Congress of the Econometric Society.
Cambridge: Cambridge University Press, 115-122.

“Big Data” Dynamic Factor Models for Macroeconomic Measurement and Forecasting

Francis X. Diebold

University of Pennsylvania and NBER

First Version, July 2000
November 28, 2000

Evolution of Big Data

2000 – 2010

- Doug Laney in 2001 coined 3 V's
- Analyst with the Meta Group (Gartner),
- *"3D Data Management: Controlling Data Volume, Velocity, and Variety*."

- The 3V's have become the most accepted dimensions for defining big data.

Application Delivery Strategies META Group

Date: 6 February 2001 File: 949
Author: Doug Laney

3D Data Management: Controlling Data Volume, Velocity, and Variety. Current business conditions and mediums are pushing traditional data management principles to their limits, giving rise to novel, more formalized approaches.

META Trend: During 2001/02, leading enterprises will increasingly use a centralized data warehouse to define a common business vocabulary that improves internal and external collaboration. Through 2003/04, data quality and integration woes will be tempered by data profiling technologies (for generating metadata, consolidated schemas, and integration logic) and information logistics agents. By 2005/06, data, document, and knowledge management will coalesce, driven by schema-agnostic indexing strategies and portal maturity.

The effect of the e-commerce surge, a rise in merger/acquisition activity, increased collaboration, and the drive for harnessing information as a competitive catalyst is driving enterprises to higher levels of consciousness about how data is managed at its most basic level. In 2001/02, historical, integrated databases (e.g., data warehouses, operational data stores, data marts), will be leveraged not only for intended analytical purposes, but increasingly for intra-enterprise consistency/coordination. By 2003/04, these structures (including their associated metadata) will be on par with application portfolios, organization charts, and procedure manuals for defining a business to its employees and affiliates. Data records, data structures, and definitions commonly accepted throughout an enterprise reduce fiefdoms pulling against each other due to differences in the way each perceives where the enterprise has been, is presently, and is headed. Readily accessible current/historical records of transactions, affiliates (partners, employees, customers, suppliers), and business processes (or rules), along with definitional and navigational metadata (see ADS Delta 896, 7 Aug 2000) enable employees to paddle in the same direction. Conversely, application-specific data stores (e.g., accounts receivable versus order status), geographic-specific data stores (e.g., North American sales vs. International sales), offer conflicting or insular views of the enterprise that, while important for feeding transactional systems, provide no "single version of the truth," giving rise to inconsistency in the way enterprise factions function. While enterprises struggle to consolidate systems and collapse redundant databases to enable greater operational, analytical, and collaborative consistencies, changing economic conditions have made this job more difficult. E-commerce, in particular, has exploded data management challenges along three dimensions: volumes, velocity, and variety. In 2001/02, IT organizations must compile various approaches to have at their disposal for dealing with each.

Data Volume. E-commerce channels increase the depth/breadth of data available about a transaction (or any point of interaction). The lower cost of e-channels enables an enterprise to offer its goods or services to more individuals or trading partners, and up to 10x the quantity of data about an individual transaction may be collected — thereby increasing the overall volume of data to be managed. Furthermore, as enterprises come to see information as a tangible asset, they become reluctant to discard it. Typically, increases in data volume are handled by purchasing additional online storage. However, as data volume increases, the relative value of each data point decreases proportionately — resulting in a poor financial justification for merely incrementing online storage. Viable alternates/supplements to hanging new disk include:

- Implementing tiered storage systems (see SIS Delta 860, 19 Apr 2000) that cost-effectively balance levels of data utility with data availability using various media

Business Impact

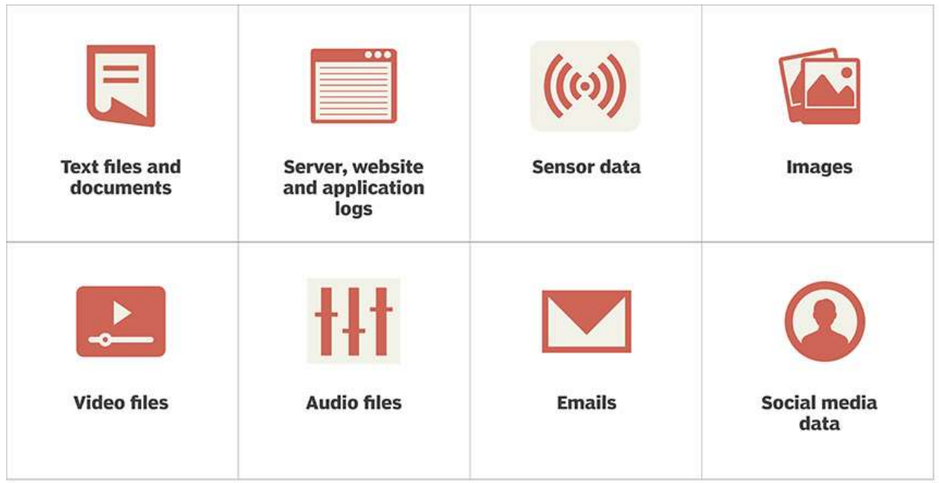
Attention to data management, particularly in a climate of e-commerce, and greater need for collaboration can enable enterprises to achieve greater returns on their information assets.

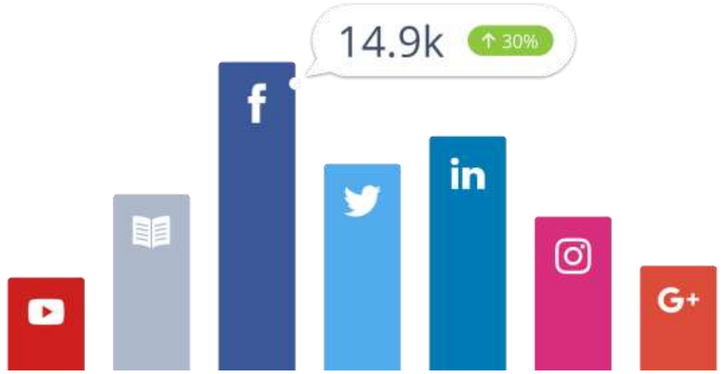
META Delta

Evolution of Big Data

2000 – 2010

- Web Based Unstructured Content
- Information retrieval and extraction
- Opinion Mining
- Web Analytics
- Social Media Analysis
- Social Network Analysis
- Web traffic and online stores

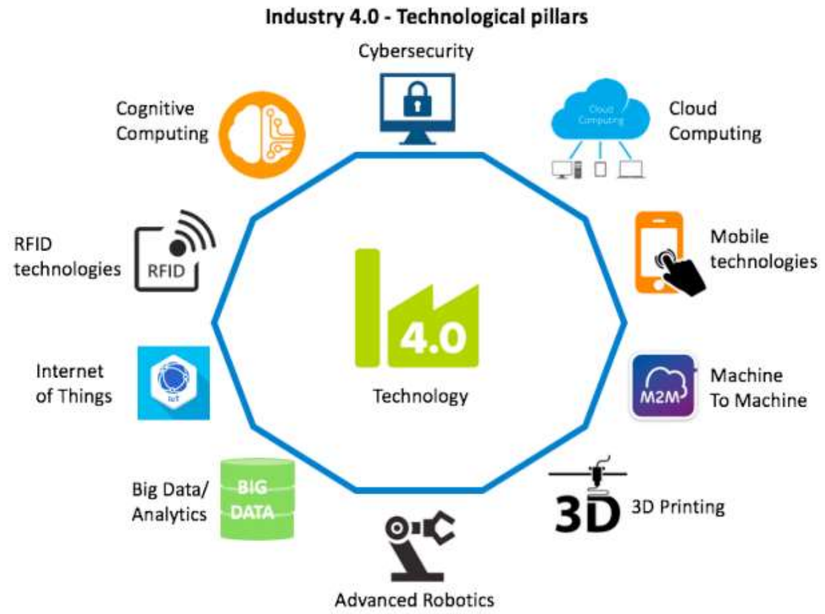




Evolution of Big Data

2010 – Present

- Mobile data
  - Location based data
  - Behavioral data
- IoT with Sensors
- Industry 4.0
- AI, Cognitive Computing



THANK YOU