

Engineering in the Context of Big Data

Jennifer Dy, Northeastern University
Sriram Raghavan, IBM Research - India

The Big Data Phenomenon



1.3 Billion RFID tags in 2005
30 Billion RFID tags in
2010



~5 Billion mobile
phones worldwide



By 2017 annual internet traffic
expected to reach **1.4**
Zettabytes



Google processes
> 24 Petabytes of data
in a single day



Facebook processes
10 Terabytes of data every
day



Twitter processes
7 Terabytes of data every day
250,000,000 tweets

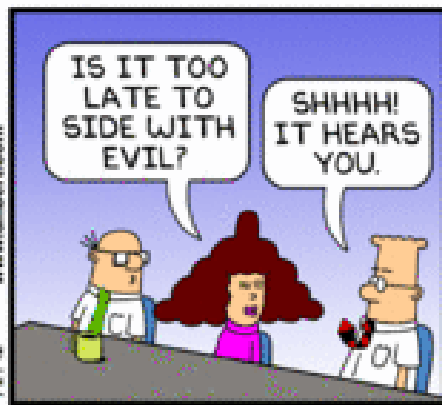
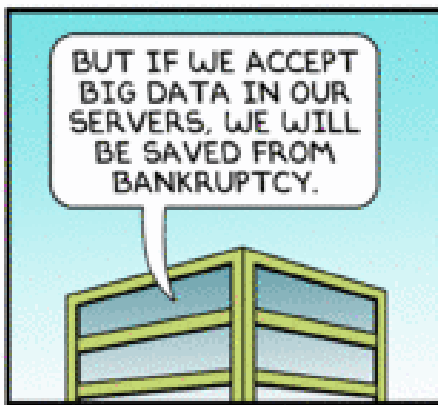
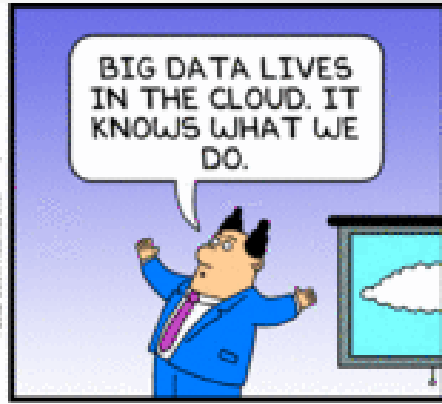
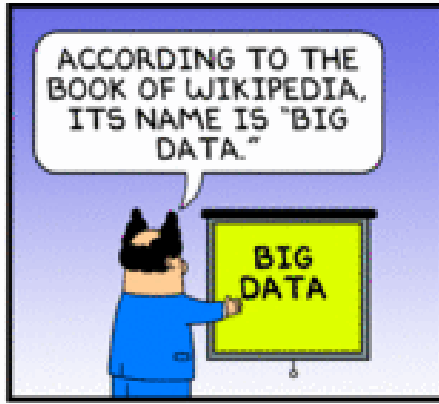


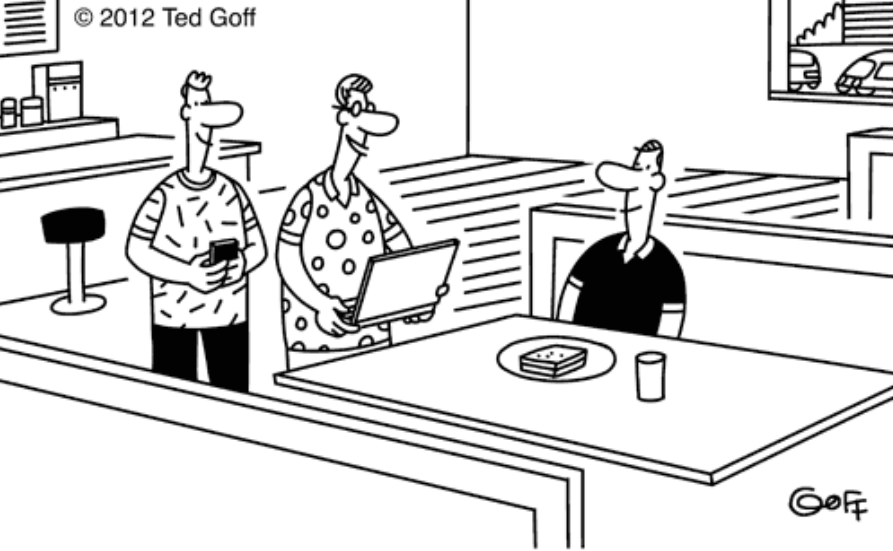
Hadron Collider at CERN
generates **40 Terabytes**
of data / sec



For every session, NY Stock
Exchange captures **1 Terabyte**
of trade information

“...Every two days, we now generate the equivalent of all of the data that existed up to 2003. And thanks to advanced computation and analytics, we have the tools to turn that data into insight, knowledge and better decisions...”





“Twitter and Facebook can’t predict the election, but they did predict what you’re going to have for lunch: a tuna salad sandwich. You’re having the wrong sandwich.”



Big Data Analytics is touching every industry



Travel & Transport

- Congestion modeling
- Predictive Maintenance Analytics
- Capacity & Pricing Optimization
- Logistics optimization



Energy & Utilities

- Smart Meter Analytics
- Distribution Load Forecasting/Scheduling
- Condition Based Maintenance
- Flexible targeted Customer Offerings



Telco

- Pro-active Call Center
- Network Analytics & Optimization
- Location Based Services



Automotive

- Advanced Condition Monitoring
- Actionable Customer Intelligence
- Warranty analytics



Health & Life Sciences

- Modeling population health outcomes
- Medical record analytics, genetic analysis
- Drug discovery, safety, and effectiveness
- Epidemiological modeling & disease surveillance



Government

- Citizen services & voice-of-citizen analysis
- Public safety and law enforcement



Chemical & Petroleum

- Operational Surveillance, Analysis & Optimization
- Data-driven exploration, condition-based monitoring, and health and safety monitoring
- Basin modeling, reservoir simulation, etc.



Banking

- Optimizing Offers and Cross-sell
- Customer Service and Call Center Efficiency
- Fraud Detection & Investigation
- Credit & Counterparty Risk



Consumer Products

- Shelf availability & Merchandising compliance
- Promotional spend optimization
- Promotion exceptions & alerts
- Dynamic pricing
- Merchandise optimization



Media & Entertainment

- Audience and marketing optimization
- Multi-channel enablement
- Social media monitoring & campaign management
- Digital commerce and optimization



Climate

- Extreme event prediction
- Climate change, impact and sustainability analysis



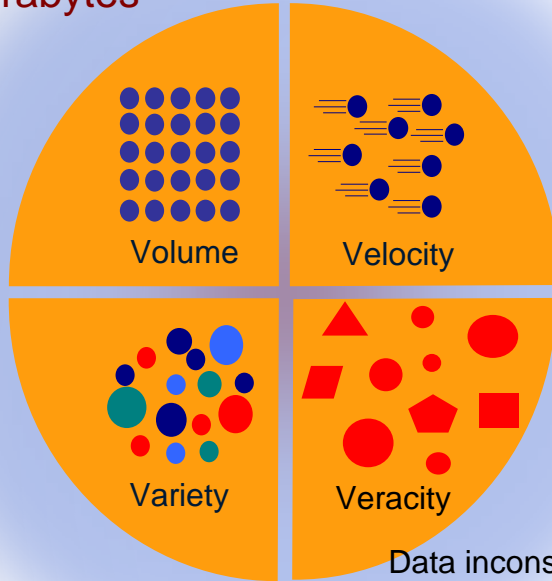
Space

- Discovery of stars, exoplanets
- Spectral analysis of rocks, soil, atmospheric composition

4V's of Big Data

12+ terabytes

of Tweets
create daily.



Volume

Velocity

Variety

Veracity

5+ million

trade events
per second.

100's

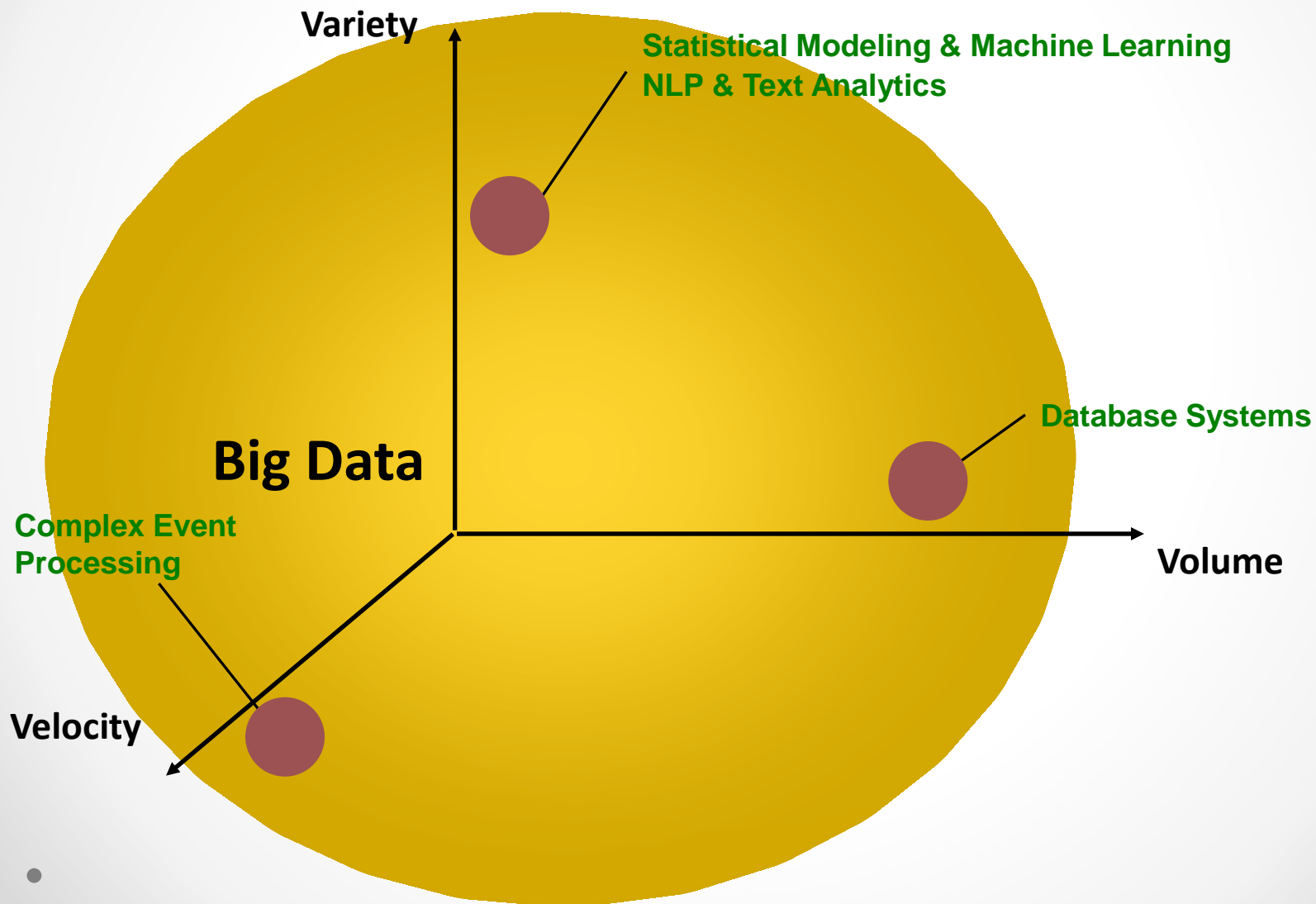
of different types of data.

Data inconsistency,
incompleteness,
ambiguity, latency,
deception, modeling
errors, acquisition
errors, ..

Analytics

Inform decisions
Optimize processes
Drive new models

Big Data: Convergence of different technical disciplines within Computer Science



Session Schedule

- ***The Internet of Manufacturing Things***
Athulan Vijayaraghavan, System Insights
- ***Predictive Analytics for Industrial Applications***
Pankaj Dayama, IBM Research
- ***At the Intersection of Healthcare, Drug Discovery, and Big Data***, Nirmal Keshava, AstraZeneca PLC
- ***Facilitating Discovery in Big Data Sets***
Kiri Wagstaff, Jet Propulsion Laboratory
- **Panel Discussion**