Big Data Analytics and Business Intelligence

1

Course description

- The main goal of this course is to help students understand and solve problems of big data and applications of big data analytics to business intelligence.
- The main objective of the course: the concept of big data overview, algorithms and data analytics technology on big data support for business.

2

Course materials

1. Vijayan Sugumaran, Arun Kumar Sangaiah, Arunkumar Thangavelu, "Computational intelligence applications in business and big data analytics", Taylor & Francis, (2017)
2. Daniel O'Reilly, Python for Data Science: The Ultimate Step-by-Step Guide to Python Programming. Discover How to Master Big Data Analysis and Understand Machine Learning, ISBN: 979-8719424248, (2021).
3. Michael Minelli, Michele Chambers, Ambiga Dhiraj, "Big data,big analytics - emerging business intelligence and analytic trends for today's businesses", John Wiley & Sons, (2013)
4. Steve Williams, "Business intelligence strategy and big data analytics: a general management perspective", Elsevier, (2016)
5. David Dietrich, Barry Heller, Beibei Yang, "Data science and big data analytics", Wiley, (2015) Oracle, "Data Mining Concepts", 18c, E83730-03, 2018
6. https://www.ibm.com/analytics/hadoop/big-data-analytics

3

Assessment methods

- Mini project: 35%
- Presentation: 15%
- Final exam: 50%

4

Content

Week	Content
1	Big data overview
2,3	Basics in big data analytics
4,5	Big data analytics
6	Business intelligence system
7,8	Big data analytics supports business
9,10,11,12	Student group reports on major subject matter topics

5

Lecturer 1

BIG DATA OVERVIEW

6

What's Big Data?

No single definition; here is from Wikipedia:

Big data is the term for a collection of data sets so large and complex that it becomes difficult to process using on-hand database management tools or traditional data processing applications.
The challenges include capture, curation, storage, search, sharing, transfer, analysis, and visualization.
The trend to larger data sets is due to the additional information derivable from analysis of a single large set of related data, as compared to separate smaller sets with the same total amount of data, allowing correlations to be found to "spot business trends, determine quality of research, prevent diseases, link legal citations, combat crime, and determine real-time roadway traffic conditions."

7

7

Definition and Characteristics of Big Data

"*Big data is high-**volume**, high-**velocity** and high-**variety** information assets that demand **cost-effective**, **innovative** forms of information processing for **enhanced insight and decision making**.*" -- Gartner

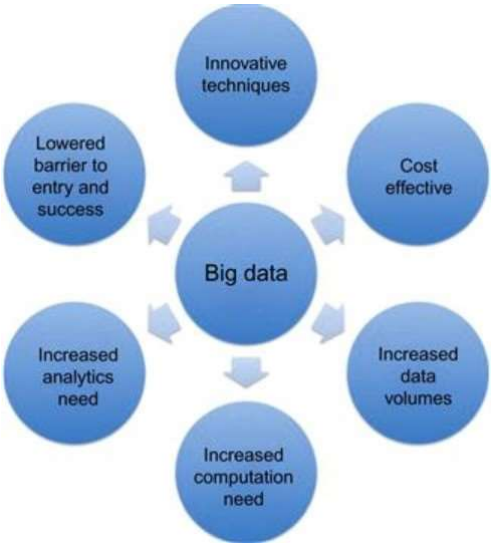
which was derived from:

"*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 much compile a variety of approaches to have at their disposal for dealing each.*" – Doug Laney

8

8

What made Big Data needed?



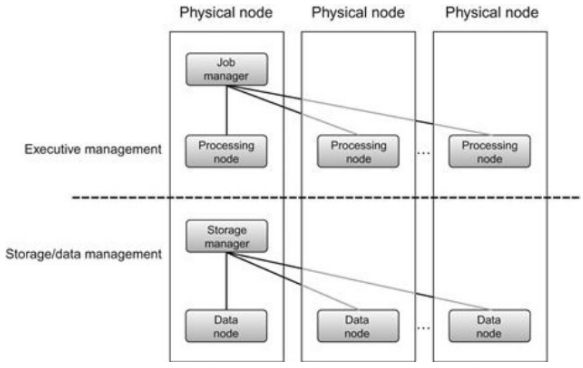
"Big Data Analytics", David Loshin

9

9

Key Computing Resources for Big Data

- Processing capability: CPU, processor, or node.
- Memory
- Storage
- Network



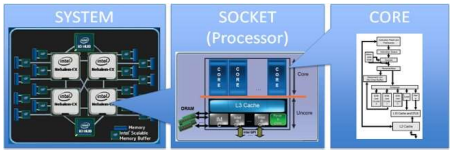
"Big Data Analytics", David Loshin

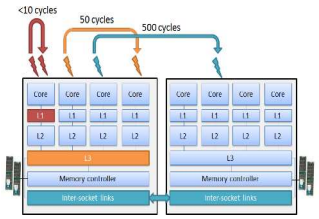
10

10

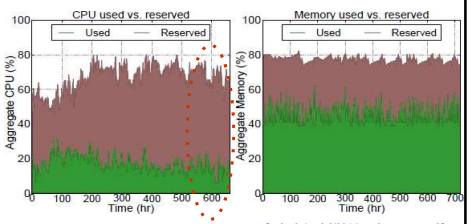
Scalability — Scale Up & Scale Out

- Scale out
  - Use more resources to distribute workload in parallel
  - Higher data access latency is typically incurred
- Scale up
  - Efficiently use the resources
  - Architecture-aware algorithm design





Example: Resource utilization for a large production cluster at Twitter data center



www.stanford.edu/~cdel/2014.asplos.quasar.pdf

- For independent data ==> scale up may not have obvious advantage than scale out
- For linked data ==> utilizing scale up as much as possible before scale out

5

11

Contrasting Approaches in Adopting High-Performance Capabilities

Aspect	Typical Scenario	Big Data
Application development	Applications that take advantage of massive parallelism developed by specialized developers skilled in high-performance computing, performance optimization, and code tuning	A simplified application execution model encompassing a distributed file system, application programming model, distributed database, and program scheduling is packaged within Hadoop, an open source framework for reliable, scalable, distributed, and parallel computing
Platform	Uses high-cost massively parallel processing (MPP) computers, utilizing high-bandwidth networks, and massive I/O devices	Innovative methods of creating scalable and yet elastic virtualized platforms take advantage of clusters of commodity hardware components (either cycle harvesting from local resources or through cloud-based utility computing services) coupled with open source tools and technology
Data management	Limited to file-based or relational database management systems (RDBMS) using standard row-oriented data layouts	Alternate models for data management (often referred to as NoSQL or "Not Only SQL") provide a variety of methods for managing information to best suit specific business process needs, such as in-memory data management (for rapid access), columnar layouts to speed query response, and graph databases (for social network analytics)
Resources	Requires large capital investment in purchasing high-end hardware to be installed and managed in-house	The ability to deploy systems like Hadoop on virtualized platforms allows small and medium businesses to utilize cloud-based environments that, from both a cost accounting and a practical perspective, are much friendlier to the bottom line

"Big Data Analytics", David Loshin

6

12

Techniques towards Big Data

- Massive Parallelism
- Huge Data Volumes Storage
- Data Distribution
- High-Speed Networks
- High-Performance Computing
- Task and Thread Management
- Data Mining and Analytics
- Data Retrieval
- Machine Learning
- Data Visualization

➔ Techniques exist for years to decades. Why is Big Data hot now?

13

13

Why Big Data now?

- More data are being collected and stored
- Open source code
- Commodity hardware / Cloud

14

14

Why Big Data now?

- More data are being collected and stored
- Open source code
- Commodity hardware / Cloud

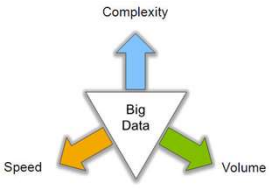
➔
- High-Volume
- High-Velocity
- High-Variety

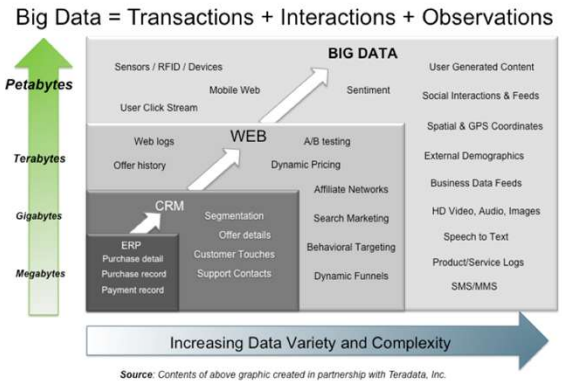
➔ Artificial Intelligence

15

15

Big Data: 3V's





16

16

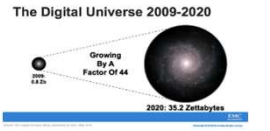
Volume (Scale)

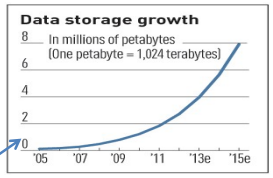
Data Volume

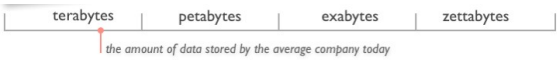
44x increase from 2009 2020

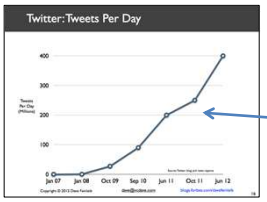
From 0.8 zettabytes to 35zb

Data volume is increasing exponentially







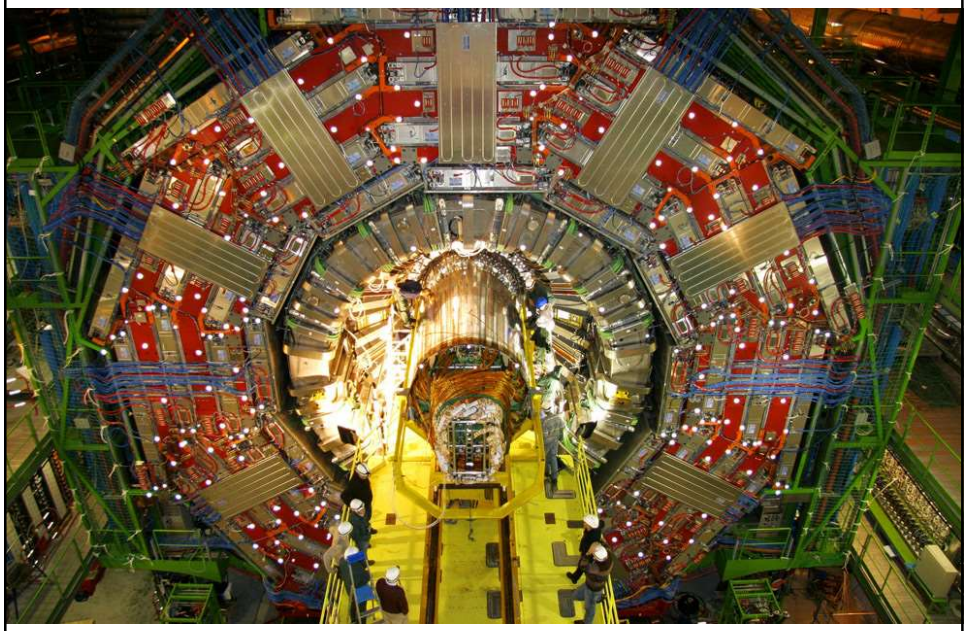


Exponential increase in collected/generated data

17

17

18



CERN's Large Hydron Collider (LHC) generates 15 PB a year

19

The Earthscope

- The Earthscope is the world's largest science project. Designed to track North America's geological evolution, this observatory records data over 3.8 million square miles, amassing 67 terabytes of data. It analyzes seismic slips in the San Andreas fault, sure, but also the plume of magma underneath Yellowstone and much, much more. (http://www.msnbc.msn.com/id/44363598/ns/technology_and_science-future_of_technology/#.TmetOdQ--ul)

20

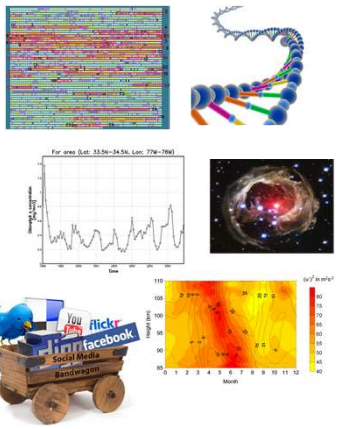
Variety (Complexity)

Relational Data (Tables/Transaction/Legacy Data)
Text Data (Web)
Semi-structured Data (XML)
Graph Data
    Social Network, Semantic Web (RDF), …

Streaming Data
    You can only scan the data once

A single application can be generating/collecting many types of data

Big Public Data (online, weather, finance, etc)

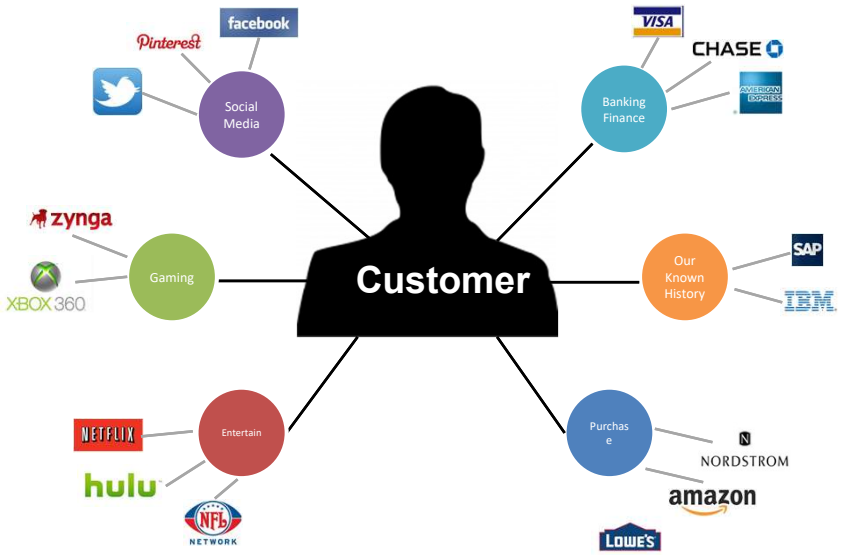


To extract knowledge➔ all these types of data need to linked together

21

21

A Single View to the Customer



22

Velocity (Speed)

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 histor
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

23

23

Real-time/Fast 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)

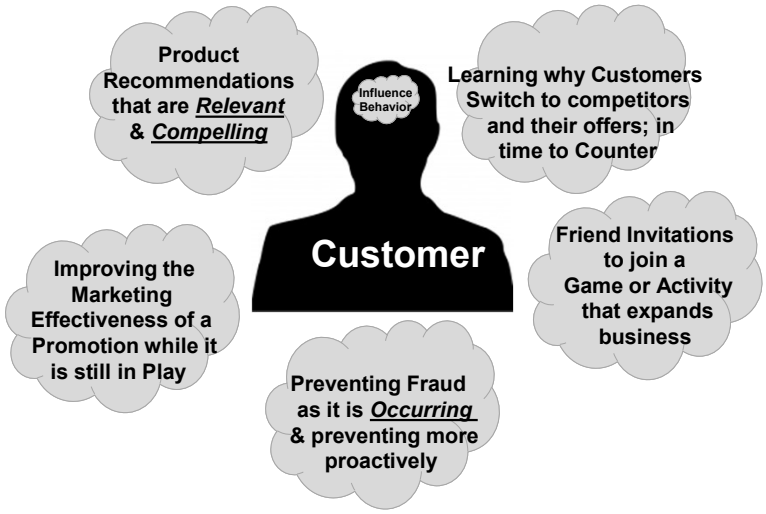
Sensor technology and networks
(measuring all kinds of data)

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

24

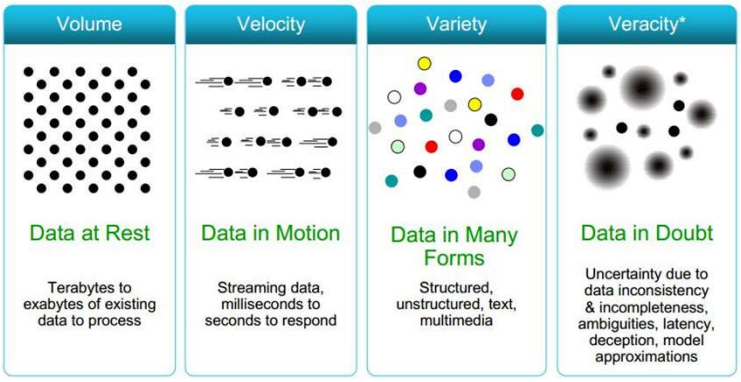
24

Real-Time Analytics/Decision Requirement



25

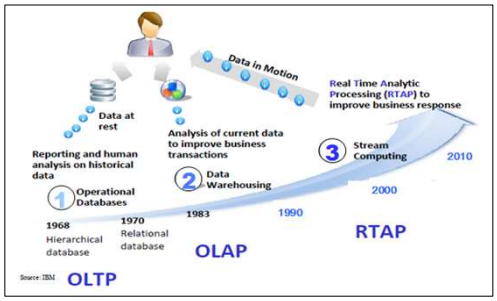
Some Make it 4V's



26

26

Harnessing Big Data



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

27

27

The Model Has Changed…

The Model of Generating/Consuming Data has Changed

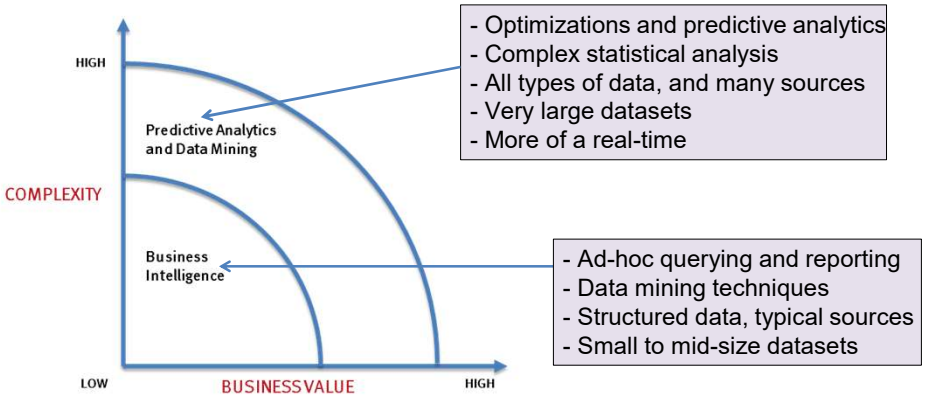
Old Model: Few companies are generating data, all others are consuming data

New Model: all of us are generating data, and all of us are consuming data

28

28

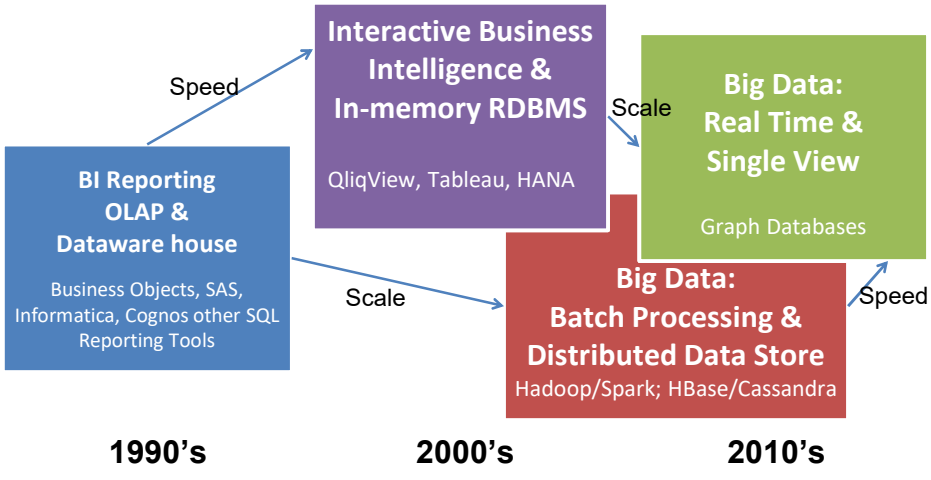
What's driving Big Data



29

29

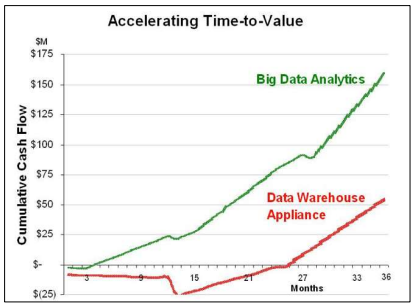
The evolution of business intelligence



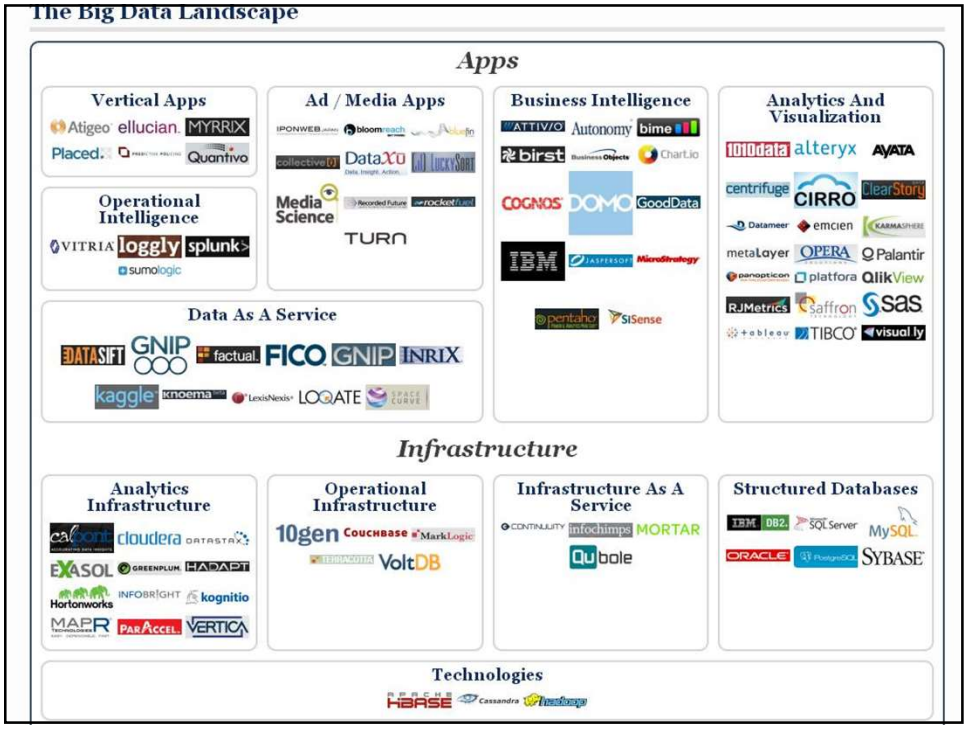
30

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

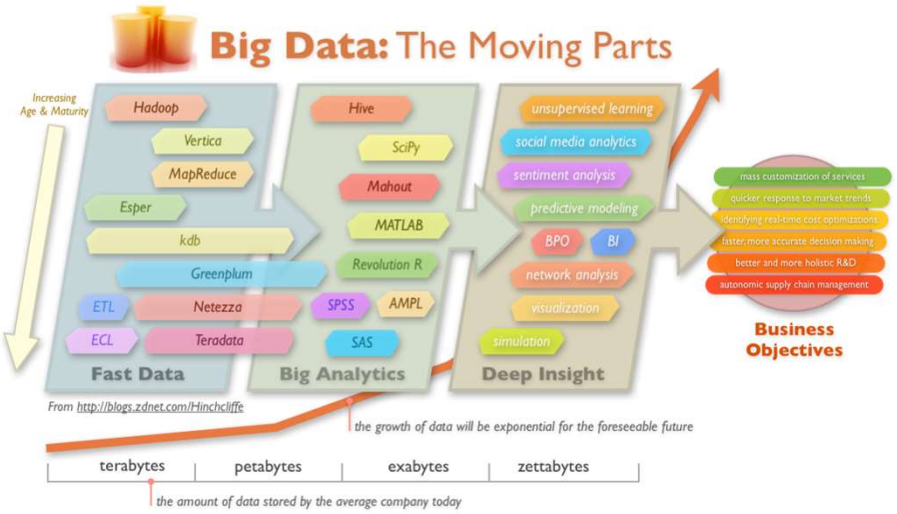


31



32

Big Data Technology



33

33

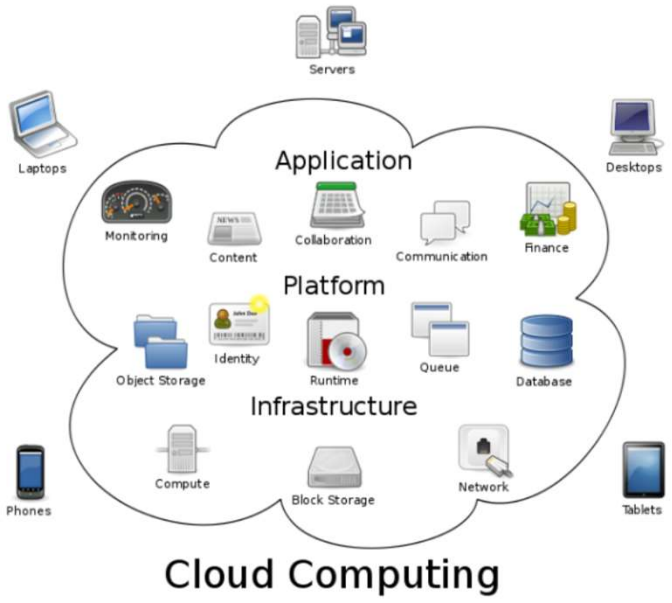
Cloud Computing

IT resources provided as a service
- Compute, storage, databases, queues

Clouds leverage economies of scale of commodity hardware
- Cheap storage, high bandwidth networks & multicore processors
- Geographically distributed data centers

Offerings from Microsoft, Amazon, Google, …

34



wikipedia:Cloud Computing

35

Benefits

Cost & management
- Economies of scale, “out-sourced” resource management

Reduced Time to deployment
- Ease of assembly, works “out of the box”

Scaling
- On demand provisioning, co-locate data and compute

Reliability
- Massive, redundant, shared resources

Sustainability
- Hardware not owned

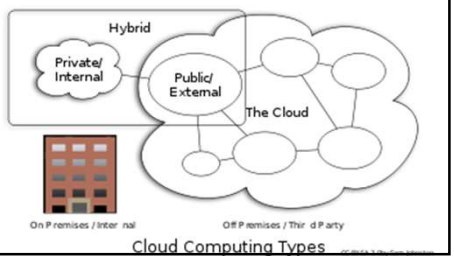
36

Types of Cloud Computing

Public Cloud: Computing infrastructure is hosted at the vendor's premises.
Private Cloud: Computing architecture is dedicated to the customer and is not shared with other organisations.
Hybrid Cloud: Organisations host some critical, secure applications in private clouds. The not so critical applications are hosted in the public cloud

Cloud bursting: the organisation uses its own infrastructure for normal usage, but cloud is used for peak loads.

Community Cloud



Cloud Computing Types

37

Classification of Cloud Computing based on Service Provided

Infrastructure as a service (IaaS)

Offering hardware related services using the principles of cloud computing. These could include storage services (database or disk storage) or virtual servers.

Amazon EC2, Amazon S3, Rackspace Cloud Servers and Flexiscale.

Platform as a Service (PaaS)

Offering a development platform on the cloud.

Google's Application Engine, Microsofts Azure, Salesforce.com's force.com .

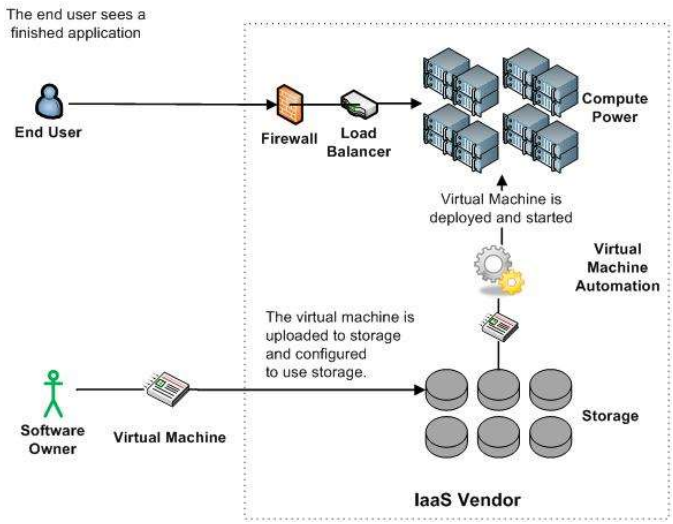
Software as a service (SaaS)

Including a complete software offering on the cloud. Users can access a software application hosted by the cloud vendor on pay-per-use basis. This is a well-established sector.

Salesforce.coms' offering in the online Customer Relationship Management (CRM) space, Googles gmail and Microsofts hotmail, Google docs.

38

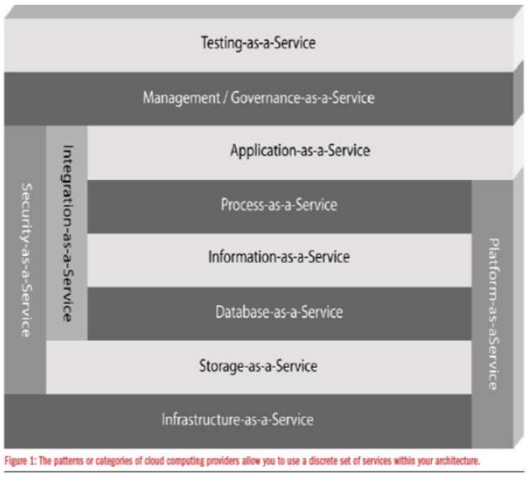
Infrastructure as a Service (IaaS)



39

More Refined Categorization

Storage-as-a-service
Database-as-a-service
Information-as-a-service
Process-as-a-service
Application-as-a-service
Platform-as-a-service
Integration-as-a-service
Security-as-a-service
Management/
Governance-as-a-service
Testing-as-a-service
Infrastructure-as-a-service



InfoWorld Cloud Computing Deep Dive

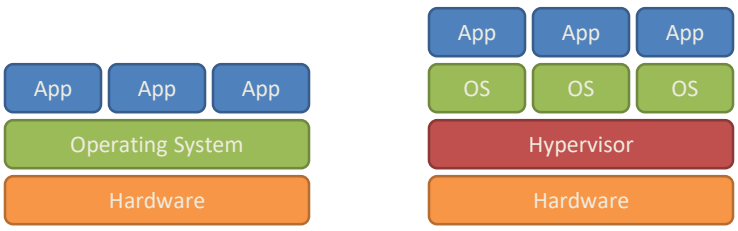
40

Key Ingredients in Cloud Computing

Service-Oriented Architecture (SOA)
Utility Computing (on demand)
Virtualization (P2P Network)
SAAS (Software As A Service)
PAAS (Platform AS A Service)
IAAS (Infrastructure AS A Servie)
Web Services in Cloud

41

Enabling Technology: Virtualization



42

Everything as a Service

Utility computing = Infrastructure as a Service (IaaS)
- Why buy machines when you can rent cycles?
- Examples: Amazon's EC2, Rackspace

Platform as a Service (PaaS)
- Give me nice API and take care of the maintenance, upgrades, …
- Example: Google App Engine

Software as a Service (SaaS)
- Just run it for me!
- Example: Gmail, Salesforce

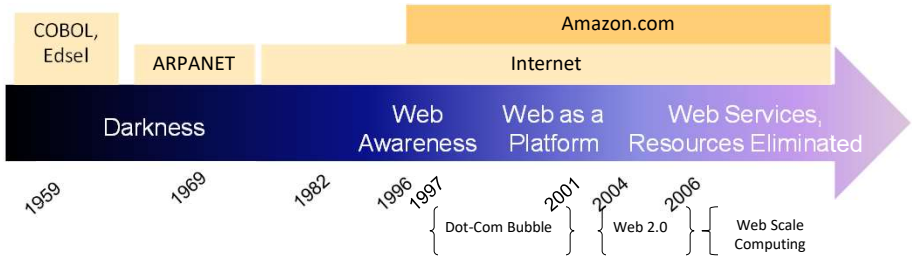
43

Cloud versus cloud

Amazon Elastic Compute Cloud
Google App Engine
Microsoft Azure
GoGrid
AppNexus

44

The Obligatory Timeline Slide
(Mike Culver @ AWS)



45

AWS

Elastic Compute Cloud – EC2 (IaaS)
Simple Storage Service – S3 (IaaS)
Elastic Block Storage – EBS (IaaS)
SimpleDB (SDB) (PaaS)
Simple Queue Service – SQS (PaaS)
CloudFront (S3 based Content Delivery Network – PaaS)
Consistent AWS Web Services API

46

What does Azure platform offer to developers?

47

Google's AppEngine vs Amazon's EC2

Python
BigTable
Other API's

VMs
Flat File Storage

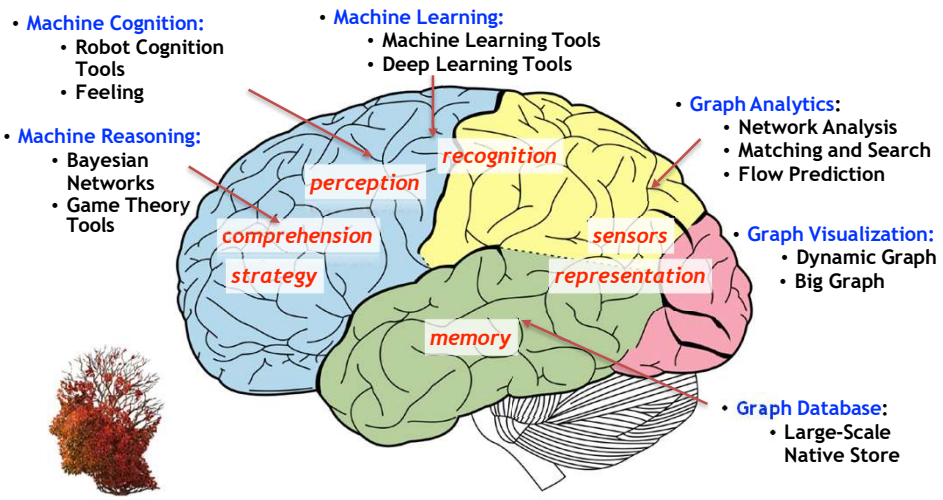
AppEngine:
Higher-level functionality
(e.g., automatic scaling)
More restrictive
(e.g., respond to URL only)
Proprietary lock-in

EC2/S3:
Lower-level functionality
More flexible
Coarser billing model

Go
ogl

48

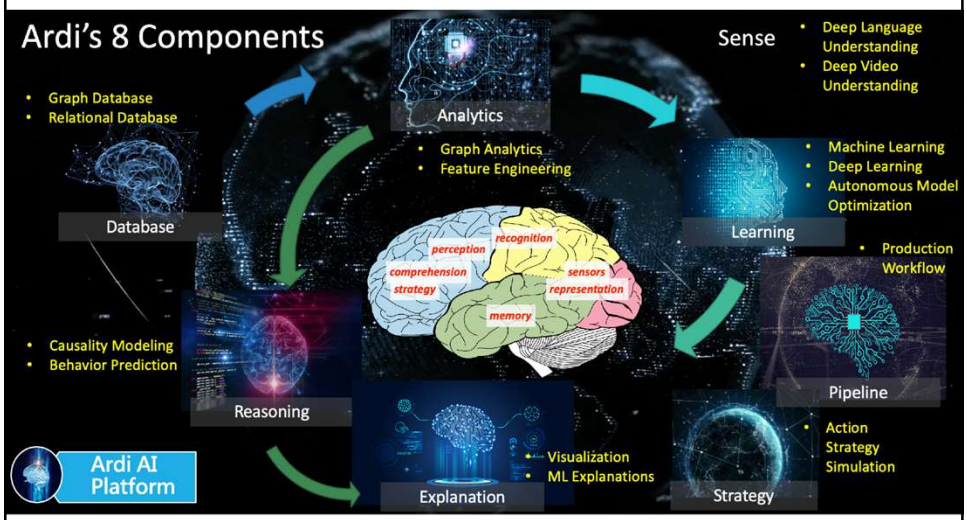
Human brain is a graph/network of 100B nodes and 700T edges.



49

49

Big Data AI Platform Example: Graphen Ardi



50

50

Why you want to take this class

- **Key Differentiator of this class:** Focusing on building a full-spectrum understanding of the latest Big Data Analytics technologies and using them to build real industry real-world solutions.
- **Sapphire Big Data Analytics Open Source Applications:** Create a Big Data open source toolsets for various industries (and disciplines)

- **Dataset and Use Cases:** Welcome!!

22

51

5 Example Big Data Use Case Categories

Big Data Exploration
Find, visualize, understand all big data to improve decision making

Enhanced 360º View of the Customer
Extend existing customer views (MDM, CRM, etc) by incorporating additional internal and external information sources

Security/Intelligence Extension
Lower risk, detect fraud and monitor cyber security in real-time

Operations Analysis
Analyze a variety of machine data for improved business results

Data Warehouse Augmentation
Integrate big data and data warehouse capabilities to increase operational efficiency

30

52

Big Data Examples -- Application Use Cases

1. Expertise Location
2. Recommendation
3. Commerce
4. Financial Analysis
5. Social Media Monitoring
6. Telco Customer Analysis
7. Healthcare Analysis
8. Data Exploration and Visualization
9. Personalized Search
10. Anomaly Detection
11. Fraud Detection
12. Cybersecurity
13. Sensor Monitoring (Smarter another Planet)
14. Cellular Network Monitoring
15. Cloud Monitoring
16. Code Life Cycle Management
17. Traffic Navigation
18. Image and Video Semantic Understanding
19. Genomic Medicine
20. Brain Network Analysis
21. Data Curation
22. Near Earth Object Analysis



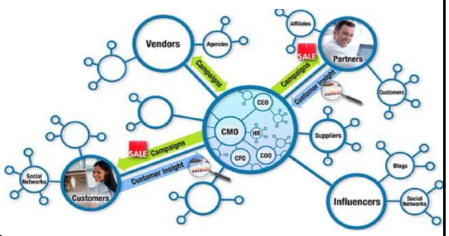


31

53

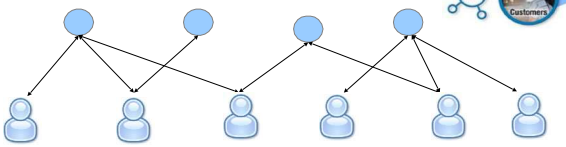
Category 1: 360º View

Recommendation



item

user



Enhancing:

Coremetrics

unica

Graph Visualizations

Communities	Graph Search	Network Info Flow	Bayesian Networks
Centralities	Graph Query	Shortest Paths	Latent Net Inference
Ego Net Features	Graph Matching	Graph Sampling	Markov Networks

Middleware and Database

54

54

Use Case 1: Social Network Analysis in Enterprise for Productivity

Production Live System used by IBM GBS since 2009 – verified ~$100M contribution

15,000 contributors in 76 countries; 92,000 annual unique IBM users

25,000,000+ emails & SameTime messages (incl. Content features)

1,000,000+ Learning clicks; 14M KnowledgeView, SalesOne, …, access d

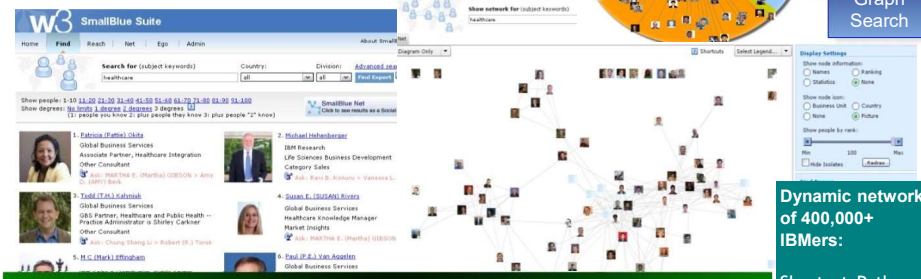
1,000,000+ Lotus Connections (blogs, file sharing, bookmark) data

200,000 people's consulting project & earning d

Shortest Paths

Centralities

Graph Search



- On BusinessWeek four times, including being the Top Story of Week, April 2009
- Help IBM earned the 2012 Most Admired Knowledge Enterprise Award
- Wharton School study: $7,010 gain per user per year using the tool
- In 2012, contributing about 1/3 of GBS Practitioner Portal $228.5 million savings an
- APQC (WW leader in Knowledge Practice) April 2013:
  "The Industry Leader and Best Practice in Expertise Location"

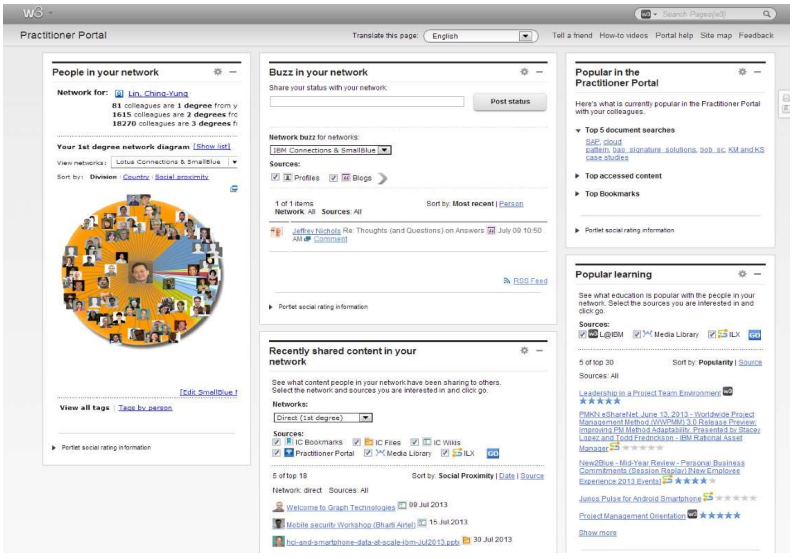
Dynamic networks of 400,000+ IBMers:

Shortest Paths
Social Capital
Bridges
Hubs
Expertise Search
Graph Search
Graph Recomm.

55

55

Use Case 2: Personalized Recommendation

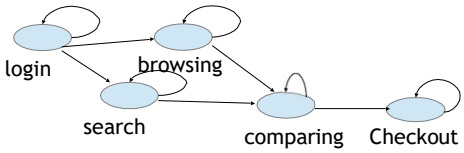


56

56

Use Case 3: Customer Behavior Sequence Analytics

Markov Network | Latent Network | Bayesian Network



- Behavior Pattern Detection
- Help Needed Detection

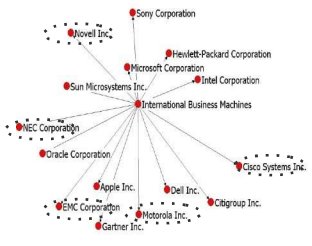
57

57

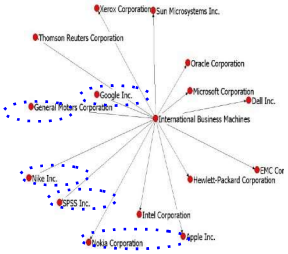
Use Case 4: Graph Analytics for Financial Analysis

Goal: *Injecting Network Graph Effects for Financial Analysis. Estimating company performance considering correlated companies, network properties and evolutions, causal parameter analysis, etc.*

- IBM 2003



- IBM 2009

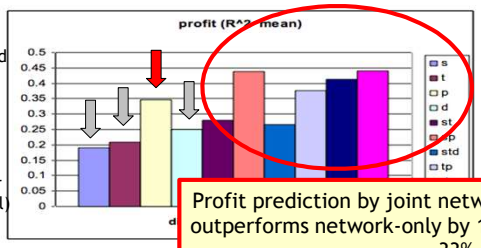


- Data Source:
  - Relationships among 7594 companies, data mining from NYT 1981 ~ 2009

Targets: 20 Fortune companies' normalized Profits

Goal: Learn from previous 5 years, and predict next year

Model: Support Vector Regression (RBF kernel)



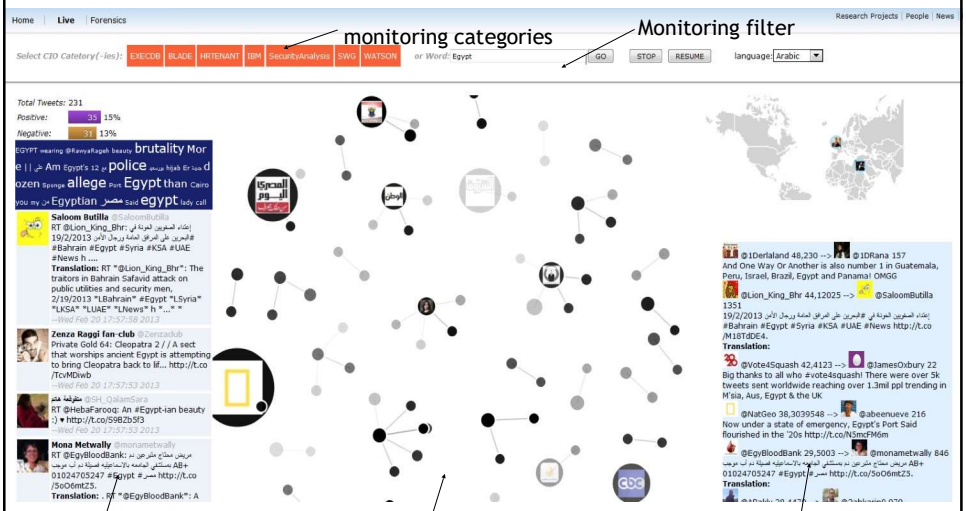
Network feature:
s (current year network feature),
t (temporal network feature),
d (delta value of network feature)
Financial feature:
p (historical profits and

Profit prediction by joint network and financial analysis outperforms network-only by 130% and financial-only by 33%.

36

58

Use Case 5: Social Media Monitoring



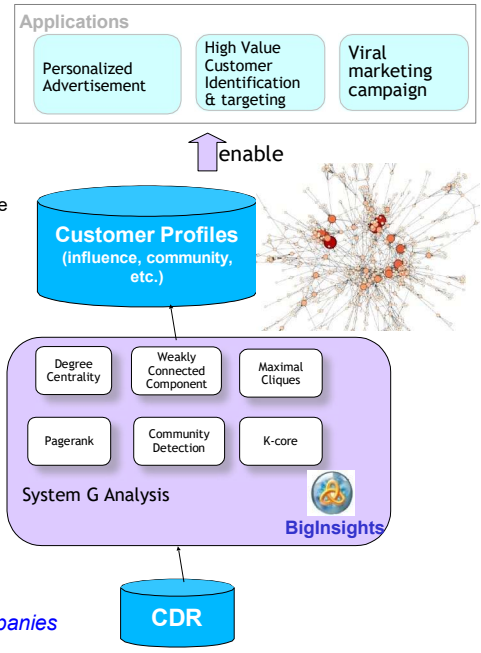
59

59

Use Case 6: Customer Social Analysis for Telco

Goal: Extract customer social network behaviors to enable Call Detail Records (CDRs) data monetization for Telco.

- Applications based on the extracted social profiles
  - Personalized advertisement (beyond the scope of traditional campaign in Telco)
  - High value customer identification and targeting
  - Viral marketing campaign
- Approach
  - Construct social graphs from CDRs based on {caller, callee, call time, call duration}
  - Extract customer social features (e.g. influence, communities, etc.) from the constructed social graph as customer social profiles
  - Build analytics applications (e.g. personalized advertisement) based on the extracted customer social profiles

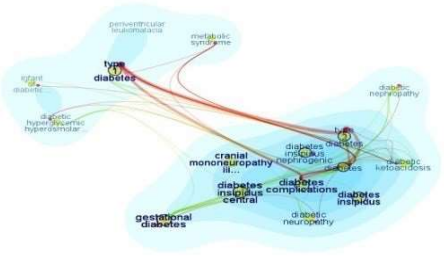


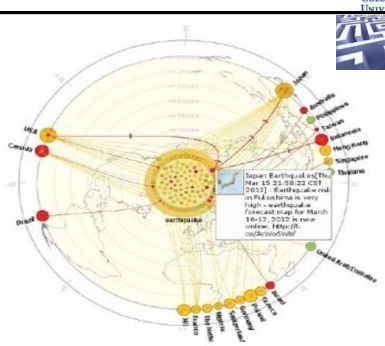
PoCs with Chinese and Indian Telecomm companies

60

60

Category 2: Data Exploration





Enhancing:

Vivísimo

CÚRAM SOFTWARE

Huge Network Visualization | Network Propagation | I2 3D Network Visualization | Geo Network Visualization | Graphical Model

Communities | Graph Search | Network Info Flow | Bayesian Networks

Centralities | Graph Query | Shortest Paths | Latent Net Inference

Ego Net Features | Graph Matching | Graph Sampling | Markov Networks

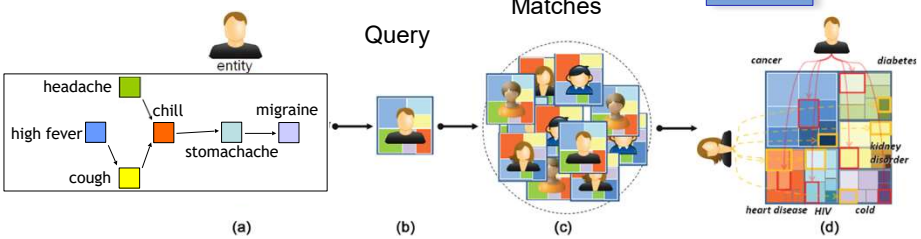
Middleware and Database

61

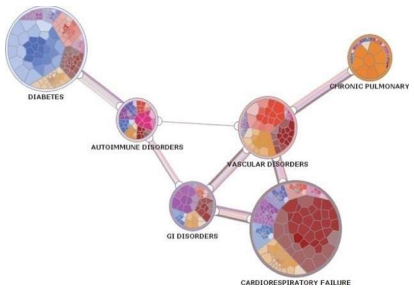
61

Use Case 7: Graph Analytics and Visualization

Graph Matching



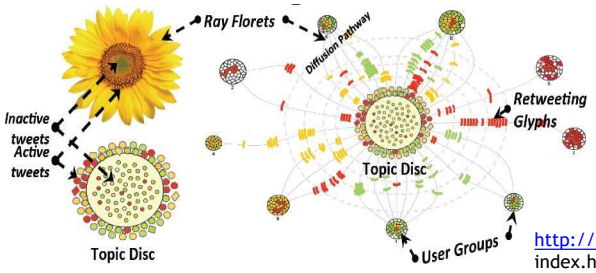
Graph Communities



62

62

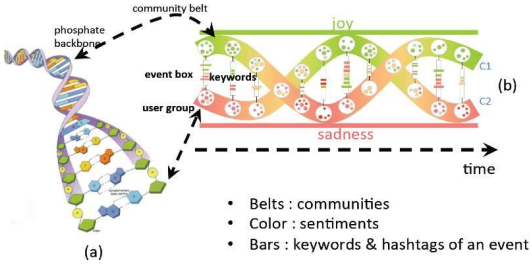
User Case 8: Visualization for Navigation and Exploration



Whisper : Tracing the information diffusion in Social Media

http://systemg.ibm.com/apps/whisper/index.html

SocialHelix: Visualizaiton of Sentiment Divergence in Social Media

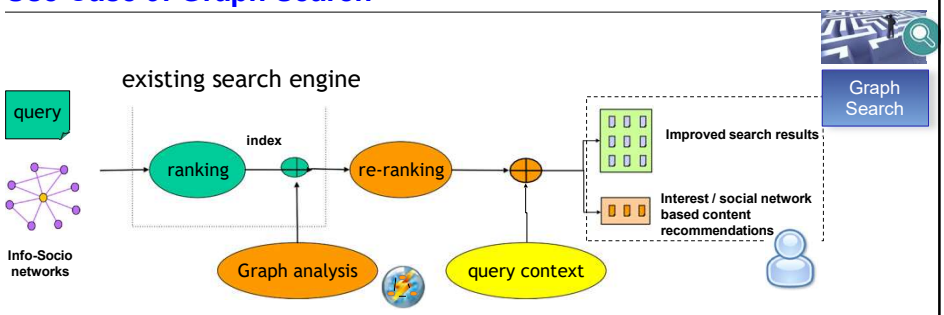


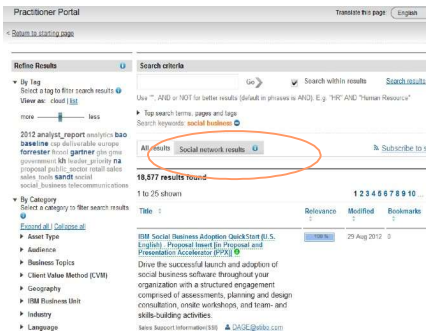
- Belts : communities
- Color : sentiments
- Bars : keywords & hashtags of an event

63

63

Use Case 9: Graph Search





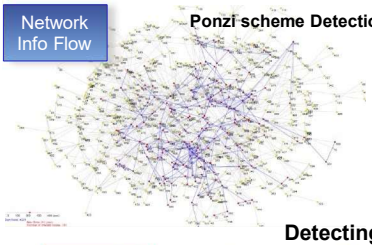
64

64

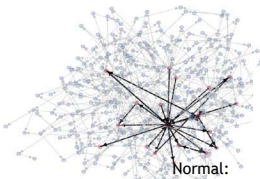
Category 3: Security

Network Info Flow

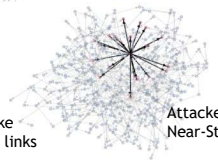
Ponzi scheme Detection



Ego Net Features

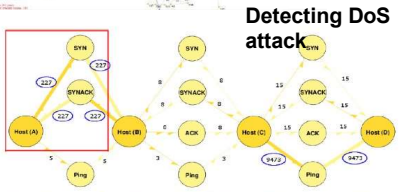


Normal:
(1)Clique-like
(2)Two-way links

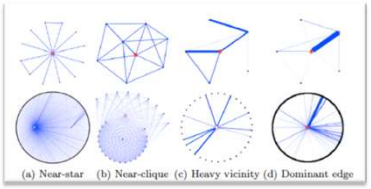


Attacker:
Near-Star

Detecting DoS attack



(a) Single large graph representing TCP SYN and ICMP PING network traffic, with two Denial of Service (DoS) attacks taking place.



(a) Near-star (b) Near-clique (c) Heavy vicinity (d) Dominant edge

Graph Visualizations

Communities	Graph Search	Network Info Flow	Bayesian Networks
Centralities	Graph Query	Shortest Paths	Latent Net Inference
Ego Net Features	Graph Matching	Graph Sampling	Markov Networks

Middleware and Database

65

65

Use Case 10: Anomaly Detection at Multiple Scales

Based on President Executive Order 13587

Goal: System for Detecting and Predicting Abnormal Behaviors in Organization, through large-scale social network & cognitive analytics and data mining, to decrease insider threats such as espionage, sabotage, colleague-shooting, suicide, etc.

"Enterprise Information Leakage Impacted economy and jobs" Feb 2013

"What's emerged is a multibillion dollar detective industry"
npr Jan 10, 2013

Emails
Instant Messaging
Web Access
Executed Processes
Printing
Copying
Log On/Off

Social sensors
Click streams capturer
Feed subscription
Database access

Graph analysis
Behavior analysis
Semantics analysis
Psychological analysis

Multimodality Analysis

Detection, Prediction & Exploration Interface

Infrastructure + ~ 490 Analytics

66

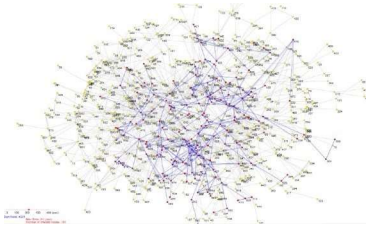
66

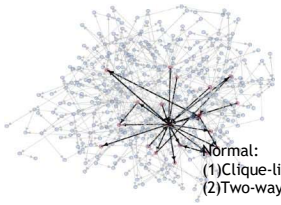
Use Case 11: Fraud Detection for Bank

Network Info Flow

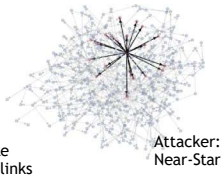
Ego Net Features

Ponzi scheme Detection

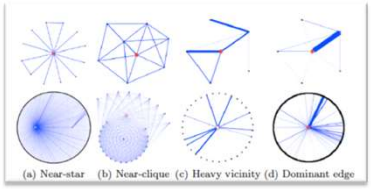




Normal:
(1)Clique-like
(2)Two-way links



Attacker:
Near-Star



67

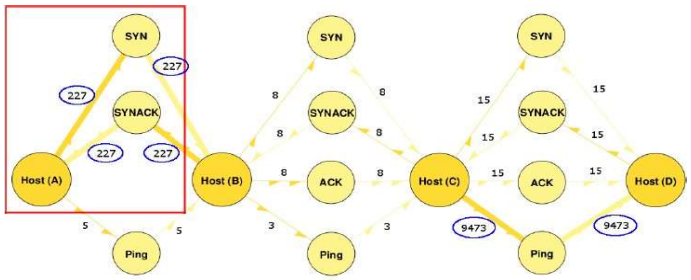
67

Use Case 12: Detecting Cyber Attacks

Network Info Flow

Ego Net Features

Detecting DoS attack

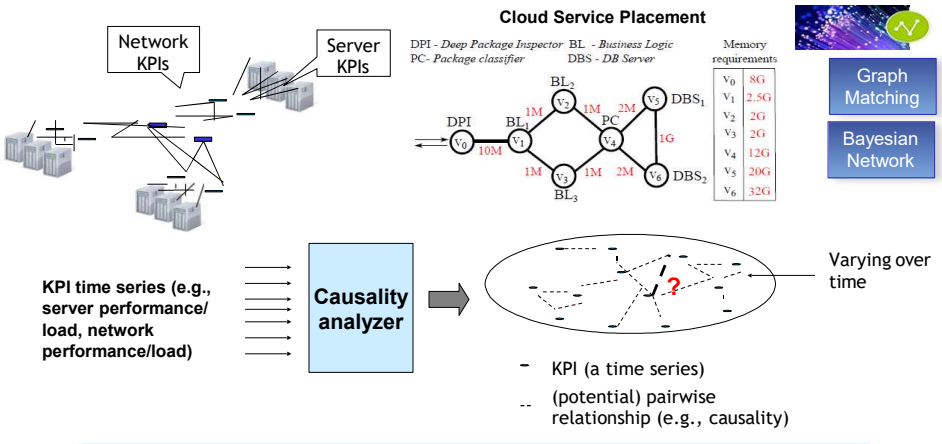


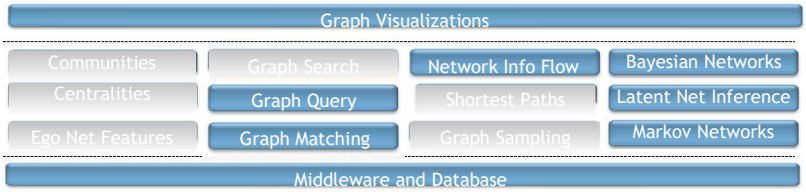
(a) Single large graph representing TCP SYN and ICMP PING network traffic, with two Denial of Service (DoS) attacks taking place.

68

68

Category 4: Operations Analysis





69

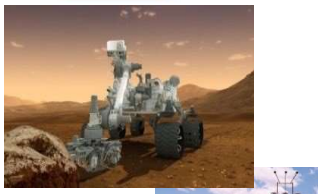
69

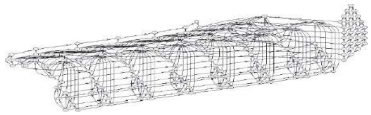
Use Case 13: Smarter *another* Planet

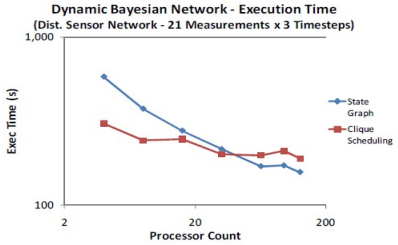
Bayesian Network

Goal: Atmospheric Radiation Measurement (ARM) climate research facility provides *24x7 continuous field observations* of cloud, aerosol and radiative processes. **Graphical models** can automate the validation with improvement efficiency and performance.

Approach: BN is built to represent the dependence among sensors and replicated across timesteps. BN parameters are learned from over *15 years* of ARM climate data to support distributed climate sensor validation. Inference validates sensors in the connected instruments.







Bayesian Network
* 3 timesteps * 63 variables
*3.9 avg states * 4.0 avg indegree
* 16,858 CPT entries

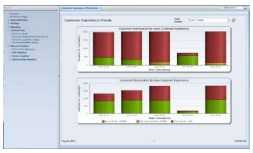
Junction Tree
* 67 cliques
* 873,064 PT entries in cliques

70

70

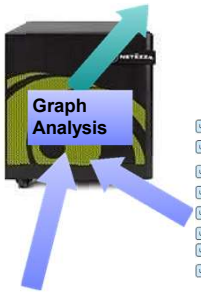
Use Case 14: Cellular Network Analytics in Telco Operation

Goal: Efficiently and uniquely identify *internal* state of Cellular/Telco networks (e.g., performance and load of network elements/links) using probes between monitors placed at selected network elements & endhosts



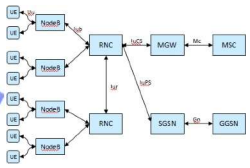
Network load level report

- Applied Graph Analytics to telco network analytics based on CDRs (call detail records): estimate traffic load on CSP network with low monitoring overhead
  - (1)CDRs, already collected for billing purposes, contain information about voice/data calls
  - (2)Traditional NMS* and EMS** typically lack of end-to-end visibility and topology across vendors
  - (3)Employ graph algorithms to analyze network elements which are not reported by the usage data from CDR information
- Approach
  - Cellular network comprises a hierarchy of network elements
  - Map CDR onto network topology and infer load on each network element using graph analysis
  - Estimate network load and localize potential problems



CDR

Network topology



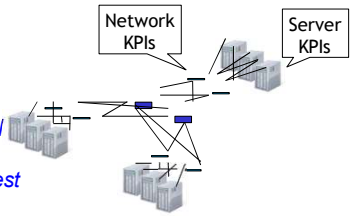
71

71

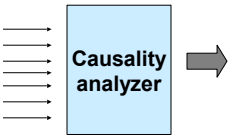
Use Case 15: Monitoring Large Cloud

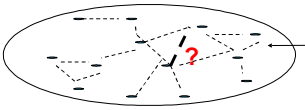
Goal: *M*onitoring technology that can track the time-varying state (e.g., causality relationships between KPIs) of a large Cloud when the processing power of monitoring system cannot keep up with the scale of the system & the rate of change

- *Causality relationships (e.g., Granger causality) are crucial performance monitoring & root cause analysis*
- *Challenge: easy to test pairwise relationship, but hard to test multi-variate relationship (e.g., a large number of KPIs)*



KPI time series (e.g., server performance/load, network performance/load)





KPI (a time series)

(potential) pairwise relationship (e.g., causality)

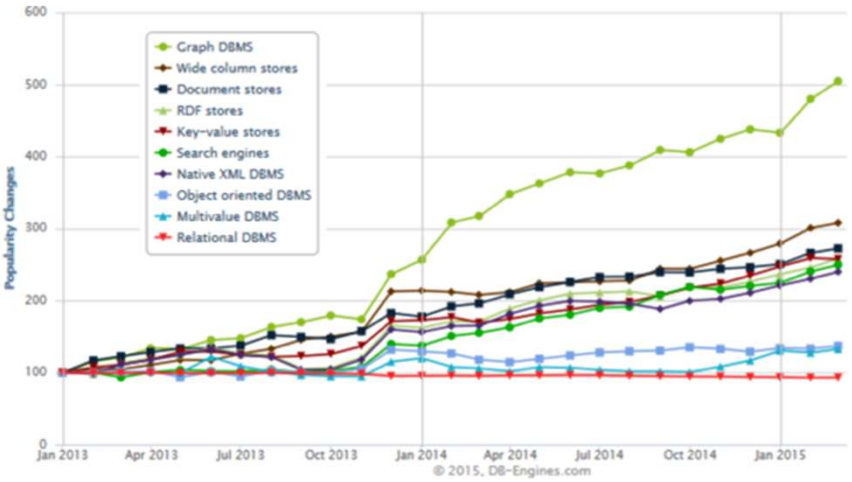
Our approach: Probabilistic monitoring via sampling & estimation

Select KPI pairs (sampling)→ Test link existence → Estimate unsampled links based on history → Overall graph

50

72

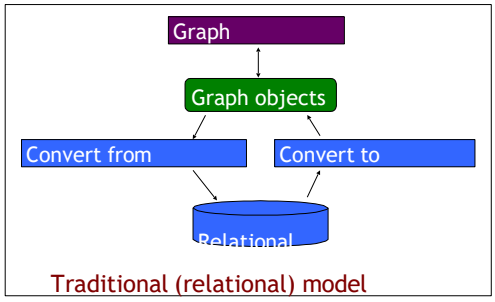
Category 5: Data Warehouse Augmentation

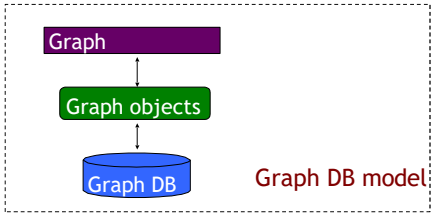


73

73

Use Case 16: Code Life Cycle Improvement





- Advantages of working directly with graph DB for graph applications
  - (1) Smaller and simpler code
  - (2) Flexible schema ! easy schema evolution
  - (3) Code is easier and faster to write, debug and manage
  - (4) Code and Data is easier to transfer and maintain

74

74

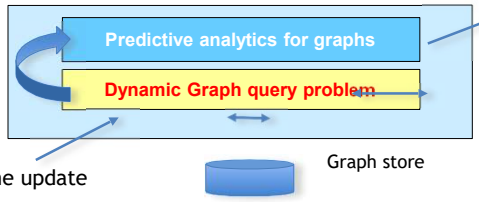
Use Case 17: Smart Navigation Utilizing Real-time Road Information

Goal: *Enable unprecedented level of accuracy in **traffic scheduling** (for a fleet of transportation vehicles) and navigation of individual cars utilizing the **dynamic real-time information** of changing road condition and predictive analysis on the data*

- *Dynamic graph algorithms implemented in System G provide **highly efficient graph query computation** (e.g. shorted path computation) on time-varying graphs (order of magnitudes improvement over existing solutions)*

- *High-throughput **real-time predictive analytics** on graph makes it possible to estimate the future traffic condition on the route to make sure that the decision taken now is optimal overall*

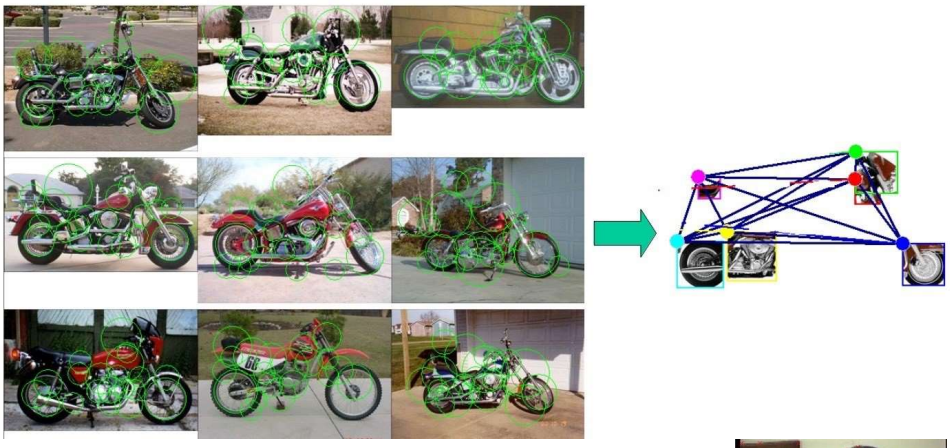
Our approach: Querying over dynamic graph + predictive analytics on graph properties

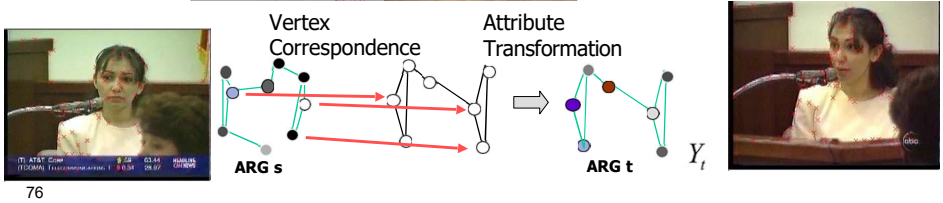


75

75

Use Case 18: Graph Analysis for Image and Video Analysis





76

76

Use Case 19: Graph Matching for Genomic Medicine

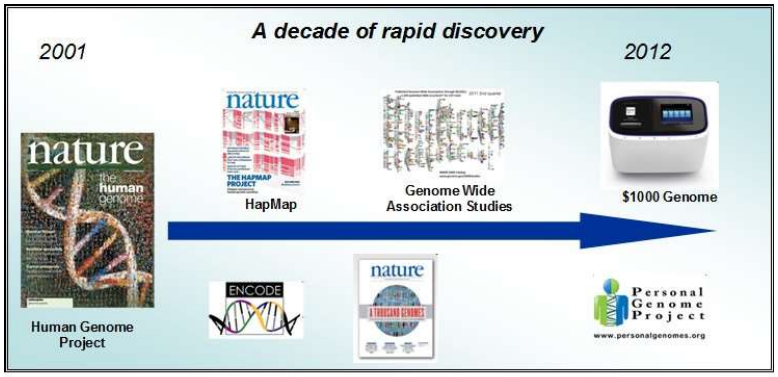
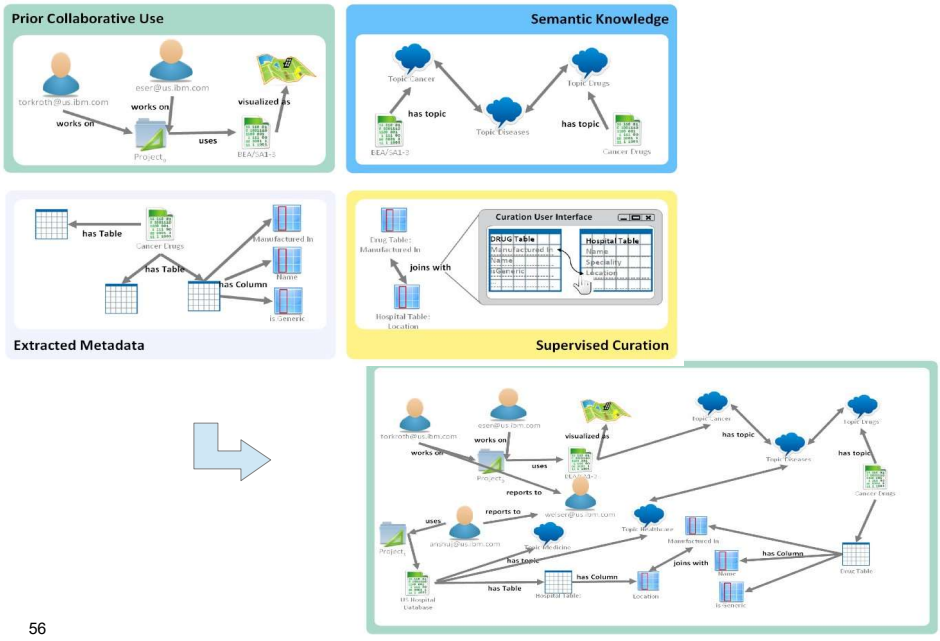


Figure 1: Since the Human Genome Project, various projects have started to reveal the mysteries of genomes and the $1000 Genome is almost reality.

77

77

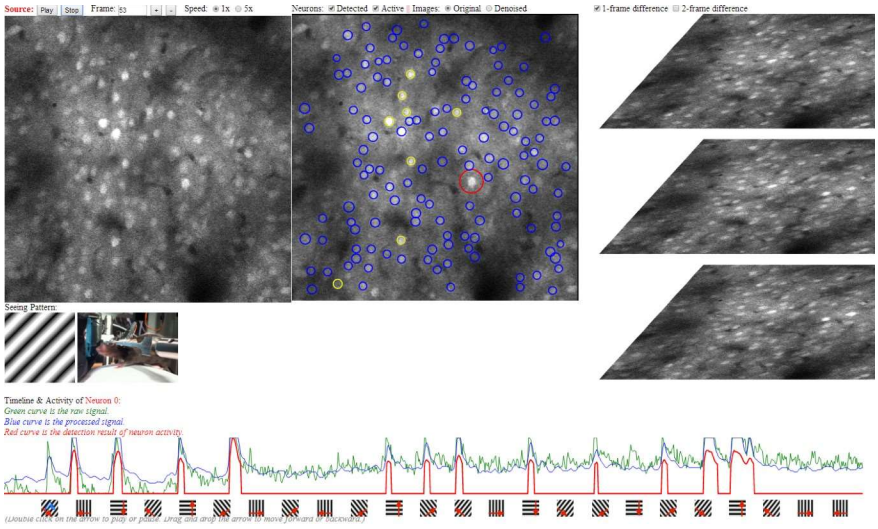
Use Case 20: Data Curation for Enterprise Data Management



56

78

Use Case 21: Understanding Brain Network



79

79

Use Case 22: Planet Security

- Big Data on Large-Scale Sky Monitoring

80

80