IIT KHARAGPUR

NPTEL ONLINE CERTIFICATION COURSES

IIoT: Industrial Internet of Things – Part I

Dr. Sudip Misra

Associate Professor

Department of Computer Science and Engineering

IIT Kharagpur

Email: smisra@sit.iitkgp.ernet.in

Website: http://cse.iitkgp.ac.in/~smisra/

"IoT as a concept has crossed the chasm from slideware to reality with many industries implementing IoT solutions."

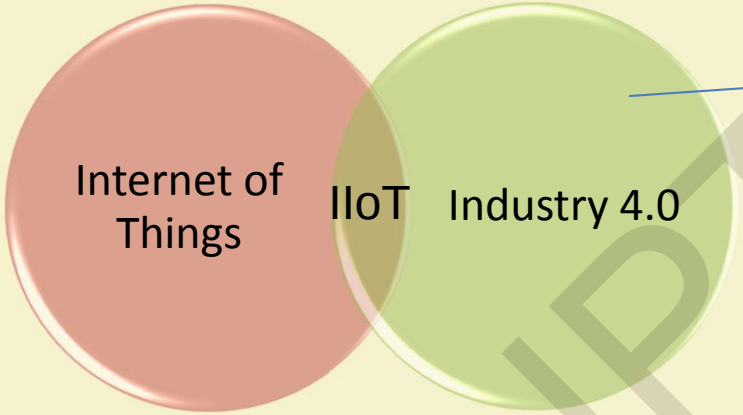
- Paul Howarth, Senior Manager, Corporate Development, CISCO

Source : http://www.mcrockcapital.com

Introduction

- ✓ The main aim of Internet of Things (IoT) is
  - ✓ to globally connect smart ‘things’ or ‘objects’ .
  - ✓ objects are uniquely identified.
  - ✓ interoperability among the objects.
- ✓ The Industrial Internet of Things (IIoT) is an application of IoT in industries to modify the various existing industrial systems. IIoT links the automation system with enterprise, planning and product lifecycle.

Introduction (contd.)



- Automation and data exchange in manufacturing technologies
- Cyber-physical systems, the Internet of things and cloud computing
- Smart factory

Fig 1(a) : IIoT as an intersection of industries and IoT

Introduction (contd.)

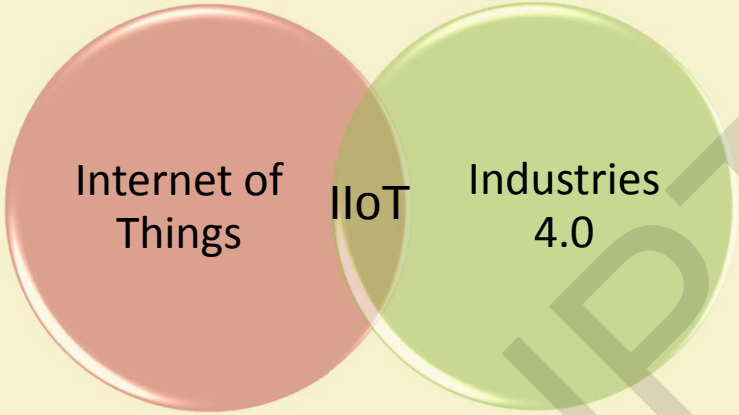


Fig 1(a) : IIoT as an intersection of industries and IoT

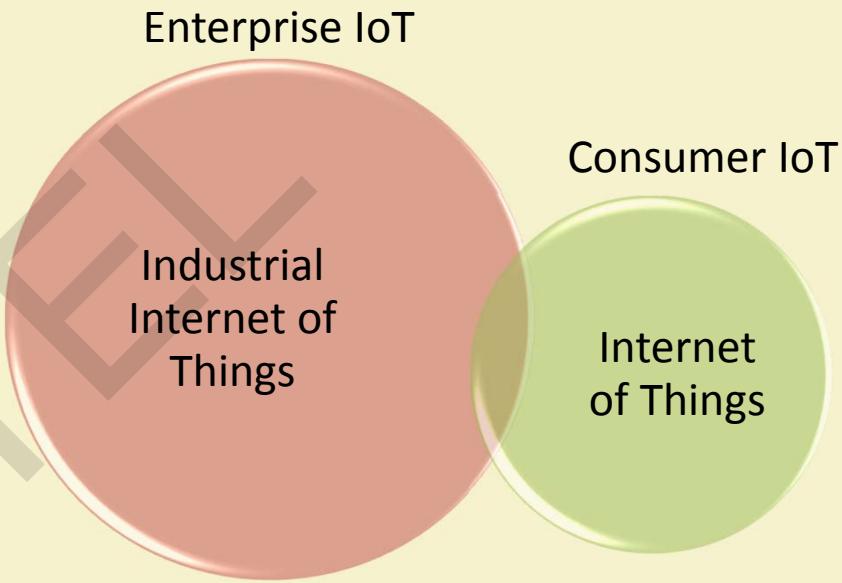


Fig 1(b) : IIoT ≠ IoT

Fig 1 : IIoT Platform

Introduction (contd.)

- ✓ IIoT includes –
  - ✓ machine learning
  - ✓ big data technology
  - ✓ machine - to - machine interaction (M-2-M)
  - ✓ automation.
- ✓ IIoT is supported by huge amount of data collected from sensors. It is based on “wrap & re-use” approach, rather than “rip & replace” approach.

(Source : http://www.mhi.org)

Introduction (contd.)

- ✓ 1st Industrial Revolution : Mechanized production
- ✓ 2nd Industrial Revolution : Mass production
- ✓ 3rd Industrial Revolution : Internet evolution and automation
- ✓ 4th Industrial Revolution : IIoT

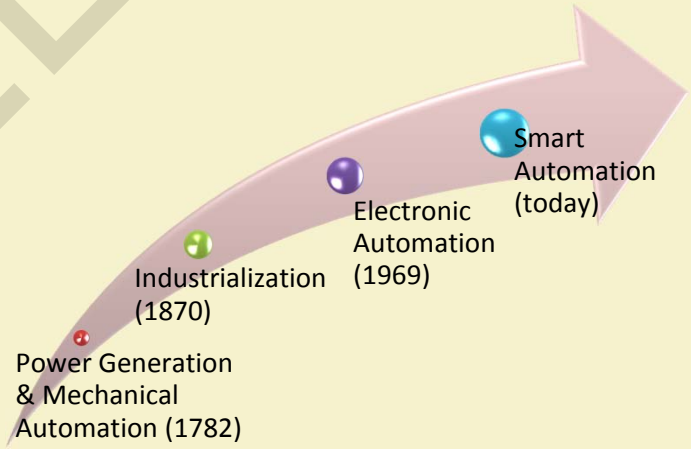


Fig 2: Industry 4.0

Source: http://www.industry40wood.com

Introduction (contd.)

Source: https://www.artika.info

Source: http://www.rehm-group.com

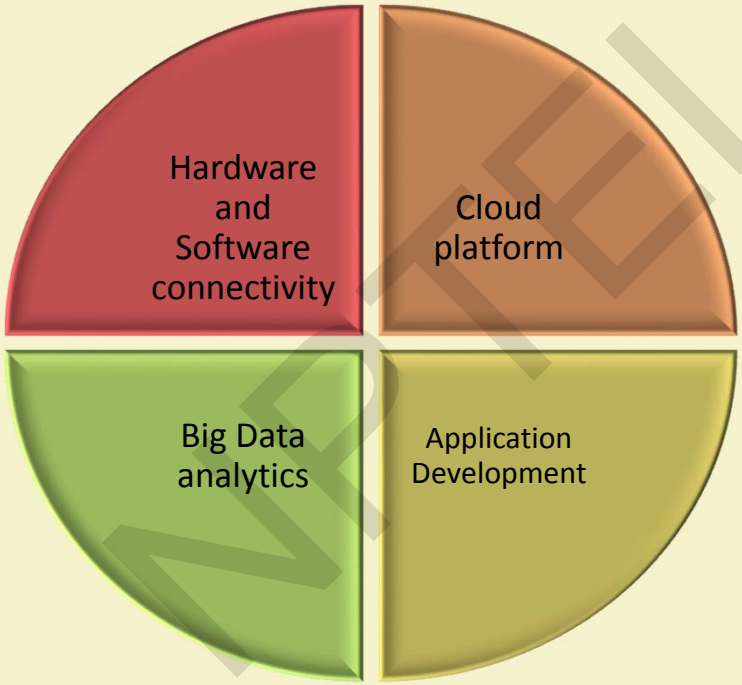
Cloud computing

IIoT : 2nd generation of Internet evolution and 4th Industrial Automation

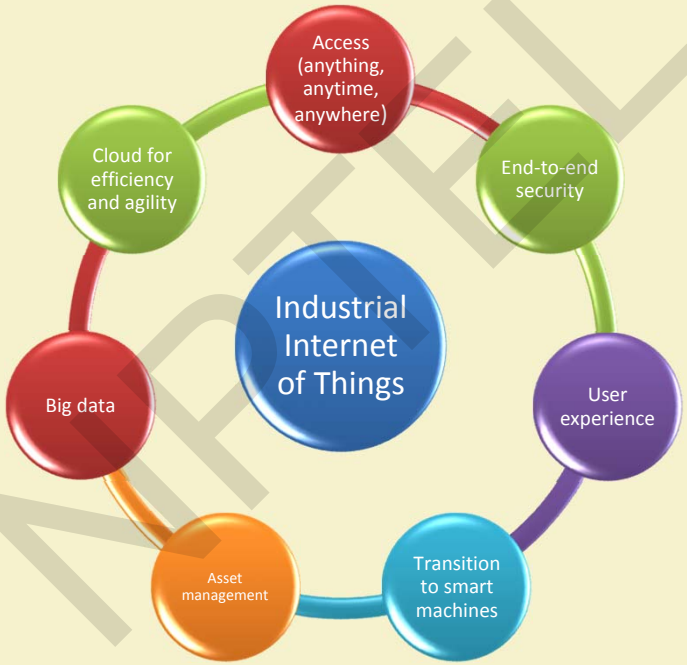
Introduction (contd.)

- ✓ IIoT is a network of
  - ✓ physical objects
  - ✓ systems
  - ✓ platforms
  - ✓ applications
- ✓ These networks can communicate with each other, external environment and other people.
- ✓ The acquisition of IIoT has led to availability and affordability of sensors, processors, and other technologies which facilitates capture and access to real-time information

IIoT Requirements



IIoT Requirements (contd.)



IIoT Requirements (contd.)

Physical Plant

Sensor readings

Virtual Plant

Machine instructions

Design Considerations

- ✓ To use an IoT device for industrial applications, the following design objectives are to be considered –
  - ✓ Energy : Time for which the IoT device can operate with limited power supply.
  - ✓ Latency : Time required to transmit the data.
  - ✓ Throughput : Maximum data transmitted across the network.
  - ✓ Scalability : Number of devices supported.
  - ✓ Topology: Communication among the devices, i.e. interoperability.
  - ✓ Safety and Security: Degree of safety and security of the application.

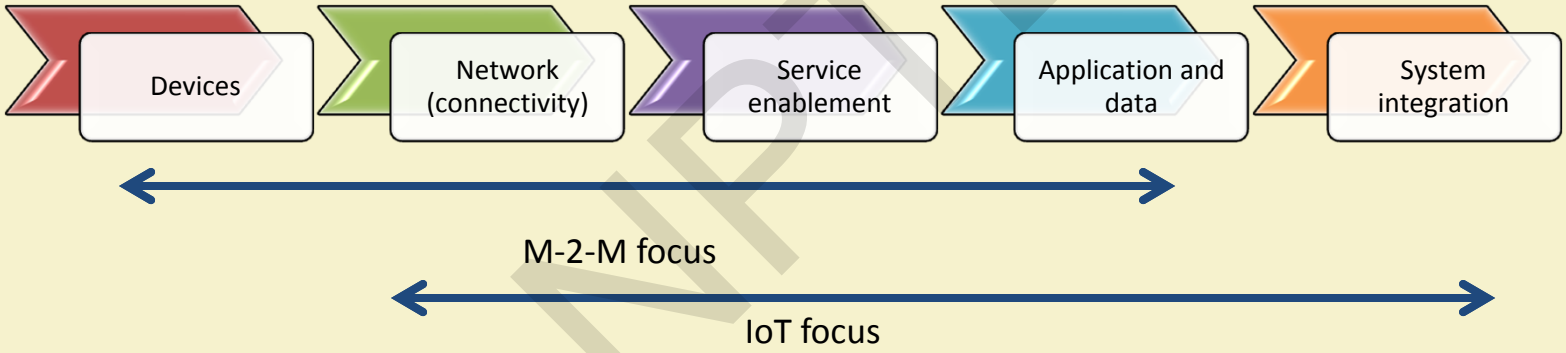
Difference between IoT and IIoT

Difference between IoT and IIoT

The main differences between IoT and IIoT are :

IoT	IIoT
• Focused on convenience of individuals • M-2-M communication: Limited • Applications areas are at consumer-level	• Focused on efficiency, safety and security of the operation. • M-2-M communication: Extensively. • Application areas are at industries.

Difference between IoT and IIoT (contd.)



Service Management in IIoT

- ✓ "Service management refers to the implementation and management of the quality of services which meets the end-users demand"
- ✓ "Service is a collection of data and associated behaviors to accomplish a particular function or feature of a device or portions of a device".

Source: Ning Lu, Nan Cheng, Ning Zhang, Xuemin Shen, Jon W. Mark, Connected Vehicles : Solutions and Challenges, IEEE Internet of Things Journal, Vol. 1, No. 4, August 2014.

Service Management in IIoT

- ✓ Service can be of two types, which are -
  - ✓ Primary service - The basic services which are responsible for the primary node functions are termed as primary service.
  - ✓ Secondary service - The auxiliary functions which provide services to the primary service or secondary services are termed as secondary service.

Thank You!!

NPTEL ONLINE
CERTIFICATION COURSES

IIoT: Industrial Internet of Things – Part II

Dr. Sudip Misra
Associate Professor
Department of Computer Science and Engineering
IIT Kharagpur
Email: smisra@sit.iitkgp.ernet.in
Website: http://cse.iitkgp.ac.in/~smisra/

Applications of IIoT

- ✓ The key application areas of IIoT are -
  - ✓ Manufacturing industry
  - ✓ Healthcare Service industry
  - ✓ Transportation & logistics
  - ✓ Mining
  - ✓ Firefighting

Manufacturing Industry

- ✓ The devices, equipment, workforce, supply chain, work platform are integrated and connected to achieve smart production. This will led to –
  - ✓ reduction in operational costs
  - ✓ improvement in the productivity of the worker
  - ✓ reduction in the injuries at the workplace
  - ✓ resource optimization and waste reduction
  - ✓ end-to-end automation.

Healthcare Service Industry

- ✓ Patients can be continuously monitored due to the implanted on-body sensors. This has led to –
  - ✓ improved treatment outcome
  - ✓ costs has reduced
  - ✓ improved disease detection
  - ✓ improved accuracy in the collection of data
  - ✓ improved drugs management.

Transportation & logistics

- ✓ To improve safety, efficiency of transportation, Intelligent Transportation system (ITS) is developed which consists of connected vehicles. ITS provides –
  - ✓ Vehicle – to – sensor connectivity
  - ✓ Vehicle – to – vehicle connectivity
  - ✓ Vehicle – to – internet connectivity
  - ✓ Vehicle – to – road infrastructure
- ✓ Dedicated short-range communications (DSRC) is the key enabling technology for V2V and V2R communications.

Transportation & logistics

- ✓ In IIoT scenario the physical objects are provided with
  - ✓ bar codes
  - ✓ RFID tags

  hence, real-time monitoring of the status and location of the physical objects from destination to the origin, across the supply chain is possible.
- ✓ Security and privacy of the data should be maintained.

Mining

- ✓ To prevent accidents inside the mines - RFID, Wi-Fi and other wireless technologies are used, which
  - ✓ provides early warning of any disaster
  - ✓ monitors air-quality
  - ✓ detects the presence of poisonous gases inside the mines
  - ✓ oxygen level inside the mines.

Firefighting

- ✓ Sensor networks, RFID tags are used to perform
  - ✓ automatic diagnosis
  - ✓ early warning of disaster
  - ✓ emergency rescue
  - ✓ provides real-time monitoring

  Hence, improves public security.

Examples of IIoT

- ✓ Examples of IIoT are -
  - ✓ unmanned aerial vehicles (UAVs) to inspect oil pipelines.
  - ✓ monitoring food safety using sensors.
  - ✓ minimizing workers' exposure to noise, chemicals and other hazardous gases.
  - ✓ unmanned marine vehicle which can collect data up to a year without fuel or crew.

Connected Ecosystems in IIoT scenario

- ✓ Traditional supply chains in industries are linear in nature.
- ✓ To shift the business focus from products to outcomes, new ecosystem should be followed.
- ✓ Digital ecosystems progress at a much faster rate than physical industries. Hence, it can quickly adapt to the changes in the external environments.

Integration of Digital and Human Workforce

- ✓ In IIoT, machines become more intelligent. Hence, the automated tasks can be done in the industries at lower costs and higher quality level.
- ✓ Humans will work with machines, the outcome will be higher overall productivity.
- ✓ IIoT will reform and redefine the skills of the workers.

Creation of New Jobs

- ✓ The creation of new composite industries, such as precision agriculture, digital healthcare system, digital mines etc., will lead to development of new job opportunities.
- ✓ Highly automated machines will require lesser number of unskilled workers, but will require skilled experts with digital and analytical skills.

Reformation of Robots

- ✓ In IIoT environment, robots are featured with three capabilities : sensing, thinking and acting. They will be reformed with the ability to carry out repetitive tasks.
- ✓ Robots will be more intelligent but will work under the supervision of human beings. Their availability will increase.
- ✓ Robots will be reprogrammable to perform new tasks. They have the capability to 'learn' faster.

Challenges in IIoT

✓ Primary challenges

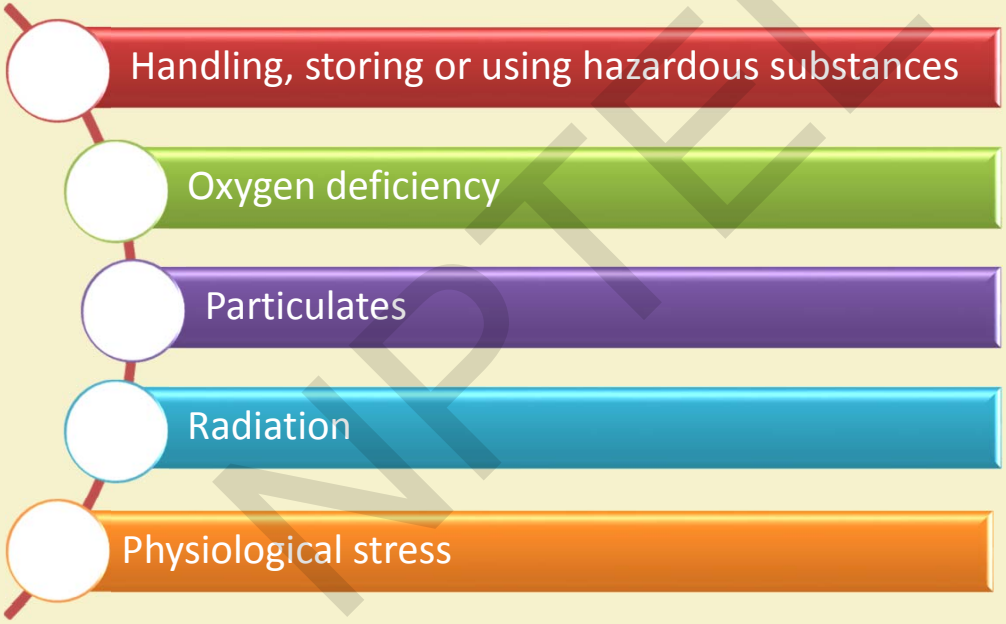
- Identification of objects or things
- Manage huge amount of data
- Integrate existing infrastructures into new IIoT infrastructure
- Enabling data storage

Challenges in IIoT(contd.)

✓ Safety Challenges

Challenges in IIoT(contd.)

✓ Hazards (related)



Challenges in IIoT(contd.)

Standardization

- ✓ Standardization plays an important role in the development of the system.
- ✓ Goal: To improve the interoperability of the different systems/ applications and allow the products/services to perform better.

Challenges in IIoT(contd.)

Standardization

- ✓ The problems related to standardization are:
  - ✓ Interoperability
  - ✓ Semantic interoperability (data sematics)
  - ✓ Security and privacy
  - ✓ Radio access level issues.

Challenges in IIoT(contd.)

Privacy and security issues

- ✓ The two most important concerns related with IIoT are -
  - ✓ information security
  - ✓ data privacy protection
- ✓ The devices/things can be tracked, monitored and connected. So there are chances of attack on the personal and private data.

Challenges in IIoT(contd.)

Privacy and security issues

- ✓ Examples –
  - ✓ Healthcare industry – the medical data of a patient must not be tampered, or altered by any person in the middle.
  - ✓ Food industry – the deterioration of any food item being sent to the company must be kept confidential as it will affect the reputation of the company.

Risks associated with IIoT in Manufacturing

- ✓ Though IIoT provides new opportunities, but few factors may cause hindrance in the path to success, which are :
  - ✓ lack of vision and leadership
  - ✓ lack of understanding of values among management employees
  - ✓ costly sensors
  - ✓ inadequate infrastructure.

Meet the challenges: Sensor improvement

- ✓ Improvement in sensor technologies –
  - ✓ miniaturization
  - ✓ performance
  - ✓ cost and energy consumption.

Meet the challenges : Manufacturing

- ✓ Manufacturers use software capabilities to improve operational efficiency through –
  - ✓ predictive maintenance
  - ✓ savings on scheduled repairs
  - ✓ reduced maintenance costs
  - ✓ reduced number of breakdowns.

Case study : Rt Tech Software

- ✓ Rt Tech particularizes in software which –
  - ✓ improves industrial facilities' efficiency
  - ✓ improves productivity.
- ✓ Energy management solution, which leads to reduction in the plant's highest variable cost.
- ✓ Rt Tech automates the process of mapping and managing energy consumption.

Source : http://www.mcrockcapital.com

PRODUCTS DEVELOPED

- ✓ M-2-M communication : Intelligent Radio Modem (IRM)
  - ✓ IRM 1500 & ACE 1000 - IRM
    - ✓ simple
    - ✓ M-2-M connectivity
    - ✓ data transmission
  - ✓ These devices provide easy maintenance and installation. They can be connected to IP and non-IP serial devices to extend the capability to monitor and communicate with other technologies.

Source : https://www.motorolasolutions.com

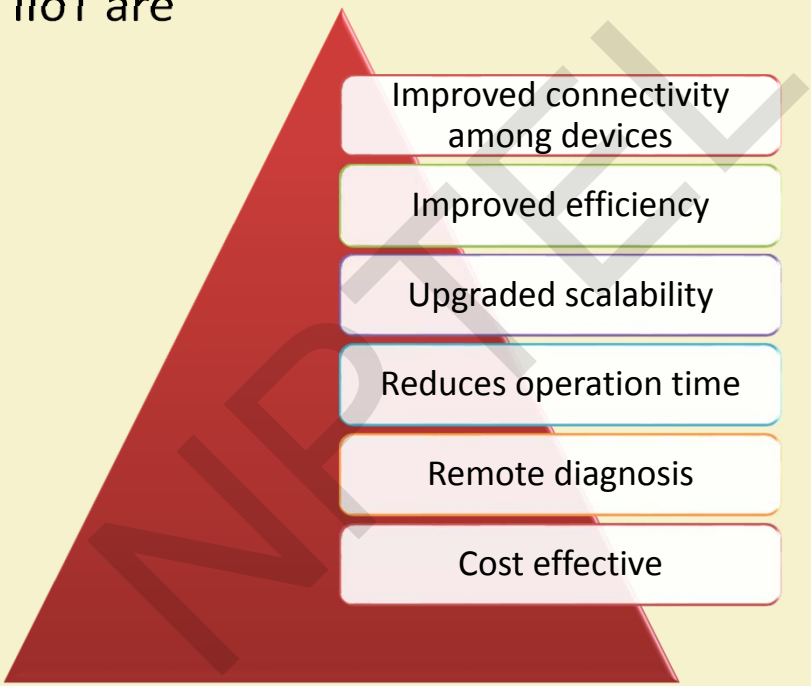
PRODUCTS DEVELOPED (contd.)

- ✓ Comtrol – IO Link Master Gateway
  - ✓ It can be easily integrated into the industrial network with existing and new installations.
  - ✓ It supports Ethernet/IP, PROFINET (PNIO) and Modbus TCP.

Source :
http://pdfserv.maximintegrated.com
http://www.comtrol.com

Benefits of IIoT

✓ The benefits of IIoT are



Recent Research trends in IIoT

- ✓ Recent research challenges in IIoT are -
  - ✓ To improve the communications among the different things or objects.
  - ✓ To develop energy-efficient techniques so as to reduce power consumption by sensors.
  - ✓ To develop context-aware IoT middleware for better understanding of the sensor data.
  - ✓ To create smart objects with larger memory, processing and reasoning capabilities.

Conclusion

- ✓ IIoT system requires the following :
  - ✓ Smaller, less expensive sensors which makes them easily accessible.
  - ✓ Distributed control of assembly line, automated monitoring, control and maintenance.

References

- ✓ Daniele Miorandi, Sabrina Sicari, Francesco De Pellegrini, Imrich Chlamtac, Internet of things: Vision, applications and research challenges, Ad Hoc Networks, Volume 10, Issue 7, September 2012.
- ✓ http://internetofthingsagenda.techtarget.com/definition/Industrial-Internet-of-Things-IIoT.
- ✓ Ning Lu, Nan Cheng, Ning Zhang, Xuemin Shen, Jon W. Mark, Connected Vehicles : Solutions and Challenges, IEEE Internet of Things Journal, Vol. 1, No. 4, August 2014.
- ✓ Zhibo Pang, Qiang Chen, Junzhe Tian, Lirong Zheng and E. Dubrova, Ecosystem analysis in the design of open platform-based in-home healthcare terminals towards the internet-of-things, 2013, 15th International Conference on Advanced Communications Technology (ICACT), PyeongChang, 2013.
- ✓ Wei Qiuping, Zhu Shunbing, Du Chunquan, Study On Key Technologies Of Internet Of Things Perceiving Mine, Procedia Engineering, Volume 26, 2011.
- ✓ Bill Karakostas, A DNS Architecture for the Internet of Things: A Case Study in Transport Logistics, Procedia Computer Science, Volume 19, 2013.
- ✓ Ying-cong Zhang, Jing Yu, A Study on the Fire IOT Development Strategy, Procedia Engineering, Volume 52, 2013.

References

- ✓ J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, Internet of things(IoT): A vision, architectural elements, and future directions, Future Gen. Comput. Syst., vol. 29, no. 7, 2013 .
- ✓ D. Bandyopadhyay and Jaydip Sen, Internet of things: Applications and challenges in technology and standardization, Wireless Personal Communications 58.1 (2011).
- ✓ Industry 4.0, The Industrial Internet of Things, by Alasdair Gilchrist
- ✓ http://pdfserv.maximintegrated.com
- ✓ http://www.comtrol.com
- ✓ http://www.mcrockcapital.com
- ✓ http://web.stanford.edu
- ✓ http://www.accenture.com

Thank You!!

IIT KHARAGPUR

NPTEL ONLINE CERTIFICATION COURSES

Data Handling and Analytics – Part I

Data is Precious

Dr. Sudip Misra
Associate Professor
Department of Computer Science and Engineering
IIT KHARAGPUR
Email: smisra@sit.iitkgp.ernet.in
Website: http://cse.iitkgp.ac.in/~smisra/

What is Data Handling

- ✓ Data handling
  - ✓ Ensures that research data is stored, archived or disposed off in a safe and secure manner during and after the conclusion of a research project
  - ✓ Includes the development of policies and procedures to manage data handled electronically as well as through non-electronic means.

- ✓ In recent days, most data concern –
  - ✓ Big Data
    - ✓ Due to heavy traffic generated by IoT devices
    - ✓ Huge amount of data generated by the deployed sensors

What is Big Data

- ✓ *"Big data technologies describe a new generation of technologies and architectures, designed to economically extract value from very large volumes of a wide variety of data, by enabling the high-velocity capture, discovery, and/or analysis."*

  [Report of International Data Corporation (IDC)]

- ✓ *"Big data shall mean the data of which the data volume, acquisition speed, or data representation limits the capacity of using traditional relational methods to conduct effective analysis or the data which may be effectively processed with important horizontal zoom technologies."*

  [National Institute of Standards and Technology (NIST)]

Types of Data

- ✓ Structured data
  - ✓ Data that can be easily organized.
  - ✓ Usually stored in relational databases.
  - ✓ Structured Query Language (SQL) manages structured data in databases.
  - ✓ It accounts for only 20% of the total available data today in the world.
- ✓ Unstructured data
  - ✓ Information that do not possess any pre-defined model.
  - ✓ Traditional RDBMSs are unable to process unstructured data.
  - ✓ Enhances the ability to provide better insight to huge datasets.
  - ✓ It accounts for 80% of the total data available today in the world.

Characteristics of Big Data

- ✓ Big Data is characterized by 7 Vs –
  - ✓ **V**olume
  - ✓ **V**elocity
  - ✓ **V**ariety
  - ✓ **V**ariability
  - ✓ **V**eracity
  - ✓ **V**isualization
  - ✓ **V**alue

Characteristics of Big Data (Contd.)

- ✓ Volume
  - ✓ Quantity of data that is generated
  - ✓ Sources of data are added continuously
  - ✓ Example of *volume* -
    - ✓ 30TB of images will be generated every night from the Large Synoptic Survey Telescope (LSST)
    - ✓ 72 hours of video are uploaded to YouTube every minute

Characteristics of Big Data (Contd.)

- ✓ Velocity
  - ✓ Refers to the speed of generation of data
  - ✓ Data processing time decreasing day-by-day in order to provide real-time services
  - ✓ Older batch processing technology is unable to handle high velocity of data
  - ✓ Example of *velocity* –
    - ✓ 140 million tweets per day on average (according to a survey conducted in 2011)
    - ✓ New York Stock Exchange captures 1TB of trade information during each trading session

Characteristics of Big Data (Contd.)

- ✓ Variety
  - ✓ Refers to the category to which the data belongs
  - ✓ No restriction over the input data formats
  - ✓ Data mostly unstructured or semi-structured
  - ✓ Example of *variety* –
    - ✓ Pure text, images, audio, video, web, GPS data, sensor data, SMS, documents, PDFs, flash etc.

Characteristics of Big Data (Contd.)

- ✓ Variability
  - ✓ Refers to data whose meaning is constantly changing.
  - ✓ Meaning of the data depends on the context.
  - ✓ Data appear as an indecipherable mass without structure
  - ✓ Example:
    - ✓ Language processing, Hashtags, Geo-spatial data, Multimedia, Sensor events
- ✓ Veracity
  - ✓ Veracity refers to the biases, noise and abnormality in data.
  - ✓ It is important in programs that involve automated decision-making, or feeding the data into an unsupervised machine learning algorithm.
  - ✓ Veracity isn't just about data quality, it's about data understandability.

Characteristics of Big Data (Contd.)

- ✓ Visualization
  - ✓ Presentation of data in a pictorial or graphical format
  - ✓ Enables decision makers to see analytics presented visually
  - ✓ Identify new patterns

- ✓ Value
  - ✓ It means extracting useful business information from scattered data.
  - ✓ Includes a large volume and variety of data
  - ✓ Easy to access and delivers quality analytics that enables informed decisions

Data Handling Technologies

- ✓ Cloud computing
  - ✓ Essential characteristics according to NIST
    - ✓ On-demand self service
    - ✓ Broad network access
    - ✓ Resource pooling
    - ✓ Rapid elasticity
    - ✓ Measured service
  - ✓ Basic service models provided by cloud computing
    - ✓ Infrastructure-as-a-Service (IaaS)
    - ✓ Platform-as-a-Service (PaaS)
    - ✓ Software-as-a-Service (SaaS)

Data Handling Technologies (Contd.)

- ✓ Internet of Things (IoT)
  - ✓ According to Techopedia, IoT *"describes a future where every day physical objects will be connected to the internet and will be able to identify themselves to other devices."*
  - ✓ Sensors embedded into various devices and machines and deployed into fields.
  - ✓ Sensors transmit sensed data to remote servers via Internet.
  - ✓ Continuous data acquisition from mobile equipment, transportation facilities, public facilities, and home appliances

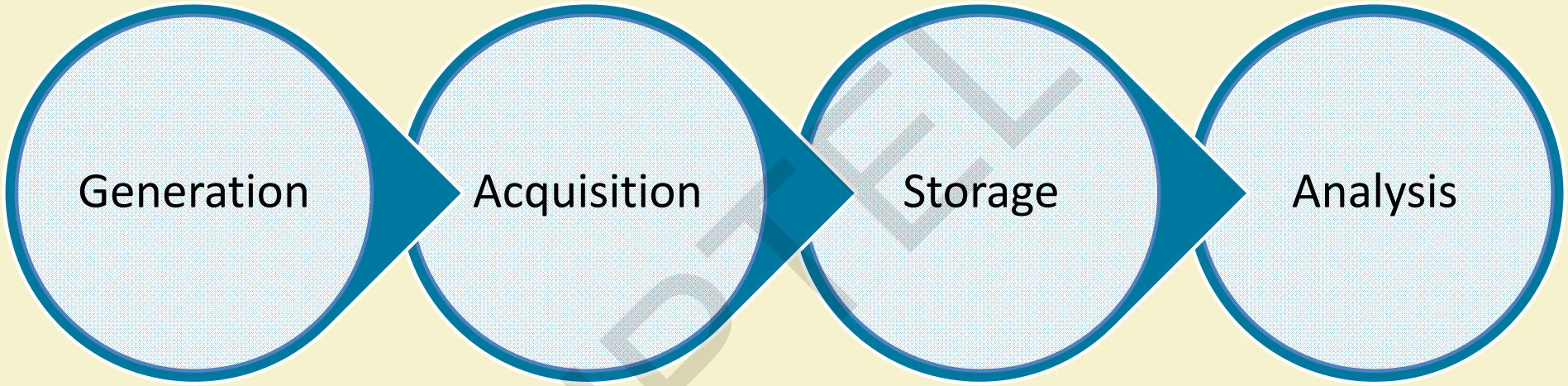
Data Handling Technologies (Contd.)

- ✓ Internet of Things (IoT)
  - ✓ According to Techopedia, IoT *"describes a future where every day physical objects will be connected to the internet and will be able to identify themselves to other devices."*
  - ✓ Sensors embedded into various devices and machines and deployed into fields.
  - ✓ Sensors transmit sensed data to remote servers via Internet.
  - ✓ Continuous data acquisition from mobile equipment, transportation facilities, public facilities, and home appliances

Data Handling Technologies (Contd.)

- ✓ Data handling at data centers
  - ✓ Storing, managing, and organizing data.
  - ✓ Estimates and provides necessary processing capacity.
  - ✓ Provides sufficient network infrastructure.
  - ✓ Effectively manages energy consumption.
  - ✓ Replicates data to keep backup.
  - ✓ Develop business oriented strategic solutions from big data.
  - ✓ Helps business personnel to analyze existing data.
  - ✓ Discovers problems in business operations.

Flow of Data



- ✓ Enterprise data
- ✓ IoT data
- ✓ Bio-medical data
- ✓ Other data

- ✓ Data collection
- ✓ Data transportation
- ✓ Data pre-processing

- ✓ Hadoop
- ✓ MapReduce
- ✓ NoSQL databases

- ✓ Bloom filter
- ✓ Parallel computing
- ✓ Hashing and indexing

Data Sources

- ✓ Enterprise data
  - ✓ Online trading and analysis data.
  - ✓ Production and inventory data.
  - ✓ Sales and other financial data.
- ✓ IoT data
  - ✓ Data from industry, agriculture, traffic, transportation
  - ✓ Medical-care data,
  - ✓ Data from public departments, and families.
- ✓ Bio-medical data
  - ✓ Masses of data generated by gene sequencing.
  - ✓ Data from medical clinics and medical R&Ds.
- ✓ Other fields
  - ✓ Fields such as – computational biology, astronomy, nuclear research etc

Data Acquisition

- ✓ Data collection
  - ✓ Log files or record files that are automatically generated by data sources to record activities for further analysis.
  - ✓ Sensory data such as sound wave, voice, vibration, automobile, chemical, current, weather, pressure, temperature etc.
  - ✓ Complex and variety of data collection through mobile devices. E.g. – geographical location, 2D barcodes, pictures, videos etc.
- ✓ Data transmission
  - ✓ After collecting data, it will be transferred to storage system for further processing and analysis of the data.
  - ✓ Data transmission can be categorized as – Inter-DCN transmission and Intra-DCN transmission.

Data Acquisition (Contd.)

- ✓ Data pre-processing
  - ✓ Collected datasets suffer from noise, redundancy, inconsistency etc., thus, pre-processing of data is necessary.
  - ✓ Pre-processing of relational data mainly follows – integration, cleaning, and redundancy mitigation
  - ✓ Integration is combining data from various sources and provides users with a uniform view of data.
  - ✓ Cleaning is identifying inaccurate, incomplete, or unreasonable data, and then modifying or deleting such data.
  - ✓ Redundancy mitigation is eliminating data repetition through detection, filtering and compression of data to avoid unnecessary transmission.

Data Storage

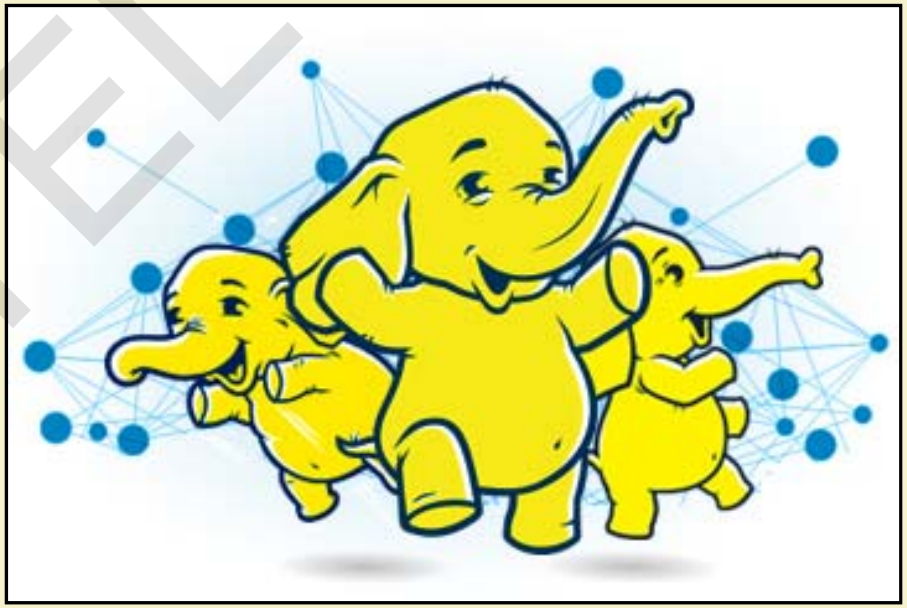
- ✓ File system
  - ✓ Distributed file systems that store massive data and ensure – consistency, availability, and fault tolerance of data.
  - ✓ GFS is a notable example of distributed file system that supports large-scale file system, though it's performance is limited in case of small files
  - ✓ Hadoop Distributed File System (HDFS) and Kosmosfs are other notable file systems, derived from the open source codes of GFS.
- ✓ Databases
  - ✓ Emergence of non-traditional relational databases (NoSQL) in order to deal with the characteristics that big data possess.
  - ✓ Three main NoSQL databases – Key-value databases, column-oriented databases, and document-oriented databases.

Data Handling Using Hadoop

Reliable, scalable, distributed data handling

What is Hadoop

- ✓ Hadoop is a software framework for distributed processing of large datasets across large clusters of computers.
- ✓ Hadoop is open-source implementation for Google ‘s GFS and ***MapReduce.***
- ✓ Apache Hadoop's Map Reduce and Hadoop Distributed File System (HDFS) components originally derived respectively from Google's MapReduce and Google File System (GFS) .



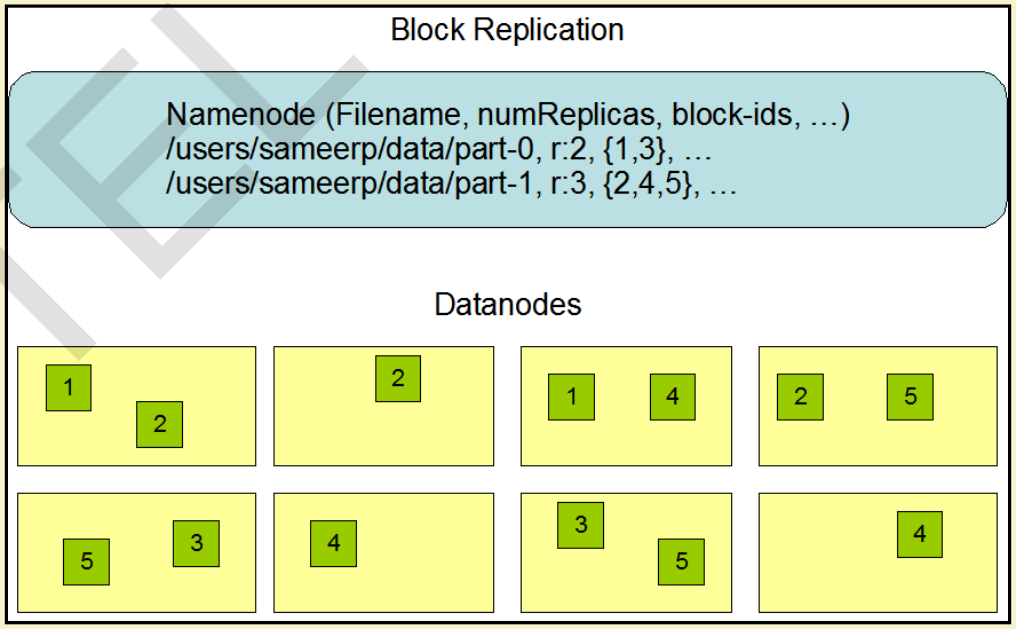
Source: https://www.cloudnloud.com/hadoop-hdfs-operations/

Building Blocks of Hadoop

- ✓ Hadoop Common
  - ✓ A module containing the utilities that support the other Hadoop components
- ✓ Hadoop Distributed File System (HDFS)
  - ✓ Provides reliable data storage and access across the nodes
- ✓ MapReduce
  - ✓ Framework for applications that process large amount of datasets in parallel.
- ✓ Yet Another Resource Negotiator (YARN)
  - ✓ Next-generation MapReduce, which assigns CPU, memory and storage to applications running on a Hadoop cluster.

Hadoop Distributed File System (HDFS)

- ✓ Centralized node
  - ✓ Namenode
  - ✓ Maintains metadata info about files
- ✓ Distributed node
  - ✓ Datanode
  - ✓ Store the actual data
  - ✓ Files are divided into blocks
  - ✓ Each block is replicated



Source: http://hadoop.apache.org/docs/r1.2.1/hdfs_design.html

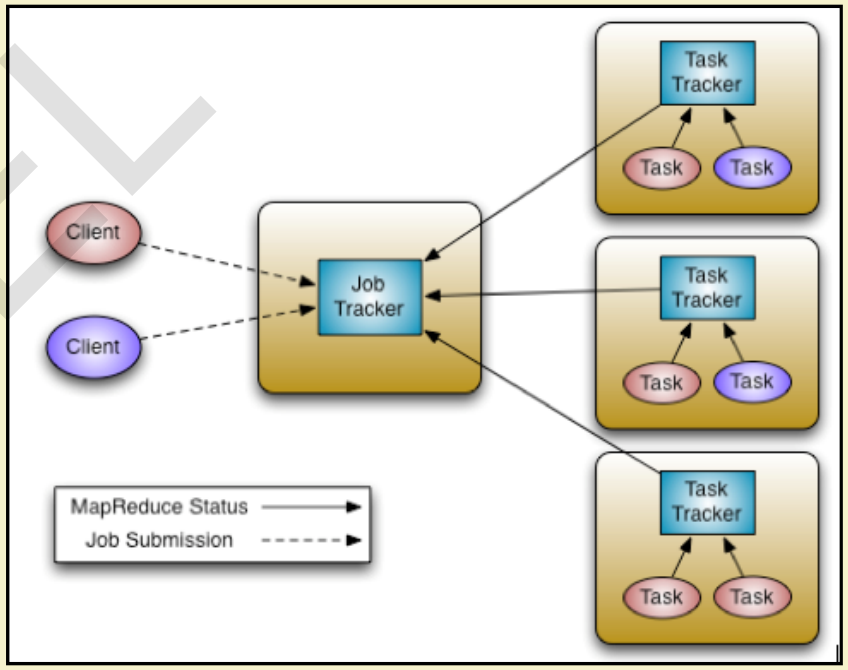
Name and Data Nodes

- ✓ Namenode
  - ✓ Stores filesystem metadata.
  - ✓ Maintains two in-memory tables, to map the datanodes to the blocks, and vice versa

- ✓ Datanode
  - ✓ Stores actual data
  - ✓ Data nodes can talk to each other to rebalance and replicate data
  - ✓ Data nodes update the namenode with the block information periodically
  - ✓ Before updating datanodes verify the checksums.

Job and Task Trackers

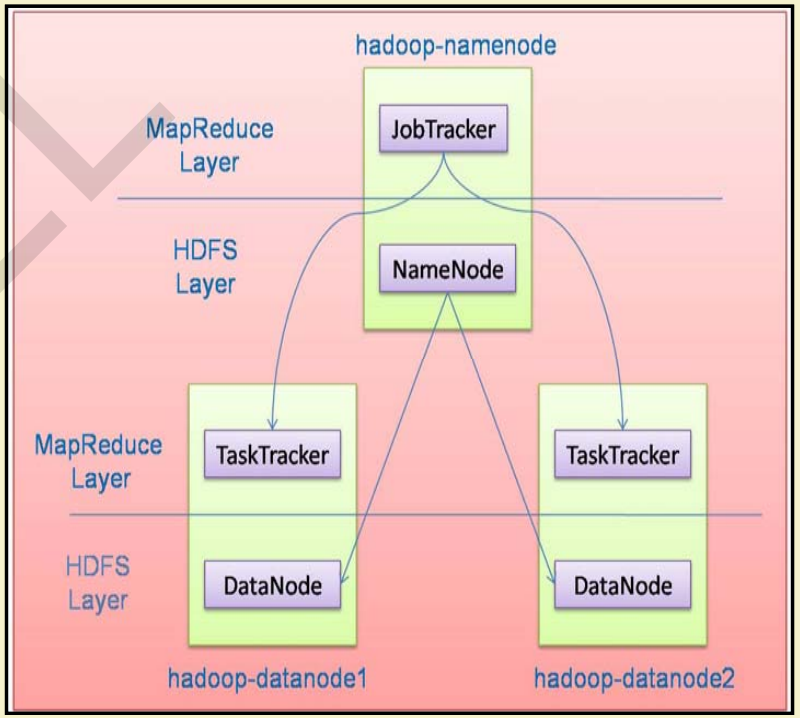
- ✓ Job Tracker –
  - ✓ Runs with the Namenode
  - ✓ Receives the user's job
  - ✓ Decides on how many tasks will run (number of mappers)
  - ✓ Decides on where to run each mapper (concept of locality)
- ✓ Task Tracker –
  - ✓ Runs on each datanode
  - ✓ Receives the task from Job Tracker
  - ✓ Always in communication with the Job Tracker reporting progress



Source: http://developeriq.in/articles/2015/aug/11/an-introduction-to-apache-hadoop-for-big-data/

Hadoop Master/Slave Architecture

- ✓ *Master-slave shared-nothing* architecture
- ✓ Master
  - ✓ Executes operations like opening, closing, and renaming files and directories.
  - ✓ Determines the mapping of blocks to Datanodes.
- ✓ Slave
  - ✓ Serves read and write requests from the file system's clients.
  - ✓ Performs block creation, deletion, and replication as instructed by the Namenode.



Source: http://ankitasblogger.blogspot.in/2011/01/hadoop-cluster-setup.html

References

- ✓ R. Ahmed and G. Karypis, “Algorithms for Mining the Evolution of Conserved Relational States in Dynamic Networks,” Knowledge and Information Systems, vol. 33, no. 3, pp. 603-630, Dec. 2012.
- ✓ M.H. Alam, J.W. Ha, and S.K. Lee, “Novel Approaches to Crawling Important Pages Early,” Knowledge and Information Systems, vol. 33, no. 3, pp 707-734, Dec. 2012.
- ✓ S. Aral and D. Walker, “Identifying Influential and Susceptible Members of Social Networks,” Science, vol. 337, pp. 337-341, 2012.
- ✓ A. Machanavajjhala and J.P. Reiter, “Big Privacy: Protecting Confidentiality in Big Data,” ACM Crossroads, vol. 19, no. 1, pp. 20-23, 2012.
- ✓ S. Banerjee and N. Agarwal, “Analyzing Collective Behavior from Blogs Using Swarm Intelligence,” Knowledge and Information Systems, vol. 33, no. 3, pp. 523-547, Dec. 2012.
- ✓ E. Birney, “The Making of ENCODE: Lessons for Big-Data Projects,” Nature, vol. 489, pp. 49-51, 2012.
- ✓ S. Borgatti, A. Mehra, D. Brass, and G. Labianca, “Network Analysis in the Social Sciences,” Science, vol. 323, pp. 892-895, 2009.
- ✓ J. Bughin, M. Chui, and J. Manyika, Clouds, Big Data, and Smart Assets: Ten Tech-Enabled Business Trends to Watch. McKinSey Quarterly, 2010.
- ✓ D. Centola, “The Spread of Behavior in an Online Social Network Experiment,” Science, vol. 329, pp. 1194-1197, 2010.
- ✓ http://hadoop.apache.org/

Thank You!!