May 4th, 10:00 AM - 11:30 AM

Factors Affecting Big Data Technology Adoption

Nayem Rahman

Portland State University, nayem.rahman@yahoo.com

Let us know how access to this document benefits you.

Follow this and additional works at: http://pdxscholar.library.pdx.edu/studentsymposium

Part of the Computer and Systems Architecture Commons, and the Data Storage Systems Commons

Factors Affecting Big Data Technology Adoption

May 4, 2016

Nayem Rahman

Department of Engineering and Technology Management

Overview

- Introduction
- Big Data Characteristics
- Literature Review and Research Question
- Big Data Adoption Factors
- Research Model
- Hypotheses Development
- Research Method
- References

Introduction

- Traditional data management software has been in the market for more than four decades since Codd's introduction of relational data model theory (Codd, 1970)

- In the early 21st century a new kind of data called 'Big Data' have come into picture

- Big data are mostly unstructured, generated in large volumes, and many cases data need to be captured in near real-time

- To handle big data a completely new set of tools and technologies have emerged

- A non-profit organization, Apache Foundation has provided a handful of open source big data tools and technologies

- The inventors, contributors and early adopters of big data tools and technologies include leading high tech companies
 - Google, Yahoo, Facebook, Microsoft, Amazon, Intel and IBM

Big Data Industry Outlook

- IDC (2015) forecasts big data market is growing over 23.01% per year during 2014-2019 period with annual spending reaching 48.6 billion in 2019

- The IDC (2015) projected annual big data market growth in three areas consisting of infrastructure at 21.7%; software at 26.2%; and services at 22.7%

- Gartner (2015) research updates on big data suggested that "more than three-quarters of companies are investing or planning to invest in big data in the next two years"

- Forrester Research firm (Gualtieri and Curran, 2015) suggested that big data has provided new opportunities to perform predictive analytics

Big Data Characteristics [1]

Characteristic	Description	Influencer
Volume	Grows from a few terabytes to hundreds of terabytes to petabytes of data that need to be captured, processed, stored and analyzed.	Data volume keeps growing faster in source.
Velocity	The data flows in today's digital era are being produced real-time and around the clock. Large volume of data needs to be captured real-time, stored, processed, and displayed faster for right time business intelligence and decsion making.	Data flows in real-time and large volume. Improved computing, processing, BI & Visualization technologies.
Variety	Originates from a variety of data sources with unstructured, semistructured, and structured data. More than 90% data are unstructured.	RFID, Sensors, social networks, digital pictures, video, transaction records, and communication surveillance.
Veracity	In most cases data is unstructured and hence data consistency is an issue. Data and findings extracted from subjective comments and opinions are difficult to predict.	Data-based decisions require traceability and justification.
Value	Provides greater and new insights to generate business value.	Corporate business value; new and untapped business opportunities.

Hadoop Components & Ecosystem [3]

6
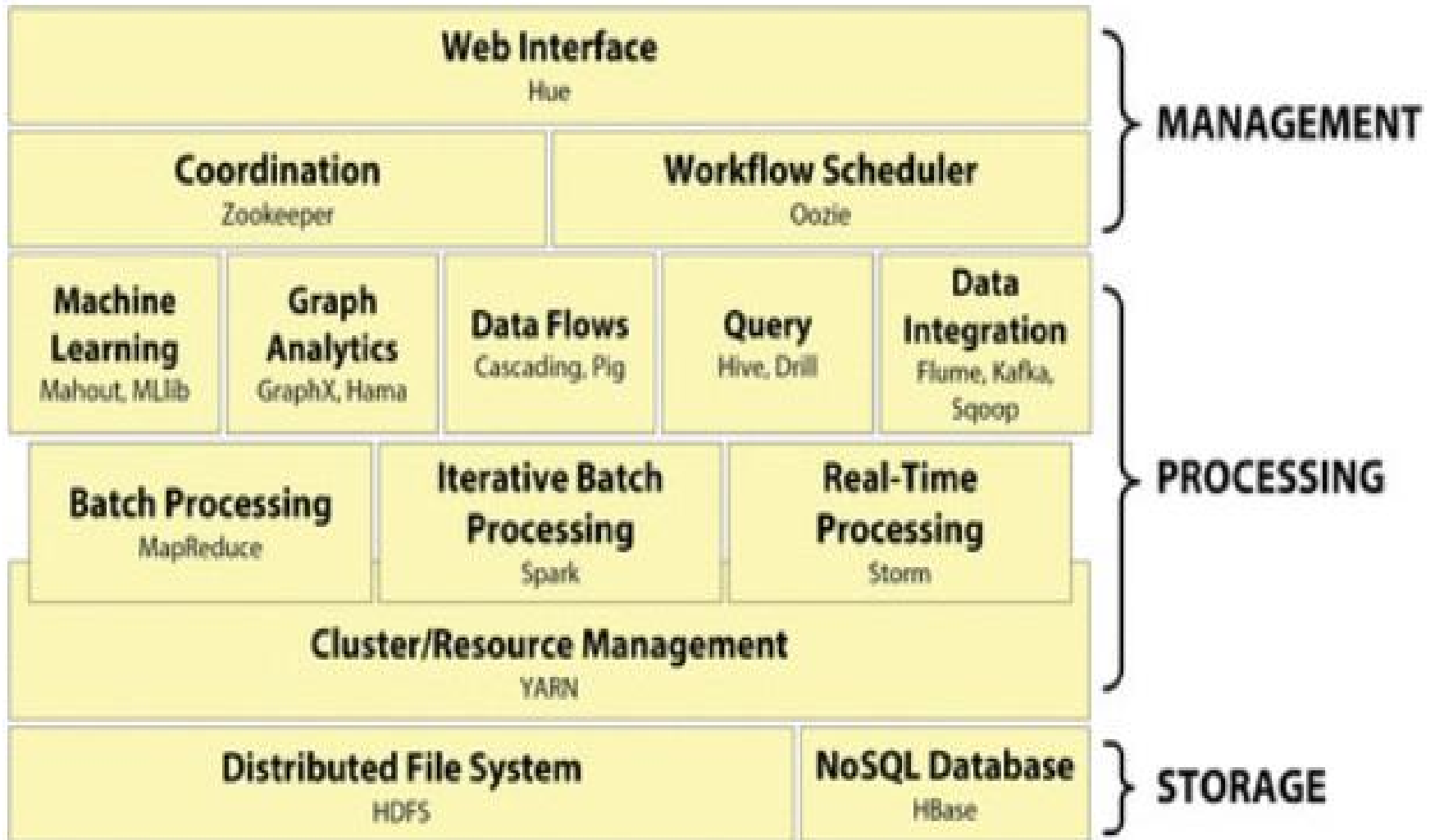
Big Data Literature Review

- 200+ Journal Articles; 50+ Conference Papers
- Most of the research work tried to address challenges relating big data
- Literature reported that a host of big data tools and technologies have emerged to deal with five characteristics of big data – volume, velocity, veracity, variety, and value
- These tools and technologies have emerged to handle data generation, data acquisition, data storage, and data analytics

Big Data Research Gaps

- Exiting literature highlighted results of case studies, data mining techniques, predictive analytics, statistical modeling, and experiments

- But these literature did not provide insights on the overall usage rate of big data tools and technologies

- **One journal paper reported research results (RBV) relating to acquisition intention of big data analytics but, that research concentrated on data quality management in big data (Kwon et al., 2014)**

- It is important to conduct a holistic study to understand acceptance of big data technology from practitioners and researchers' standpoint

Big Data: Research Gaps & Questions

Research Gaps	Research Goals	Research Questions
Since inception (mid last decade) big data researchers and computer scientists focused on technical aspects (technical algorithms, machine learning, etc) and systems developments (Kwon et al., 2015). It is about time to extend the research on big data in terms of its acceptance by industry users. There is a lack of an indepth analysis of the factors that influence the adoption of big data (Hadoop) technology and ecosystem.	Analyze the factors that affect big data technology adoption. A good number of big data dimensions, attributes, and characteris have been identified based on literature review.	What are the key factors associated with industry users behavioral intention to adopt and use big data tools and technologies?
Significant amount of research in terms of individual case studies and use cases have been conducted to experiment different aspects of big data tools and technolgoies. These research findings provided positive results in terms of capability and performance.	A holistic study needs to be conducted on Hadoop and its Ecosystem to understand overall impression of big data technologies from industry users.	How could users' experience of big data tools and technologies be improved?

Technology Acceptance Model (TAM)

- Technology adoption is considered to be the determinants of the success or failure of a product or technology

- Technology Acceptance Model (TAM)/ Theory has been developed by Fred D. Davis (1986) as part of his doctoral dissertation at MIT Sloan School of Management

- TAM comprises two beliefs, the perceived utilities and the perceived ease of application, which determine attitudes to adopt new technologies. The attitude toward adoption will decide about the adopter's positive or negative behavior in the future concerning new technology

- TAM suggests that when users are presented with a new technology, a number of factors determine their decision about how and when they will use it

Factors Influencing Big Data Adoption

- #1. Relative Advantage (analogous to Perceived Usefulness)
  - An innovation is considered to have relative advantage if it is technically superior in terms of cost and functionality than the technology it supersedes (Fichman and Kemerer, 1993)
  - Big Data technologies are open-source, scalable, cheaper to store and process complex and large volume of data
- #2. Compatibility
  - An innovation needs to be compatible "with existing values, skills, and work practices of potential adopters" (Fichman and Kemerer, 1993)
  - Existing data management tools, technologies, and skill set are good for 'normal data', that is, transactional data as opposed to big data

Factors Influencing Big Data Adoption (2)

- #3. Ease of Use (Reverse: Complexity)
  - Big Data concept takes a position of storing data in HDFS in raw format (Liu et al., 2013)
  - Reduced data movement once data lands in HDFS from heterogeneous sources
  - All analytical, machine learning, and reporting tools run against HDFS

- #4. Flexibility
  - Big data tools and technologies provides greater flexibility to consolidate data from various sources into one singe place (HDFS)
  - Consolidated data in to a single platform provides improved data mining and business intelligence capabilities

Factors Influencing Big Data Adoption (3)

- #5. Real-Time Decision Making Abilities
  - One of the key capabilities of big data technologies is analyzing data in real-time, or close to real time (Goes, 2014; Posey, 2013)
  - Big data captures data real-time with streaming technologies, processes and performs predictive analytics with machine leaning tools

- #6. Security and Privacy Considerations
  - Data privacy is one of the concerns with big data
  - Richards and King (2014) stated that big data has issues in terms of privacy, confidentiality, and identity
 - Need to see how these factors impact big data technology adoption

Factors Influencing Big Data Adoption (4)

- #7. Scalability and Latency
    - Scalability is one of the core competency features for data warehousing success (Sen & Jacob, 1998; Sen & Sinha, 2005)
    - One of the key benefits to Hadoop clusters is scale-out storage system (Posey, 2013; Aye & Thein, 2015)
    - Big data technologies provides scalability in terms of storage, data processing, and scalable mining algorithms (MLlib)

- #8. Professional Vendor Support
    - Professional vendor support is very important because most of the big data tools and technologies are open source
    - Organizations might want to make sure vendor support is available to help them get their big data platforms running even when new versions of tools and technologies arrive

Factors Influencing Big Data Adoption (5)

- #9. Organizational size
  - Organizational size in terms workforce in IT and business might have an impact on new technology acceptance
  - Given big data has a handful of tools and technologies learning those tools and maintaining the tools and data requires workforce and other resources
  - A company needs to have resources, technical expertise and willingness to maintain a new technology
  - Big data consists of a good number of tools and technologies

- #10. Data Storage and Processing
  - Hadoop is highly scalable in terms of storage and data processing
  - It would be interesting to see if Hadoop's storage capacity and data processing have influence in big data acceptance

Research Model
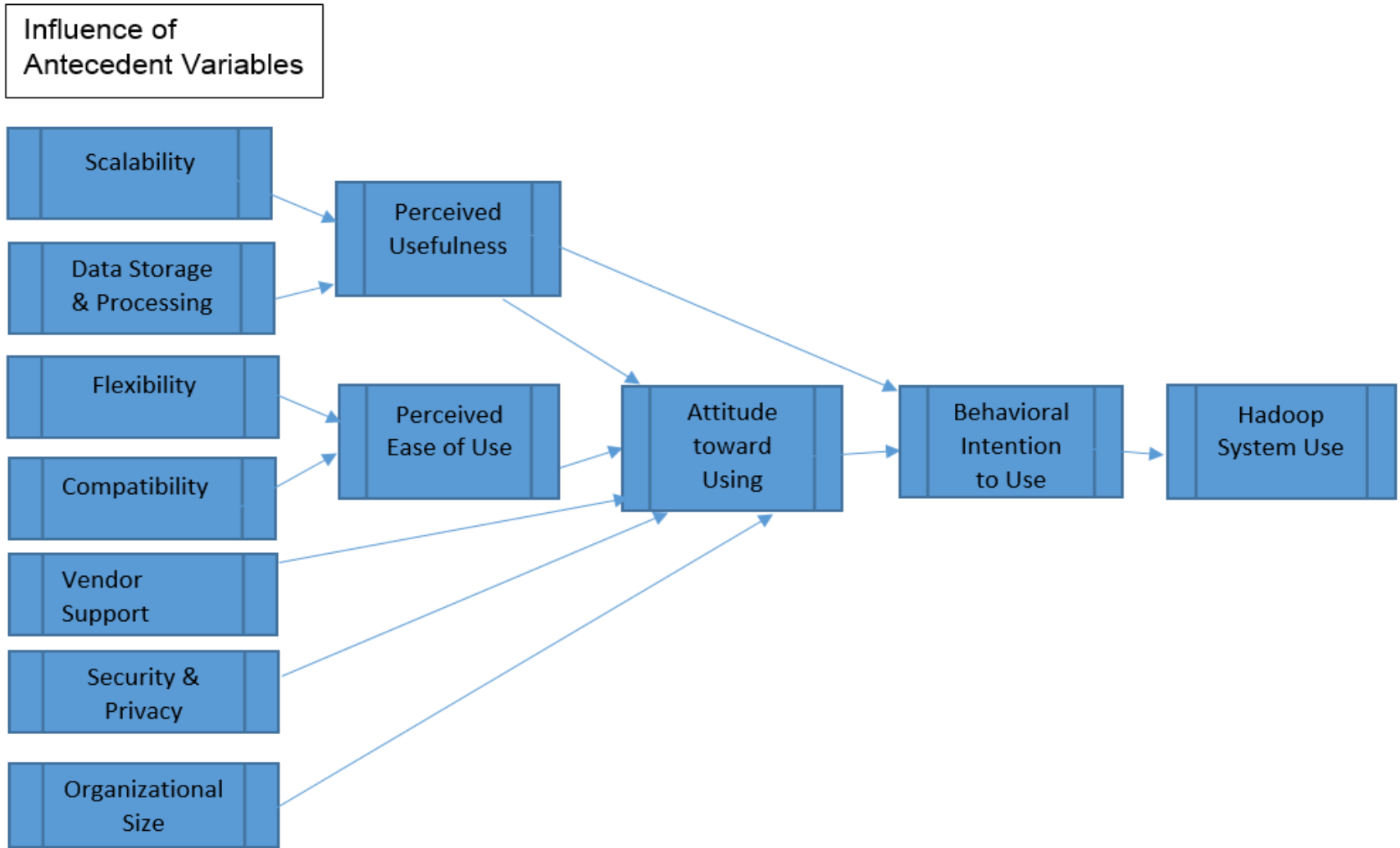
Validating Model

- Evaluation of validity of variables is judgmental and subjective (Straub et al., (2004)

- Content validation or revalidation to be done by consulting with a few industry experts who have implemented Hadoop

- Likert scale will be used to measure each independent variable in the research model

Hypotheses Development

- All required hypothesis will be developed based on observed and latent variables finalized by expert panel
  - H1 (Scalability): Scalability in terms of Hadoop scale-out-storage system will have a positive effect on perceived usefulness
  - H2 (Data Storage and Processing): Real-time streaming, processing, storage, and mining will have a positive effect on perceived usefulness
  - H3 (Flexibility): Hadoop's flexibility to consolidate data from various sources to single place (HDFS) will have a positive effect on perceived ease of use
  - H4 (Compatibility): New set of big data tools' compatibility with existing data management tools, technologies, and skill set has a positive effect on perceived ease of use

Hypotheses Development (Cont'd)

- Hypotheses…
  - H5 (Support): Availability of professional vendor support has a positive effect on attitude toward using Hadoop
  - H6 (Security and Privacy): Privacy and confidentiality of big data have positive effect on attitude toward using Hadoop
  - H7 (Organizational Size): Need for workforce, technical expertise and willingness to handle a big data tools and technologies are positively related toward using Hadoop

Validating Survey Instruments

- Questionnaire will be developed based on previous research work in related topics
- New questions will be developed to make the questionnaire suitable to get meaningful results
- By using expert panels and focused group questionnaire will be finalized. Need to make sure questions are appropriate for the respondents
- Pilot test the survey questions. Colleagues to be asked to complete the survey and provide feedback on the survey questionnaire
- Later a focus group consisting of about 30 to 40 experts will be used to revalidate the modified survey questionnaire

Data Collection

- The survey development will be based on previous research questions and new questions as needed

- Survey Respondents' contact address from Hadoop Vendors

- Planning on conducting email survey to collect data

Validities and Data Analysis

- The Research model will be tested using Structural Equation Model (SEM)
  - AMOS Software
- Structural equation modeling will be used to apply multivariate analysis combining regression analysis to examine relationships and factor analysis
- The reliability and validity of the measurement instrument will be conducted using reliability and factor analysis

References

- [1] Rahman, N., & Aldhaban, F. (2015). "Assessing the Effectiveness of Big Data Initiatives." *Proceedings of the IEEE Portland International Center for Management of Engineering and Technology (PICMET 2015) Conference*, Portland, Oregon, USA. August 2 - 6, 2015, pp. 478-484.

- [2] Rozados, I.V., & Tjahjono, B. (2014). "Big Data Analytics in Supply Chain Management: Trends and Related Research." *Proceedings of the 6th International Conference on Operations and Supply Chain Management*, Bali, 2014.

- [3] Landset, S., Khoshgoftaar, T.M., Richter, A.N., & Hasanin, T. (2014). "A Survey of Open Source Tools for Machine Learning with Big Data in the Hadoop Ecosystem." *Journal of Big Data*, 2(24), 1-36.

- [4] Rahman, Nayem; and Rutz, Dale (2015). "Building Data Warehouses Using Automation." *International Journal of Intelligent Information Technologies (IJIIT)*, 11(2), 1-22.

23

References (2)

- [5] Kwon, O., Lee, N., & Shin, B (2014): "Data quality management, data usage experience and acquisition intention of big data analytics", *International Journal of Information Management*, 34, 387-394.

- [6] Codd, E.F. (1970). A relational model of data for large shared data banks. Communications of the ACM, 13(6), 377-387

- [7] IDC (2015). New IDC Forecast Sees Worldwide Big Data Technology and Services Market Growing to $48.6 Billion in 2019, Driven by Wide Adoption Across Industries. International Data Corporation (IDC). Retrieved on 02/28/2015 from: http://www.idc.com/getdoc.jsp?containerId=prUS40560115

- [8] Gartner (2015). Gartner Survey Shows More Than 75 Percent of Companies Are Investing or Planning to Invest in Big Data in the Next Two Years. Gartner, Inc. September 16, 2015. Retrieved on 02/28/2016 from: http://www.gartner.com/newsroom/id/3130817

References (3)

- [9] Fichman, R.G., & Kemerer, C.E. (1993). Adoption of software engineering process innovations: The case of object orientation. Sloan Management Review. 34(2), 7-22.

- [10] Liu, J., Byna, S. and Chen, Y. (2013): "Segmented Analysis for Reducing Data Movement", In Proceedings of the 2013 IEEE International Conference on Big Data (IEEE BigData 2013), October 6-9, 2013, Santa Clara, CA, USA, pp. 344-349.

- [11] Hartmann, P.M., Zaki, M., Feldmann, N., and Neely, A. (2014): "Big Data for Big Business? A Taxonomy of Data-driven Business Models used by Start-up Firms", University of Cambridge, pp. 1-29.

- [12] Russom, P. (2013): "Managing Big Data", TDWI Best Practices Report, TDWI Research, pp. 1-40, 2013

References (4)

- [13] Goes, P.B. (2014). Editor's Comments - Big Data and IS Research. MIS Quarterly, 38(3).

- [14] Richards, N.M., and King, J.H. (2014): "Big Data Ethics", Wake Forest Law Review, Vol. 49, pp. 393-432.

- [15] Aye, K.N., & Thein, T. (2015). A platform for big data analytics on distributed scale-out storage system. International Journal of Big Data Intelligence, 2(2), 127-141.

- [16] Lamba, H.S., & Dubey, S.K. (2015). Analysis of requirements for Big Data Adoption to maximize IT Business Value. In Proceedings of the 2015 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO). September 2-4, 2015. Noida, India

- [17] Davis, F.D. (1986). A Technology Acceptance Model for Empirically Testing New End-User Information Systems: Theory and Results, Doctoral Dissertation, MIT Sloan School of Management, Cambridge, MA, USA.

Backup

Sources of Big Data in SCM [2]
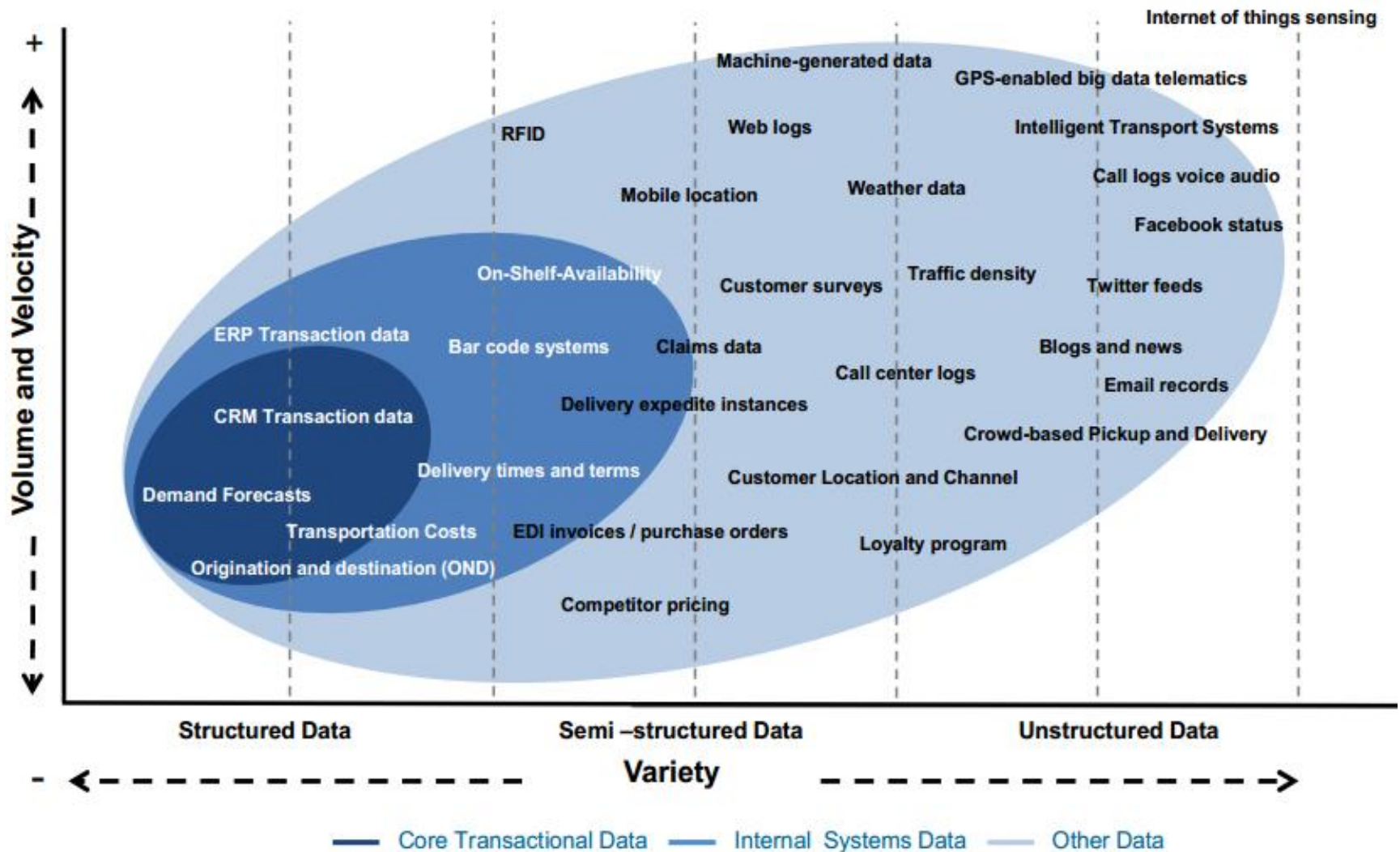
Hadoop: Open Source & Vendors

- Java-based Open source platform
    - Processes large data over thousands of distributed nodes

- Hadoop's base components
    - MapReduce/ Spark (distributed programming framework)
    - HDFS (Hadoop Distributed File System)

- Hadoop clusters has linear scalability in terms of
    - Performance and storage capacity as node count increases

- Apache Hadoop is based on Intel® Xeon® processors

- Vendor-specific distributions based on Hadoop
    - Cloudera
    - Hortonworks
    - MapR