SONA COLLEGE OF TECHNOLOGY
An endearing shrine for excellence in education
| An Autonomous Institution |

Department of Electronics and Communication Engineering

Embedded and IoT

2016-17

Junction Main Road
Suramangalam (PO)
Salem-636 005. TN. India
www.sonatech.ac.in

EDITOR'S BOARD

Editorial Head

Dr.R.S.Sabeenian,
Professor &Head, Dept of ECE, Head R&D Sona SIPRO.

Editorial members:

1. Prof.S.Deepa,
   Associate Professor

1. Mr.A.Ayub Khan,
   Assistant Professor

Magazine Coordinator

M.Senthil Vadivu,
Assist.Professor

Student Members:

1. Gokul Ravi Kumar

2. Krishna Kumar

PREFACE

Embedded system is basically the study of how to setup a device that is hardware or software or both that is embedded in a larger system and is mostly a real time system. An embedded system usually consists of a microcontroller programmed to do a specific job.

Internet of things is how these devices communicate with each other directly and indirectly to serve a specific purpose. Directly is when two devices or more talk peer to peer. And decide actions based on what the other device says. Indirect is when all of these devices are connected to a single node and the node receives and transmits signals to the devices and intercommunicate is thus established.

The “Internet” side of IoT is about processing the huge amount of data that can be collected by devices and extract the useful bit of information that can improve the way we use many services and devices today. The applications of embedded systems have increased drastically over the past years. Multi-core technologies are being appreciated and are now in great demand across various industry verticals.

BIG DATA ANALYTICS

Durga Devi – IV Year

Despite the rising potential of big data, a few studies have been conducted to examine it in the supply chain field. This article gives an overview of big data use in this field and underlines its potential role in the supply chain transformation by leading a systematic literature review. The results show that the big data analytics techniques can be categorized into three types: descriptive, predictive, and prescriptive and these in turn influence supply chain processes and creates value. We conclude by highlighting future research directions. The rise of social media, mobile devices, and Internet of Things involves massive in real-time data generation. More than 1200 Exabyte's of new data is brought about each year from a variety of sources. 80% of this data is unstructured and is hard to be stored, processed and analyzed with commonly used tools. The concept of big data that emerged in this context is often characterized by the five Vs: Volume, Velocity, Variety, Veracity, and Value. Fosso Wamba et al.defines big data as "a holistic approach to manage, process and analyze 5 Vs in order to create actionable insights for sustained value delivery, measuring performance and establishing competitive advantages." Analytics allow making sense of big data by transforming into intelligent information. Overall process of extraction from big data consists of two subprocesses: data management and analytics. Data management involves "processes and supporting technologies to acquire and store data and to prepare and retrieve it for analysis" whereas analytics refers to "techniques used to analyze and acquire intelligence from big data". Gartner outlines that only 15% of Fortune 500 companies will be able to exploit big data for creating value and only 8% of them are currently using big data analytics . In the face of volume, velocity, and variety of big data, many organizations seek to capitalize on analytics for gaining a competitive advantage [8]. Big data analytics is a domain which begins to proliferate in the field of supply chain. A supply chain is defined as "a bidirectional flow of information, products and money between the initial suppliers and final customers through different organizations"; supply chain management includes planning, implementing and controlling of this flow. In this context, big data analytics are expected to take supply chain transformation. The supply chain transformation is defined as "the degree to which a firm engages in a series of actions intended to bring significant changes in the supply chain, including changes in the firm's supplier or customer portfolios, supplier/customer development, and the coordination of the flow of goods" .It focuses on "information and collaboration-driven extended supply chain integration … to improve collaboration and the quality of decisions"]. Big data analytics is used as a driver for improving operational performance and supporting decision making for better matching supply and demand .Then, the challenge is to identify analytics techniques or methods used in supply chain for analyzing big data and creating business value. Thereby, the main objective of this research is to give a comprehensive overview of big data use in supply chain and underline its potential role in supply chain transformation. For this to happen, we decided to run a systematic literature review in order to bring answers to the following questions: - What are big data analytics techniques which are used in supply chain field? - What is the impact of big data analytics on the supply chain management? - How can big data analytics create business value in supply chain? The rest of this paper is structured as follow.It describes the research methodology .It presents the results and discusses the key findings. Finally, it provides the conclusion including future research directions.

Methodology

The aim of this study is to explore the big data phenomenon in the supply chain field drawing on a literature review. More generally, there are mainly four types of literature reviews: narrative review, systematic review, meta-analysis and meta-synthesis. For this research, we used a systematic literature review (SLR). It is a rigorous approach which relies on the ability of digital research libraries for discovering, classifying and synthesizing existed works in research domain.

Research identification

The idea here is to examine and evaluate research on big data and supply chain. For that, we investigated the above research questions.

Search strategy

Our search strategy consists first in deriving major terms related to the research questions, and then identifying alternative spelling and synonyms for these terms by leading a pilot test. Later, we use the Boolean operators (OR; AND) for connecting the founded terms. This resulted in the following used strings for automated search: ("Big data" OR "business analytics") and ("supply chain" OR "supply chain management" OR "supply network" OR "supply chain network"). The search of articles was conducted regardless of time limitation of publications by using “Science direct” database. The search of articles has begun on May 7th, 2015. The articles identified from these databases were downloaded into a references management tool called EndNote.

Study selection

In this step, we define selection criteria to determine which studies are to be included or excluded. Indeed, to be included, the paper has to be written in English, published in a scientific journal and approaches big data and supply chain or their synonyms. Therefore, master and doctoral theses, proceedings or conference articles, working papers and textbooks were excluded for this review. It is the case also of articles which are not accessible. The final list of included publications was 320. Two scholars carried out the selected study process independently. Each reviewer performed the screening of the results based on title and abstract for each publication according to the inclusion and exclusion criteria. Then, a comparison of screening results is realized. In case of difference, verification is jointly made to reach a consensus. At the end of this process, 293 articles were excluded and 27 articles were kept for the quality assessment step.

Quality assessment

In this step, the quality criteria are defined to evaluate the rigor and credibility of the selected articles. The evaluation requires the complete review of the paper. Based on the works of some researchers, we defined the following quality stated criteria as questions: Is there an adequate description of the context in which the research was carried out? Is there a clear statement of research aims? Does the paper describe an explicit research question? Is the research design appropriate to address the research aims? Is the literature review adequate? Is the collected data in a way of addressed research issue? Is the data analysis sufficiently rigorous? Is there a clear statement of findings? Is the study valuable for research or practice? Does the paper discuss limitations or validity?

Each question has four possible options:

(0) issue is not mentioned at all,

(1) little mentioned,

(2) adequately addressed and

(3) completely addressed .

Hence, we used a four points Likert scale for collecting answers. The articles with the average quality score lower than 1 was removed. At the end of this process 15 articles were qualified to be analyzed for the data extraction step. Figure 1 presents the literature search, selection and assessment process.

Data extraction

In this step, we extracted data from the qualified articles. The types of data are: publications details (title, authors, journal, year of publication, country), research methodology, industry, and number of items related with big data in supply chain era.

Data synthesis and analysis

At the end, some results came out of the extracted data. The data synthesis includes a descriptive analysis to provide a background about the included articles and an analysis of their findings in order to underline the future directions of research.

Future research

A holistic big data approach is proposed to define and excavate the trajectory from massive RFID-enabled shop-floor logistics data for supporting production logistics decision-makings.

- This study develops and tests an analytic infrastructure framework based on the deduction graph technique in big data environment. This approach could be used by firms to gain competitive advantage by enhancing their supply chain innovation capabilities.
- The implications of consumption data from Internet Connected Objects on strategic choices of product variety are studied. An analytical model of producer's choice between "tailoring strategy" and "platform strategy is proposed. This study shows that the "platform strategy" is more profitable when contextual variety is higher.
- An enterprise networks integration architecture in big data environment with relative enabling technologies is developed for solving problems of product lifecycle management and omni-channel marketing management.
- A framework for analytics applications in SCM from business and IT perspectives is proposed. These applications were structured along three functional domains (operations management, sales management, and integrated business management) and four use cases ("Monitor and navigate", "Sense and respond", "Predict and act", and "Plan and optimize").
- A framework for implementing big data analytics projects is proposed and explored at a manufacturing company, Ramco Cements Limited, in India. The case study provides a roadmap in conceptualizing, planning and successfully implementing such a project.
- A framework for analyzing supply chain tweets is proposed. This study presents insights into the potential role of twitter analytics for supply chain practice and research.
- This study describes "the application of advanced analytics techniques to supply chain management. The applications are categorized in terms of descriptive, predictive, and

prescriptive analytics and along the supply chain operations reference (SCOR) model domains plan, source, make, deliver, and return."

- This study proposes and evaluates a process management method for RFID data mining. This method enables to perform real time process control efficiently.
- This study describes the differences between forecasting existing products and new product across data, analytics dimensions.
- This study proposes methods for monitoring and controlling data quality in big data environment and shows its importance in supply chain practice and research.
- This study explores big data analysis in supply chain management context and proposes a tool of simulation to analyze business processes and a set of key performance indicators useful for business automation.
- This study shows that the use of advanced analytics improve a firms' operational performance through SCM initiatives such as TQM, JIT, and statistical process control.
- This study assesses how the Internet of Things can enhance virtualization of supply chain in floriculture sector. A framework for analysis of virtual supply chains is developed and applied in the Dutch floriculture.
- A simulation study is developed in order to analyze the characteristics and value of RFID in retail store operations.

BIG DATA ANALYTICS TRANSFORMATION

Mani Raja – IV Year

In assessing the grounds on why several organizations are gravitating towards Big Data analytics, concrete understanding of traditional analytics is necessary. Traditional analytical methods include structured data sets that are periodically queried for specific purposes. A common data model used to manage and process commercial applications is the relational model; this Relational Database Management Systems (RDBMS) provides “user-friendly query languages” and provides simplicity that other network or hierarchical models are not able to deliver. Within this system are tables, each with a unique name, where related data is stored in rows and columns. These streams of data are obtained through integrated databases and provide leverage on the intended use of those data sets. They do not provide much advantage for the purpose of creating newer products and/or services as Big Data do – leading to the transformation of Big Data analytics. Frequent usage of mobile devices, Web 2.0, and growth in the Internet of Things are among a few to mention in reasoning behind organizations looking to transform analytical processes. Organizations are attracted to big data analytics as it provides a means of obtaining real time data, suitable for enhancing business operations.

Along with providing parallel & distributed processing architectures in data processing, big data analytics also enables the following services: “after-sales service, searching for missing people, smart traffic control system, customer behavior analytics, and crisis management system”. Research conducted by Kyounghyun Park and his colleagues of the Big Data SW Research Department in SouthKorea proved to develop a platform with the purpose of establishing Big Data as a Service. In facilitating the services mentioned above, traditional methods require various distinctive systems to perform various tasks such as “data collection, data pre-processing, information extraction, and visualization”. On the contrary, this web-based platform provides developers & data scientist an environment to “develop cloud services more efficiently”, view shared data, “supports collaborative environments to users so that they can reuse other’s data and algorithms and concentrate on their own work such as developing algorithms or services”.Unlike traditional methods where databases are accessible by all persons, this platform and similar big data analytic platforms supports restricted access on different datasets.

This big data platform and the like prove their cost efficiency by:

- Centralizing all aspects of storage and processing procedures in big data onto one platform.
- Ensures easy and rapid access to view other data; allowing developers to focus on their own work (algorithms/service).
- Privacy coverage Currently there are over 4.6 billion mobile phone subscribers. In addition, there are over 1 to 2 billion persons accessing the Internet at any given time. Two of the most widely known social platforms.
- Facebook has over 1 billion active users monthly, accumulating over 30 billion bits of shared content.
- Twitter also has mass amounts of data, serving as a platform for over 175 million tweets a day. With approximately 2.5 quintillion bytes of data being created everyday, it is understandable why Business Intelligence (BI) is drifting towards analytical processes that involve extraction of larger data sets, as traditional management systems are unable to fathom these amounts of data. “Data are being constantly collected through specially designed devices to help explore various complex systems”. These streams of data when analyzed properly using big data methods will “help predict the possibility to increase productivity, quality and flexibility”. The power of big data is its ability to bring forth much more intelligent measures of formulating decisions.

Map reduce

The advancements in technology within the last few decades has caused an explosion of data set sizes , though there’s now more to work with the speed at which these volumes of data are growing exceeds the computing resources available. MapReduce, a programming paradigm utilized for processing large data sets in distributed environments” is looked upon as an approach to handle the increased demand for computing resources.

There are two fundamental functions within the paradigm, the Map and the Reduce function. The Map function executes

sorting and filtering, effectively converting a data set into another & the Reduce function takes the output from the Map function as an input, then completes grouping and aggregation operations to combine those data sets into smaller sets of tuples.

The Map function above effectively takes a document as an input, split the words contained in a document (i.e. a text file), and creates a (Key, Value) pair for each word in the document.

In the Reduce function the list of Values (partialCounts) are worked on per each Key (word). “To calculate the occurrence of each word, the Reduce function groups by word and sums the values received in the partialCounts list”.As the final output, a list of words displaying their occurrences in the document is displayed. The MapReduce paradigm’s main perk is its scalability, it “allows for highly parallelized and distributed execution over a large number of nodes”. An open source implementation of MapReduce is Hadoop. Within this implementation, tasks in the Map or Reduce function are separated into various jobs. These jobs are assigned to nodes in the network, and assigned to other nodes if an initial node fails its jobs. Hadoop permits the distribution of big data processes across machines using simple programming models not suitable for big data.

Though MapReduce provides significant progress in datastorage, a few uncertainties persist within this paradigm:

1) Absence of a standardized SQL query language:

(i) Current solution, providing SQL on-top of Map Reduce Apache Hive - stipulates an SQL-like language on top of Hadoop.

(ii) Deficiency in data management features like advanced indexing and a complex optimizer.

(iii) NoSQL Solutions - MongoDB & Cassandra enable queries similar to SQL HBase uses Hive.

2) “Limited optimization of MapReduce jobs”

i. “Integration among MapReduce, distributed file system, RDBMSs and NoSQL stores”.

Data storage

Data storage has always been an area of concern in the knowledge management domain. Shortly after the Information Explosion was conceptualized in the 1930s, focus was driven to understand how to manage this everlasting growth of data and information. Referring to libraries, the first source of data organization and storage, a sign of data storage overload was reported in 1944 when Fremont Rider, a librarian from Wesleyan University, estimated that “American University libraries were doubling in size every 16 years”. With this estimation, it was necessary to change the methods of storing and retrieving data, not only in respect to libraries but all sectors concerned with knowledge management. Before analyzing the importance of effective data storage, a better understanding in the evolution of data storage is briefly documented:

A. Evolution of Data Storage

- Late 1920s – IBM takes successfully redesigns Basile Bouchon's punch card invention, generating 20% of their revenue in the 1950s.
- 1952 – IBM announces first magnetic tape storage unit; standard data storage technology in the 1950s and still in use in the entertainment industry.
- 1956 – IBM invents first hard drive first hard drive capable of holding up to 5MB, pushing to 1GB in 1982, and a few TB currently.
- 1967 – First floppy disk is created, initially storing up to 360KB on a 5.25-inch disk leading to a 3.5-inch disk capable of storing 1.44MB.
- 1982 - Concept of the compact disk (CD) is invented in Japan, CD-ROM is later developed with storage capacity of 650MB to 700MB; Equivalent to 450 floppy disks.
- 2000 – USB flash drives debuts; similar to the floppy disks, data storage capacity improved overtime and continually improves.
- 2010s – "The Cloud is estimated to contribute more than 1 Exabyte of data".

In the early 1950s Fritz-Rudolph Guntsch developed the concept of virtual memory, allowing finite storage to be treated as infinite. Guntsch's concept permitted the processing of "data without the hardware memory constraints that previously forced the problem to be partitione" in other words, focus was on the hardware architecture not the data itself. Derek Price, information scientist known as the father of scientometrics, agreed with Rider's statement on storage overload by observing and expressing that "the vast amount of scientific research was too much for humans to keep abreast of". The evolution of data storage has shown to continually improve techniques of storage, effectively storing 'more on less', now dealing away with most in-house hardware's and focusing on cloud storage. However this method of mass data storage may be problematic if not prepared for accordingly. Data storage is now extremely complex. Big data or 'total data landscape' as 451 Researchers call it reflect complications not only with storage techniques, but also with the transformation of those data-sets to map them into knowledge-space in order to drive some value to a business as well as the costs involved.

As Professor Hai Zhuge of Aston University mentioned, "If a problem is unclear or unable to be represented, problem-solving is not meaningful". This is a fundamental issue with big data analytics. Viewing the Evolution of Data Storage documented earlier it's evident that engineers have always been eager to compress more data into smaller storage units, and now into the cloud. Data capture has escalated from manual inputs in the early years of technology to now attaining data sets from social media platforms, IoT devices, and the like. Availability for the 'general population' to create data through their mobile devices and computers strike as an opportunity for businesses to improve their business models, the storage mediums earlier available in the early 2000s were problematic.

Modern technology now allows approximately 4 terabyte (102) per disk, equating to 25,000 disks for 1 terabyte (1018). Some view this as being problematic in the life science space as they involve large datasets;

Looking to the example below a systematic analysis will be shown as to why this point of view is not accurate.

BIG DATA TECHNOLOGIES

Sunodh – IV Year

Apache flume

Apache Flume is a distributed, reliable, and available system for efficiently collecting, aggregating and moving large amounts of log data from many different sources to a centralized data store. Flume deploys as one or more agents, each contained within its own instance of the Java Virtual Machine (JVM). Agents consist of three pluggable components: sources, sinks, and channels. Flume agents ingest incoming streaming data from one or more sources. Data ingested by a Flume agent is passed to a sink, which is most commonly a distributed file system like Hadoop.

Multiple Flume agents can be connected together for more complex workflows by configuring the source of one agent to be the sink of another. Flume sources listen and consume events. Events can range from newline-terminated strings in stout to HTTP POSTs and RPC calls — it all depends on what sources the agent is configured to use.

Flume agents may have more than one source, but at the minimum they require one. Sources require a name and a type; the type then dictates additional configuration parameters. Channels are the mechanism by which Flume agents transfer events from their sources to their sinks. Events written to the channel by a source are not removed from the channel until a sink removes that event in a transaction. This allows Flume sinks to retry writes in the event of a failure in the external repository (such as HDFS or an outgoing network connection). For example, if the network between a Flume agent and a Hadoop cluster goes down, the channel will keep all events queued until the sink can correctly write to the cluster and close its transactions with the channel.

Sink is an interface implementation that can remove events from a channel and transmit them to the next agent in the flow, or to the event's final destination and also sinks can remove events from the channel in transactions and write them to output. Transactions close when the event is successfully written, ensuring that all events are committed to their final destination.

Apache SQOOP

Apache Sqoop is a CLI tool designed to transfer data between Hadoop and relational databases. Sqoop can import data from an RDBMS such as MySQL or Oracle Database into

HDFS and then export the data back after data has been transformed using MapReduce. Sqoop also has the ability to import data into HBase and Hive. Sqoop connects to an RDBMS through its JDBC connector and relies on the RDBMS to describe the database schema for data to be imported. Both import and export utilize MapReduce, which provides parallel operation as well as fault tolerance.

During import, Sqoop reads the table, row by row, into HDFS. Because import is performed in parallel, the output in HDFS is multiple files.

Apache pig

Apache's Pig is a major project, which is lying on top of Hadoop, and provides higher-level language to use Hadoop's MapReduce library. Pig provides the scripting language to describe operations like the reading, filtering and transforming, joining, and writing data which are exactly the same operations that MapReduce was originally designed for. Instead of expressing these operations in thousands of lines of Java code which uses MapReduce directly, Apache Pig lets the users express them in a language that is not unlike a bash or Perl script.

Pig was initially developed at Yahoo Research around 2006 but moved into the Apache Software Foundation in 2007. Unlike SQL, Pig does not require that the data must have a schema, so it is well suited to process the unstructured data. But, Pig can still leverage the value of a schema if you want to supply one. PigLatin is relationally complete like SQL, which means it is at least as powerful as a relational algebra.

Turing completeness requires conditional constructs, an infinite memory model, and looping constructs.

Big data inconsistencies

With the emergence of the Information Era, computational ability to capture data has increased considerably; sources of these data streams come from all sectors: automobile, medical, IT, and more. Majority of data used in Big Data analytics come in unstructured formats, primarily obtained from the Internet (cloud computing) and social media. According to industry consultant leaders at CapGemini, organizations looking to optimize their big data processes need to refer to tweets, text messages, blogs, and the like to recognize response on specific products or services that may aid in discovering new trends.

Mapping into knowledge-space

After data is captured, processed and analyzed, how are they mapped into other dimensions? Dimensions within the big data environment are methods for observing, classifying, and understanding a space. Knowledge space acquires the structure of a semantic network where its vertices represent knowledge modules and its relations represent relations amongst two knowledge modules.

A multi-dimensional resource space model was proposed to manage knowledge from multiple dimensions. It provides a way to divide big data into a multi-dimensional resource space through multi-dimensional classifications.

To bridge the gap between machines and humans in order to map data into knowledge space appropriately, a cyber-space infrastructure must be formulated. The United States defines this infrastructure as "the converged information technologies including the Internet, hardware and software that support a technical platform where users can use globally distributed digital resources".

This infrastructure provides cross-connections in the space - interlocking different modules within the different dimensions.

Representation of knowledge from big data analytics requires multiple links throughout various spaces (i.e. physical, social, knowledge, etc.) to not only link different symbols but also to differing individuals. Vannevar Bush's establishment of the theoretical Memex machine awakened further study and research into interlinking various spaces through the cyber-space.

Within the cyber-space, continuous tests must be conducted to ensure the capability of modules being derived from other modules, no conflicts between modules, and no partial modules as this could lead to ambiguous results. A number of softwares have been developed to accommodate the issues brought forth by establishing a well-formulated cognitive cyber-infrastructure. Globus, for example, has been developed to accommodate the integration of "higher-level services that enable applications to adapt to heterogeneous and dynamically changing meta-computing environments".

Knowledge-space has strong impacts to scientific conceptions that may deliver tools to be used in identifying strong points. Value of statements can be tested as a result of the knowledge-space.

Intelligent manufacturing is a result of such space, humans mapping the representation of the external world into their mind (i.e. social & economical factors not computable), converging it onto the virtual space, and allowing the different modules to be interlocked, displaying weak-points and strong-points, bringing forth the 4th industrial revolution.

Future developments

The concept of Big Data analytics is continually growing.

Its environment demonstrates great opportunities for organizations within various sectors to compete with a competitive advantage, as shown in the examples mentioned earlier. The future of medical science is changing dramatically due to this concept, scientist are able to access data rapidly on a global scale via the cloud, and these analytics contribute to the development of predictive analytic tools (i.e. facilitating predictive results at primary stages). However as mentioned (Section IV), there are inconsistencies and challenges within Big Data privacy: sufficient encryption algorithms to conceal raw data or analysis, reliability & integrity of Big Data, data storage issues and flaws within the MapReduce paradigm.