Advances in Internet of Things, 2014, 4, 29-35
Published Online July 2014 in SciRes. http://www.scirp.org/journal/ait
http://dx.doi.org/10.4236/ait.2014.43005

Scientific Research

Managing Computing Infrastructure for IoT Data

Sapna Tyagi[1], Ashraf Darwish[2], Mohammad Yahiya Khan[3]

[1]Institute of Management Studies, Ghaziabad, UP, India
[2]Faculty of Science, Helwan University, Cairo, Egypt
[3]College of Science, King Saud University, Riyadh, Saudi Arabia
Email: ashraf.darwish.eg@ieee.org

Received 23 May 2014; revised 23 June 2014; accepted 22 July 2014

Abstract

Digital data have become a torrent engulfing every area of business, science and engineering disciplines, gushing into every economy, every organization and every user of digital technology. In the age of big data, deriving values and insights from big data using rich analytics becomes important for achieving competitiveness, success and leadership in every field. The Internet of Things (IoT) is causing the number and types of products to emit data at an unprecedented rate. Heterogeneity, scale, timeliness, complexity, and privacy problems with large data impede progress at all phases of the pipeline that can create value from data issues. With the push of such massive data, we are entering a new era of computing driven by novel and ground breaking research innovation on elastic parallelism, partitioning and scalability. Designing a scalable system for analysing, processing and mining huge real world datasets has become one of the challenging problems facing both systems researchers and data management researchers. In this paper, we will give an overview of computing infrastructure for IoT data processing, focusing on architectural and major challenges of massive data. We will briefly discuss about emerging computing infrastructure and technologies that are promising for improving massive data management.

Keywords

Big Data, Cloud Computing, Data Analytics, Elastic Scalability, Heterogeneous Computing, GPU, PCM, Massive Data Processing

1. Introduction

Internet technology has become ubiquitous within our society which is infiltrating all aspects of our lives, and it

How to cite this paper: Tyagi, S., Darwish, A. and Khan, M.Y. (2014) Managing Computing Infrastructure for IoT Data. *Advances in Internet of Things*, **4**, 29-35. http://dx.doi.org/10.4236/ait.2014.43005

is better to call it as necessity rather than a convenience. This widespread use of cell phones and other mobile communication devices like laptops, notebooks, dongles, tablet PC, ebook reader like kindle and nook, GPS system, in-vehicle info displays like BMW idrive, has brought with it an increasing acceptance of their use in virtually all social situations and facilitates "always on world". A mobile phone is a constant companion that accompanies a person throughout their daily life and allows them the convenience of easy communication and access to information. The future, *i.e.*, a world dominated by the "Internet of Things" will be an ecosystem where every tangible identity will talk to you and all your objects will be controlled through a touch on your PDA or at the click of a mouse. Everyday objects ranging from electrical appliances to what you wear, what/where you drive, what you read/see, and anything humanly perceptible will be more addressable and controllable through the Internet. The way that people access information and communicate is radically changing, right before our eyes, in many ways that are not yet readily apparent. As the cost to mobilize the devices continues to drop, and wireless networks become faster, ubiquitous and cheaper, it is easy to see a near future where almost everything and everyone are wirelessly online, 24 × 7. In the future where everything is Web-connected, mobile phones will serve as the hub, or the remote control, for all the things around us which is broadly termed as Internet of Things.

Internet of Things (IoT) is an integrated part of future Internet and could be defined as a dynamic global network infrastructure with self-configuring capabilities based on standard and interoperable communication protocols where physical and virtual "things" have identities, physical attributes, and virtual personalities and use intelligent interfaces, and are seamlessly integrated into the information network [1]. And in this new network, where inanimate objects are Internet-enabled, our mobile phone will sit in the center of this Web of things. It will help you orchestrate the interactions of the things around you and provide real-time access to all sorts of information, including the people you meet, the places you go and the content that's available there. Some research estimates that the number of connected objects will reach 50 billion as early as 2020 [2]. The IoT promises humans to live in a smart, highly networked world, which allows for a wide range of interactions with this environment. The phone is the key to authenticating with these connected devices and taking their content with you, wherever you go and use the concept of Object hyperlinking. Object hyperlinking is a neologism that usually refers to extending the Internet to objects and locations in the real world [2]. The current Internet does not extend beyond the electronic world. Object hyperlinking aims to extend the Internet to the real world by attaching object tags with URLs as meta-objects to tangible objects or locations. Most of them rely on some kinds of unique marker integrated in or attached to the object. Some of these markers can be analyzed using different kinds of wireless near field communication (for instance RFID tags [3] or Bluetooth beacons [4]), and others are visual markers and can be analyzed using cameras, for instance standard 1D-barcodes [5] or their modern counterparts, the 2D codes [6]. These object tags can then be read by a wireless mobile device and information about objects and locations retrieved and displayed [7]. Using radio frequency identification (RFID), every real object in the analogue world could have a unique identifying number, like an IP address.

The paper is organized as follows. Section I introduces the Internet of things. Section II explores the data emerging through Internet of Things. This section explores the characterstics and challenges held with such large data sets. Section III describes the related technologies. Section IV presents the architectural framework to manage the large data set. A conclusion is presented in the last section.

2. Data Management in IoT

The physical world is becoming a type of information system. In Internet of Things, sensors, actuators, RFID Tags are embedded in physical objects—from roadways to pacemakers and placed on products moving through supply chains, thus improving inventory management while reducing working capital and logistics costs are linked through wired and wireless networks, often using the same Internet Protocol (IP) that connects the Internet. These networks churn out huge volumes of data that flow to computers for storage and analysis. Inventory goods are monitored using RFID tags, hospital patients are managed using RFID tags, and parking place availability is managed using a range of sensors. Internet-connected cars, sensors on raw food products, sensors on packages of all kinds, data streaming in from the unlikeliest of places: restrooms, kitchens, televisions, personal mobile devices, cars, gasoline pumps, car washes, refrigerators, vending machines, and SCADA systems. Professionals are highly motivated to harness this big data into an asset for the organization. This requires tremendous storage and computing resources linked with advanced software systems that generate a variety of

graphical displays for analyzing data—rise accordingly.

The Internet of Things (IoT) is causing the number and types of products to emit data at an unprecedented rate. Companies use this data to analyze and improve processes, predict trends and failure. The data can also provide insights for product development, customer support, operations, and sales teams who use the information to improve their features, increase revenues, lower costs and more. Our solution is built on a proven framework and tested methodology. It has generated tremendous results for leading manufacturing, high technology, energy and telecommunications companies.

A) Characteristics of Data in "Internet of things"

The volume of information deriving from the tags is substantial generating large data. Velocity suggests that information is being generated at a rate that exceeds those of traditional systems. There are a variety of different types of information available to monitor. Variety is indicative of there being multiple emerging forms of data that are of interest to enterprises [8]. For example, Twitter and other social media have become a source of big data. In mid-2010, Twitter tweets hit 65 million per day and there were 190 million users [2] [9]. The "Internet of Things" can generate large data for a number of reasons. The volume of data attributable to the "Internet of Things" is substantial. As sensors interact with the world, Things such as RFID tags generate volumes and volumes of data. As a result, digital processing becomes a requirement of feasibility. The velocity of data associated with the "Internet of Things", compared with traditional transaction processing, explodes as sensors can continuously capture data. The variety of data associated with the "Internet of Things" also depends on as the types of sensors and the different sources of data expand. Variety deals with the complexity of large data and information and semantic models behind these data. Thus data collected is in the form of structured, unstructured, semi-structured, and a mixed data. Data variety imposes new requirements to data storage and database design which should dynamic adaptation to the data format. Veracity ensures that the data used are trusted, authentic and protected from unauthorised access and modification. The data must be secured during the whole their lifecycle from collection from trusted sources to processing on trusted compute facilities and storage on protected and trusted storage facilities. The veracity of data in the "Internet of Things" may also be improved as the quality of sensor and other data improves over time. For example, use of RFID tags generates much more reliable information than a decade ago. Such high volumes of data, coupled with an increasing velocity of data, along with an increased variety of data generates large amount of raw data which need analytical processing to create its value. Variability or data dynamicity refers to change in data while processing or analyzing.

B) Challenges

Data produced through sensors is increasing at very exponential rate from sensors in IoT. Heterogeneity, scale, timeliness, complexity, and privacy problems with large set of data impede progress at all phases of the pipeline that can create value from data. As data is increasingly becoming more varied, more complex and less structured, it has become imperative to process it quickly. Meeting such demanding requirements poses an enormous challenge for traditional databases. It Need to consolidate e-Infrastructures platforms to ensure research continuity and cross-disciplinary collaboration, deliver/offer persistent services, with adequate governance modelIt required to upgrade the architectures that address these needs. Such enormous data fundamentally requires massively distributed architectures and massively parallel processing to manage and analyze data. The three main categories under which management of this massive data lies are populating this huge IoT data, querying the database and managing the database **Figure 1**. Above this one more challenge which lies above all is com-

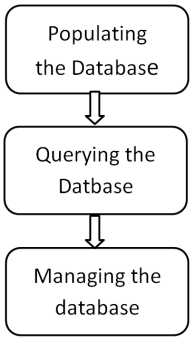

Figure 1. Main components of challenges.

munication of the data. Communication cost is much higher than processing cost. The challenge here is to minimize that communication cost while satisfying the additional storage and data requirement. Bandwidth and latency are the two major network features that will affect the communication between the clients and the data server.

3. Related Technologies to Manage the Massive Data

In the aspect of massive data management, a lot of work has been done. Gonzalez [10] proposed a model called RFID-Cuboids to store massive RFID data. Authors in [11] proposed a one-to-one model and a many-to-many model storing WSN data. For the purpose of managing heterogeneous data from different devices, a scheme of IoT data management based on SOA was proposed in [12]. But the efficiency of the application will be reduced because of the SOA. Besides, the major shortcoming of the above solutions is that they are only for certain data format and they lack systematicness. Recently, there is a little work has focused on systematically dealing with challenges from massive IoT data. Ding [13] proposed a massive IoT data oriented framework to support massive IoT data management. However, the core of Ding's solution is RDBMS (Relational Database Management System), though the *join* operation is avoided as all the data is stored in one table, the concurrency is not well supported because of the lock mechanisms adopted by the RDBMS. Tingli [11] have proposed An Internet of Things storage management architecture name IOTMDB based on NoSQL is proposed to meet the needs of IoT data storage. The IOTMDB not only concerns about how to reasonably and effectively store massive IoT data, but also cares for data sharing and collaboration. Combined with a public service platform for the Internet of Things named RNS and the data abstraction based on ontology, we are able to easily search and locate data and finally realizing data sharing between different IoT applications. The IoT data storage strategies are proposed including a preprocessing mechanism to meet both common and specific needs, a unified data expression form and data distribution strategies. These strategies are beneficial to improve the performance of the cluster and store data effectively.

Apache Hadoop & NoSQL: The dominant Massive Data technologies in use today commercially are Apache's Hadoop and No-SQL databases.

No-SQL databases are typically part of the real-time event detection process deployed to inbound channels (discussed in more detail in the section: "Big Data needs Big-Execution and Agile IM") but can also be seen as an enabling technology behind analytical capabilities such as contextual search applications. These are only made possible because of the flexible nature of the No-SQL model where the dimensionality of a query is emergent from the data in scope, rather than being fixed by the developer in advance. For the Data Scientist and Business Analysts in particular, this more agile approach can often lead to earlier insights into the data that may otherwise be obscured or constrained by a more formal development process.

Hadoop is a software framework for data intensive distributed applications and was developed from a number of academic papers published by Google who were researching in the area of parallel processing.

Hadoop has two major components:

1. The Hadoop File System (HDFS). A highly scalable and portable file system for storing the data;
2. Map-Reduce. A programming model for processing the data in parallel.

The Map-Reduce framework allows analysis to be brought to the data in a distributed and highly scalable fashion, and the Hadoop ecosystem includes a wide range of tools to simplify analysis or manage data more generally. These tools create Map-Reduce programs which are then executed on top of HDFS. Analysis tools of note include:

1. Apache Hive which provides a simple SQL-like interface;
2. Apache Pig which is a high level scripting language;
3. Apache Mahout for Data Mining.

Hadoop is designed for large volumes of data and is batch oriented in nature—even a simple query may take minutes to come back. In a typical Big Data oriented analysis scenario a Data Scientist may start by selecting a much smaller set of data and transforming it in some fashion and then combining this with the relational data from the Data Warehouse for analysis in a range of tools. Big Data analysis is also typically very explorative and iterative in nature so significantly more freedom is required than may traditionally be the case in Information Management. This is discussed in more detail in subsequent sections in this white paper.

While Hadoop offers native capabilities in the form of the Map-Reduce framework to analyse data as well as

a wide range of analysis tools, Hadoop is more typically used as a preparatory step within an analysis process. Hadoop's low cost data storage model lends itself to providing a broad pool of data each item of which may be of limited value to the organization, but for any given business problem may complete a missing link. Data may be selected, transformed and enhanced before it is moved to a relational setting and combined with additional corporate data where a more interactive analysis can be performed.

Given Hadoop is (currently at least) batch oriented other technologies are required in order to support real-time interactions. The most common technologies currently in use within this area are Complex Event Processing (CEP), In-Memory Distributed Data Grids, In-Memory Databases and traditional relational databases. These may be supported by other related technologies such as No-SQL databases, either sitting on top of a Hadoop cluster or using a specific data storage layer.

4. Proposed Framework

A) IoT Basic Data Transformation Model: **We have presented the data transformation model as shown in Table 1.**

B) Basic Framework for IoT Data Manage.

The framework should support the whole IoT data lifecycle and explore the benefit of the data storage/preservation, aggregation and provenance in a large scale and during long/unlimited period of time. Important is that this infrastructure must ensure data security (integrity, confidentiality, availability, and accountability), and data ownership protection. With current needs to process big data that require powerful computation, there should be a possibility to enforce data/dataset policy that they can be processed on trusted systems and/or complying other requirements.

We divide an IoT data management system into various layers as depicted in **Table 2**. Layer 1 interacts directly with the interconnected IoT objects and sensors. Data production involves sensing, collecting and sending data by the Things within the IoT framework and reporting this data to Pre-processing layer. Layer 2 involves Data Pre-processing. IoT data will come from different sources with varying formats and structures. Data may need to be pre-processed to handle missing data, remove redundancies and integrate data from different sources into a unified schema before being committed to storage. Moreover brute-force fitting of all the data into a fixed relational (tables) schema cannot be done with IoT data, but rather a more abstract definition of a consistent way to access the data without having to customize access for each source's data format(s). Layer 3 is storage-intensive; involving the mass storage of produced data for later processing and analysis and more in-depth queries. Layer 4 is dedicated for applying data mining techniques or partitioning schemes in order to easily managing the data. It also involves data curation which includes retirement and other clean-up of unwanted data. Layer 5 incorporates Analysis and Real time event processing. Layer 6 relates with data visualization techniques and layer 7 is application layer. Above Presented Framework should incorporate following features **Table 3**.

5. Conclusion

We have entered in IoT world which is generating lot of data. This Massive data technology era creates lots of exciting opportunities and research challenges. The paper has presented an architectural framework for managing IoT data. It has also discussed the problems and challenges of IoT Data. We have also presented an overview of computing infrastructure for IoT data processing, focusing on architectural and major challenges of

Table 1. IoT data transformation.

Step	Stages	Type of data
1	Data Source	Raw Data
2	Data Collection and Registration	Non Structured Data
3	Data Processing (Filtering, Enriching)	Classified Structured Data
4	Data Analysis (classification, Prediction)	Processed Data
5	Data Delivery & Visualization	Viewable Data
6	Consumer	Information

Table 2. Framework for managing IoT data.

Layer 7	Application
Layer 6	Visualization techniques
Layer 5	Data Analysis & Real Time event Processing
Layer 4	Specialized Data storage (Data Mining techniques, partitioning. Data curation techniques)
Layer 3	Data storage/Management & Archival
Layer 2	Data Pre-processing (Cleaning, Removal of duplicates)
Layer 1	Data Acquisition/Production (RFID Tags, Sensors etc.)

Table 3. Characteristics of framework.

Support for multiple data types
Handle batch processing and/or real time data streams
Utilize what already exists in your environment
Support NoSQL and other newer forms of accessing data
Overcome low latency
Provide cheap storage
Integrate with cloud deployments

massive data. We will briefly discuss about emerging computing infrastructure and technologies that are promising for improving massive data management. In future we will incorporate partitioning schemes for achieving parallelism and scalability.

References

[1] Internet of Things—Strategic Research Roadmap. http://ec.europa.eu/information_society/policy/rfid/documents/in_cerp.pdf

[2] Schonfeld, E. (2010) Costolo: Twitter Now Has 190 Million Users Tweeting 65 Million Times a Day. http://techcrunch.com/2010/06/08/twitter-190-million-users/

[3] Want, R. (2004) Rfid—A Key to Automating Everything. Scientific American. http://dx.doi.org/10.1038/scientificamerican0104-56

[4] Fuhrmann, T. and Harbaum, T. (2003) Using Bluetooth for Informationally Enhanced Environments. *Proceedings of the IADIS International Conference e-Society* 2003, Lisbon, 2003.

[5] Adelmann, R., Langheinrich, M. and Floerkemeier, C. (2006) A Toolkit for Bar Code Recognition and Resolving on Camera Phones—Jump Starting the Internet of Things. Workshop Mobile and Embedded Interactive Systems (MEIS 2006) at Informatik.

[6] Rohs, M. and Gfeller, B. (2004) Using Camera-Equipped Mobile Phones for Interacting Real-World Objects. In: Ferscha, A., Hoertner, H. and Kotsis, G., Eds., *Advances in Pervasive Computing*, Autrian Computer Society (OCG).

[7] http://edutechwiki.unige.ch/en/Internet_of_things

[8] Zikopoulos, P., DeRoos, D., Parasuraman, K., Deutsch, T., Corrigan, D. and Giles, J. (2013) Harness the Power of Big Data McGraw-Hill.

[9] Spangler, S., Chen, Y., Proctor, L., Lelecu, A., Behal, A., He, B. and Davis, T. (2009) COBRA—Mining Web for Corporate Brand and Reputation Analysis. *Web Intelligence and Agent Systems*, **7**, 243-254.

[10] Gonzalez, H., Han, J.W., Li, X.L., *et al*. (2006) Warehousing and Analyzing Massive RFID Data Sets. *Proceedings of the* 22*nd International Conference on Data Engineering*, Atlanta, 83-92.

[11] Li, T.L., Liu, Y., Tian, Y., Shen, S., and Mao, W. (2012) A Storage Solution for Massive IoT Data Based on NoSQL. *IEEE International Conference on Green Computing and Communications*, 20-23 November 2012, Besancon, 50-57.

[12] Fan, T.R. and Chen, Y.Z. (2010) A Scheme of Data Management in the Internet of Things. *Proceedings of ICNIDC*-2010, Beijing, 24-26 September 2010, 110-114.

[13] Ding, Z.M. and Gao, X. (2012) A Database Cluster System Framework for Managing Massive Sensor Sampling Data in the Internet of Things. *Chinese Journal of Computers*, **35**, 1175-1191.

Scientific Research Publishing (SCIRP) is one of the largest Open Access journal publishers. It is currently publishing more than 200 open access, online, peer-reviewed journals covering a wide range of academic disciplines. SCIRP serves the worldwide academic communities and contributes to the progress and application of science with its publication.

Other selected journals from SCIRP are listed as below. Submit your manuscript to us via either submit@scirp.org or Online Submission Portal.