The Cyber Time

An initiation towards cyber security...

Editorial

WELCOME TO ANOTHER ISSUE OF *The Cyber Time*.

Another month, another issue. This issue fully dedicated to Big Data and related Technologies. Every day, we create 2.5 quintillion bytes of data .This data comes from everywhere: sensors used to gather climate information, posts to social media sites, digital pictures and videos, purchase transaction records, and cell phone GPS signals to name a few. This data is big data. It has become a challenge to store, retrieve and analysis of a huge amount of data. So this issue will give some insights to big data.

Hope you will enjoy the reading. All the best, and keep in touch!

Wish you very happy Holi

Hetram Yadav

spu1311211@policeuniversity.ac.in

Highlights

- Big Data: An Introduction
- Hadoop
- Big Data: the Future of Healthcare

Recent Achievement

- Meet with Delegates of ICSI
- Successful Celebration of Anunaad-2015
- Sports and Cultural Events

Big Data

-Nishant Grover

Big data is a 'small term' used to describe a massive volume of both structured and unstructured data that is so big that it is difficult to process using traditional database style, software tools and techniques. In most enterprise scenarios the data is either too big or it moves too fast or it exceeds current processing capacity of processor. Big data has the potential to help companies improve their operations, make them faster and allows them to make more intelligent decisions.

An example of big data might be petabytes (1,024 terabytes) or Exabyte's (1,024 petabytes) of data consisting of billions to trillions of records of millions of people all from different sources For example, Web, sales, customer contact center, social media, mobile data. The data is typically loosely structured data that is often incomplete and inaccessible until preprocessed. When dealing with larger datasets, organizations face difficulties in being able to create, manipulate, and manage big data. Big data is particularly a problem in business analytics because standard tools and procedures are not designed to search and analyze massive datasets.

Big data analytics refers to the process of collecting, organizing and analyzing large sets of data i.e. to discover patterns and other useful information out of it. Big data analytics will help organizations to better understand the information contained within the data and will also help identify the data that is most important to the business and future business decisions. Big data analysts basically want the knowledge that comes from analyzing the data. For most organizations, big data analysis is a challenge. Consider the sheer volume of data and the many different formats of the data (both structured and unstructured data) collected across the entire organization and the many different ways different types of data can be combined, contrasted and analyzed to find patterns and other useful information.

The first challenge is in breaking down data silos to access all data an organization stores in different places and often in different systems.

A second big data challenge is in creating platforms that can pull in unstructured data as easily as structured data. This massive volume of data is typically so large that it is difficult to process using traditional database and software methods.

For instance, Netflix mined its subscriber data to put the essential ingredients together for its recent hit House of Cards, and subscriber data also prompted the company to bring Arrested Development back from the dead.

Enterprises are increasingly looking to find actionable insights into their data. Many big data projects originate from the need to answer specific business questions. With

the right big data analytics platforms in place, More accurate analyses can be made which lead to more confident decision making. And better decisions can mean greater operational efficiencies, cost reductions and reduced risk.

Hadoop

-Pragya Johari

Hadoop is an open source software framework for storing and processing big data in a distributed way on large clusters. It is free and it is based on JAVA. Hadoop is based on simple data model, i.e. "any data will fit." It is part of the Apache project sponsored by the Apache Software Foundation.

Hadoop makes it possible to run applications on systems with thousands of nodes involving thousands of terabytes. Its distributed file system facilitates rapid data transfer rates among nodes and allows the system to continue operating uninterrupted in case of a node failure. This approach lowers the risk of catastrophic system failure, even if a significant number of nodes become inoperative. Hadoop was inspired by Google's MapReduce, a software framework in which an application is broken down into numerous small parts. Any of these parts (also called fragments or blocks) can be run on any node in the cluster.

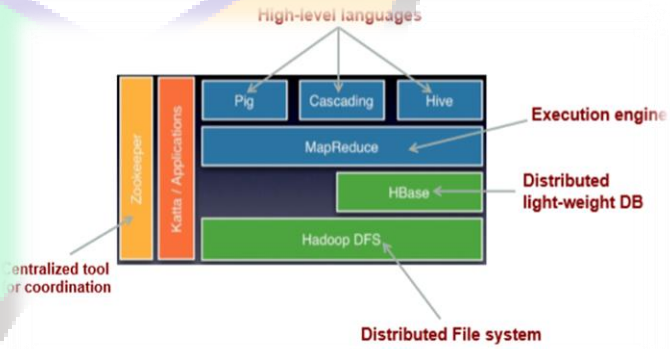
It mainly accomplishes two tasks: massive data storage and faster processing. Hadoop is a very famous technology. One of the top reasons is its ability to handle huge amounts of data- any kind of data- quickly. Data growing each day in volume and in verities, especially from social media and automated sensors, that's a key consideration for most organization. Other reasons include:

- Low Cost
- Computing Power
- Scalability
- Storage Flexibility

What is in Hadoop?

Three basic components are included in Hadoop from the Apache Software Foundation. These are:-

- **HDFS (Hadoop Distributed File System)–** the Java-based distributed file system that can store all kinds of data without prior organization.
- **MapReduce –** a software programming model for processing large sets of data in parallel.
- **YARN–** A resource management framework for scheduling and handling resource requests from distributed applications.



- **Pig –** a platform for manipulating data stored in HDFS. It consists of a compiler for MapReduce programs and a high-level language called Pig Latin. It provides a way to perform data extractions, transformations and

loading, and basic analysis without having to write MapReduce programs.

- **Hive –** a data warehousing and SQL-like query language that presents data in the form of tables. Hive programming is similar to database programming. (It was initially developed by Facebook.)
- **HBase –** a nonrelational, distributed database that runs on top of Hadoop. HBase tables can serve as input and output for MapReduce jobs.
- **Zookeeper –** an application that coordinates distributed processes.
- **Ambari –** a web interface for managing, configuring and testing Hadoop services and components.
- **Flume –** software that collects, aggregates and moves large amounts of streaming data into HDFS.
- **Sqoop –** a connection and transfer mechanism that moves data between Hadoop and relational databases.
- **Oozie –** a Hadoop job scheduler.

Q. A functional dependency between two or more nonkey attributes is called?

A. Transitive dependency

B. Partial transitive dependency

C. Functional dependency

D. Partial functional dependency

Big Data: the Future of HealthCare

-Vikas Yadav (Editor)

Big data, a general term for the massive amount of digital data being collected from all sorts of sources, is too large, raw, or unstructured for analysis through conventional relational database techniques. big data as "a term that describes large volumes of high velocity, complex, and variable data that require advanced techniques and technologies to enable the capture, storage, distribution, management, and analysis of the information.

Big data goes beyond size and volume to encompass such characteristics as variety, velocity, and, with respect specifically to health care, veracity.

Volume: Big data means there is a lot of data terabytes or even petabytes. This is perhaps the most immediate challenge of big data, as it requires scalable storage and support for complex, distributed queries across multiple data sources. While many organizations already have the basic capacity to store large volumes of data, the challenge is being able to identify, locate, analyze and aggregate specific pieces of data in a vast, partially structured data set.

Variety: Big data is an aggregation of many types of data, both structured and unstructured, including multimedia, social media, blogs, Web server logs, financial transactions, GPS and RFID tracking information, audio/video streams and Web content. While standard techniques and technologies exist to deal with large volumes of structured data, it becomes a

significant challenge to analyze and process a large amount of highly variable data and turn it into actionable information. But this is also where the potential of big data potential lays, as effective analytics allow you to make better decisions and realize opportunities that would not otherwise exist.

Velocity: While traditional data warehouse analytics tend to be based on periodic daily, weekly or monthly loads and updates of data, big data is processed and analyzed in real or near real time. This is important in healthcare for areas such as clinical decision support, where access to up-to-date information is vital for correct and timely decision-making and elimination of errors.

Big Data = Big Opportunities

Benefits to Health Care

By digitizing, combining and effectively using big data, health care organizations ranging from Single physician offices to multi-provider groups, large hospital networks, and accountable care organizations stand to realize significant benefits. These can include improving the quality and efficiency of health care delivery; detecting diseases at earlier stages, when they can be treated most successfully: managing specific health populations and individuals and detecting health care fraud more quickly and efficiently.

Health Care Quality and Efficiency

As of 2010, national health expenditures represented 17.9% of gross domestic product, up from 13.8% in 2000. Concurrently, the prevalence of chronic diseases like diabetes is growing and consuming a greater percentage of health care resources. Electronic health records (EHRs), coupled with new analytics tools, open the door to mining information for the most effective outcomes across large populations. Using de-identified information, researchers can look for statistically valid trends and provide assessments based upon true quality of care.

Earlier Disease Detection

Electronic sensors are increasingly being employed to monitor key biochemical markers, with real-time analysis taking place as the data streams from individual patients to HIPAA-compliant analysis systems. Analytics like these can alert specific individuals and their providers to potentially adverse events, such as side effects to medications, the early development of infection, and allergic reactions.

Fraud Detection

Big data is widely expected to fundamentally transform medical claims payment systems, resulting in reduced submissions of improper, erroneous or fraudulent claims. For example, a significant challenge confronting the Centers for Medicare and Medicaid Services (CMS) is managing improper payments under the Medicare Fee-For-Service Program (FFS). Until recently, claims were manually reviewed against medical documentation submitted by providers to verify compliance with

Medicare policies. Using powerful, new big data tools, techniques and governance processes, CMS is dramatically improving its fraud-detection efforts. In fiscal year 2011, for the second year in a row, CMS anti-fraud activities resulted in more than $4 billion in recoveries, an all-time high, owing in large part to big data-based detection and analytics tools.

Big data also can be used to improve such endeavors as population health management, the identification and measurement of more accurate quality metrics, the management of capitated populations, and treatment protocols for a wide range of chronic conditions such as diabetes and congestive heart failure (CHF). Through the use of data mining techniques, big data can be used to identify patients and populations at risk for various conditions and diseases, identify adverse drug events, and improve selection of candidates for patient- centered interventions and identify costly procedures, waste and delays. All told, the health care industry can potentially realize as much as $300 billion in annual value by effectively leveraging big data.

Some examples of early success capturing value from big data in health care:

1. The University of Ontario's Institute of Technology partnered with a prominent technology firm IBM to develop Project Artemis, a highly flexible platform that leverages streaming analytics to monitor new born in the neonatal intensive care unit of a hospital. Using these technologies, the hospital was able to predict the onset of nosocomial infections 24 hours before symptoms appeared. The hospital also tagged all time-series data that had been modified by software algorithms. In case of a lawsuit or medical inquiry, the hospital felt that it had to produce both the original and modified readings. Plus, the hospital established policies around safeguarding protected health information.

2. The California-based integrated managed-care consortium Kaiser Permanente connected clinical and cost data to provide a crucial dataset that led to the discovery of adverse drug effects and subsequent .withdrawal of the drug Vioxx from the market.

Future Applications: Individual and Population Health Management

1. Individual patient care: One of the goals of health care transformation is to garner the ability to personalize care for an individual patient. Big data is poised to make a significant contribution toward that goal in the coming years. For example, each diabetic patient has a combination of characteristics that can identify her within the larger population. Clinical decision support (CDS) systems already include computerized physician order-entry capabilities that analyze entries and compare them against medical guidelines to alert for such potential errors as adverse drug reactions. By deploying these systems, providers can potentially reduce adverse reactions and lower treatment error rates and liability claims, especially those arising from clinical mistakes.

2. Population health management: not only individual patient care but also help in

Population health management. The United Kingdom's National Health Service (NHS) has announced plans to sequence the entire genomes of up to 100,000 patients over three to five years to build a national research database. The database will help clinicians and researchers better understand the genetic causes of cancer and other rare conditions. Based on this data, it is believed that new drugs, treatments, and therapies can be developed and patients can receive targeted therapies that may prove more effective. Genetic sequencing will be voluntary, and the information in the database will be kept anonymous.

Reference: http://ihealthtran.com/

Recent Achievement

Meet with Delegates of ICSI: Delegates from Israeli College for Security and Investigations (ICSI), Israel interacted delegates of SPUP, Jodhpur to discuss about possible collaborations between SPUP and ICSI on 25th February, 2015 in Ahmadabad.

Successful Celebration of Anunaad-2015

The university celebrated its second foundation day, Anunaad-2015 from 2nd Feb to 4th Feb, A number of competitions like rangoli designing on themes of environment and social relevance, pot painting, collage and poster making, dance, singing, food without flame, antakshari etc. were organized in which the students showcased their immense talent and creativity. The minor prize distribution function and cultural program took place in the

evening of 3rd, which was graced by Shri Harendra Kumar Mahawar, IPS S.P (Rural) Jodhpur. An Inter University quiz, an annual feature of the celebration was organized on the morning of the 4th. The celebration culminated with the prize distribution for winners of 1st prize + trophies in the different sports and cultural events and the evening was graced by the distinguished presence of Shri Manoj Bhatt, IPS, DG, and ACB.

Sports and Cultural Events: SPUP, Jodhpur will be celebrating Sports events from 29 - 31st Jan, 2015 and various Cultural Events on 1-3 Feb, 2015. Get Ready for Some Excitement.

Q. A type of query that is placed within a WHERE or HAVING clause of another query is called?

A. Super query

B. Sub query

C. Master query

D. Multiquery

Q. What is the purpose of the PSH flag in the TCP header?

A. Typically used to indicate end of message

B. Typically used to indicate beginning of message

C. Typically used to push the message

D. Typically used to indicate stop the message

Call for articles:

Students are invited to get involved in the TechNewsletter activities by providing articles and other related materials. Suggestions and feedbacks for the improvement of the newsletter are most welcome and contributions are invited from the faculty and students of the department. Contributions can be from any of the whole gamut of activities in the department like any special achievement, an admirable project, a publication, and Cyber Crime case, Quiz, puzzles or even the fun section material like jokes, cartoons, interesting facts or poems. You can also report any interesting workshops or talks taking place in the department.

You can send your material on: - **editors@policeuniversity.ac.in**

Note: - If any of the article is found to be copied, the writer himself/herself will be responsible for copyright issues. Editor or University will not be liable for any issue.

Brought out by the Department of Computer Science & Cyber Security

Sardar Patel University of Police, Security & Criminal Justice, Jodhpur