A SEMINAR REPORT

on

"BIG DATA TO AVOID WEATHER RELATED FLIGHT DELAYS "

SUBMITTED TO THE SAVITRIBAI PHULE PUNE UNIVERSITY,PUNE

In Partial Fulfillment of
T.E. Computer Engineering
T.E. Semester II

By

Mr. PUSHKAR GIRISH ZAGADE
Exam Seat No: T120404353

UNDER THE GUIDANCE OF
PROF. NEELIMA R. SATPUTE

DEPARTMENT OF COMPUTER ENGINEERING
JSPM's Jayawantrao Sawant College of Engineering
Hadapsar, Pune-028.
[2014-15]

CERTIFICATE

DEPARTMENT OF COMPUTER ENGINEERING

JSPM's Jayawantrao Sawant College of Engineering
Hadapsar, Pune-028.

This is to certify that the Seminar Report entitled

"BIG DATA TO AVOID WEATHER RELATED FLIGHT DELAYS "

Submitted by

Mr. PUSHKAR GIRISH ZAGADE

Exam Seat No:T120404353

is a bonafide work carried out under the supervision of Prof. NEELIMA S. SATPUTE and it is submitted towards the partial fulfilment of the requirement of Savitribai Phule Pune University,Pune.

(Prof. Neelima Satpute) (Prof.Neeta Saware)
Internal Guide External Examiner

(Prof. A.S.Devare) (Prof. H.A.Hingoliwala)
Seminar Coordinator H.O.D.

(Dr.M.G.JADHAV)
Principal

Place: Pune

Date:

Acknowledgments

It is our proud privilege and duty to acknowledge the kind of help and guidance received from several people in preparation of this report. It would not have been possible to prepare this report in this form without their valuable help, cooperation and guidance.

First and foremost, we wish to record our sincere gratitude to our beloved Principal, **Dr. M. D. Jadhav, Principal, Jayawantrao Sawant College Of Engineering ,Pune** for his constant support and encouragement in preparation of this report and for making available library and laboratory facilities needed to prepare this report.

Our sincere thanks to **Prof. H.A.Hingoliwala, Head, Department of Computer Science and Engineering, JSCOE** for his valuable suggestions and guidance throughout the period of this report.

We express our sincere gratitude to our guide, **Asst. Prof. Neelima R. Satpute, Department of Computer Science and Engineering, JSCOE, Pune** for guiding us in investigations for this seminar and in carrying out experimental work. Our numerous discussions with his were extremely helpful. We hold his in esteem for guidance, encouragement and inspiration received from his.

The seminar on **Big Data ToAvoid Weather Related Flight Delays** was very helpful to us in giving the necessary background information and inspiration in choosing this topic for the seminar. Our sincere thanks to **Prof. A.S.Devare , Seminar Coordinator** for having supported the work related to this project. Their contributions and technical support in preparing this report are greatly acknowledged.

(Mr.PUSHKAR GIRISH ZAGADE)

Exam no: T120404353

Batch[2014-15]

Abstract

This paper identifies key aviation data sets for operational analytics, presents a methodology for application of big-data analysis methods to operational problems, and offers examples of analytical solutions using an integrated aviation data warehouse. Big-data analysis methods have revolutionized how both government and commercial researchers can analyze massive aviation databases that were previously too cumbersome, inconsistent or irregular to drive high-quality output. Traditional data-mining methods are effective on uniform data sets such as flight tracking data or weather. Integrating heterogeneous data sets introduces complexity in data standardization, normalization, and scalability. The variability of underlying data warehouse can be leveraged using virtualized cloud infrastructure for scalability to identify trends and create actionable information. The applications for big-data analysis in airspace system performance and safety optimization have high potential because of the availability and diversity of airspace related data. Analytical applications to quantitatively review airspace performance, operational efficiency and aviation safety require a broad data set. Individual information sets such as radar tracking data or weather reports provide slices of relevant data, but do not provide the required context, perspective and detail on their own to create actionable knowledge. These data sets are published by diverse sources and do not have the standardization, uniformity or defect controls required for simple integration and analysis. At a minimum, aviation big-data research requires the fusion of airline, aircraft, flight, radar, crew, and weather data in a uniform taxonomy, organized so that queries can be automated by flight, by fleet, or across the airspace system.

Contents

Chapter 1

INTRODUCTION

1.1 BIG DATA

Recent years have witnessed a dramatic increase in our ability to collect data from various sensors, devices, in different formats, from independent or connected applications. This data ood has outpaced our capability to process, analyze, store and understand these datasets. Consider the Internet data. The web pages indexed by Google were around one million in 1998, but quickly reached 1 billion in 2000 and have already exceeded 1 trillion in 2008. This rapid expansion is accelerated by the dramatic increase in acceptance of social networking applications, such as Facebook, Twitter, Weibo, etc., that allow users to create contents freely and amplify the already huge Web volume.

Furthermore, with mobile phones becoming the sensory gateway to get realtime data on people from different aspects, the vast amount of data that mobile carrier can potentially process to improve our daily life has significantly outpaced our past CDR (call data record)-based processing for billing purposes only. It can be foreseen that Internet of things (IoT) applications will raise the scale of data to an unprecedented level. People and devices (from home cofee machines to cars, to buses, railway stations and airports) are all loosely connected. Trillions of such connected components will generate a huge data ocean, and valuable information must be discovered from the data to help improve quality of life and make our world a better place. For example, after we get up every morning, in order to optimize our commute time to work and complete the optimization before we arrive at once, the system needs to process information from trafic, weather construction, police activities to our calendar schedules, and perform deep optimization under the tight time constraints.

In all these applications, we are facing significant challenges in leveraging the vast amount of data, including challenges in (1) system capabilities (2) algorithmic design (3) business models.

1.2 RESEARCH IN THE UNIVERSITY OF MICHIGAN

The students from the University of Michigan have started a new research which helps in understanding the weather of a particular place. They have taken data of the weather of the past 10 years. The analysis of this data helps in understanding the patterns in the weather. This is a very creative and new process. It could lead to understanding similarities in the weather in the past years. It could be of help in predicting the weather in the future. This can be very helpful for flights. With the help of this data, the flights can be cautious of bad weather in advance. So it will be usefull.

1.3 MORE ABOUT BIG DATA FOR PREDICTIVE ANALYTICS

The data used in this research is available publicly. Since it the hourly data of the last ten years, the data is huge in quantity. Hence, it has to be managed cleverly and all of it must be taken into consideration. The study of the weather is carried out keeping the flights and their journeys in mind. This enables the researchers to understand the effect of weather on a particular journey. This is a very unique study. It will help in predicting the delays or preventing them in certain cases.

Chapter 2

DISSERTATION PLAN

This topic is generally belongs to weather forecasting that is how we implement Big Data computing for future weather prediction.The Objective and Aim of this report is to help Airlines by providing information of future weather which help to avoid flight delays and cancellation of flights.

This will help to solve social issues of flight delays by using BIG DATA computing method. As we know because of bad weather everyday lots of Flights has been canceled or dalayed. This is a big SOCIAL ISSUES we need to solve to avoid fight delays. As a Software Engineer , Engineers from University of Michigan developed a Big Data Computing method to predict the future weather . By using Big Data Computing method , they try to predict the future weather to avoid weather related fight delay.

Chapter 3

STARTING WITH BIG DATA

3.1 WHAT IS BIG DATA...?

Big data is a very popular term that is used to describe the large growth and availability of data, which can be both structured and unstructured. And big data are very important to business and society as well as the Internet which has become popular.

3.2 THREE V'S OF BIG DATA

In year 2001, industry analyst Doug Laney (currently with Gartner) had articulated the now mainstream definition of big data as the three Vs of big data which are volume of data, velocity of data and variety of data.

- **Volume** Many factors contribute to the increase in data volume of data. The transaction based data storage through many years. Unstructured data are coming from social media. Increase in amounts of sensor and machine to machine data gonna be collected. In the past years, excessive data volume was a big storage issue. But with the decreasing costs of storage, other issues emerge, including how to determine relevance within large volume of data and how to use analytics to create value from relevant data.

- **Velocity** Data is streaming in at unprecedented speed and must be dealt with in a timely manner. RFID tags, sensors as well as smart metering are driving the need to deal with torrents of data in real time. Reacting quickly enough to deal with data velocity is a big challenge for most organizations.
- **Variety** Data today comes in all types of formats that means Structured, numeric data in databases. Information created from line of business applications. Unstructured email, text documents,video, audio, stock ticker data and financial transactions. Managing, modifying and governing different varieties of data in many organizations still grapple with.

3.3 ADDITIONAL TWO DIMENTIONS OF BIG DATA

We consider two additional dimensions when thinking about big data:

- **Variability** In addition to the increasing in velocities and varieties of stored data, data flows can be highly inconsistent with periodic peaks. Is something really trending in social media? Daily, seasonal and event triggered peak data loads can be challenging to manage. Even more with unstructured data involved.

- **Complexity** Today's data comes from various and multiple sources which need to be analysed . And it is still an undertaking to the link, match, cleanse and transform data across this systems. However, it is necessary to connect and correlate relationships, hierarchie as well as multiple data linkages or your data can quickly spiral out of control.

3.4 PROPERTIES OF BIG DATA

TEN PROPERTIES OF THE PERFECT BIG DATA STORAGE ARCHITECTURE

- **Be Scalable** Any big data storage system should be scalable. What capacity will meet to your requirements? The problem with simply adding disks are that this model is not scalable in the number of ways. Scalability is not just all about size of data storage,but it has wider implications. The throughput and the speed of access should be scalable. In addition, the system should able to scale , that is, to grow quite large without a huge increase in staff.

- **Provide Tiered Storage** When it comes to storing and retrieving data, the first question is: How long can I wait to get the data I need? However, you also need to factor in the cost of storing the data on different tiers. An optimal big data storage architecture stores the data you need and archives what you dont need right away at the lowest possible cost. The best systems support a lifecycle that provides a home for data flows in each stage of the lifecycle, from creation through archiving.

- **Be Self Managing** Storage systems almost always have so many users. Most of the time, those users will be applications that are placing data in some type of storage and alerting the system when data needs to be moved back and forth between the tiers just mentioned. This sort of communication is hardwired into the applications via APIs. The apps tell the storage system what to do.

- **Ensure Content Is Highly Available** As petabyte-sized information stores increasingly become a key source of business advantage, there is a corresponding desire to keep this data forever while ensuring that it is highly available. Customers need to accomplish this objective without growing administrative or backup staff at the same rate data grows. Well-architected storage systems leverage their internal policy engine to automatically make copies of newly stored data across media and sites to assure basic data availability on top of traditional RAID architectures. But as data growth continues to outpace traditional approaches, this availability model is being challenged, particularly in customer environments that require complete reliance on disk storage.

- **Ensure Content Is Widely Accessible** Just as the increasing value of big data has made high availability a critical factor, it has also driven the need for content to be widely and quickly accessible as more users want to leverage the data to extract value. Often, these users are geographically dispersed and can even include suppliers and partners. As a result, distributing data geographically so that it is closer to users has become more important. In fact, says Lee, cloud storage is as much about providing wider access to data as it is about outsourcing the management of that data. So , One of the other advantages of wide area storage over RAID-based architectures is the greater geographic distribution and, in the case of most wide area storage technologies, a cloud-ready interface.

- **Support Both Analytical And Content Application** In the past, almost all the data companies had to deal with was records in databases. Each unit of data was small and the trick was sifting through huge collections, usually stored in SQL databases, to find the records you wanted.But in the modern world, the analysis model has been dramatically extended. Some of the most valuable analysis being done these days is massive parallel analysis of big unstructure files, whether huge web logs, FINANCIAL data, or sensor information. In some cases this is the same data being shared by human users in a content management application. However, the data performance requirements for these two uses are diametrically opposed; the best performance for human analysis requires an extreme service level for delivery of a single file or set of files to a single user so that even the most efficient granular, high performance dataset can be delivered with integrity while the best performance for computational analytical environments (like Hadoop) are instead reliant on the simultaneous movement of many streams of dataeach one perhaps a bit slower, but with the highest overall parallel throughput.

- **Support Workflow Atomation** Smaller unstructured data, typically end user productivity files, is typically for the use of single users. Big unstructured data is almost always driven by a set of data-sharing applications. Big data must be delivered to users in context of a workflowthe transfer of information from application to application and user to user. For this reason, a big data storage architecture must support easy integration of workflow. This may include a specialized professional application such as a content asset manager, a laboratory information management system, as well as a broadcast information system. Alternatively, it may be driven by customer-written applications or scripts.

- **Integrate With Legac Application** With the dramatic changes in both big data requirements and technologies, as outlined above, customers need the ability to leverage the latest big data technology (such as wide area storage). However, frequently vendors offer these new technologies only if the user is willing and able to forklift upgrade his or her prior system. Lee notes that customers deserve a better productization experience than this offers.

- **Enable Integration With Public, Private And Hybrid Cloud Ecosystem** Many of the storage tiers mentioned so far are going to be profoundly influenced by the cloud. It is possible to imagine huge networks of cloud-based computer memory, banks of flash drives, and wide area storage. As already mentioned, moving data to and from clouds is crucial. The ideal big data storage system must be built from the ground up to be cloud enablednot only for public clouds but also for private and hybrid cloud environments.

- **Be Self Healing** The largest scale websites and applications, those created by Facebook, Twitter,Google, and AMAZON, all have the built-in ability to handle failure. When a server in this cloud environment fails, there is no apparent outage, and this particular server is never fixed. Work is automatically redirected to another resource, with the failing server automatically taken offline for later removal. A well designed big data storage system must work in exactly this same model; it must accommodate component failures and heal itself without customer intervention.

3.5 APPLICATIONS OF BIG DATA

- Understanding and Targeting Customers
- Understanding and Optimizing Business Processes
- Personal Quantification and Performance Optimization
- Improving Healthcare and Public Health
- Improving Sports Performance
- Improving Science and Research
- Optimizing Machine and Device Performance
- Improving Security and Law Enforcement.

- Improving and Optimizing Cities and Countries

- Financial TRADING

Chapter 4

BIG DATA FOR WEATHER PREDICTION TO AVOID FLIGHT DELAYS

4.1 INTRODUCTION

Weather forecasting has been one of the most scientifically and technologically challenging problems around the world in the last century. This is due mainly to two factors: first, its used for many human activities and secondly, due to the opportunism created by the various technological advances that are directly related to this concrete research field, like the evolution of computation and the improvement in measurement systems . To make an accurate prediction is one of the major challenges facing meteorologist all over the world. Since ancient times, weather prediction has been one of the most interesting and fascinating domain. Scientists have tried to forecast meteorological characteristics using a number of methods, some of these methods being more accurate than others.

Weather forecasting entails predicting how the present state of the atmosphere will change. Present weather conditions are obtained by ground observations, observations from ships and aircraft, radiosondes, Doppler radar, and satellites. This information is sent to meteorological centers where the data are collected, analyzed, and made into a variety of charts, maps, and graphs. Modern high-speed computers transfer the many thousands of observations onto surface and upper-air maps. Computers draw the lines on the maps with help from meteorologists, who correct for any errors. A final map is called an analysis. Computers not only draw the maps but predict how the maps will look sometime in the future. The forecasting of weather by computer is known as numerical weather prediction.

4.2 WORKING OF BIG DATA FOR WEATHER PREDICTION TO AVOID FLIGHT DELAYS

Next year's holiday so many travelers may see fewer delays thanks to research now being conducted by a team of Engineers from Michigan University. They have gathered more than 10 to 15 years of hour by hour data of weather observations as well as domestic fight data, and they are using advanced data analytics to spot pattern and also help airlines manage more efficiently.

While the project uses public data that has been available for so many years, its size and scope make it unique, says Brian Lemay, a U-M doctoral student in industrial and operations engineering who leads the project.

"We are the first people who gather this data in one place and apply this level of computing to it," Lemay said. "That enables us to do a very very sophisticated analysis of how weather as well as flight delays are connected and also go far beyond individual airports. "We know that how the weather in Atlanta always affect flight operations in Detroit later in the day, or exactly how a delayed plane on the West Coast ripples through the system to California."

The chief goal is always to enable airlines to anticipate as well as to deal with delays of fliight before they happen, says Amy Cohn, a U-M associate professor in industrial and operations engineering who researches airline industry operations.

Today, most airlines compensate for delays by adding too much slack to the system. They may schedule it in extra flight time during the winter or try to keep additional staff members on call. But generally, they don't look at large scale weather patterns when they are building initial flight schedules. And their ability to shuffle resources to deal with weather patterns is limited.

"Airlines generally deal with weather delays after they happen," Cohn said. "We want to give them the ability to be a bit more proactive. When they're able to predict delays further in advance, they'll be able to do a much better job of communicating with passengers and optimizing resources." Cohn said the data from the project may be used to build computer-modeling software that could predict the outcome of an infinite number of hypothetical flight and weather scenarios, helping airlines spot likely weather delays in advance.

That knowledge could enable airlines to adjust their schedules to account for weather patterns. It may also lead to new options for passengers. For example, airlines could look several steps ahead to predict a future flight delay, then offer passengers a pre-emptive re-booking to avoid it. "Imagine you're scheduled to fly out of Detroit four hours from now and there's a storm in Atlanta." Cohn said. "The airline could use this data to determine that the storm in Atlanta is likely to delay your plane. They could then contact you and offer you a seat on an alternate flight. You save time, and the airline doesn't have to accommodate you on a later flight after the delay happens." Airlines could also use the advance warning to allocate their own resources more efficiently, shuffling ground crews, flight crews and other assets to minimize disruption.

The project draws on resources from a wide swath of disciplines including engineering, computer science and others, says George Tam, a U-M industrial and operations engineering graduate student. He says that breadth of knowledge has brought a fresh perspective that could hold farreaching benefits for both airlines and passengers. "Aeronautics experts think about airplanes and meteorologists think about the weather," he said. "But our background in industrial engineering and computer science enables us to put existing data together in new ways and ask a whole new set of questions. For me, that's what has been really exciting about this research."

Some of the first changes that passengers see are likely to be simple ones, like tweaks to flight times and more proactive communication. "I think these new analytics will enable passengers as well as airlines to better manage the whole travel process," Cohn said. "If airlines can offer more options and passengers can educate themselves on how to use those options, we'll see fewer delays and a less stressful travel experience in the years to come."

Chapter 5

DATA MINING IN FIELD OF BIG DATA

5.1 DATA MINING

Generally, data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information - information that can be used to increase revenue, cuts costs, or both. Data mining software is one of a number of analytical tools for analyzing data. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases.

5.2 DATA WAREHOUSE

Dramatic advances in data capture, processing power, data transmission, and storage capabilities are enabling organizations to integrate their various databases into data warehouses. Data warehousing is defined as a process of centralized data management and retrieval. Data warehousing, like data mining, is a relatively new term although the concept itself has been around for years. Data warehousing represents an ideal vision of maintaining a central repository of all organizational data. Centralization of data is needed to maximize user access and analysis. Dramatic technological advances are making this vision a reality for many companies. And, equally dramatic advances in data analysis software are allowing users to access this data freely. The data analysis software is what supports data mining.

5.3 WORKING OF DATA MINING TO PREDICT THE FUTURE

In this paper, there is one and only a rough step by step description of how the classification as well as prediction of weather forecasting is taking place, means a designing Classification and Prediction of Future Weather by using BackPropagation Algorithm technique is described.

The Classification and Prediction of Future Weather by using BackPropagation Algorithm is basically developed for forecasting weather and processing information.

5.3.1 STEP BY STEP DESCRIPTION OF OPERATION

- **Data Collection**

  The different sensors like rain sensor, wind sensor, and thermo-hygro sensor records different parameters like rainfall, wind, temperature and humidity. The recorded data is present in the form of datasheet. This data set is send for Pre-processing and then to the Statistical Software.

- **Pre-processing**

  The Pre processing step is used to remove the unwanted data or noise recorded by the sensors during transmission or it may refer to the selection of a particular area for consideration for prediction purpose.

- **Data Transfer**

The recorded data is transferred to the Statistica Software in order to give an input data.

- **Data Mining**

The Data Mining Technique is to be applied to the transferred data in order to validate data. This technique will be implemented by using Statistical Data Miner Software and by quantitative analysis. Quantitative Analysis is the process of presenting and interpreting numerical data. It can allow for greater objectivity and accuracy of results. Generally, quantitative methods are designed to provide summaries of data that support generalizations about the phenomenon under study. In order to accomplish this, quantitative research usually involves few variables and many cases, and employs prescribed procedures to ensure validity and reliability

- **Prediction of Future Weather using ANN by Back Propagation Algorithm**

In order to perform a BackPropagation Algorithm a program or logic must has to be created. What will be the change on other parameters by changing any one parameter, will be observed.

- **Classification**

After predicting data, what will be the weather in upcoming future after some periods, the Classification will take place. In Classification , it will display what will be the future weather, whether it will be sunny day or rainy or cloudy day what will be the change in speed of wind, humidity etc. the Classification Technique will help for taking some prevention from the climatic hazard.

Chapter 6

CONCLUSION

It concludes that the new technology Big data Computng can be used for weather forecasting process. Data Mining in field of big data compute accurate future weather. The system increases the accuracy ,reliability and consistency of identification and interpretation of weather . It also concludes that the BackPropagation Algorithm can also be applied on the forecasting weather data. Neural Networks are capable of modeling a weather forecast system. Which overall help airlines to avoid flight delays and cancellation of flight .

Bibliography

[1] Manyika JChui MBrown Bet al Big data: The next frontier for innovation,competition, and productivity McKinsey Global Institute.

[2] Research Trends Issue 30 September 2012.

[3] https://www.linkedin.com/pulse/20131113065157-64875646-the-awesome-ways-big-data-is-used-today-to-change-our-world

[4] Big Data for Development: From Information- to Knowledge Societies Martin Hilbert (Dr. PhD.) Hilbert, Big Data for Dev.; pre-published version, Jan. 2013

[5] S. Consulting. The New York City Taxicab Fact Book.[Online]. Available: http://www.schallerconsult.com/taxi/taxifb.pdf, accessed 2006.

[6] Taxi of Tomorrow Survey, New York City Taxi and Limousine Commission, New York, NY, USA, 2011.

[7] W. Wu, W. S. Ng, S. Krishnaswamy, and A. Sinha, To taxi or not to taxi? Enabling-personalised and real-time transportation deci- sions for mobile users, in Proc. IEEE 13th Int. Conf. Mobile Data Manage. (MDM), Jul. 2012, pp. 320 323.

[8] Y. Ge, H. Xiong, A. Tuzhilin, K. Xiao, M. Gruteser, and M. Paz-zani, An energy effcient mobile recommender system, in Proc. 16th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD), 2010,pp. 899 908.

[9] H. Yang, C. S. Fung, K. I. Wong, and S. C. Wong, Nonlinear pricing of taxi services,Transp. Res. A, Policy Pract., vol. 44, no. 5, pp. 337v348, 2010.

[10] K. Yamamoto, K. Uesugi, and T. Watanabe, Adaptive routing of cruising taxis by mutual exchange of pathways, in Knowledge-Based Intelligent Information and Engineering Systems. Berlin, Germany: Springer-Verlag, 2010.

[11] http://www.csmonitor.com/USA/Society/2013/0811/The-new-age-of-algorithms-How-it-aects-the-way-we-live/(page)/3 viewed 9 Sep 2013.

[12] APA citation: How big data reduce weather-related fight delays (2014, December 23) retrieved 20 February 2015 from http://phys.org/news/2014-12-big-weather-related-fight.html

[13] http://www.gartner.com/it-glossary/big-data/, viewed 15 Oct 2013.

[14] Data Mining with Big Data, Xindong Wu, Fellow, IEEE, Xingquan Zhu, Senior Member, IEEE, Gong-Qing Wu, and Wei Ding, Senior Member, IEEE

[15] Cross-platform aviation analytics using big-data methods Larsen, T.

[16] IntelBigthinkersonBigData, http://www.intel.com/content/www/us/en/bigdata/big-thinkers-on-bigdata. html, 2012.

[17] S. Consulting. The New York City Taxicab Fact Book. [Online]. Available: http://www.schallerconsult.com/taxi/taxifb.pdf, accessed 2006.

[18] http://www.masflight.com/masflight-news/wp-content/uploads/pdf/ICNS

[19] https://www.deepdyve.com/lp/institute-of-electrical-and-electronics-engineers/cross-platform-aviation-analytics-using-big-data-integration-methods-4Qqz2VTEMn

[20] http://ieeexplore.ieee.org/xpl/login.jsp?tp=arnumber=6548579url=http