**Logo, company name

Description automatically generated**

**BIG DATA IN WEATHER FORECAST SECTOR**

7BDIN006W - BIG DATA THEORY AND PRACTICE

Assessment 002- Technical Report (Group) Coursework

Module Leader: Tasos Ptohos

**Student ID’s (Group A)**

Dhivya Rekha Manikandan (w1871422)

Mohammad Ashraful Islam (w1861890)

Malaya Kumar Mishra (w1852625)

Hela Bouslama (w1700741)

Nirav Pandya (w1882379)



**Table of Contents**

[Abstract](#_ww7elx2d4mcu) 3

[Introduction](#_xs92aar5z30j) 3

[Discussions - An insight into weather forecasting](#_hiekqcidpj3s) 4

[Approaches in Weather forecasting](#_y5ddg6lhi2tz) 4

[Big Data Techniques in weather forecasting](#_a9vcdx6w6osu) 4

[Big Data Applications and Players](#_amfykje4zyui) 6

[Industries benefited from big data weather forecast models](#_dx4nofjtlzgd) 9

[Data Governance and ROI](#_1hgo1uzfha15) 10

[Challenges and Solutions](#_xrtepvr7821n) 11

[Bibliography](#_67ownl2oxith) 12

# 

# **Abstract**

Weather forecasting is critical for providing information to individuals and organisations that may be utilised to decrease weather-related losses and improve societal gains, like life and property protection, public health and safety, and economic well-being and standard of living. Accurate weather forecasting requires processing, collection, and storage of vast amounts of disparate weather data, so using big data for weather forecasting can provide many benefits, such as saving lives, improving quality of life, reducing risks, and increasing profitability benefitting industries as well. This paper discusses the big data technologies that are used in weather forecasting and how it benefits industries like agriculture etc. and the challenges faced while using the technology.

**Keywords**

Big data, Weather forecasting, Analyse data

# **Introduction**

Weather is an essential part for any living organism on this planet.Weather is immensely powerful such that it can influence permanent changes to an individual, society or the environment as a whole. Predicting weather is highly beneficial in many aspects. Weather forecasting makes a difference when natural catastrophes such as tornadoes, floods and hurricanes etc., occur allowing individuals and organisations to proactively plan for emergencies. In this internet era, data is omnipresent and it is collected from radars, radiosondes, weather satellites, buoys and other instruments. Data collected could come in formats such as JSON, XML, txt, csv files, image files, HTML, videos, sensor data, statistical data and traditional weather forecasting systems simply lack the speed and potential to process it. Every month a single user generates 40 Exabytes (approximately) of data and this increases exponentially if it is calculated for all users of the world. This demonstration of rapid data growth is handled using Big data. Big data is a concept used to describe large-scale structured, semi-structured or unstructured data derived from several places. The goal of our report is to analyse big data in weather forecast modelling and how it impacts industries. The first part of the paper focuses on approaches to big data in weather forecast, then second part about applications, players and Industries benefited from big data weather forecast models and last part of the report it is about Challenges, Complexity, Maintenance, ROI and Data Governance.

# **Discussions - An insight into weather forecasting**

## **Approaches in Weather forecasting**

The following are the different methods used in forecasting:

* Weather Trends - Future forecasting is based on historical and present weather conditions.
* Synoptic Forecasting - Synoptic, which means summary, is a method used by forecasters for predicting weather based on empirical laws derived from a series of synoptic charts gathered over time. These guidelines aid in determining the rate and direction of weather system migration.
* Weather Prediction Using Numerical Methods -Numerical weather prediction (NWP) is a weather forecasting technique that uses a set of equations to predict weather based on current weather . Before being performed across a domain, these equations are converted into computer code and include governing equations, numerical techniques, parameterizations of various physical processes, and beginning and boundary conditions (geographic area).
* Statistical Forecasting - Statistical approach is used along with numerical weather prediction to predict weather based on the physical interactions between the elements. It uses regression equations between climatic conditions and weather elements.

## **Big Data Techniques in weather forecasting**

Even though we have approaches like weather trends, synoptic forecasting, numerical weather predictions, and statistical forecasting, the level of accuracy is a concern. For example, weather trends can be suitable for regions like Ecuador where weather change is ceaseless but not for regions like Birmingham, UK which faces unpredictable weather. Similarly numerical weather predictions are a time-consuming process and directly proportional to the equation used for analysis, hence a slight mistake can disturb the accuracy.

When it comes to weather data, the data must be used within minutes. So, in order to generate relevant data, data must flow in and out swiftly and be recycled quickly, within minutes. The Big data revolution with its massive storage farms and super-computers brings hyper real-time forecasting a reality.

## **Implementation using Map-Reduce:**

Apache Hadoop is an open source framework consisting of Hadoop Distributed File System (HDFS) and Map Reduce (Vaishnavi,Rizwan and Dechakka, 2020).

HDFS handles file distribution in clusters among nodes and follows master-slave system, Map Reduce handles processing large data. HDFS consists of a single name node synchronised with a backup node. Hadoop when writing a file to HDFS breaks and makes replicas of the data files and stores them in different data nodes for fault tolerance and name nodes will contain metadata about them.

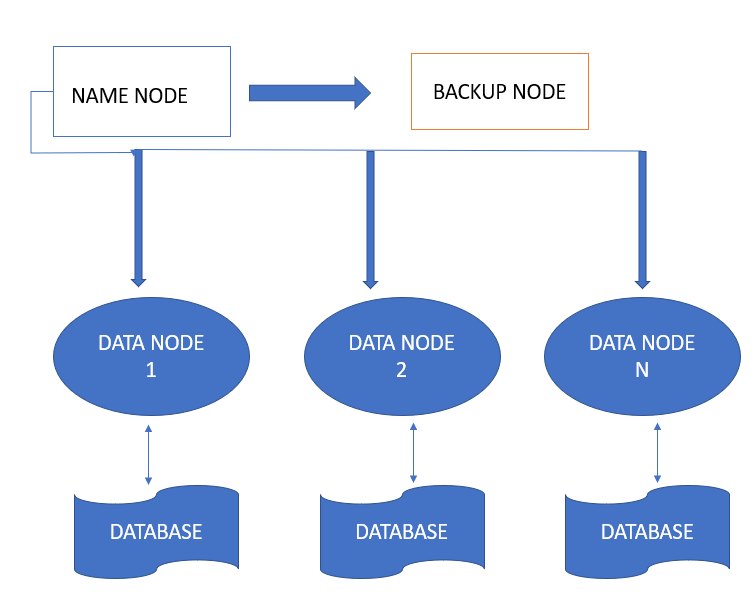


Figure: 01

Map-reduce algorithm splits the input weather predictions into multiple blocks of data which are then sorted and matched against other data sets. The process of mapping data into key-value pairs is called ‘Mapping’ and the data is rearranged and sorted in the process of ‘shuffling’. After shuffling, the data is reduced as individual blocks and the final forecast will be the data with the highest occurrences.

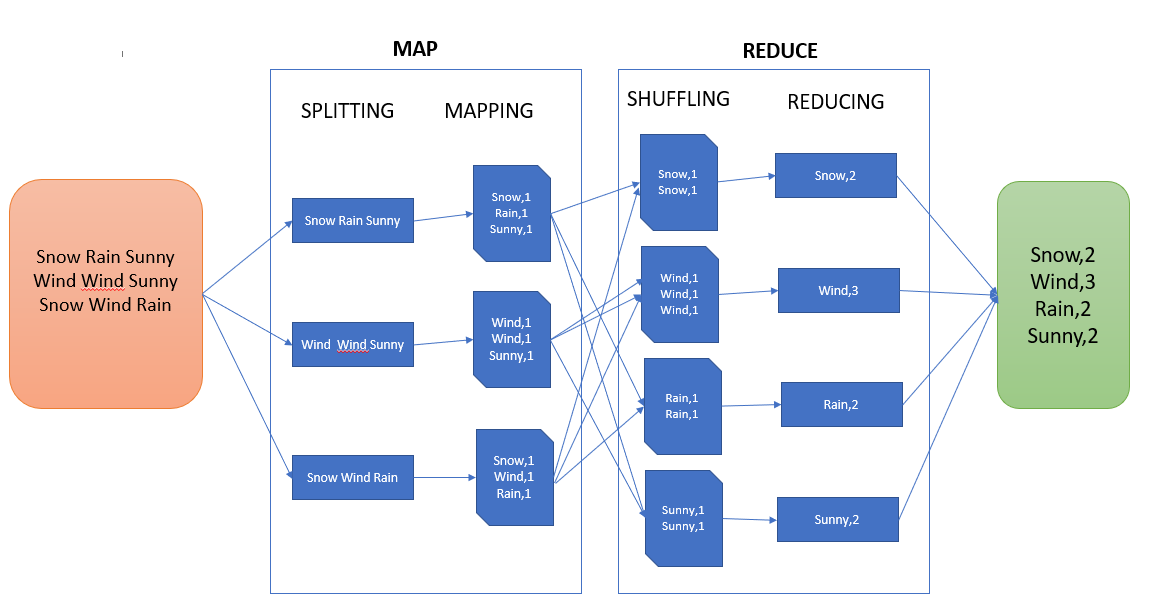


Figure: 02

The above diagram shows a map-reduce process on sample data. The highest occurrence ‘Wind’ would be presented as output.

Hadoop, being a parallel data processing framework may run map-reduce jobs for minutes or hours. The same implementation when using Apache Spark improves the performance ten times more as it runs on top of the Hadoop model that can be used for real-time streaming data processing (B Sri Divija et al,2020). Spark also has in-memory data storage capabilities meaning it stores the intermediate data in the RAM instead of writing to a disk. Resilient Distributed Datasets is another advantage implying it can hold any type of data.

M.P Ramos et al ,2022 have discussed implementing ensemble members method along with big data technologies for data processing like Python, Apache Hadoop, Apache HIVE and Optimised Row Columnar format.

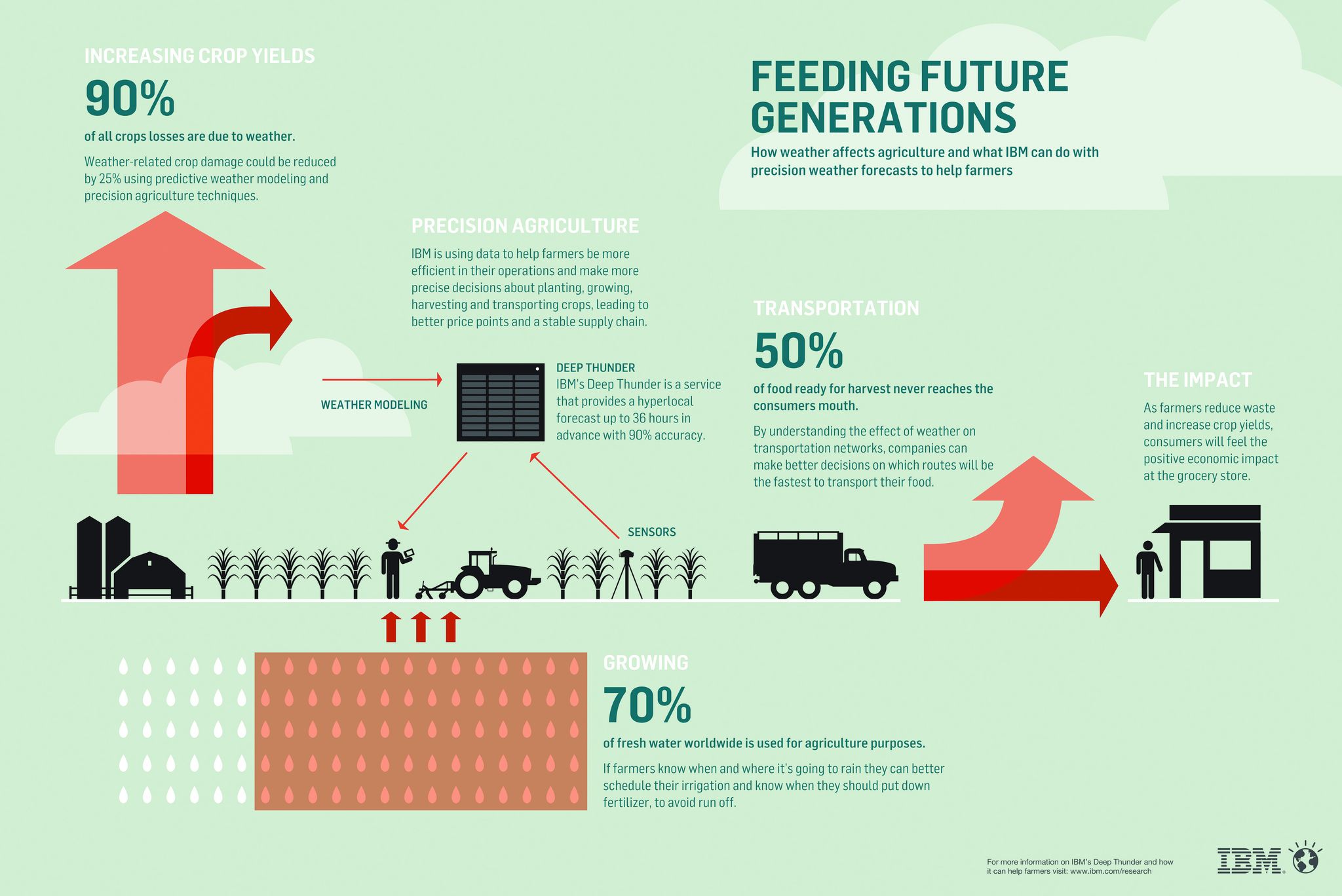
Wei Fang et al,2014 have implemented meteorological data analysis using Mapreduce for distributed computing and K-Means clustering algorithm to cluster large-scale weather data.

## **Big Data Applications and Players**

In an era where millions of petabytes of data is produced, organisations are harnessing data insights to improve decision-making, enter new markets and deliver better customer experiences. Here we have discussed IBM's Weather company, the largest network in the world of personal weather stations, offering about 26 million daily forecasts on weather and Weather Decision Technologies Inc.

## **IBM Deep Thunder for hyper local prediction**

IBM's Deep Thunder is a service demonstrating the viability of cloud-scale weather modelling to predict a hyper local forecast up to 36 hours in advance with 90% accuracy. The following figure illustrates how IBM Deep Thunder benefits agriculture and transportation.

Figure: 03

**IBM Deep Thunder in Agriculture**

Food is a basic necessity and 90% of all crop losses are due to weather resulting in famine all over the world. Weather-related crop damage could be reduced by 25% using predictive weather modelling and precision agriculture techniques enabling farmers to be more efficient in planning when to irrigate to avoid run-off, and when to fertilise to increase yield. By understanding the effect of weather on transportation networks, companies can make better decisions on which routes will be fastest to transport their food. Positive economic impact may be experienced by consumers due to reduction in wastage and increment in crop yields.

**Architecture**

Data acquired from various sources like ‘The Weather Company’ , satellites, sensor data from NOAA, NASA, the US Geological Survey, and the European Space Agency are used along with IBM Watson and cloud platform solutions to provide weather forecasting resolution in the USA.

Around 400 TB data is processed generating tens of millions of forecasts around the globe within microseconds at 15 minute intervals. IBM GRAF, an advanced version was launched in 2019 which can process data volumes of up to 3.5 petabytes.

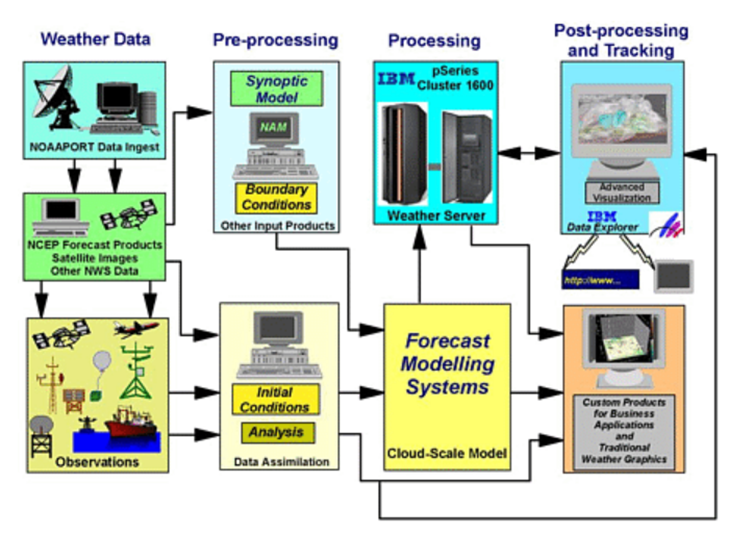
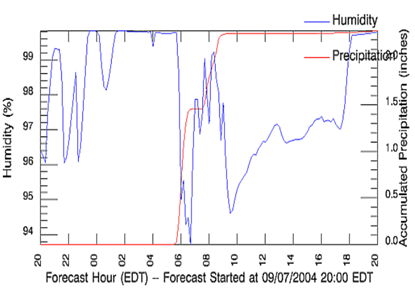
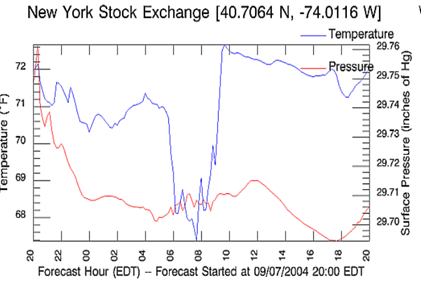


Figure: 04

1. A non-hydrostatic, terrain following coordinate system with 3-way nested configuration is utilised. Data received from various sources is pre-processed and converted to usable formats for analysis and visualisation.
2. IBM RS/6000 scalable power parallel(eleven nodes of four 375 MHz Power3 processors and one node of eight 222 MHz Power3 processors) and IBM pSeriesCluster 1600(five nodes of four 1.7 GHz Power4 processors and one node with two 1.2 GHz Power4 processors) is employed for parallelised simulation.
3. The modelling software is parallelised using Scalable Modelling System/Nearest-Neighbour Tool. Modelling domain for all nests is decomposed to be used by processors and mapped to MPI tasks within a node . An additional processor collects results from MPI tasks and performs disk output.
4. Parallel model used in post processing generates 2 outputs- an hourly comprehensive of variable files and output at every 10 mins.
5. Perl scripting is used as the master script and Unix crontab initiates the script.



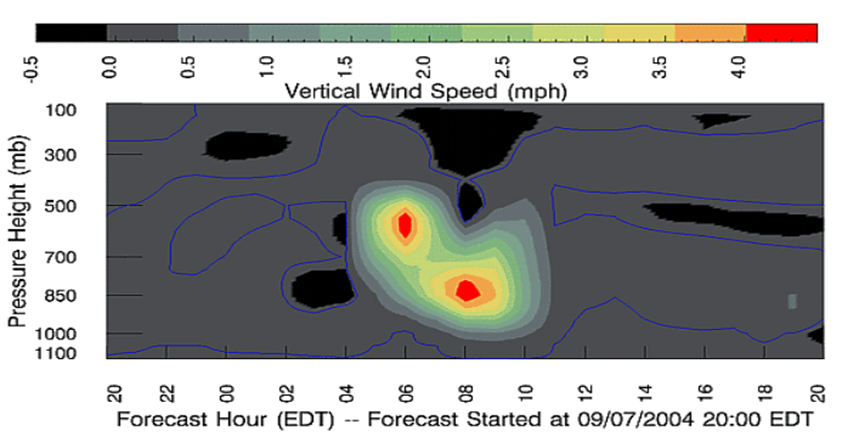


Figure: 05

Since the precipitation is accumulated through the model run, the slope of the curve will be indicative of the predicted rate of precipitation. Therefore, when the slope is zero, it is not raining (or snowing).[1D]

**Weather Decision Technologies**

Weather Decision Technologies, Inc. (WDT), an Esri partner, employs advanced GIS (geographic information system) technology to better organise and analyse this massive amount of collected data. The organisation makes use of ArcGIS Amazon System Image (AMI), which allows it to run many ArGIS server instances on a single machine. ArGIS server facilitates the publication of real-time, dynamic data as a map service for usage by other applications. Users may cycle over the previous 60 minutes of radar data to observe where a storm has been and where it is heading using Amazon Web Services, which helps with dependability, redistribution, and load balancing between servers.

Some other companies that aggregate weather data are [Skymet Weather Services](https://www.cbinsights.com/company/skymet-weather-services) used in India, providing farmers with short-range, medium-range, and seasonal weather forecasts; [Axelspace](https://www.cbinsights.com/company/axelspace), and [Earth Networks](https://www.cbinsights.com/company/earth-networks) to name a few.

## **Industries benefited from big data weather forecast models**

Weather forecasting models are computer programs that can help anticipate future weather at any time from one hour to ten days or even months in advance. Following are some of the industries that benefit from weather forecasting using big data technologies:

**Transportation**

Airlines can plan flight paths to catch a current of air that is fuel efficient and therefore cost effective, while ensuring that turbulence is avoided.

**Mining**

Mining industry uses explosives to blast soil and rock, thus sending acoustic waves, dust and sediments damaging the hearing abilities of workers and also damage to properties. If the wind speed is high, workers are at risk of illness if this dust is inhaled in large quantities. Accurate wind, temperature and lightning forecasts are essential to plan activities accordingly when the weather conditions are favourable.

**Construction**

Using cranes to do the heavy lifting or planning for a concrete placement overnight expecting huge downpour the next day - these are some of the reasons that can be overcome with weather forecasting.

**Retail**

Retail industries rely on seasonal changes to purchase the type of merchandise. If an early spring is expected, customers would be interested in buying more garden essentials and swim-wear than buying winter clothes. If the shelves are not stocked accordingly it may result in a dip in sales. Earlier anticipation of the seasonal change will help retail industries to stock their shelves for appropriate seasonal wear.

**Sports**

Extreme heat, strong winds, heavy rain, and thunderstorms all affect sports events. With millions of spectators watching the event, how big or small, changes in weather will adversely affect the athletes and spectators. Weather prediction comes to rescue to plan international sporting events like the Olympics.

**HealthCare**

We can also predict asthma attacks using weather data. The inhalers that patients use collect data related to temperature and dust particles present in air predicting the occurrence of an asthma attack.

## **Data Governance and ROI**

Data governance solutions provide security and trust of the organisation’s data. Companies manage and control their data using proper models and make data comparisons based on a unique truth point, such as by comparing contact variations to their correct spelling in the postal system database as well as making matches and fusing records from the same entity. Businesses need rules for data cleansing and processing and using tools like IBM data Pak etc. This allows weather forecast companies to monitor data to keep it up to date, document data for compliance, collect and enhance the data. IBM solution is data catalogue. These data catalogues are driven by machine learning algorithms and enable faster collection of metadata and cultivating data assets. The security infrastructure implemented by IBM's latest technology, assured that all the data and privacy are protected. But as much as data governance is important, investing in big data should provide better ROI for businesses. In fact the insights acquired from Big data analysis have steered companies public relation campaigns to better estimate its business strategies. According to statista, revenue from big data is estimated to grow as much as $274.3 billion in 2022. So it is clearly evident that organisations are coming up with algorithms to give much accurate and faster predictions.

## **Challenges and Solutions**

As with any technology using big data also has some challenges. The availability of vast raw information might not be sufficient - a good example might be the availability of data for earthquakes, but there isn't a reliable model for accurate prediction. Also, there is a serious lack of professionals to work with giant data sets in the field. Due to evolving data tools, data scientists are also required to update their knowledge and insufficient understanding can break the game. The Insufficient availability of historical weather data may also cause problems for machine learning. The geographical and temporal resolutions of weather forecast models are determined by computing capacity, which limits the model's ability to reflect small-scale and fast-moving weather phenomena. These are some of the challenges faced in weather forecasting models. However, the above problems may be overcome by using machine learning, using proper data cleaning techniques, prioritising security and intelligent approaches in proper big data architecture that best suits the organisation.

# **Conclusion**

This report was presented to analyse big data in weather forecast modelling and how it impacts industries. Data is like energy. It is essential to prepare the appropriate data in valid format for making decisions and generating useful information. To predict the weather it is mandatory to collect the exact location and time.

In terms of computer technology, the volume of data is rapidly increasing, and there are different definitions of big data. According to one of these notions, big data necessitates the use of new tools, analytics, and technological architectures in order to create high-value sources for organisations and to gather large amounts of hidden information in analytics.

There is a growing appreciation for the value that may be derived from weather-driven mathematical models for forecasting and explaining. Such models necessitate a clear understanding of the distinctions between using actual and forecast data, as well as an effective selection of geographical and temporal resolutions and feature engineering based on domain knowledge that provides key

information about the objects or phenomena under study. Such principles establish the groundwork for building solid models that provide actual insight, rain or shine.

# 

# 

# **Reference**

“The World’s Most Accurate Weather Forecasting System.” *IBM Nordic Blog*, 4 Dec. 2019, https://www.ibm.com/blogs/nordic-msp/the-worlds-most-accurate-weather-forecasting-system/.

Schwartz, Scott W. *Deep Thunder: The Rise of Big Meteorology | Platypus*. https://blog.castac.org/2019/04/deep-thunder-the-rise-of-big-meteorology/. Accessed 30 Mar. 2022.

*IBM100 - Deep Thunder*. 7 Mar. 2012, http://www-03.ibm.com/ibm/history/ibm100/us/en/icons/deepthunder/.

“Will IBM’s Deep Thunder Finally Be the Key to Accurate Weather Forecasting?” *Technology and Operations Management*, https://digital.hbs.edu/platform-rctom/submission/will-ibms-deep-thunder-finally-be-the-key-to-accurate-weather-forecasting/. Accessed 30 Mar. 2022.

Casals, Claudi. “Big Data and the Future in Weather Forecasts.” *Meteorología En Red*, 10 July 2017, https://www.meteorologiaenred.com/en/big-data-meteorology.html.

H. Jain and R. Jain, "Big data in weather forecasting: Applications and challenges," *2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC)*, 2017, pp. 138-142, doi: 10.1109/ICBDACI.2017.8070824.

DataFlair. (2019). *Data Science for Weather Prediction - The Prerequisite to all Natural Disasters*. [online] Available at: https://data-flair.training/blogs/data-science-for-weather-prediction/.

‌Big Data in Weather Forecasting. (n.d.). [online] Available at: https://www.ijert.org/research/big-data-in-weather-forecasting-IJERTCONV8IS13047.pdf.

‌“Big Data and Business Analytics Revenue 2022.” *Statista*, https://www.statista.com/statistics/551501/worldwide-big-data-business-analytics-revenue/. Accessed 30 Mar. 2022.

Anusha, K. and Usha, K. (2017). Big Data Techniques for Efficient Storage and Processing of Weather Data. *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*, [online] 887, pp.2321–9653. Available at: https://www.ijraset.com/fileserve.php?FID=8986.

‌Casals, C. (2017). Big Data and the future in weather forecasts. [online] Meteorología en Red. Available at: https://www.meteorologiaenred.com/en/big-data-meteorologia.html#Como\_funciona\_el\_Big\_Data [Accessed 14 Mar. 2022].

Chouksey, P. and Chauhan, A.S. (2017). A Review of Weather Data Analytics using Big Data. IJARCCE, 6(1), pp.365–368.

‌UKEssays. November 2018. Big Data Frameworks for Weather Prediction Analysis. [online]. Available from:

*Big Data Frameworks for Weather Prediction Analysis*. https://www.ukessays.com/essays/computer-science/big-data-frameworks-for-weather-prediction-analysis.php. Accessed 30 Mar. 2022.

Tyagi, Neelam. *Weather Forecasting: How Does Big Data Analytics Magnify It ? | Analytics Steps*. https://www.analyticssteps.com/blogs/weather-forecasting-how-do-big-data-analytics-magnify-it. Accessed 30 Mar. 2022.

Army, Prep. “What Is Weather Forecasting and Its Importance.” *PrepArmy*, 25 Apr. 2017, https://preparmy.com/featured/weather-forecasting-importance/.

“Are You Using Weather Data to Forecast or to Explain?” *Smith Institute*, 14 Mar. 2018, https://www.smithinst.co.uk/insights/the-use-of-weather-data-in-modelling-and-forecasting/.

UKEssays. November 2018. Big Data Frameworks for Weather Prediction Analysis. [online]. Available from: https://www.ukessays.com/essays/computer-science/big-data-frameworks-for-weather-prediction-analysis.php?vref=1 [Accessed 29 March 2022].

Naga, S., Ch, J., Divija, B., Tharun, K., Reddy, S. and Ahmed, M. (n.d.). *Weather Data Analysis Using Hadoop MapReduce and Spark*. [online] Available at: https://www.xajzkjdx.cn/gallery/311-april2020.pdf.

‌Big Data in Weather Forecasting. (n.d.). [online] Available at: https://www.ijert.org/research/big-data-in-weather-forecasting-IJERTCONV8IS13047.pdf.

‌

J.H. Pae, D.R. Lehmann

Multigeneration innovation diffusion: The impact of intergeneration time

J. Acad. Mark. Sci., 31 (2003), p. 40-45

Technological forecasting at the stock market

Technol. Forecast. Soc. Chang., p. 198-202

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

https://journalofbigdata.springeropen.com/articles/10.1186/s40537-022-00580-9