RESEARCH ARTICLE OPEN ACCESS

Data Structure for Representation of Big Data of Weather Forecasting: A Review

Dhanashri V. Sahasrabuddhe [1], PallaviP.Jamsandekar [2]
Bharati Vidyapeeth Deemed University
Institute of Management and Rural Development Administration, Sangli
Maharashtra - India.

ABSTRACT

The paper takes review for data structures used for big data with respect to weather forecasting. Results and accuracy of Weather forecasting plays a vital role in daily routine of a layman, businesses and their decisions. The process of weather forecasting is developing as the effect of advancement in technology. Right from the realisation of increasing size of data, Weather forecasting was found to be based on big data. The researchers have taken review with the objective to study the current forecasting process and methods, and the need of a data structure is recognized for handling the weather data, which is big in size, used for the process of weather forecasting.

Keywords:-Weather Forecasting, Numerical Weather Prediction, Big Data, Data Structures.

I. INTRODUCTION

The paper takes review of the literature for data structures used for Big Data with respect to a special case of weather forecasting. Today the big data has become a buzz word, and still in developing stage. Weather forecasting, basically the problem of initial value, is considered by researcher as a case of big data, which will help to improve the accuracy of forecasting. For handling this voluminous data require for weather forecasting, there is a need of an efficient data structure.

Through this paper researcher discuss process of weather forecasting, different approaches used for forecasting, review of Big Data and role of Big Data in weather forecasting, review of Data structures used for Big Data as well as weather forecasting. Numerical Weather Prediction (NWP) is the desirable technique for weather forecasting. The data structures available till now has some limitations to apply for weather data, hence researcher plan to design a new data structure which will store the weather data efficiently.

II. WEATHER FORECASTING

Since from the time when the cultivation had started we are interested in knowing about the weather changes. Different methods were developed to forecast the weather changes, some were intuition based while some were scientific. Always we look for accuracy of forecast. This section discusses development of weather forecasting techniques and the Numerical Weather Prediction as a scientific and mathematical technique of weather forecasting.

A. History of Weather Forecasting

Weather plays important role in daily routine of a layman or a businessman, which insists a person to know about the future conditions of weather (1). For weather forecasting different methods are being used, some of which are intuition based, while some has scientific and mathematical base.

The Babylonians (650 BC) predicted the weather from cloud patterns as well as astrology. During 340 BC Aristotle described weather patterns in Meteorologica. Later, the Ophrastus compiled a book on weather forecasting, called the Book of Signs. During 300 BC Chinese and around the same time ancient Indian astronomers developed weather-prediction methods. 904 AD- Ibn Wahshiyya's Nabatean for the purpose of agriculture discussed the weather forecasting. He tried to forecast the future weather conditions by observing the atmospheric changes and signs from the planetary astral alterations. For example signs of rain based on observation of the lunar phases the movement of winds.

Real development in weather forecasting has taken place during 1835. In this year a well known British sailor Admiral Robert FitzRoy created the weather charts, which were generated based on the current weather data collected using electric telegraph from 160km/d area wide. Publically for the first time the daily weather forecast appeared on the Times, dated 1 August 1861 the first ever daily weather forecast. In 1890 the American meteorologist Cleveland Abbe come up with a mathematical approach for forecasting. He was a strong believer that ''meteorology is essentially the application of hydrodynamics and thermodynamics to the atmosphere''. In his paper, "The physical basis of long-range weather forecasting", he has stated that the atmospheric scientists would understand the urge and would model the weather conditions graphically, analytically, or numerically. (2)

Before development of numerical forecasting the "analog approach" was at the process of operational forecasting where weather maps were used to match with current maps (3). In 1922 the possibility of NWP was proposed by Lewis Fry Richardson. Using differential equations Richardson represented the atmospheric motions. When values of the tendencies of various field variables, defined in the set of algebraic equations which represents the atmospheric motions, are provided and the equations are solved the estimated values for the same variables can be calculated.

During his experiment he came to know that there were large number of computations and without a computing machine it was not possible to forecast weather. In 1950 the first computerized weather forecast was performed by a team led by the mathematician John von Neumann and in 1955 practical use of numerical weather prediction began spurred by the development of programmable electronic computers. Presently for forecasting current (short term) as well as future (long term) weather conditions different weather models are being developed and used with the prime objective of accuracy.

B. Basic Process of Weather forecasting

In 1950, when for the first time weather forecasting was found as initial value problem (2), the basic process of weather forecasting was defined by (4) which is carried out in following three steps-

- Collection of observational data-
  The readings of different parameters (temperature, humidity, pressure and airflow) for current weather conditions are collected from different atomized centres. World Meteorological Organization (WMO) is responsible for gathering, plotting and distributing weather data over the world.WMO acts to standardize the instruments as well as the observing practices and timings for collection of data.
- Transmission –
  The observation stations either reports the observed data hourly in METAR reports or six hourly in SYNOP reports.
- Compiling the large observed data –
  Data assimilation technique (5)works as a quality control phase in the forecasting process. The accuracy of the forecast generated depends on the initial data which depicts the past and current state of weather. In this phase the irregularly spaced observations are processed using data assimilation and objective analysis techniques.
- Using Models –
  After the analysis phase the data is ready for the forecasting purpose. The observations are then provided as an initial input to various models (generally computerized) for the further process.

C. Different approaches used for forecasting-

The approaches used in modern weather forecasting include:(6)(4)

- Synoptic Weather Forecasting –
  It is oldest and the first method used for the weather forecasting and was used until the late 1950s. The process of forecasting involves the analysis of synoptic (summary) weather charts, employing several empirical rules.
- Numerical Weather Prediction –
  It is a modern weather forecasting based on the fact that –
  'The changes in the atmosphere obey many known physical principles and the same laws can be used to predict the future state of the atmosphere based on the current conditions.'
  For the forecasting NWP uses a number of highly refined computer models that attempt to mimic the behaviour of the atmosphere.
- Statistical Weather Forecasting Methods -
  In conjunction with NWP, using past weather data statistical methods predict future events. The analog method, one of the statistical approaches, examines past weather records to find ones that come close to duplicating current conditions. The results generated by statistical approach are usually correct 40-50% of the time.
- Short-Range Forecasting Techniques –
  The simplest short-range forecasting techniques, called persistence forecasts, predict that the future weather will be the same as the present conditions. Another technique, often called nowcasting, uses radar and geostationary satellites to quickly forecast severe weather events, such as thunderstorms, tornadoes, hail storms, and microbursts.
- Long Range Forecasting Techniques-

It is an area that relies heavily on statistical averages obtained from past weather events, also referred to as climatic data. Weekly, monthly, and seasonal weather outlooks prepared by the National weather Service are not weather forecasts in the usual sense. They indicate only whether the region will experience near-normal precipitation and temperatures or not.

D. Numerical Weather Prediction (NWP)

The basic idea of NWP is to sample the state of the fluid at a given time and use the equations of fluid dynamics and thermodynamics to estimate the state of the fluid at some time in the future.(7)(8)(9)(10)

But during that time it was not possible to carry out the processing manually. In 1950 with the development of the machines with high computing power it was possible to forecast the weather conditions using NWP. With the development of more powerful computers and better modeling techniques numerical weather prediction has returned to models that are quite similar to Richardson's model and are more accurate.

According to Hamill, Tom the stages of NWP process are as defined as follows (11)–

1. Gather Observations
2. Data Assimilation
3. NWP Models execution
4. Forecast Post processing
5. Issue forecasts, Evaluate

The first attempt towards NWP Forecasting is shown in following table-

Level (km)	LFR	MOD	DFI
1 (11.8)	48.3	48.5	−0.2
2 (7.2)	77.0	76.7	−2.6
3 (4.2)	103.2	102.1	−3.0
4 (2.0)	126.5	124.5	−3.1
Surface	145.1	145.4	−0.9

Table 1:Six-hour changes in pressure (units: hPa/6 h).
[LFR:Richardson; MOD: Model; DFI:Filtered]
Source :*The origins of computer weather prediction and climate modeling.*
Lynch, Peter

The NWP approach to weather forecasting is a technology based approach and since the initial stage of development it is continuously evolving due to (10):

–Introduction of new modeling techniques for better accuracy
–New parameterization schemes
–Continuously evolving computing resources being used for execution of NWP forecasting process

E. Mathematical Approach of NWP

The great American meteorologist Cleveland Abbe recognized that meteorology is essentially the application of hydrodynamics and thermodynamics to the atmosphere, and he identified the system of mathematical equations that govern the evolution of the atmosphere.(3)

In the above table 1, column LFR are the values calculated by Richardson, column MOD are the values calculated by computer model which follows the same equations given by Richardson, while the third column shows Digital Filter Initialization. Though the predictions were unrealistic, it was due to the imbalance in initial data.

To forecast the change in pressure, Richardson used the continuity equation, employing precisely the method that Margules (Max Margules, almost twenty years before Richardson's forecast, pointed to serious problems with Richardson's methodology examined the relationship between the continuity equation (which expresses conservation of mass) and changes in surface pressure) had shown to be seriously problematical. The resulting prediction of pressure change was completely unrealistic.

He went on to speculate that the vertical component of vorticity or rotation rate of the fluid might be a suitable prognostic variable. This was indeed a visionary adumbration of the use of the vorticity equation for the first successful numerical integration in 1950. The continuity equation is an essential component of primitive equation models which are used in the majority of current computer weather prediction systems.

Mathematically Richardson represented process of NWP using different differential equations, of which the basic form can be given as (2) -

$$\frac{\Delta A}{\Delta t} = F(A)$$

Fig2: Basic equation
Source : Numerical Weather Prediction (10)

These equations are known as primitive equations. For different forecasting variables like temperature (T), humidity (q), two horizontal wind components (u and v) and - surface pressure (p) different differential equations are defined, which changes according to model (10).

At that time though, Richardson has given the equations, it was not possible to predict the forecast due to following four reasons-

1. Observations of the three-dimensional structure of the atmosphere were available occasionally

2. Instabilities in numerical algorithms for solving atmospheric equations
3. The balanced nature of atmospheric flow was inadequately understood, and the imbalances arising from observational and analysis errors confounded Richardson's forecast.
4. The massive volume of computation required to advance the numerical solution

The need of large computing power was recognized by Richardson, and after the development of ENIAC electronic device the initiative was taken by John Von Neumann to execute the forecasting atmospheric equations.

F. Process of NWP

The figure given below explains the process of weather forecasting using NWP. The lecture notes on NWP has explained different models used in process.(10)

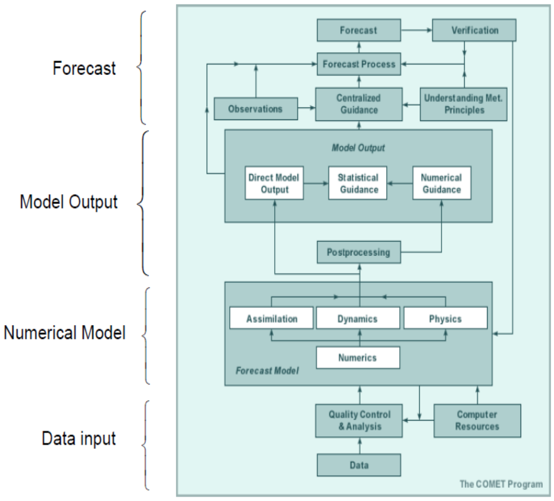

Fig 1: NWP Process

Source : Numerical Weather Prediction (10)

After pre-processing (Data Assimilation) the historical weather data collected from different centres, it is given as input to primitive equations which are executed iteratively.

For different parameters there are equations, which accept and give forecasted values for the same parameters like temperature, humidity etc.

III. BIG DATA AND WEATHER FORECASTING

Big data follows the distributed approach and different sources of big data can be given as SNS, Cloud Computing, Large Processing Technology and Mobile Data. The big data architecture typically divided in three segments Storage, Processing and Analysis.(12)

According to Biswas Ranjit to deal with big data (expanding very fast in 4Vs: Volume, Variety, Velocity and Veracity) he has introduced 4 N's-

- New logical and physical storage structures,
- New heterogeneous data structures,
- New mathematical theories
- New models;

According to him, existing data structures or architecture of existing distributed systems is not capable of handling this speedily increasing and changing data. These problems give rise to need for New Data Structure and New Architecture for distributed system which will have a tremendous extent of mutual compatibility between them, with a facility of unlimited scalability. (13)

Improvements to NWP models have been identified by the National Centers for Environmental Prediction (NCEP) using historical data archives that in turn better protect life and property through improvements of forecasting skill.(14)

Today, NOMADS (NOAA Operational Model Archive and Distribution System) provides access to approximately 650 terabytes of model data arranged in more than 50 million files on tape. Approximately 1.2 million of these files are arranged on disk for immediate access, and a sub-set of those are pre-aggregated by variables in a time-series to promote easier access for users. Some data need to be pre-staged from tape prior to access — a very slow and tedious process. During 2013, NCDC NOMADS serviced more than 850 Terabytes from 116 million requests by 130,000 separate users.

Current technology trends have centered on the use of cloud computing and technologies such as Apache Hadoop as "big data" solutions. Much of this, however, assumes specific use cases and algorithms that are optimized for these types of environments and technologies.

According to Bryson "Weather is the original big data problem". It has been discussed earlier though any approach is followed; weather forecasting is initial value problem. Size of initial data increases, accuracy of forecasting increases. (15)

Nick Wakeman with reference to Hurricane Sandy stated in his blog the importance of Big Data in weather forecasting. With the help of available data, three-days out, forecasters predicted within 10 miles where landfall would occur.

According to author it was possible only because of rapidly growing speed and power of computers, and the ability to collect and analyze data faster and more accurately than before, an even bigger disaster was averted.(16)

According to Nancy Grady the velocity of weather data plays role in the development of economy as a case. This weather data can be used by combining it with other disciplines which can generate new opportunities to businesses. (17)

Weather, air travel, safety, financial, health, agriculture entrepreneurs are leveraging weather and climate data to build previously impossible business. The example of a climate corporate is given by the author who sells bad weather insurance to farmers. (18)

To succeed in enabling the Nation's weather prediction goals, NASA research activities need to be integrated with NOAA and Navy operational requirements. These include:

1) Techniques developed within NASA have to fit into the operational time allowance for real-time weather forecasting. This requires the end-to-end forecast-assimilation system to run in less than 2 hours of wall time.

2) Attributes of NASA algorithms, e.g., resolution, data usage, parameterizations, must be consistent with the state-of-the-art operational systems. Otherwise, the impact of NASA research will be diminished.

The computational requirements for the anticipated forecast-assimilation system for 2010 are currently driven by data volume. A data volume of 1011 observations per day is chosen based on projections of satellite observing systems that are planned to be operational and relevant to weather forecasting. As a baseline, the current horizontal resolution is order 100 km, with 50 vertical levels, and current data usage is order 106 observations per day. The literature insists that the weather data volume should be reduced prior to assimilation. (15)

IV. DATA STRUCTURES USED FOR BIG DATA AND WEATHER FORECASTING

Pedja Bogdanovich et.al. has proposed the ATree Data Structure for Large datasets, which stores abstract of data, which is less in magnitude and capable of satisfying most of the queries of user as well as original data. The ATree is extension of region quadtree data structure. Regular databases are incapable of handling the data generated by satellites and super computer simulations and other monitoring devices, and had proposed ATree data structure for handling such large datasets. Global Climate Model is used for the research for the reason firstly it is MDD (Raster Data) and climate model data is large in size and didn't have easy access to it.

The Data structure had been explained under four headings- Subdivision Scheme, Location Codes, Transformation Function, Sub division Criteria

Algorithms for ATree – Two types of build algorithms are used with ATree. The role of build algorithm, which accepts raw data in standard formats like DRS, NetCDF, HDF, and builds ATree according to the parameters which are shape, location code scheme, transformation function, subdivision criteria and metadata function. The ATree is build using two different approaches – Top Down Approach and Bottom Up Approach.

Main Features are given as-

- Data is partitioned, and as per the need data is chosen
- Data partitions are organized on archival storage according to the expected usage of the data.
- The data structure allows for hierarchical compression techniques to be applied (both lossy and non-lossy).
- The ATree data structure can store multi resolution data.
- The data structure allows for non-spatial information to be included, which facilitates quick retrieval of data based on non-spatial attributes (e.g., crop types).(19)

Huuhka et al. has taken review of the data structures used by powerful search engine 'Google' which handles large datasets. The search engine uses data structures optimized for large data sets and adding more servers to the system easily increases its performance of answering queries. Google's PageRank algorithm uses the link structure of the WWW to calculate the qualities of web pages. Using the link structure in the ordering of the search results makes it harder for anyone to manipulate Google's search results but not impossible. Google also has some means to prevent manipulation and maintain the quality of the search results.

Data structures used by the Google are- GFS (Google File System, store files in distributed manner which breaks file in different chunks stored on different servers in three copies on three different servers for reliability), Repository (Stores addresses of all web pages used while searching, makes use of stack), Hit Lists (hit corresponds to the words used in website, stores 16 bit information about word), Forward Index (The forward index consists of barrels and each barrel has its own range of wordIDs, A record in a barrel consists of a docID followed by wordIDs with their hit lists.), Inverted Index (used to find all documents that contain a given word, and is made from the forward index by sorting the contents of the barrels in the order of wordIDs.) are major data structures used by the Google.

The PageRank is used to calculate a quality rank among nodes in a graph, or in the case of WWW among web pages in a collection.

A simplified equation of PageRank is

$$PR(u) = c\sum_{v \in B_u} \frac{PR(v)}{N_v} \quad (1)$$

where PR(u) is the PageRank of a page u, Bu is a set of pages pointing to u, Nv is the number of links on page v and c is a normalization factor. Factor c is used so that the total sum of rank on the graph is constant.(20)

Michael A. Bender et.al. has stated there is a need of structure for data which helps to carry out different operations on data efficiently (space and time)(21). The DBMSs has to perform different operations on data like storing, indexing and querying, and for the purpose it uses different data structures like B+ trees, hashing etc.(22). With increasing size of data the data structures used by the conventional DBMSs has to be modified, which has been suggested and used by TokuDB. The table below shows necessity of data structures for handling large datasets.

Better data structures may be a luxury now, but they will be essential by the decade's end.

Example: Time to fill a disk in 1973, 2010, 2022.

Year	Size	Band width	Access Time	Time to log data on disk	Time to fill disk using a B-tree (row size1 K)	Time to fill using Fractal tree* (row size 1K)
1973	35 MB	835 KB/s	25ms	39s	975s	200s
2010	03T B	150 MB/s	10ms	5.5h	347d	33h
2022	220 TB	1.05 GB/s	10ms	2.4d	70y	23.3d

Table2: Time consumed for accessing data over the period
Source :Data Structures and Algorithms for Big DBs

Log Structured Merge (LSM) have become popular and many databases dealing with big data like Accumulo, Bigtable, bLSM, Cassandra, HBase, Hypertable, LevelDB are LSM trees (or borrow ideas).Looking in all those trees is expensive, but can be improved by

• Caching is warm, small trees are cached

• Bloom filters, avoid point queries for elements that are not in a particular B-tree and

• Fractional cascading, helps to reduce the cost in each tree

Instead of avoiding searches in trees, we can use a technique called fractional cascading to reduce the cost of searching each B-tree to O(1).With forward pointers(forwarding pointers to the first tree, we can jump straight to the node in the second tree, to find c Remove the redundant ones.) and ghosts (need a forwarding pointer for every block in the next tree, even if there are no corresponding pointers in this tree. Add ghosts.), LSM trees require only one I/O per tree, and point queries cost only $O(\log_R N)$This data structure no longer uses the internal nodes of the B-trees, and each of the trees can be implemented by an array. The problem with big data is microdata. Sometimes the right read optimization is a write-optimization. As data becomes bigger, the asymptotics become more important (21).

Yadav, Chanchal has reviewed various algorithms developed and used during 1994-2013 for handling large data sets. These algorithms define various structures and methods implemented to handle Big Data, also in the paper various tool are listed were developed for analyzing it. (12)

The author had given different algorithms used which has used different data structures like – R tree, R*tree, Nearest neighbor search, decision tree learning, GA tree (decision tree+ GA algorithm) , hierarchical neural network etc. These different algorithms were defined for different purposes to handle the large data sets. Every algorithm has its own efficiency and application. In comparative study it was found that recently GA Tree and Hierarchical Neural Network were found to be more efficient.

Manda Sai Divya et.al. had taken review of the **ElasticSearch** application which is horizontally-scalable, distributed database built on Apache's Lucene (is an open-source Java library for text search) that delivers a full-featured search experience across terabytes of data with a simple yet powerful API. The application is preferred for searching data in applications which are easy to set up, scalable and built for cloud. Basic features of ElasticSearch are given as – REST API, Key Value Store, Multi Tenancy and Mapping, Sharding and Replication. It also provides all functionalities which enables to build frontend application on the top of it as well as supports complex data models which supports applications. It also supports multiple indices (databases) and multiple mappings (tables) per index.(23)

Like Google as discussed above, ElasticSearch also uses Inverted index for creating indexes. While for querying it uses JSON syntax.

Biswas Ranjit has come up with two new data structures namely **r-train** and **r-atrain**, to be used for the storage of distributed big data. According to the author, for storage and process the big data there is a need of new architecture as well as new scalable data structure which can cope up with increasing data size. Author has suggestedtwo data structures viz. r-train and r-atrain which were used with homogeneous and heterogeneous data.

Along with these data structures author has proposed unitier and multitier distributed architecture called Atrain Distributed Architecture(ADS) useful for processing big data.(13)

Naseer, Aish aet. al. has recognized the need of the enterprises whose business has spread over the world and maintains large data of their customers which is big in size, needs to utilize this data to link them with customers to provide support and services. The project was carried out at the Fujitsu Laboratories of Europe Limited. The main focus was on the data itself –and sharing of this data across the enterprise, thus unlocking hidden information that might be of interest to the enterprises.(24)

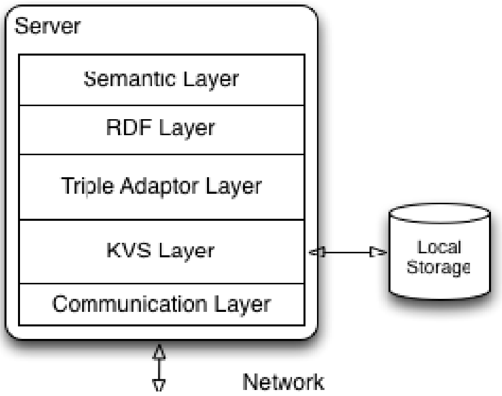

Fig : Architecture of BigGraph
Source : *Enterprise BigGraph.* (24)

BigGraph is based on RDF (Resource Description Framework), which connects different and isolated data silos and transform into a connected data graph. Due to the fact that RDF could be used as a schema-less data representation format, it enables BigGraph to satisfy the important requirements of flexibility and dynamic extensibility such as adding new data easily and coping with constantly changing data,. A schema-less data representation format enables any data to be added and linked together without the need to constantly make changes to any schema. From the architectural point of view, BigGraph is a graph-based semantic platform over a distributed, ordered Key-Value-Store, KVS , where data is distributed across multiple machines. The distributed architecture enables to achieve a high level of scalability in terms of the data stored and an efficient approach of querying the data. The architecture of a single server is illustrated in Figure 3. Each server is composed by a Semantic Layer, an RDF Layer, a Triple Adaptor Layer, a KVS Layer and a Communication Layer. The Semantic Layer interacts with end-users and relays user queries to the RDF processor. The semantic layer explicitly captures meaningful relationships among data, which are represented at conceptual and model level (as groups of related RDF triples). In this layer, semantic operations such as data reconciliation and social analysis are performed.

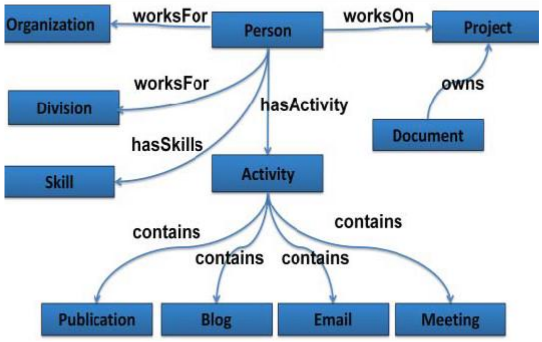

Fig: Ontology of BigGraph
Source : *Enterprise BigGraph.* (24)

There are three important aspects that contribute to the success of BigGraph- the use of Linked Data, the graph technologies, and the social analysis capability. By using Linked Data, BigGraph provides a unique infrastructure that allows storing and managing data, facilitates data integration, and makes the enterprise data more accessible and easier to use externally. BigGraph has the ability to create dynamically the structure of enterprise data, without having prior knowledge, and accommodates rapid changes. BigGraph offers more than a graph database by providing the whole infrastructure for managing, interlinking information, analyzing, and visualizing very large complex datasets.

GRIB (**GRIdded Binary** or **General Regularly-distributed Information in Binary form**) is a concise data format commonly used in meteorology to store historical and forecast weather data. It is standardized by the World Meteorological Organization's (WMO) Commission for Basic Systems, known under number GRIB FM 92-IX, described in WMO Manual on Codes No.306. Currently there are three versions of GRIB.(25) Version 0 was used to a limited extent by projects such as TOGA, and is no longer in operational use. The first edition (current sub-version is 2) is used operationally worldwide by most meteorological centers, for NWP output. A newer generation has been introduced, known as GRIB second edition, and data is slowly changing over to this format. Some of the second-generation GRIB are used for derived product distributed in Eumetcast of Meteosat Second Generation.(26)

Most GRIB files are actually a collection of individual self-containing records, and can be appended to each other or broken down easily.(27)

GRIB Structure for a GRIB record is composed of 6 GRIB sections for one parameter. Section 4 from record contains the data itself, while the other sections give the information required to read the GRIB record.

It is an extremely useful and cost effective tool to store and access weather data. GRIB information is without human assessment, hence no quality control nor do any guarantee that the data are correct.

The another format used to store weather data is *Binary Universal Form for the Representation of meteorological data* (BUFR) file format is widely used even in satellite meteorology. The format is used mostly for satellite sounder data, like NOAA AMSU and MHS and Metop IASI, which traditionally have been the primary satellite data going into the (NWP) models.(28)

There are some analog methods used for Weather Forecasting. M.Sudha et.al.have proposed model estimates the enhancement achieved in spatial reduction and classifier accuracy using Rough Set Attribute Reduction Technique (RSART) and data mining methods.(29)

In paper L. Al-Matarneh et.al. the researchers have used time series method for forecasting.(30)

Wang, Qinhas proposed forecasting model using data mining method. For Data Mining he has used Rough Sets. (31)

The paper by Abhishek Agrawal et.al. utilizes Artificial Neural Network (ANN) simulated in MATLAB to predict two important weather parameters i.e. maximum and minimum temperature. The model has been trained using past 60 years of data (1901-1960) and tested over 40 years to forecast maximum and minimum temperature.(32)

V. CONCLUSION

The above sections has discussed and reviewed data structures used for representing big data for weather forecasting. Though the weather forecasting is evolving since historical time, it lacks accuracy for long run forecast. Big data platforms may help to improve the accuracy for long run forecasting. But due to lack of well defined data structures which will support the historical weather data, the accuracy and duration matters. In the future work the researcher will try to represent the big size weather data in memory to improve the efficiency and accuracy of weather forecasting.

REFERENCES

[1] The Birth of the Weather Forecast. *Magazine.* s.l. : BBC , April 30, 2015.

[2] *The origins of computer weather prediction and climate modeling.* **Lynch, Peter.** 2007, Journal of Computational Physics, pp. 3431–3444.

[3] **Lynch, Peter.** From Richardson to early numerical weather prediction. October 27, 2010.

[4] **Schultz, Dr. R. B.** Weather Analysis and Forecasting.

[5] Weather Forecasting. *www.wikipedia.com.* [Online] [Cited: July 16, 2015.]

[6] Methods Of Weather Forecasting. *http:\\www.ww2010.com.* [Online] University of llinois. [Cited: February 15, 2014.]

[7] https://en.wikipedia.org/wiki/Numerical_weather_prediction. *https://en.wikipedia.org.* [Online] [Cited: July 16, 2015.]

[8] **Saulo, Juan Ruiz and Celeste.** Numerical Weather Prediction. *Trainning Workshop on Nowcasting Techniques.* 2013.

[9] *Anticipated Advances In Numerical Weather Prediction (NWP), And The Growing Technology Gap In Weather Forecasting.* Geneva 2, Switzerland : World Meteorological Organization.

[10] fall03Pt7.pdf. *http://rams.atmos.colostate.edu.* [Online] November 06, 2003. [Cited: May 06, 2014.] http://rams.atmos.colostate.edu/at540/fall03/fall03Pt7.pdf.

[11] **Hamill, Tom.** Introduction to Numerical Weather Prediction and Ensemble Weather Forecasting. *Climate Diagnostic Center.*

[12] *Algorithm and Approaches to Handle Large Data - A Survey.* **Yadav, Chanchal.** June 2013, International Journal of Computer Science and Network, Volume 2, Issue 3,, pp. 37-41.

[13] *Processing of Heterogeneous Big Data in an Atrain Distributed System (ADS) Using the Heterogeneous Data Structure r-Atrain.* **Biswas, Ranjit.** 2014, Int. Journal of Computing and Optimization, pp. 17 - 45.

[14] *Improving NWP Models and data access Latencies.* **Glenn K. Rutledge, Daniel J. Crichton, Jordan Alpert.** 2014, IEEE Oceanic Engineering Society.

[15] **Bryson Koeler, CIO Weather Channel,.** [Online] 2013. http://www.nasdaq.com/article/big-data-delivers-fewer-hunchesmore-facts-to-weather-channel-20130211-00763#.UUxqs1s4VIh.

[16] **Nick Wakeman, Washington Technology.** http://washingtontechnology.com/blogs/editorsnotebook/2012/11/sandy-big-data-opportunity.aspx. [Online] 2012.

[17] **Nancy Grady, David Green, Troy Anselmo.** Big Data Effects on Weather and Climate, Informal Discussions on The New Economics. s.l. : SAIC, 2013.

[18] *Report from the Earth Science Enterprise Computational Technology Requirements Workshop*. Washington, DC : NASA, April 30 - May 1, 2002.

[19] *The ATree: A Data Structure to Support Large Scientific Databases.* **Pedja Bogdanovich, Hanan Samet.** 1999, Springer-Verlag Berlin Heidelberg, pp. 237-248.

[20] *Google: data structures and algorithms.* **Huuhka, Petteri.**

[21] **Michael A. Bender, Bradley C. Kuszmaul.** Data Structures and Algorithms for Big Databases. s.l. : Stony Brook, MIT, Tokutek, 2012.

[22] **Hammer, Joachim.** Data Structures for Databases. 2001, pp. 60-1 to 60-20.

[23] *ElasticSearch- An advanced and quick search technique to handle voluminous data*. **Manda Sai Divya, Shiv Kumar Goyal.** June-2013 , COMPUSOFT, An international journal of advanced computer technology, 2 (6), (Volume-II, Issue-VI), pp. 171-175.

[24] *Enterprise BigGraph*. **Naseer, Aisha.** s.l. : IEEE, 2013. 46th Hawaii International Conference on System Sciences. pp. 1005-1014.

[25] What is GRIB-EnvironmentCanada.html. *Government Of Canada*. [Online] August 11, 2015. [Cited: 09 28, 2015.] https://weather.gc.ca/grib/what_is_grib_e.html.

[26] grib. *www.wikipedia.com.* [Online] April 18, 2013. [Cited: February 06, 2014.] "http://en.wikipedia.org/w/index.php?title=GRIB&oldid=551055204".

[27] BADC_GRIB_Documentation.html. [Online] October 24, 2015. [Cited: September 28, 2015.]

[28] WMO_file_formats.html. *pytroll*. [Online] 2011-14. [Cited: sept 28, 2015.]

[29] *Rainfall Forecast Analysis using Rough Set Attribute Reductionand data mining Methods.* **M.Sudha, B. Valamarthi.** 2014, Agris on-line Papers in Economics and Informatics Volume IV, pp. 145-154.

[30] *Development of Temperature-based Weather Forecasting Models Using Neural Networks and Fuzzy Logic.* **L. Al-Matarneh, A. Sheta, S. Bani-Ahmad, J.Alshaer and I.Al-oqily.** 2014, International Journal of Multimedia an d Ubiquitous Engineering Vol. 9, No. 12, pp. 343-366.

[31] *Cumulonimbus Forecasting Based on Rough Set and Artificial Immune Algorithm.* **Wang, Qin.** s.l. : IEEE, 2010. 2010 Sixth International Conference on Natural Computation. pp. 2856-2860.

[32] *AN APPLICATION OF TIME SERIES ANALYSIS FOR WEATHER FORECASTING.* **Abhishek Agrawal, Vikas Kumar,Ashish Pandey,Imran Khan.** 2012, International Journal of Engineering Research and Applications (IJERA) Vol. 2, Issue 2, pp. 974-98