Journal of Information and Operations Management
ISSN: 0976–7754 & E-ISSN: 0976–7762 , Volume 3, Issue 1, 2012, pp-107-110
Available online at http://www.bioinfo.in/contents.php?id=55

CONCEPTUAL MODEL FOR DEVELOPING METEOROLOGICAL DATA WAREHOUSE IN UTTARAKHAND- A REVIEW

PRITI DIMRI* AND HARSHAL GUNWANT

CSED, Pauri, Garhwal (Uttarakhand), India
*Corresponding Author: Email- [1]pdimri1@gmail.com, [2]harshal.gunwant@gmail.com

Received: December 12, 2011; Accepted: January 15, 2012

Abstract- Data warehouse is a new generation Decision Support System (DSS) tool. Data warehouse technology has grown up to voluminous data having there size in terabytes range or higher; data is stored from different meteorological stations situated in uttarakhand, analyzed or mined and kept in records for future references as well. The purpose of this paper is to develop a conceptual model for data warehouse technology in the meteorological research area starting with uttarakhand. Natural disasters and calamities throw up major challenges and landslides have become of common occurrence in the region, repeatedly taking a heavy toll of life and property. Uttarakhand could be an area of intense research, resulting in the development of many new and advanced systems which could be helpful in early warning, forecasting, and mitigating the impact of natural disasters. Efficient data storage and manipulation is a prerequisite in the meteorological and climatology domain.
Keywords- Meteorological data warehousing, meteorological data report, On-Line Analysis processing, Data Mining

Citation: Priti Dimri and Harshal Gunwant (2012) Conceptual Model for Developing Meteorological Data Warehouse in Uttarakhand- A Review. Journal of Information and Operations Management ISSN: 0976–7754 & E-ISSN: 0976–7762, Volume 3, Issue 1, pp-107-110.

Introduction

Barry Devlin, IBM Consultant on practitioner's viewpoint defines a data warehouse simply as a single, complete, and consistent store of data obtained from a variety of sources and made available to end users in a way they can understand and use it in a Business Context

The application of data warehouse technology in the domain of meteorology/Climatology data can be advantageous for manipulating large quantities of manual, digital and sensor data, performing statistical analysis and extracting meaningful trends and patterns. During the past years relational database management systems have been gradually introduced in many meteorological organizations to substitute proprietary file-based application storage concepts. The use of meteorological data is mainly twofold:

Weather forecasting- where quick access to actual data is important, and Climatology- where flexible access to high quality information about past weather is important .As per the report of Germany's National meteorological services Climate change has taken precedence as an international priority involving extremely complex political and socioeconomic challenges. Efficient data storage and manipulation is a prerequisite in the meteorological and climatology domain. This paper will articulate the benefits that are derived from data warehousing today in meteorological field. Using On-Line Analysis Processing (OLAP) and by generating multidimensional report, we get the relevant data for real-time analytics and weather predictions.

Climatic Elements for Meteorological Observations and there Instruments

Meteorological calculations are recorded as manual, digital and via satellites. The Table 1 describes Meteorological Instruments and the parameters they measure:

Automated weather systems (AWS) and meteorological instruments with no access to land lines for power or communication, the system was set up with a solar panel for power and a GPRS modem for wireless communication [1]. In recent years, with the development of the digital satellite cloud images, the quantitative analysis on satellite cloud images has become an important research direction to some of the meteorological researchers [2] [3], whereas Profilers, and Storm lightening locater are new generation tools that are yet to arrive in uttarakhand. GPS Meteorology is a

body of science and technology which makes use of the Global Positioning System (GPS) for active remote sensing of the Earth atmosphere [4].

Table 1- Meteorological instruments

Instruments	Parameter Measured
Standard Raingauge	Rainfall
Automatic Raingauge	Continuous record of rainfall, storm
Cupcounter Anemometer	Windrun
Campbell, stoke sunshine recorder	Sunshine hours
Evaporation Pan	Evaporation
Stephenson Screen	Housing for instruments
Dry & Wet bulb Thermometer	Dry & Wet bulb temperature
Thermohydrograph	Temperature & Humidity
Soil Thermometers	Soil temperature at different Depths
Wind Vane	Wind Direction

Building Meteorological Data Warehouse from Meteorological Data Observed at Forest Research Institute (FRI) Dehradun

Data acquisition/collection

The weather data used for the data warehousing application described in this paper was acquired at meteorological observatory FRI, Dehradun. The area is situated between $30^{o}19'55"$ and $30^{o}21'16"$ North latitude and $70^{o}58'40"$ and $77^{o}1'$ East longitude and 640.08 meters from the main sea level. Systematic Climatic elements observations on daily basis are taken here, although compilation of temperature and rainfall data was started in 1931, while wind speed, wind direction, sunshine hours, dew condensation etc. are being compiled and published annually since 1967[5]. The weather data is then copied to Excel spreadsheets and archived on daily basis as well as monthly basis to ease data identification and manipulation.

Data cleaning/scrubbing

Data warehouses requires and provide extensive support for data cleaning as it is responsible for loading huge amount of data and some weather sources may contain noisy data (random error), inconsistencies, duplicated values etc., that should be removed to avoid wrong conclusions and weather predictions. Cross checking and discrepancy detection may be useful [6].

T*able 2- Meteorological data of Dehradun (uttarakhand)*

Month	Rainfall (mm)	Relative Humidity %	Temperature		
			Max	Min	Ave.
Jan	46.9	91	19.3	3.6	10.9
Feb	54.9	83	22.4	5.6	13.3
Mar	52.4	69	26.2	9.1	17.5
Apr	21.2	53	32	13.3	22.7
May	54.2	49	35.3	16.8	25.4
Jun	230.2	65	34.4	29.4	27.1
Jul	630.7	86	30.5	22.6	25.1
Aug	627.4	89	29.7	22.3	25.3
Sep	261.4	83	29.8	19.7	24.2
Oct	32	74	28.5	13.3	20.5
Nov	10.9	82	24.8	7.6	15.7
Dec	2.8	89	21.9	4	12
Average Annual	2051.4	76	27.8	13.3	20

Discrepancies raised includes data decays (outdated addresses), human error on data entry, errors in weather instrumentation. Data integrity is maintained after removing such bugs/errors.

Data extraction

Only few attributes from the operational database out of the several weather parameters are expected to be useful in decision making thus they are extracted for the experimentation purpose and bringing it into the data warehouse.

Extraction process includes files, tables to be accessed, selected fields to be extracted, format of target and resulting database and schedule to repeat extraction process [7].

It depicts the average of Rainfall, Relative humidity, Maximum, Minimum and average temperatures of last ten years in Dehradun, Uttarakhand

Data warehousing provides enterprise with memory while data mining provides the enterprise the intelligence. Weather data mining is a form of data mining concerned with finding hidden patterns inside largely available meteorological data, so that the information retrieved can be transformed into usable knowledge. Meteorology is one of the domains, where data mining can improve the productivity of its analysts tremendously by transforming their voluminous, unmanageable and prone to ignorance information into usable pieces of knowledge. Following tables describes the mining the above data using k-means algorithm, dividing data into clusters for easy manipulations and finding hidden patterns and forecasting related information in an easy retrievable form.

Following tables depicts setting of parameters for this algorithm in our software

Attribute	Target	Input	Illustrative	Attribute standardization	
				Src att	New att
Month	-	-	-		
Rainfall(mm)	-	yes	-	Rainfall(mm)	std_Rainfall(mm)_1
Relative Humidity(%)	-	yes	-	Relative Humidity(%)	std_Relative Humidity(%)_1
Max temp	-	yes	-	Max temp	std_Max temp_1
Min temp	-	yes	-	Min temp	std_Min temp_1
Ave. temp	-	yes	-	Ave. temp	std_Ave. temp _1

Fig. 1- Attribute and standardizing them for applying K-means

A variety of data mining tool and techniques are available in the industry, but their use is limited for meteorologic data. It is a methodology designed to perform knowledge-discovery expeditions over the database data with minimal end-user intervention Weather forecasters may use many mining methods like classification, constructing decision tree, artificial neural networks, genetic algorithms, clustering, etc. for predicting, comparing, detecting weather patterns irrespective of its format that may vary from spatial databases to flat files or to semi structured repositories such as WWW. Clustering analysis is one of the main analytical methods in data mining. K-means is the most popular and partition based clustering algorithm. But it is computationally expensive and the quality of resulting clusters heavily depends on the selection of initial centroid and the dimension of the data. K-means [8] is an iterative clustering algorithm in which items are moved among sets of clusters until the desired set is reached. The cluster means

$K_i = \{t_{i1}, t_{i2}, t_{13} \ldots\ldots t_{im}\}$ is defined as:

$$m_i = \frac{1}{m} \sum_{j=1}^{m} t_{ij}$$

K-Means parameters	
Clusters	2
Max Iteration	10
Trials	5
Distance normalization	none
Average computation	McQueen
Seed random generator	Standard

Fig. 2- Passing K-means parameters

Results

Description of "Cluster_KMeans_1"

Cluster_KMeans_1=c_kmeans_1				Cluster_KMeans_1=c_kmeans_2			
Examples			[58.3 %] 7	Examples			[41.7 %] 5
Att - Desc	Test value	Group	Overral	Att - Desc	Test value	Group	Overral
Continuous attributes : Mean (StdDev)				Continuous attributes : Mean (StdDev)			
Ave. temp	2.92	24.33 (2.15)	19.98 (5.84)	Relative Humidity(%)	1.32	82.80 (8.61)	76.08 (14.23)
Max temp	2.76	31.46 (2.56)	27.90 (5.05)	Rainfall(mm)	-1.65	33.58 (24.74)	168.75 (230.32)
Min temp	2.68	19.63 (5.77)	13.94 (8.34)	Min temp	-2.68	5.98 (2.35)	13.94 (8.34)
Rainfall(mm)	1.65	265.30 (265.98)	168.75 (230.32)	Max temp	-2.76	22.92 (2.68)	27.90 (5.05)
Relative Humidity(%)	-1.32	71.29 (16.05)	76.08 (14.23)	Ave. temp	-2.92	13.88 (2.70)	19.98 (5.84)
Discrete attributes : [Recall] Accuracy				Discrete attributes : [Recall] Accuracy			
Month =Aug	0.85	[100.0 %] 14.3 %	8.3 %	Month =Dec	1.18	[100.0 %] 20.0 %	8.3 %
Month =Sep	0.85	[100.0 %] 14.3 %	8.3 %	Month =Nov	1.18	[100.0 %] 20.0 %	8.3 %
Month =Oct	0.85	[100.0 %] 14.3 %	8.3 %	Month =Jan	1.18	[100.0 %] 20.0 %	8.3 %
Month =May	0.85	[100.0 %] 14.3 %	8.3 %	Month =Feb	1.18	[100.0 %] 20.0 %	8.3 %
Month =Apr	0.85	[100.0 %] 14.3 %	8.3 %	Month =Mar	1.18	[100.0 %] 20.0 %	8.3 %
Month =Jul	0.85	[100.0 %] 14.3 %	8.3 %	Month =Sep	-0.85	[0.0 %] 0.0 %	8.3 %
Month =Jun	0.85	[100.0 %] 14.3 %	8.3 %	Month =Oct	-0.85	[0.0 %] 0.0 %	8.3 %
Month =Mar	-1.18	[0.0 %] 0.0 %	8.3 %	Month =May	-0.85	[0.0 %] 0.0 %	8.3 %
Month =Jan	-1.18	[0.0 %] 0.0 %	8.3 %	Month =Apr	-0.85	[0.0 %] 0.0 %	8.3 %
Month =Feb	-1.18	[0.0 %] 0.0 %	8.3 %	Month =Jun	-0.85	[0.0 %] 0.0 %	8.3 %
Month =Nov	-1.18	[0.0 %] 0.0 %	8.3 %	Month =Aug	-0.85	[0.0 %] 0.0 %	8.3 %
Month =Dec	-1.18	[0.0 %] 0.0 %	8.3 %	Month =Jul	-0.85	[0.0 %] 0.0 %	8.3 %

Fig. 3- Dividing data into two clusters for fast evaluation

Data transformation

Data transformation [7] includes

- Character sets must be converted ASCII to EBCDIC, or vice versa.
- Mixed-case text may have to be converted to all uppercase consistency.
- Numerical data, in formats from fixed decimal to floating-point binary, may have to be converted to a consistent data type.
- Time dimensions must be converted into a common representation in data warehouse system.
- Measurements have to be converted in accordance to time metric, zone, unit etc.

Life Cycle of Meteorological/ Climatology Data

The life cycle of meteorological/climatology is shown schematically in the figure below, "The large rectangle area denotes the parts of the process that are covered by the data warehouse system". The life cycle of meteorological/climatology is shown schematically in the figure below, "The large rectangle area denotes the parts of the process that are covered by the data warehouse system". Data are fetched from various observatories or from some reliable measuring sources located in uttarakhand. The raw manual/digital or sensor data are then transformed to a common format using extractor/ monitor classes. Data from various sites is collected by a polling system.

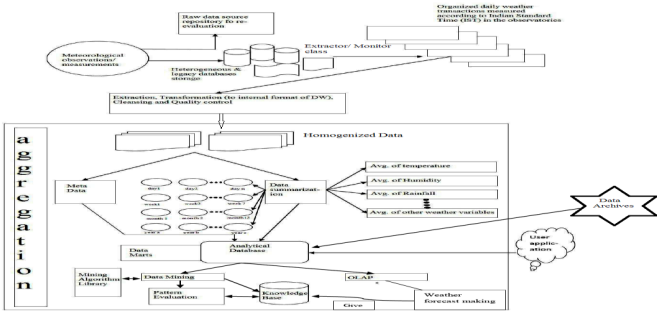

Fig. 4- Our Conceptual model for the Meteorological warehousing system

Like Indian meteorological department pune, keeps track of data observed from various laboratories included that are established in various regions in uttarakhand, calibration constants carry out the transformation thereafter. The raw data repository is also maintained for the 'reevaluation' if anything goes wrong. The organized data obtained from different polling systems are integrated and loaded into the data warehouse system. The core of this transformation step is the 'data cleansing. Cleansing of data removes the redundant, blank and duplicated values that are being added to our storage file. Cleansing of meteorological data may affect also historical data that are used for error detection and to fetch some other additional information. After stepping into the above predefined steps homogeneous series of data is obtained. For faster access and enhancing the knowledge of forecasters the data is thoroughly maintained via data summarization step that could ease in accumulation of new knowledge, predicting weather patterns and faster data comparisons with historical perspectives of data as well. The Metadata corresponds of keeping track of station IDs with their present and historical context, Meteorological Instruments details used by the observatories etc. The information from Analytical databases reinforced by Data Mining and OLAP tools. The main functionality of data mining model design tool is to select a subset or samples from the data warehouse, analyze them using some pre defined algorithms and generate interactive mathematical model [9].

Schema Generation

Data warehousing generates three schemas-snowflake, snowflake and fact constellation and various variants of these three. We are considering Snowflake schema here for our meteorological database in uttarakhand. Meteorological data warehouse contains varieties of meteorological data[10].

As in the figure below DAILYINFO represents the fact table and consists of meteorological data values including the average of temperature, humidity, wind speed, bright sunshine hours etc., while DIMSTATION and DAILYINFO represents dimension tables.

Large fact table creation is demonstrated, which subsequently allows for the development of meaningful queries and cross tab analysis utilizing pivot tables. Fact table sizes of one million records can be iteratively developed and quickly imported into databases such as Microsoft Access or MySQL, while dimension tables contains particular information such as keeping records of meteorologi-

cal stations, time zones etc. and whose coding can be illustrated as:

define cube dailyinfo_snowflake [TimeID, StationID, MeasurementID]
define dimension TimeID as (day,week,month,year)
define dimension StationID as (longitude, latitude, zone, city (city, state)
define dimension MeasurementID as (Thermohydrograph (Temperature (Maximum, Minimum, Average), Humidity)

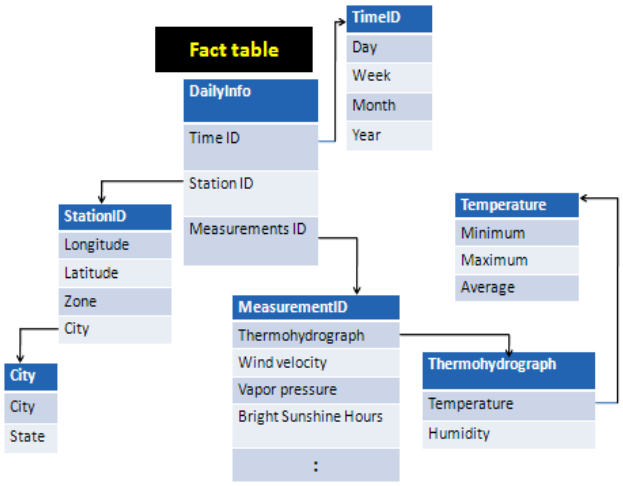

Fig. 5- The fact table and the dimension tables in snowflake Schema

OLAP

OLAP [11] is a category of software technology that enables analysts, managers and executives to gain insight into data through fast, consistent, interactive access to a wide variety of possible views of information that has been transformed from raw data to reflect the real dimensionality of the enterprise as understood b y the user. OLAP queries include ROLL UP that summarizes data along a dimension hierarchy, if we have temperature data per city it can be aggregated to location to obtain sales per state. SLICE and DICE it is beneficial in selection and projection with decreased number of dimensions, Humidity measure of last 3 months. RANKING query deals with selection of first n elements (e.g. select 5 heavy rainy days). PIVOT query deals with re-orientation of cube for cross tabulation Pivot function allows meteorological data observed to view multi-dimensionally from different angles while slicing and dicing may be useful in abstracting particular data like humidity of last three years, heavy rainfall occurrences and temperature transition in uttarakhand via meteorological database.

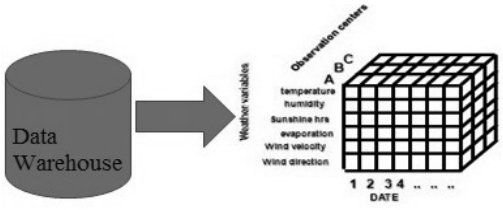

Fig. 6- Data warehouse for multidimensional analysis

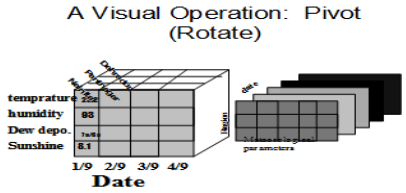

Fig. 7- Applying pivot (rotate) function on meteorological data

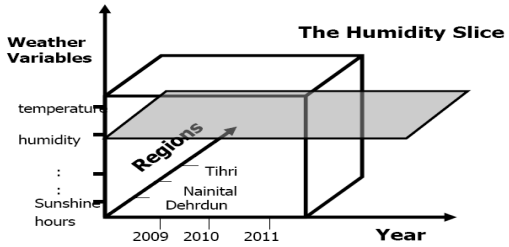

Fig. 8- Applying slicing function on meteorological data

Conclusion

In this paper we prepared a conceptual model of meteorological data warehouse using snowflake schema and demonstrated OLAP queries to analyze multi-dimensional data.

Acknowledgement

Dr. (Mrs.) laxmi rawat, Scientist-F, Head Ecology and Environment Division, Forest Research Institute for providing knowledge of Meteorological calculations and Databases.

References

[1] Zhou Joe and Zhang Jason *Beijing Techno Solutions, Campbell Scientific systems monitor environmental conditions and water quality*.
[2] Qina Kun, Xua Min, Dub Yi, Yuea Shuying (2008) *Remote Sensing and Spatial Information Sciences.* Vol. XXXVII. Part B2.
[3] Wang Y.L., Zhang R., Sun Z.B., Niu S.J., Wang Q.L., Liang J.Y. (2005) *Advances in Marine Science,*23(2), pp.219-226.
[4] GPS/Met University Corporation for Atmospheric Research (UCAR)
[5] Rawat Laxmi (1998) *changing facets of weather and climate in Doon Valley, FRI.*
[6] Sharma Gajendra (2008) *Data mining,data warehousing and OLAP (Second Edition), Katson Books.*
[7] Mallach G. Efrem (2002) *Decision support and data warehouse sytems. Tata McGraw-Hill.*
[8] Tajunisha Saravanan (2011) *International Journal of Database Management Systems,* Vol. 3, No. 1.
[9] WANG Shi Huai (2011) *IEEE.*
[10]Ma Nan, Yuan Mei, Bao You Wen, Jin Zong Min, Zhou He (2010) *Second International Conference on Information Technology and Computer Science, IEEE.*
[11]OLAP Council (1995) *The Guide to OLAP technology.*