

Introduction

There are massive amounts of weather data generated in real-time on a daily basis. Archives of predicted and observed data can easily become enormous. Both researchers and end-users of weather information require tools that can perform rapid analysis and easily share these data.

Atmospheric scientists often store meteorological data in GRIB (GRIdded Binary), which is the WMO standard format for transmitting gridded data. Data sets stored in the GRIB format may be split across multiple files, each file containing compressed, binary data, which adds complexity to querying meteorological data across multiple variables. To simplify the process of analyzing meteorological data across multiple dimensions, we developed a tool that uses a generalized format for each meteorological data point to store and query a collection of data.

Our tool, known as Wintx, permits queries focused on a location over a duration of time within a single command. Performing a similar query with data in the GRIB format would require analyzing multiple GRIB files, expanding their compressed binary data, filtering the geospatial grids on multiple levels, then referencing the geospatial locations with the variable data. Adding additional constraints with Wintx, such as limiting the data in the query to a temperature range, is also done in the same query command. In order to add the same constraint for data in the GRIB format, additional filters are required to accommodate each file.

Requirements

Our requirements were formed to satisfy the need for scalability, space efficiency, and geospatial searches:

- Generalized format for storing and retrieving meteorological data
- Interface to store and query data across multiple variables
- Stores data in a scalable architecture
- Allow geospatial queries from shape files
- Imports the data to keep up with daily weather prediction operations
- Efficiently utilizes storage space
- Provides access to data through web based services

Visualization of Weather Information Using a Scalable Database

Michael Baldwin, Seth Cook, Jaewoo Shin, and Stephen Harrell

Purdue University

Wintx Design

Wintx is a Python library written to provide an abstract interface for Python scripts to use. The interface encapsulates interactions with a sharded database, allowing developers to write programs that interact with their data without using database or sharding functions. A RESTful^[2] service has been developed using this interface to expand Wintx's functionality to web based services. The generalized format, developed to interact with meteorological data, allows for the database to change without impacting the user's interaction with Wintx.



Figure 1: The Wintx workflow model

A NoSQL database (MongoDB[3]) and a relational database (MySQL Fabric[1]) were tested as potential candidates for Wintx. MongoDB was tested with and without indexing. NoSQL databases store uncompressed metadata with each data point, unlike relational databases which can normalize data and compress the duplicate metadata. As seen in Figure 2, the MySQL Fabric solution stores a data point in the same space as the MongoDB without indexing and in less space than the MongoDB with indexing. MySQL Fabric is the ideal solution in terms of storing data efficiently. The original data no longer needs to be maintained once ingested into Wintx. Only database sizes are taken into consideration for these tests.

Figure 2: Amount of space (in gigabytes) used on disk by each database. Each GRIB file tested contained # records.

Web App Description

WintxApp is a web-based application to visualize meteorological data in Google Maps using Wintx. Using WintxApp, users can interact with underlying data easily. They can choose parameters, such as date range, aggregate function, area and variable name, based on what they want to see. Once parameters are set, the application builds a request URI to access to the Wintx RESTful service, then receive meteorological results from the server. Subsequently, the result data is parsed and processed to make new layer on specific geospatial locations with colored tiles, and finally the result layer is added on Google Maps.

In the WintxApp, users can draw a box on the map or select a specific Figure 3: The web application for Wintx INDOT area to be probed so it can filter out the whole area into a smaller one. It also offers features to aggregate / compare results data. By selecting a specific function, such as sum, max, min, mean or standard deviation, results are aggregated based on the function and then the aggregated results are visualized on the map. Moreover, it can animate results to show how results have been changed over a duration of selected time (time-series).

Database Disk Space Usage





Wintx was created to satisfy the need for scalability, space efficiency, and geospatial searching with meteorological data. Our solution, Wintx, provides a generalized method for interacting with data. This service provides users an interface that is simple to interact with and quick to develop applications against. It uses a storage facility that is capable of scaling with the user's data while storing the data within a manageable amount of resources.

- [1] Oracle Corporation. MySQL Fabric.
- [2] Todd Fredrich.
- [3] MongoDB Inc. MongoDB.

This work was supported by the Indiana Department of Transportation. The contents of this paper reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented herein, and do not necessarily reflect the official views or policies of the sponsoring organizations. These contents do not constitute a standard, specification, or regulation.

Mike Baldwin <baldwin@purdue.edu> Seth Cook <sethcook@purdue.edu> Jaewoo Shin <shin152@purdue.edu> Stephen Harrell <slh@purdue.edu>





Conclusion

References

REST API Tutorial, August 2013.

Acknowledgements

Contact Information

