Large-Scale Data Analysis and Visualization for ARM Using NoSQL Technologies

Bhargavi Krishna[μ], Kyle Dumas[μ], William Gustafson[ω], Andrew Vogelmann[ρ], Tami Toto[ρ], and Giri Prakash[μ]

[μ]Oak Ridge National Laboratory　　[ω]Pacific Northwest National Laboratory　　[ρ]Brookhaven National Laboratory

1. Objective

Provide near real-time analytics and visualizations for ARM data such as from LES ARM Symbiotic Simulation and Observation (LASSO[1]), radar, and best estimate value added products
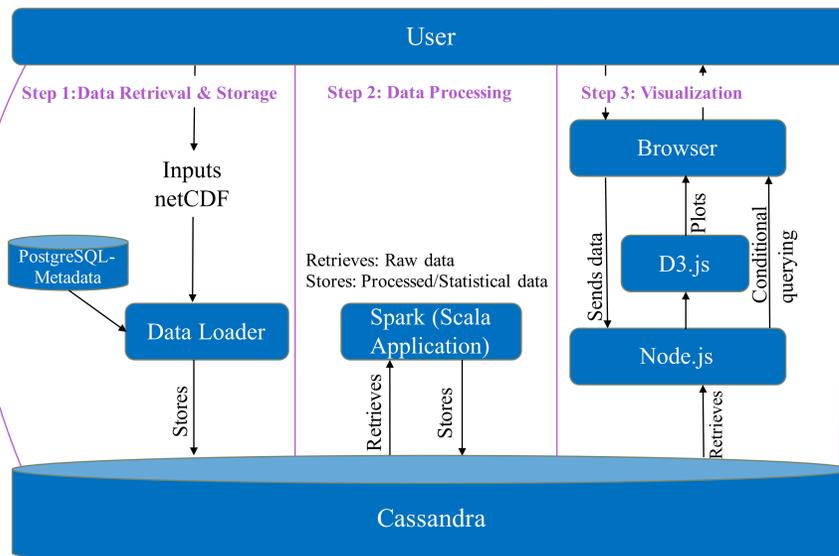
2. Introduction

I. LASSO:

- Lays the groundwork to generate regular LES modeling at the ARM Southern Great Plains (SGP).
- Result: library of simulations
- A "Data Bundle" for LASSO combines ARM observations and high-resolution model output to provide a highly detailed description of the atmosphere.

II. Conditional Querying:

- Current ARM data searches are performed by using its metadata. NoSQL can improve the data searching capabilities by including measurement values.
- Data: ARMBEATM[2] provides best estimate of selected atmospheric state profiles and surface quantities averaged every hour.
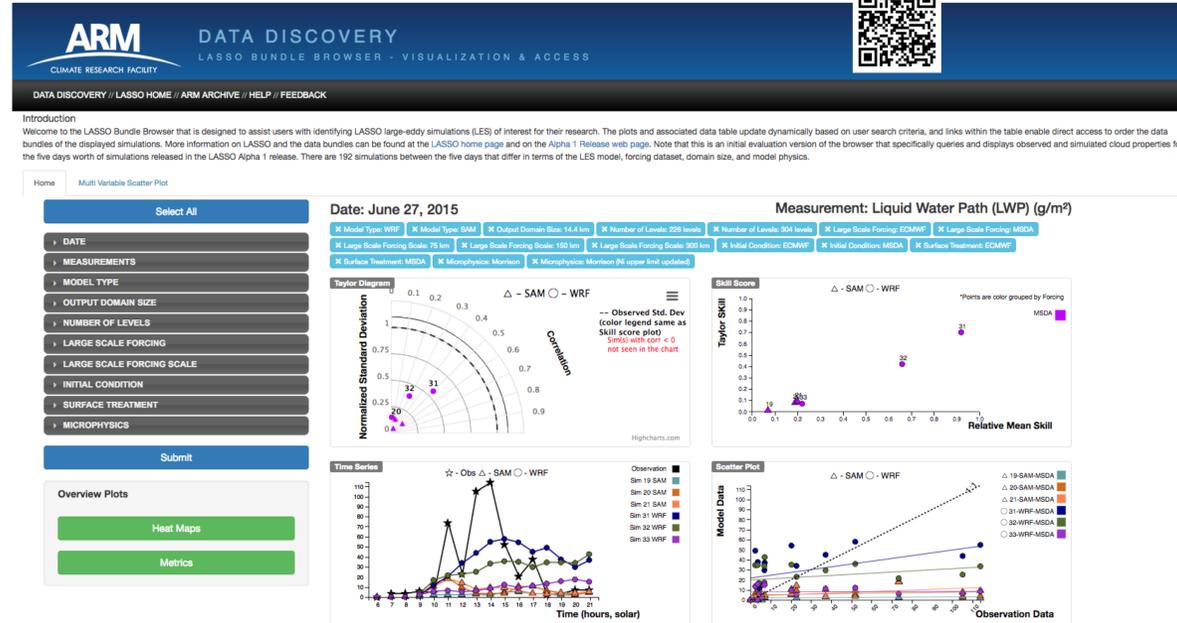
3. Workflow

4. Technology

Cassandra	Spark
NoSQL Database	**Processing Framework**
Distributed Environment	Distributed Environment
High Availability	In-Memory Processing
Elastic Scalability	Master-Worker Architecture
Column Store	Cassandra Compatible
Easy to Use	Written in Scala
Good Documentation	Good Documentation
Production Cluster Details	
Number of nodes - 5	Number of nodes - 5 (1 master and 4 workers)
RAM: 256GB	RAM: 256GB
CPU Cores: 32	CPU Cores: 32
Storage: 2 nodes -> 3TB SSD	Storage: 2 nodes -> 3TB SSD
3 nodes -> 3TB HDD	3 nodes -> 3TB HDD

5. Visualization

5.1 Interactive Visualizations for LASSO and ARM Observational Data

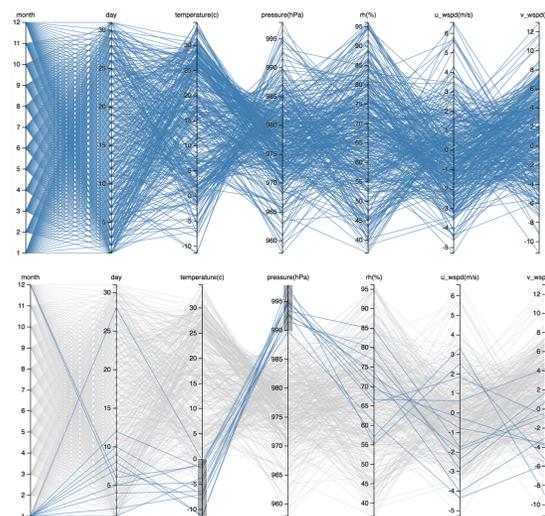http://archive.arm.gov/lassobrowser

5.2 Interactive Visualizations for ARMBE data
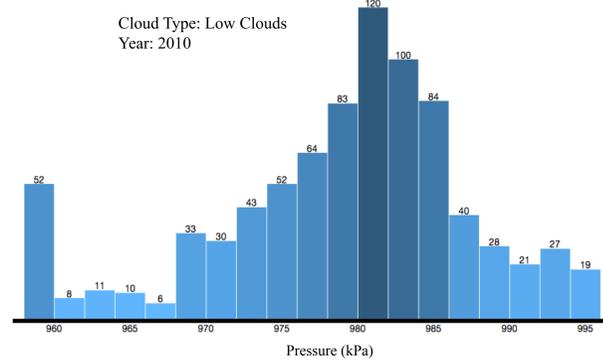
Parallel Coordinates

Each measurement/dimension is a coordinate which allows for selecting ranges and are interchangeable.

Histogram

Visualization shows histograms for different variables from armbe data based on cloud types.

The cloud types values were provide by Laura Riihimaki in the same time sample as ARMBE data.

5.3 Interactive Visualizations Using RADAR Data

The two figures below represent a sample statistical summary of radar data as a map and multiline time series plot. The data was provided by Scott Collis and Jonathan Helmus.
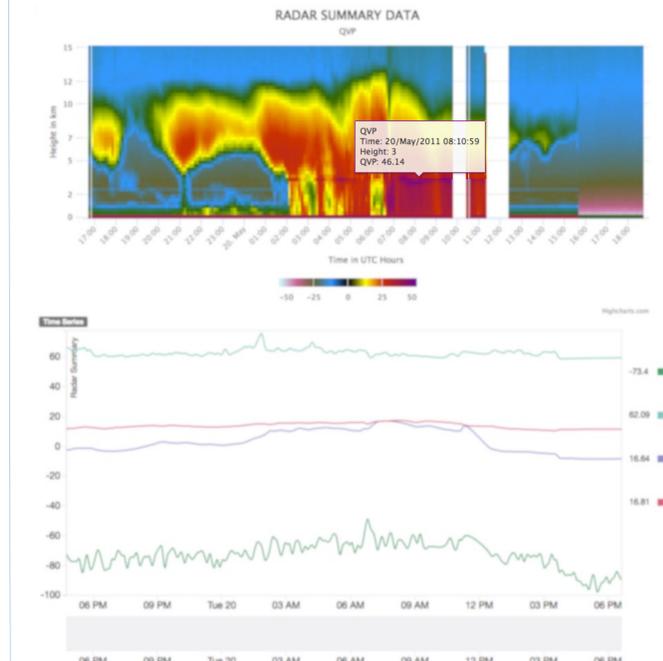
6. Data Retrieval, Processing, and Storage

- ARMBEATM[2] datastream generated use cases. Spark Scala Application generated outputs for conditional querying and statistics.
- Data are retrieved dynamically based on user selection.

Conditional querying:

Use Case : The table below shows the days in which surface temperature was less than 0 °C/273.15 K in 2012 at SGP

7. "We'd like to hear from you…."

Bhargavi Krishna, Ph.D.

Email: krishnab@ornl.gov

Phone: 865-574-8264

Giri Prakash

Email: palanisamyg@ornl.gov

Phone: 865-241-5926

8. References

1. https://www.arm.gov/capabilities/modeling/lasso/
2. http://www.arm.gov/data/vaps/armbe/armbeatm
3. https://github.com/mbostock/d3/wiki/Tutorials