20160804 Meeting with Brazil

Attendees:

Joao, Bebo, Les (SLAC), Maria, Marcos, Antonio, Renan, Claudio (Brazil)

Admin

Contacts see Contacts for Brazil.

Topics

Introduction to PingER if necessary, n.b. Maria, Les, Joao, Bebo, Antonio are already familiar with PingER

- http://www-iepm.slac.stanford.edu/pinger/
- http://www.slac.stanford.edu/comp/net/wan-mon/tutorial.html
- http://www-iepm.slac.stanford.edu/paperwork/ieee/ieee.pdf
- PingER Papers and Presentations
  - Applying Data Warehousing and Big Data Techniques to Analyze Internet Performance, Thiago Barbosa, Renan Souza, Sergio Serra, R. Les Cottrell presented at NetApps2015, Dec 2015

There appears to be considerable interest in the Brazilians to develop new ways to visualize the PingER data. This would continue after Joao leaves at the end of August. This would use a warehouse to access the data via a Hadoop cluster etc. Initially this would use the warehouse that Joao has installed at SLAC. It could also be a warehouse elsewhere since Joao now keeps the FTP hourly data up to date daily and it is possible to just get the new data (see https://confluence.slac.stanford.edu/display/IEPM/Public+FTP+daily+update)

We are thinking of a brain-storming session with the Brazilians, (cc'd) Bebo, Joao and myself. This should be soon, i.e. before Joao leaves. We were thinking of this coming Thursday, would you be available, even better if could be at SLAC. Since they are 4 hours ahead it could be daytime.

We are thinking of an extension to pingtable.pl to make the menu much more powerful, e.g. provide the standard ticks(like we do today) plus the ability to select times and dates, allow selection of multiple metrics, allow selection of multiple from and to, plus allow selection of multiple visualization (time series, bar charts, maps etc.) One place to look for graphs that we need today but require exporting to Excel etc to create would be the yearly PingER status reports (e.g http://www.slac.stanford.edu/xorg/icfa/icfa-net-paper-jan15/report-jan15.docx). Other visualizations would be to make comparisons versus economic (HDI, IDI, corruption ...) or development indices (e.g. correlation plots a la http://www-iepm.slac.stanford.edu/pinger/pinger-metrics-motion-chart.html) or even versus other data available via the semantic web

Some examples of required graphics

Accessing data

Hourly data is accessible via FTP at ftp://ftp.slac.stanford.edu/users/cottrell/ as tar files by metric. This is updated daily. See Archiving PingER data by tar for retrieval by anonymous ftp for how to access the complete set. Also can just few (up to 60) days in case you missed some days (see PingER Public FTP Archive daily update). Could use this to populate one's own warehouse.

How do we move forward

Program that supported Renan, Thiago and Joao has finished.

Complete SLAC warehouse

- Measure performance
- Is SLAC warehouse accessible from Brazil?
  - Use SLAC warehouse
  - Can we keep it running?
  - Need to get into production, document and get performance timings
  - Create own warehouse?

Provide web access to select data using a URL or menu to choose:

- one or more metrics,
- pre-set time ticks,
- Select time window
- select one or more from hosts or aggregates
- select one or more to hosts or aggregates
- select visualizations: map, time series, bar charts

What resources are Brazil thinking of engaging?

- Students, BS, MS, PhD
- Courses

What are interests and for whom:

- Warehouse
  - Maria suggests two undergraduate students put up an OLAP application on top of Joao's Data Warehouse.
  - Maria already has a an ETL (Extraction Transforming Loading) process for a small Data Warehouse they had at UFRJ (done Cristiane). This could be adapted to maintain the SLAC Data Warehoues from the FTP hourly PingER data.
  - One of Maria's students (Rafaelo) has a cluster of four Impala/Hadoop nodes. He can use SPARQL to make analytical queries (not sure if data is PingER or other)
  - Renan interested in speeding up queries from the Warehouse
- Semantic web
  - Bebo, Renan
- Visualization & analysis
  - Marcos, Claudio and Antonio
    - One thing they are looking at is real-time visualization, which needs fast response. From Renan's work a couple of years ago response time to OLAP queries using Impala of 100 million tuples was ~ 30 seconds.

Next meeting

Thursday 8/19/2016 4:00pm SLAC, 8:00pm Rio.

Atendees: Les (SLAC), Bebo (SLAC), João (SLAC), Guto (Rio), Renan (Rio), Maria Luiza (Rio)

Some slides about the meeting: PingER Vis.pdf

Also look at meeting Brazil in early September after Joao returns.