Request For Quotation (RFQ)

MUSTANG PostgreSQL Database Query Optimization, Performance Issues, and Upgrade (9.5 to 11.4)

August 2019

IRIS Data Management Center
1408 NE 45th St.
Seattle, WA 98105

After September 1, 2019:
IRIS Data Management Center
9706 4th Ave. NE, Suite 303
Seattle, WA 98115

1.0 INTRODUCTION

Incorporated Research Institutions for Seismology (IRIS) is a consortium of over 126 universities with the purpose of archiving, managing, and distributing seismic data. The IRIS Data Management Center (DMC) in Seattle, Washington is affiliated with the University of Washington, which serves as the host institution.

The IRIS DMC MUSTANG system calculates seismic data quality metrics and power spectral measurements, stores data in a PostgreSQL 9.5 database, and makes data available to the seismological community. The values in the production database are accessible by the public 24/7 through web services at http://service.iris.edu/mustang.

1.1 Project Description

The database is currently 11 TB in size, with 2.5 TB of quality metrics and 8.5 TB of spectra. There is also a smaller 143 GB schema that supports our scheduler system. The largest table is more than 105 million rows and 6.8 TB. This is a database that is always changing; metrics and spectra for newly archived data are calculated daily. The values are also recalculated when data or metadata changes. The database is growing by 500GB each year and this rate is expected to increase. Through the web services, we deliver 500-1000 GB of metrics/spectra to outside users each year.

The MUSTANG database server is running on CentOS 7 OS with VMWare hypervisor (8 core/24 GB RAM). It has NFS mounted cluster tablespaces on 2xNetApp. The production database is copied to a backup off-site database using streaming physical replication to an identical remote server in standby mode. Indexing is used for the metrics tables and there is no partitioning currently. Interaction with the database is largely through Java code and employs Spring Framework JDBC templates. In addition, there is also a small 10 GB test database that has the metrics schema but not the spectra schema.

1.2 Observations and Issues

a) Slow queries

IRIS is encountering performance problems with large queries for quality metrics. For example, requesting the Global Seismic Network (_GSN) percent_availability values for a single day can take 5-10 seconds or more:
http://service.iris.edu/mustang/measurements/1/query?metric=percent_availability&network=_GSN&channel=?H?&format=text&timewindow=2019-07-01,2019-07-02&nodata=404

Another example is a request for the USArray Transportable Network (TA):
http://service.iris.edu/mustang/measurements/1/query?metric=sample_mean&net=TA&cha=?H?&format=xml&start=2014-07-01&end=2014-07-02

Requesting metric values for a time period of a year often results in a timeout exception:
http://service.iris.edu/mustang/measurements/1/query?metric=percent_availability&network=_GSN&channel=?H?&format=text&timewindow=2018-07-01,2019-07-02&nodata=404

http://service.iris.edu/mustang/measurements/1/query?metric=sample_mean&net=TA&cha=?H?&format=xml&start=2014-07-01&end=2015-07-02

For spectra, we observe recent intermittent slowdowns in web service returns. IRIS believes this might be related to autovacuuming processes running or high I/O loads, but our team has not been able to verify this. Below is an example of a Probability Density Function (PDF) plot from the power spectral values:
http://service.iris.edu/mustang/noise-pdf/1/query?net=TA&sta=109C&loc=--&cha=BHZ&quality=M&format=plot&plot.interpolation=bicubic&nodata=404

PDFs and Power Spectral Density (PSD) values make up a schema that is very large: over 6TB in size, much of which is array data that is stored as TOAST tables. There is a single table that incorporates a majority of the space and it is not partitioned. We want to explore the most sensible options to make this continually growing schema more performant.

b) Database upgrade

The other issue is that IRIS is using PostgreSQL version 9.5 and IRIS needs to upgrade to version 11.4. This upgrade needs to be performed on the production database, the off-site backup database, and the small test database. IRIS is interested in implementing new features from versions 9.6/10/11 that might assist with the performance issues.

2.0 SCOPE OF WORK

- Diagnose causes of slow query return and implement fixes.
- Advise on query optimization, indexing, partitioning, other database design issues, scaling, autovacuum tuning, configuration changes, or hardware recommendations.
- Plan and perform upgrade from PostgreSQL 9.5 to 11.4 for the test, production, and backup MUSTANG database.
- If the estimated database downtime needed for the upgrade is greater than 4 hours, provide an estimate of downtime.
- Document all changes made to the database.

Note: Work can be performed during normal business hours with access to the database available through SSH.

3.0 SUBMISSION REQUIREMENTS

3.1 Vendors **must** provide the following:

1. Time and cost estimates per item identified in Section 2.0 – Scope of Work

2. Total projected cost estimate

3. Dates of availability to conduct work

Note: Please include estimates for both a standard hourly service and any special pricing arrangements that might be offered such as a pre-paid retainer contract or a monthly service agreement. If on-site work is an option, please include cost estimates for on-site services in addition.

3.3 PROPOSAL SUBMISSION DETAILS

Vendors should submit responses via email to spo@iris.edu. Vendor responses are due no later than **5:00 p.m. ET** on **Wednesday, August 28, 2019**.

All questions should be directed to: SPO@iris.edu.

All vendor questions and answers will be made available to all potential bidders.