Monitoring Ozone Levels with PostgreSQL

Alex Lai, Marty Brandon
May 2012
Overview

- Background (Marty)
- Ingestion & Processing (Alex)
- Challenges (Alex)
- Ongoing Development (Marty)
Background
Data Downloading
Earth Science Data Types

- Interface between the raw science data and the data management system
- > 900
EOS Aura Atmospheric Profile Measurements

OMI also measures UVB flux, cloud top/cover, and column abundances of O₃, NO₂, BrO, aerosol and volcanic SO₂
TES also measures several additional 'special products' such as ClONO₂, CF₂Cl₂, CFC₃, N₂O and volcanic SO₂

HIRDLS: High Resolution Dynamics Limb Sounder
OMI: Ozone Monitoring Instrument
MLS: Microwave Limb Sounder
TES: Tropospheric Emission Spectrometer
Ingestion & Processing
Ingestion Processing

Auto Ingest

Legacy Ingest Process

Ingest Planner

OMI Operators

Sync Database Process

MDS Ingest Process

Postgres Legacy Metadata Server

Postgres MDS Metadata Server

ARCHIVE SERVER

Archive Node

Archive Node

Archive Node

Thursday, May 17, 12
Challenges
Backup and Asynchronous Replication

Master Server
- Legacy/MDS Database

Slave Server
- Legacy/MDS Database

Monitor/Remove Applied Wal files
- Wal files Archive Disk
  - 20 days dumps
    - pg_dump
  - scp/rsync

ADNET server
- 20 days dumps
  - OFF-SITE ADNET'S SERVER

TLCF server
- 20 days dumps
  - GSFC/NASA TLCF CLUSTER

GSFC/NASA

TLCF server
- ADNET server
- Legacy/MDS Database

ADNET's server
- GSFC/NASA
- TLCF cluster
- Archive Disk
- Monitor/Remove Wal files
- Applied Wal files
Monitor WAL is the key to ensure the replication is in good status?

- The cause of WAL not properly apply to Slave
  - Network unavailable
  - WAL files were deleted
  - While COPY took too long to finish that cause the apply WAL file in wait status (set Max_standby_streaming_delay = -1)
- Slave hangs for any reasons
Monitor WAL is the key to ensure the replication is in good status?

- Slave unable to connect and copy wal file from replication archive disk
- ps aux shows waiting status in slave
- Pg log also show waiting no recovery state for a long period of time
Monitor WAL is the key to ensure the replication is in good status?

- Query using functions:
  - `pg_current_xlog_location()` (master only)
  - `Pg_last_xlog_receive_location()`,
  - `Pg_last_xlog_replay_location()`
- When slave hangs for any reasons, an monitoring script will trigger an alert email to DBA
Monitor WAL is the key to ensure the replication is in good status?

- What and how to monitor the WAL files?
  - Monitor the disk space available for WAL archive
  - Monitor the last applied WAL file in Slave
  - Monitor number of purging old applied WAL file
  - Monitor any wait states
  - Monitor dead lock situation
Monitor WAL is the key to ensure the replication is in good status?

- Email Notification to DBA for
  - Disk storage down to warning level
  - The last applied WAL file not changed longer than the statement_timeout in SLAVE
- Network unavailable between master and slave
- Log the long queries that exceed a long time like 20 minutes
Monitor WAL is the key to ensure the replication is in good status?

• Too many processes in wait state
• When Database has slow performance for a period of time
HOW much data do we have?

- Our current database implementation includes 10 clusters, each running Postgres 9.0.6 and divided into three production levels: development, testing, and operations.
- On average about 200 commits per second
- The largest table contains over 160 millions rows
- Total data increase from 40 – 60 % in the last three years.
Number of records (Legacy)

- **FILE**
  - 2010 Jan-Dec: 
  - 2011 Jan-Dec: 
  - 2012 Jan-April: 

- **FILEMETA**
  - 2010 Jan-Dec: 
  - 2011 Jan-Dec: 
  - 2012 Jan-April: 

- **OTHERS**
  - 2010 Jan-Dec: 
  - 2011 Jan-Dec: 
  - 2012 Jan-April: 

- **TOTAL**
  - 2010 Jan-Dec: 
  - 2011 Jan-Dec: 
  - 2012 Jan-April: 

**Units:** millions

**Legend:**
- Blue: 2010 Jan-Dec
- Red: 2011 Jan-Dec
- Green: 2012 Jan-April
Number of records (MDS)

- FILE
- FILEMETA
- OTHERS
- TOTAL

- 2010 Jan-Dec
- 2011 Jan-Dec
- 2012 Jan-April

millions
Synchronization between Legacy and MDS

MDS – Always has long delay
Also has to resolve record locking
OMI LEGACY DATABASE SCHEMA
# Setting of Postgresql.conf

<table>
<thead>
<tr>
<th>Variables</th>
<th>MASTER</th>
<th>SLAVE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>wal_level</strong></td>
<td>hot_standby</td>
<td>hot_standby</td>
</tr>
<tr>
<td>(enable read-only queries</td>
<td></td>
<td></td>
</tr>
<tr>
<td>on a standby server)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>hot_standby</strong></td>
<td>off</td>
<td>on</td>
</tr>
<tr>
<td></td>
<td>(master MUST be off)</td>
<td>(enable read-only queries</td>
</tr>
<tr>
<td></td>
<td></td>
<td>on a standby server)</td>
</tr>
<tr>
<td><strong>wal_keep_segment</strong></td>
<td>32</td>
<td>32</td>
</tr>
<tr>
<td><strong>wal_max_senders</strong></td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>
Setting of Postgresql.conf

<table>
<thead>
<tr>
<th>Variables</th>
<th>MASTER</th>
<th>SLAVE</th>
</tr>
</thead>
<tbody>
<tr>
<td>archive_mode</td>
<td>on</td>
<td>off</td>
</tr>
<tr>
<td>Archive_timeout</td>
<td>30s</td>
<td>30s</td>
</tr>
<tr>
<td>archive_command</td>
<td>'cp %p /your_path/%f'</td>
<td>'cp %p /your_path/%f'</td>
</tr>
<tr>
<td>Max_standby_archive_delay</td>
<td>300s</td>
<td>1200s</td>
</tr>
<tr>
<td>Max_standby_streaming_delay</td>
<td>300s</td>
<td>-1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Wait for any queries</td>
</tr>
<tr>
<td></td>
<td></td>
<td>before apply for the</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WAL file</td>
</tr>
<tr>
<td></td>
<td></td>
<td>This avoid pg_dump</td>
</tr>
<tr>
<td></td>
<td></td>
<td>COPY fail.</td>
</tr>
</tbody>
</table>
## Setting of Recovery.conf on Slave

<table>
<thead>
<tr>
<th>Variables</th>
<th>SLAVE</th>
</tr>
</thead>
<tbody>
<tr>
<td>standby_mode</td>
<td>‘on’</td>
</tr>
</tbody>
</table>
| primary_conninfo   | ‘host=ip_of_Master
port=5432
user=user_has_all_access_to_all_database
pw=user_password’ |
| trigger_file       | ‘/path/to/trigger’                                          |
| restore_command    | ‘cp /path/to/archive_files/%f “%p”’                         |
Ongoing Work
• Web Framework
• ORM
• Database Patches
Web Framework

View ➔ Model

Controller

GET /person/(*key)
POST /person
DELETE /person/(*)key)
ORM

Server Start:
- read config file
- register database
- generate CRUD methods
Database Patches

Git Repo - Component X
  code
  schema.sql
  patch1.sql

Patch2.sql

Instance - Component X

Thursday, May 17, 12
Database Patch Details

Git Repo
- base.sql
- patches_applied.txt
- patch1.sql
- patch2.sql
- ...

Instance
- patches_applied

base.sql = schema dump + new patches
Perl Modules

- Mojolicious
- Rose::DB
- Module::Build
Monitoring

Database Clusters 18:40:27 (checking every 5 minutes)

- acpsdb1
- acpsas1
- devarch11
- testdb2
- testmds1
- devdb2
- testarch1
- devmds1
- acpsdb2

Thursday, May 17, 12
devmds1 18:34:48 (checking every 60 seconds)

ssh -N -L3002:acpsdb1:5433 acpsdb1.omisips1.eosdis.nasa.gov
databases: omiops, metamine, policy
roles: omiops, omiopsdb

Connectivity
- ssh connection verified
- psql connection verified

Load
- CPU load is %10

Space
- 10 GB of 100 GB is unused

Replication
- master WAL file: AF4322BADD3
- slave WAL file: AF4322BADD2