Infrastructure - Task #3244

Implement PostgreSQL PITR streaming backups with Barman

2020-06-18 10:02 - Brett Cornwall

<table>
<thead>
<tr>
<th>Status:</th>
<th>In Progress</th>
</tr>
</thead>
<tbody>
<tr>
<td>Priority:</td>
<td>Normal</td>
</tr>
<tr>
<td>Assignee:</td>
<td>Guilhem Moulin</td>
</tr>
<tr>
<td>Category:</td>
<td></td>
</tr>
<tr>
<td>Target version:</td>
<td>Q2/2021</td>
</tr>
<tr>
<td>Tags:</td>
<td>Salt</td>
</tr>
<tr>
<td>Start date:</td>
<td></td>
</tr>
<tr>
<td>Due date:</td>
<td></td>
</tr>
<tr>
<td>% Done:</td>
<td>0%</td>
</tr>
<tr>
<td>URL:</td>
<td></td>
</tr>
</tbody>
</table>

Description

PostgreSQL maintains a Write-ahead log (WAL), which is a record of transactions that can be "replayed" on top of an older backup to recover data.

Barman is an application for physical backups of PostgreSQL servers that manages base backups and WAL archiving.

Features include:

- Retention policies (e.g. `recovery window of 30 days` or `redundancy 3`):
- GPL3
- Debian inclusion
- Backed by a company affiliated with core PostgreSQL development
- Focuses on backup/recovery procedures
- Promotes testing backups
- Geographical redundancy reduces single points of failure (e.g. backup/prod being in the same datacenter).

A reasonable way to start modelling disaster recovery architecture is to design a couple of possible architectures in respect to PostgreSQL and Barman, such as evaluating the pros and cons of:

- Same-data-center backups
- Different data center backups in the same metropolitan area
- Different data center backups

Evaluate the single points of failure (SPOF) of the system, with a cost-benefit analysis.

Having said this, a very common setup for Barman is to be installed in the same data center where the PostgreSQL servers are. In this case, the single point of failure is the data center. Fortunately, the impact of such a SPOF can be alleviated thanks to two features that Barman provides to increase the number of backup tiers.

Planning for disk space is important: Plan the required disk space based on the size of the cluster, number of WAL files generated per day, frequency of backups, and retention policies.

Plan for having an identical recovery environment that are as similar as possible, if not identical, to the original server. Doing so eases burdens of maintenance.

Streaming backup is recommended from PostgreSQL 9.4 or higher and is easier to set up.

History

#1 - 2020-06-18 15:28 - Guilhem Moulin
- Status changed from New to In Progress
- Assignee set to Guilhem Moulin

#2 - 2020-08-04 18:13 - Florian Effenberger
- Target version set to Q4/2020

#3 - 2021-03-01 10:56 - Guilhem Moulin
- Target version changed from Q4/2020 to Q2/2021
We'll reuse the old Pootle box (upgraded to XXL) for that, Brett is taking care of the initial deployment and will hook a RDBMS or two.