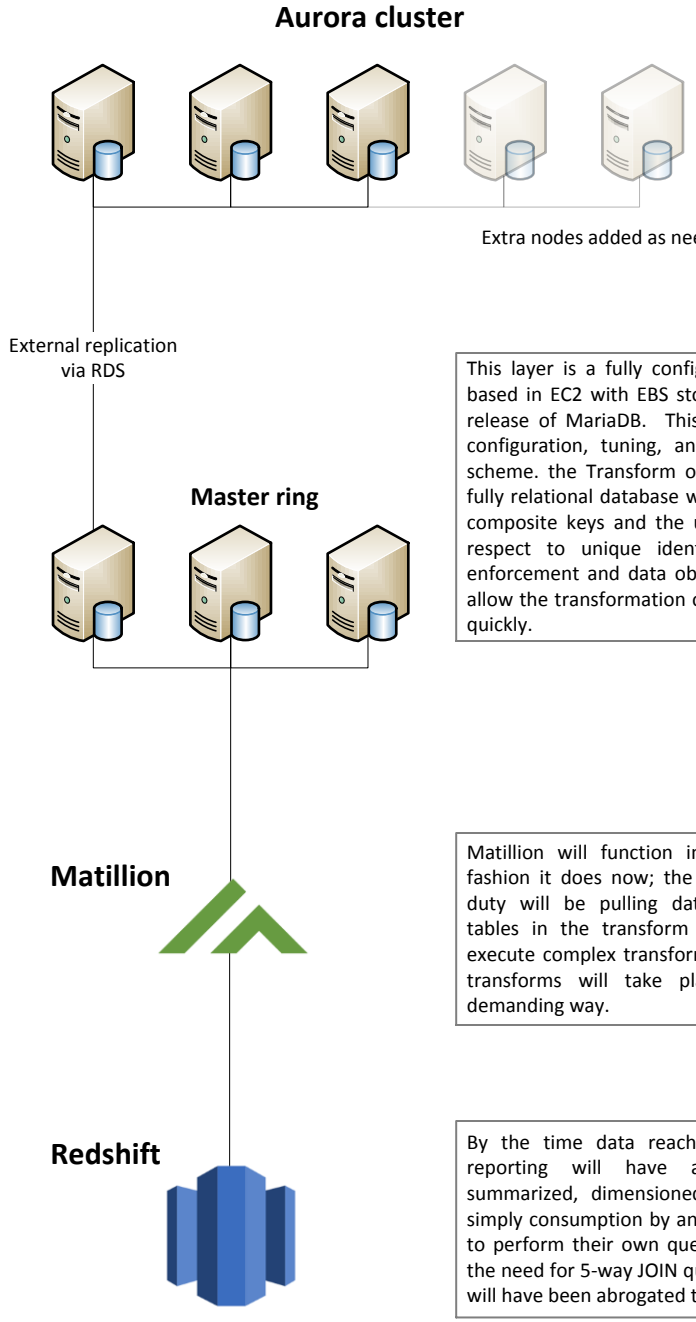
Proposed rearchitecture of the current data tier

Divide and Conquer



OLTP layer

The Aurora cluster will serve as a drop in replacement for the current OLTP monolith. Aurora provides for both multiple entrypoints for data, as opposed to the classic Master/Slave/Slave architecture (with some caveats) and is both geographically disparate and with all the benefits of an RDS instance, such as snapshotting, DR, etc.

Transform layer

This layer is a fully configurable MariaDB master ring, based in EC2 with EBS storage, running the current GA release of MariaDB. This will allow for full control of configuration, tuning, and a much more robust ACL scheme. the Transform operations will take place in a fully relational database with the benefits of simple and composite keys and the utilities we already have with respect to unique identfiers for data, foreign key enforcement and data obfuscation if needed. This will allow the transformation of data to proceed much more quickly.

These three MariaDB instances will be slaves to the Aurora cluster. In addition to having this group of instances doing the “heavy lifting,” they can also serve as an immediate backup, especially if configured to start and stop replication at intervals to allow the BI transforms to have a steady-state.

Extract/Load layer

Matillion will function in this scenario in the same fashion it does now; the difference is that its primary duty will be pulling data from already-dimensioned tables in the transform layer rather than having to execute complex transforms on-the-fly. Certainly some transforms will take place, but not in the same demanding way.

OLAP layer

By the time data reaches Redshift, therefore, most reporting will have already been appropriately summarized, dimensioned, and cubed and ready for simply consumption by analysts. The ability for analysts to perform their own queries will certainly remain, but the need for 5-way JOIN queries and abstruse operations will have been abrogated to a great extent.

Richard Byrd and Mark Butler
2019-02-01