

Hybrid Transactional Analytical Processing (HTAP) for Amazon Redshift

Whitepaper

Thank you for downloading this AWS and Heimdall Whitepaper! Carahsoft is the distributor for AWS public sector solutions available via GSA, NASPO, The Quilt and other contract vehicles.

To learn how to take the next step toward acquiring AWS's solutions, please check out the following resources and information:



For additional resources:
carah.io/AWS-Resources



For upcoming events:
carah.io/AWS-Events



For additional AWS solutions:
carah.io/AWS-Solutions



For additional Marketplace solutions:
carah.io/AWS-Marketplace



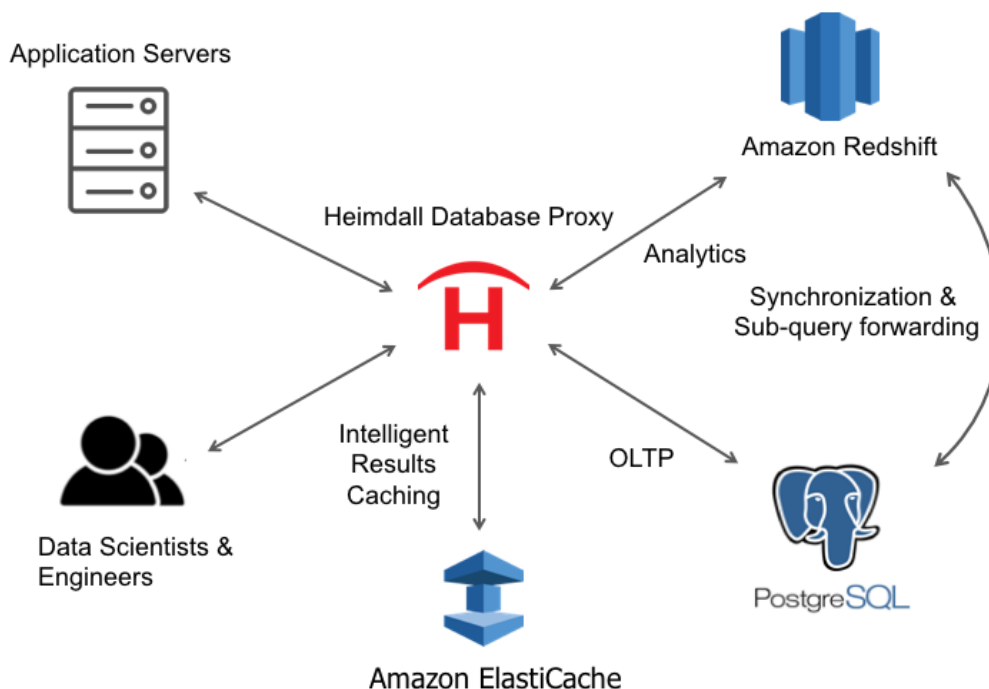
To set up a meeting:
AWS@carahsoft.com
888-662-2724



To purchase, check out the contract vehicles available for procurement:
carah.io/AWS-Contracts

For more information, contact Carahsoft or our reseller partners:
AWS@carahsoft.com | 888-662-2724

Hybrid Transactional Analytical Processing (HTAP) for Amazon Redshift



Companies deploy Amazon Redshift for large-scale MPP (Massively Parallel Processing) data analytics. These systems apply parallel compute resources to answer queries quickly. However, there are still performance challenges:

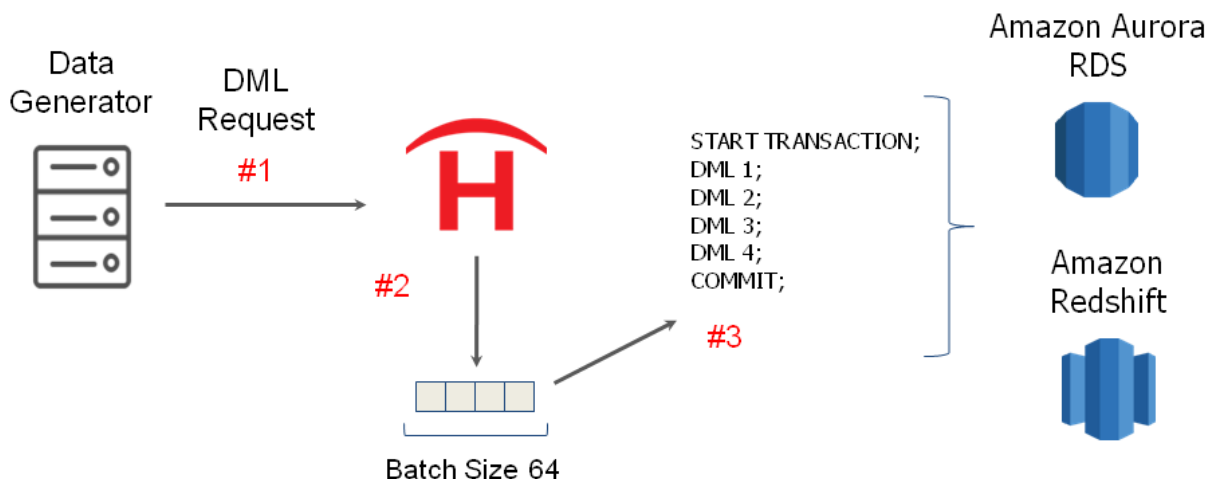
- Transactional overhead: Singleton DML operations result in slow application response times
- Higher latency: Distributed queries means that many nodes have to coordinate to generate an answer
- No materialized views: If a query result is used for multiple following queries, additional work is required to preserve the result or repeated calls to the same base query will be required

These challenges can be solved by modifying the application. However, the ideal solution requires no code changes. Heimdall Data is that solution providing Amazon Redshift users the ability to:

- Batch singleton, DML operations
- Fast materialized views for results for Amazon Redshift
- Automated SQL caching for Amazon ElastiCache

Hybrid Transactional Analytical Processing (HTAP) for Amazon Redshift

Asynchronous Batching Processing



Heimdall Data improves database write performance by batching DML operations against a table and putting them onto a separate connection under a single transaction. Batching DML operations result in:

- Improved application response times due to fewer commits
- Improved DML scale

Ideal use case for batch processing:

Insert/update/delete a large amount of data at once on a thread. Heimdall can process it all at once much faster than if individual queries outside of a transaction were completed.

Not so ideal use case for batch processing:

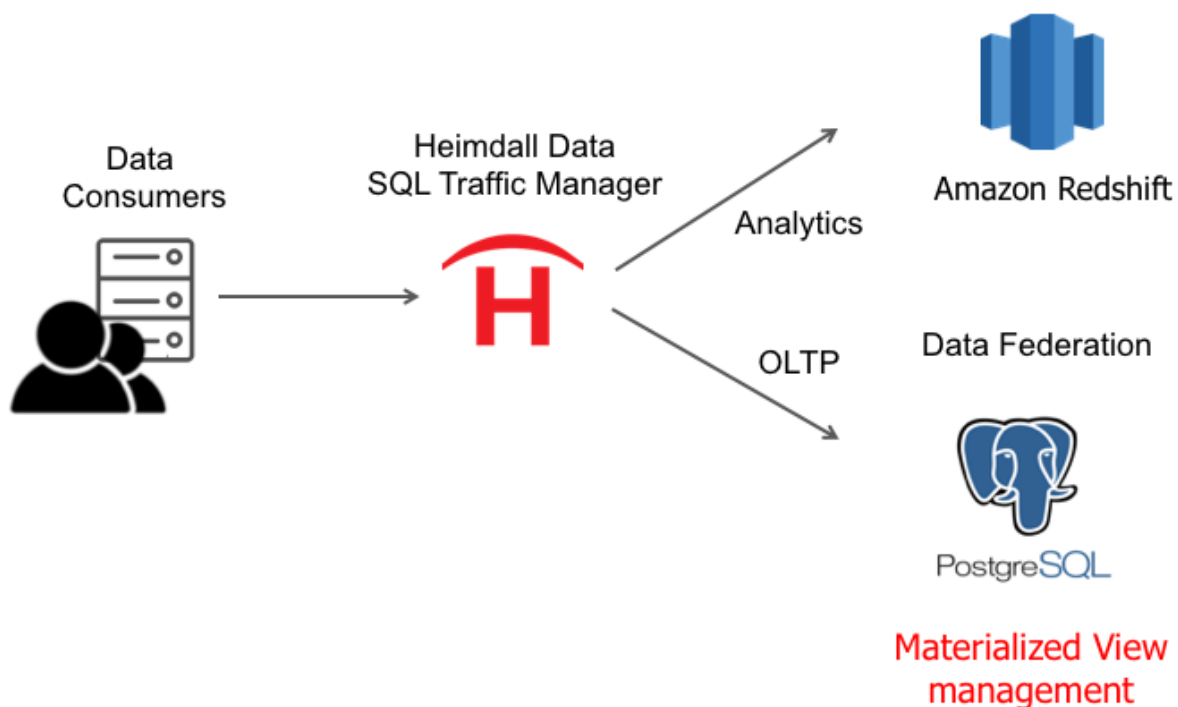
If there are concurrent writes and reads against the same table, on the same thread, as everything will just block until the DML operation is completed anyway.

If an exception occurs in the transaction, Heimdall determines which query resulted in the transaction, and will remove the query from the list, report the exception in our logs, and reprocess the batch without the query in place.

Hybrid Transactional Analytical Processing (HTAP) for Amazon Redshift



Fast materialized views for results for Amazon Redshift

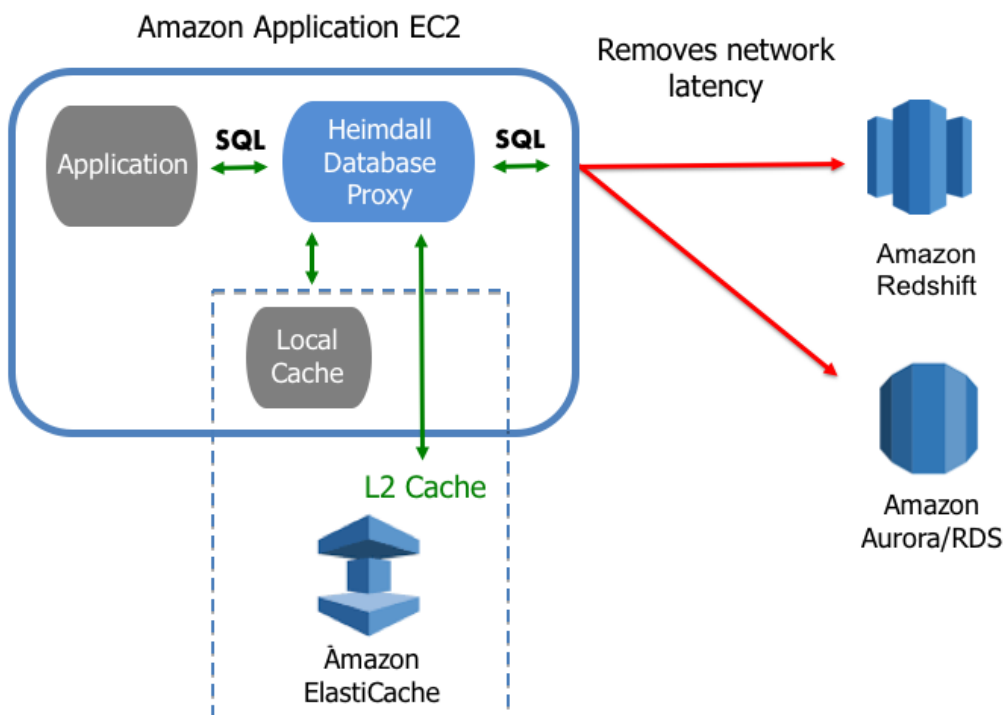


Fast materialized views are very important in analytics environments. When reports are generated, a subset of data is pulled from the back-end data store, then various operations are performed on that data. Heimdall provides the following functionality:

- Queries against a materialized view can be routed to an alternate database, typically Postgres, which acts on behalf of Amazon Redshift. Postgres answers queries offloading Amazon Redshift.
- Heimdall triggers a refresh of the view automatically. Heimdall is aware updated views from Amazon Redshift and when data was loaded that may impact the view. The net result is faster reports and a lighter load on Redshift, allowing the processing of other queries to be faster and more scalable.

Automated SQL Caching for Amazon ElastiCache

Hybrid Transactional Analytical Processing (HTAP) for Amazon Redshift



The fastest query is the one that does not have to be executed. Heimdall's intelligent auto-caching and auto-invalidation works together with Amazon Redshift's query caching, but in the EC2 application tier, removing network latency. This distributed architecture allows caching to be scalable, while acting as one cache cluster. Result sets are cached in tandem from local memory and Amazon ElastiCache and are invalidated upon writes to the table. Best of all, Heimdall deployment requires zero code changes.

The Heimdall architecture was designed for ease of deployment without the need to modify the application or database. Configuration changes are updated at runtime without restarting the application. The net result is a platform that can be updated with 100% application uptime.

Heimdall Data is an AWS Technology partner providing SQL visibility and performance optimization as a Database Proxy. Available in the AWS Marketplace, Heimdall is transparently integrated with Amazon Redshift, Aurora, RDS, ElastiCache and CloudWatch.

For more information about how Heimdall Data can help improve your AWS environment, contact us at info@heimdalldata.com.