Real-World Performance 2018

HVR Real-World Performance

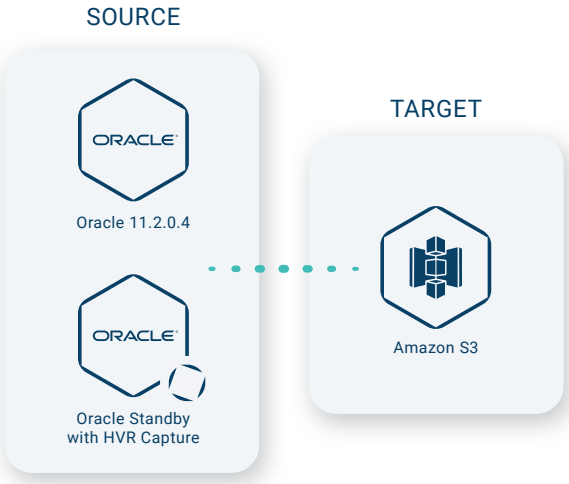
TESTING MAX LOAD ON YOUR SYSTEMS

Occasionally organizations will aim to find the limits of what the technology can achieve with the ability to generate a representative load up to the maximum level their configuration can handle. This paper discusses such case.

To ensure organizations feel comfortable with the performance and impact on the systems we always recommend customers and prospects to run a representative test in their environment to experience what the load would be for their implementation.

Performance Test

Below is a real-world performance scenario for one of HVR's highest volume customers based on a single Oracle Database source.



CAN DATA REPLICATION TECHNOLOGY KEEP UP WITH PRODUCTION WORKLOAD?

A common question prospects and customers ask is whether data replication technology can keep up with the production workload, and what resources (CPU, memory, and to a lesser extent IO and storage space) are required to keep up. The answer to such question always starts with "it depends" due to the specific nature of every workload. Variables that influence the answer include:

- The number of tables to be replicated
- Data types involved
- The transaction mix (concurrency, small frequent transactions versus longer, larger transactions)
- The percentage of the transaction log containing information that must be replicated, and more.

SYSTEM CONFIGURATION

- 4 x Intel® Xeon® Processor E7-8894 v4 (96 cores total, 192 threads)
- 512 GB RAM.
- 10 TB of SSD storage for database transaction logs and HVR transaction files
- ASM storage for Oracle Database configured for maximum throughput at 10 GB/s

SETUP

Source: Oracle Database 11.2.0.4 8-node RAC running Oracle eBusiness Suite in single instance standby Active Data Guard setup, with supplemental logging enabled on all columns for tables included in the replication. HVR capture runs on the standby database capturing and merging changes from all 8 RAC threads.

Target: S3 file system

- Tab-delimited files, no enclosures, backslash as escape character, and substitution character sequences for characters that would otherwise be hard to process (e.g. carriage return, new line, and others).
- Using client-side encryption with KMS i.e. using server CPU to encrypt the data.
  Not using compression.
- Only after-row image is published unless the row identifier changed in which case both before and after image are published.
- Extra metadata columns are populated on the target:
    - commit timestamp
    - database user performing the change
    - unique integration key (36 bytes)
    - operation type (single byte to indicate insert, update before/after, or delete)
- Changes are pushed once per minute to limit the number of files created.
- File names include:
    - table name
    - commit timestamp down to the minute (to enable consistent data publication per minute of processed data)

Performance Metrics – Refresh

Refresh performance for the largest table XLA_DISTRIBUTION_LINKS:

3289 *BYTES*	Max Source Row Size
68K ROWS / *SECOND*	Single Data Load Performance (includes KMS client-side encryption)
20+ MB / *SECOND*	Data Load into S3 (after encryption, based on actual row size, plus additional metadata)
100% CPU	During the load the single HVR process is running at 100% CPU (KMS encryption introduces ~50% overhead).
12 BILLION ROWS	With ~12 Billion rows in the table XLA_DISTRIBUTION_LINKS this table is refreshed in parallel (as well as the other top 30 biggest tables).

Performance Metrics – Log-Based Change Data Capture (CDC)

Redo log generation on a typical day is 1-2 TB, with 3+ TB on a busy day. HVR keeps up with redo log generation to ensure the system stays well within the 30 minute SLA. Typical end-to-end latency is typically between one and two minutes because changes are pushed at a one-minute interval (to limit the number of files that get created).

HVR can keep up with the current peak volume and anticipated growth (30-40% growth YOY) extensive testing was performed on key tables in the Oracle eBusiness Suite Database.

Table name	Hourly production volume in millions of rows per hour (max)	HVR CDC speed in millions of rows per hour (max)
GL_JE_LINES	40	114
AP_INVOICE_DISTRIBUTIONS_ALL	13.5	38
AP_INVOICES_ALL	3	47.5
AP_UNSELECTED_INVOICES_ALL	1.4	308
RA_CUSTOMER_TRX_LINES_ALL	1.35	86
RA_CUSTOMER_TRX_ALL	1.9	50

Conclusion

HVR refresh (initial load) performance can be scaled out by running in parallel.

The performance of HVR's CDC solution is more than sufficient for current and projected loads. In an instance in which faster CDC performance is required, then a setup with capture per thread could be used to skip the resource-intensive step of merging changes from all threads. Such setup introduces extra complexity and requires data consistency enforcement downstream.

ABOUT HVR

We accelerate data movement so that you can revolutionize your business. HVR is designed to move large volumes of data FAST and efficiently in modern environments for real-time updates.