Actian Hybrid Data Conference

Introduction to HVR

Zulfikar Qureshi, Solution Architect

November 2017
Agenda

- Introduction
- HVR Overview
- Technology
- Use Cases
- Case Studies
- Q&A
Deploying Around the World

- > 150 CUSTOMERS
- > 1500 DEPLOYMENTS
- > 30 COUNTRIES
Why HVR?

- **Complexity**
  - Heterogeneity
  - Multi-Directional Movement
  - Conflict Detection
  - Data Validation
  - Reporting and Monitoring

- **Scalability**
  - Move high volumes of data
  - Distributed Architecture
  - Easy to add sources/targets

- **Performance**
  - Fast
  - Bulk-compression mode
  - Network efficient
  - Log-based Change data Capture

HVR customers continually process over one terabyte of transaction log data/day/database
TECHNOLOGY

Overview
HVR Distributed Architecture

Initial Load (Refresh)
Incremental Changes (Capture & Integrate)
Data Validation (Compare)

Source

Efficient Secure
Hub
Filter, Transform, Schedule & Monitor

Target

Efficient Secure
Agent

File
salesforce

RDBMS

SAP HANA

kafka

DWH

RDBMS

File
salesforce
HVR Hub

- It’s part of the small download and controls all endpoints.
- It’s where log, checkpoint, runtime, and transaction files will be automatically maintained.
- It requires access to a database to store configuration info.

The source, target or a middle-tier system can be designated as the HVR Hub.
If a database involved in replication is not located on the hub machine, HVR can connect to it in two ways:

**Via the network database's own protocol (e.g. Ingres/Net)**

Remote HVR connections require the same small installation as the Hub but will only be running the HVR Agent, sometimes called the HVR remote listener. It does not require a license file.

**Via an HVR remote (relative to the Hub) connection (recommended).**
Ease of Manageability for Complex Environments

Peer-to-Peer vs Central Hub Architecture
Topologies

**UNI-DIRECTIONAL**
Real-Time Reporting
Migrations (reverse post migration)

**BI-DIRECTIONAL**
Active / Active Standby
High Availability

**PEER-TO-PEER**
Multi-way Active / Active
Geographical Distribution

**BROADCAST**
Data Distribution

**INTEGRATION / CONSOLIDATION**
Data Warehouse / Data Lake

**CASCADEING**
Data Marts
Real-Time Capture Sources

Files: Avro, CSV, HDFS, S3, XML
Destination Support

- Microsoft Azure
- Microsoft SQL Server
- Microsoft SQL Azure
- Kafka
- Hadoop
- AWS Marketplace
- ORACLE
- MySQL
- INGRES
- HBase
- SharePoint
- MariaDB
- PostgreSQL
- Teradata
- ParAccel
- IBM DB2
- Greenplum
- Amazon Redshift
- Salesforce
- Files: Avro, CSV, HDFS, SS, XML
- Vector
Real-Time Change Data Capture

- **Real-time**: Are not scheduled source table Selects or triggers

- **Complete**: Capture all changes including Deletes

- **Low impact**: Reads all table rows that were affected by single transaction from binary database recovery files

- **Access to metadata**: Capture transaction ID, who made the change, commit timestamp, DDL changes and more
Data Validation and Repair

- Protect against human error
- Validate the data before you complete your migration
- Compare against mixed technologies
Monitoring and Alerting

- Real-time monitoring of HVR processes
- Automatic alerting and notification
  - Email, SNMP, Slack
- Integration with enterprise monitoring solutions
- Report on replication statistics

<table>
<thead>
<tr>
<th>Job</th>
<th>State</th>
<th>Retries</th>
<th>Recent Error</th>
<th>Latency</th>
</tr>
</thead>
<tbody>
<tr>
<td>orcl2ms-cap-orcl</td>
<td>SUSPEND</td>
<td></td>
<td></td>
<td>42m 5s</td>
</tr>
<tr>
<td>orcl2ms-cmp-orcl-mssql</td>
<td>PENDING</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>orcl2ms-integ-mssql</td>
<td>SUSPEND</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Job</th>
<th>State</th>
<th>Retries</th>
<th>Recent Error</th>
<th>Latency</th>
</tr>
</thead>
<tbody>
<tr>
<td>orcl2ms-cap-orcl</td>
<td>RUNNING</td>
<td></td>
<td></td>
<td>1s</td>
</tr>
<tr>
<td>orcl2ms-integ-mssql</td>
<td>RUNNING</td>
<td></td>
<td></td>
<td>3s</td>
</tr>
</tbody>
</table>
All in One Box

Less than 100MB - our streamlined download* includes everything:

- Comprehensive platform support, secure, high speed real-time database, big data and file replication
- Heterogeneous initial loads, schema creation, auto mapping
- Heterogeneous DDL replication
- Database data validation and repair
- Graphical real-time monitoring, alerting and scheduling

Learn once, use everywhere
HVR Use Cases

Overview
# What We Help You Achieve

## Real-Time Data Integration at Scale

<table>
<thead>
<tr>
<th>Data Lake/Warehouse</th>
<th>Cloud</th>
<th>Real-Time Reporting</th>
<th>Geographical Distribution</th>
</tr>
</thead>
</table>

Deliver **continuous operations** for mission-critical applications

Improve **efficiencies** through real-time data distribution for better application performance

Reduce **risk** by ensuring data integrity while improving availability between sources and targets

**Better business insights** by enabling a consolidated, real-time Data Warehouse or Data Lake on a performance-optimized environment
HVR Use Case

Data Lake/Warehouse
Real-Time DWH/Data Lake

- **Solution**
  - Data consolidation from various heterogeneous environments
  - Optimized, parallel data integration
  - Supports multiple Big Data platforms
  - File replication

- **Benefits**
  - 360 degree view of the data
  - Scalable data retrieval
  - Low-impact and real-time onboarding of data
HVR Use Case

Cloud
Cloud Hybrid Integration

- Available where you need it
  - On AWS and Azure Marketplace
- Always in sync – low latency
  - Non-intrusive real-time log-based change data capture
- Secure
  - Encrypted network traffic
- Low migration risk
  - Synchronize on-premises and cloud for testing and initial setup
  - Synchronize cloud and on-premises as a fallback option
  - Inter-cloud data integration
HVR Use Case

Real-Time Reporting
Real-Time Reporting

**Solution**

- Non-intrusive log-based CDC from major mega vendors
- Filter on data that is only required for reporting
- Delivery to low costing platform
- Optimized target structures for faster reporting

**Benefits**

- Access to real-time updated data to make quicker decisions
- Remove vendor lock-in and reduce license costs
HVR USE CASE

Geographical Distribution
Geographical Distribution

- **Solution**
  - Log-based CDC and integration at multiple sites
  - Efficient network utilization
  - Active-Active with conflict detection and resolution
  - Filter data for different requirements
  - Centralized design, monitoring and manageability

- **Benefits**
  - Data available where you need it, when you need it
  - Real-time data, synchronized for 24x7 business continuity
HVR Use Case

High Availability
High Availability and Disaster Tolerance

- **Solution**
  - Committed transactions are only replicated to the target
  - Active-Active to fully utilize DR and load balance workload
  - Conflict detection and resolution to support active sites

- **Benefits**
  - No distance restrictions
  - Different database versions
  - Different database vendors
  - Target is fully read/write
HVR USE CASE

Zero Downtime Migrations/Upgrades
Zero Downtime Migrations

**Solution**

- Heterogeneous initial load and table creation for instantiation of target
- Change data capture to keep source systems on-line
- Mitigate risk with failback option
- Data validation and repair to verify data consistency
- Active-Active for phased switchover and keep both systems synchronized

**Benefits**

- Reduce and even eliminate downtime
- All in one solution
Migration Deployment Options

**Uni-directional**
- Minimize downtime by keeping source on-line
- Allows for extended testing period
- Downtime incurred only when switching users to new system
- Data validation before switchover

**Uni-directional with Failback**
- Minimize downtime by keeping source on-line
- Allows for extended testing period
- Failback to old system to mitigate risk of failure on new system
- Downtime incurred only when switching users to new system
- Data validation before switchover

**Active-Active**
- Eliminate downtime and keep both systems on-line
- Allows for extended testing period
- Phased migration approach, move users across in batches
- Account for conflicts and resolution
- No downtime incurred
- On-line data validation
Objective: Distribute flight planning data

Critical: airplanes don’t take off without it

- Over (60) Airlines use LIDO, including, China Airlines, Emirates, British Airways, Air France, UPS

Secure and reliable: Confidential data cannot be shared

Customer Benefits

- Global real-time data replication
- Complex routing and filtering
Case Study | GE Corporate Data Lake

- **Objective:** Cross-divisional Data Lake
- Dozens of mostly ERP sources
  - SAP on Oracle and DB2i
  - Oracle Applications
  - Industry-specific applications
- Target Greenplum *(could be Vector)*

**Customer Benefits**
- Efficient and flexible in a complex setup
- Performs well for MPP target

<table>
<thead>
<tr>
<th>Industry</th>
<th>Data Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Others</td>
<td>Data Source</td>
</tr>
<tr>
<td>Transportation</td>
<td>Data Source</td>
</tr>
<tr>
<td>Aviation</td>
<td>Data Source</td>
</tr>
<tr>
<td>Oil &amp; Gas</td>
<td>Data Source</td>
</tr>
<tr>
<td>Power</td>
<td>Data Source</td>
</tr>
</tbody>
</table>
Objective: Moving Data for Real-Time Reporting and Analytics

Source: 40+ Oracle systems; 8i > 12c

Single connection into AWS using HVR proxy on-premise

Target: Multiple uses; S3 for historical data, EMR for analytical purposes, and Oracle for downstream reporting

Customer Benefits

Easy to implement and manage

High performance loading into S3
All in one Box - Summary

Less than 100MB - our streamlined download* includes everything:

- **Comprehensive**
  platform support, secure, high speed real-time database, big data and file replication

- **Heterogeneous initial loads**, schema creation, auto mapping

- Heterogeneous **DDL replication**

- **Database data validation and repair**

- **Graphical real-time monitoring, alerting and scheduling**

Learn once, **use everywhere**
QUESTIONS
And Answers
GET FREE TRIAL

https://www.hvr-software.com/free-trial/