COMPARISON REPORT

SingleStore vs Oracle

2019

SingleStore

Table of Contents

Summary

Oracle is a legacy architecture with separate transactional, operational, and data warehouse requirements resulting in data duplication, complexity, and maintenance overhead. Customers running Oracle Exadata or RAC may struggle with growing scale and cost challenges. Legacy platforms like Oracle are also not suited to meet modern requirements for businesses that are on the path to digital transformation.

SingleStore, a shared-nothing natively distributed architecture can ingest data and run analytics on a single platform, without compromising on performance, and while eliminating legacy platform issues with data duplication, complexity, and high maintenance overhead. SingleStore can scale-out efficiently and quickly respond to growing workloads leveraging commodity hardware without add-ons or specialized tuning expertise. SingleStore's modern cloud native architecture is capable of handling both OLAP and OLTP workloads in a single system.

Following key features and performance improvements added to our column store technology in v7.0 release for SingleStore delivers row-store like capabilities in our column store, resulting in dramatic cost savings:

- Performance improvements for bulk inserts, large insert/selects, LOAD DATA, and pipelines to columnstore tables (particularly for wide columnstore tables).
- Much improved performance of synchronous replication and durability
- Queries on columnstore tables can now reorder filters to decrease execution time. Reordering occurs automatically and allows filters that are more selective to be evaluated first.
- Hash indexes on columnstore tables are now supported. CREATE TABLE and ALTER TABLE now support adding and dropping hash indexes on columnstore tables.

- Column statistics are now automatically gathered on rowstore tables. This is in addition to column statistics being automatically gathered on columnstore tables, which was done previously.

SingleStore Overview

SingleStore is a distributed, highly-scalable SQL database that can deliver maximum performance for transactional and analytical workloads with familiar relational data structures. SingleStore is built on a cloud-native data platform designed for today's most demanding applications and analytical systems.

Architecture

Designed for high performance general purpose database workloads including operational reporting, ad hoc queries, real-time analytics, and AI/ML applications

- Designed for high performance general purpose database workloads including operational reporting, ad hoc queries, real-time analytics, and AI/ML applications
- In-memory transactional storage for low latency, high throughput queries
- In-memory row store for OLTP and highly efficient on-disk column store for OLAP
- Memory-optimized skip list indexes combined with full Multi-Version Concurrency Control (MVCC) delivers durable, lock-free streaming ingest performance
- Lock-free data structures leveraging skip list indexes with dynamic SQL interpretation accelerates query execution
- ACID compliant transactions
- Integrated transactional storage engine for low latency and high throughput queries (read + write)
- Execute SQL queries utilizing 100s of nodes, 1000s of cores, TBs of memory and 100s of TBs of disk for interactive analytics
- In SingleStore 7.0 release, two new capabilities move SingleStore much further to handle system of record (SoR) transactional workload: by delivering very fast *synchronous replication* – including a second copy in the initial write operation,

atomically – and *incremental backup*, which offers increased flexibility and reliability

- In SingleStore Phase 1, shipping as part of SingleStore 7.0 (currently in beta), rowstore tables get null compression, lowering TCO in many cases by 50% and columnstore tables get seekable columnstores, which support fast seeks and updates, giving columnstore tables many of the performance and usability features of rowstore tables

Deploy Anywhere

- SingleStore is available as a fully-managed cloud service. Our on-demand, elastic cloud database-as-a-service offering deployed across public cloud providers and in availability zones around the world
- SingleStore is a fully cloud-native database that supports on-premise and multi-cloud/hybrid cloud architectures
- Hybrid and private cloud deployments of SingleStore are possible through our certified Kubernetes operator
- SingleStore can be deployed on a variety of hardware for maximum flexibility in on-premises infrastructure

Query Optimization

- Mature query optimizer performs complete TPC-H and TPC-DS query plans without extensions or customizations
- Optimized queries including compiled queries and single instruction multiple data (SIMD) deliver highly efficient performance
- Code-generation in the query processor performs fast repeat queries leveraging fewer CPU instructions

Performance

- Millions of writes per second with full durability
- Millions of low latency reads a second
- Fast, scalable standard ANSI SQL support

- Breakthrough performance proven through analytical benchmarks:
    - SingleStore 6.7 Performance Summary
    - SingleStore shatters trillion rows per second barrier
- Optimized data layout delivers low latency queries by avoiding data movement across the network

Data Ingestion

- Integrated pipelines deliver streaming ingest with scalable parallel continuous data loading from other distributed systems (S3, HDFS, Kafka), and supports data in a broad variety of formats (CSV, JSON, Avro, Parquet)
- Continuous ingest into columnstore tables leverages an in-memory rowstore to maximize throughput and scale for streaming ingestion into the columnstore

High Availability and Fault Tolerance

- Transparent high availability with disaster recovery
- Self healing functions detect node failures within seconds to keep clusters online and avoid query failures
- Failed nodes automatically recover when they become available and are rejoined to the cluster to begin serving queries
- Online rolling upgrades
- Flexible data type storage across JSON, geospatial, and full-text

Security

- Enterprise grade security with role based access control(RBAC), authentication (PAM, SAML, GSSAPI/Kerberos), administration separation of concerns, SSL
- Full audit logging for tracking user activity
- Strict mode setting separates administrator controls from gaining access to data

Cluster Operations

- Dynamic Scalability
- Integrated backup and restore
- CPU and Memory resource governor optimizes performance under extreme load and concurrency
- Add/Remove new compute nodes to the cluster as needed
- Add/remove/rebalance nodes while cluster is online
- Online ALTER TABLE (fully parallel add index, add column, etc.)

Ecosystem Support

- MySQL wire protocol allows many mysql applications to run against SingleStore without any changes or with minimal changes (MySQL Client drivers, SQL IDEs, ETL tools, dashboards, etc.)
- SingleStore supports a variety of analytics & business intelligence tools
    - Tableau
    - Microstrategy
    - Looker
    - Zoomdata
    - Informatica
- SingleStore supports all of the popular streaming and CDC platforms, including Apache Kafka and Apache Spark

Comparing Oracle to SingleStore

Oracle Database was first released in 1979 and is currently in its nineteenth version. Oracle Database was a leader in the relational database market with respect to structured data. However, it is not designed or able to manage semi-structured and unstructured data well enough to deliver on today's requirements, such as systems of engagement.

I - Legacy Architecture

- Traditional Oracle databases do not support mixed workloads (such as concurrent transactions and analytics) well. As a result, most deployments of Oracle in the enterprise space have a clear separation between transactions and analytics. This separation adds to the overall cost and complexity of Oracle data platforms.
- Oracle also does not integrate with data sources such as Kafka, S3, or Hadoop as an out-of-the-box function. Most such functionality is licensed separately. This results in additional cost and time delays to integrate Oracle within any ecosystem that has both transactional and modern data sources.

II - Performance and Scalability Limitations

- Oracle was designed and developed on the computers of the 1970s and 1980s - standalone, single-processor systems with limited memory. As a result, Oracle is not built from the ground up as a scale-out database.
- Oracle has since released a scale-out version of the database with a feature called Real Application Clusters (RAC). RAC uses a shared storage model, which limits horizontal scale-out. Instead, vertical scale-out tends to be the predominant model. Vertical scale-out has several limitations. Database software and the operating system has to scale to take advantage of the increased CPU and memory on the system. Tall vertical systems needed for large and demanding workloads are also non-standard hardware models requiring additional data center considerations.
- Oracle has released an appliance, Exadata, to address some of the scalability and performance issues that are not met by RAC. However, this is highly expensive, and adds complexity and maintenance challenges to the overall platform. And because Exadata is an appliance, it is proprietary as well.
- Oracle has released an in-memory option with 12C (IMDB) and higher versions in order to provide capabilities including columnar performance on queries and

address some of Oracle's limitations on analytics. This option is an add-on and licensed separately. Being a bolt-on, it has several restrictions on the size of the data, and it requires maintaining multiple copies of data in synchronization.

III - Not Cloud-Friendly nor Elastic

- With regards to deploying in the cloud, Oracle has limited options with RAC deployments (only supported on Oracle Cloud). Single-instance databases are otherwise supported in the cloud.
- Oracle RAC is not elastic-friendly by design. Most enterprise Oracle environments are always-on cost.

IV - AI/ML Capabilities as an Additional Add-On and Licensed Separately

- AI/ML capabilities are available as part of Oracle Data Mining, included with the Oracle Advanced Analytics Package. Oracle Data Mining allows customers to build, train and deploy models.
- This is a separate licensed component and requires Oracle Enterprise Edition.

V - Management and Support Challenges

- Oracle has generally required a dedicated DBA team to install, maintain, and manage. It is not DevOps-friendly. A typical RAC installation requires a team of system, storage, and database administrators to setup the environment prior to installation. An install process can take several hours depending on configuration. Exadata systems are even harder and more expensive to set up and manage.
- Single-instance Oracle systems are easier to install; however, due to the complexity of Oracle, even single-instance systems require dedicated and well-trained resources to manage them.
- Performance optimization and troubleshooting require specialized skills that tend to be expensive, especially expensive especially for complex workloads.

VI - Cost Inefficiencies

- With Oracle, enterprise features are licensed separately. The Oracle licensing model can be considered restrictive, wherein an end user is required to track which features are used, and by whom, in order to comply with the licensing agreement. Oracle licensing audits are a tactic used by Oracle to pressurize users to comply with licensing restrictions or pay for additional licensing options.

- Scalability requires purchasing Oracle RAC licenses or upgrading to Exadata. RAC workloads require hardware to be Oracle-compliant, as these workloads use shared storage (generally a storage area network) and require specialized skills to install and manage.

Feature List

Architecture	SingleStore	Oracle
Distributed Shared Nothing Scale out Architecture	●	○
Transaction Management		
- Multi-statement transaction	●	●
- Lock-free synchronization	●	○
- Multi-version concurrency control	●	●
- SELECT FOR UPDATE	●	●
Deployment flexibility (Multi-Cloud and On-Premises)	●	◑
Compressed columnar indexes for OLAP	●	◑
Memory optimized row-store for OLTP/real time applications	●	◑

Query		
SQL-92	●	●
SQL-99 OLAP Extensions	●	●
SQL-2003 extensions	●	●
Vectorization and single instruction, multiple data (SIMD)	●	●
Bloom filter pushdown for hash join	●	●
Local join support (hash, merge, nested loop)	●	●

Distributed join support (broadcast, reshuffle, centralized join, push down join, reference table join)	●	●
Optimization and Statistics	●	●
Adaptive query compilation and execution	●	○
Interpreted compiled to machine code	●	○
Native Just-in-time (JIT) compilation	●	○
Compiled execution against in-memory data for 10x performance gain	●	◑
Storage		
Rowstore In-memory	●	◑
Columnstore compressed	●	◑
Sharding	●	◑
Database replication across geographies to enable disaster recovery	●	●
Temporary tables	●	●
Data types		
JSON	●	●
Relational	●	●
Full text	●	●
Geospatial	●	●
Performance		
Compile queries to low level machine code for real time analytics	●	○
Performance benchmarking results in milliseconds	●	◑
In-built query optimizer for best execution plan	●	◑
Ingest and transactional performance	●	◑

Ingestion		
Native pipeline for data ingestion	●	○
LOAD DATA: One-time Bulk Load	●	●
Native Parallel Ingest from many data sources (Linux File System, S3, Azure Blob Store, HDFS, Kafka)	●	◑
Supports most popular formats (CSV, Avro, JSON)	●	●
Native internal transforms	●	○
Native pipelines to stored procedures/pipelines with transform scripts	●	○

Cluster Operations		
Rolling production updates	●	
Data centre replication	●	●
Automated full backup	●	●
Point-in-time recovery	●	●
Query while ingest	●	◑
Snapshot	●	●
High write throughput	●	◑
Transactional	●	◑
Management and Monitoring tools	●	●
Online operational capabilities (add/remove nodes/re-balance etc.)	●	●
Automated workload management	●	●

Security		
Auditing	●	●
Encryption	●	●
RBAC	●	●

Authentication (GSSAPI/Kerberos)	●	●

Ecosystem		
Looker	●	●
Zoomdata	●	●
Tableau	●	●
Streamsets	●	●
SAS Access	●	●
Informatica	●	●
Data Virtuality pipes	●	●
Talend	●	●
Power BI	●	●

Client drivers		
JDBC	●	●
OBDC	●	●