

SQream DB on PowerEdge R940xa

Abstract

SQream DB on Dell EMC PowerEdge R940xa offers complete integration of computing, networking, and storage resources, and provides high performance storage and advanced BI analytics with scalable and NVMe storage.

August 2019

Revisions

Date	Description
August 2019	Initial release

Acknowledgements

This paper was produced by the following members of the Dell EMC SIS team and SQream:

Author: Arnon Shimoni [SQream], Bhavesh Patel [Dell EMC]

Support: Nihit Bhavsar [Dell EMC]

The information in this publication is provided “as is.” Dell Inc. makes no representations or warranties of any kind with respect to the information in this publication, and specifically disclaims implied warranties of merchantability or fitness for a particular purpose.

Use, copying, and distribution of any software described in this publication requires an applicable software license.

© August 2019 Dell Inc. or its subsidiaries. All Rights Reserved. Dell, EMC, Dell EMC and other trademarks are trademarks of Dell Inc. or its subsidiaries. Other trademarks may be trademarks of their respective owners.

Dell believes the information in this document is accurate as of its publication date. The information is subject to change without notice.

Table of contents

- Revisions..... 2
- Acknowledgements..... 2
- Executive Summary 4
 - SQream DB Overview 4
 - CHALLENGES WITH MODERN DATA WORKLOADS..... 4
 - How does SQream DB address these challenges? 4
- What are some of the use cases? 6
- What does the solution look like? 7
 - DELL EMC POWEREDGE R940xa 7
 - SQREAM DB: A GPU-ACCELERATED DATA WAREHOUSE 8
- How does SQream DB perform on R940xa?..... 9
 - R940xa Configuration Details..... 9
 - Performance 9
- CONCLUSION 10
- FOR MORE INFORMATION 11

Executive Summary

SQream DB on Dell EMC PowerEdge R940xa offers complete integration of computing, networking, and storage resources, and provides high performance storage and advanced BI analytics with scalable NVMe storage. Validated with industry-accepted benchmarks, the SQream DB and Dell EMC PowerEdge R940xa integrated analytics platform is a flexible performance-optimized system that reduces integration and deployment risk and provides a predictable customer deployment.

SQream DB Overview

SQream DB is ANSI SQL compliant, and is tightly integrated with industry-standard tools for ETL, BI, and more. SQream DB also works directly with Hadoop framework Parquet files for fast loading at over ten million records per second (or 3.3TB/hour) with a single GPU.

SQream DB on Dell EMC PowerEdge R940xa provides uncompromising performance and reliability. In industry-accepted tests, the SQream DB solution loaded 7.5 million records per second, translating to a throughput of over 3.3TB per hour with a single GPU. The joint system provides shared accelerated NVMe storage that delivers high-throughput performance in one box, eliminating the need to replicate data across BI servers for data locality and availability.

The SQream DB analytics data warehouse is designed to perform fast, flexible, and cost-efficient analysis for huge data stores, empowering data analysts and data scientists, enabling and in fact encouraging them to broaden their ad-hoc and analytic queries.

CHALLENGES WITH MODERN DATA WORKLOADS

Many organizations today experience some issue with their data warehouse. Either the data warehouse cannot support their data workloads, or there is just too much data to affordably store. With no trivial solution, many companies are cutting down on their data pipelines. Each level of technology further limits the flow of information, which reduces the return on investment of storing the data in the first place. Unfortunately, this phenomenon sends BI and data scientists the message, “Don’t run complex ad-hoc queries or expand the timeframe of analysis – because our current system won’t keep up, and each insight will be expensive to our bottom line.” When success and growth are intrinsically linked to data, and information and insights from this data are the new gold – modern data analytics solutions must adapt to the changing business demands.

How does SQream DB address these challenges?

Created to empower data analysts and data scientists, SQream DB is a purpose-built accelerated data warehouse designed for complex big data workloads. At its core, SQream DB is a fast SQL analytics database. SQream DB adopts the ANSI SQL standard and augments it with useful functionality such as window function analysis, live aggregations, ultra-fast

mathematical operations, and joining on any number and types of keys. These functions have been fine-tuned to handle a wide range of complex big data queries, across several industries. These capabilities enable actionable insight and immediate action. The ability to ask any kind of question about a wide variety of information while it is still relevant and valuable distinguishes SQream DB from other solutions, especially when the data produced is in the terabyte to petabyte range.

Some of the highlights are:

- Enabling and empowering BI analysts and data scientists by treating the root cause of data warehousing problems first.
- Companies should have the freedom to decide how much capacity they require, both in storage and in compute. SQream DB detaches the traditional relationship between storage and compute, allowing for independent scaling without affecting the database, and with little to no downtime.
- For the highest return-on-investment, a data warehouse should allow any user to query and explore the data at any time, while remaining cost-effective. Traditional data warehouses don't compress data, and force DBAs to prepare cubes, projections, indexes, partitions, distribution keys, and materialized views in order to perform adequately. In contrast, SQream DB's high-throughput architecture and best-in-class compression lets you focus on your data rather than the infrastructure. Just load the data, and it's ready to be queried immediately. No pre-computations necessary. SQream DB's transparent and automatic metadata collection and always-on adaptive GPU compression takes care of the rest, enabling BI analysts and data scientists to query any piece of data immediately.

What are some of the use cases?

The SQream DB analytics database is successfully deployed in the following use cases for telecom analytics, financial services, IoT, marketing and ad-tech, retail, and other industries.

- Customer experience improvement: Understand where customers are, what they do, how to address them, and how they behave. Use this understanding to reduce churn, improve QoS, and upsell with relevant, targeted offers.
- Risk and fraud detection, prevention, and management: Meet regulatory demands like FRTB and Basel IV by detecting, preventing, and managing fraud with analysis on large historical and fresh datasets.
- Self-service BI: Let data consumers and even customers access data, generate reports, get near real-time alerts, and explore their own historical data.

What does the solution look like?

SQream DB on Dell EMC PowerEdge R940xa is an integrated advanced analytics solution that helps organizations make use of their full data stores. The power of the joint solution comes from its ability to deliver complex analytics to users from any part of the organization, regardless of query complexity and data scale.

DELL EMC POWEREDGE R940xa

The Dell EMC PowerEdge R940xa is a four-socket enterprise server designed to drive GPU database acceleration for business-critical applications with a large capacity internal storage. The 4U unit is flexible and adapts to changing business demands. Its large internal storage gives you plenty of space to grow as your databases increase in complexity and size.

PowerEdge R940xa brings out SQream DB's full potential, with up to 112 CPU processor cores, 4 NVIDIA Tesla GPUs, and fast NVMe-based storage. The PowerEdge R940xa system allows SQream DB to saturate GPU and CPU compute for the high-throughput computations common with data science, BI reporting, and ad-hoc analytics on very large data.



Figure 1 Front and rear views of the PowerEdge R940xa

In the R940xa each CPU has direct PCIe connectivity to a single GPU which results in extremely fast data processing in all 4 CPU-GPU subsystems and minimizes latencies to a level that was impossible to achieve earlier.

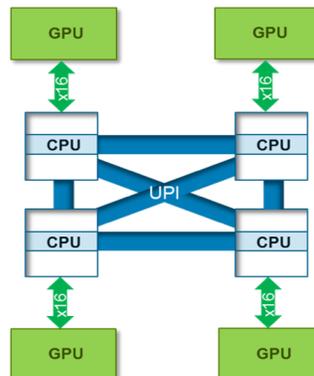


Figure 2: CPU: GPU High Level block diagram

SQREAM DB: A GPU-ACCELERATED DATA WAREHOUSE

SQream DB is an enterprise-ready solution for deploying analytics on Hadoop data. SQream DB provides data professionals a revolutionary way to explore and rapidly analyze their vast data.

- **Reduced requirement for data preparation.** SQream DB runs on any structured dataset in its raw form, reducing the need for indexing, pre-aggregations, etc..
- **Nested GPU-accelerated JOINS** with any number of tables and keys enable data consumers to **correlate huge amounts of data**.
- **Dynamic workload management** supports scheduled and ad-hoc analytics workloads from a variety of stakeholders, including dedicated performance for the most important users and processes.
- **Fast loading and querying.** SQream DB helps ensure high query concurrency while simultaneously loading new data, at up to 3.3TB/hour.
- **Columnar engine and compression.** SQream DB lets you gain significant advantage for I/O, performance, and storage by efficiently storing data in a columnar, compressed format.
- **Auto-tuning database.** SQream DB requires little to no maintenance. Most operations are completely automatic (with overrides for professionals), so data consumers can simply load and query data immediately.

SQream DB and Dell PowerEdge R940xa storage reference platform consists of the following components:

1. One Dell EMC PowerEdge R940xa with 4 CPUs and 4 NVIDIA Tesla GPUs
2. Expansions provided by compute-only Dell C4140 1U blades with 4 NVIDIA Tesla GPUs and NVLink
3. ToR 10GigE ethernet switch for the BI fabric
4. 100GigE or InfiniBand switches for the storage fabric for expansion



Figure 3: SQream DB on Dell PowerEdge R940xa reference architecture featuring additional PowerEdge C4140 compute nodes for scale-out

How does SQream DB perform on R940xa?

R940xa Configuration Details

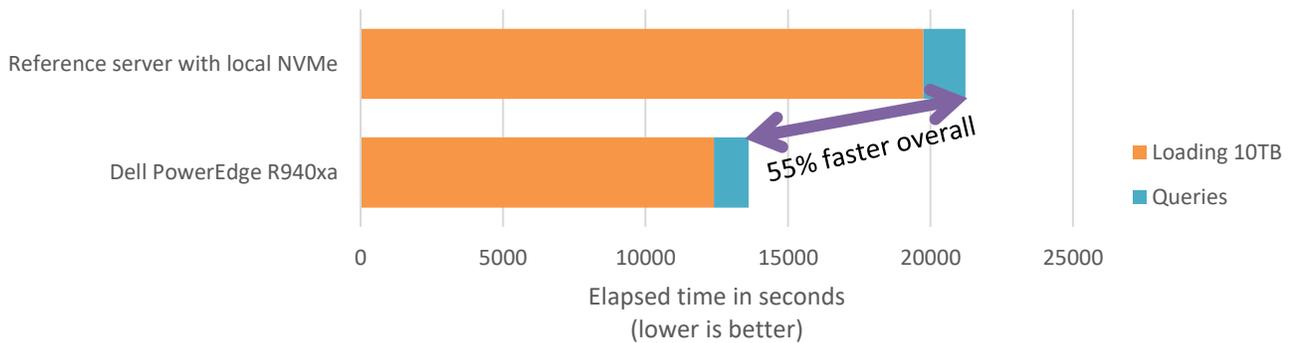
Component	Details
Server Model	PowerEdge R940xa
Processor	4x Intel® Xeon® Platinum 8180M CPU @2.50GHz
Memory	3TiB (64GiB x48) - DIMM DDR4 Synchronous LRDIMM 2666 MHz
Disks	NVMe SSD 6.4TB- Total capacity 24TB
Out of Band Management	iDRAC9 Enterprise with Lifecycle controller
Power Supplies	Titanium 2400W, Platinum
Storage Controllers	PowerEdge RAID controller (PERC)

Performance

20% FASTER QUERIES, 60% FASTER LOAD

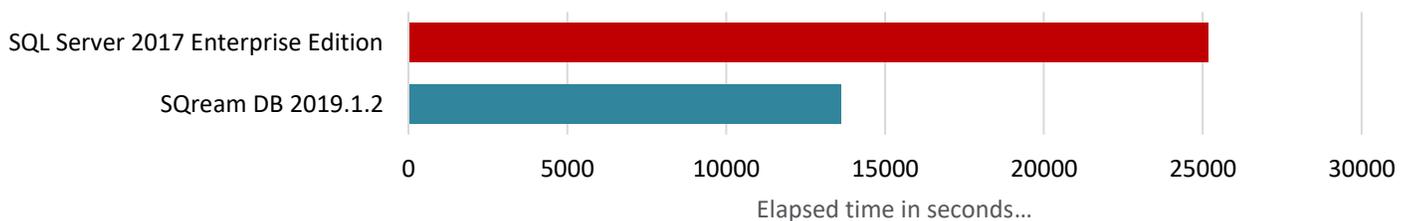
The joint system performs up to 60% faster than a standalone SQream DB instance, according to an industry-standard benchmark.

Dell PowerEdge R940xa Overall Test Performance



SQream DB also beats overall test performance on a similarly configured machine, according to official benchmarks for SQL Server 2017 Enterprise Edition.

SQream DB Overall Test Performance



CONCLUSION

The Dell EMC PowerEdge R940xa integrated database platform builds upon years of experience in handling big data for analytics. The SQream DB and PowerEdge R940xa solution is designed, tested, and validated for enterprises and companies to scale to very large data sets while lowering TCO, and opening up new revenue streams from the vast volumes of data.

FOR MORE INFORMATION

- SQream DB's reference architecture for big data – <https://sqream.com/product/hardware>
- Try SQream DB – <https://sqream.com/try-sqream-db>
- SQream - <https://www.sqream.com>
- Dell PowerEdge R940xa - <https://www.dell.com/en-us/work/shop/povw/poweredge-r940xa>
- Dell - <https://dell.com>