FlashStack Converged Infrastructure Solution

Analytics Design Guide using SAP HANA® SPS10 and SAP® Lumira on Pure Storage®

March 2016

Contributors: Krishna Satyavarapu
Pure Storage, Inc. 650 Castro Street, Mountain View, CA 94041
http://www.purestorage.com

Table of Contents

Introduction

FlashStack Converged Infrastructure (CI) is a flexible, all-flash converged infrastructure solution from Pure Storage that brings the flash revolution to your data center, faster. It combines the latest in compute, network, storage hardware and virtualization software into a single, integrated architecture that reduces time to deployment, lowers overall IT costs and reduces deployment risk. Highly efficient components reduce the costs associated with power, cooling and data center space. Based on 100% flash storage, FlashStack CI provides the performance and reliability business-critical applications demand.

This document describes a FlashStack reference architecture for deploying analytics solution using SAP HANA® SPS10 and SAP Lumira Server and operating on SUSE Linux Enterprise 11 Service Pack 3. Pure Storage has validated this reference architecture within its lab environment. This document presents the advantages of deploying SAP Lumira server on SAP HANA® analytics solution on FlashStack CI. It covers the best practices guidance for SAP HANA® Basis adminstrators, Storage administrators and especially for SAP HANA® modellers and SAP Lumira business users.

Goals and Objectives

SAP HANA® is an in-memory data platform that is deployable as an on-premise appliance, or in the cloud. It is a revolutionary platform that's best suited for performing real-time analytics, and developing and deploying real-time applications. Customers always face a big dilemma for chosing the infrastructure for SAP HANA® and understand the importance of enterprise storage.

The typical challenges faced by SAP customers are:

- Time and effort spent in deployment/maintenance of SAP HANA® landscapes.
- Time and effort spent in performing SAP HANA® systems backup and recovery and space required to store SAP HANA® copies/clones for testing and development purposes.
- Performance of real-time reporting with respect to Business queries runtime, downtime.

The goal of this document is to address these challenges faced by customers by deploying SAP HANA® on a highly available and performant FlashStack CI. The benefits of deploying SAP on FlashStack include:

- Up to 50% reduction in deployment time and day-to-day maintenance.
- Reduced TCO: With best in class data reduction built into Pure Storage FlashArray //m, FlashStack customers can significantly reduce the foot print of a single SAP instance and can effectively store multiple copies of the SAP HANA® Instance, making it very economical to store multiple SAP system copies.
- Built-in DR: Built-in snapshot capabilities of FlashStack significantly reduce the total time and effort required to complete an SAP HANA® system copy or backup even when database is a few hundred TBs in size.

- Consistent Performance: SAP Lumira real-time reporting on SAP HANA® deployed on FlashStack will be faster even when under the high volume of real-time replication as FlashStack can handle delta-merges much more efficiently.

  The downtime for business users using SAP Lumira would be minimal as SAP HANA® restarts, fail-overs and first time column store loads are extremely fast.

In the sections that follow, we will further highlight the benefits of running SAP Lumira on SAP HANA® deployed on FlashStack improving the SAP Lumira reporting performance and how it can accelerate typical SAP HANA® operations such as backup/recovery.

Audience

The target audience for this document includes storage architects, data center architects, SAP HANA® Basis administrators , and especially for SAP HANA® modellers and SAP Lumira business users who want to implement SAP HANA® real-time analytics systems based on SAP Lumira server. A working knowledge of server, storage, networks and data center design is assumed but is not a prerequisite to read this document.

Reference Architecture Design Principles

The guiding principles for implementing this reference architecture are:

- Repeatable: Create a scalable building block that can be easily replicated at any customer site. Publish the version of various firmware under test and weed out any issues in the lab before customers deploy this solution.
- Available: Create a design that is resilient and not prone to failure of a single component. For example, we include best practices to enforce multiple paths to storage, multiple NICs for connectivity, and high availability (HA) clustering.
- Efficient: Build a solution that leverages efficient benefits of an all-flash architecture.
- Simple: Simplify deployment and ongoing maintenance tasks via automation.
- Scalable: Create a design that can start small, but can easily grow to the meet the needs of a growing enterprise.

FlashStack Introduction

FlashStack CI is available from accredited FlashStack Partners who help provide an excellent converged infrastructure ownership experience. FlashStack Partners have the knowledge and experience necessary to help streamline the sizing, procurement, and delivery of your entire system.

Key Benefits of the FlashStack solution are:

1. Consistent Performance and Scalability

   Consistent sub-millisecond latency with 100% flash storage.
   Consolidate 100s of enterprise-class applications in a single rack.
   Scale easily, without disruption.
   Repeatable growth through multiple FlashStack CI deployments.

2. Operational Simplicity

   Fully tested, validated, and documented for rapid deployment
   Reduced management complexity
   Auto-aligned 512b architecture eliminates storage alignment headaches
   No storage tuning or tiers necessary

3. Lowest TCO

   Dramatic savings in power, cooling and space with 100% Flash.
   Industry leading data reduction

   Free controller upgrades with three year maintenance renewal under Forever Flash
   Free FlashArray controller upgrades every three years with Forever Flash™

4. Enterprise Grade Resiliency

   Highly available architecture and redundant components
   Non-disruptive operations
   Upgrade and expand without downtime or performance loss
   Native data protection: snapshots and replication

FlashStack Reference Architecture Components Introduction

FlashArray //m

The FlashArray//m expands upon the FlashArray's modular, stateless architecture, designed to enable expandability and upgradability for generations. The FlashArray//m leverages a chassis-based design with customizable modules, enabling both capacity and performance to be independently improved over time with advances in compute and flash, to meet your business' needs today and tomorrow.

FlashArray//m makes server and workload investments more productive, while also lowering storage spend. With FlashArray//m, organizations can dramatically reduce the complexity of storage to make IT more agile and efficient, accelerating your journey to the cloud.

FlashArray//m's performance can also make your business smarter by unleashing the power of real-time analytics, driving customer loyalty, and creating new, innovative customer experiences that simply weren't possible with disk. All by Transforming Your Storage with FlashArray//m.

FlashArray//m enables you to transform your data center, cloud, or entire business with an affordable all-flash array capable of consolidating and accelerating all your key applications.

Mini Size—Reduce power, space and complexity by 90%

- 3U base chassis with 15-120+ TBs usable
- ~1kW of power
- 6 cables

Mighty Performance—Transform your datacenter, cloud, or entire business

- Up to 300,000 32K IOPS
- Up to 9 GB/s bandwidth
- <1ms average latency

Modular Scale—Scale FlashArray//m inside and outside of the chassis for generations

- Expandable to ~½ PB usable via expansion shelves
- Upgrade controllers and drives to expand performance and/or capacity

Meaningful Simplicity—Appliance-like deployment with worry-free operations

- Plug-and-go deployment that takes minutes, not days
- Non-disruptive upgrades and hot-swap everything

- Less parts = more reliability

The Pure Storage FlashArray//m is ideal for:

Accelerating Databases and Applications Speed transactions by 10x with consistent low latency, enable online data analytics across wide datasets, and mix production, analytics, dev/test, and backup workloads without fear.

Virtualizing and Consolidating Workloads Easily accommodate the most IO-hungry Tier 1 workloads, increase consolidation rates (thereby reducing servers), simplify VI administration, and accelerate common administrative tasks.

Delivering the Ultimate Virtual Desktop Experience Support demanding users with better performance than physical desktops, scale without disruption from pilot to >1000's of users, and experience all-flash performance for under $100/desktop.

Protecting and Recovering Vital Data Assets Provide an always-on protection for business-critical data, maintain performance even under failure conditions, and recover instantly with FlashRecover.

Pure Storage FlashArray sets the benchmark for all-flash enterprise storage arrays. It delivers:

Consistent Performance FlashArray delivers consistent <1ms average latency. Performance is optimized for the real-world applications workloads that are dominated by I/O sizes of 32K or larger vs. 4K/8K hero performance benchmarks. Full performance is maintained even under failures/updates.

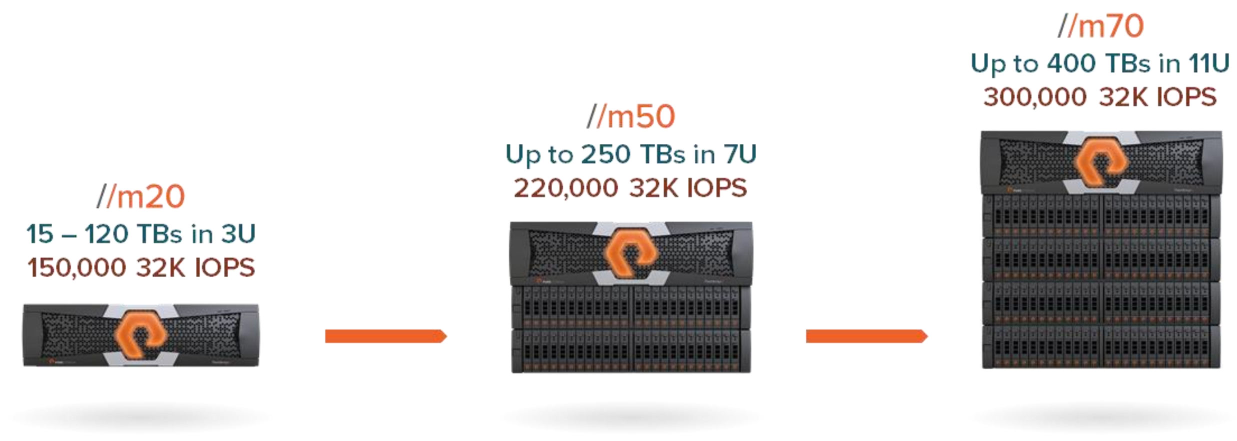
Less Cost than Disk Inline de-duplication and compression deliver 5 – 10x space savings across a broad set of I/O workloads including Databases, Virtual Machines and Virtual Desktop Infrastructure.

Mission-Critical Resiliency FlashArray delivers >99.999% proven availability, as measured across the Pure Storage installed base and does so with non-disruptive everything without performance impact.

Disaster Recovery Built-In FlashArray offers native, fully-integrated, data reduction-optimized backup and disaster recovery at no additional cost. Setup disaster recovery with policy-based automation within minutes. And, recover instantly from local, space-efficient snapshots or remote replicas.

Simplicity Built-In FlashArray offers game-changing management simplicity that makes storage installation, configuration, provisioning and migration a snap. No more managing performance, RAID, tiers or caching. Achieve optimal application performance without any tuning at any layer. Manage the FlashArray the way you like it: Web-based GUI, CLI, VMware vCenter, Rest API, or OpenStack.

FlashArray//m Specifications



	//m20	//m50	//m70
Capacity	• Up to 120+ TBs effective capacity* • 5 – 40TBs raw capacity (base chassis)	• Up to 250+ TBs effective capacity* • 30 – 88TBs raw capacity (w/shelves)	• Up to 400+ TBs effective capacity* • 44 – 136TBs raw capacity (w/shelves)
Performance	• Up to 150,000 32K IOPS** • <1ms average latency • Up to 5 GB/s bandwidth	• Up to 220,000 32K IOPS** • <1ms average latency • Up to 7 GB/s bandwidth	• Up to 300,000 32K IOPS** • <1ms average latency • Up to 9 GB/s bandwidth
Connectivity	• 8 Gb/s Fibre Channel • 10 Gb/s Ethernet iSCSI • Management and Replication ports	• 16 Gb/s Fibre Channel • 10 Gb/s Ethernet iSCSI • Management and Replication ports	• 16 Gb/s Fibre Channel • 10 Gb/s Ethernet iSCSI • Management and Replication ports
Physical	• 3U • 742 Watts (nominal draw) • 110 lbs. (49.9 kg) fully loaded • 5.12” x 18.94” x 29.72” FlashArray//m chassis	• 3U – 7U • 1007 - 1447 Watts (nominal draw) • 110 lbs. (49.9 kg) fully loaded + 44 lbs. per expansion shelf • 5.12” x 18.94” x 29.72” FlashArray//m chassis	• 5U – 11U • 1439 – 2099 Watts (nominal draw) • 110 lbs. (49.9 kg) fully loaded + 44 lbs. per expansion shelf • 5.12” x 18.94” x 29.72” FlashArray//m chassis

* Effective capacity assumes HA, RAID, and metadata overhead, GB-to-GiB conversion, and includes the benefit of data reduction with always-on inline deduplication, compression, and pattern removal. Average data reduction is calculated at 5-to-1, below the global average of the FlashArray//m user base.

** Why does Pure Storage quote 32K, not 4K IOPS? The industry commonly markets 4K IOPS, but real-world environments are dominated by IO sizes of 32K or larger.FlashArray//m adapts automatically to 512B-32KB IO for superior performance, scalability, and data reduction.

Table 1. Pure Storage FlashArray//m Series.

Purity Operating Environment

Purity implements advanced data reduction, storage management and flash management features, and all features of Purity are included in the base cost of the FlashArray//m.

Storage Software Built for Flash—The FlashCare technology virtualizes the entire pool of flash within the FlashArray, and allows Purity to both extend the life and ensure the maximum performance of consumer-grade MLC flash.

Granular and Adaptive—Purity Core is based upon a 512-byte variable block size metadata layer. This fine-grain metadata enables all of Purity's data and flash management services to operate at the highest efficiency.

Best Data Reduction Available—FlashReduce implements five forms of inline and post-process data reduction to offer the most complete data reduction in the industry. Data reduction operates at a 512-byte aligned variable block size, to enable effective reduction across a wide range of mixed workloads without tuning.

Highly Available and Resilient—FlashProtect implements high availability, dual-parity RAID-3D, non-disruptive upgrades, and encryption, all of which are designed to deliver full performance to the FlashArray during any failure or maintenance event.

Backup and Disaster Recovery Built In—FlashRecover combines space-saving snapshots, replication, and protection policies into an end-to-end data protection and recovery solution that protects data against loss locally and globally. All FlashProtect services are fully-integrated in the FlashArray and leverage the native data reduction capabilities.

Pure1

Pure1 Manage—By combining local web-based management with cloud-based monitoring, Pure1 Manage allows you to manage your FlashArray wherever you are – with just a web browser.

Pure1 Connect—A rich set of APIs, plugin-is, application connectors, and automation toolkits enable you to connect FlashArray//m to all your data center and cloud monitoring, management, and orchestration tools.

Pure1 Support—FlashArray//m is constantly cloud- connected, enabling Pure Storage to deliver the most proactive support experience possible. Highly trained staff combined with big data analytics help resolve problems before they start.

Pure1 Collaborate—Extend your development and support experience online, leveraging the Pure1 Collaborate community to get peer-based support, and to share tips, tricks, and scripts.

Experience Evergreen™ Storage

Tired of the 3-5 year array replacement merry-go-round? The move to FlashArray//m can be your last data migration. Purchase and deploy storage once and once only – then expand capacity and performance incrementally in conjunction with your business needs and without downtime. Pure Storage's vision for Evergreen Storage is delivered by a combination of the FlashArray's stateless, modular architecture and the ForeverFlash business model, enabling you to extend the lifecycle of storage from 3-5 years to a decade or more.

Cisco Unified Computing System

The Cisco Unified Computing System™ (Cisco UCS™) is a next-generation data center platform that unites compute, network, storage access, and virtualization into an organized structure aimed to reduce total cost of ownership and introduce vastly improved infrastructure deployment mechanisms at scale. UCS incorporates a unified network fabric with scalable, modular and powerful x86-architecture servers. With an innovative and proven design, Cisco UCS delivers an architecture that increases cost efficiency, agility, and flexibility beyond what traditional blade and rack-mount servers provide. Cisco makes organizations more effective by addressing the real problems that IT managers and executives face and solves them on a systemic level.

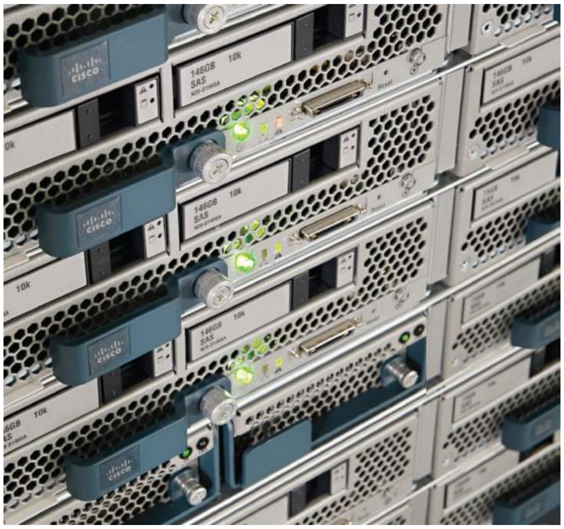


Figure 1: Cisco Unified Computing System

Greater Time-on-Task Efficiency

Automated configuration can change an IT organization's approach from reactive to pro-active. The result is more time for innovation, less time spent on maintenance, and faster response times. These efficiencies allow IT staff more time to address strategic business initiatives. They also enable better quality of life for IT staff, which means higher morale and better staff retention—both critical elements for long-term efficiency.

Cisco UCS Manager is an embedded, model-based management system that allows IT administrators to set a vast range of server configuration policies, from firmware and BIOS settings to network and storage connectivity. Individual servers can be deployed in less time and with fewer steps than in traditional environments. Automation frees staff from tedious, repetitive, time-consuming chores that are often the source of errors that cause downtime, making the entire data center more cost effective.

Easier Scaling

Automation means rapid deployment, reduced opportunity cost, and better capital resource utilization. With Cisco UCS, rack-mount and blade servers can move from the loading dock and into production in a "plug-and-play" operation. Automatically configure blade servers using predefined policies simply by inserting the devices into an open blade chassis slot. Integrate rack-mount servers by connecting them to top-of-rack Cisco Nexus® fabric extenders. Since policies make configuration automated and repeatable, configuring 100 new servers is as straightforward as configuring one server, delivering agile, cost-effective scaling.

Virtual Blade Chassis

With a separate network and separate management for each chassis, traditional blade systems are functionally an accidental architecture based on an approach that compresses all the components of a rack into each and every chassis. Such traditional blade systems are managed with multiple management tools that are combined to give the illusion of convergence for what is ultimately a more labor-intensive, error-prone and costly delivery methodology. Rack-mount servers are not integrated and must be managed separately or through additional tool sets, adding complexity, overhead, and the burden of more time.

Architecturally, Cisco UCS blade and rack-mount servers are joined into a single virtual blade chassis that is centrally managed yet physically distributed across multiple blade chassis, rack-mount servers, and even racks and rows. This capability is delivered through

Cisco® fabric interconnects that provide redundant connectivity, a common management and networking interface, and enhanced flexibility. This larger virtual chassis, with a single redundant point of management, results in lower infrastructure cost per server, with fewer management touch points, and lower administration, capital, and operational costs.

SAP HANA® – Real time analytics platform

SAP HANA® is an in-memory data platform that is deployable as an on-premise appliance, or in the cloud. It is a revolutionary platform that's best suited for performing real-time analytics, and developing and deploying real-time applications.

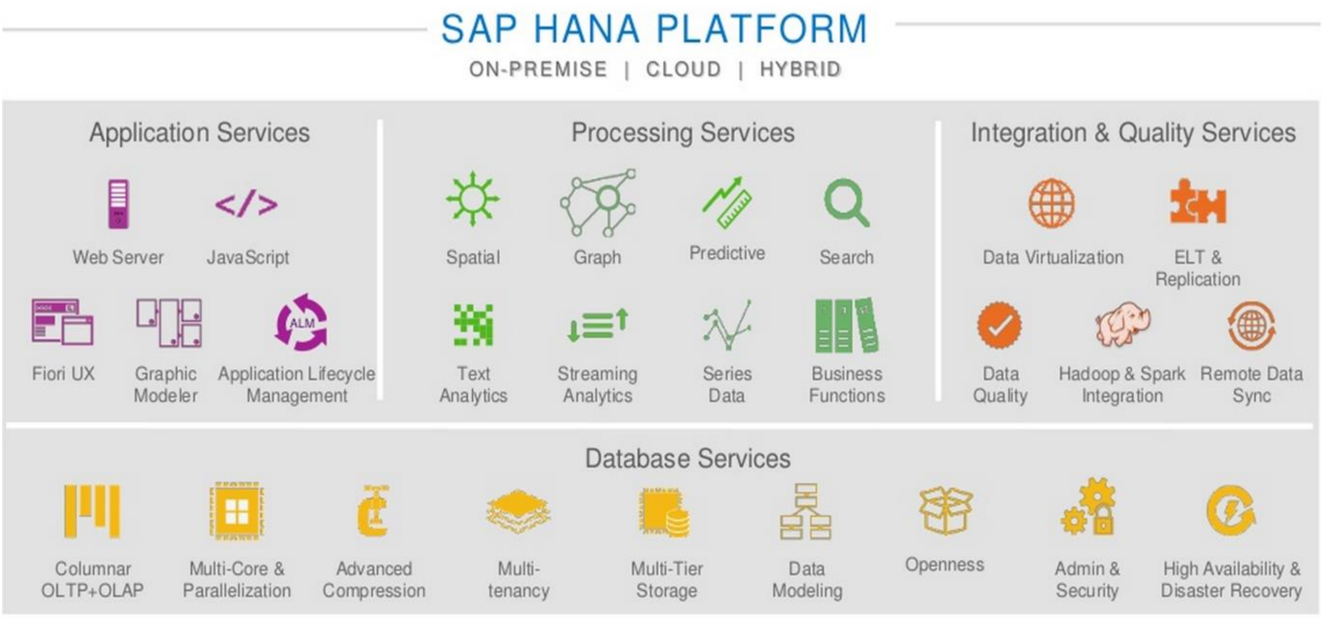


Figure 2: SAP HANA® Platform

SAP HANA® is an in-memory database:

- It is a combination of hardware and software made to process massive real time data using In-Memory computing.
- It combines row-based, column-based database technology.
- Data now resides in main-memory (RAM) but it still needs a very good enterprise storage to operate efficiently.
- It's best suited for performing real-time analytics, and developing and deploying real-time applications.

An in-memory database means all the data is stored in the memory (RAM), but it would still need an extremely fast enterprise storage for persistence, delta merges etc. The speed advantages offered by this RAM storage system are further accelerated by the use of multi-core CPUs, and multiple CPUs per board, and multiple boards per server appliance. Complex calculations on data are not carried out in the application layer, but are moved to the database.

SAP HANA® is equipped with multiengine query processing environment which supports relational as well as graphical and text data within same system. It provides features that support significant processing speed, handle huge data sizes and text mining capabilities.

SAP Lumira

SAP Lumira server for teams allows your team or department to visualize, collaborate on, and better understand their data all on a lightweight in-memory server. With SAP Lumira server for teams, you can quickly analyze large volumes of data, create data stories that captivate audiences, and speed up time to insight. Basically, SAP Lumira enables business users to access, transform, and visualize data of any size in a repeatable and self-service manner. The tool emphasizes a simple user-friendly interface and creates very captivating visualizations which allow users to swiftly analyze data without the need for scripting.

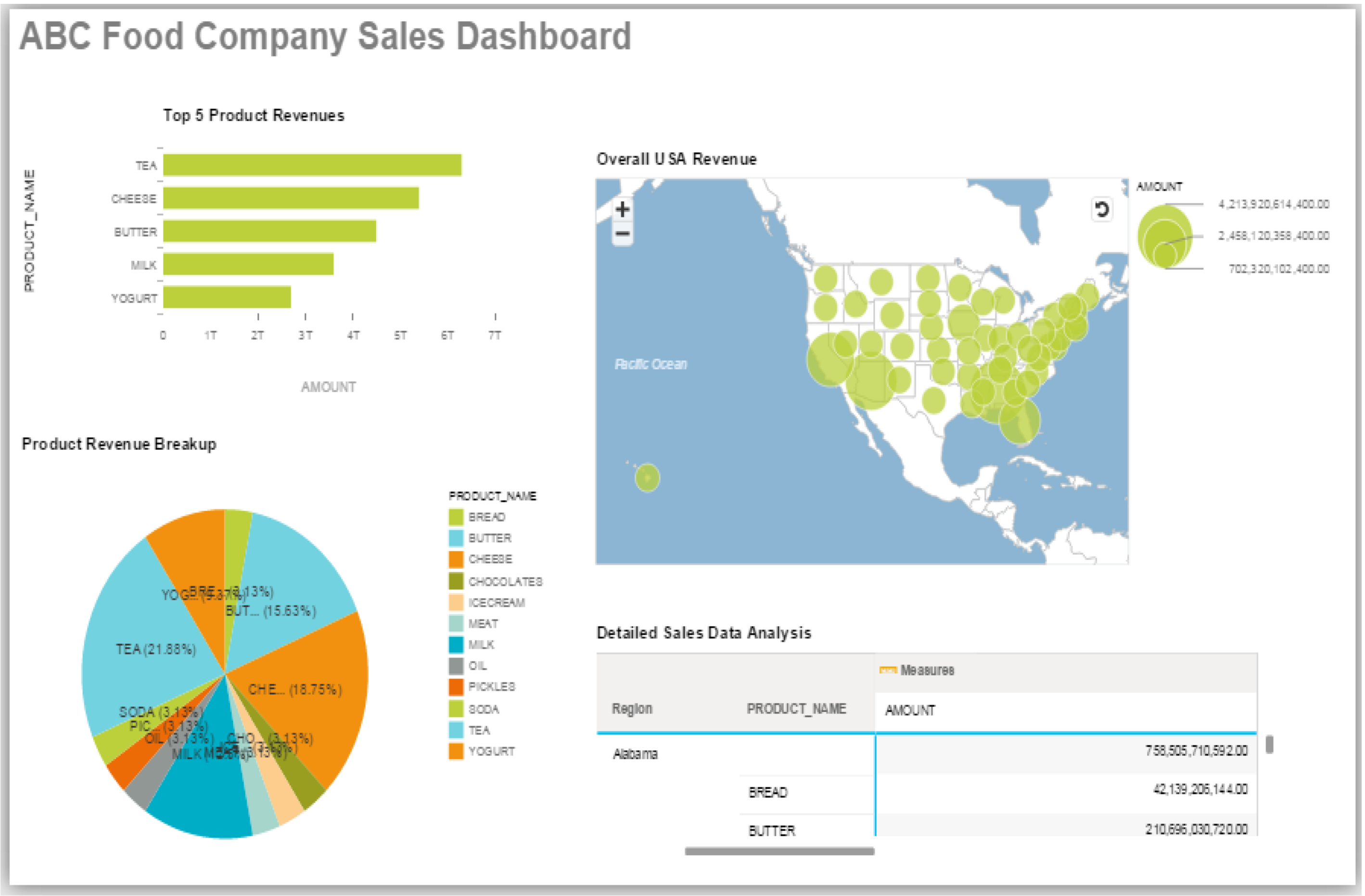


Figure 3: SAP Lumira dashboard

SAP Lumira's drag-and-drop interface is easy to become familiar with, and with a few clicks data from multiple sources can be gathered, combined, and visualized. After the tool is installed and system access is setup, training is really not needed, but some basics by someone who knows the tool can be helpful.

- Maximizing business knowledge by integrating wide-scale insights and drill-down to granular details
- Accelerating decision-making with immediate fact-based solutions to intricate business questions by avoiding list tables and fixed format reports. Interaction is key.
- Increasing self-service data usage without increasing the workload of the IT department
- Visualizing any amount of data in real time using SAP HANA® and simple deployment to mobile devices.

Design Guide Solutions Overview

FlashStack consists of a combined stack of hardware (storage, network and compute) and software (Cisco UCS Manager, SAP HANA® SPS10 (Scale out 2+1 nodes cluster), SAP Lumira server and Pure Storage GUI)

- **Network:** Cisco Nexus 5500UP and Cisco UCS Fabric Interconnect 6248UP for external and internal connectivity of IP and FC network.
- **Storage**: Pure Storage FlashArray//m 50 with Fibre Channel connectivity
- **Compute**: Cisco UCS B440 M2 Blade Server

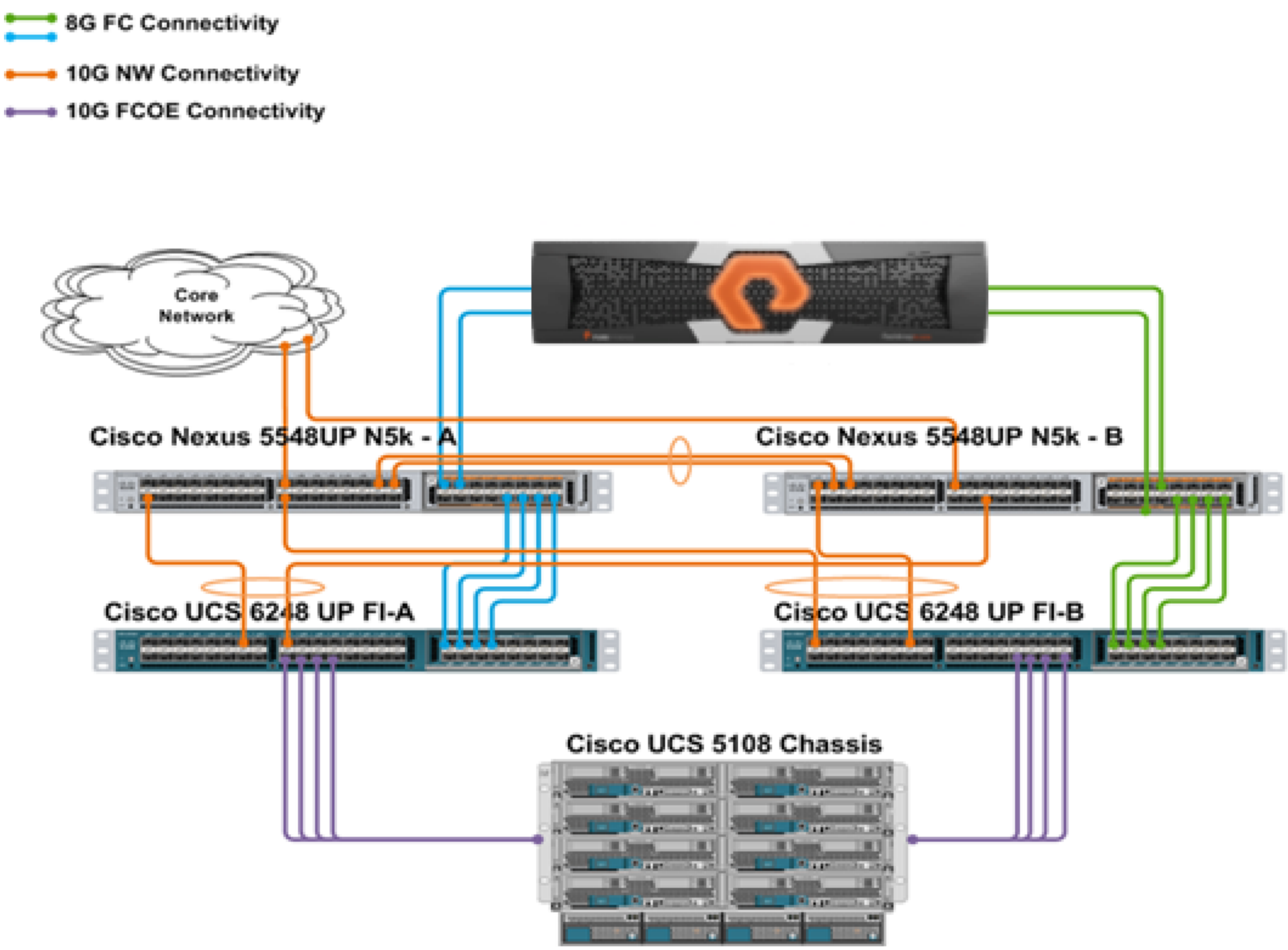


Figure 4: FlashStack connectivity diagram

Figure 4 shows a detailed topology of the reference architecture configuration. A major goal of the architecture is to build out a highly redundant and resilient infrastructure. Thus, we used powerful servers with dual Fibre Channel ports connected redundantly to two SAN switches that were connected to redundant FC target ports on the FlashArray//m. The servers also have redundant network connectivity. It is configured in End-Host Mode.

Cisco UCS Server Configuration

A pair of Cisco UCS Fabric Interconnects 6248UP, and 3 Cisco UCS B-series B440-M4 blade servers were deployed for hosting SAP HANA® SPS10 (Scale out 2+1 nodes cluster), SAP Lumira server. The UCS manager, UCS Fabric Interconnects and the components in the chassis were upgraded to 2.2.3f firmware level.

The server has Cisco VIC 1280 cards and they were connected through four ports from each Cisco Fabric extender of the Cisco UCS chassis to the Cisco Fabric Interconnect, they were in turn connected to Cisco Nexus 5548UP Switch for upstream connectivity to access the Pure Storage FlashArray//m LUNs. The server configuration is described in Table 2.

Component	Description
Processor	4X Intel(R) Xeon(R) E74870 @ 2.4GHz, 40 Cores
Memory	512 GB @ 1600 MHz, regular voltage
HBA	4 X 8G ports on Cisco UCS VIC 1280 (UCSB-MLOM-40G-03)
NIC	4 X 10G ports on Cisco UCS VIC 1280 (UCSB-MLOM-40G-03)
Application	SAP Lumira Sever 1.29
Database/Platform	SAP HANA® SPS10 (Scale out 2+1 nodes cluster)

Table 2. Cisco UCS Server configuration.

Cisco UCS Service Profile configuration

In order to facilitate rapid deployment of UCS servers, a service profile template was created with the following characteristics [more details can be found in the deployment guide].

1. We configured boot from SAN policy so that the server booted from a Pure Storage boot LUN
2. We kept every other setting to the default, we didn't tweak any parameters
3. The Ethernet and FC adapter policy was set to Linux.
4. The BIOS defaults were used for the B440-M2 blade servers
5. We configured two vHBA FC adapters and four vNIC Eth adapters on the Cisco VIC cards to avoid any single point of failure.
6. We deployed three service profiles from the template and associated it with the blade servers in.

Cisco Networking

Fabric Interconnect Configuration (Cisco UCS 6248UP)

The Cisco UCS 6248UP 48-Port Fabric Interconnect is a core part of Cisco Unified Computing System. It is usually deployed in redundant pairs, the Cisco UCS 6248UP Fabric Interconnects provide uniform access to both networks and storage. Here we are configuring it for the End-Host Mode.

Figure 5: Cisco UCS 6248UP

Cisco Nexus 5500UP Switches

Cisco Nexus 5500UP switches, using cut-through architecture, supports line-rate 10 Gigabit Ethernet on all ports while maintaining consistently low latency independent of packet size and services enabled. It supports a set of network technologies known collectively as Data Center Bridging (DCB) that increases the reliability, efficiency, and scalability of Ethernet networks. These features allow the switches to support multiple traffic classes over a lossless Ethernet fabric, thus enabling consolidation of LAN, SAN and cluster environments. Its ability to connect Fibre Channel over Ethernet (FCoE) to native Fibre Channel protects existing storage system investments while dramatically simplifying in-rack cabling. The Nexus 5500UP series switch we used was a Nexus 5548UP (see Figure 5 below), which is a 1RU 10 Gigabit Ethernet, Fibre Channel, and FCoE switch offering up to 960 Gbps of throughput and up to 48 ports. The switch has 32 unified ports and one expansion slot.

Figure 6: Cisco Nexus 5500UP switch

Installation/Setup of SAP HANA® and SAP Lumira server

SAP HANA® SPS10 (scale out 2+1 node cluster) and SAP Lumira server

The SAP HANA® certification infrastructure uses a 2+1 cluster configuration. Three Cisco UCS B440 M2 blades are used with SUSE 11 SP3 OS, including High Availability Option and OCFS2 to create the environment and on which SAP Lumira server is installed.

The Pure Storage FlashArray used in this configuration is the FlashArray//m50. There are four 8GB Fibre channel ports from the array attached to the SAN Fabric. Each node has two Fibre channel ports for storage and four 10 GBPS ports for network.

All the servers boot from SAN with dedicated drive for their OS related partitions. All other LUNs are shared across the nodes. The /hana/shared partition is used for storing SAP binaries and is mounted in parallel simultaneously.

The HANA® clustering controls the DATA and LOG LUNs that are needed by the master and the worker nodes. These are mounted in the appropriate node with the role of master or worker.

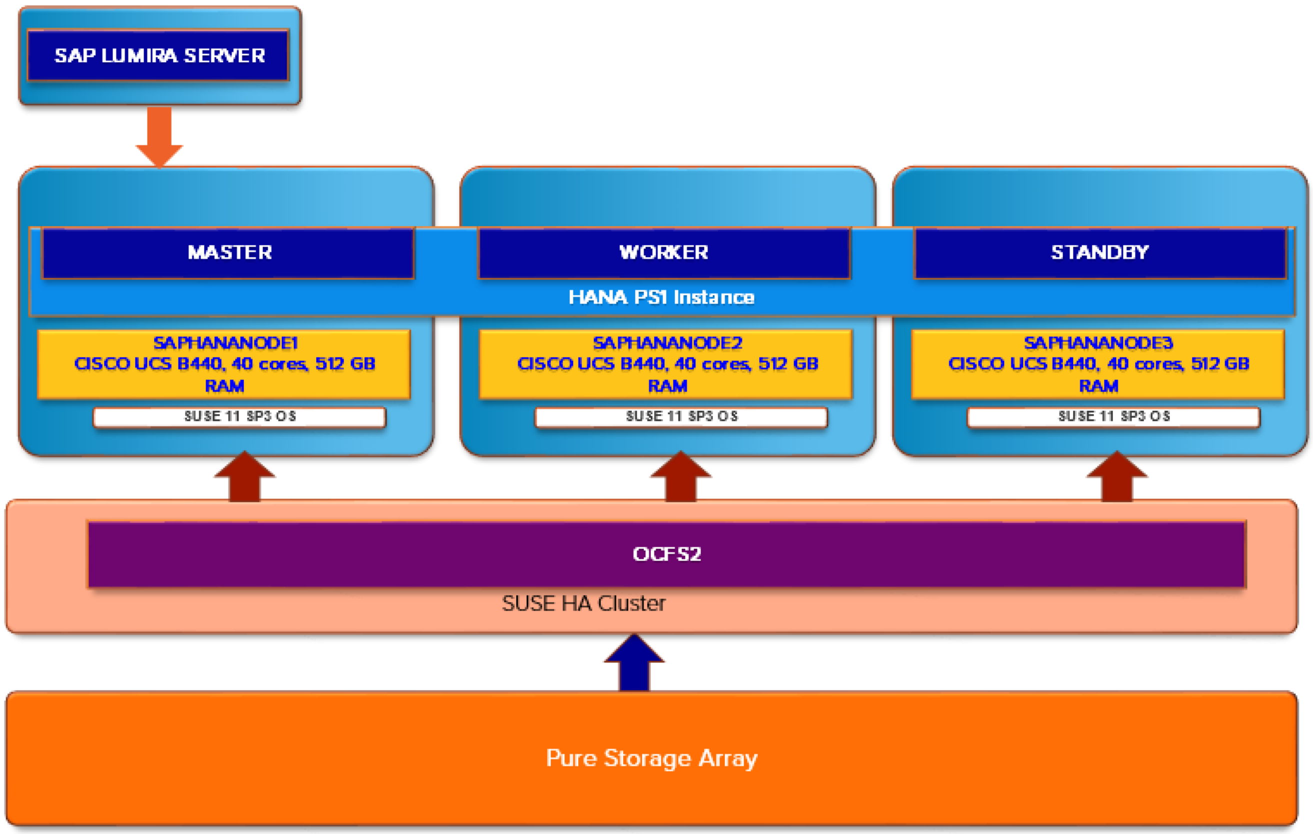


Figure 7: Logical architecture of the Pure Storage HANA® setup

Software Matrix used for SAP System

Slno	Component	Software
1.	Database/Platform	SAP HANA® SPS10
2.	Operating System	SUSE Linux Enterprise 11 Service Pack 3
3.	Reporting Application	SAP Lumira server
4.	File System	XFS, /hana/shared (OCFS2)

Table 3: Software Matrix used for SAP HANA® system

Value propositions of Deploying analytics SAP HANA® solution on FlashStack

Data reduction

SAP HANA® has both row stores and column stores. Column stores are used for storing 90% of data in a typical SAP HANA® system. The column store allows for the efficient compression of data. This makes it less costly for the SAP HANA® database to keep data in main memory. This cannot be turned off, it is always on.

Data in column tables can have a two-fold compression:

- Dictionary compression

  This default method of compression is applied to all columns. It involves the mapping of distinct column values to consecutive numbers, so that instead of the actual value being stored, the much smaller consecutive number is stored.

- Advanced compression

  Each column can be further compressed using different compression methods, namely prefix encoding, and run length encoding (RLE), cluster encoding, sparse encoding, and indirect encoding. The SAP HANA® database uses compression algorithms to determine which type of compression is most appropriate for a column.

Row stores are not compressed on SAP HANA®, hence more row store the more data reduction on FlashArray//m.

To test data reduction on SAP HANA®, we are using 500GB of TPC-H data. Here initially we loaded all the tables as column stores and then will convert them in to Row stores using SQL

➔ ALTER TABLE <table_name> ROW

This would increase the amount of row store and as you can see the data reduction increases as the amount of Row store increases.

Initially with all the 500GB column store (100% column store) the data reduction was somewhere from **1.9-2.3** across the two data volumes. Now using the above SQL we will convert it to row store and see that the data reduction increases when the row store percentage increases.

Below are the analysis and impact on data reduction when the row store size is increased. As mentioned, data reduction increases as the row store percentage increases. See below is the table and graph on SAP HANA® data reduction analysis.

Percentage of Column store (%)	Percentage of Row store (%)	Minimum Data Reduction	Maximum Data reduction
100	0	1.9	2.3
90	10	2.0	2.4
80	20	2.1	2.5

Table 4: Data reduction analysis

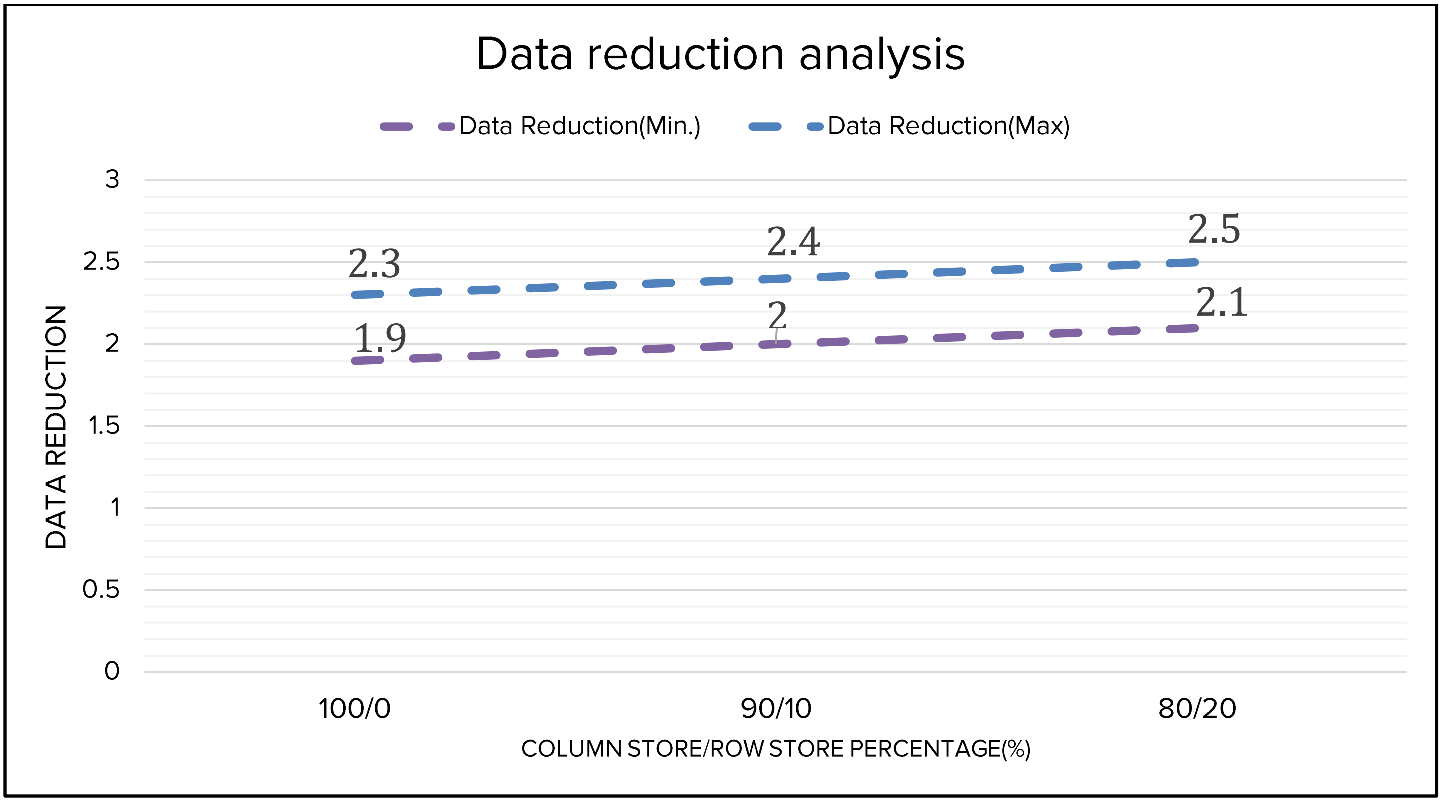


Figure 8: Data reduction analsis graph

Summary

FlashArray//m reduces the data even further on top of SAP HANA®'s compression. Data reduction of around 1.9 to 2.3 is seen when 100% column stores are present. Data reduction increases when the percentage of row stores increases in SAP HANA®. It increases even further when SAP HANA® system copies and clones are done.

Improved Restart time

SAP HANA® restarts are very frequent, and this causes temporary downtime for business users using SAP Lumira reports. The advantage of deploying on FlashStack to real-time analytics SAP HANA® solution on FlashStack is reduction of downtime as SAP HANA® restarts are much faster on FlashStack due to high read performance on FlashArray//m.

During SAP HANA® restarts the following activities are performed by restart agent of persistence layer.

- Row store tables are loaded into memory of master node of SAP HANA®.
- Column store tables which are critical for reporting are loaded into memory across SAP HANA® nodes.

To test the restart performance on FlashStack we have tested the following ways:

- Row store test case: By populating SAP HANA® master node with 50 GB to 350 GB of Row store data.
- Column store test case: By populating SAP HANA® nodes Column stores from 50 GB to 250 GB per active nodes. These column stores were modified to be loaded on restart with preload flag using "ALTER TABLE <TABLE_NAME> PRELOAD ALL".

IO pattern of table reload in data volume: READ

4 KB – 16 MB blocks, up to 64 MB (clustered Row Store super blocks)

Row store test case on Master node

Populated SAP HANA® master node with only row store tables and tested with different data volume sizes from 50GB to 350GB data and checking the restart times.

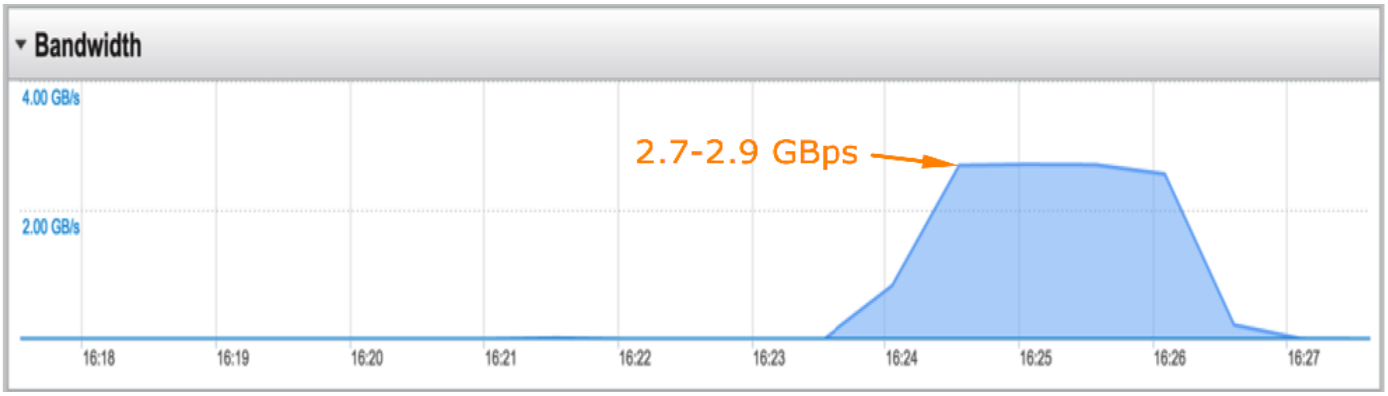


Figure 9: SAP HANA restart run with 350GB row store

Above is the 350 GB Row store restart run on FlashArray//m50. As you can see we were able to achieve read bandwidth of **2.7-2.9 GB/second on just master node** and entire loading of row store into memory took less than 3 minutes and 30 seconds.

Analysis of Row store test case

Below is the analysis of row store load time (seconds) vs amount of row store present in master node, while we scaled it from 50GB to 350GB. SAP HANA® after restart takes around 90 seconds to 2 minutes to reach the file IO operations when it starts loading SAP HANA® row store. So total restart time to load 350GB would be just around 5 minutes.

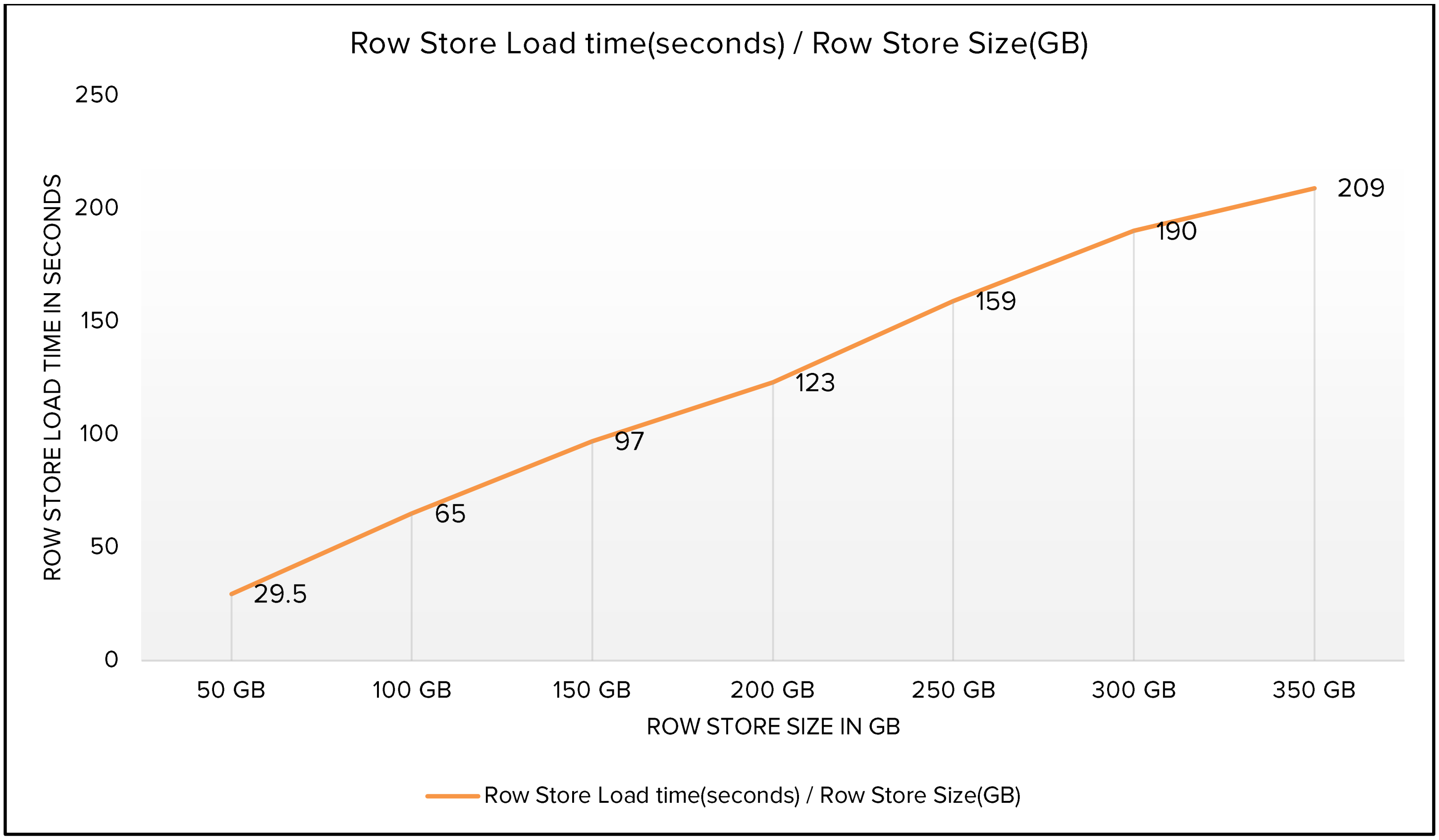


Figure 10: Row store load time(Seconds) – Row store size(GB) analysis

Column store test case on two active nodes

In order to test restart times for two SAP HANA® nodes in parallel and to showcase rapid restart time of SAP HANA® on FlashStack, SAP HANA® active nodes were populated with column stores from 50 GB to 250 GB per active node. These column stores were modified to be loaded on restart with preload flag using "ALTER TABLE <TABLE_NAME> PRELOAD ALL".

Below is the peak bandwidth achieved in each of the active nodes when it was restarting with 400 GB column store across 2 nodes.

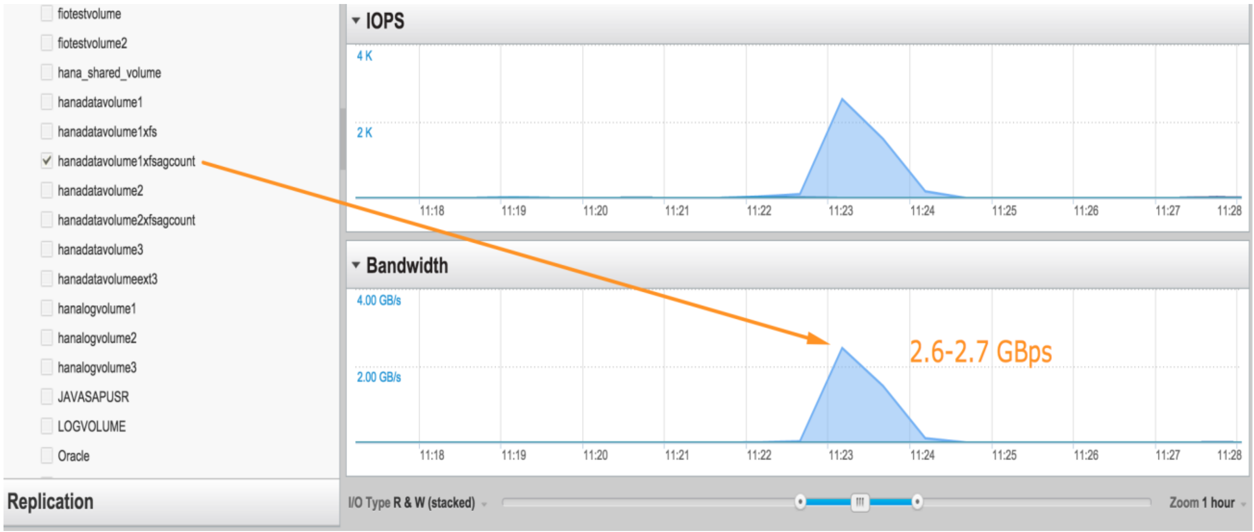


Figure 11: SAP HANA restart with 200GB column store on master node

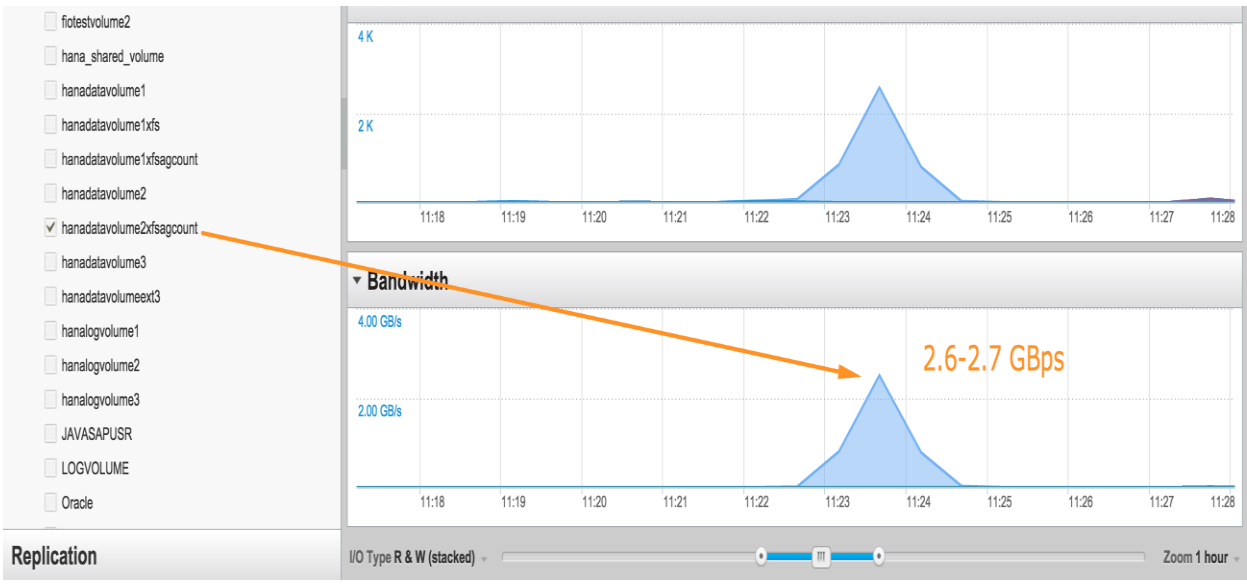


Figure 12: SAP HANA restart with 200GB column store on worker node

Below is the 400 GB Column store restart run on FlashArray//m50. As you can see we were able to achieve read bandwidth of **4.5 GB/second on just two nodes** on FlashArray//m50. Complete loading of column store into memory took less than 2 minutes.

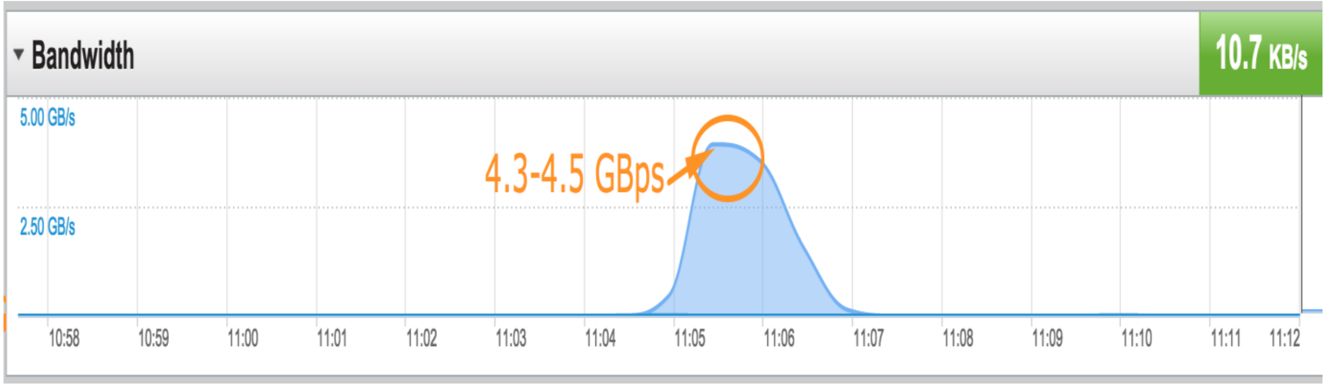


Figure 13: SAP HANA restart with 400GB column store on two nodes

Analysis of Column store test case

Below is the graphical analysis of two nodes SAP HANA® restart times with only column store tables from 50GB to 250GB data per node mapped against the restart times for each of these runs. SAP HANA® during restart takes around 90 seconds to 2 minutes to reach the file IO operations when it starts loading SAP HANA® column store. So total restart time to load 500 GB would be just around 5 minutes.

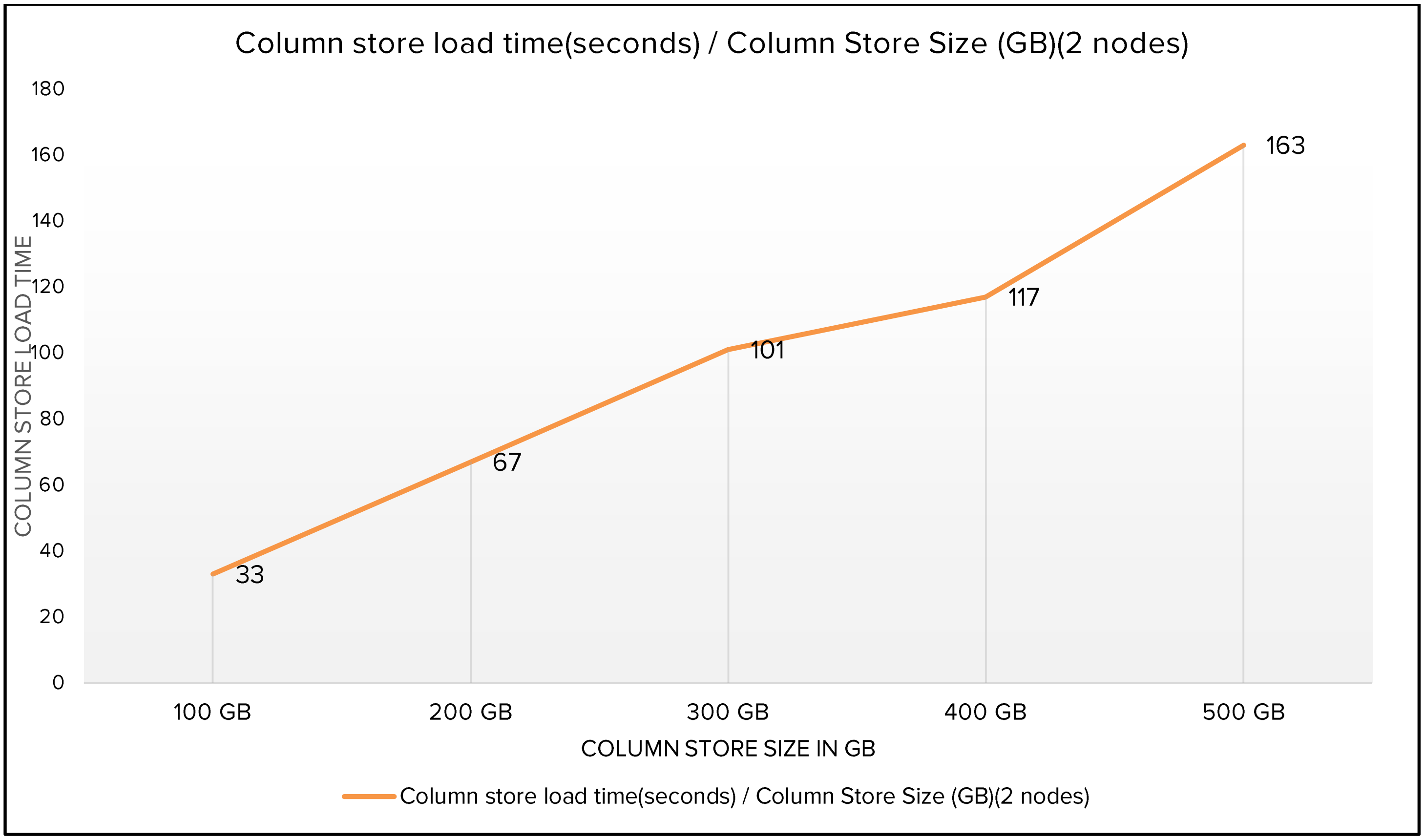


Figure 14: Column store load time(seconds) – column store size(GB) (2 nodes) graph analysis

Below is the analysis of the number of active SAP HANA® nodes against the peak bandwidth achieved by FlashArray//m50. FlashArray//m50 can reach a maximum of 6.3 GBps when three SAP HANA® nodes are restarting.

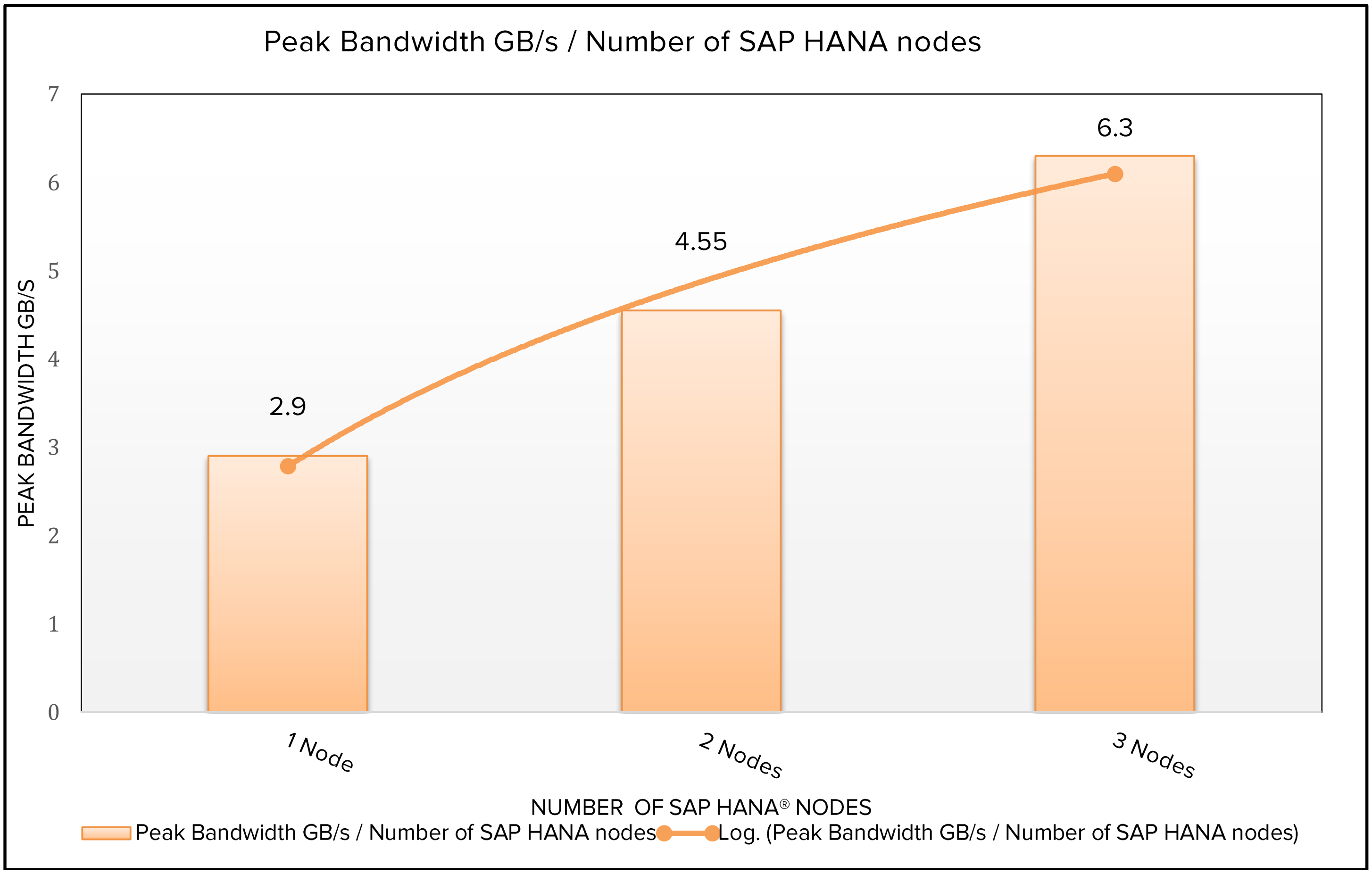


Figure 15: Peak bandwidth GB/s – Number of SAP HANA nodes graph analysis

Summary

Restarts are accelerated on FlashStack thereby reducing the downtime for business users no matter how much row store or preloaded column stores are present. This is very important advantage as SAP HANA® restarts are very frequent. This would reduce your downtimes by at least 50%.

> Important: It is extremely important to fine tune File system IO for SAP HANA® using hdbparams.Only then you will get the best performance as mentioned in the section. This is described in detail in the SAP HANA® best practices guide(See appendix for URL)

Rapid Failover recovery (Fail-Over time)

SAP HANA® host can fail due to hardware, software or network failure and this results in temporary downtime for business users using SAP Lumira reports. The advantage of deploying real-time analytics SAP HANA® solution on FlashStack is reduction of downtime. SAP HANA® host auto-failovers are much faster on FlashStack due to high read performance on FlashArray//m.

Here we have 2 active nodes + 1 standby node SAP HANA® scale our cluster. For detailed configuration please see the SAP HANA® best practices guide on Pure Storage (See appendix for URL to this document). Once the active node fails the index server on the selected standby host takes over the database volumes of the failed index server and loads the data into memory by executing the restart sequence.

To test the failover recovery performance on FlashStack we have tested by failing the SAP HANA® master node which is populated SAP HANA® master node with 50 GB to 350 GB of Row store data. The master node was de-activated using "echo b > /proc/sysrq-trigger"

IO pattern of table reload in data volume: READ

4 KB – 16 MB blocks, up to 64 MB (clustered Row Store super blocks)

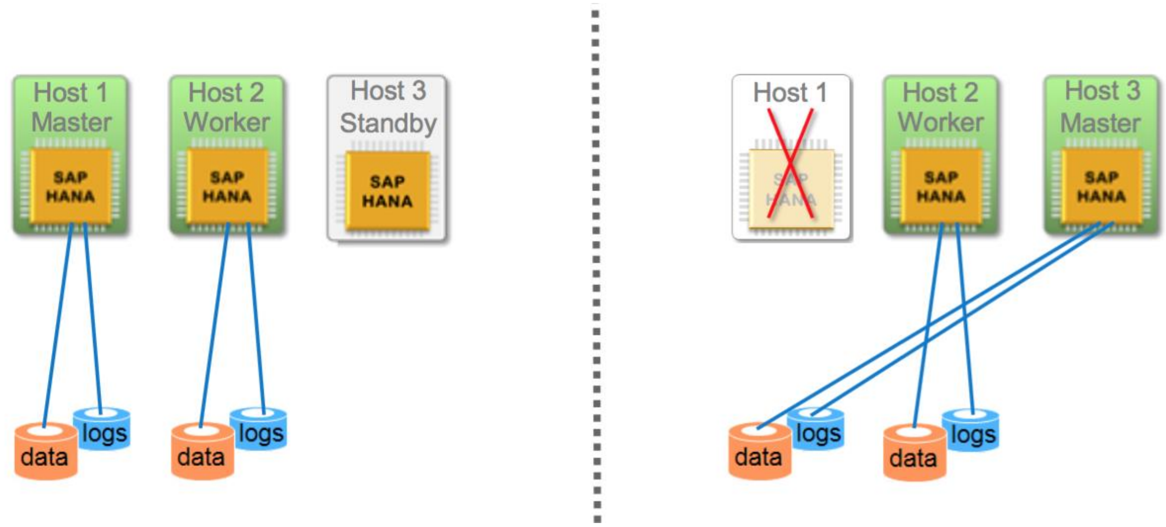


Figure 16: Master node fail-over

Master node fail-over with Row store test case

Populated SAP HANA® master node with only row store tables from 50GB to 350GB data and checking the fail-over times.

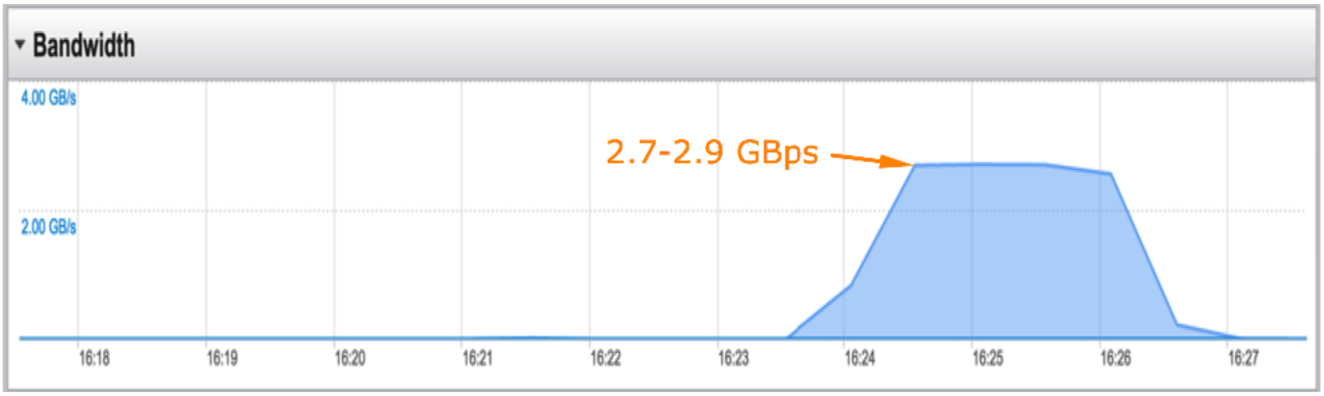


Figure 17: SAP HANA master node fail-over with 350 GB row store

Above is the run of 350 GB Row store fail-over run on FlashArray//m50. As you can see we were able to achieve read bandwidth of **2.7-2.9 GB/second on just standby master node** and whole loading of row store into memory of standby node took less than 3 minutes and 30 seconds.

Analysis of Master node fail-over with Row store test case

Below is the graphical analysis of SAP HANA® master node fail-over times with only row store tables (50GB to 350GB data). SAP HANA® after failure would take around 30 seconds to detect failure and then it will start the restart procedure on standby node. Then it would start the file IO operations when it starts loading row store.

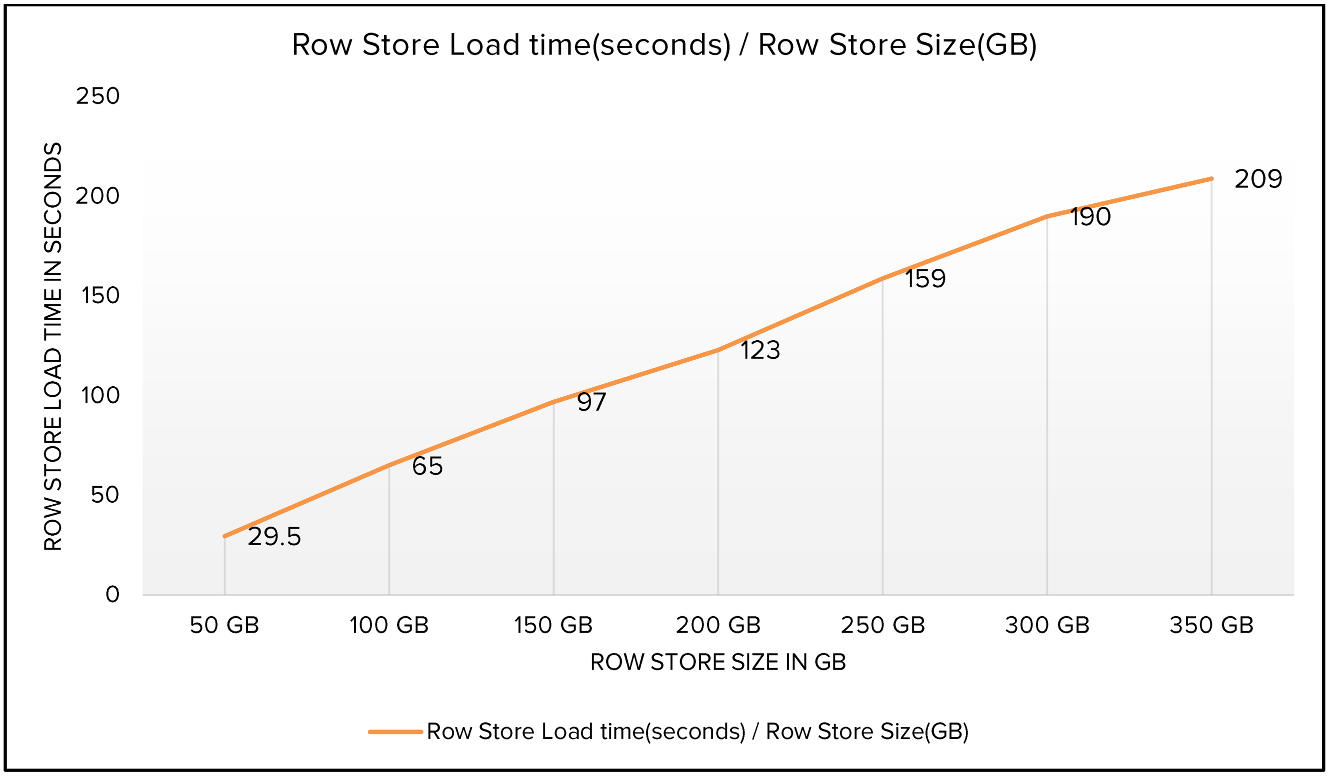


Figure 18: Row store load time(Seconds) – Row store size(GB) analysis

Summary

Failover recovery is rapid on FlashStack thereby reducing the downtime for business users no matter how much row store or preloaded column stores are present. This would reduce your fail-over related downtimes by at least 50%.

Important: It is extremely important to fine tune File system IO for SAP HANA® using hdbparams.Then only you will get the best performance as mentioned in the section. This is described in detail in the SAP HANA® best practices guide(See appendix for URL)

Rapid backup and recovery

Performing a full backup and recovery process can be a time-consuming process for SAP HANA® BASIS administrators. They spend a considerable amount of time and effort preparing the software environment and going through complicated procedures to ensure that the backup process is carried out with minimum disruption to the business and in accordance with best practices.

This section describes the speed and efficiency of performing a backup and recovery of SAP HANA® system FlashStack using Pure Storage snapshots.

FlashArray//m snapshots are instantaneous, even with large data volumes. Therefore, the entire process of a system backup and recovery of SAP HANA® is accelerated. This saves SAP Basis consultants a considerable amount of time and effort.

One of the cornerstones of FlashStack Pure Storage design principles is simplicity. This is clearly evident in how easy it is to snapshot/clone databases on a Pure Storage Array. This section we will attempt to showcase this feature for a very common use case – backup and recovery of a database instance. The database in this case is SAP HANA®. We will go over the necessary steps to take a snapshot a SAP HANA® instance for backup purposes and how to restore the database from the snapshot.

Preparing SAP HANA® to take a snapshot

SAP HANA® best practice dictates using the PREPARE statement prior to taking a backup copy. This step creates a database internal snapshot based on a system wide savepoint executed during the PREPARE. During this period no other backup should run, but log backup can continue as configured.

Step 1: PREPARE statement

From SAP HANA® Studio, perform the following steps:

In Backup node → Context menu → Manage Storage Snapshot→ Select Prepare

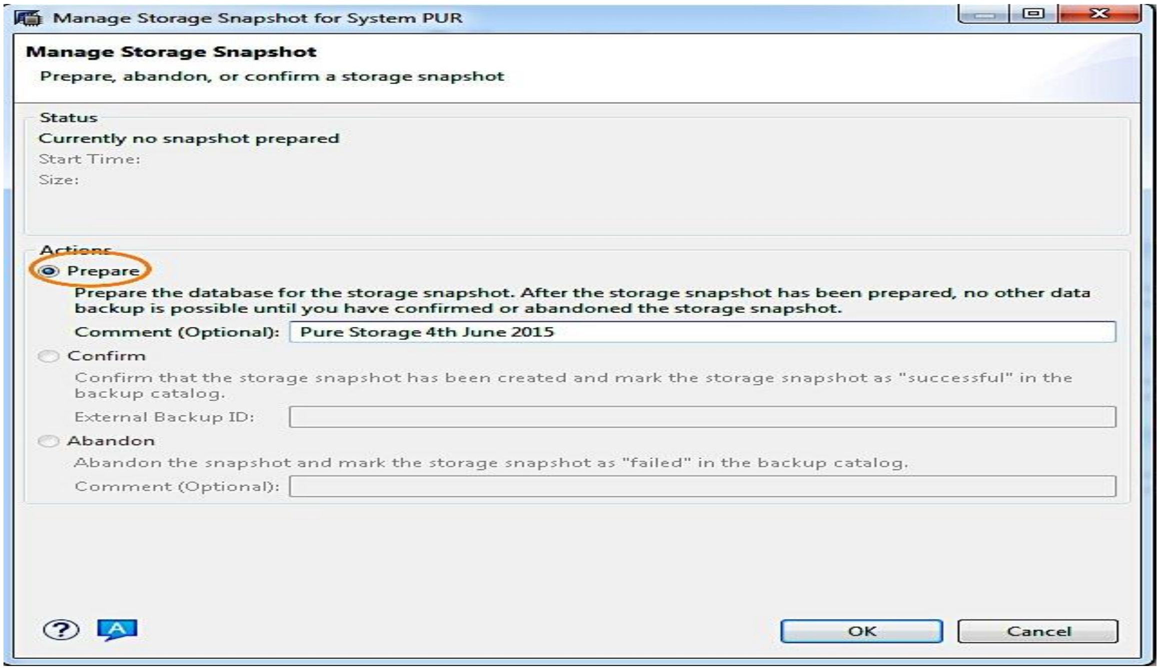


Figure 19: SAP HANA Studio: Manage storage snapshot

Alternatively, this step can also be performed with this SQL statement on the index server (The database user as BACKUP OPERATOR authorization):

BACKUP DATA CREATE SNAPSHOT BACKUP OPERATOR

Now the database is prepared for the storage snapshot. Verify the backup details, either through the backup catalog tab in SAP HANA® studio (make sure you capture important details, like ID):

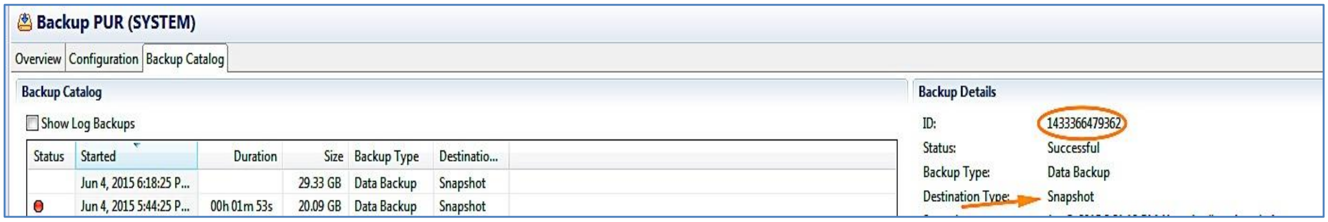


Figure 20: Backup catalog

Or by issuing the following SQL statement to get the backup details using the view M_BACKUP_CATALOG

select * from M_BACKUP_CATALOG where "ENTRY_TYPE_NAME" ='data snapshot'

→ Check the data volumes mount point and see if there is a file created as you see below snapshot_databackup_0_1

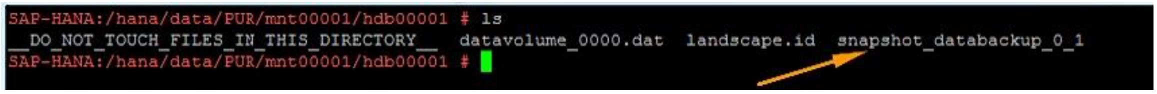


Figure 21: Linux shell data persistance main location

But be sure to freeze your file system (XFS) using xfs_freeze before you perform the above operation: xfs_freeze –f <Data Volume mount point>

And unfreeze it using: xfs_freeze –u <Data volume mount point>

Step 2: Create Pure Storage Snapshot

There are 2 methods for taking a snapshot on a Pure Storage array – from the Pure Storage array GUI or from the CLI. We will cover both methods.

For taking a snapshot using the Pure Storage GUI, go to the Storage tab and select the host associated with the SAP HANA® instance and select the data volume for the snapshot.

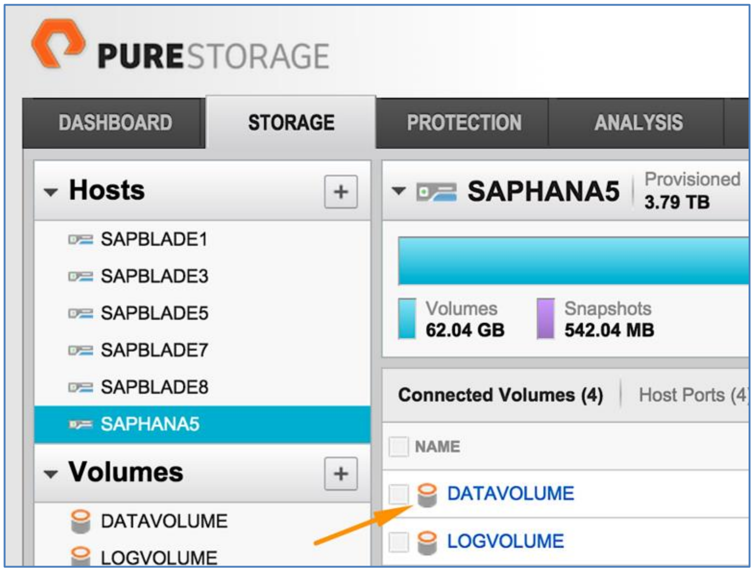


Figure 22: FlashArray//m GUI-Volumes

Click on the Gear on the right and Select "Create Snapshot." This creates a near instantaneous copy regardless of the size of the volume.

Figure 23: FlashArray//m GUI - snapshots

Snapshots can also be created via the following CLI command:

purevol snap <NAME OF THE DATA VOLUME>

Step 3: Confirm or Abandon the snapshot in SAP HANA®

After creating the snapshot, SAP HANA® has to be made aware that it has been created in order to confirm or abandon the snapshot. All the storage snapshots are recorded in SAP HANA® in the backup Catalog with the status successful or failed. This is very important, as SAP HANA® has to clear its internal state once it is confirmed. This action can either be done in SAP HANA® studio or via SQL statements.

Using SAP HANA® Studio:

Go to Manage Storage Snapshot and this time you will see only two options: Confirm or Abandon. Confirm the storage snapshot by using the External Backup ID generated. This is the same ID that was captured in Step 1.

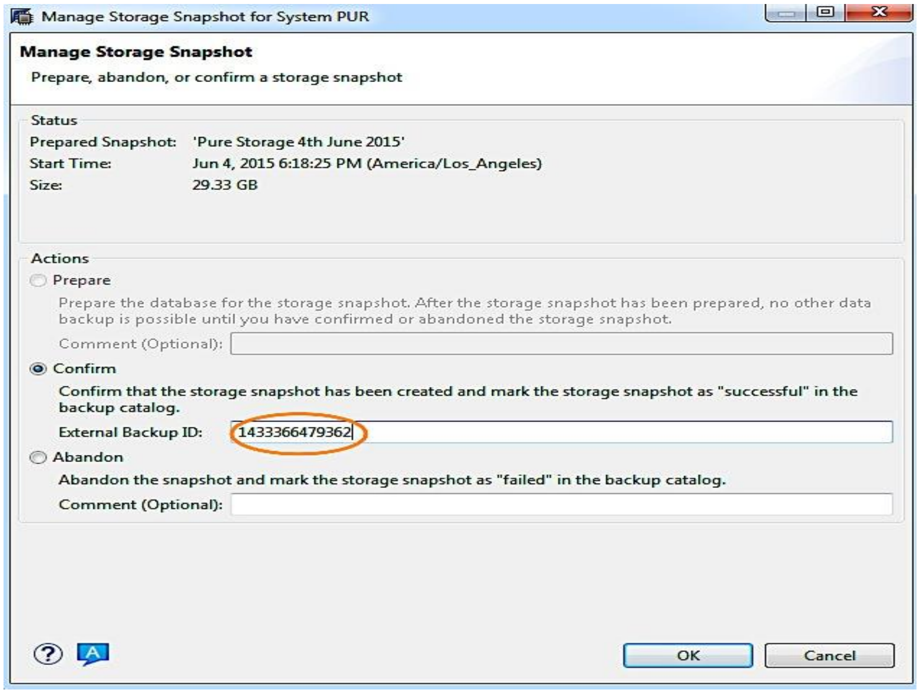


Figure 25: SAP HANA Studio: Manage storage snapshot

Using SQL Statements:

You can use the following SQL statement to perform the same action as above:

BACKUP DATA CLOSE SNAPSHOT BACKUP_ID <BACKUP_ID> SUCCESSFUL

Recovering SAP HANA® from a snapshot

After a storage snapshot is created, it can be used to recover the SAP HANA® instance. There are multiple options for recovering an instance using a storage snapshot:

1. Recover the database to its most recent state (Only available option for Multitenant database containers)

- Uses the most recent storage snapshot
- Log backups made since the most recent data backup

- Log area

2. Recover the database to a point in time

    - Uses the last storage snapshot before the specified point in time
    - Log backups made since the data backup to be used until the desired point in time
    - Log area

3. Recover the database to a specific storage snapshot

    - Uses a specific storage snapshot for recovery
    - All logs will be lost; it is like it begins a new database lifecycle

4. Recover the database to a specific log position

    - The most recent storage snapshot available before the specified log position
    - Log backups made since the storage snapshot to be used
    - Log area

In this document, we will demonstrate recovery of an SAP HANA® instance to a particular point in time.

Recovering SAP HANA® to a point in time

Step 1: SAP HANA® in recovery mode

In SAP HANA® studio, from the main context menu select → Backup and Recovery → Recover System. This will shut down SAP HANA® database.

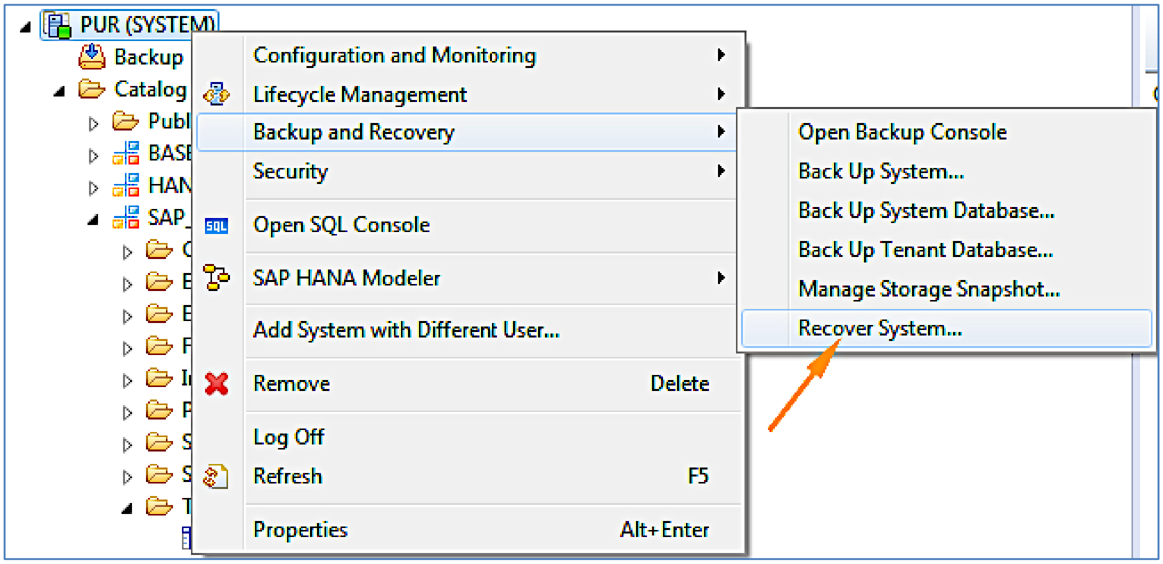


Figure 26: SAP HANA Studio: Recover System…

Step 2: Restore the data volume from the snapshot

The recover from snapshot process can be done either from the Pure Storage GUI of from the CLI.

To initiate the process from the GUI, Select the option "Restore volume from snapshot" in the Storage tab

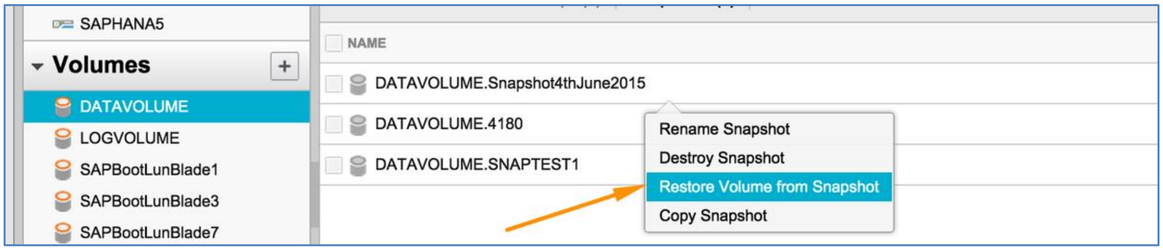


Figure 27: FlashArray//m GUI

To perform the same action from the CLI, issue the following command

purevol recover

Once the volume has been recovered, perform the following steps –

- Unmount the data volume (umount /hana/data)
- Mount –a (to refresh the data volume to reflect the recovery from snapshot)
- Mount again (mount </dev/mapper/ID> /hana/data

Once recovery of the data volume is complete, ensure that the file is present

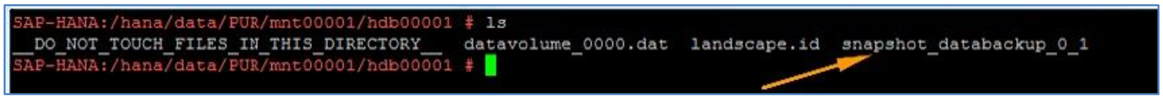


Figure 28: Linux shell data persistance main location

Step 3: Recover SAP HANA®

Select the date and time to the point in time you want to recover SAP HANA®.

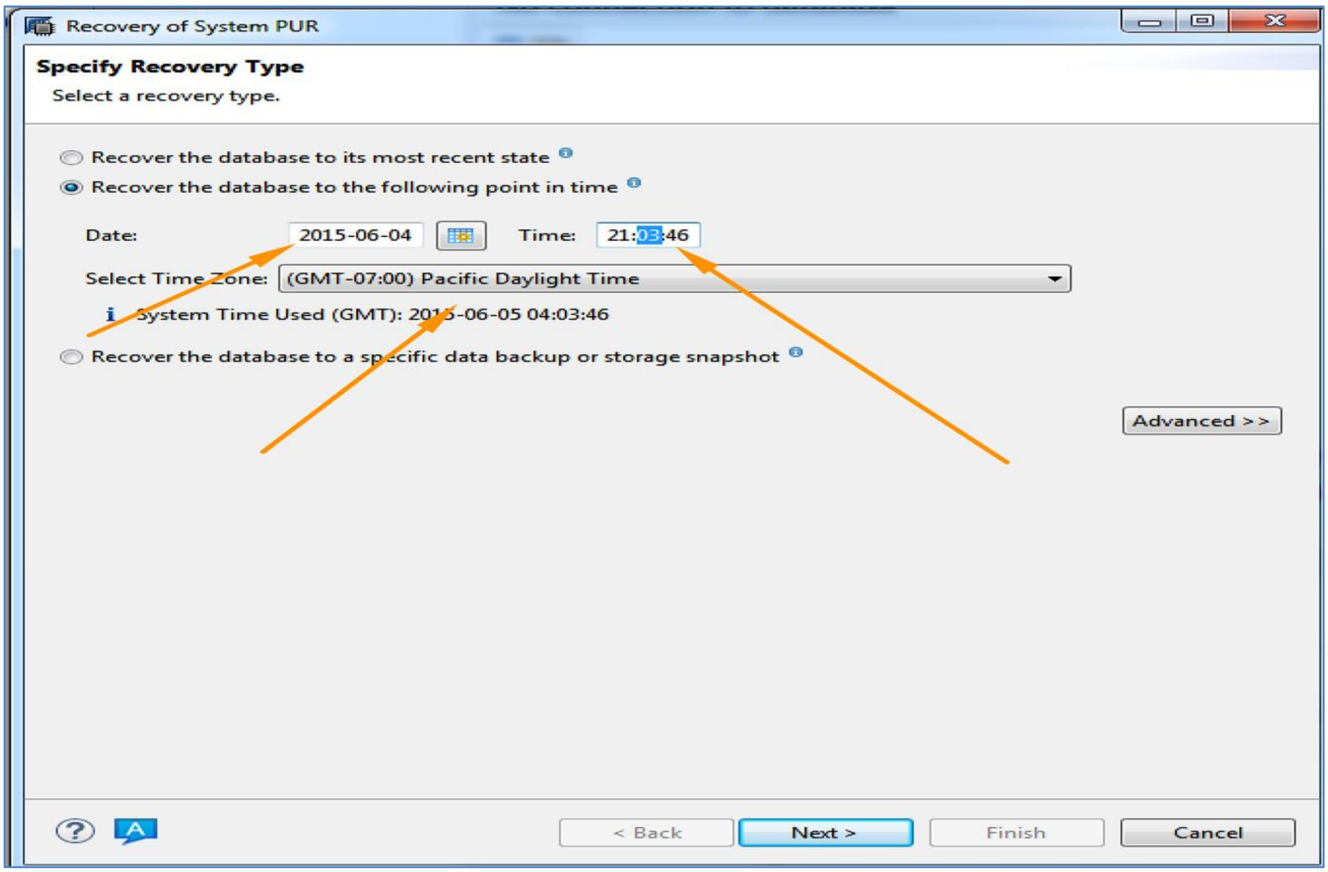


Figure 28: SAP HANA studio: Recovery of System

Enter all the locations of log backups then click “Next”

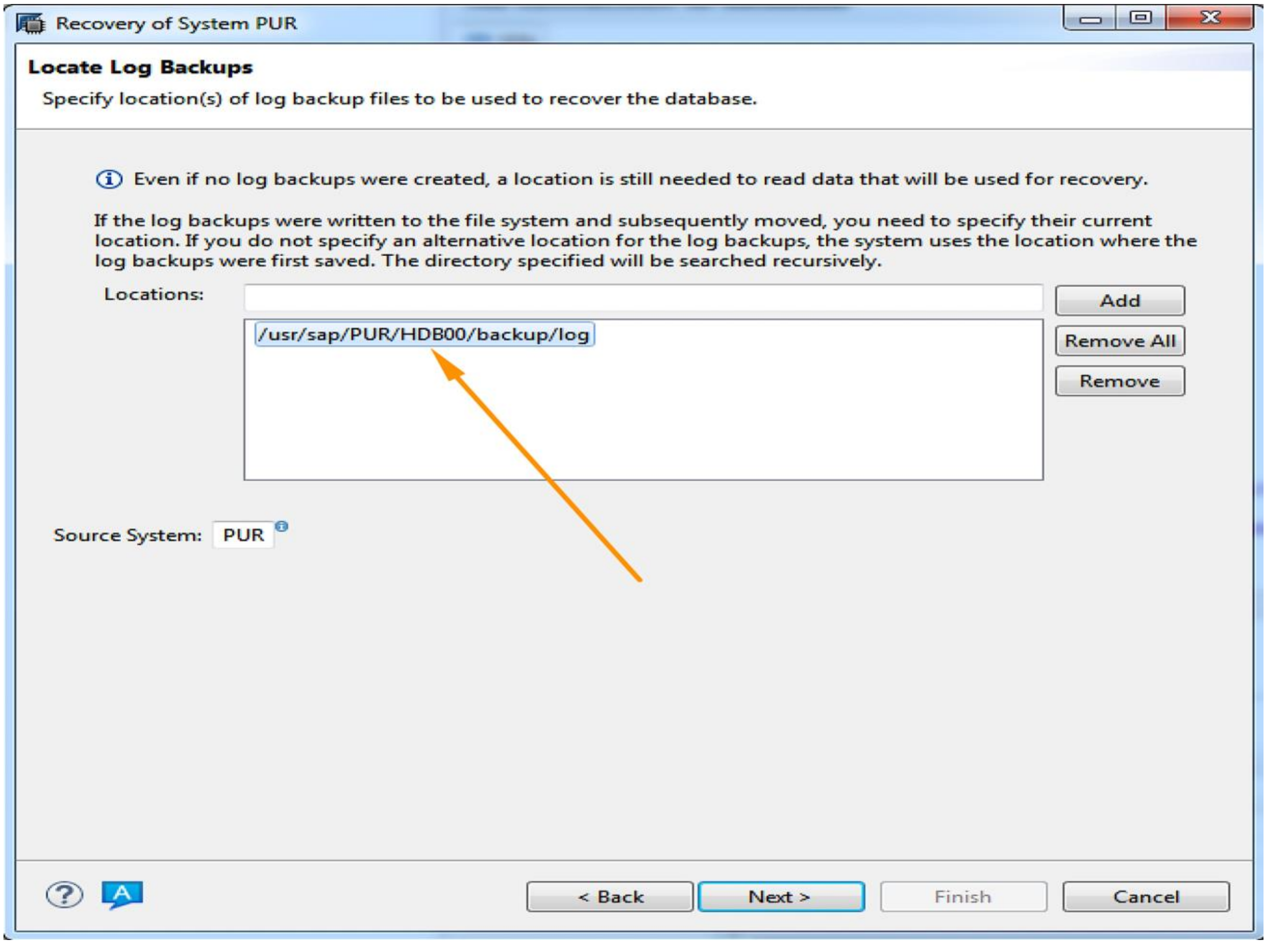


Figure 29: SAP HANA studio: Locate Log Backups

Based on the time entered for recovery, SAP HANA® will select/suggest the best backup using which it can be restored in shortest time, then click "Next"

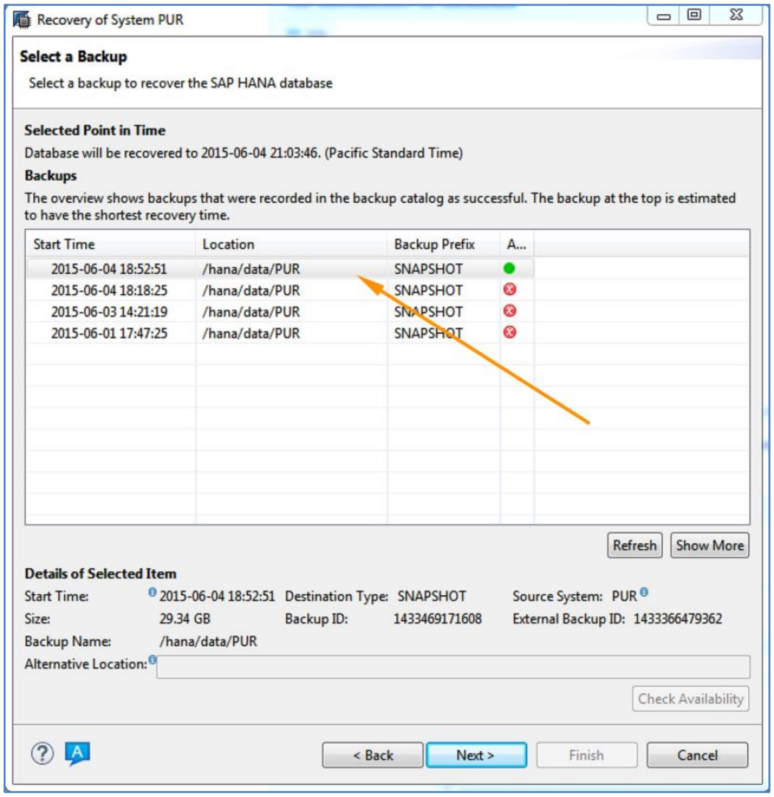


Figure 30: SAP HANA studio:Backups catalog

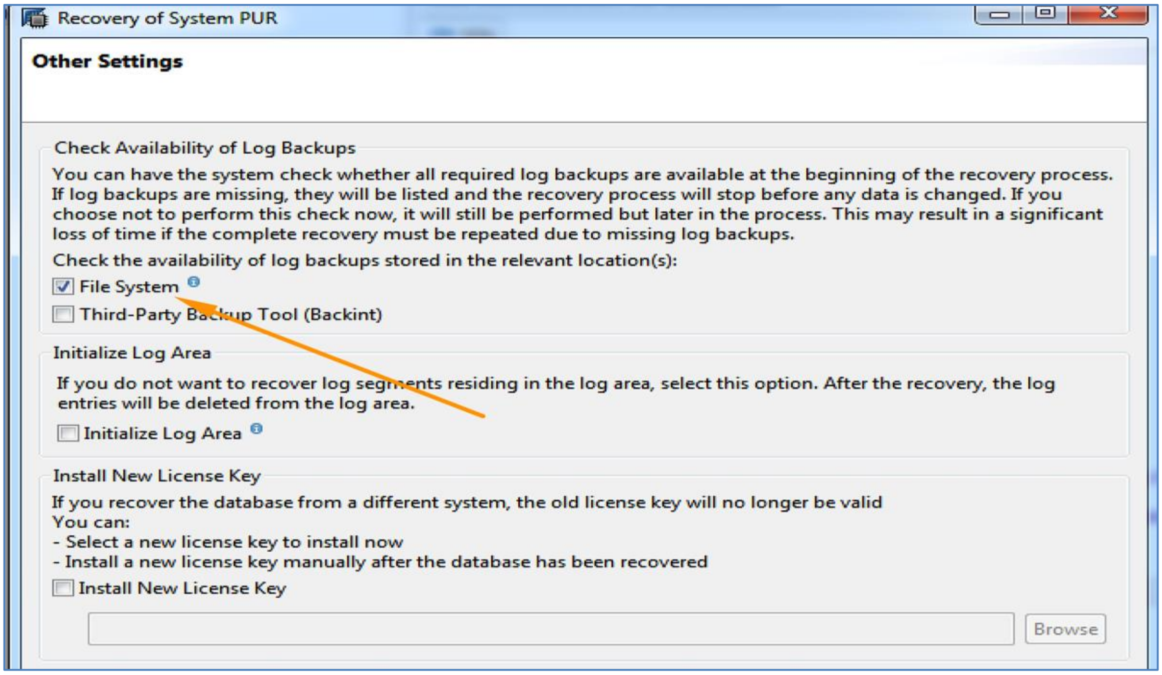


Figure 31: SAP HANA studio:Recovery of system

Review and complete recovery process

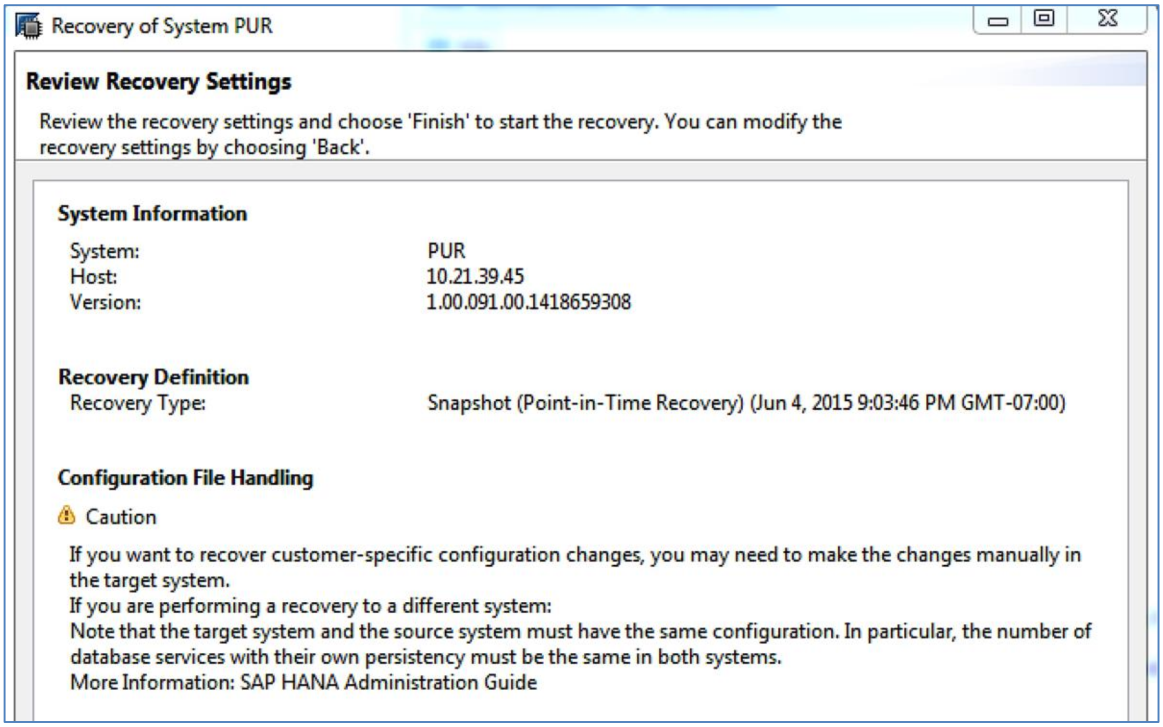


Figure 32: SAP HANA studio:Recovery of system

Now SAP HANA® will recover Data, Logs and Restart all the configured services.

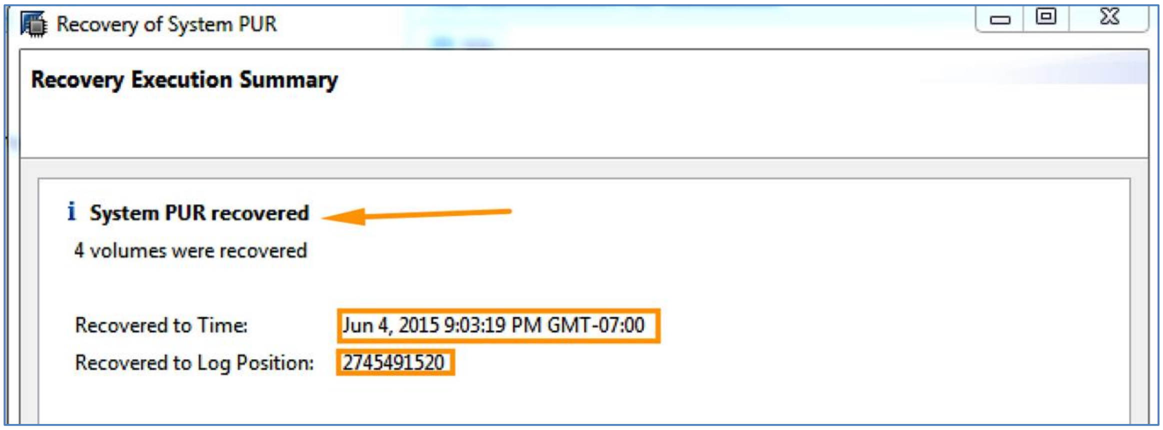


Figure 33: SAP HANA studio:Recovery of system

The same result can be achieved by the using the following SQL statement:

RECOVER DATABASE UNTIL TIMESTAMP '2015-06-05 04:03:46' USING DATA PATH ('/usr/sap/PUR/HDB00/backup/data') USING LOG PATH ('/usr/sap/PUR/HDB00/backup/log') USING SNAPSHOT CHECK ACCESS USING FILE

Summary

Backup and recovery using Pure Storage FlashRecover snapshots is extremely fast and efficient. The whole process can be automated and there is no need to invest on Backint 3rd party SAP certified tools. As mentioned this whole process can be automated very easily via a script and can be scheduled on a regular basis. This process allows customers to achieve very low RTO and RPO and save on costs of Backint tools.

The below graphic shows how fast backup and recovery is on FlashArray//m when compared to a legacy storage. Backup is instantaneous due to Pure Storage FlashRecover snapshots and recovery is also 3-4 times faster due to our superior read performance. Overall we are 6-8 times faster for backup and recovery process on FlashStack.

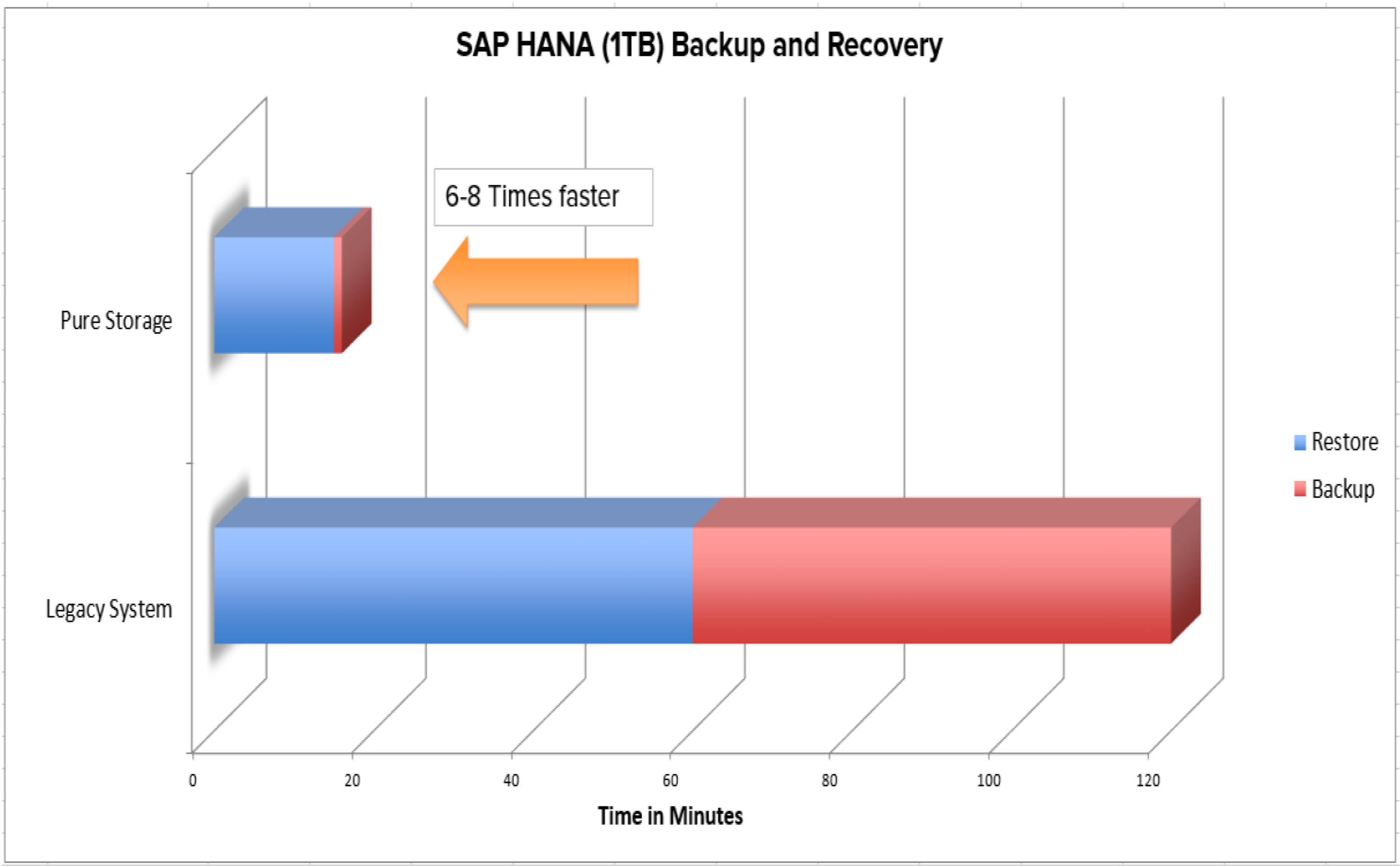


Figure 34: SAP HANA Backup and recovery comparision (Legacy/Pure Storage)

SAP Lumira Dashboard query performance

In this section we will dissect the query performance of sales dashboard built using SAP Lumira deployed on SAP HANA®. Here we will focus on how FlashArray//m makes a difference to query runtimes.

SAP HANA® is an in-memory database but still query performance depends a lot on the storage. Two important factors which contribute query performance are

- Loading of tables into memory
- Delta-merge of column tables

In order to test the query performance based on above contributors we have created a sales dashboard in SAP Lumira which belong to imaginary company ABC incorporated. This sales dashboard is used by Head of sales department very frequently for doing real-time analysis of company's best products, top sales teams, top regions etc. The SAP HANA® system is connected to a CRM application which is replicating the data to SAP HANA® in real-time.

Below is the sales dashboard composed of 4 analytical queries. We will do performance test of individual queries on tables which are not loaded into memory and also in partial delta merge situations. This will show us the importance of storage and how well FlashArray//m fares in such situations.

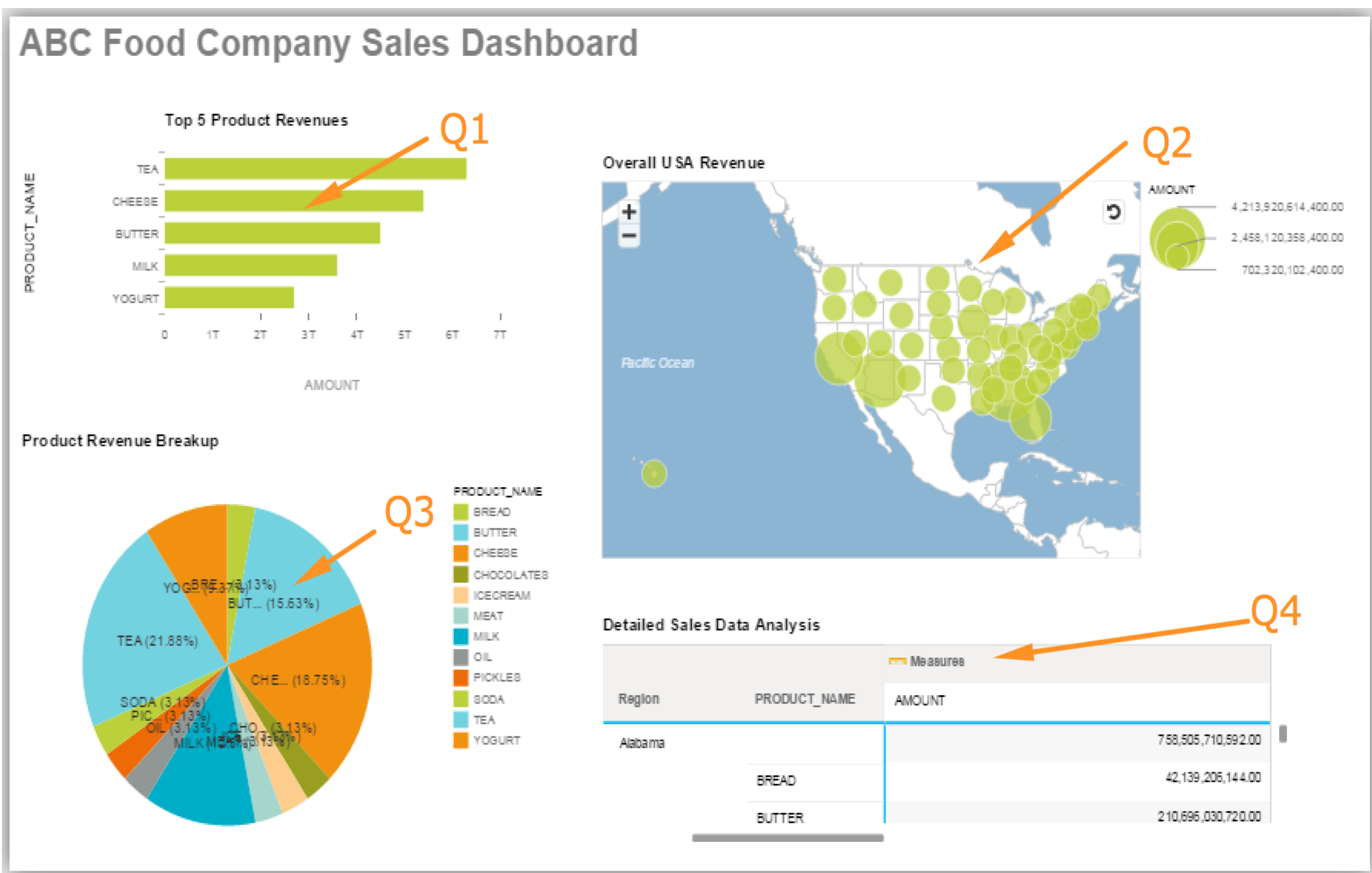


Figure 35: SAP Lumira sales dashboard

Dashboard queries

Below are the breakdown of SAP Lumira dashboard queries hitting SAP HANA® Analytical view. The same column table "SALES_DATA_1_BILLION" is partitioned multiple times. Firstly this column table's partitioning criteria is satisfying the query dimensions, then it was repartitioned not to satisfy these analytical queries dimensions. All the below queries do the entire 1 billion records scan for the respective columns.

Query ID	Query (From SAP HANA® SQL Plan Cache)	Analytics View	Column Table
Q1(Top 5 Product Revenues)	select top 5 "PRODUCT_NAME" ,"AMOUNT" FROM (SELECT "C2" AS "PRODUCT_NAME","M1" AS "AMOUNT" FROM (SELECT DISTINCT sum("AMOUNT") AS M1, "PRODUCT_NAME" AS C2 FROM "_SYS_BIC"."PURE/AV_SALES_DATA_FINAL" GROUP BY "PRODUCT_NAME" ORDER BY M1 desc, C2 asc LIMIT 10000))	PURE/AV_SALES_DATA_FINAL	SALES_DATA_1_BILLION
Q2(Overall USA revenue)	SELECT DISTINCT "STATE" AS C1, sum("AMOUNT") AS M2 FROM "_SYS_BIC"."PURE/AV_SALES_DATA_FINAL" GROUP BY "STATE" ORDER BY C1 asc LIMIT 10000	PURE/AV_SALES_DATA_FINAL	SALES_DATA_1_BILLION
Q3(Product Revenue Breakup)	SELECT DISTINCT "PRODUCT_NAME" AS C1, sum("AMOUNT") AS M2 FROM "_SYS_BIC"."PURE/AV_SALES_DATA_FINAL" GROUP BY "PRODUCT_NAME"ORDER BY C1 asc LIMIT 100000	PURE/AV_SALES_DATA_FINAL	SALES_DATA_1_BILLION
Q4(Detailed Sales Data analysis)	SELECT DISTINCT "YEAR" AS C1, "STATE" AS C2, "PRODUCT_NAME" AS C3, sum("AMOUNT") AS M4 FROM "_SYS_BIC"."PURE/AV_SALES_DATA_FINAL" GROUP BY "YEAR", "STATE", "PRODUCT_NAME" ORDER BY C1 asc, C2 asc, C3 asc LIMIT 100000	PURE/AV_SALES_DATA_FINAL	SALES_DATA_1_BILLION

Table 5: Analytical sales queries

Column table information

To test the delta merge and first time loading of table scenarios we have created a Column table `SALES_DATA_1_BILLION` with 1.144 billion records. As mentioned before it is partitioned multiple times one to satisfy the analytical queries and repartitioned to not satisfy analytical queries dimensions.

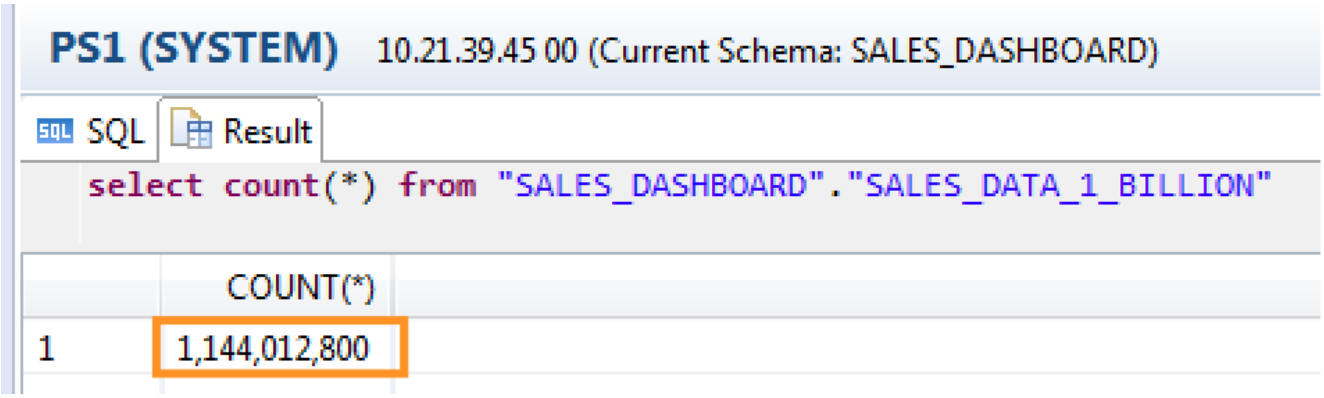


Figure 36: SAP HANA studio: column table information

This table was partitioned by both hash and range partitioning with range partitioning on “YEAR”. The hash partitioning was done multiple times.

For e.g. as seen below this table was hash partitioned on PRODUCT to satisfy the queries based on PRODUCT like Q1 and Q2 above. Then it was later repartitioned not to align with any query dimensions.

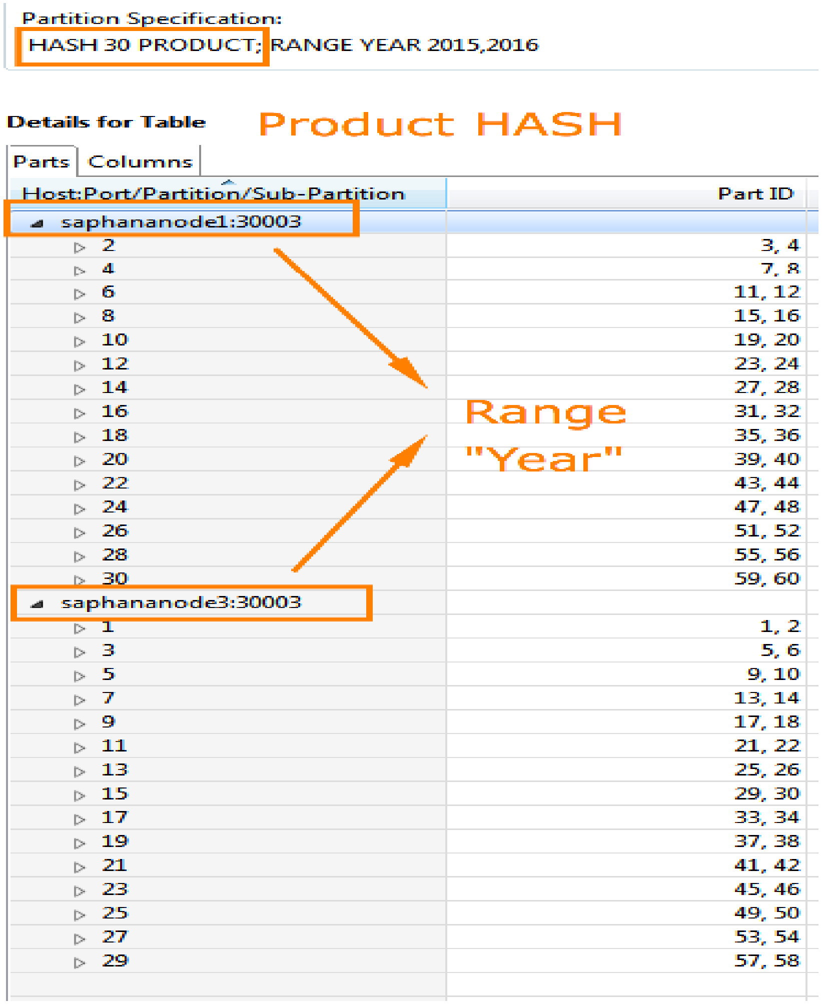


Figure 37: SAP HANA studio: column table partitioning information

Analytics configuration in SAP HANA®

An analytical view was created on top of the column store for analytical queries. This enables analytical queries to use OLAP engine in SAP HANA®. As seen below is the Analytical view on SALES_DATA_1_BILLION column table. Also attribute views have been created to join master data with factual data. Attribute views are joined to Analytical view using the left outer join also known as star join or star schema.

SAP Lumira Sales dashboard's queries are built using the below analytical view which uses the SALES_DATA_1_BILLION column table as data foundation.

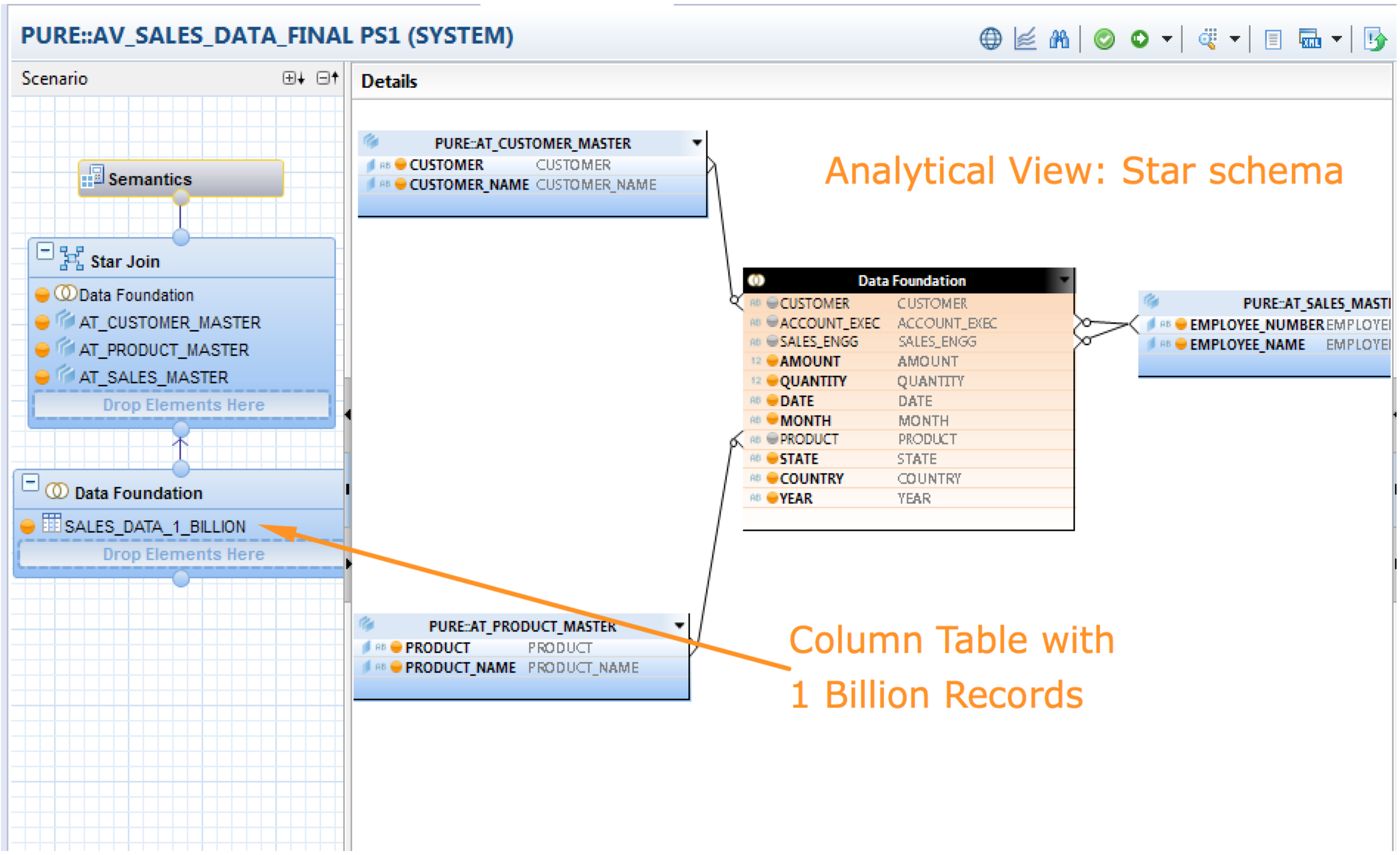


Figure 38: SAP HANA studio: Analytical view information

Loading of tables into memory - Query performance affect

One of the important factor to have the best query performance to satisfy business users is that all the data for the query should be in memory. This is not often the case though as after a SAP HANA® restart column stores will not be loaded in to memory if it is not set for preload flag and also another reason is SAP HANA® offloads to tables to storage if SAP HANA® is running out of memory. SAP HANA® unloads the data from memory into storage and this happens quite frequently.

Unloading process can be monitored by activating the unload_trace configuration for the index server. This can be traced using the → indexserver_<hostname>.<port>.unloads.<file_id>.trc or it can queried on the table "M_CS_UNLOADS".

When one of these situations happens immediately the query performance takes a hit and your entire dashboard slows down.

Running your analytics solution on FlashStack will not slow your dashboard by a great extent and for this we will analyze the runtime of the queries in the dashboard when the data is NOT present in the memory and we will compare it with runtime of the same query when the data is present in the memory.

For this we will use the sales dashboard with queries mentioned previously and we will check the runtime of each of these queries which are based on the 1.1 billion records column table.

Every time to test the data not present in the memory, the column store is unloaded from memory using UNLOAD TABLE statement.

IO pattern of table reload in data volume: READ

4 KB – 16 MB blocks, up to 64 MB (clustered Row Store super blocks)

Scenario 1: Query dimensions align with Partition criteria (Best case scenario)

Testing query runtimes when the partitions aligns with query dimensions. All these queries need to perform entire table scan for 1 Billion records as they were ran without any filter. SAP HANA® loads only the columns and partitions which are addressed by query when table in unloaded. So when query dimensions aligns with partition criteria, the query runs much faster as loading is much more parallelized.

Query Name	Query Description	Runtime(Data Not present in memory)	Runtime(Data present in memory)
Q1	Top 5 Product Revenues	1.1 seconds	0.479 second
Q2	Overall USA revenue	1.2 seconds	0.503 second
Q3	Product Revenue Breakup	1.2 seconds	0.569 second
Q4	Detailed Sales Data analysis	2.42 seconds	0.772 second

Table 6: Analytical sales queries runtimes comparision (Best case scenario)

Scenario 2: Query dimensions does not align with Partition criteria (Worst case scenario)

Testing query runtimes when the partitions do not align with query dimensions. All these queries need to perform entire table scan for 1 Billion records as they were ran without any filter. SAP HANA® loads only the columns and partitions which are addressed by query when table in unloaded. So when query dimensions do not align with partition criteria, the slower the query runs slower as loading partitions not aligned with query dimensions has a negative impact. This is because it creates an additional overhead as multiple partitions needs to be searched.

Query Name	Query Description	Runtime(Data Not present in memory)	Runtime(Data present in memory)
Q1	Top 5 Product Revenues	2.071 seconds	0.639 second
Q2	Overall USA revenue	2.3 seconds	0.721 second
Q3	Product Revenue Breakup	2.47 seconds	0.753 second
Q4	Detailed Sales Data analysis	4.87 seconds	0.872 second

Table 7: Analytical sales queries runtimes comparision (Worst case scenario)

Below is the graphical analysis showing the comparison of dashboard queries runtime performance when the data is in memory against when the data is not in memory for both the scenarios. As seen the best case scenario the query performance is so close to the query performance to when the data is in memory. In the worst case scenario when partitions don't align with the query dimensions then also the runtimes are comparable and does not slow down considerably. This shows the superior read performance of the FlashArray//m.

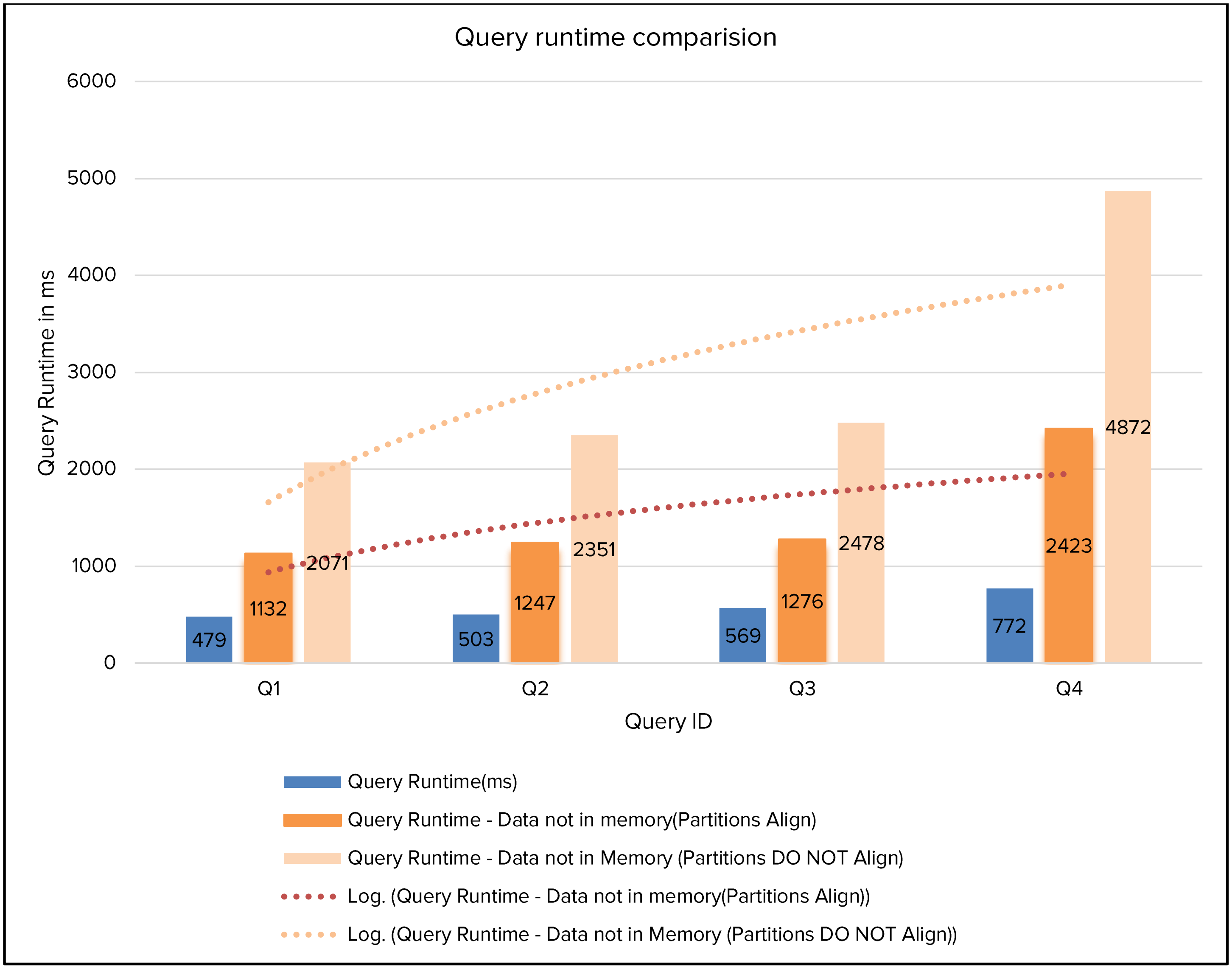


Figure 39: Query runtime comparision graph analysis

Summary

SAP HANA® analytics solution on FlashStack does not affect the query performance of real-time reporting performance by a great extent. As seen above the query performance when the data was not present in memory was quite close to the query performance of the data present in memory due to FlashArray//m amazing read performance. It is even better performance when the tables are partitioned and aligned to queries dimensions as seen above

Delta merge – Impact on Query performance

A column store table is made of two stores types, for each column a main store and a delta store. The delta storage is optimized for write operations and the main storage is optimized in terms of read performance and memory consumption. The use of the delta tables addresses the performance issues of loading directly to compressed columns.

As described in the Figure "Read and Write Operations on a Column Store Table" below, read operations are performed on both parts whereas write operations only affect the delta part. In order to optimize query execution performance of the system and to ensure optimum compression, the system needs to transfer the data from the delta part into the main part. This process is called delta merge.

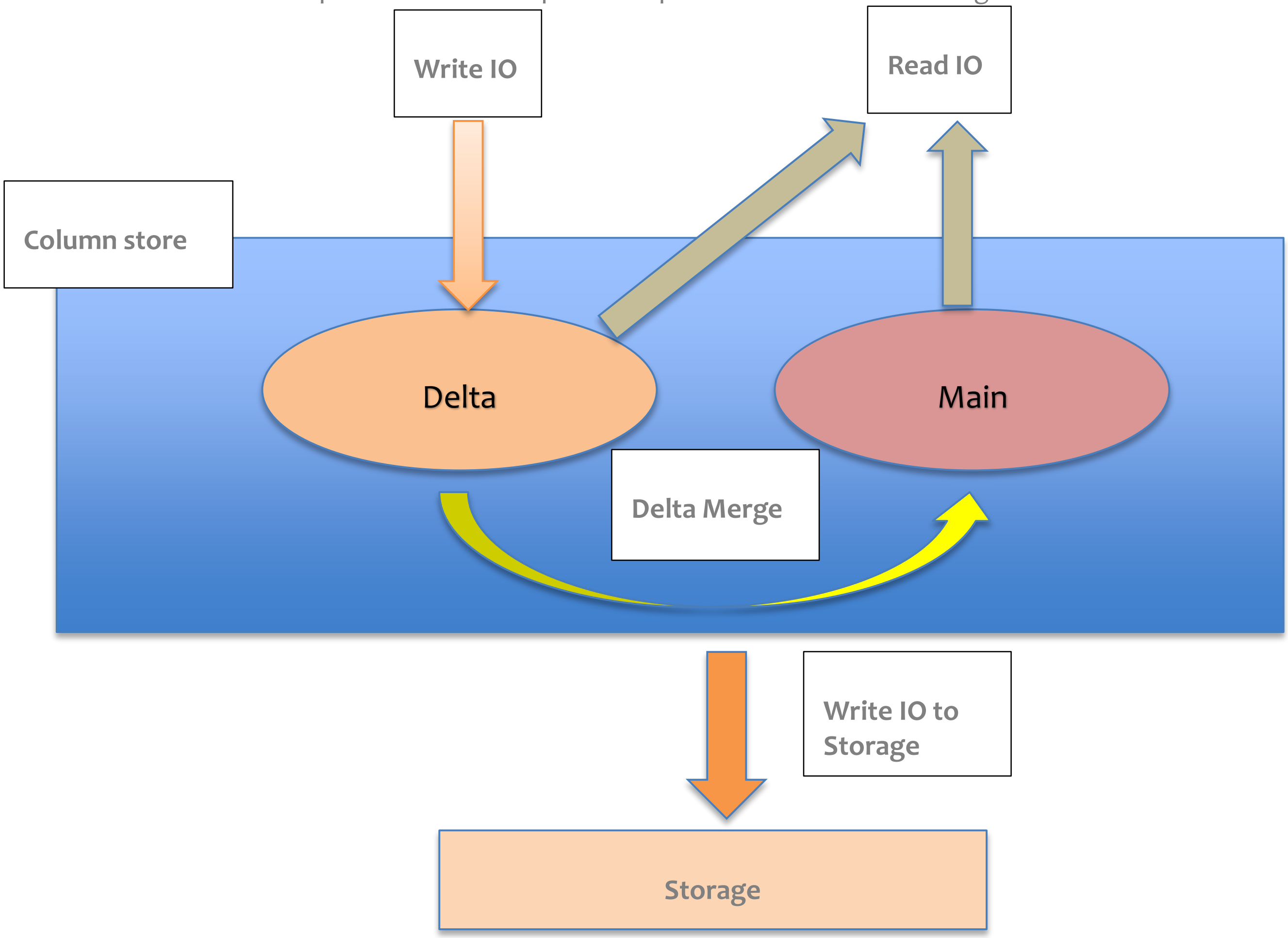


Figure 40: Delta merge

Delta merge is an extremely important operation as it affects query performance drastically. This is because if delta merge does not happen then query has to read data from delta which is read optimized and it slows down the query. After each delta merge there is a huge write IO to storage, as it persists the main store.

IO pattern of delta merge to data volume: WRITE

4 KB – 16 MB asynchronous bulk writes

If storage cannot handle delta merge fast enough and efficiently then it has two negative effects:

- Query performance decreases drastically.
- Compression is SAP HANA® decreases as delta stores are not compressed. This increases the memory utilization.

Now let us take the above Sales Dashboard and see the effects of query performance when the column store is not completely delta merged. Below table shows the detailed runtimes of a partition aligned queries when the column store is not completely delta merged. Also we will track the increase in memory footprint for the column store table when delta merge is not completely done.

Query Name	Query Description	Query Runtime(ms) 100% Delta merged	Query Runtime(ms) 95% Delta merged	Query Runtime(ms) 90% Delta merged	Memory footprint increase 95% Delta merged	Memory footprint increase 90% Delta merged
Q1	Top 5 Product Revenues	479 ms	565 ms	622 ms	~1.2 times more	~1.35 times more
Q2	Overall USA revenue	503 ms	583 ms	644 ms	~1.2 times more	~1.35 times more
Q3	Product Revenue Breakup	569 ms	654 ms	705 ms	~1.2 times more	~1.35 times more
Q4	Detailed Sales Data analysis	772 ms	837 ms	949 ms	~1.2 times more	~1.35 times more

Table 8: Partial delta merge impact analysis

As seen above even if column table is 90-95% delta merged that means only 5-10% of 1.144 Billion rows in delta store then also the query performance drops by 15 to 30% and also the memory consumption increases from 1.2 to 1.35 times. This shows how important delta merge is and it is extremely important that storage can handle delta merges efficiently.

Delta Merge performance

To show the performance of delta merge on FlashArray//m we will not consider the other bottlenecks like CPU (Merge Token) or memory. Here we will only consider the performance of FlashArray//m to see how efficiently it can absorb a delta merge. The tables used to test the delta merge performance are having 23 columns and structure width of 576 Bytes

To make sure that CPU or merge token and memory does not became the bottleneck we have configured the mergedog (Merge monitor) in SAP HANA® with enough tokens or CPU and also increased the delta memory size. Here we tested how fast delta merge can be absorbed using the multiple tables in parallel sales column stores each having 100 million records in delta store.

These column store tables are triggered in parallel using a Java multi-threaded program to issue "Merge delta" so that it SAP HANA® can execute delta merge in parallel as seen below. We were able to scale from 500 million delta store records across 5 tables to 2 billion records across 20 tables present on two active nodes of SAP HANA®.

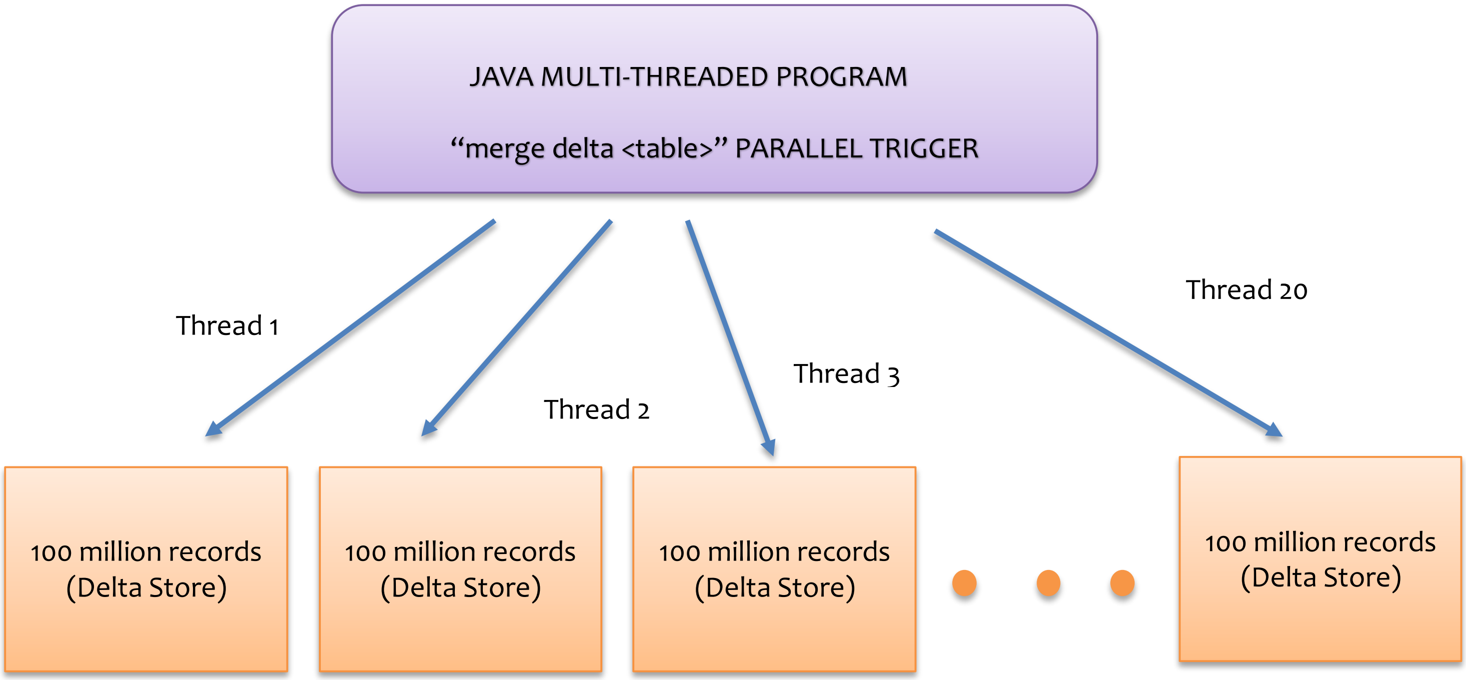


Figure 40: Delta merge performance testing (Parallel trigger using Java program)

The performance of delta merge was monitored using the statistics table “M_DELTA_MERGE_STATISTICS”, as seen below the Start_time column shows when all the delta merges are activated and execution time column shows the delta merge duration for each of the tables.

SQL | Result

```
select * from M_DELTA_MERGE_STATISTICS where "SCHEMA_NAME" = 'DELTA_MERGE_TESTING' and "SUCCESS" = 'TRUE' ORDER BY "START_TIME" DESC
```

Parallel Trigger

SCHEMA_NAME	TABLE_NAME	PART_ID	HISTORY	MEMORY_...	PASSPORT	START_TIME	EXECUTION_TIME	MOTI...	SUCCESS	MERGED_DELTA_RECORDS
DELTA_MERGE_TESTING	SALES_DATA_DM_12	0	FALSE	FALSE		Feb 6, 2016 9:57:49.197 PM	17,979	HARD	TRUE	100,000,000
DELTA_MERGE_TESTING	SALES_DATA_DM_14	0	FALSE	FALSE		Feb 6, 2016 9:57:49.194 PM	18,178	HARD	TRUE	100,000,000
DELTA_MERGE_TESTING	SALES_DATA_DM_15	0	FALSE	FALSE		Feb 6, 2016 9:57:49.194 PM	18,177	HARD	TRUE	100,000,000
DELTA_MERGE_TESTING	SALES_DATA_DM_11	0	FALSE	FALSE		Feb 6, 2016 9:57:47.689 PM	18,257	HARD	TRUE	100,000,000
DELTA_MERGE_TESTING	SALES_DATA_DM_1	0	FALSE	FALSE		Feb 6, 2016 9:57:47.686 PM	19,116	HARD	TRUE	100,000,000
DELTA_MERGE_TESTING	SALES_DATA_DM_13	0	FALSE	FALSE		Feb 6, 2016 9:57:46.82 PM	17,558	HARD	TRUE	100,000,000
DELTA_MERGE_TESTING	SALES_DATA_DM_3	0	FALSE	FALSE		Feb 6, 2016 9:57:46.815 PM	18,584	HARD	TRUE	100,000,000
DELTA_MERGE_TESTING	SALES_DATA_DM_2	0	FALSE	FALSE		Feb 6, 2016 9:57:46.815 PM	17,863	HARD	TRUE	100,000,000
DELTA_MERGE_TESTING	SALES_DATA_DM_4	0	FALSE	FALSE		Feb 6, 2016 9:57:46.815 PM	17,336	HARD	TRUE	100,000,000
DELTA_MERGE_TESTING	SALES_DATA_DM_5	0	FALSE	FALSE		Feb 6, 2016 9:57:46.81 PM	18,159	HARD	TRUE	100,000,000

Figure 41: Delta merge statistics table

In the table and graph below the execution time of delta merge is shown when we scaled it from 5 tables with 500 million records aggregated in delta store to 2 billion records in delta store across 20 tables.

As seen below in the table/graph the delta merge execution time was flat as we scaled it from 500 million records to 2 billion records. Delta merge is very efficiently absorbed by FlashArray//m. Even for 2+ billion records it just took less than 22 seconds to perform a delta merge.

Column Store	Number of Records	Number of tables	Delta Merge duration
SALES_DATA_DM*	500 million	5	17 seconds
SALES_DATA_DM*	1000 million	10	19.3 seconds
SALES_DATA_DM*	1.5 billion	15	21 seconds
SALES_DATA_DM*	2 billion	20	22.5 seconds

Table 9: Delta merge performance test results

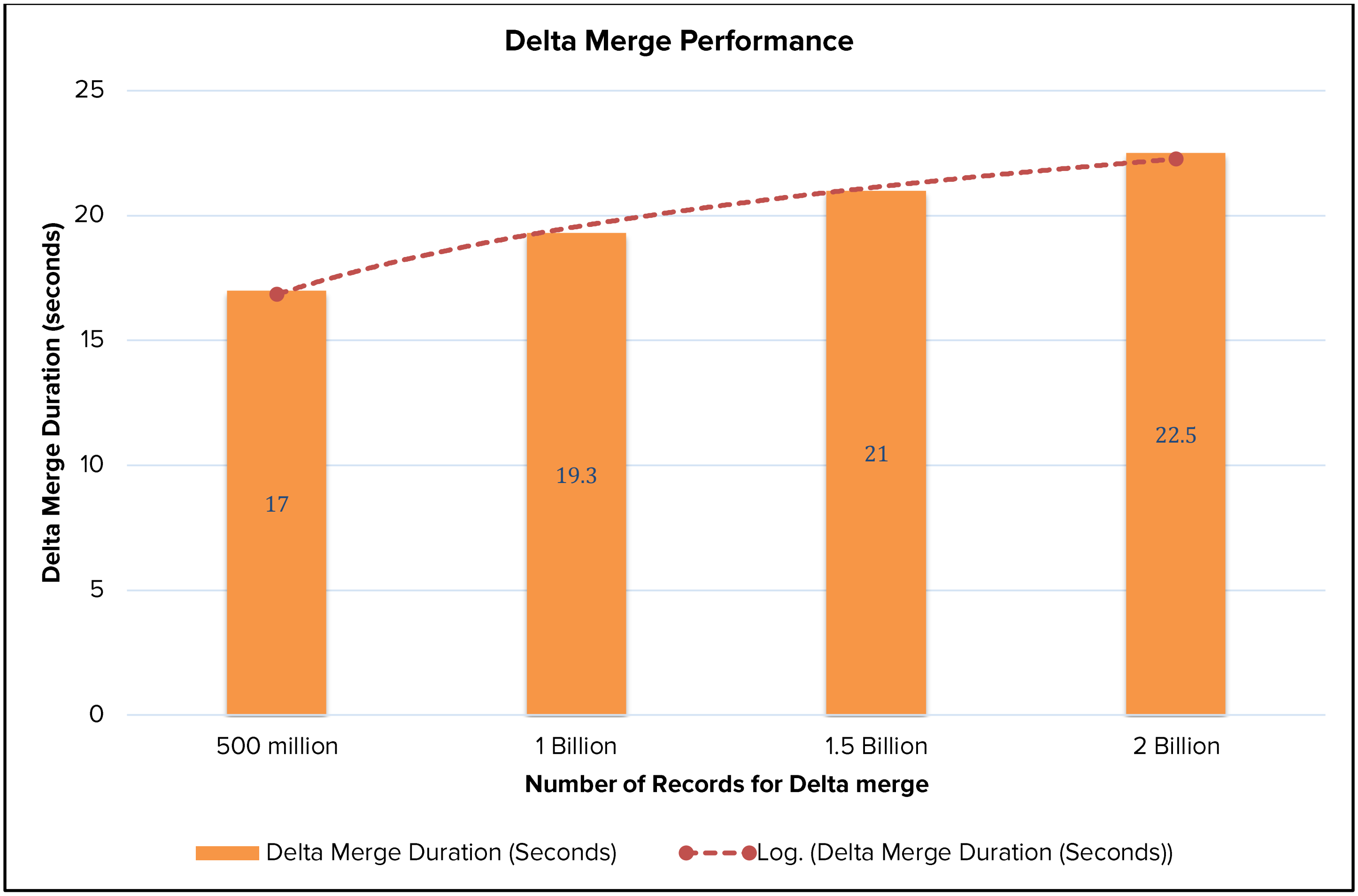


Figure 42: Delta merge performance graph analysis

Summary

SAP HANA® analytics solution on FlashStack does not affect the query performance of real-time reporting performance by a great extent as delta merge lagging will not occur. As seen above the delta merge are 2-3 times faster and efficient on FlashArray//m.

Real-time data load or replication performance

Redo log entries are persisted in log volumes before the write transaction completes. Hence the write performance of log volumes of SAP HANA® is extremely important especially when there is huge write load on SAP HANA®. This kind of huge write load can happen on SAP HANA® during an SAP SLT side car approach where data from SAP ECC is replicated to SAP HANA® in real-time or on BW on HANA® when the daily-nightly load happens.

Here to simulate such scenarios we have used Apache JMeter tool to do the load testing by creating multiple threads loading tables similar to delta merge testing tables with structure width of 576 Bytes and 23 columns. As shown below we created a continuous write load with 6 tables and targeting 5 threads for per table doing 80% inserts, 15% updates, 5% deletes with batch commit of 1000 records per thread.

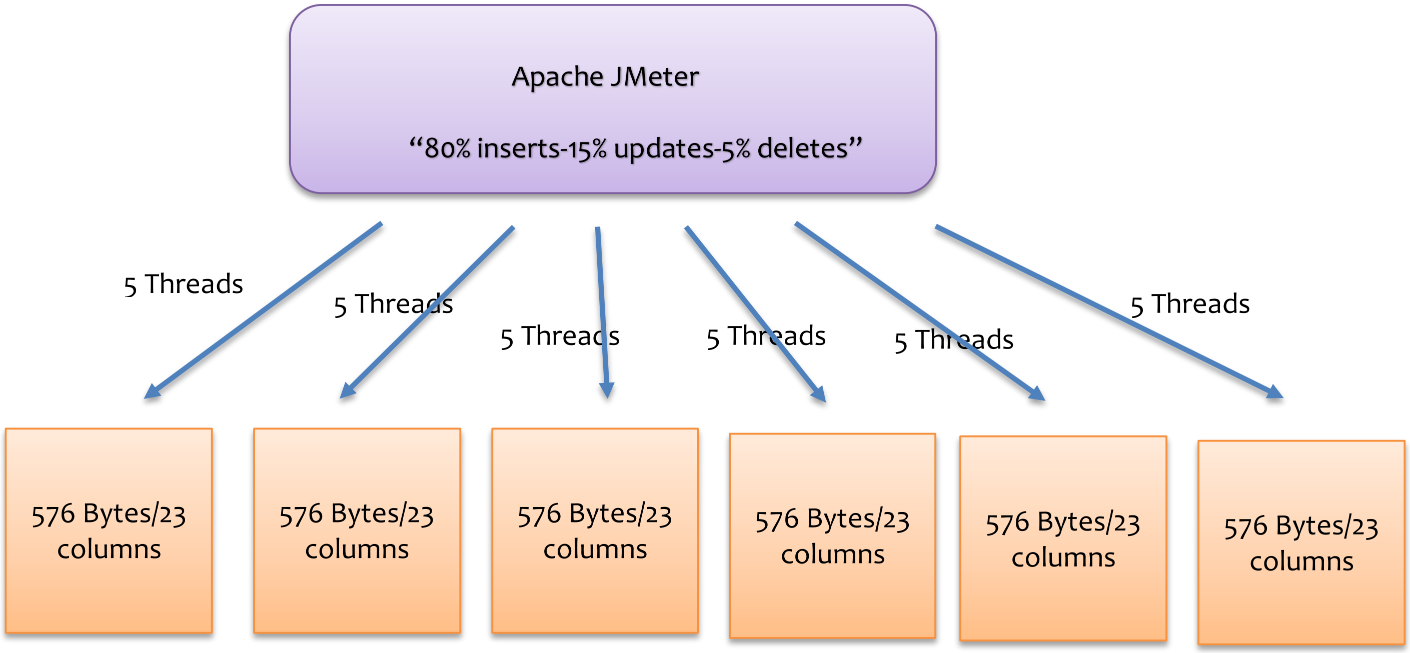


Figure 43: Log volumes performance testing (Continuous heavy write using JMeter)

As seen below from FlashArray//m GUI the above setup generated a load from 275 MB/sec to 600 MB/sec where it was writing somewhere from 500,000 to 1 million records per second on an average.

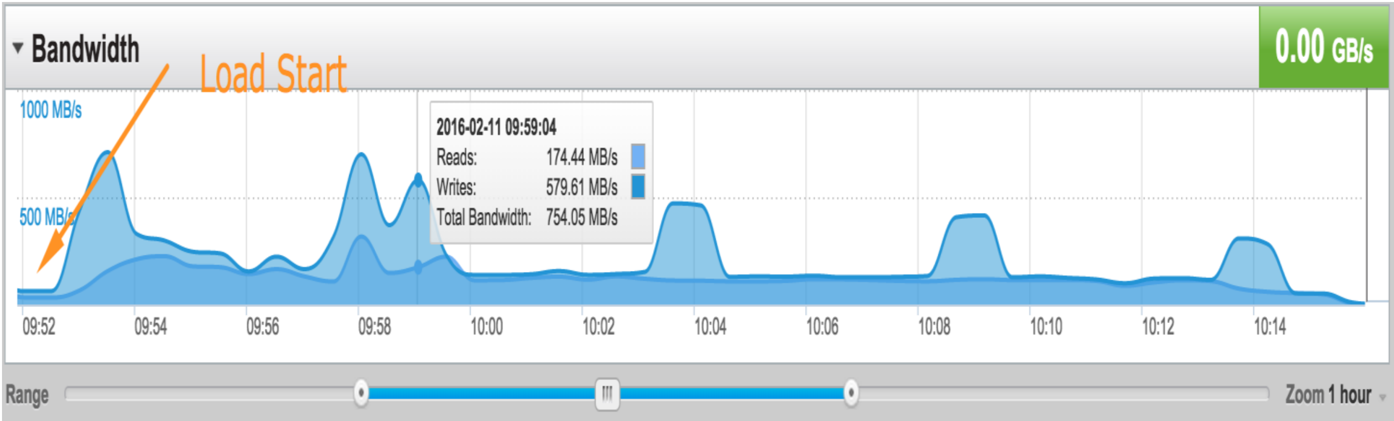


Figure 44: Mixed Write load generated by Apache JMeter

Below is the snapshot of latency for the log volume during the same write load. During this load latency was always sub-millisecond throughout this huge aggressive write load 25 minutes run.

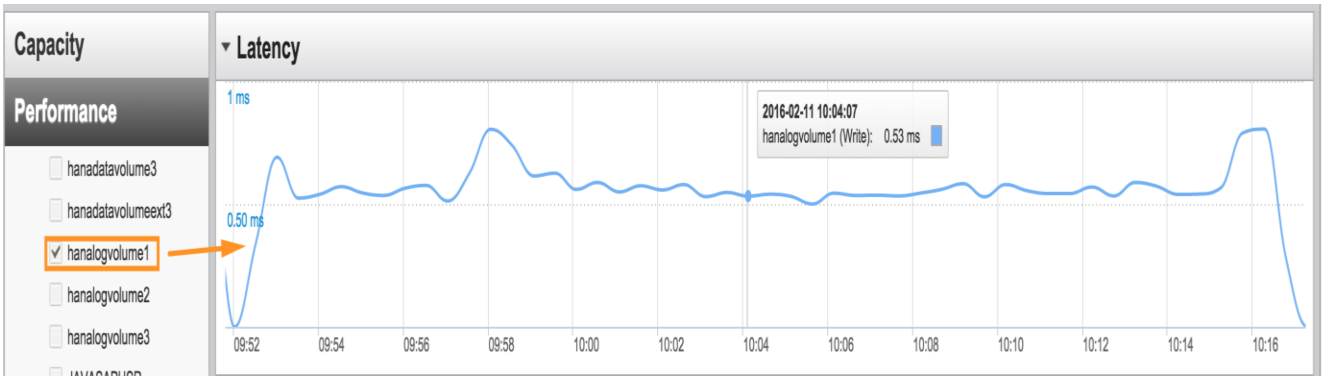


Figure 45: Redo log latency

Summary

SAP HANA® analytics solution on FlashStack can do real-time reporting even under heavy write load. As seen the log volumes write latency under heavy write load is always sub-millisecond.

FlashArray//m sizing for SAP HANA®

FlashArray//m sizing for SAP HANA® needs to be done in three stepped process:

- Throughput sizing
- Capacity sizing
- Apply data reduction of FlashArray//m

Both throughput sizing and capacity sizing should be considered for SAP HANA® sizing. Start the sizing calculation with number of SAP HANA® nodes and then perform the capacity sizing.

Throughput sizing

Throughput sizing depends on number of SAP HANA® nodes, below is the sizing recommendation based on number of SAP HANA® nodes

FlashArray//m Controller Model	Scale out SAP HANA® nodes
//m20	6
//m50	9
//m70	12

Table 10: Throughput sizing recommendation

Capacity sizing

Once the throughput sizing is performed the capacity sizing of FlashArray//m storage sizing calculation is based on the overall amount of memory needed by the SAP HANA® database which in turn is the result of conducting a memory sizing for a given system.

Two types of data for which memory is required are differentiated:

Static data: Memory is required to hold the database tables. It is calculated by applying a compression factor to the size of the source database and also data reduction done on FlashArray//m. The value of the compression factor depends on the application scenario, for example OLTP (like Suite-on-HANA®) or OLAP (like BW- on-HANA®), and on the source database type (some source databases already apply a certain compression).

Dynamic data: Memory also needs to hold objects created dynamically at runtime, e.g. when loading new data or when executing queries.

It is highly recommended to reserve the same amount of memory for dynamic data as for static data.

FlashArray//m is SAP HANA® TDI certified and it can be deployed on hardware infrastructures set up following the SAP HANA® Tailored Datacenter Integration (TDI). The sizing formulas in the following sections are valid for SAP HANA® TDI environments.

Data volumes sizing

A Savepoint or a Snapshot is created or a delta merge is performed, data is persisted from memory to the data volume under `/hana/data/<sid>`. For the hosts altogether, the recommended size is calculated as follows:

Size$_{data}$ = 1 x RAM (= 2 x overall table size)

RAM = Size of memory of SAP HANA® system.

For migration purposes of a non-HANA® database to SAP HANA®, the system may temporarily need more disk space for data than calculated in the sizing phase.

Redo Log Volumes sizing

The minimum size of the log volume depends on the number of data changes occurring between two SAP HANA® Savepoints which per default are created every 5 minutes. The more data changes are executed by write transactions in that period of time, the more redo log segments are written to the log volume under `/hana/log/<sid>`. When sizing the log volume, the following points have to be considered:

The redo log must not be overwritten before a Savepoint entry is available in the data volume; otherwise, the SAP HANA® database may become unable to restart.

If "log_mode = normal" is set the redo log must not be overwritten before a backup took place. Therefore, it is recommended to have some extra space available for situations where incidents or faults may interrupt the backup process. That extra space should allow for system administrators to fix and finish the backup process before the log volume runs full.

Below is the formula that calculates the log volume size depending on the RAM

[systems <= 512GB] Size redo log = 0.5 x RAM (= 1 x overall table size)

[systems > 512GB] Size redo log(min) = 512GB

For the high availability configuration of storage replication or SAP HANA® system replication the same applies as for the data volumes – the storage size must be the same on the secondary site.

Shared Volume sizing

All binary, trace and configuration files are stored on a shared file system that is exposed to all hosts of a system under `/hana/shared/<sid>`. If a host needs to write a memory dump (which can reach up to 90% of the RAM size), it will be stored in this file system. Thus, additional space is required for the traces written by the compute node(s) of the SAP HANA® database.

SAP HANA® Shared volume = 1 x RAM

Backups

A complete data backup contains the entire payload of all data volumes. The size required by the backup directory not only depends on the total size of the data volumes, but also on the number of backup generations kept on disk and on the frequency with which data is changed in the SAP HANA® database. For example, if the backup policy requires to perform complete data backups on a daily basis and to keep those backups for one week, the size of the backup storage must be seven times the size of the data area.

In addition to data backups, backup storage for log backups must be reserved to provide the possibility for a point-in-time database recovery. The number and size of log backups to be written depend on the number of change operations in the SAP HANA® database.

Size of Backups >= 0.5 x RAM (= 1 x overall table size)

Total capacity size

Calculate the total capacity size of SAP HANA® based on above information

SAP HANA® Total Capacity size =

Data volumes size + Redo log volumes size + Shared volume size + Backup size

Data reduction of FlashArray//m

Once the throughput calculation and capacity calculation is completed refer to data reduction topic in this document under Value propositions section to determine the data reduction and apply it on capacity size to get the capacity size.

Final SAP HANA® Total Capacity size = Data Reduction * SAP HANA® Total Capacity size

Conclusions

In this paper, we have demonstrated that FlashStack provides the ultimate infrastructure platform for analytics solutions on SAP HANA®.There are significant benefits in deploying SAP on a FlashStack converged infrastructure solution. The benefits being:

- Simplicity: FlashStack eliminates the need for space management of a SAP system thus reducing the operational overhead on the SAP administrator. Deployment times are reduced via pre-tested, validated designs.

- Built-in Data Protection: With built in snapshot technology, FlashStack allows SAP admins can backup SAP HANA® data in a matter of seconds even when the database is a few hundred TBs large.

- Superior read performance: FlashStack ensures superior read bandwidth which helps with rapid/accelerated restarts and fail-overs which reduces the overall downtime even when there a lot of row store or pre-loaded column tables.

  This also helps with faster first time loading of tables which does not impact by a great extent the overall query performance even when data is not present in memory.

- Superior write performance: FlashStack ensures superior write bandwidth and consistent sub-millisecond response time helping SAP HANA® analytical solutions with faster delta merges operations which makes sure that business critical queries are not slowed down due to delta merge lag.

  This also helps with faster and continuous heavy write when there is a real-time replication to SAP HANA® for reporting or BW nightly data loads.

- Lower Cost: FlashStack can dramatically reduce the SAP HANA® solutions footprint via best in class data reduction on top of SAP HANA® compression. FlashStack is also an ideal platform for consolidating multiple system copies efficiently and economically as data reduction increases even further. Furthermore, a smaller footprint means less power, less datacenter space, and lower cooling requirements, which makes the TCO for large scale SAP deployments even more attractive.

References

SAP HANA® best practices on pure storage guide-

http://blog.purestorage.com/in-memory-database-meets-mighty-performance-storage-flasharraym-is-certified-for-sap-hana/

Pure Storage FlashArray Data Sheet - http://www.purestorage.com/content/dam/purestorage/pdf/datasheets/PureStorage_FlashArraym-Brochure.pdf

Nexus 5000 series Configuration Guide: http://www.cisco.com/c/en/us/td/docs/switches/datacenter/nexus5000/sw/configuration/guide/cli/CLIConfigurationGuide.pdf

SAP Systems Copy Guide: https://websmp201.sap-ag.de/~sapidb/011000358700001324472012E

Pure Storage, Inc.
Twitter: @purestorage
www.purestorage.com

650 Castro Street, Suite #260
Mountain View, CA 94041

T: 650-290-6088
F: 650-625-9667

Sales: sales@purestorage.com
Support: support@purestorage.com
Media: pr@purestorage.com
General: info@purestorage.com