SAP HANA Data Management and Performance on IBM Power Systems

Dino Quintero
Damon Bull
Vinicius Cosmo Cardoso
Cleiton Freire
Eric Kass
Note: Before using this information and the product it supports, read the information in “Notices” on page v.

First Edition (April 2020)

This edition applies to:
SAP HANA V2.0 SPS4 R44
SUSE Linux Enterprise Server V15 SP1
Red Hat Enterprise Linux V8.2
Hardware Management Console V9.1.940.0

© Copyright International Business Machines Corporation 2020. All rights reserved.
Note to U.S. Government Users Restricted Rights -- Use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM Corp.
Contents

Notices...........................................................................v
Trademarks ........................................................................ vi

Preface................................................................................ vii
Authors.............................................................................. vii
Now you can become a published author, too!.......................... viii
Comments welcome................................................................ viii
Stay connected to IBM Redbooks........................................... ix

Chapter 1. Introduction....................................................... 1
1.1 How to approach SAP HANA on IBM Power Systems........... 2
  1.1.1 Memory footprint.................................................... 3
  1.1.2 Startup times.......................................................... 4
  1.1.3 Backup................................................................. 4
  1.1.4 High availability.................................................... 5
  1.1.5 Exchanging hardware.............................................. 5
  1.1.6 Remote database connectivity................................... 5
  1.1.7 Conclusion............................................................. 5

Chapter 2. SAP HANA data growth management..................... 7
2.1 The challenge of data growth management and the data temperature concept..... 8
2.2 SAP HANA data tiering options for SAP HANA databases........... 9
  2.2.1 Near-Line Storage with SAP IQ.................................. 9
  2.2.2 SAP HANA Extension Node...................................... 11
  2.2.3 SAP HANA Native Storage Extension......................... 12
  2.2.4 SAP HANA Dynamic Tiering.................................. 14
  2.2.5 SAP HANA on IBM Power Systems and SAP HANA Dynamic Tiering solutions.... 18
  2.2.6 SAP HANA Dynamic Tiering: Hands-on demonstration............ 19

Chapter 3. SAP HANA I/O performance enhancements............... 35
3.1 Virtual persistent memory............................................ 36
  3.1.1 SAP HANA usage of vPMEM volumes.......................... 36
  3.1.2 Sizing the vPMEM volume........................................ 38
  3.1.3 Checking the LPAR hardware assignment....................... 38
  3.1.4 Creating a vPMEM device......................................... 41
  3.1.5 vPMEM affinity configuration options.......................... 43
  3.1.6 Affinity disabled.................................................... 44
  3.1.7 Enabling vPMEM use in Linux.................................... 44
  3.1.8 Configuring SAP HANA to use vPMEM volumes............. 46
  3.1.9 Specifying tables, columns, or partitions to use vPMEM volumes.......... 46
  3.1.10 Undersized vPMEM volumes.................................... 48
  3.1.11 Disabling the usage of vPMEM volumes....................... 48
  3.1.12 Benefits of using vPMEM volumes on Power Systems servers........ 48
  3.1.13 vPMEM usage considerations.................................... 48
  3.1.14 PowerVM vPMEM enablement dependencies.................. 49
  3.1.15 vPMEM-related SAP Notes..................................... 49
  3.1.16 Additional documentation that is related to vPMEM and persistent memory usage with SAP HANA.................................. 50
  3.2 RAM disk (temporary file systems): SAP HANA Fast Restart Option........ 50
3.2.1 Configuring tmpfs volumes ....................................................... 51
3.2.2 Configuring SAP HANA to use the tmpfs volumes ...................... 52
3.2.3 Comparing vPMEM and tmpfs .................................................. 52
3.2.4 SAP-related documentation .................................................... 53
3.3 Persistent disk storage by using native Non-Volatile Memory Express devices ........................................ 54
  3.3.1 NVMe device details ............................................................. 54
  3.3.2 Configuration and usage considerations ................................... 54
  3.3.3 NVMe performance characteristics ......................................... 56
  3.3.4 Striping effects for internal NVMe cards ................................. 56
  3.3.5 Summary ............................................................................. 56
3.4 NVMe Rapid Cold Start Mirror ....................................................... 57
  3.4.1 Hybrid mirrored volume ......................................................... 57
  3.4.2 Summary ............................................................................. 58
3.5 Comparing vPMEM to Intel Optane ............................................... 58
3.6 Impact to Live Partition Mobility capabilities .................................. 59

Related publications ........................................................................ 61
IBM Redbooks .................................................................................. 61
Online resources ............................................................................... 61
Help from IBM .................................................................................. 62
Notices

This information was developed for products and services offered in the US. This material might be available from IBM in other languages. However, you may be required to own a copy of the product or product version in that language in order to access it.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user’s responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not grant you any license to these patents. You can send license inquiries, in writing, to:
IBM Director of Licensing, IBM Corporation, North Castle Drive, MD-NC119, Armonk, NY 10504-1785, US

INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION “AS IS” WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some jurisdictions do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM websites are provided for convenience only and do not in any manner serve as an endorsement of those websites. The materials at those websites are not part of the materials for this IBM product and use of those websites is at your own risk.

IBM may use or distribute any of the information you provide in any way it believes appropriate without incurring any obligation to you.

The performance data and client examples cited are presented for illustrative purposes only. Actual performance results may vary depending on specific configurations and operating conditions.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

Statements regarding IBM’s future direction or intent are subject to change or withdrawal without notice, and represent goals and objectives only.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to actual people or business enterprises is entirely coincidental.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs. The sample programs are provided “AS IS”, without warranty of any kind. IBM shall not be liable for any damages arising out of your use of the sample programs.
Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corporation, registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the web at “Copyright and trademark information” at http://www.ibm.com/legal/copytrade.shtml

The following terms are trademarks or registered trademarks of International Business Machines Corporation, and might also be trademarks or registered trademarks in other countries.

AIX®
DB2®
Db2®
FlashCopy®

IBM®
IBM Cloud™
POWER9™
PowerVM®

Redbooks®
Redbooks (logo) ®
System z®
XIV®

The following terms are trademarks of other companies:

Intel, Intel logo, Intel Inside logo, and Intel Centrino logo are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

The registered trademark Linux® is used pursuant to a sublicense from the Linux Foundation, the exclusive licensee of Linus Torvalds, owner of the mark on a worldwide basis.

Windows, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

Red Hat, are trademarks or registered trademarks of Red Hat, Inc. or its subsidiaries in the United States and other countries.

Other company, product, or service names may be trademarks or service marks of others.
Preface

This IBM® Redpaper publication provides information and concepts about how to take advantage of SAP HANA and IBM Power Systems features to manage data and performance efficiently.

The target audience of this book includes architects, IT specialists, and systems administrators who deploy SAP HANA and manage data and SAP system performance.

Authors

This paper was produced in close collaboration with the IBM SAP International Competence Center (ISICC) in Walldorf, SAP Headquarters in Germany and IBM Redbooks®.

Dino Quintero is an IT Management Consultant and IBM Level 3 Senior Certified IT Specialist with IBM Redbooks in Poughkeepsie, New York. He has 24 years of experience with Power Systems technologies and solutions. Dino shares his technical computing passion and expertise by leading teams developing technical content in the areas of enterprise continuous availability, enterprise systems management, high-performance computing, cloud computing, artificial intelligence (including machine and deep learning), and cognitive solutions. He is a Certified Open Group Distinguished IT Specialist. Dino holds a Master of Computing Information Systems degree and a Bachelor of Science degree in Computer Science from Marist College.

Damon Bull is a Senior SAP Software Performance Engineer who has worked with different SAP versions from 2.2F onward. His focus is on SAP performance, and he has conducted and published over 50 SAP benchmarks over the last 25 years. Damon has experience with IBM DB2®, Oracle, SQL Server, and SAP HANA databases (DBs), and SAP Sales and Distribution (SAP SD), SAP Business Warehouse (SAP BW), SAP Transaction Banking (SAP TRBK), SAP Financial Accounting (SAP FI), Assemble-to-Order (ATO), and others running on IBM AIX®, Linux, and Windows servers. Damon holds a Bachelor of Science in Computer and Information Sciences from the University of Oregon.

Vinicius Cosmo Cardoso is an IBM Certified Expert IT Specialist and Senior SAP Basis Administrator working for IBM Brazil. He has 12 years of experience in SAP, working on complex projects like new implementations, upgrades, and platform migrations. He is an SAP certified professional for SAP NetWeaver, SAP OS/DB Migration, and SAP HANA. Vinicius holds a Master of IT Solutions Architecture from 2018 and a Bachelor of Computer Science from 2008. Throughout these 12 years, he has worked for several Brazilian and global clients in several industry areas.
Cleiton Freire is an IBM Cloud™ Solution Leader for the Europe, Latin America, and Canada cluster. Cleiton has almost 20 years in the Information Technology industry and throughout the past 14 years working in IBM. He has served as a subject matter expert on SAP on IBM Power platform migrations, installations, and administration for large customers in Brazil, the United States, and Europe. For the past 2 years, Cleiton has helped customers find the best tech solution to address their business problems. Cleiton is a certified Technology Consultant for the SAP HANA platform and holds a Bachelor of Science degree from the Faculty of Technology in Sao Paulo, Brazil.

Eric Kass is an IBM Developer Architect in Worldwide SAP Technical Enablement, which is part of an IBM Systems group that is based in Boeblingen, Germany. Eric started with IBM in 1997 in operating system (OS) network and communication development, where he wrote device control and network routing code. In 2000, he spent a year with SAP to write the DB layer so that SAP can communicate between Windows and IBM i. He became a successful IBM Developer and architect working at the SAP development lab, where he designs and codes DB drivers; SAP kernel components; and SAP and DB high availability (HA) engines, network drivers, and security protocols. Eric has over 20 patents in SAP related fields, and he supports worldwide SAP customers on IBM i and AS400 platforms.

Thanks to the following people for their contributions to this project:

Wade Wallace  
IBM Redbooks, Austin Center

Katharina Probst, Walter Orb, Tanja Scheller  
IBM Germany

Now you can become a published author, too!

Here’s an opportunity to spotlight your skills, grow your career, and become a published author—all at the same time! Join an IBM Redbooks residency project and help write a book in your area of expertise, while honing your experience using leading-edge technologies. Your efforts will help to increase product acceptance and customer satisfaction, as you expand your network of technical contacts and relationships. Residencies run from two to six weeks in length, and you can participate either in person or as a remote resident working from your home base.

Find out more about the residency program, browse the residency index, and apply online at:  
ibm.com/redbooks/residencies.html

Comments welcome

Your comments are important to us!

We want our papers to be as helpful as possible. Send us your comments about this paper or other IBM Redbooks publications in one of the following ways:

- Use the online Contact us review Redbooks form found at:  
  ibm.com/redbooks

- Send your comments in an email to:  
  redbooks@us.ibm.com
Stay connected to IBM Redbooks

➤ Find us on Facebook:
   http://www.facebook.com/IBMRedbooks
➤ Follow us on Twitter:
   http://twitter.com/ibmredbooks
➤ Look for us on LinkedIn:
   http://www.linkedin.com/groups?home=&gid=2130806
➤ Explore new Redbooks publications, residencies, and workshops with the IBM Redbooks weekly newsletter:
➤ Stay current on recent Redbooks publications with RSS Feeds:
   http://www.redbooks.ibm.com/rss.html
Chapter 1. Introduction

This chapter introduces the features of SAP HANA on IBM Power Systems that help manage data and performance efficiently.

This chapter describes the following topics:

- How to approach SAP HANA on IBM Power Systems
- Memory footprint
- Startup times
- Backup
- High availability
- Exchanging hardware
- Remote database connectivity
- Conclusion
1.1 How to approach SAP HANA on IBM Power Systems

In October 2019, the contributing authors to this publication gathered in a small office at the IBM SAP International Competence Center (ISICC) to define the content that is appropriate for the audience of this publication, which is system architects seeking guidance and best practices for designing and implementing SAP HANA on IBM Power Systems.

The authors of this publication are architects and engineers who have been implementing SAP HANA systems for years, and whom are often asked to provide their insights about designing systems. The inquiries are diverse, the conditions vary, and the answers are individual. However, in designing this book, the team anticipates the questions that are most likely asked.

The authors intend that this book be used in a task-based fashion to find answers to questions like the following ones:

- Which hardware do you choose?
- How do you plan for backups?
- How do you minimize startup time?
- How do you migrate data to SAP HANA?
- How do you connect to remote systems?
- How much hardware do you need (memory, CPU, network adapters, and storage)?
- How do you reduce the memory footprint?
- What high availability (HA) options exist?
- Where do you get support?
- How do you answer other questions?

Consider this book as a starting guide to learn what questions to ask and to answer the last question first. For any remaining questions, contact the ISICC.

The authors understand the goal of every SAP HANA installation is to be as resilient, available, inexpensive, fast, simple, extensible, and as manageable as possible within constraints, even if some of the goals are contradictory. This publication is unique in its intention to exist as a living document. SAP HANA technology and Power Systems technology supporting SAP HANA change so quickly that any static publication is out-of-date months after distribution. The development teams intend to keep this book up-to-date.

Where does one begin to define the requirements of an SAP HANA system? Before you do sizing, it is necessary to establish your business goals that come from the changing IT industry. The typical motivation for moving forward to an SAP HANA-based system is the path towards digital transformation, which requires systems to perform real-time digitalization processing. The requirements for processing pervasive user applications that use SAP Fiori and real-time analytics, for example, has a different processing footprint (Open Data Protocol (OData)-based) than a classical form-based process before output and the process after input into SAP GUI applications.

IBM bases its sizing on tables that are established by the benchmark team, who are highly proficient in SAP performance analysis. Conceptually, calculating SAP HANA sizing from a classical AnyDB system begins by determining the SAP Application Performance Standard (SAPS) (a standard measure of workload capacity) of your present system, then using that value as reference for the IBM sizing tables to determine the hardware requirements for an Power Systems server to run SAP HANA with an equivalent SAPS performance.
After you size the memory, you find that the number of hardware permutations that fits your requirements is overwhelming. To narrow down your choices, you decide whether to use a scale-up or scale-out infrastructure. As a best practice, use a scale-up infrastructure because even though there is some automation that is available to help select which data can be spanned across multiple nodes, you still need some manual management when you implement a scale-out structure.

When you use a scale-up infrastructure, consider that different hardware implementations have different performance degradation characteristics when scaling-up memory usage. As systems become larger, the memory to CPU architecture plays an important role in the following situations:
1. How distant the memory is from the CPU (affinity).
2. How proficiently the system can manage cache to memory consistency.
3. How well the CPU performs virtual address translation of large working sets of memory.

Power Systems are designed to scale up by using enterprise-grade, high-bandwidth memory interconnects to the CPU, and relatively flat non-uniform memory access (NUMA). Scaling up with Power Systems servers is a best practice to support large memory footprint SAP HANA installations (For more information, see 2.2.5, “SAP HANA on IBM Power Systems and SAP HANA Data Tiering solutions” on page 18). The SAP HANA on Power Systems scaling benefits are complemented by the advantage of consolidating multiple SAP HANA, Linux, IBM AIX, and IBM i workloads on the same machine.

Note: Acquisition costs and licensing become complicated when running mixed workloads. Have your customer representative contact the ISICC to have them help provide an optimal cost configuration. Contacting the ISICC is a good idea in general because we want to know what you have in plan, and it is our job to help make the offering as attractive as possible to fit your needs.

Customers that have SAP incidence support for SAP on IBM products continue to have support with SAP HANA. SAP HANA on Power Systems support channels are intricately integrated into SAP development and support. The people supporting SAP on AIX, SAP with IBM Db2®, or SAP on IBM i and System z® are members of the same team that support SAP HANA on Power Systems. By migrating to SAP HANA but remaining with IBM, you have not changed support teams. For questions regarding anything SAP, open an SAP support ticket with an IBM coordinated support channel, such as BC-OP-AIX, BC-OP-AS4, BC-OP-LNX-IBM, BC-OP-S390; and for issues regarding interaction with other databases (DBs), BC-DB-DB2, BC-DB-DB4, and BC-DB-DB6.

1.1.1  Memory footprint

In contrast to the disk-storage-based DB engines that are used by classical SAP Enterprise Resource Planning (ERP), a memory footprint with memory-based DB systems is a continuous issue. A memory footprint is not a static calculation because the calculation takes time. If your archiving rate cannot keep up with your data generation rate, your memory footprint tomorrow will be greater than today.

Data growth is an issue that continuously requires attention because it is likely that your application suite will change. Therefore, the characteristics of your data generation will likely change too.
An archiving strategy is your primary application-level control for data growth, but some applications like SAP Business Warehouse (BW) support near-line-storage as an alternative. For data that should not be archived for various reasons, SAP HANA supports a selection of hardware technologies for offloading data that is not used frequently (warm data or cold data) to places other than expensive RAM with expensive SAP licensing.

The available technologies to alleviate large data and data growth are SAP HANA Native Storage Extension (NSE), which is a method of persisting explicit tables on disk and loading data into caches as necessary (It is similar in function to classical DB engines), extension nodes (slower scale-out nodes), and SAP HANA Dynamic Tiering (SAP HANA to SAP HANA near-line storage).

Note: Some of the options that are offered by SAP HANA are not available for all applications. SAP S/4HANA has other restrictions than SAP BW, and both SAP S/4HANA and SAP BW have more restrictions than SAP HANA native applications.

SAP provides reports to run on your original system to provide data placement assistance, and the results are typically good suggestions. You must be prepared to distribute your data differently if prototypes demonstrate other configurations are necessary. Chapter 2, “SAP HANA data growth management” on page 7 describes options for planning SAP archiving and managing various technologies for data that is accessed at different frequencies.

1.1.2 Startup times

The SAP HANA processing model implies that data is in memory for immediate access. The cost of accessing data from storage is met when the SAP HANA system starts, during which all data is loaded into RAM. For example, if your storage is connected by 4-multipath 16 Gbps lines, loading 7 TB requires an hour. Both IBM and Intel provide solutions to minimize startup times by using persistent memory in which data survives SAP HANA restarts in specific situations.

Even though the hardware and operating system (OS) methods of retaining persistent data in memory varies, in SAP HANA persistent memory is referenced as memory mapped files. Either the OS or hardware provides a memory-based file system (like a RAM disk) that is “seen” as a file system that is mounted in a path. Instructed by a configuration parameter, SAP HANA uses as much persistent storage as possible and reverts to regular memory for the remainder. SAP HANA can be instructed to attempt to store everything in persistent memory, or each file can be altered to prefer persistent or non-persistent RAM.

Even with hardware support for a quick start, consider that a high availability (HA) solution alleviates the need to wait for SAP HANA to restart because SAP HANA is always available. With a HA solution in place, the preferred method for scheduled maintenance is to move the data to the backup site.

1.1.3 Backup

SAP HANA provides an interface for backup tools that is called Backint. SAP publishes a list of certified third-party tools that conform to that standard. If you plan on using methods like storage-based snapshot or flash copies, quiesce the SAP HANA system before taking a storage-based snapshot. A quiesce state is necessary to apply subsequent logs (the record of transactions that occur since the time of the quiesce) to an image that is restored from the IBM FlashCopy® copy.
1.1.4 High availability

SAP HANA System Replication (HSR) is the underlying foundation of DB HA. The method is based on logical replication, where individual row-level changes are recorded and transferred to the backup site (referred to as log shipping). Logical replication is the opposite of storage-based replication, where pages on disk are duplicated without considering the logical data structure of the change. SAP HSR accomplishes the task of transferring changes between hosts by transferring segments of the log (segments of the record of the changes to the DB that are used for transactional recovery). Changes that are received by backup hosts are sent to the local DB in a fashion similar to when undergoing forward recovery.

1.1.5 Exchanging hardware

Unfortunately, exchanging hardware is a problem that you encounter every few years. If your system manages to fulfill all the prerequisites for Live Partition Mobility (LPM), which is a feature of Power Systems hardware that moves running partitions between hosts, then exchanging hardware by using LPM is the least disruptive to a running SAP system, but the prerequisites are difficult to master. Many specialized hardware features, which are not virtualized and must be assigned to a physical partition, make LPM impossible.

The SAP standard method of exchanging hardware, for example, any hardware with the same endianness, is by using SAP HSR. For more information, see SAP Note 1984882.

1.1.6 Remote database connectivity

SAP provides a wealth of data connectivity options. Some methods like Remote Function Call (RFC) and Database Multiconnect are provided at the layer of the SAP system. Other methods like Smart Data Access (SDA) and Smart Data Integration (SDI) are integrated directly into SAP HANA. Extract-transform-load (ETL) methods include SAP Landscape Transformation (SLT) and the SAP Data Migration Option (DMO) for upgrades.

SDA and SDI are not specific to SAP HANA to IBM i for the connectivity function. The IBM i case is a prime example of using a generic Open Database Connectivity (ODBC) adapter for SDA, and the generic Camel JDBC adapter for SDI.

1.1.7 Conclusion

Facilitating designs of superior SAP HANA systems is the goal of this publication. Your productive, development, and test systems have different reliability, availability, serviceability (RAS); connectivity; and security requirements, but the aspects to consider are universal. Where tradeoffs must be made between cost, RAS, and complexity are decisions that are unique to your situation, but the intent of this book and the service and support you have from the ISICC and the IBM SAP development team is to help optimize your decisions so that you feel comfortable and confident with the final architecture of your design.
This chapter describes the challenges of data growth in systems and how it can impact the operations of an SAP HANA database (DB).

This chapter introduces the data temperature concept, which is used as a guide to decouple data types based on their criticality. Data temperature is helpful for companies deciding when to move their data to different but still accessible data tiers.

Different SAP data tiering solutions that are supported on IBM Power Systems servers are described in this chapter. There is an overview that can help you decide what solution is most suitable for your challenge.

The purpose of this chapter is to help you decide on the most suitable solution among the different available solutions.

This chapter describes the following topics:

- The challenge of data growth management and the data temperature concept
- SAP HANA data tiering options for SAP HANA databases
2.1 The challenge of data growth management and the data temperature concept

Companies collect more information about their business to control their day-to-day operations in real time. The more information that is collected, the more IT resources are consumed over time, which increases the costs for organizations because of the need for data scaling.

In an SAP HANA DB, main memory and disk areas are consumed, which increases the total cost of ownership (TCO) and impacts performance over time.

Before scaling up or scaling out the SAP HANA DB, think about options for decoupling your data location by defining what data always must be in memory and available with the highest performance for applications and users, and what data is less frequently accessed so that it available from a lower performance data tier with no impact to the business operations.

You can define which data is accessed infrequently so that it can be available to the users in a reasonable and cheaper performing storage tier. This concept is called the data temperature.

Figure 2-1 shows how data temperature is classified.

<table>
<thead>
<tr>
<th>HOT</th>
<th>Data read and written very frequently for mission-critical business tasks and available for real-time processing and real-time analytics.</th>
</tr>
</thead>
<tbody>
<tr>
<td>WARM</td>
<td>Data less frequently accessed with reduced performance so that it can be stored in a lower cost tier with reduced SLA.</td>
</tr>
<tr>
<td>COLD</td>
<td>Very infrequent access, historical data, typically read-only. Data stored on low cost storage tiers, like disk only, but still accessible at anytime.</td>
</tr>
</tbody>
</table>

Figure 2-1  Data temperature concept

There are benefits for using data tiering options for your SAP HANA DB:

- Reduce data volume and growth in the hot store and SAP HANA memory.
- Avoid performance issues on SAP HANA DBs because too much data must be loaded into the main memory.
- Avoid needing to scale up or scale out over time.
- Ensures lower TCO.

SAP offers four data tiering solutions that are supported on SAP HANA on IBM Power Systems:

- Near-Line Storage (NLS) (cold data)
- SAP HANA Extension Node (warm data)
- SAP HANA Native Storage Extension (NSE) (warm data)
- SAP HANA Dynamic Tiering (warm data)
2.2 SAP HANA data tiering options for SAP HANA databases

This section describes the different data tiering options for SAP HANA DBs.

2.2.1 Near-Line Storage with SAP IQ

With NLS, you can move your SAP Business Warehouse (BW) data that is classified as cold out of the primary SAP HANA DB and store it in SAP IQ.

*SAP IQ* is a column-based with petabyte scale relational DB software system that is used for business intelligence, data warehousing, and data marts. Produced by Sybase Inc., which is now an SAP company, its primary function is to analyze large amounts of data in a low-cost, HA environment.

SAP developed NLS to sue with SAP NetWeaver BW and SAP BW/4HANA *only*.

**Prerequisites for Near-Line Storage**
The required SAP NetWeaver BW versions and support package levels are:

- SAP NetWeaver BW 7.30 SPS >= 09
- SAP NetWeaver BW 7.31 SPS >= 07
- SAP NetWeaver BW 7.4 SPS >= 00
- SAP NetWeaver BW 7.5 SPS >= 00

The required SAP BW/4HANA versions and support package level is SAP BW/4HANA 1.0 SPS >= 00.

For more information about the minimum release level for SAP BW, see [SAP Note 1796393](https://support.sap.com/).

The SAP components for NLS implementation are:

- SAP NetWeaver BW on SAP HANA DB or SAP BW/4HANA
- SAP IQ DB
- SAP IQ Libraries on SAP Application Server side for connection to SAP IQ
- SAP Smart Data Access (SDA) on SAP HANA Server side for connection to SAP IQ (optional)

Implementing SDA to access the NLS data is optional. SAP HANA SDA optimizes the running of queries by moving processing as much as possible to the DB that is connected through SDA. The SQL queries the work in SAP HANA on virtual tables. The SAP HANA Query Processor optimizes the queries and runs the relevant parts in the connected DB, and then returns the result to SAP HANA and completes the operation.

Using an NLS solution with SAP HANA SDA is supported as of SAP NetWeaver BW 7.40 SP8 or higher or SAP BW/4 HANA 1.0 or higher.

**Note:** To use the SDA solution, the SAP BW application team must configure the SAP BW objects.

For more information about using SAP HANA SDA for NLS and its performance benefits, see the following SAP Notes:

- SAP Note 2165650
- SAP Note 2156717
- SAP Note 2100962
NLS implementation architecture
The architecture of NLS implementation with SAP NetWeaver BW on SAP HANA and SAP BW/4HANA without SDA is shown in Figure 2-2.

![Figure 2-2  NLS architecture without Smart Data Access](image)

The architecture of NLS implementation with SAP NetWeaver BW on SAP HANA and SAP BW/4HANA with SDA is shown in Figure 2-3.

![Figure 2-3  NLS architecture with Smart Data Access](image)

**Hint:** SAP IQ is supported on Power Systems (Big Endian only), so it is possible to reserve a logical partition (LPAR) for implementing it and take advantage of a virtual LAN (VLAN) to produce the lowest network latency possible between SAP IQ and SAP HANA.
For the SDA implementation, SAP developed and provided the packages that are supported on Power Systems servers.

For more information about the implementation of NLS for SAP BW, see SAP Note 2780668.

### 2.2.2 SAP HANA Extension Node

The SAP HANA Extension Node is based on the SAP HANA scale-out feature, and it stores the warm data of SAP BW in a separate node from the SAP HANA main nodes.

#### Sizing for the SAP HANA Extension Node

Besides using the memory area for storing static objects such as tables (for example, the data footprint), SAP HANA also uses memory for dynamic objects, for example, objects for the SAP HANA run time. The SAP recommendation for SAP HANA memory sizing is to reserve as much memory for dynamic objects as for static objects in the column store.

**Note:** As a best practice, the SAP recommendation for memory sizing is:

\[
\text{RAM}_{\text{dynamic}} = \text{RAM}_{\text{static}}.
\]

For example, if you have a footprint of 500 GB, your SAP HANA DB memory size must be at least 1 TB.

The SAP HANA Extension Node can operate twice as much data with the same amount of memory and fewer cores. For example, if you expect to have up to 1 TB of footprint in your Extension Node, you can have up to 250 GB of memory for the Extension Node.

#### Prerequisites for the SAP HANA Extension Node

For SAP BW on HANA, you need SAP HANA 1.0 SPS12 or later and SAP BW 7.4 SP12 or later. However, SAP BW 7.50 SP05 or SAP BW/4HANA is recommended. For SAP HANA native applications, you need SAP HANA 2.0 SPS3 or later.

The ideal use case for Extension Node is SAP BW on SAP HANA or SAP BW/4HANA because the SAP BW application controls the data distribution, partitioning, and access paths. For SAP HANA native applications, all data categorization and distribution must be handled manually.
SAP HANA Extension Node Deployment

A possible SAP HANA scale-out with Extension Node implementation is shown in Figure 2-4.

SAP HANA Extension Node is built into the SAP HANA platform and supported on Power Systems.

For more information about the implementation of Extension Node, see the following SAP Notes:
- SAP Note 2486706
- SAP Note 2643763

2.2.3 SAP HANA Native Storage Extension

Starting with SAP HANA 2.0 SPS4, SAP developed a new native data tiering solution for warm data called NSE. NSE can process in-memory stored data for performance of critical operations, and it has a buffer-cache managed page to process less frequently accessed data.

To activate NSE, you must configure your warm data-related tables, or partitions of columns to be page-loadable by running SQL DDL commands.

After the table, partition, or column is configured to use NSE, it is not loaded into memory anymore, but it is readable by using the buffer cache. The performance for accessing data on that table is slower than accessing it in-memory.
Total cost of ownership reduction and performance advantages
If you use NSE and make 50% of your data page-loadable, then 50% of memory is freed from
the SAP HANA main memory, so persistent area and buffer cache is used instead. This
situation can reduce TCO because most of it is related to the amount of memory (DRAM) that
is needed in the server.

Also, depending on the amount of data that is moved to NSE, SAP HANA can start faster
during scheduled and unexpected maintenance.

Supportability
NSE supports SAP HANA native applications for SAP S/4HANA and SAP Suite on SAP
HANA (SoH).

Note: SAP recommends using NSE with SAP S/4HANA or SoH with Data Aging only. If
Data Aging is used in SAP S/4HANA or SoH with SAP HANA 2.0 SPS4, NSE is used for
storing the historical partitions of tables in aging objects. To learn more about this use case
for SAP S/4HANA, see SAP Note 2816823.

The NSE Advisor
The NSE Advisor comes bundled with NSE. Use the Advisor to learn what tables, partitions,
or columns are suitable to be converted to page-loadable (to save the memory space) or
column-loadable (to improve the performance) within the recommendations view result.

Based on the workload on the table, partition, or column over time, the NSE Advisor identifies
the frequently accessed and rarely accessed objects so that the system administrators can
decide which objects can be moved to NSE.

Figure 2-5 shows the NSE architecture perspective.

NSE comes with SAP HANA 2.0 SPS4, and it is supported on Power Systems. For more
information about using NSE, see SAP Note 2799997.
2.2.4 SAP HANA Dynamic Tiering

Starting in SAP HANA 1.0 SPS09, SAP introduced *SAP HANA Dynamic Tiering*, which is another option for storing warm data on disk. The data is stored in a columnar format to dedicate more space in memory for hot data, which decreases the TCO of your SAP HANA DB.

SAP HANA Dynamic Tiering does not come with the standard installation package, so you must download an extra component and install it on SAP HANA.

In SAP HANA Dynamic Tiering, you can create two types of warm tables: extended table and multistore table. The *extended table* type is disk-based, so all data is stored in disk. The *multistore table* type is an SAP HANA partitioned table with some partitions in memory and some partitions on disk.

The distribution of the data among the in-memory store tables, the extended tables, and the multistore tables is shown in Figure 2-6.

![SAP HANA Dynamic Tiering: Distribution of tables](image)

SAP HANA Dynamic Tiering can be installed in the same server hosting SAP HANA or in a separate dedicated host server. You can also install a second SAP HANA Dynamic Tiering host as a standby host for HA purposes.

The operating system (OS) process for the SAP HANA Dynamic Tiering host is `hdbesserver`, and the service name is `esserver`.

SAP HANA Dynamic Tiering supports only low-tenant isolation. Any attempt to provision the SAP HANA Dynamic Tiering service (`esserver`) to a tenant DB with high-level isolation fails. After the SAP HANA Dynamic Tiering implementation, if you try to configure the tenant DB with high isolation, the SAP HANA Dynamic Tiering service stops working.

**SAP HANA Dynamic Tiering configuration properties**

The `esserver.ini` file stores the extended storage configuration properties. With these parameters, you can configure the SAP HANA Dynamic Tiering for memory usage, concurrent connections, concurrent queries, and so on.

Table 2-1 on page 15 shows important parameters of SAP HANA Dynamic Tiering.
Table 2-1  SAP HANA Dynamic Tiering configuration properties

<table>
<thead>
<tr>
<th>Parameter name</th>
<th>Description</th>
<th>Default value</th>
</tr>
</thead>
<tbody>
<tr>
<td>catalog_cache</td>
<td>Amount of memory that is initially reserved for caching the dynamic tiering catalog.</td>
<td>32000000 bytes</td>
</tr>
<tr>
<td>delta_memory_mb</td>
<td>Amount of memory that is available to store a delta-enabled extended table.</td>
<td>2048 MB</td>
</tr>
<tr>
<td>heap_memory_mb</td>
<td>Amount of heap memory that dynamic tiering can use. A value of 0 or empty means no limit on heap memory.</td>
<td>1024 MB</td>
</tr>
<tr>
<td>load_memory_mb</td>
<td>Maximum amount of memory the extended storage can request from the OS for temporary use.</td>
<td>2048 MB</td>
</tr>
<tr>
<td>main_cache_mb</td>
<td>Amount of memory to be used for caching dynamic tiering DB objects.</td>
<td>1024 MB</td>
</tr>
<tr>
<td>max_concurrent_connections</td>
<td>Maximum number of concurrent connections that the dynamic tiering service accepts.</td>
<td>50</td>
</tr>
<tr>
<td>max_concurrent_queries</td>
<td>Maximum number of concurrent queries that are allowed by the server.</td>
<td>32</td>
</tr>
</tbody>
</table>

For a complete list of the parameters, see *SAP HANA Dynamic Tiering: Administration Guide*.

**Deployment options for SAP HANA Dynamic Tiering**

You can deploy SAP HANA Dynamic Tiering in a dedicated host, and in the same host of SAP HANA. An SAP HANA Dynamic Tiering service can coexist in the same SAP HANA host for only one SAP HANA tenant. Therefore, if you need SAP HANA Dynamic Tiering for any additional tenants, the `esserver` service must be in a different host.
**Same host deployment**

In this host deployment, you have just one SAP HANA host with the tenant (**indexserver**) and SAP HANA Dynamic Tiering (**esserver**) services in the same SAP HANA host. Starting with SAP HANA Dynamic Tiering 2.0, the same host deployment is supported for production environments. Figure 2-7 shows the configuration scenario.

![Figure 2-7 SAP HANA Dynamic Tiering: Same host deployment](image)

**Note:** Some host deployments are designed for small, nonproduction environments, but are supported in production environments.
**Dedicated host deployment**

In a dedicated host deployment, the SAP HANA tenant service and SAP HANA Dynamic Tiering service are in different SAP HANA hosts, and the SAP HANA Dynamic Tiering service of host B is associated with the SAP HANA tenant service of host A (Figure 2-8).

![Figure 2-8 SAP HANA Dynamic Tiering: Dedicated host deployment](image)

This is a deployment where the organizations can take advantage of all Power Systems benefits with the flexibility of LPAR support, with low network latency.
More than one SAP HANA Dynamic Tiering server deployment

In this scenario, there are two SAP HANA tenant DBs running in a same SAP HANA system. In this case, if SAP HANA Dynamic Tiering is needed for the two tenant DBs, the SAP HANA Dynamic Tiering service for the second tenant DB must be in an extra host because the SAP HANA tenant service (indexserver) and the SAP HANA Dynamic Tiering service (esserver) can coexist on the same host for only one tenant. Figure 2-9 shows the configuration for this scenario.

![Figure 2-9 SAP HANA Dynamic Tiering: More than one SAP HANA Dynamic Tiering server deployment](image)

2.2.5 SAP HANA on IBM Power Systems and SAP HANA Data Tiering solutions

With Power Systems, it is possible to implement a scale-up or scale-out SAP HANA DB environment.

With support for multiple LPARs, organizations can consolidate multiple SAP HANA instances or SAP HANA nodes of the same instance (multi-host) on a single Power Systems server by leveraging its simplicity of management with a low-latency network.

By using IBM PowerVM® (the Power Systems hypervisor), you can virtualize up to 16 production SAP HANA instances on the LPARs of a single Power Systems server (IBM Power Systems E950 or IBM Power Systems E980 server). It is also possible to move memory and CPUs among the LPARs with flexible granularity (for more information, see SAP Note 2230704).

PowerVM allows more granular scaling and dynamically changing allocation of system resources. You can avoid adding new hardware that can cause higher energy, cooling, and management needs.
The SAP HANA on Power Systems solution runs the same SUSE or Red Hat Linux distributions as x86 servers, with the flexibility, scalability, resiliency, and performance advantages of Power Systems servers that help the client to accomplish the following tasks:

- Accelerate SAP HANA deployments by using the flexibility of built-in virtualization, which allows faster provisioning of SAP HANA instances and allocating capacity with granularity as little as 0.01 cores and 1 GB.
- Minimize infrastructure and simplify management by using the following capabilities:
  - Virtualization scalability of up to 24 TB in scale-up mode.
  - The ability to deploy up to 16 SAP HANA modules in a single server.
  - Shared processor pools that optimize CPU cycles across SAP HANA virtual machines (VMs) in a server.

Power Systems is the best solution for implementing SAP HANA scale-up and scale-out modes and the data tiering solutions that are shown in this section. For more information, see SAP HANA Server Solutions with Power Systems.

### 2.2.6 SAP HANA Dynamic Tiering: Hands-on demonstration

This section demonstrates the implementation of SAP HANA Dynamic Tiering.

**Before you start**

This section describes a high-level, step-by-step installation of SAP HANA Dynamic Tiering on SAP HANA 2.0 on Power Systems (Little Endian), and a demonstration about how to create an extended storage and multistore table for manipulating data on it.

You use the Console Interface (CI) to do an SAP HANA Data Tiering on a Same Host Deployment, that is, on the same server where the SAP HANA is installed with no additional SAP HANA nodes.

**Note:** Some data manipulation is demonstrated by using SAP HANA Studio.


**Preparing the installation**

This section shows you how to prepare for the installation.

**Checking the hardware and operating system requirements**

According to the *SAP HANA Dynamic Tiering Installation Guide*, SAP HANA Dynamic Tiering is available for:

- Intel -based hardware platforms
- Power Systems servers

**Note:** Power Systems environments require the appropriate IBM XL C/C++ redistributable libraries. Download and install the appropriate runtime environment for the latest updates from the supported IBM C and C++ Compilers page on the IBM Support Portal. Install the libraries on both the SAP HANA and SAP HANA Dynamic Tiering hosts. These libraries are not required for an Intel -based hardware platform environment.
Checking SAP HANA compatibility with the SAP HANA Dynamic Tiering version

For a matrix of compatible SAP HANA and SAP HANA Dynamic Tiering versions, see SAP Note 2563560.

Downloading the SAP HANA Dynamic Tiering package

To download the SAP HANA Dynamic Tiering package, go to the SAP Software Download Center and select Support Packages and Patches → By Alphabetical Index (A-Z) → H → SAP HANA DYNAMIC TIERING → SAP HANA DYNAMIC TIERING 2.0 → COMPRISED SOFTWARE COMPONENT VERSIONS → SAP HANA DYNAMIC TIERING 2.0. At the top of the download page, click LINUX ON POWER LE 64 BIT.

Now, you can download the SAP HANA Dynamic Tiering revision that is compatible with the SAP HANA 2.0 SP level you have in place or are installing.

Extracting the SAP HANA Dynamic Tiering package

After you download and copy the package to the SAP HANA host, extract it by running the following command:

SAPCAR -manifest SIGNATURE.SMF -xvf DYNTIERING20004_3-70002283.SAR

Installing SAP HANA Dynamic Tiering (same host deployment)

To install SAP HANA Dynamic Tiering by using the CI, complete the following steps:

1. Log in to the SAP HANA host by using the root ID and change to the resident hdblcm directory within the <sid> folder, for example, /hana/shared/H01/hdblcm, where /hana/shared/ is the shared resources location for the system and H01 is the HANA <sid>.

2. Start the SAP HANA platform lifecycle management tool by running the command that is shown in Example 2-1, where <full-path-option> refers to the path where SAP HANA Dynamic Tiering installation package was extracted.

   Example 2-1  SAP HANA Dynamic Tiering installation command for starting the installation

   ./hdblcm --component_dirs=/<full_path_option>

   For this demonstration, the installation package path is /usr/sap/tmp/ebf29158/HANA_DYNAMIC_TIERING.

3. At the Choose an action prompt, select Install or Update Additional Components, as shown in Example 2-2.

   Example 2-2  SAP HANA Dynamic Tiering installation: Choose an action window

<table>
<thead>
<tr>
<th>Index</th>
<th>Action</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>add_host_roles</td>
<td>Add Host Roles</td>
</tr>
<tr>
<td>2</td>
<td>add_hosts</td>
<td>Add Hosts to the SAP HANA Database System</td>
</tr>
<tr>
<td>3</td>
<td>check_installation</td>
<td>Check SAP HANA Database Installation</td>
</tr>
<tr>
<td>4</td>
<td>configure_internal_network</td>
<td>Configure Inter-Service Communication</td>
</tr>
<tr>
<td>5</td>
<td>configure_sld</td>
<td>Configure System Landscape Directory Registration</td>
</tr>
</tbody>
</table>
Enter selected action index [16]: 14

4. At the Choose components to be installed or updated prompt, select Install SAP HANA Dynamic Tiering, as shown in Example 2-3.

Example 2-3  SAP HANA Dynamic Tiering installation: Choose components to be installed or updated window

Choose components to be installed or updated:

<table>
<thead>
<tr>
<th>Index</th>
<th>Components</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>all</td>
<td>All components</td>
</tr>
<tr>
<td>2</td>
<td>es</td>
<td>Install SAP HANA Dynamic Tiering version 2.0.043.00.12711</td>
</tr>
</tbody>
</table>

Enter comma-separated list of the selected indices: 2

5. You are prompted to add another host. Enter no, as shown in Example 2-4.

Example 2-4  SAP HANA Dynamic Tiering installation: Add hosts window

Verifying files...
Do you want to add hosts to the system? (y/n) [n]:

6. You are prompted to enter the System Database User Name and password. Enter SYSTEM and its password, as shown in Example 2-5.

Example 2-5  SAP HANA Dynamic Tiering installation: User and password window

Enter System Database User Name [SYSTEM]:
Enter System Database User (SYSTEM) Password:

7. You are prompted to add the paths for SAP HANA Dynamic Tiering data and log volume paths. In this case, the paths are /hana/data/dt_es/H01 for the data volumes and /hana/log/dt_es/H01 for the log volumes, as shown in Example 2-6.

Example 2-6  SAP HANA Dynamic Tiering: Installation

Enter Location of Dynamic Tiering Data Volumes [/hana/data_es/H01]:
/hana/data/dt_es/H01
Enter Location of Dynamic Tiering Log Volumes [/hana/log_es/H01]:
/hana/log/dt_es/H01
8. You are prompted to confirm all the added parameters. Enter \textit{y}, as shown in Example 2-7.

\textbf{Example 2-7} \ SAP HANA Dynamic Tiering installation: Summary window

\begin{verbatim}
Summary before execution:
==========================
SAP HANA Database
  Update Parameters
    Remote Execution: ssh
    Update Execution Mode: standard
    System Database User Name: SYSTEM
    Location of Dynamic Tiering Data Volumes: /hana/data/dt_es/H01
    Location of Dynamic Tiering Log Volumes: /hana/log/dt_es/H01
Software Components
  SAP HANA Dynamic Tiering
    Install version 2.0.043.00.12711
    Location: /usr/sap/tmp/ebf29158/HANA_DYNAMIC_TIERING/es
Log File Locations
  Log directory:
    /var/tmp/hdb_H01_hdblcm_update_components_2019-10-25_16.29.52
  Trace location: /var/tmp/hdblcm_2019-10-25_16.29.52_2245.trc

Do you want to continue? (y/n): y
\end{verbatim}

During the installation process until it is complete, you see similar messages as shown in Example 2-8.

\textbf{Example 2-8} \ SAP HANA Dynamic Tiering installation process

\begin{verbatim}
Installing components...
Installing SAP HANA Dynamic Tiering...
  Installing jre8...
  Installing shared...
  Installing lang...
  Installing conn_lm...
  Installing conn_add_lm...
  Installing odbc...
  Installing client_common...
  Installing server...
  Installing complete - log files written to /hana/shared/H01/es/log
Restarting HANA System...
  Copying delivery unit HANA_TIERING
  Copying delivery unit HDC_TIERING
Updating Component List...
SAP HANA Database components updated
\end{verbatim}

The installation log files are in the following path:

/\texttt{var/tmp/hdb\_H01\_hdblcm\_update\_components\_<date\_and\_time>}/hdblcm.log
SAP HANA Dynamic Tiering postinstallation activities

This section describes the postinstallation steps after the SAP HANA Dynamic Tiering installation.

Adding the SAP HANA Dynamic Tiering role

To add the SAP HANA Dynamic Tiering role, complete the following steps:

1. While logged in to the SAP HANA host with root ID, go to the path \\
hana/\shared/H01/hdb1cm, run the SAP HANA Lifecycle Management tool, and select Add Host Roles, as shown in Example 2-9.

```
Example 2-9   SAP HANA Dynamic Tiering role: Running the SAP HANA Lifecycle Management tool

./hdblcm
SAP HANA Lifecycle Management - SAP HANA Database 2.00.044.00.1571081837
************************************************************************
Choose an action
Index | Action                     | Description
----------------------------------------------------------------------------------
1     | add_host_roles             | Add Host Roles
2     | add_hosts                  | Add Hosts to the SAP HANA Database System
3     | check_installation         | Check SAP HANA Database Installation
4     | configure_internal_network | Configure Inter-Service Communication
5     | configure_sld              | Configure System Landscape Directory Registration
6     | extract_components         | Extract Components
7     | print_component_list       | Print Component List
8     | remove_host_roles          | Remove Host Roles
9     | rename_system              | Rename the SAP HANA Database System
10    | uninstall                  | Uninstall SAP HANA Database Components
11    | unregister_system          | Unregister the SAP HANA Database System
12    | update                     | Update the SAP HANA Database System
13    | update_component_list      | Update Component List
14    | update_components          | Install or Update Additional Components
15    | update_host                | Update the SAP HANA Database Instance Host integration
16    | exit                       | Exit (do nothing)
```

Enter selected action index [16]: 1

2. Select the SAP HANA host to be assigned the additional SAP HANA Dynamic Tiering role. In this case, there is just one host, as shown in Example 2-10.

```
Example 2-10   SAP HANA Dynamic Tiering role: Host selection for adding role

System Properties:
H01 /hana/\shared/H01 HDB_ALONE
   HDB00
       version: 2.00.044.00.1571081837
       host: pdemo1 (Database Worker (worker))
       edition: SAP HANA Database

Select hosts to which you would like to assign additional roles

Index | System host | Roles
-----------------------------------------------
```
3. In Select additional host roles for host '<host>', select the host. In this case, there is just one host. Add the <sid>adm ID password for it, as shown in Example 2-11.

Example 2-11  SAP HANA Dynamic Tiering role: Selection of host

Select additional host roles for host 'pdemo1'

<table>
<thead>
<tr>
<th>Index</th>
<th>Additional Role</th>
<th>Role Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>extended_storage_worker</td>
<td>Dynamic Tiering Worker (extended_storage_worker)</td>
</tr>
</tbody>
</table>

Enter comma-separated list of additional roles for host 'pdemo1' [1]: Enter System Administrator (h01adm) Password:

4. You are prompted to confirm all added parameters. Confirm and enter y, as shown in Example 2-12.

Example 2-12  SAP HANA Dynamic Tiering role: Installation summary window

Summary before execution:
================================
Add Host Roles
Add Host Roles Parameters
  Do not start hosts with modified roles: No
  Remote Execution: ssh
  Auto Initialize Services: Yes
  Do not Modify '/etc/sudoers' File: No
Additional Host Roles
  pdemo1
    Current Role(s): worker
    Additional Role(s): extended_storage_worker
Log File Locations
  Log directory: /var/tmp/hdb_H01_hdblcm_add_host_roles_2019-10-25_16.52.34
  Trace location: /var/tmp/hdblcm_2019-10-25_16.52.34_828.trc

Do you want to continue? (y/n): y

At the end of the process, you see a summary of the installation as shown in Example 2-13.

Example 2-13  SAP HANA Dynamic Tiering role: Installation process

Assigning Additional Roles to the Local Host...
  Adding role 'extended_storage_worker' on local host 'pdemo1'...
  Performing esaddhost...
    esaddhost.sh: Configuring SAP ES...
  Stopping instance...
    Stopping 8 processes on host 'pdemo1' (extended_storage_worker, worker):
Stopping on 'pdemo1' (extended_storage_worker, worker): hdbdaemon, hdbcompileserver, hdbeeserver, hdbindexserver (H01), hdbnameserver, hdbpreprocessor, hdbwebdpsr, hdbxsengine (H01)

All server processes stopped on host 'pdemo1' (extended_storage_worker, worker).

Stopping sapstartsrv service...
Starting service (sapstartsrv)...
Starting instance H01 (HDB00) on host 'pdemo1'...
Starting 8 processes on host 'pdemo1' (extended_storage_worker, worker):
Starting on 'pdemo1' (extended_storage_worker, worker): hdbdaemon, hdbcompileserver, hdbeeserver, hdbindexserver (H01), hdbnameserver, hdbpreprocessor, hdbwebdpsr, hdbxsengine (H01)
Starting on 'pdemo1' (extended_storage_worker, worker): hdbdaemon, hdbeeserver, hdbindexserver (H01), hdbwebdpsr, hdbxsengine (H01)
Starting on 'pdemo1' (extended_storage_worker, worker): hdbdaemon, hdbeeserver, hdbwebdpsr, hdbxsengine (H01)
Starting on 'pdemo1' (extended_storage_worker, worker): hdbdaemon, hdbwebdpsr, hdbxsengine (H01)
Starting on 'pdemo1' (extended_storage_worker, worker): hdbdaemon, hdbwebdpsr

All server processes started on host 'pdemo1' (extended_storage_worker, worker).

Additional host roles successfully assigned

The installation log files are in the following path:
/var/tmp/hdb_H01_hdblcm_add_host_roles_<date_and_time>/hdblcm.log

5. Using SAP HANA Studio, log in to the tenant DB as the SYSTEM ID. In the Overview tab, you see the SAP HANA Dynamic Tiering status as “Installed but not running yet” for that tenant, as shown in Figure 2-10.

![Figure 2-10 SAP HANA Dynamic Tiering: Yellow status in SAP HANA Studio](image-url)
6. Click the **Landscape** tab, and you see the SAP HANA Dynamic Tiering Server service **esserver**, as shown in Figure 2-11.

![Figure 2-11 SAP HANA Dynamic Tiering: Service esserver in SAP HANA Studio](image)

**Creating extended storage**

After you are done with the SAP HANA Dynamic Tiering installation and role additions, create extended storage. The extended storage is a DB space that is the file on disk where tables and partitions store the SAP HANA Dynamic Tiering.

To create the extended storage, you must have the system privileges EXTENDED STORAGE ADMIN.

**Note:** The extended storage is created with SYSTEM ID now, but for all subsequent activities, you need another user ID with all necessary privileges for the subsequent activities, and it is the owner of the extended storage and multistore tables.

Complete the following steps:

1. On SAP HANA Studio, connect to the SAP HANA Tenant DB to which the SAP HANA Dynamic Tiering was provisioned to with the SYSTEM ID.

   **Note:** In this demonstration, there is a single tenant that is the initial tenant, and the SAP HANA instance never previously contained more tenants. So, the SAP HANA Dynamic Tiering service is automatically provisioned to the tenant DB.

2. For this demonstration, create one extended storage with 1 GB of available allocated space. Right-click the tenant database and click **Open SQL Console** to open the SQL console. Then, run the command that is shown in Example 2-14.

   **Example 2-14 SAP HANA Dynamic Tiering: Command for creating extended storage**

   ```sql
   CREATE EXTENDED STORAGE AT 'pdemo1' SIZE 1000 MB;
   ```

   Notice that pdemo1 is the location (host) that is used in this demonstration. Adjust it to your host. The result is shown in Figure 2-12 on page 27.
3. Click the **Overview** tab in SAP HANA Studio, you see the status of SAP HANA Dynamic Tiering as “Running”, as shown in Figure 2-13.

![Figure 2-13 SAP HANA Dynamic Tiering: Running status in SAP HANA Studio](image)

Your SAP HANA Dynamic Tiering is now ready for you to create an Extended Table or multistore table.

**Creating a user ID for your SAP HANA Dynamic Tiering objects**

You created your extended storage with the tenant SYSTEM ID. Now, to create your extended table and multistore tables, you need a new user ID that creates and owns these objects.

To create the user ID by using SAP HANA Studio, complete the following steps:

1. Click **Tenant → Security**, right-click **Users**, and click **New User**.
2. Define a name for the user and add the System Privileges CATALOG READ, EXTENDED STORAGE ADMIN, and IMPORT. You do not have to force the password change in the first logon if you prefer.
The parameters are shown in Figure 2-14.

In this case, the user ID is RPDT. Log in to the tenant with the user ID that was defined by you.

**Creating an extended table**

You can create an extended table in the same way you create any other column table in an SAP HANA DB; the only difference is that you must add `USING EXTENDED STORAGE` at the end of the statement, as shown in Example 2-15.

**Example 2-15  SAP HANA Dynamic Tiering: Extended table creation command-line interface**

```sql
CREATE TABLE "RPDT"."CUSTOMER_ES" (  
  C_CUSTKEY    integer     not null,  
  C_NAME       varchar(25)  not null,  
  C_ADDRESS    varchar(40)  not null,  
  C_PHONE      char(15)     not null,  
  primary key (C_CUSTKEY)  
) USING EXTENDED STORAGE;
```
The same table was created in this demonstration, as shown in Figure 2-15.

![Figure 2-15  SAP HANA Dynamic Tiering: Extended table CUSTOMER_ES creation](image)

In left pane of the window, the table is identified as an EXTENDED table in the SAP HANA Catalog for user RPDT.

**Important:** Foreign keys between two extended tables or between an extended table and an in-memory table are *not* supported.

To insert data, use the same syntax as a common in-memory column store table, as shown in Example 2-16 and in Figure 2-16.

**Example 2-16  SAP HANA Dynamic Tiering: Insert data into table CUSTOMER_ES command-line interface**

```
INSERT INTO "RPDT"."CUSTOMER_ES"
(C_CUSTKEY, C_NAME, C_ADDRESS, C_PHONE)
VALUES
(1,'CUSTOMER 1','ADDRESS 1', 19999999999);
```

![Figure 2-16  SAP HANA Dynamic Tiering: Insert data into table CUSTOMER_ES - SAP HANA Studio window](image)
To read data from the table, the SQL syntax that you use is the same for reading data from any other table. So, you can select data in the same way as any other in-memory column store table, as shown in Figure 2-17.

Figure 2-17 SAP HANA Dynamic Tiering: Contents of the extended table CUSTOMER_ES in SAP HANA Studio

Creating a multistore table

When creating a multistore table, you must split it in partitions and define which partitions are part of the default store (for example, the memory store) and which partitions are part of the extended storage.

Complete the following steps:

1. Create a table called SALES_ORDER and partition it by RANGE by using a date type field as the partition field. In the default store, the values are 2010/12/31 - 9999/12/31, and in the extended storage, the values are 1900/12/31 - 2010/12/31, as shown in Example 2-17.

Example 2-17 SAP HANA Dynamic Tiering: Creating a multistore table command-line interface

```
CREATE TABLE "RPDT"."SALES_ORDER" ( 
  S_SALESOKEY           integer           not null, 
  S_CUSTOMER             integer           not null, 
  S_VALUE                decimal(15,2)     not null, 
  S_DATE                 date              not null, 
primary key (S_SALESOKEY,S_DATE)) 
PARTITION BY RANGE ("S_DATE") 
( 
  USING DEFAULT STORAGE 
  (PARTITION '2010-12-31' <= VALUES < '9999-12-31') 
  USING EXTENDED STORAGE 
  (PARTITION '1900-12-31' <= VALUES < '2010-12-31'));
```

In SAP HANA Studio, you see that the multistore table symbol differs from the extended table symbol, as shown in Figure 2-18 on page 31.
The Data Manipulation Language (DML) operations on the table do not differ from any other table type.

2. Run a query from the TABLE_PARTITIONS table as shown in Example 2-18, and you see that the new table has two partitions: one in the default store, and another one in extended storage, as shown in Figure 2-19.

Example 2-18  SAP HANA Dynamic Tiering: Query SALES_ORDER table from the TABLE_PARTITIONS table

```
SELECT SCHEMA_NAME, TABLE_NAME, PART_ID, STORAGE_TYPE FROM TABLE_PARTITIONS
WHERE TABLE_NAME = 'SALES_ORDER' AND SCHEMA_NAME = 'RPDT'
```

<table>
<thead>
<tr>
<th>SCHEMA_NAME</th>
<th>TABLE_NAME</th>
<th>PART_ID</th>
<th>STORAGE_TYPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>RPDT</td>
<td>SALES_ORDER</td>
<td>1</td>
<td>DEFAULT</td>
</tr>
<tr>
<td>RPDT</td>
<td>SALES_ORDER</td>
<td>2</td>
<td>EXTENDED</td>
</tr>
</tbody>
</table>

Figure 2-19  SAP HANA Dynamic Tiering: Results from TABLE_PARTITIONS table

3. Insert a row into the table so that it is stored in the default store by running the command that is shown in Example 2-19.

Example 2-19  SAP HANA Dynamic Tiering: Inserting a row in the default store partition of the SALES_ORDER table

```
INSERT INTO "RPDT"."SALES_ORDER"
(S_SALESOKEY, S_CUSTOMER, S_VALUE, S_DATE)
VALUES
(1,1,120,'2011-12-11');
```
From table M_CS_TABLES (the table that shows data that is stored in the default store), which is shown in Example 2-20, you see that there is one record in the default store as part of SALES_ORDER table, as shown in Figure 2-20.

**Example 2-20  Checking the count of default store rows of the table SALES_ORDER**

```
SELECT RECORD_COUNT FROM M_CS_TABLES WHERE TABLE_NAME = 'SALES_ORDER' AND SCHEMA_NAME = 'RPDT';
```

---

![SQL Result](image)

**Figure 2-20  SAP HANA Dynamic Tiering: Results for the count of default store rows of the table SALES_ORDER**

4. Insert one more row in the extended storage of the table SALES_ORDER, as shown in Example 2-21.

**Example 2-21  Inserting a row in the extended storage partition of the SALES_ORDER table**

```
INSERT INTO "RPDT"."SALES_ORDER"
(S_SALESOKEY, S_CUSTOMER, S_VALUE, S_DATE)
VALUES
(1,1,120,'2009-12-11');
```

---

From the table M_ES_TABLES (the table that shows data that is stored in the extended storage) as shown in Example 2-22, you see that there is one record in the extended storage as part of SALES_ORDER table, as shown in Figure 2-21.

**Example 2-22  Checking the count of extended storage rows of the table SALES_ORDER**

```
SELECT * FROM M_ES_TABLES WHERE TABLE_NAME = 'SALES_ORDER' AND SCHEMA_NAME = 'RPDT';
```

---

![SQL Result](image)

**Figure 2-21  SAP HANA Dynamic Tiering: Results for the count of extended storage rows of the table SALES_ORDER**
When running a normal DML `select` command, you can see the two rows normally with no distinction between a default store and extended storage, as shown in Figure 2-22.

![Figure 2-22 SAP HANA Dynamic Tiering: Normal select statement from a multistore table](image)
SAP HANA I/O performance enhancements

As SAP HANA Online Analytical Processing (OLAP) databases (DBs) become larger, restarting them takes more time, which impacts the availability of applications and the efficiency of your business.

This chapter describes different solutions that can help speed up the starting of large SAP HANA DBs to help minimize downtime.

This chapter describes the following topics:
- Virtual persistent memory
- RAM disk (temporary file systems): SAP HANA Fast Restart Option
- Persistent disk storage by using native Non-Volatile Memory Express devices
- NVMe Rapid Cold Start Mirror
- Comparing vPMEM to Intel Optane
- Impact to Live Partition Mobility capabilities
3.1 Virtual persistent memory

In the quest to provide better performance for applications, new technologies across the computer industry are developed to help mitigate the inherent slowness of many of today's persistent disk-based storage solutions. IBM has a new PowerVM Persistent Memory architecture that is implemented at the hypervisor level, and IBM is developing several solutions to address this need. The first solution that was released is a new virtual persistent memory (vPMEM) solution that provides for high-speed access to critical business data. This vPMEM technology is integrated into the IBM PowerVM hypervisor for POWER9™ systems, and it provides a high-speed persistent RAM disk storage solution for applications that persist across operating system (OS) and logical partition (LPAR) restarts.

The PowerVM Persistent Memory architecture allows for multiple types of memory to be defined and deployed for different use cases. Currently, the vPMEM solution creates persistent storage volumes from standard system DRAM, which provides high-speed persistent access to data for applications running in an LPAR. For this solution, no special memory or storage devices are required, just unused available system DRAM. There is the intention to support enhancements that will allow access to other types of memory to be used for different use cases.

The vPMEM solution is incorporated into the PowerVM hypervisor management interface for POWER9 and later systems. vPMEM volumes are created as part of a specific LPAR definition. Each defined LPAR on a system can have a dedicated vPMEM volume. Individual vPMEM volumes are not sharable between LPARs, and vPMEM volumes are not transferable to another LPAR.

The PowerVM hypervisor allocates an independent segment of system memory for the vPMEM volume and associates it to the LPAR. This system memory segment is separate from the DRAM memory that is defined for the LPAR to use for the OS to run applications, as defined in the LPAR profile. After the application uses this persistent system memory volume as a disk resource, any data that is stored in the vPMEM device persists if the LPAR is restarted.

Access to vPMEM volumes by the Linux OS is provided by the standard non-volatile memory device (libnvdimm) subsystem in the Linux kernel corresponding ndctl utilities. The resulting vPMEM volumes are then mounted on the Linux file system as Direct Access (DAX) type volumes.

3.1.1 SAP HANA usage of vPMEM volumes

On POWER9 systems, SAP HANA2 SPS04 revision 44 and later can use DAX-based persistent memory volumes to store column store table data for fast access. Persistent memory DAX file systems bypass the traditional file system page cache mechanisms, which increases the access speed to the data that is stored on those volumes compared to disk-based file systems.

When SAP HANA detects the presence of DAX vPMEM volumes, it starts the copy of main column store table data in these defined persistent volumes. Through its default settings, SAP HANA attempts to copy all compatible column store table data into the vPMEM volumes, maintaining a small amount of space in the LPAR DRAM for column store table metadata. This situation creates a persistent memory copy of the table data that SAP HANA then uses for query and transactional processing. SAP HANA can also be configured to copy in to vPMEM-only column store tables, or even specific partitions of individual column store tables.
To access the column store table data on the vPMEM volumes, SAP HANA creates memory-mapped pointers from the DRAM memory structures to the column store table data. Considering these vPMEM volumes are allocated from memory, accessing the table data is done at memory speeds with no degradation in performance compared to when the column store data is stored without vPMEM in DRAM.

Any changed or added data to tables loaded in to the vPMEM device is synchronized to disk-based persistent storage with normal SAP HANA save point activity. When SAP HANA is shut down, all unsaved data that is stored in the LPAR DRAM and the vPMEM volumes is synchronized to persistent disk.

When SAP HANA shuts down, the column store table data persists in the vPMEM volumes. The next time SAP HANA starts, it detects the presence of the column store table data in the vPMEM volumes and skips loading that data. SAP HANA re-creates memory structure pointers to the columnar table data that is stored on the vPMEM volumes, which results in a reduction in SAP HANA startup times, as shown in Figure 3-1.

This chart shows that substantial time savings for SAP HANA start for a large OLAP DB when all of the DBs column store data is allocated in vPMEM volumes. There are also time savings for SAP HANA if shut down because not as much DRAM memory is required to be programatically tagged as freed for ceding control to the OS.
3.1.2 Sizing the vPMEM volume

Before creating the vPMEM volume in the Hardware Management Console (HMC), you should know the total required size of the column store table data. A simple query to the DB can give an estimate of the amount of memory in use by all column store tables and the estimated size that is required of a vPMEM volume, as shown in Example 3-1.

Example 3-1  Querying the amount of memory in use

```sql
select to_decimal(round(sum(memory_size_in_total/1024/1024),2,ROUND_UP),20,2) COL_STORE_MB from m_cs_tables where schema_name = '<database schema name>'
```

Add an extra 10 - 15% memory than what is required to the vPMEM volume to allow for growth of the DB over time.

If less LPAR DRAM memory is used by SAP HANA to store the column store table data that is on the vPMEM volume, the RAM memory allocation for the LPAR can be reduced by a similar amount to avoid using more system memory that is required for the LPAR. This memory definition adjustment is done in the LPAR profile on the HMC.

3.1.3 Checking the LPAR hardware assignment

It is important to have a vPMEM target LPAR assigned memory and CPU resources evenly across non-uniform memory access (NUMA) resource domains. SAP supports only LPAR and subsequent vPMEM allocations that are evenly distributed across the NUMA domains because PowerVM allocates memory resources to the LPAR that are in proportion to the CPU allocation across the NUMA domains. If the LPAR is assigned CPU resources of 10 cores on one NUMA domain, say NUMA0, but only 5 on another domain, say NUMA1, the memory allocation follows this ratio, where the memory allocated on NUMA0 is twice the size as on NUMA1. This imbalance is also reflected in the size of the vPMEM volumes when they are created, which is a configuration that is not supported.

The resource allocation of CPU and memory for a system’s LPARs can be queried by performing a resource dump from the HMC. There are two methods:

- Within the HMC, which is described in How to Initiate a Resource Dump from the HMC - Enhanced GUI.
- By logging on to the LPAR’s HMC command-line interface (CLI) with a privileged HMC user account, like hscroot, and running the command that is shown in Example 3-2.

Example 3-2  Starting a resource dump from the HMC command-line interface

```bash
startdump -m <system name> -t resource -r 'hvlpconfigdata -affinity -domain'
```

Substitute `<system_name>` with the system name that is defined on the HMC that is hosting the LPAR.

Both methods create a resource dump file in the `/dump` directory that is timestamped. Depending on the size of the system, it can take a few minutes before the dump file is ready.

You can view the list of resource dump files on the HMC listed in chronological order by running the command that is shown in Example 3-3.

Example 3-3  Listing the resource dump files in the `/dump` directory

```bash
ls -ltr /dump/RSCDUMP*
```
The last file in the list can be viewed by running the `less` command, as shown in Example 3-4.

**Example 3-4  Viewing the contents of the RSCDUMP file**

```bash
less /dump/RSCDUMP.<system serial number>.<auto generated dump number>.<date stamp>
```

The `less` command determines that the file can be in binary form because some of the data in the file is in this format. The details are in text form, as shown in Example 3-5, for an IBM Power System E950 system with four sockets and 2 TB of RAM, with a single LPAR running that has been allocated 48 cores and 1 TB of the available memory.

**Example 3-5  Main section of the RSCDUMP file list CPU and memory resources that are assigned to LPARs**

<table>
<thead>
<tr>
<th>Domain</th>
<th>Procs Units</th>
<th>Memory</th>
<th>Proc Units</th>
<th>Memory</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEC</td>
<td>PRI</td>
<td>Total</td>
<td>Free</td>
<td>Total</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1200</td>
<td>0</td>
<td>2048</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1200</td>
<td>0</td>
<td>2048</td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>1200</td>
<td>0</td>
<td>2048</td>
</tr>
<tr>
<td>0</td>
<td>3</td>
<td>1200</td>
<td>0</td>
<td>2048</td>
</tr>
</tbody>
</table>

In Example 3-5, the columns of data that are of interest in this context have the following meanings:

**Domain SEC**

The socket number in which the cores and memory are installed. In Example 3-5, the system has four sockets, 0 - 3.

**Domain PRI**

The NUMA domain number. Example 3-5 has four NUMA domains, 0 - 3, and each NUMA domain aligns to a socket number. Some Power Systems servers have two NUMA domains per socket.
| **Procs Total** | The number of processors in a 1/100th of a processor increment. Because PowerVM can allocate subprocessor partitions in 1/100th of a single core, this number is 100 times larger than the actual number of cores on the NUMA domain. Example 3-5 on page 39 shows each socket has a total of 12 cores per socket. |
| **Procs Free / Units Free** | The total number of 1/100th of a core processor resources that are available. |
| **Memory Total** | The total amount of memory that is available on the NUMA domain. This number is four times larger than the actual memory in gigabytes that is available. Example 3-5 on page 39 shows each socket has 512 GB of RAM installed, for a total system capacity of 2 TB. |
| **Memory Free** | The amount of memory that is not in use by assigned LPARs on the NUMA domain. Again, this value is four times larger than the actual memory in gigabyte of available memory. This measurement is an important detail in determining the amount of memory that can be used for a vPMEM volume because this value decreases after the creation of the vPMEM volume. Example 3-5 on page 39 shows sockets 0, 2, and 3 all have about 75 GB of available memory, and socket 1 has about 65 GB of available memory. This system has a vPMEM volume of 700 GB that is assigned to the running LPAR. |
| **LP** | The LPAR number as defined in the HMC. Example 3-5 on page 39 shows only one LPAR running, LPAR number 1, which is assigned all CPU resources, and a subset of the available RAM resources. |
| **Proc Units Tgt** | The number of subprocessor processor units that is assigned to the LPAR from the NUMA domain. They are allocated by using the value from the Procs Total column. Example 3-5 on page 39 shows that the target allocation of processing units is 1200. |
| **Proc Units Aloc** | The number of subprocessor units that are allocated to the LPAR. Example 3-5 on page 39 shows that all 1200 units per socket are assigned and activated to the LPAR across all four NUMA domains or sockets. |
| **Memory Tgt** | The amount of memory that is assigned to the LPAR's DRAM configuration as defined in the LPAR profile. Again, this value is four times larger than the actual memory (gigabytes) that is assigned, and the hypervisor allocates this memory per NUMA domain in the same ratio as the processing unit assignment to the rest of the processing unit allocation across the other assigned NUMA domains. Example 3-5 on page 39 shows around 256 GB is targeted to be allocated to each NUMA domain, and in the same ratio as processing units. The memory is evenly distributed just as the processing units are evenly distributed. |
| **Memory Aloc** | The real allocation of memory to the LPAR per NUMA domain. Example 3-5 on page 39 shows all that memory that has requested is allocated to the LPAR. Summing up these values across the system reflects the LPAR DRAM memory allocation as seen by the OS. |
If the system has vPMEM volumes assigned, this memory allocation is not explicitly listed in this output. The memory values for the LPARs are the ones that are assigned to the LPAR’s memory allocation in the profile. To determine the approximate amount of memory vPMEM volumes are taking on a specific socket, add up the memory allocations for the LPARs on that socket and subtract that value from the Memory Total. Taking this result and subtracting the value from Memory Free shows the amount of RAM that is used by the vPMEM volume, as shown in Example 3-6.

Example 3-6   General vPMEM memory allocation for a single socket

\[
\text{vPMEM memory allocation GB} = \left( \frac{(\text{Memory Total} - \text{sum(All LPAR MemoryAloc)} - \text{Memory Free})}{4} \right)
\]

Example 3-5 on page 39 for socket 0 used the values that are shown in Example 3-7.

Example 3-7   Calculating the vPMEM memory allocation for socket 0

\[
\text{socket0 vPMEM memory} = \left( \frac{2048 - 1023 - 311}{4} \right) = 178. \text{ GB}
\]

Considering the same memory allocation is assigned across all four nodes, the total vPMEM device that is allocated to this LPAR is about 714 GB.

3.1.4 Creating a vPMEM device

Creating a vPMEM memory device for an LPAR is done on the HMC by modifying the properties of the LPAR:

1. Click **System definition** in the HMC to get a list of available LPARs, and then click the LPAR’s name to get the general details of the partition. Then, click the **Persistent Memory** property to show the details for the vPMEM volumes, as shown in Figure 3-2.

![Figure 3-2 General properties: Persistent Memory options](image)
By default, the list of vPMEM volumes is empty, as shown in Figure 3-3.

![Persistent Memory pane: Empty list of defined vPMEM volumes](image)

Figure 3-3  Persistent Memory pane: Empty list of defined vPMEM volumes

2. Click **Add** to add a vPMEM device, as shown in Figure 3-4.

![Adding a 1TB vPMEM volume](image)

Figure 3-4  Adding a 1TB vPMEM volume

3. Add a descriptive volume name, the total size in megabytes of the vPMEM device, and select the **Affinity** check box. Click **OK**, which creates a single vPMEM device for the LPAR.
### 3.1.5 vPMEM affinity configuration options

When the vPMEM volume is configured for affinity as shown in Figure 3-5, the hypervisor divides the total allocated memory into equal parts that fit into the system memory of each of the LPAR's assigned NUMA memory nodes. From the OS level, this division results in multiple `/dev/nmemX` devices that are defined, with one device for each NUMA node that is allocated to the LPAR.

<table>
<thead>
<tr>
<th></th>
<th>2TB</th>
<th>2TB</th>
<th>2TB</th>
<th>2TB</th>
</tr>
</thead>
<tbody>
<tr>
<td>LPAR0 vPMEM</td>
<td>/dev/nmem0 1TB</td>
<td>/dev/nmem1 1TB</td>
<td>/dev/nmem2 1TB</td>
<td>/dev/nmem3 1TB</td>
</tr>
<tr>
<td>LPAR0 RAM</td>
<td>512GB DRAM</td>
<td>512GB DRAM</td>
<td>512GB DRAM</td>
<td>512GB DRAM</td>
</tr>
<tr>
<td>NUMA0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NUMA1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NUMA2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NUMA3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Figure 3-5  An 8 TB system with memory that is allocated across four NUMA domains*

Figure 3-5 shows an 8 TB system with memory that is allocated across four NUMA domains. Creating a 4 TB vPMEM device with NUMA affinity creates one vPMEM device per NUMA node, each 1 TB.

The benefit of dividing the vPMEM volume into segments and affinitizing them to NUMA boundaries enables applications to access data in physically aligned NUMA node memory ranges. Considering data is accessed sequentially, storing data NUMA optimized is best for throughput and access latency performance.

Affinitized vPMEM volumes is the only option that is supported by SAP HANA.
3.1.6 Affinity disabled

When the vPMEM device is activated without affinity, the hypervisor allocates a single vPMEM memory segment from the unused pool of system memory, and divides this single memory region over all NUMA nodes to which the LPAR is assigned. When this single vPMEM device is activated at the OS level, one \( /\text{dev/nmem} \) device is created, and all data that is copied to it is also divided over the NUMA nodes to which the LPARs are assigned. This information is less than desirable after all queries to the SAP HANA table data are retrieved across all NUMA nodes, slightly increasing latency of data access when compared to the affinitized option of the vPMEM volume creation.

![Diagram showing memory allocation across NUMA nodes](image)

Figure 3-6 shows an 8 TB system with memory that is allocated across four NUMA nodes, which creates a 4 TB non-affinitized vPMEM device that results in a single 4 TB device that is stripped across all NUMA nodes.

Currently, this vPMEM device option is not supported for SAP HANA.

3.1.7 Enabling vPMEM use in Linux

With Linux kernel 4.2, support for \libnvdimm was introduced, which provides access to persistent memory volumes. When persistent memory volumes are activated by the OS, raw memory devices are created in \( /\text{dev/nmem} \), with one for each active persistent memory device. For Power Systems vPMEM volumes that are created with the Affinity option, there is one vPMEM volume for each NUMA node that is assigned to the LPAR.

The persistent memory volumes are then initialized, enabled, and activated with the standard OS non-volatile DIMM control (\texttt{ndctl}) commands. These utilities are not provided by default in a base level Linux installation, but are included in the Linux distribution. Install them by running the application repository commands, for example, for Red Hat, install it by running the command that is shown in Example 3-8 on page 45.
Example 3-8  Red Hat command-line interface installation of the ndctl package

yum install ndctl

For SUSE Enterprise Linux Server, install it by running the command that is shown in Example 3-9.

Example 3-9  SUSE Enterprise Linux Server command-line interface installation of the ndctl package

zypper install ndctl

On Power Systems servers, each vPMEM volume is initialized and activated automatically. There is a corresponding number of /dev/nmem and /dev/pmem devices as there are NUMA nodes that are assigned to the LPAR, as shown in Example 3-10.

Example 3-10  Listing of raw persistent memory devices

```
# ls -l /dev/*[np]mem*
```

```
crw------- 1 root root  241,  0 Jan 10 17:32 /dev/nmem0
crw------- 1 root root  241,  1 Jan 10 17:32 /dev/nmem1
crw------- 1 root root  241,  2 Jan 10 17:32 /dev/nmem2
crw------- 1 root root  241,  3 Jan 10 17:32 /dev/nmem3
brw-rw---- 1 root disk 259,  0 Jan 10 17:38 /dev/pmem0
brw-rw---- 1 root disk 259,  1 Jan 10 17:38 /dev/pmem1
brw-rw---- 1 root disk 259,  2 Jan 10 17:38 /dev/pmem2
brw-rw---- 1 root disk 259,  3 Jan 10 17:38 /dev/pmem3
```

If the /dev/pmem devices are not created automatically by the system during the initial OS start, then they must be created. Example 3-11 shows the set of ndctl commands to initialize the raw /dev/nmem device, where X is the device number (for example, /dev/nmem0).

Example 3-11  Creating /dev/pmem devices

```
# ndctl disable-region regionX  # remove any previously defined regions
# ndctl zero-labels nmemX  # clear any previously defined devices
# ndctl init-labels nmemX  # initialize the new device
# ndctl enable-region regionX  # enable the device region
# ndctl create-namespace -r regionX  # create the region and /dev/pmem device
```

New device definitions are then created as /dev/pmemX. These are the new disk devices that must be formatted, as shown in Example 3-12.

Example 3-12  Creating the XFS file system on the vPMEM device

```
# mkfs -t xfs -b size=64k -s size=512 -f /dev/pmemX
```

When mounting the vPMEM volumes, use the /dev/disk/by-uuid identifier for the volumes. These values are stable regarding OS renaming of devices on restart of the OS. Also, these volumes must be mounted with the -o dax option, as shown in Example 3-13.

Example 3-13  Manual mounting of a vPMEM device to file system directory

```
# mount -o dax /dev/disk/by-uuid/34cb1120-1a61-47e5-9bcc-5b60e6d8e1d
/path/to/vPMEM/directory
```
To mount the volumes automatically on system restart, add an entry to the `/etc/fstab` file for each vPMEM volume by using the corresponding UUID name and adding the option `dax` in the options column. Using the UUID name of the volume guarantees correct remounting of the `/dev/pmemX` volume number after an OS restart. Example 3-14 shows an entry for the `fstab` file for one vPMEM volume.

**Example 3-14  Adding vPMEM devices into the `/etc/fstab` file for automatic mounting on OS start**

```
/dev/disk/by-uuid/34cb1120-1a61-47e5-9bcc-5b60e6d8e1d /hana/data/vPMEM/vPMEM0 xfs
defaults,dax 0 0
```

vPMEM volumes are not traditional block device volumes. Therefore, normal block device disk monitoring tools, for example, `iostat` and `nmon`, cannot monitor the I/O to the vPMEM devices. But, a normal director monitoring tool, for example, `du`, works because the files use the available storage space of the vPMEM volume.

### 3.1.8 Configuring SAP HANA to use vPMEM volumes

The SAP HANA instance configuration must be updated to use the new vPMEM volume. Update the `global.ini` file to add the file system directory paths to the `basepath_persistent_memory_volumes` parameter in the `[persistence]` section, with each directory separated by a semi-colon, as shown in Example 3-15.

**Example 3-15  Entry in the `global.ini` file defining the paths to the vPMEM volume directories**

```
[persistence]
basepath_persistent_memory_volumes =
/path/to/first/directory;/path/to/second/directory;...
```

This parameter option is an offline change only, which requires the restart of SAP HANA to enable it.

On first startup, SAP HANA by default copies all column store table data, or as much as possible, from persistent disk into the newly configured and defined vPMEM volumes. With partitioned column store tables, SAP HANA assigns in a round-robin fashion, partitions to vPMEM volumes to distribute evenly the column store table partitions across the entire vPMEM memory NUMA assignment. When all column store data for a table is loaded into the vPMEM volumes, SAP HANA maintains a small amount of column store table metadata in normal LPAR DRAM memory.

### 3.1.9 Specifying tables, columns, or partitions to use vPMEM volumes

SAP HANA can also be configured to populate the vPMEM volumes with specifically defined column store tables, columns, and partitions. First, change the `indexserver.ini` parameter to turn off the loading of all column store tables into the persistent memory, as shown in Example 3-16.

**Example 3-16  Changing the default behavior of SAP HANA to not load all tables on SAP HANA startup**

```
[persistent memory]
table_default = OFF
```
The `table_default` parameter default value is `DEFAULT`, which is synonymous with the value of `ON`. This parameter, along with the `global.ini` parameter `basepath_persistent_memory_volumes`, make the SAP HANA default behavior to load all table data in to the vPMEM devices.

The `table_default` parameter is dynamic. If there is column store table data in the vPMEM volumes, then performing an SAP HANA UNLOAD of the unneeded columnar table data removes the table data from the persistent device, after which a LOAD operation is needed to reload the table data into system DRAM, or a shutdown of SAP HANA can be done so that the old column store table data can be removed from the vPMEM volumes. Then, on the next startup, SAP HANA loads all column store table data into DRAM.

Individual column store tables that are to use the vPMEM volumes can be moved by running the SQL command `ALTER TABLE`, as shown in Example 3-17.

**Example 3-17   Altering a table move the column store table to PMEM**

```
ALTER TABLE "<schema_name>.<table_name>" PERSISTENT MEMORY ON IMMEDIATE CASCADE;
```

This command immediately moves the table from LPAR DRAM to the vPMEM devices, and it persists for all future restarts of SAP HANA.

For columns and partitions, the only way to load data into vPMEM volumes is by running one of the `CREATE COLUMN TABLE` commands, as shown in Example 3-18.

**Example 3-18   Specifying column store table columns or partitions in vPMEM**

```
# create a table that the named column will use persistent memory
CREATE COLUMN TABLE ... <column> ... PERSISTENT MEMORY ON

# create a table that the named partitions will use persistent memory
CREATE COLUMN TABLE ... PARTITION .. PERSISTENT MEMORY ON
```

Column store table data can also be unloaded from the vPMEM devices and the data removed from the vPMEM volumes. Example 3-19 shows the commands that remove the column store table data from the vPMEM volume and unload the data from all memory.

**Example 3-19   Altering a column store table to stop using vPMEM**

```
ALTER TABLE <table_name> PERSISTENT MEMORY OFF IMMEDIATE UNLOAD <table_name> DELETE PERSISTENT MEMORY
```

After this command runs, as shown in Example 3-19, the column store table data is no longer loaded into any memory areas (DRAM or vPMEM). Table data must be reloaded either by future query processing or by manually running the SQL command that is shown in Example 3-20.

**Example 3-20   Reloading table data from disk-based persistent storage**

```
LOAD "<table_name>" ALL
```

If the table persistent memory setting is changed to either `OFF` or `ON`, it can be reset to `DEFAULT` by running the SQL `ALTER TABLE` command that is shown in Example 3-21.

**Example 3-21   Changing the persistent memory setting**

```
ALTER TABLE "<table_name>" PERSISTENT MEMORY DEFAULT
```
3.1.10 Undersized vPMEM volumes

If the vPMEM volumes are too small to hold all of the column store table data, SAP HANA loads the remaining column store tabled data into LPAR DRAM. Queries still process normally, accessing data either from DRAM or from the vPMEM volumes. On an SAP HANA restart, SAP HANA loads the non-vPMEM stored column table data into LPAR DRAM from persistent disk.

3.1.11 Disabling the usage of vPMEM volumes

If it is necessary to stop using the vPMEM volumes, shut down SAP HANA and comment out the parameter `basepath_persistent_memory_volumes` and the vPMEM volumes in the `global.ini` file. On restart, SAP HANA loads all tables into the LPAR DRAM. The contents of the vPMEM volumes can then be cleaned out by either reformatting them or deleting the contents from the directories. The vPMEM volumes can then be redeployed.

3.1.12 Benefits of using vPMEM volumes on Power Systems servers

Here are the benefits of using vPMEM volumes:

- Faster restart of SAP HANA because column store table data is persistently loaded into vPMEM volumes.
- No special hardware is required to create and use vPMEM volumes for SAP HANA. The vPMEM volumes are created out of spare system memory.
- Access to data on vPMEM volumes is at RAM speeds, with no loss in performance compared to accessing table data in the traditional LPAR DRAM configuration.
- Multiple LPARs on a system can be assigned vPMEM volumes so that multiple applications on a system can use vPMEM and its benefits. Intel Optane is not supported in virtualized environments, so only a single application per system is supported.
- SAP HANA attempts to align full columnar table data to a single vPMEM device, which aligns the table data to the memory of a single NUMA and speeds up access to that data.
- Partitioned table data is assigned in a round-robin fashion to NUMA nodes, keeping all data within the partition range localized to a single NUMA node.
- After SAP HANA shuts down, deleting an existing vPMEM device does not affect any data in the DB because SAP HANA synchronizes and saves new data to the persistent disk layer on shutdown.
- The maximum vPMEM segment per NUMA node is 4 TB.
- The vPMEM configured SAP HANA instances are supported and transparent to SAP HANA System Replication (HSR) / Distributed Statistics Records (DSR) configurations.

3.1.13 vPMEM usage considerations

Here are some vPMEM considerations:

- vPMEM device creation or deletion can be done only when the LPAR is inactive.
- vPMEM volumes are static in capacity. After they are created and assigned to an LPAR, the volume size cannot be changed.
- If a volume of a different size is required, the existing volume must be deleted, and a new volume of a different size can be created.
When you create a vPMEM volume, the new /dev/nmem devices must be reinitialized, and the resulting /dev/pmem devices must be formatted and mounted before SAP HANA starts. Then, SAP HANA repopulates the vPMEM volumes with columnar data during startup.

A defined vPMEM volume is assigned to a single LPAR and is not reassignable, which helps secure the data that is stored in the vPMEM volumes for the application running on the assigned LPAR.

If the NUMA configuration for an LPAR that is assigned a vPMEM volume changes (for example, changes the number of cores), then the previous vPMEM volume NUMA assignment does not match that of the new LPAR NUMA resource allocation. To maintain NUMA affinity, the vPMEM volume must be deleted and re-created to match the new NUMA alignment of the LPAR.

vPMEM volumes at initial introduction are not Live Partition Mobility (LPM) capable. This feature is intended to be enabled as the technology matures.

All /dev/pmem devices must be used directly. Partitioning these devices or creating LVM volumes is not supported because SAP HANA queries the device tree to understand the NUMA alignment of the devices. Any volumes that are created from the pmem device do not return valid NUMA alignment details, which cause SAP HANA to not start.

SAP supports only one vPMEM device per NUMA node per SAP HANA instance.

Multiple tenants of an SAP HANA instance can share vPMEM volumes. Each tenant must have a separate directory structure for the vPMEM volumes to be mounted.

Only SAP HANA column store data is stored in vPMEM volumes. Row store tables and LOB column store tables are stored in DRAM memory.

Log volume data is still stored on persistent disk for the protection of transactional data.

Any changes to column store table data that is accessed from vPMEM volumes are protected by persistent disk through SAP HANA's save point synchronization mechanism.

### 3.1.14 PowerVM vPMEM enablement dependencies

To use vPMEM on Power Systems, there are dependencies on several components of the system, as shown in Table 3-1.

<table>
<thead>
<tr>
<th>Component</th>
<th>Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>Power Systems Firmware</td>
<td>Version 9.40 or later</td>
</tr>
<tr>
<td>HMC</td>
<td>Version 9.1.940.0 or later</td>
</tr>
<tr>
<td>OSs</td>
<td>SUSE Enterprise Linux Server 15 SP1 or later, or Red Hat 8.2 or later</td>
</tr>
<tr>
<td>SAP HANA</td>
<td>Version 2.0 SPS4 rev44 or later</td>
</tr>
</tbody>
</table>

### 3.1.15 vPMEM-related SAP Notes

Here are some vPMEM-related SAP Notes:

- SAP Note 2700084
- SAP Note 2618154
- SAP Note 2786237
- SAP Note 2813454
3.1.16 Additional documentation that is related to vPMEM and persistent memory usage with SAP HANA

This section provides additional documentation for vPMEM and persistent memory usage.

- IBM product documentation
  Managing persistent memory volumes
- SAP product documentation
  “Persistent Memory” in SAP HANA Administration Guide for SAP HANA Platform

3.2 RAM disk (temporary file systems): SAP HANA Fast Restart Option

Temporary file systems (tmpfs) volumes are emulated file systems that are created from the available memory that is allocated to the OS of an LPAR. All modern OSs can create and use memory-based tmpfs volumes. SAP HANA can use tmpfs volumes to store columnar table data in the same way it uses vPMEM volumes: The tmpfs memory volumes are mounted on the OSs file system directory structure for storage of table data for fast access. SAP refers to the tmpfs volumes solution as the SAP HANA Fast Restart Option.

As with vPMEM persistent memory volumes, SAP HANA can use tmpfs volumes to store columnar table data in configured LPAR DRAM volumes. The mechanisms by which SAP HANA is configured to use tmpfs volumes are identical to vPMEM volume usage.

Like vPMEM volumes, access to the files that are stored in a tmpfs file system are performed at DRAM speeds. In this regard, accessing data from tmpfs and vPMEM volumes has the same performance. Also, no special hardware is needed to create tmpfs volumes because the DRAM memory that is allocated to the LPAR is used.

Unlike vPMEM, because the memory that is allocated to tmpfs volumes is allocated within the memory that is allocated to an LPAR, tmpfs volumes are not persistent through OS or LPAR restarts. Data that is stored in tmpfs LPAR DRAM disks is volatile, and the contents are erased when the OS is shut down or restarted. Upon restart of the OS, the RAM disk volumes must be re-created, and SAP HANA columnar table data is reloaded from persistent disk volumes.

Also, unlike vPMEM volumes that are created with the Affinity option, creating a tmpfs volume is not automatically aligned to any specific NUMA node. NUMA node memory alignment details are gathered as a preparatory step and used in creating the tmpfs volumes at the OS level.

One benefit of tmpfs file systems have over vPMEM volumes is that tmpfs volumes can be created to grow dynamically as the tmpf volumes fills, so the volumes can accommodate larger than expected data growth. But, this dynamic characteristic of tmpfs file systems has the side effect of using more LPAR DRAM memory than expected, which steals memory away from applications that need that memory to function. Hence, correctly sizing the tmpfs volumes is still important. Alternatively, the tmpfs volumes can be created to use a fixed amount of LPAR DRAM.
3.2.1 Configuring tmpfs volumes

Creating tmpfs file systems involves using LPAR DRAM memory to create volumes to store application data. Because memory is typically non-uniformly distributed across NUMA nodes, you should understand the memory architecture of the system before allocating memory to tmpfs file systems to ensure that the memory allocation of the tmpfs file systems can be allocated within NUMA nodes for fastest access.

A quick check at the OS shows the available NUMA nodes and how much memory is allocated to each node, as shown in Example 3-22.

**Example 3-22  Determining the amount of RAM that is allocated to each NUMA node of an LPAR**

```bash
grep MemTotal /sys/devices/system/node/node*/meminfo
```

The command in Example 3-22 produces an output that shows the amount of memory that is available on each of the NUMA nodes that the OS assigned, as shown in Example 3-23.

**Example 3-23  Example output from the grep MemTotal command**

<table>
<thead>
<tr>
<th>Path</th>
<th>MemTotal</th>
</tr>
</thead>
<tbody>
<tr>
<td>/sys/devices/system/node/node0/meminfo</td>
<td>510299264 kB</td>
</tr>
<tr>
<td>/sys/devices/system/node/node1/meminfo</td>
<td>510931712 kB</td>
</tr>
<tr>
<td>/sys/devices/system/node/node2/meminfo</td>
<td>511427328 kB</td>
</tr>
<tr>
<td>/sys/devices/system/node/node3/meminfo</td>
<td>511452992 kB</td>
</tr>
</tbody>
</table>

In this output (Example 3-23), the system has four NUMA nodes, each installed with roughly 512 GB of system DRAM.

To create tmpfs devices in Example 3-23, you want to allocate four different tmpfs devices, with one for each NUMA node. The `mount` command has an option that can assign the memory for the tmpfs to a named NUMA node. Example 3-23 shows four directories to where the tmpfs files systems will be mounted. Example 3-24 shows how to create the file systems by using the `mount` command options.

**Example 3-24  Mounting the file systems**

```bash
mount <tmpfs file system name> -t tmpfs -o mpol=prefer:X /<directory to mount file system>
```

In Example 3-24:

- `<tmpfs file system name>`: The operating device name. Use any descriptive name.
- `-t tmpfs`: The file system type, and in this case, of type tmpfs.
- `-o mpol=prefer:X`: The NUMA node number to assign the memory for the tmpfs.
- `/<directory to mount file system>`: The location on the OS file system path to mount the tmpfs file system. This directory is accessible and readable from the OS level. Check that this directory exists as for any normal `mount` command.
In Example 3-23 on page 51, the system has four NUMA nodes, so creating four directories and mounting four different tmpfs file systems can be done, as shown in Example 3-25, by substituting the `<SID>` with the SID of the SAP HANA DB.

**Example 3-25**  Sample script to create, mount, and make available the tmpfs volumes for use by SAP HANA

```bash
for i in 0 1 2 3; do
    mkdir -p /hana/data/<SID>/tmpfs$i
    mount tmpfs_<SID>_${i} -t tmpfs -o mpol=prefer:${i} /hana/data/<SID>/tmpfs$i
done
chown -R <db admin user>:<db group> /hana/data/<SID>/tmpfs*
```

Using these options (Example 3-25), the amount of memory that is allocated to each tmpfs is dynamically sized based on what is being stored by SAP HANA in the file systems. This option is the preferred one because the file system grows as SAP HANA table data is migrated from RAM to the tmpfs file system.

If you must statically allocate an amount of memory to the tmpfs file system, use the `-o size=<size in GB>` option to allocate statically a fixed size of LPAR DRAM to the tmpfs file systems.

**Note:** The tmpfs volumes are not formatted as an XFS file system like for vPMEM volumes, and are the volumes are not mounted by using the `-o dax` option because of the differences in file system format for tmpfs volumes and the ability of SAP HANA to differentiate and support both types of file system formats to persistently store columnar table data.

### 3.2.2 Configuring SAP HANA to use the tmpfs volumes

Configuring SAP HANA to use tmpfs volumes is the same as with vPMEM memory:

1. Edit the `global.ini` file to specify the paths to the directories where the tmpfs volumes are mounted.
2. Verify that the `indexserver.ini` value for `table_default` is set to `DEFAULT` or `ON`.
3. Restart SAP HANA to populate the tmpfs volumes with column store table data.

### 3.2.3 Comparing vPMEM and tmpfs

Using tmpfs for holding column store table data has many of the same benefits as using vPMEM volumes. Table 3-2 outlines some of the benefits of using persistent memory solutions on Power Systems servers.

<table>
<thead>
<tr>
<th>Benefit</th>
<th>vPMEM</th>
<th>tmpfs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Faster restart of SAP HANA because column store table data is persistently loaded into mounted memory volumes.</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Memory volumes persist across OS and LPAR restarts.</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>

Table 3-2 Benefits of using persistent memory solutions
### Benefit

<table>
<thead>
<tr>
<th>Benefit</th>
<th>vPMEM</th>
<th>tmpfs</th>
</tr>
</thead>
<tbody>
<tr>
<td>No special hardware is required for SAP HANA to use mounted memory volumes for columnar table data.</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Access to data on mounted memory volumes is at RAM speeds with no loss in performance when compared to accessing table data in the traditional DRAM configuration.</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Table data that is designated as non-preload on startup is accessed at DRAM speeds, which increases query and application performance.</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Multiple LPARs on a system can be assigned memory volumes.</td>
<td>Yes, and volume DRAM taken from spare system memory</td>
<td>Yes, tmpfs volume created within each LPAR memory</td>
</tr>
<tr>
<td>Automatic detection and alignment of memory volume to NUMA node.</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>SAP HANA attempts to align full columnar table data to a single memory volume.</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Partitioned columnar table data is assigned in a round-robin fashion to memory volume.</td>
<td>Yes, with Affinity enabled</td>
<td>Yes, when each file system aligned to NUMA nodes in the <code>mount</code> command</td>
</tr>
<tr>
<td>After SAP HANA shutdown, deleting an existing vPMEM device does not affect any data of the DB because SAP HANA synchronizes and saves new data to the persistent disk layer on shutdown.</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Memory volume is Live Partition Migration (LPM) capable.</td>
<td>Not yet, but coming in a future release</td>
<td>Yes</td>
</tr>
<tr>
<td>Dependency on system firmware, HMC code, or OS release.</td>
<td>Yes, as outlined in the previous section</td>
<td>Can be used with any POWER9 supported versions.</td>
</tr>
<tr>
<td>SAP HANA Release Requirements.</td>
<td>&gt;= 2.00.035</td>
<td>&gt;=2.00.040</td>
</tr>
</tbody>
</table>

### 3.2.4 SAP-related documentation

This section provides more related documentation:
- “Live Partition Migration” in *SAP HANA Administration Guide for SAP HANA Platform*
- SAP Note 270084
3.3 Persistent disk storage by using native Non-Volatile Memory Express devices

Non-Volatile Memory Express (NVMe) adapters, with their high-speed flash memory, have become a popular technology to store data that needs low latency and high throughput. The NVMe architecture and its access protocols have become an industry standard, making it easy to deploy through a device driver addition to the OS. When you use NVMe in this context, you use the flash modules to store the DB persistently as though they are disk-based storage. SAP HANA reads data from these devices on start as they do from regular storage area network (SAN) -based storage. The benefit is that the read operations from NVMe are faster than SAP disk-based storage, providing for faster SAP HANA startup and for persistence of data across SAP HANA, OS, and system restarts.

Here are some of the key benefits NVMe devices provide over SAN disk-based storage devices:

- Increased queue depth, which provides decreased I/O times.
- Lower latencies for read and write operations.
- Higher I/O throughput than traditional disk fabric systems (for example, Fibre Channel) because of the location of the adapters on a PCIe adapter slot.

3.3.1 NVMe device details

NVMe devices are presented to the OS through the traditional block device interface. For Linux, these devices are listed in the /dev directory along other disk-based block devices. Therefore, NVMe devices can be inserted into an LPAR as regular storage devices.

NVME adapters are made up of multiple individual flash memory modules. This architecture allows the OS to access multiple storage modules per NVMe adapter independently. Figure 3-7 shows a sample output listing the devices on an individual NVMe adapter.

![Figure 3-7 NVMe adapter module listing](image)

Figure 3-7 shows that there are four modules with 745 GB per module, and 3 TB total storage for the adapter.

3.3.2 Configuration and usage considerations

The flash NVMe adapters that are assigned to the LPAR may be used in several different ways:

- Directly used as an independent storage device.
- RAID arrays can be created to span multiple modules on an individual NVMe adapter.
- RAID arrays can span modules across other NVMe adapters.
Figure 3-8, Figure 3-9, and Figure 3-10 show a few examples.

Because flash modules are subject to long-term degradation in high-write and data-change environments, be careful about which data is assigned to flash storage volumes. Preference must be given to data profiles that do not have excessive data modification rates. For environments that must use flash modules with high-write and change activity characteristics, mirroring a NVMe volume with a traditional disk-based storage volume can help preserve data integrity in the case of a NVMe flash wear failure.
3.3.3 NVMe performance characteristics

Tests that are conducted by using SAP HANA and high I/O workloads have shown that using NVMe adapters for SAP HANA volumes can decrease latency by up to 9x over traditional storage technologies (for example, IBM SAN Volume Controller that uses 20 GB cache and IBM XIV® storage).

For read performance, testing shows that NVMe adapter volumes are up to 2 - 4x faster than disk-based storage solutions, depending on I/O block sizes, as shown in Table 3-3.

<table>
<thead>
<tr>
<th>Block size</th>
<th>64 KB</th>
<th>128 KB</th>
<th>256 KB</th>
<th>1 MB</th>
<th>16 MB</th>
<th>64 MB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Performance increase by using NVMe</td>
<td>2.2x</td>
<td>2.7x</td>
<td>4.0x</td>
<td>4.5x</td>
<td>4.0x</td>
<td>4.1x</td>
</tr>
</tbody>
</table>

Table 3-4  VVMe write performance characteristics

<table>
<thead>
<tr>
<th>Block size</th>
<th>4 KB</th>
<th>16 KB</th>
<th>64 KB</th>
<th>128 KB</th>
<th>256 KB</th>
<th>1 MB</th>
<th>16 MB</th>
<th>64 MB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Performance increase initial write while using NVMe</td>
<td>N/A</td>
<td>1.1x</td>
<td>1.4x</td>
<td>1.7x</td>
<td>1.9x</td>
<td>1.8x</td>
<td>1.7x</td>
<td>1.7x</td>
</tr>
<tr>
<td>Performance increase overwrite while using NVMe</td>
<td>1.5x</td>
<td>1.8x</td>
<td>2.0x</td>
<td>1.8x</td>
<td>1.7x</td>
<td>1.9x</td>
<td>1.6x</td>
<td>1.6x</td>
</tr>
</tbody>
</table>

3.3.4 Striping effects for internal NVMe cards

Creating a RAID 0 volume over multiple NVMe adapters increases the throughput on read operations on block sizes greater than or equal to 256 KB nearly by a factor of 2.

Creating a RAID 0 volume over multiple NVMe devices increases the throughput on write operations on block sizes greater than or equal to 64 KB nearly by a factor of 1.7. On block sizes greater than 256 KB, the factor is nearly 2.

Creating a RAID 0 volume on the multiple memory modules of one NVMe device has no positive effect on performance over storage on a single memory module.

3.3.5 Summary

NVMe adapters with their fast flash modules can provide an increase in I/O performance for SAP HANA compared to storage on traditional SAN storage options, which can be either solid-state drives (SSDs) or hard disk drives (HDDs). The increase in speed and lower latencies is provided by much faster flash technology in the NVMe adapter versus SSD flash storage and by accessing the NVMe storage through PCIe 3.0 protocol versus over Fibre Channel protocols.
3.4 NVMe Rapid Cold Start Mirror

Another option for NVMe adapter flash storage is to mirror the volumes to external SAN disk-based persistent storage for high availability (HA), and faster start of large SAP HANA DBs.

A large DB takes some time to load from SAN disk-based storage through connectivity solutions like Fibre Channel that are connected to HDD-based or SSD-based volumes. NVMe storage that is hosted at the host level provides faster access to the SAP HANA data for loading and writing.

But flash modules are subject to wear due to data writes or changes. To protect valuable data, a hybrid of the traditional SAN disk-based storage and NVMe storage can be used to provide faster SAP HANA startup along with protection of the data on disk-based storage.

3.4.1 Hybrid mirrored volume

The solution that is shown in Figure 3-11 creates RAID 0 arrays from the NVMe FLASH modules, and a RAID 1 mirror to the storage that is created on the SAN disk-based storage solution.

In Figure 3-11, the NVMe adapter flash modules (inside the green box) are all added into a RAID 0 array. On the SAN side, RAID 0 arrays of the same size as the NVMe RAID 0 array are created (inside the red box). Each of these four RAID 0 volumes are represented with blue boxes on their respective storage platforms.

Then, a mirrored RAID 1 volume is created by assigning one SAN RAID volume to one NVMe RAID 0 volume, represented by each of the gray boxes.
When creating the RAID 1 mirror between the NVMe volumes and the SAN storage volumes, you can set the preference for the OS to write to the NVMe volumes by passing the option **--write-mostly** of the `mdadmin` array utility. In this case, the preference is to assign it to the RAID 0 device name for the external SAN volume. Example 3-26 shows the Linux man page excerpt for `mdadm`.

**Example 3-26** The `mdadm --write-mostly` option to prefer one RAID device for reading

```bash
-W, --write-mostly
subsequent devices listed in a --build, --create, or --add command will be
flagged as 'write-mostly'. This is valid for RAID 1 only and means that the
'md' driver will avoid reading from these devices if at all possible. This
can be useful if mirroring over a slow link.
```

In this case, for the RAID 1 device, specify the SAN storage device as **--write-mostly** in the `mdadm --create` command, as shown in Example 3-27.

**Example 3-27** Specifying the SAN storage device

```bash
mdadm --create --verbose --/dev/<new RAID device name> --level=1 --raid-devices=2
/dev/<NVMe device name --write-mostly /dev/<SAN storage device name>
```

### 3.4.2 Summary

Mirroring an internally installed NVMe adapter to an external SAN volume of the same size can provide the benefits of a rapid startup of SAP HANA by reading data from the NVMe adapter with the RAID 1 mirroring of the data to an external SAN disk for ultimate data protection.

### 3.5 Comparing vPMEM to Intel Optane

The Optane persistent memory solutions that are available on the Intel platform constrict technology to the persistent memory solutions vPMEM and tmpfs for fast restart. The Optane memory Data Center Persistent Memory Module (DCPMM) solution is based on a new memory DIMM architecture that allows for persistent storage of data that is stored to the DCPMMs, which allows that data to survive OS and system restarts. vPMEM and tmpfs for fast restart also provide resiliency of the data from an SAP HANA restart. In addition, for vPMEM for an LPAR restart, data that is stored in these solutions must be restored in the event of a full system restart.

Optane DCPMM memory is implement by installing new Optane DCPMM memory cards into existing system DRAM DIMM slots. Real DRAM capacity must be sacrificed to use the Optane memory solution. In contrast, the vPMEM option for Power Systems serves uses standard system DRAM that is already installed in the system. Implementation of vPMEM is as simple as defining the vPMEM memory segment, and starting the LPAR, configuring SAP HANA to use the persistent memory segments. No hardware downtime is necessary unless more DRAM is required to support the vPMEM solution.

Optane memory capacities are provided in 128 GB, 256 GB, and 512 GB sizes per PCDMM module, which is much higher capacities than what exists for DIMM memory modules, which have a maximum capacity of 64 GB per DIMM.
Rules for using PCDMM modules are complicated and vary depending on the memory mode that is used, but with a maximum of 12 DIMM modules per socket using six DIMM modules of 64 GB and the maximum of six 512 GB DCPMM memory modules, a socket maximum memory configuration is 3.4 TB. This is compared to a maximum memory configuration of 1.5 TB when using only DIMM memory modules. POWER9 systems support up to 4 TB of DIMM system memory per socket by using 128 GB DIMMs. Future DIMM sizes will increase this memory footprint.

From a memory latency point of view, Optane DCPMM memory modules have a higher read and write latency compared to standard DIMM memory technologies because of the technology that is implemented to provide data persistency in the DCPMM module. Higher latencies can have an impact on application performance and must be evaluated when implementing the Optane solution.

POWER9 vPMEM and tmpfs use DIMM-backed RAM and perform read and write operations at full DIMM throughput capabilities.

Optane has three memory modes that the DCPMM modules can use:

- **Memory Mode**: In this mode, the DCPMM memory modules are installed along standard DIMM modules and are used as a regular memory device. One advantage of using the DCPMM modules in this mode is that greater overall memory capacities can be achieved over standard DIMM modules that are available for x86 systems. But, enabling the PCDMM modules in this mode puts the regular DIMMs in the system into a caching function, which makes their capacity invisible to the host OS. Therefore, only the capacity of the DCPMM memory can be used by the host OS, and the regular DIMM memory capacity is unavailable for OS and application use.

- **App Direct Mode**: In this mode, the DCPMM memory modules are used as persistent storage for OSs and applications that can take advantage of this technology. The DCPMM memory modules are recognized by the OS as storage devices and are used to store copies of persistent disk data, making access to that data faster after an OS or system restart. The standard DIMMs are used normally as available RAM to the OS.

- **Mixed Mode**: This mode is a mixture of the Memory and App Direct modes. When using DCPMM modules in this mode, a portion of the capacity of the module is used as memory for the host OS, and the remaining capacity is used for persistent storage. But, as in Memory Mode, any DIMM memory is unavailable for use by the host OS and is instead converted into a memory cache subsystem.

Currently, Optane persistent memory is not supported for use by SAP HANA in virtualized environments.

### 3.6 Impact to Live Partition Mobility capabilities

PowerVM LPM technology allows for the migration of LPARs from one system to another, either in offline mode or live. This technology allows LPARs to be moved to another system to avoid downtime to the running applications.

Using tmpfs for persistent memory uses the memory that is assigned to the LPAR and available to the LPAR. Therefore, moving an LPAR from one system to another preserves the use of the tmpfs persistent memory volumes at the destination system.
vPMEM volumes that are assigned to an LPAR are defined outside the LPAR configuration. Due to this implementation, LPM operations are not supported for vPMEM-enabled LPARs. Before the LPAR migration, the vPMEM device must be removed from the LPAR. Then, at the destination system, a new vPMEM volume can be created to support the application. vPMEM LPM is intended to be supported in a future firmware release.
Related publications

The publications that are listed in this section are considered suitable for a more detailed description of the topics that are covered in this paper.

IBM Redbooks

The following IBM Redbooks publications provide additional information about the topics in this document. Some publications that are referenced in this list might be available in softcopy only.

- *IBM Power Systems Security for SAP Applications*, REDP-5578
- *IBM Power Systems Virtualization Operation Management for SAP Applications*, REDP-5579
- *SAP HANA on IBM Power Systems: High Availability and Disaster Recovery Implementation Updates*, SG24-8432
- *SAP Landscape Management 3.0 and IBM Power Systems Servers*, REDP-5568

You can search for, view, download, or order these documents and other Redbooks, Redpapers, web docs, drafts, and additional materials, at the following website:

ibm.com/redbooks

Online resources

These websites are also relevant as further information sources:

- Guide Finder for SAP NetWeaver and ABAP Platform
  https://help.sap.com/viewer/nwguidefinder
- IBM Power Systems rapid cold start for SAP HANA
  https://www.ibm.com/downloads/cas/WQDZWBYJ
- SAP Support Portal
- Software Logistics Tools
- Welcome to the SAP Help Portal
  https://help.sap.com
Help from IBM

IBM Support and downloads
ibm.com/support

IBM Global Services
ibm.com/services