April 2020

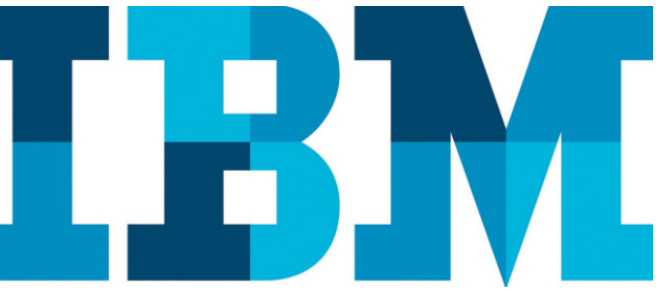
IBM Power Systems for SAS Viya 3.5 deployment guide

Options and best practices for deploying SAS Viya 3.5 on IBM POWER9 servers

Version 2.0

Table of contents

Planning your deployment 2
POWER9 deployment options 2
Network configuration options 8
POWER9 hardware minimum configuration options 16
Storage considerations 21
Sizing considerations 22
SAS Viya installation on POWER9 23
Tuning best practices 24
Summary 27
Get more information 27
About the authors 27
Revision history 28



SAS Viya is a complex set of artificial intelligence (AI) and analytics solutions that require a properly planned infrastructure to meet the needs of the data scientists, business analysts, and application developers who will use Viya capabilities in their daily work activities. Regardless of the user role, the underlying infrastructure matters to ensure performance expectations and service level agreement (SLA) requirements are met or exceeded. Although the general planning process is similar for deploying SAS Viya on any platform, key IBM® POWER9™ differentiators should be considered in order to ensure that an optimized infrastructure deployment is achieved. This guide provides useful information needed during the planning, sizing, ordering, installing, configuring, and tuning phases of your SAS Viya deployment on POWER9 processor-based servers.

Planning your deployment

The primary step in planning your deployment is to understand your end users' requirements. It is important to know which SAS Viya capabilities will be used, who will use them, and what data will be used. These are the fundamentals to determine the optimal infrastructure environment to support the users. Another important aspect to factor in the decision is whether the SAS Viya environment needs to integrate with an existing SAS 9.4 environment.

In addition, this information along with the users' performance expectations help in determining the size and configuration of each server, and how many servers are needed for an initial SAS Viya deployment.

SAS provides a deployment guide for SAS Viya on Linux®. This guide should be read and used during the planning phase to understand the Viya deployment modes and other software and environmental requirements. You can find the latest guide at: https://bit.ly/2UAIkP1

Beyond this, IBM Power Systems™ offer several additional flexible deployment options for SAS Viya. The next section outlines these options. While the deployment options provide examples as to the server options for the SAS Viya workload, clients should engage their SAS Account Team for assistance in sizing their workload from the SAS Enterprise Excellence Center (EEC) sizing team. This ensures that the hardware environment is properly sized for the client's specific SAS workload.

POWER9 deployment options

The IBM Power System portfolio of servers enables flexible deployment options for running SAS Viya. The IBM portfolio offers ultra-flexible systems with the highest reliability[1]. In addition, the Enterprise and scale-out servers can run Linux in addition to IBM AIX® and IBM i and allow for the consolidation of multiple workloads on a single system which can reduce data center footprint. An example of this is deploying multiple SAS workloads on the same system. Customers can also take advantage of advanced virtualization features, such as capacity on demand (CoD), to dynamically activate processor capacity and memory resources for ease of scalability and flexibility in adapting to growing demands on your infrastructure.

[1] ITIC Global Server Hardware, Server OS Reliability Survey

SAS Viya can run in two modes, symmetric multiprocessing (SMP) and massively parallel processing (MPP). The SMP mode uses multiple cores or processors on just one node or server. SMP deployment options are appropriate when doing a proof-of-concept (POC) or when the Viya workload requirements do not initially require a cluster. SMP mode deployments are also great options for programming-only environments and application development. There are two SMP mode deployment options for IBM Power Systems.
The MPP mode uses a cluster of servers or partitions to run SAS Viya. In the MPP mode, SAS Cloud Analytics Services (CAS) Workers will offload the analytics and spread the data out to allow parallel loading of data and processing on multiple worker nodes. MPP offers a growth path and maximizes scaling which you will see in the next section.

SMP mode deployment options

Figures 1 and 2 show the simplest options for deploying SAS Viya on IBM Power. These are SMP mode deployments. There are two variations that use different types of Power Systems.

Figure 1 shows a scale-out single-server deployment where SAS Viya and all its complementary software are installed on one POWER9 server. An IBM Power System AC922 server model is shown as an example. The Power AC922 server is a powerful system built for analytics and can run CPU and GPU workloads. This server supports NVIDIA GPUs plus an NVIDIA technology called NVLink 2 which is unique to IBM POWER9 processor-based servers and is embedded within the NVIDIA GPUs. NVLink 2 provides 5.6 times data bandwidth acceleration when software switches between CPU and GPU.

Figure 2 shows a scale-up single-partition deployment, where SAS Viya including CAS and all its complementary software are installed on one POWER9 partition. In the POWER9 processor-based server family there are several scale-up server models to choose from, the IBM Power System E950 being one of them. An IBM Power System E980 server could also be used. These enterprise-class servers are large; you can use IBM PowerVM® virtualization to create a separate Linux partition, or a logical partition (LPAR), for installing SAS Viya. This SMP deployment option does not require the entire large server, so it is ideal for customers who currently run SAS 9.4 in an IBM AIX partition where they can use untapped resources (such as cores and memory) for deploying SAS Viya.

Example: IBM Power System AC922

Figure 1. SMP mode: Scale-out single-server

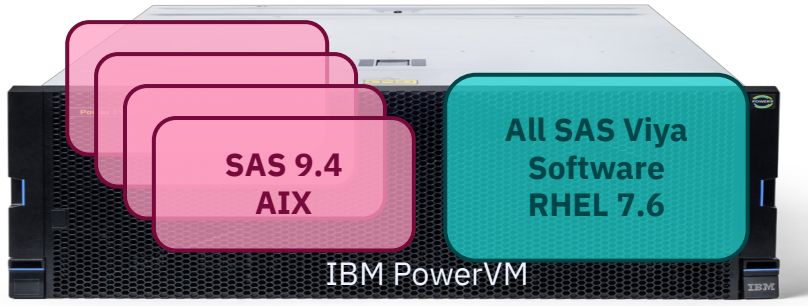


Figure 2. SMP mode: Scale-up single-partition

MPP mode deployment options

In MPP mode, CAS supports workers to help offload the analytics and spread the data out to allow for parallel processing on multiple workers. There are three MPP options.

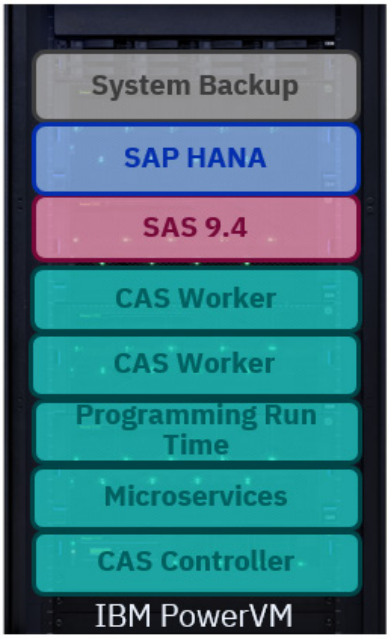
Figure 3 shows a scale-out multi-server deployment, and is where SAS Viya, including the CAS controller and CAS workers, and all its complementary software are installed across a cluster of POWER9 scale-out servers. This example shows a cluster of Power AC922 servers. As growth is experienced, you can scale out by adding more servers to the cluster.

Figure 4 shows another MPP mode deployment, where the analytics and data are spread out to allow for parallel processing on the multiple worker nodes. This example is for a scale-up multi-partition deployment, where all the same SAS Viya software is installed across a virtual cluster of POWER9 partitions. In this example, you can see the set of PowerVM partitions on a Power E980 server for running the Viya software in green. These partitions can be added to an existing server where other workloads have already been running. Some workload examples may be SAS 9.4 on an AIX partition or SAP HANA in a Linux partition, or a partition that runs system backups. These new Viya partitions are separate and coexist with the other partitions. The advantage of this scale-up environment is that you can grow without necessarily having to order more hardware. You can activate unused cores and memory, and either add more resources to an existing LPAR or add more LPARs, such as additional worker nodes. You also benefit from a simplified network configuration between the nodes because all nodes are on the same server. This scale-up MPP deployment option would be applicable for CPU-based workloads.

Figure 3. MPP mode: Scale-out multi-server

Figure 4. MPP mode: Scale-up multi-partition

Figure 5 shows an MPP SAS Viya deployment in a mixed environment. SAS Viya is installed across a set of partitions on a scale-up server and one or more scale-out servers. In this case, the CAS controller and management type services would run in the partitions on the left, and CAS worker nodes could run either on partitions (for non-GPU workloads) or on scale-out servers (for CPU or GPU workloads). Again, you can see the SAS Viya components in green and non-Viya software in the other colors. In this environment, you gain a combined set of the benefits for scaling, supporting parallel loading and processing of data, a growth path, scaling by using unused resources (which allows adding more partitions, resources to existing partitions or worker nodes), simplified networking between LPAR nodes, supporting GPU-enabled workloads as they are added, and performance benefits with NVLink 2.0 technology.



Example: IBM Power System E980

Example: GPU-enabled AC922 servers

Figure 5. MPP mode: Mixed environment

Table 1 shows a summary of the different deployment options along with a summary of the key differences and the use cases that are good fits.

Number	POWER9 deployment options	Description
1	Single-server on accelerated compute server	• Is ideal for POCs, small workloads that do not require a cluster, or when a few application developers use a SAS Viya programming-only environment • Provides options for bare-metal, and CPU and GPU
2	Single-partition on a PowerVM-based server	• Is ideal for POCs, small workloads that do not require a cluster, or when a few application developers use a SAS Viya programming-only environment • Is also ideal for clients already running SAS 9 on AIX who want to integrate with SAS Viya and leverage unused POWER9 system resources • Supports PowerVM virtualization and CPUs only
3	Multi-servers on accelerated compute servers	• Is ideal for SAS Viya full deployment environments of any size • Offers a growth path and maximizes scaling • Supports parallel loading of data and processing. Options for bare-metal, and CPU and GPU
4	Multi-partitions on a PowerVM-based server	• Is ideal for SAS Viya full deployment environments • Offers a growth path and maximizes scaling leveraging existing unused server resources • Is also ideal for clients already running SAS 9 on AIX who want to integrate with SAS Viya • Has simplified networking between nodes • Supports parallel loading of data and processing • Supports PowerVM virtualization and CPUs only
5	Mixed environment multi-servers with both PowerVM-based and accelerated compute servers	• Is ideal for SAS Viya full deployment environments of any size • Offers a growth path and maximizes scaling by leveraging existing unused server resources and adding more servers as needed (for example, when GPU-enabled workloads are needed)

Table 1. Summary of POWER9 deployment options for SAS Viya

Network configuration options

Each deployment option shown in the POWER9 deployment options section can be configured using either internal storage devices or external storage solutions for the workload source data. In either case, the servers used for a SAS Viya deployment need to be networked together. Figures 6 to 14 provide a network configuration example for each deployment option, illustrating how the servers and storage can be networked and connected. The following types of networks can be considered in your network design:

- Basic network: This provides users access to the system and provides a path for administrators to perform systems management.
- Node network: This is typically a private network that is used to provide high-speed, high-bandwidth, and low-latency communication between the nodes in the cluster. This needs to be a minimum of 10 Gb network.
- External data network: This provides a separate connection for data transmission between the nodes and the external data source.

With the single-server and single-partition deployments, the network design is simpler because it does not require internode connectivity. With the multi-partition deployments, the internode connectivity is simplified because all partitions are on the same physical server.

Figure 6 is an example of a single-server deployment with a bare metal Power AC922 server running Red Hat® Enterprise Linux 7.6. For internal storage, solid-state drive (SSD) or NVMe devices (shown in green) can be used in the POWER processor-based server. The figure also illustrates the use of a two-port 32Gb Fibre Channel card (shown in orange) if workload source data will be accessed from an external storage solution such as IBM FlashSystem®. The on-board 1GbE Ethernet adapter can be used for basic network access.

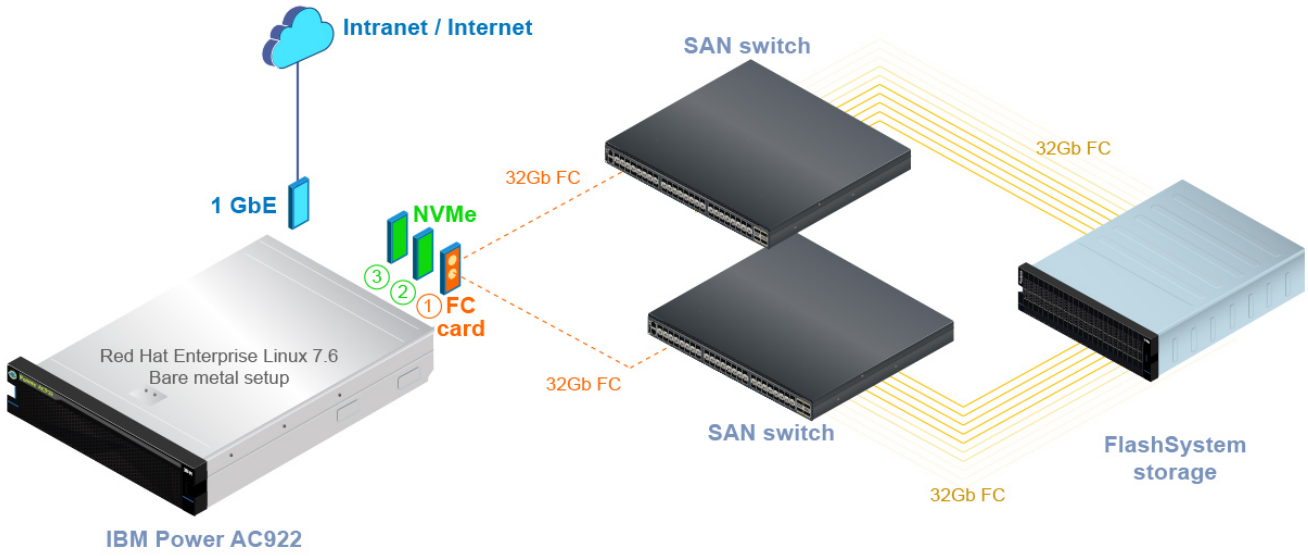


Figure 6. Single-server (internal and external storage)

Figure 7 illustrates an example of a single-partition deployment using one RHEL LPAR on an IBM Power E980 server. For internal storage, SSD or NVMe devices can be used in the POWER processor-based server. This example illustrates the use of a two-port 32Gb Fibre Channel card, dedicated to the LPAR, if workload source data will be accessed from an external storage solution such as IBM FlashSystem. This example uses VIOS for virtualized network and internal storage. A 1GbE or 10GbE Ethernet adapter can be used for basic network access.

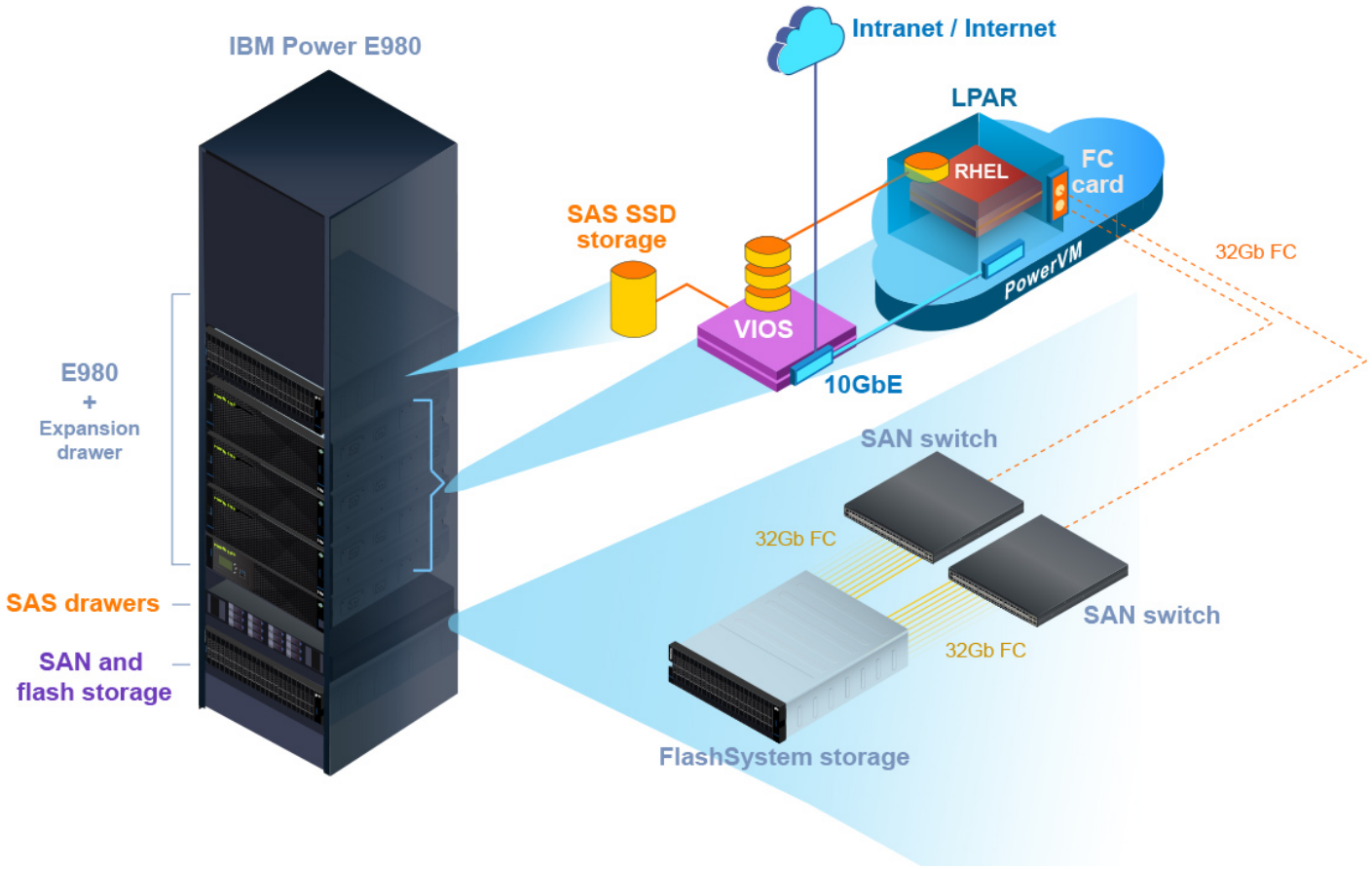


Figure 7. Single-partition (internal and external storage)

Figure 8 is an example of a multi-server deployment with a set of bare metal Power AC922 servers. This example shows the use of two NVMe adapters (shown in green) in the POWER processor-based server. One 10 GbE or 100 GbE two-port adapter on each server can be used for internode communications. The on-board 1GbE Ethernet adapters can be used for basic network access. **Figure 9** illustrates that you can add two-port 32Gb Fibre Channel cards (shown in orange) if workload source data will be accessed from an external storage solution such as IBM FlashSystem.

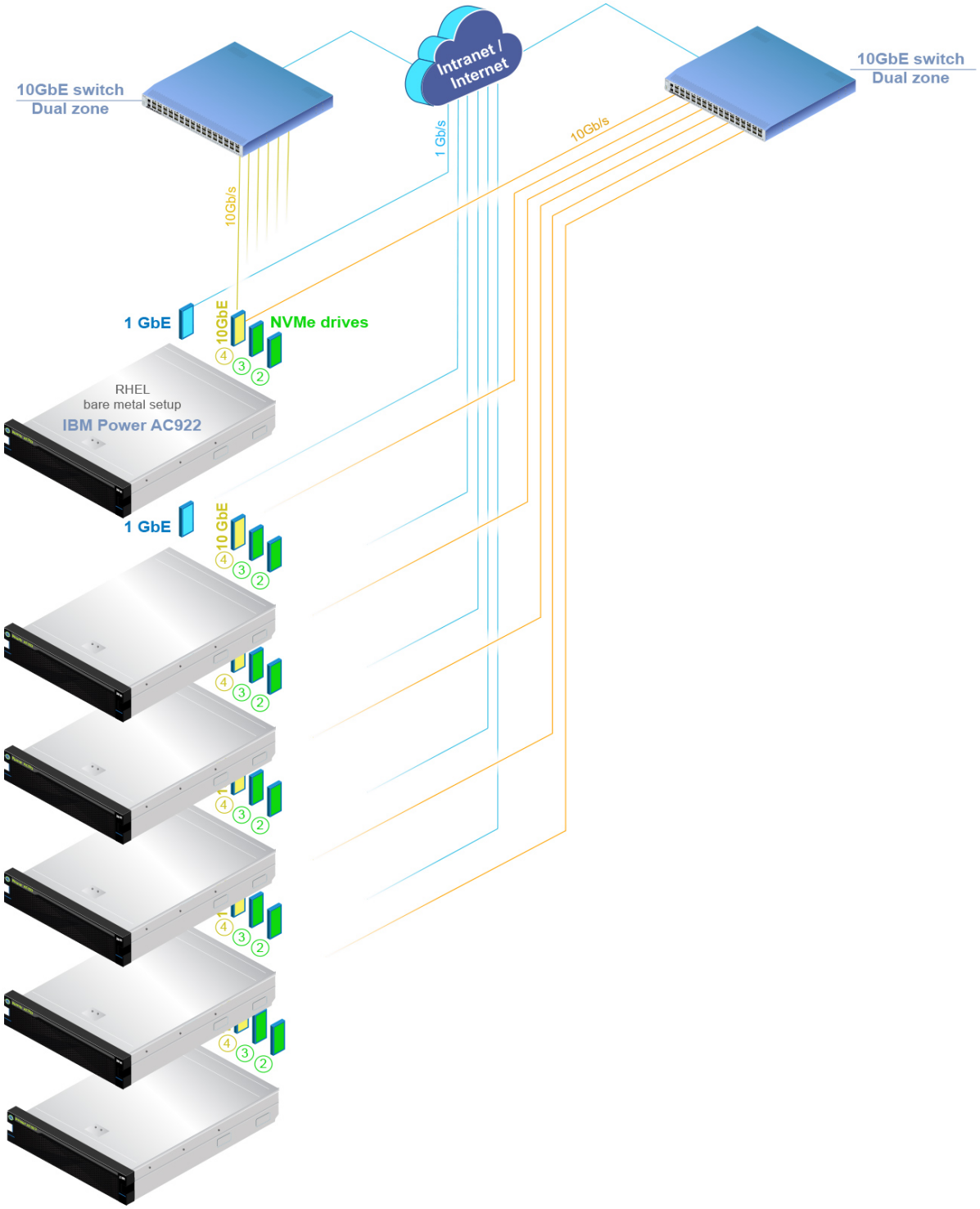


Figure 8. Multi-server with internal storage

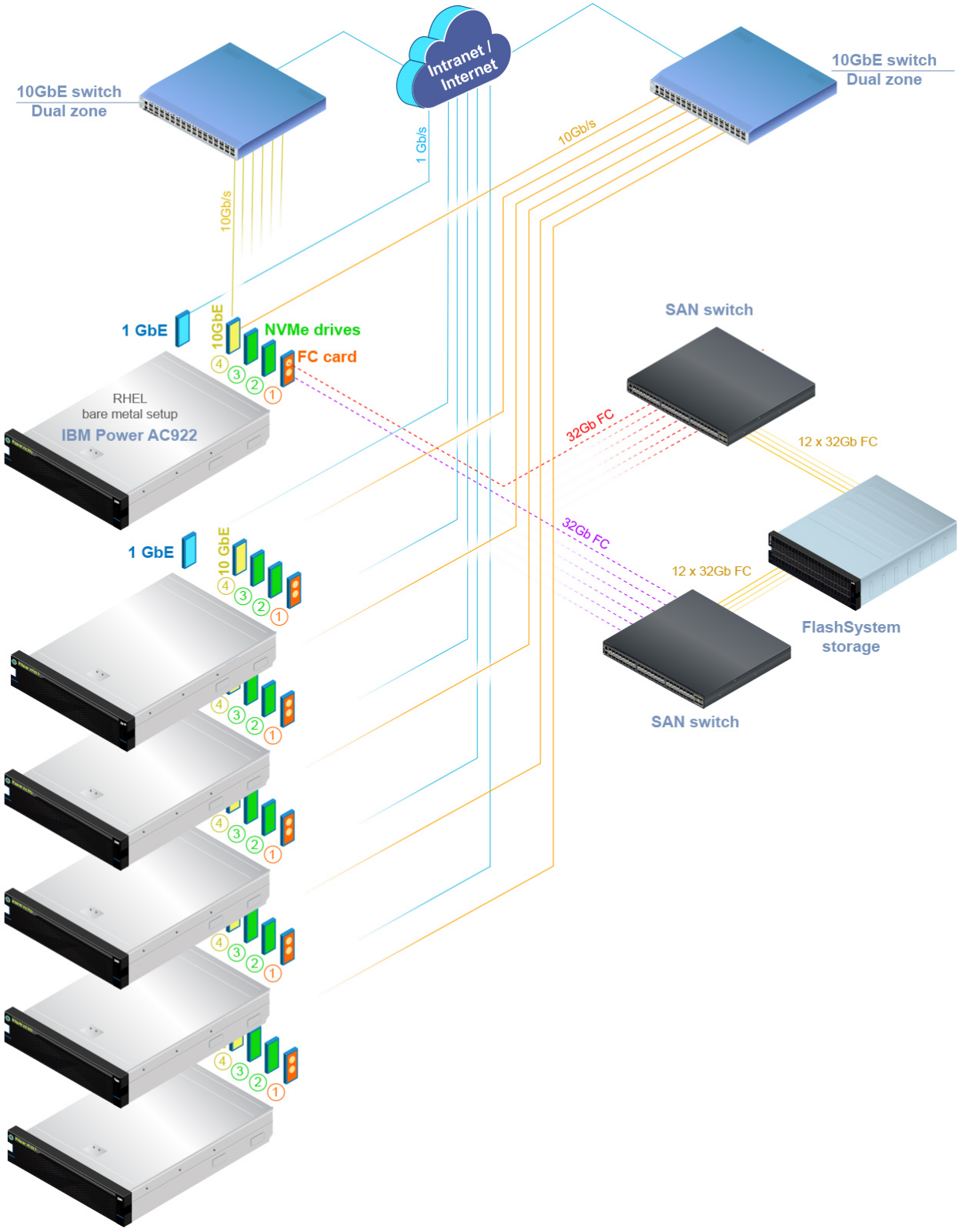


Figure 9. Multi-server with external storage

Figure 10 illustrates an example of a multi- partition deployment using a set of RHEL LPARs on an IBM Power E980 server. This example shows the use of one NVMe adapter (shown in green) in each LPAR. One 10GbE or 100GbE two-port adapter can be configured to each LPAR for internode communications. The VIOS is used for virtualized network and internal storage. 10GbE Ethernet adapters can be used for basic network access. **Figure 11** illustrates the use of a two-port 32Gb Fibre Channel cards (shown in orange) if workload source data will be accessed from an external storage solution such as IBM FlashSystem.

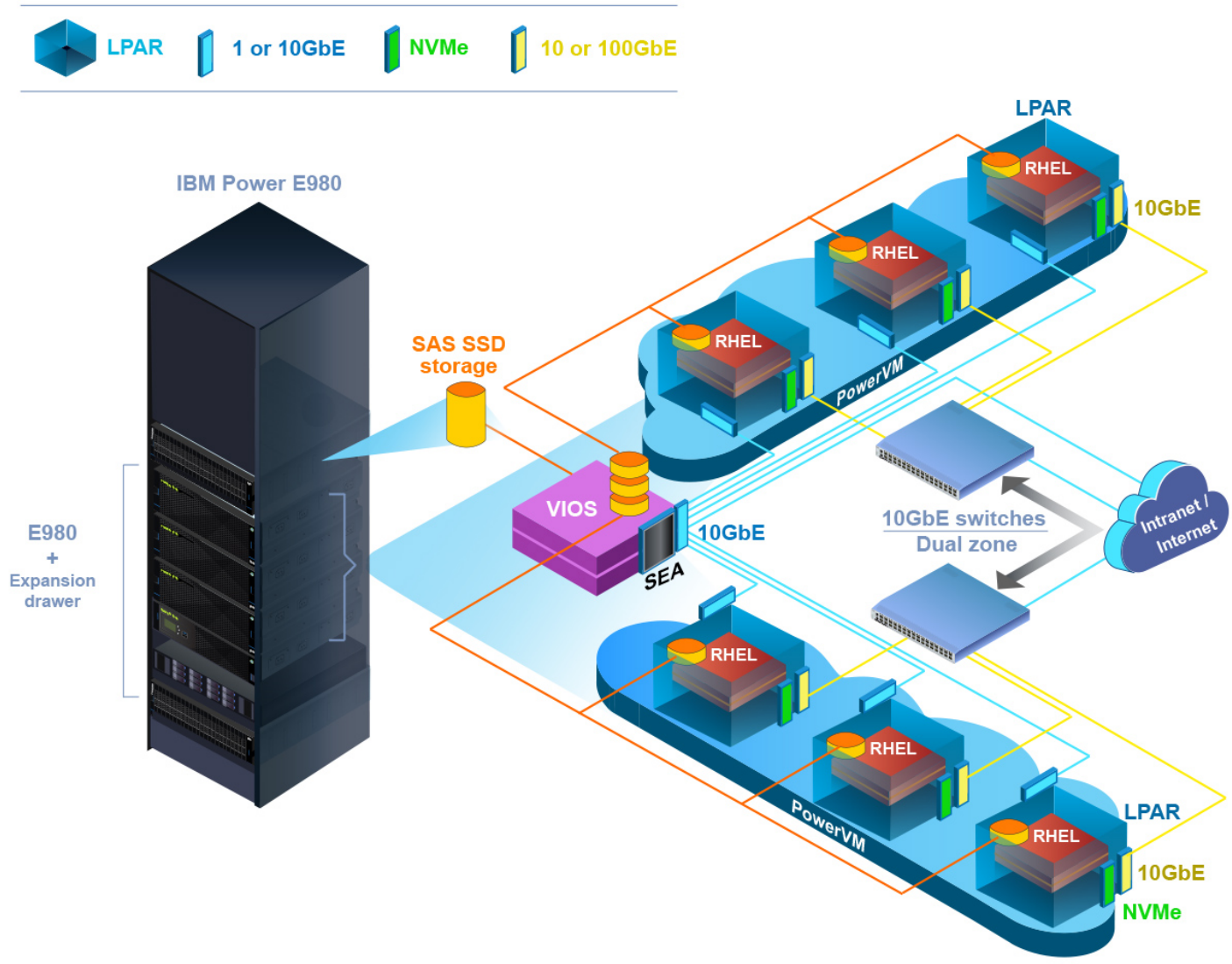


Figure 10. Multi-partition with internal storage

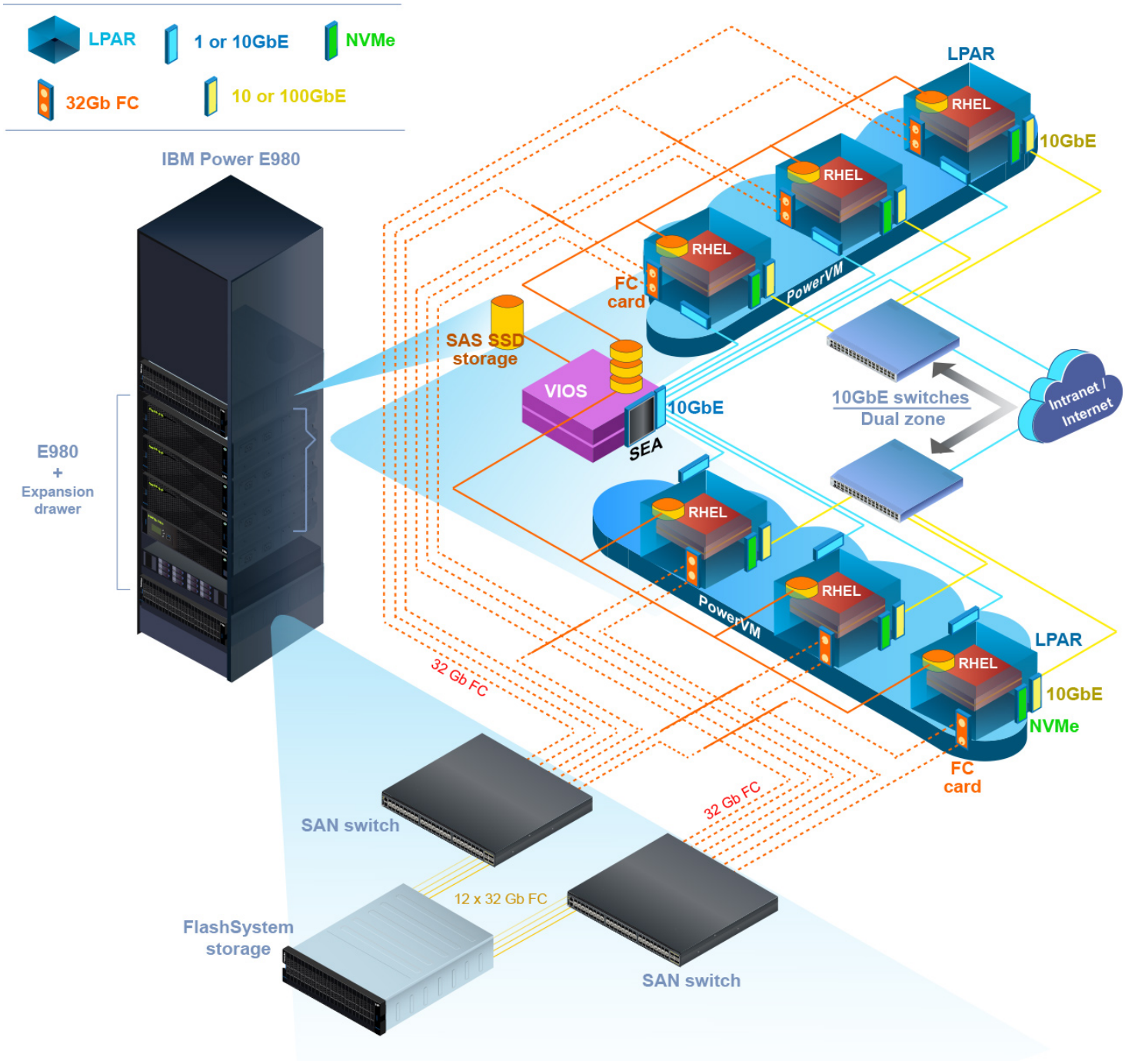


Figure 11. Multi-partition with external storage

Figure 12 illustrates an example of a mixed server deployment using a set of RHEL LPARs on an IBM Power E980 server and a set of bare metal Power AC922 servers. Each LPAR and server uses one or more NVMe adapter (shown in green) for internal storage. One 10GbE two-port adapter is configured to each LPAR and each server for internode communications. The VIOS is used for virtualized network and internal storage on the Power E980 server. Either 1GbE or 10GbE Ethernet adapters can be used for basic network access. **Figure 13** illustrates the use of two-port 32Gb Fibre Channel cards (shown in orange) if workload source data will be accessed from an external storage solution such as IBM FlashSystem. **Figure 14** illustrates the use of two-port 10GbE or 100GbE Ethernet adapters (shown in orange) if workload source data will be accessed from an external storage solution such as IBM Elastic Storage® System (ESS) 3000.

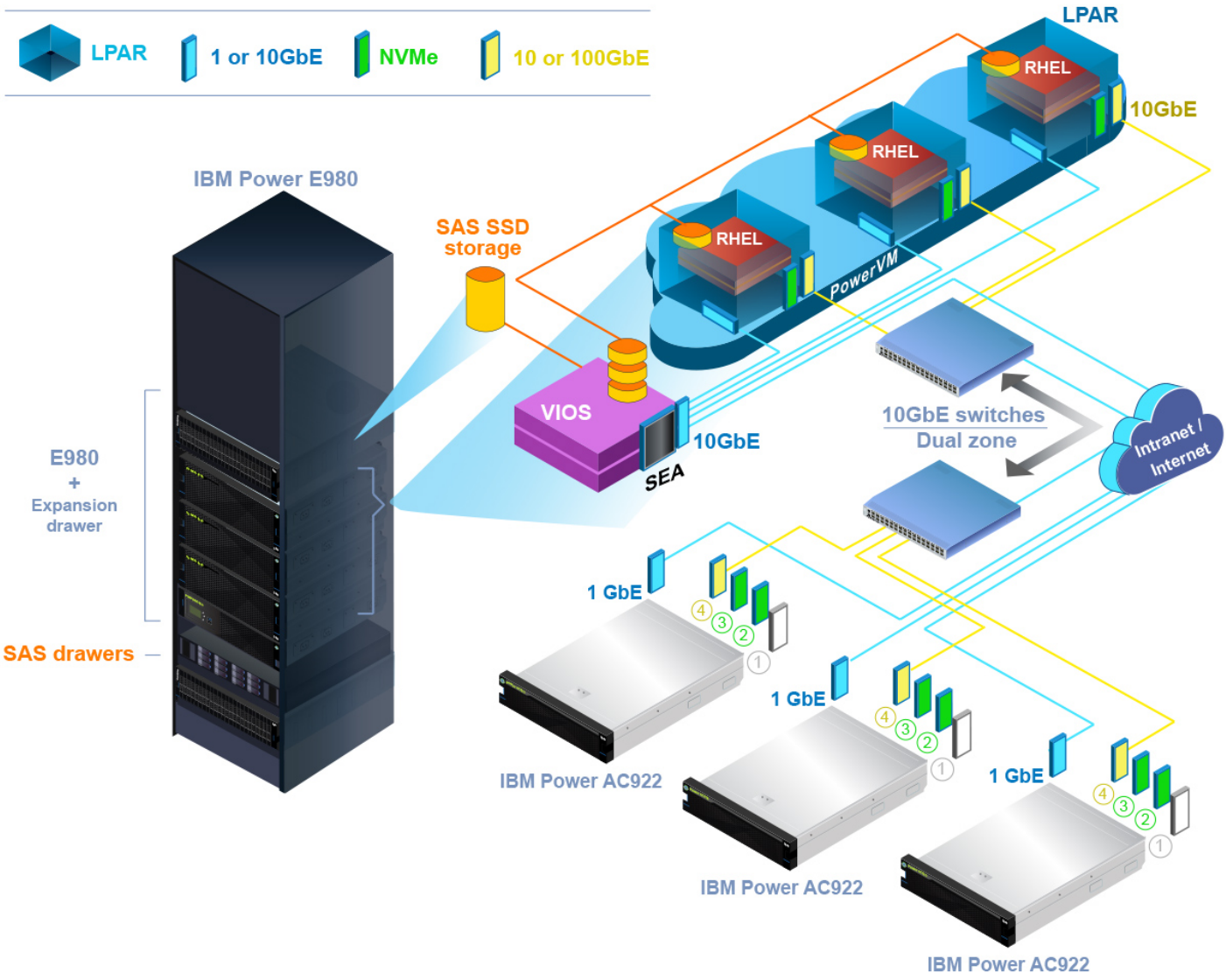


Figure 12. Mixed-environment with internal storage

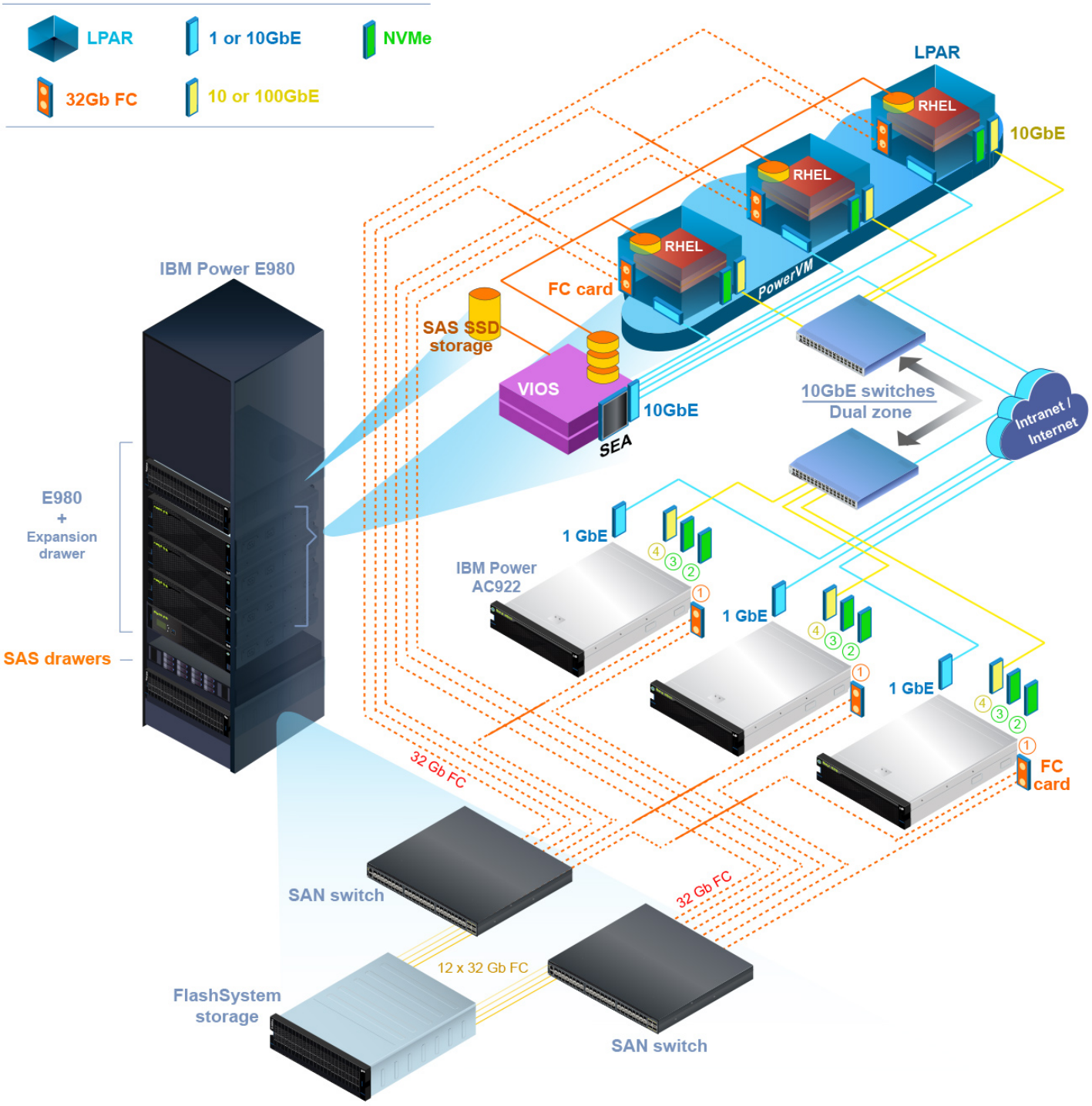


Figure 13. Mixed-environment with external storage

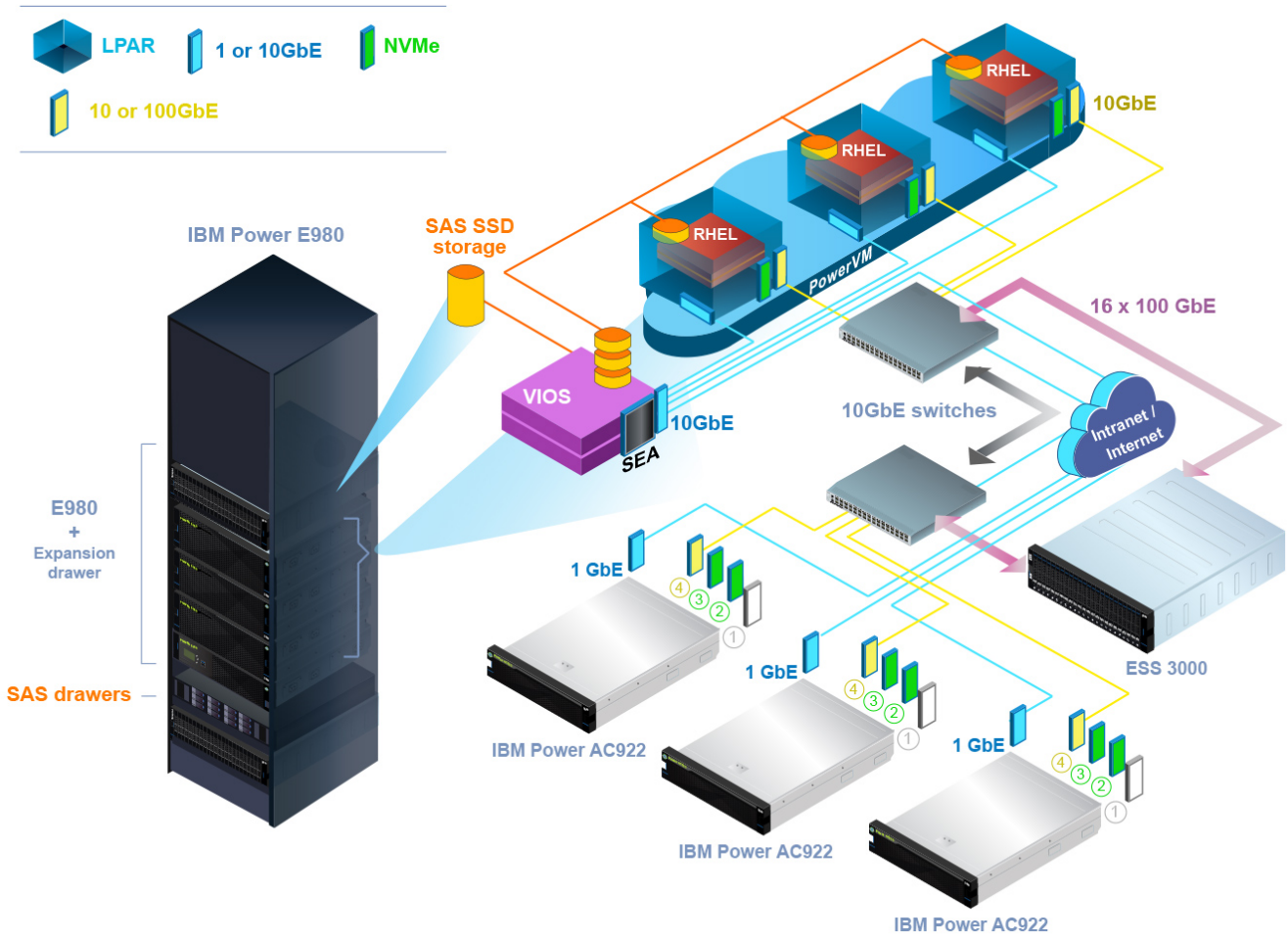


Figure 14. Mixed-environment with network attached storage appliance

POWER9 hardware minimum configuration options

This section describes the minimum hardware configurations for the various deployment options. When ordering POWER9 servers for a SAS Viya deployment, the SAS section in the IBM POWER eConfigurator tool reflects the minimum configuration options documented here. Table 2 shows the minimum configuration for an IBM Power AC922 server when it is used in a single-server deployment. The minimum configuration is dictated by a combination of the hardware configuration minimum of the server components and the SAS Viya software needs. Memory needs to be at least 16 GB per core unless more is needed to load the data into memory.

There are three types of storage required. The Viya management services (non-worker node functions) need two drives and solid-state drives (SSDs) are sufficient. The CAS_Disk_Cache (CDC) requires internal temporary storage. The workload source data requires permanent storage and it is best to have at least two drives for striping the data. For a configuration that allows for growth, consider a larger number of cores per processor and more memory. With the AC922 8335-GTH type and model, a maximum of two CPUs can be ordered (with up to 20 cores each) as well as up to four GPUs if the client's Viya workloads can make use of GPUs. If six GPUs are needed in one server, the server will need to be the water-cooled model.

For more information about storage, refer to the “Storage considerations” section. For more information about networking options, refer to the “Network configuration options” section.

Single-server hardware configurations

IBM Power AC922	Minimum configuration	Growth configuration	Notes
POWER9 server • Type-Model	IBM Power AC922 • 8335-GTH • Air-cooled	IBM Power AC922 • 8335-GTH • Air-cooled	
Processors, cores per processor	Two, 16-cores	Two, 20-cores	
Memory	512 GB	Up to 2 TB	16 GB per core minimum
Storage • Permanent space: Management services (internal) • Temporary space: CDC (internal) • Permanent space: Workload data (internal or external)	Two SSDs One SSD or NVMe Two SSD or NVMe	Two SSDs One SSD or NVMe Two SSD or NVMe	Min goal is 25MB/sec/core. Prefer stripe across more than 1 drive.
Network adapter • Basic access • External data (optional)	One 1 GbE One 32 Gb HBA or 100 GbE NIC card	One 1 GbE Up to two 32 Gb HBA or 100 GbE NIC cards	NIC card bandwidth is important.
Optional			
GPUs: NVIDIA Tesla V100 GPU accelerators with NVLink	Two	Four	Six GPUs require water-cooling GTH model. Viya 3.5 may not take advantage of more than four GPUs.
GPU cache	16 GB	32 GB	

Table 2. Use cases: POCs, small projects, programming-only, application development

Table 3 shows the minimum configuration for an enterprise-class server when it is used in a single-partition deployment. This configuration requires the use of PowerVM to create the partition where SAS Viya will be deployed. The minimum number of cores required by SAS is 16. The enterprise-class server models offer large configuration sizes (for cores and memory) that allow for growth and allow room for other workloads in other partitions on the same server.

Single-partition hardware configurations

IBM POWER9 enterprise-class server	Minimum configuration	Growth configuration
Processors, cores per processor	Two, 8-cores	Up to two 96-cores
Memory	512 GB	Up to 64 TB
Storage • Permanent space: Management services (internal) • Temporary space: CDC (internal) • Permanent space: Workload data (internal or external)	Two SSDs One SSD or NVMe Two SSD or NVMe	Two SSDs Up to 16 NVMe cards
Network adapter • Basic access • External data (optional)	One 1 GbE One 32 Gb HBA or 100 GbE NIC card	One 1 GbE One 32 Gb HBA or 100 Gb NIC card
Virtualization hypervisor	IBM PowerVM	IBM PowerVM

Table 3. Use cases: POCs, small projects, programming-only, application development

Table 4 shows the minimum configuration for a cluster of IBM Power AC922 servers. The top row shows the different types of SAS Viya nodes. The three management type nodes require less memory than the worker nodes and will need internal drives for storage management related data. There are three different types of worker nodes. One worker node must be designated as a secondary CAS controller. One type of worker node supports CPU-based workloads, and the third type supports GPU-based workloads such as Visual Data Mining and Machine Learning (VDMML). The GPU-based worker nodes are optional and are currently only taken advantage of if VDMML is used.

Multi-server minimum hardware configuration

SAS Viya node type	CAS Controller (primary)	Viya Microservices and other processes	SAS Programming Runtime	CAS Worker Workloads: CPU-enabled + CAS Controller (secondary)	CAS Worker Workloads: CPU-enabled	CAS Worker Workloads: VDMML/ GPU-enabled
POWER9 server • Type-model	AC922 • 8335-GTH • Air-cooled	AC922 • 8335-GTH • Air-cooled	AC922 • 8335-GTH • Air-cooled	AC922 • 8335-GTH • Air-cooled	AC922 • 8335-GTH • Air-cooled	AC922 • 8335-GTH • Air-cooled
Minimum quantity	1	1	1	1	1	0
Processors, cores per processor	Two, 16-cores	Two, 16-cores	Two, 16-cores	Two, 16-cores	Two, 16-cores	Two, 16-cores
Memory	512 GB	512 GB	512 GB	512 GB	512 GB	512 GB
Storage for management services*	Two SSD	Two SSD	Two SSD	N/A	N/A	N/A
Storage* • CDC (temporary) • Workload data (or external)	One SSD or NVMe Two SSD or NVMe	N/A	N/A	One SSD or NVMe Two SSD or NVMe	One SSD or NVMe Two SSD or NVMe	One SSD or NVMe Two SSD or NVMe
Network adapter • Basic access • Node inter-connectivity • External data (optional)	One 1 GbE One 10 GbE	One 1 GbE One 10 GbE	One 1 GbE One 10 GbE	One 1 GbE One 10 GbE One 32 Gb HBA or 100 GbE NIC card	One 1 GbE One 10 GbE One 32 Gb HBA or 100 GbE NIC card	One 1 GbE One 10 GbE One 32 Gb HBA or 100 GbE NIC card
GPUs: NVIDIA Tesla V100 GPUs with NVLink	N/A	N/A	N/A	N/A	N/A	Two
GPU cache	N/A	N/A	N/A	N/A	N/A	16 GB

Table 4. Use cases: Visual analytics [CPU-enabled, VDMML (GPU-enabled)]

*The storage is assumed to be internal storage for permanent purposes unless otherwise noted.

Table 5 shows the minimum configuration for a cluster of enterprise-class server partitions. Because of the flexibility of partition configurations, the number of cores and amount of memory can be smaller to match the minimum required by the SAS Viya software. Because GPUs are not available on enterprise-class servers, there is no GPU-based worker node type. VDMML workloads can still run on enterprise-class servers using CPUs for processing.

Multi-partition minimum hardware configuration

IBM POWER9 enterprise-class server	CAS Controller (primary)	Viya Microservices and other Processes	SAS Programming Runtime	CAS Worker Workload: CPU-enabled + CAS Controller (secondary)	CAS Worker Workload: CPU-enabled
Minimum quantity of partitions	1	1	1	1	1
Processors, cores per processor	Two, 8-cores	Two, 8-cores	Two, 8-cores	Two, 8-cores	Two, 8-cores
Memory	512 GB	512 GB	512 GB	512 GB	512 GB
Storage for management services*	Two SSD	Two SSD	Two SSD	N/A	N/A
Storage* • CDC temporary) • Workload data (or external)	One SSD or NVME Two SSD or NVMe	N/A	N/A	One SSD or NVMe Two SSD or NVMe	One SSD or NVMe Two SSD or NVMe
Network adapter • Basic access • Node interconnectivity • External data (optional)	One 1 GbE One 10 GbE	One 1 GbE One 10 GbE	One 1 GbE One 10 GbE	One 1 GbE One 10 GbE One 32Gb HBA or 100 GbE NIC card	One 1 GbE One 10 GbE One 32 Gb HBA or 100 GbE NIC card
Virtualization hypervisor	PowerVM	PowerVM	PowerVM	PowerVM	PowerVM

Table 5. Use cases: Visual analytics, VDMML

*The storage is assumed to be internal storage for permanent purposes unless otherwise noted.

For mixed environment deployments where the management nodes are installed on a set of enterprise-class server partitions and worker nodes are deployed on enterprise-class server partitions and/or Power AC922 servers, use Table 4 and Table 5 for configuration guidance.

Storage considerations

SAS Viya requires storage for proper functioning. It allows some flexibility with the type and location of the storage drives or devices, while also recommending internal drives for some specific elements. SAS documents several key storage considerations in the CAS storage performance consideration document with a focus on making the best storage choice to ensure that optimal Viya performance is achieved. It is important to read and apply those recommendations.

IBM Power Systems offer a wide variety of SSD and NVMe devices that can be used in SAS Viya deployments. The device sizes vary depending on the specific server type and model and provide the performance and capacity required by SAS Viya.

There are three primary types of Viya data-related storage entities. These types are mentioned in tables 2 to 5 in the “POWER9 hardware minimum configuration options” section of this document. Data and software associated with Viya-related management services, such as CAS controller and microservices, should reside on the internal storage (that is, front-facing devices). Generally, two or more write-intensive, high-performance SSDs are sufficient. The CDC is an on-device extension of memory mapped files, and also acts as a backing store for memory mapped file segments. The CDC requires high bandwidth and low latency. It is best provisioned as multiple write-intensive, high-performance SSDs or NVMe devices. Enterprise-class storage devices with the largest drive writes per day (DWPD) metric should be considered. The number and architecture of the devices (file system, parity, and striping choices) should be guided by the SAS Infrastructure Assessment activities to fit the EEC sizing or hardware estimate workload guidelines.

The SAS Programming Run-time (SPRE) node needs persistent storage for files that need to persist between runs. The SPRE node needs a SAS WORK file system, similar to SAS 9.4.

Persistent storage is where all Viya workload source and result data persistently resides. While this can be internal (collocated) storage, it is often external storage due to the volume of data. A high-performance Fibre Channel or Ethernet attached storage solution may be used for CASLIB PATH serial or distributed NFS (DNFS) parallel access to CAS nodes for processing.

IBM offers several storage solutions that are great choices for SAS Viya environments. The IBM FlashSystem 9100 system is a great choice for Fibre Channel attached storage with high bandwidth and low latency. You can see example network configurations for a SAS Viya deployment that leverages IBM POWER9 server and FlashSystem 9150 in the “Network configuration options” section of this document.

IBM ESS 3000 is a network storage appliance and comes with high performance IBM Spectrum Scale software. An example network configuration for a SAS Viya deployment that leverages IBM POWER9 server and ESS 3000 can be found in Figure 14 in the “Network configuration options” section of this document.

SAS Viya can be used with a variety of file systems. For example, on RHEL 7, **XFS** is a high performance, scalable file system and is the default. **ext4** is also supported. Both have extent-based allocation schemes which help in overall file system throughput.

When deploying CAS on an MPP system with a secondary CAS controller, a shared file system is needed. It is also needed when using a DNFS type of data source (introduced for accessing data from conventional storage solutions) where CAS nodes have concurrent access to the same data to support parallel loading.

IBM Spectrum Scale clustered file system is a great choice for this purpose. GFS2 can also be used.
When a Hadoop cluster and SASHDAT (SAS proprietary file format) files are used with SAS Viya, Hadoop Distributed File System (HDFS) is the natural choice especially when Hadoop data nodes are co-located with the SAS Viya CAS nodes.

With the CAS_disk_cache, only local file systems such as XFS or ext4 are supported. Cluster file systems are not supported.

To learn more about file system requirements, refer to the SAS Viya deployment guide at:
https://support.sas.com/en/documentation/install-center/viya/deployment-guides.html

Sizing considerations

The client requirements for analytics workloads and usage will influence the decision for which POWER9 deployment option and hardware configuration is best to deploy. The sizing process is critical for assessing the requirements and making that determination. The sizing process involves experts within SAS and IBM. Key inputs are needed from the client to start the sizing process. Table 6 shows a list of the types of input that should be gathered before starting the sizing process. After the key inputs are known, engage with your SAS account team to submit a sizing request to the SAS EEC Sizing team.

Input required	Description	Server specifications influenced
Workload type	Type of applications, analytics CPU or GPU-enabled	• Server model • Number of servers • Number of GPUs
SAS products to be installed	Type of workloads allowed to be processed	• Server model • Number of servers
Number of users/concurrent users	Estimated number of authorized SAS users by workload type, if known	• Number of servers • Overall size of server
Data set size	GB, TB, PBs of data to be ingested and analyzed	• Amount of memory • Type of storage (for example, internal versus external) • Speed of storage (for example HDD, SSD, NVMe)
Data set growth rate in 1-3 years	Helps select the right initial configuration that will enable easy scaling and growth	• Number of servers • Amount of memory • Type of storage (for example, internal versus external)
Location of data	Where will workload data be saved	• Type and number of adapters • Type of storage (for example, internal versus external)
Interested in using available space on an existing enterprise server	Helps determine if available resources are sufficient for planned usage	Server model

Table 6: Examples of sizing input needed

SAS Viya installation on POWER9

SAS Viya 3.5 should be installed following the instructions documented in the SAS Viya 3.5 for Linux Deployment Guide. You can find that guide at: https://bit.ly/2UAIkP1

The steps are identical for installing on an IBM POWER9 processor-based system running RHEL. If Viya will be installed in one or more PowerVM partitions, those partitions should be created before the Viya software installation process.

SAS Viya is only supported on RHEL for IBM Power LE (POWER9) 7.6 (sometimes referred to as *RHEL 7.6alt*). SAS Viya cannot be installed on RHEL 7.6.3.10 classic as it does not fully support the POWER9 instruction set.

Tuning best practices

There are several recommendations for running SAS Viya on IBM POWER9 servers.

Large memory capacity and high data bandwidth

The POWER9 processor has exceptional cache, memory capacity, and interconnect bandwidths that makes it a perfect environment for SAS Viya's in-memory analytics.

The IBM Power System E980 server has up to 64 TB memory capacity, and the IBM Power System AC922 server has up to 2 TB memory capacity. When equipped with a large memory capacity, it can support a much larger in-memory database than other platforms that are limited by memory capacity.

The POWER9 processor, with enhanced cache hierarchy, has a total of 120 MB of L3 caches with twelve 20-way associative regions and advanced placement policies. The L3 caches are fed by a 7 TBps on-chip bandwidth. Along with its peak I/O bandwidth of 80 GBps, as powered by PCIe Gen4 (with Gen3 compatibility), POWER9 provides a very high data bandwidth environment that is superior in the server market.

Fast storage devices for optimal performance

The recommendation is to use fast storage options, such as SSDs and NVMe devices that would provide a faster I/O speed that matches with the superior computing power on the POWER9 processor-based servers. Installing multiple SSDs or NVMe devices allows multiple storage devices to be accessed concurrently thus increasing total I/O write and read throughput. Creating a striped logical volume is a good method for balancing I/O across multiple SSDs or NVMe devices with the following Linux command examples:

Linux command to create LVM physical volumes:
pvcreate /dev/nvme0n1 /dev/nvme1n1

Linux command to create volume group:
vgcreate <striped_vol_group> /dev/nvme0n1 /dev/nvme1n1

Linux command to create striped logical volume:
lvcreate -i2 -L<volume size> -n <striped_logical_volume> <striped_vol_group>

Refer to the following Red Hat documentation on striped logical volume implementation:
https://red.ht/2JuoyOQ

Turbo frequency

Set the frequency performance mode to turbo for best performance. In a bare metal environment, the CPU energy governor is controlled by the operating system (OS). Use the `cpupower` command to verify and set the frequency as follows:

```
cpupower -c all frequency-info
cpupower -c all frequency-set -g performance
```

Simultaneous multithreading

Simultaneous multithreading (SMT) enables a single physical processor core to simultaneously dispatch instructions from more than one hardware thread context. With SMT, each POWER9 core can present up to eight hardware threads. Because there are multiple hardware threads per physical processor core, more instructions can run at the same time. SMT8 is the default for IBM Power System E980, and SMT4 is the default for the IBM Power System AC922. While POWER9 processor-based servers offer up to eight threads per core, Intel processor-based servers provide only two threads per core.

However, SAS Viya automatically determines parallelization based on the number of available hardware threads, and the application can over-parallelize to have more than 1000 threads running concurrently. In certain workloads, large numbers of runnable threads may cause lock contention and longer runtime. In such situations, reducing SMT to four or two may alleviate the lock contention and improve runtime. Use the ppc64_cpu command to set the SMT mode to 4 or 2 as follows:

```
ppc64_cpu --smt=4
ppc64_cpu --smt=2
```

```
procs         -----------memory---------- ---swap--    -----io----     -system-- ------cpu-----
r     b   swpd      free   buff     cache   si   so      bi     bo     in     cs us sy id wa st
1303  2      0 131319808    512 248346368    0    0 1890659     57 106992 136739 83 17  0  0  0
1180  1      0 127116992    512 258590144    0    0 1944883    258 104255 143950 88 12  0  0  0
1073  4      0 118337728    512 266641664    0    0 1553706 319264 101604 105751 79 21  0  0  0
1506  5      0 107629248    512 275541312    0    0 1631779 182636 119852 192827 78 21  1  0  0
```

Example 1. Output of the vmstat command showing high number of runnable threads in the first column

```
# Overhead  Command          Shared Object            Symbol
# ........  ...............  .......................  .............................................
#
    39.16%  cas              [kernel.kallsyms]        [k] _raw_spin_lock
     5.93%  cas              airlev.so                [.] IPRA.$GetClasVals
     5.58%  cas              libessl.so.1             [.] datb17v
     5.24%  cas              airlev.so                [.] IPRA.$MakeOneEffectXBuf
```

Example 2. Profile from the perf tool showing lock contention at the top of the list

Minimize NUMA effects in multi-partition configuration

Most multi-socket servers experience nonuniform memory access (NUMA) effects that can affect performance. When a partition is created or deployed, the PowerVM hypervisor chooses the physical resources to assign to the partition. The hypervisor has information about the dual inline memory modules (DIMMs), processor sockets, I/O devices, and so on and uses this location information to optimize the resources assigned to partitions. But over time, when there are multiple partitions defined on a server, especially with the use of functions such as dynamic LPAR (DLPAR), the assignment of resources may no longer be optimal.

It is a good practice to use the `numactl` Linux command to check the NUMA nodes assignment in each partition to ensure optimal memory assignment:
numactl --hardware

In the situation where a partition is unnecessarily assigned across multiple NUMA nodes, a reconfiguration of memory assignment may improve the performance of the partition.

In a lab situation where a power off on the server is possible, the following steps can reconfigure the memory assignment to all the partitions on the server for the most optimal assignment:

1. Make any profile updates that you need to make to the partitions.
2. Power off all the defined partitions.
3. Activate (power on) all the partitions specified in the current configuration with the necessary attributes. The partition needs to be activated to at least the SMS prompt.
4. After all partitions are simultaneously activated with the updated profile, start powering off all the partitions.
5. Power off the server.
6. Power on the server.
7. Activate the partitions from the management console. Order of activation is not important.

When in a production situation, there are often other existing workloads and users running on the same server which makes a power-off on the server not possible. The Dynamic Platform Optimizer (DPO) is a function provided as part of PowerVM that allows for the dynamic optimization of the resources assigned to active and inactive partitions to minimize these NUMA effects without interrupting the availability of the existing workloads.

One way to determine if a DPO operation should be initiated is to use the `lsmemopt` Hardware Management Console (HMC) command-line interface (CLI) command. Its output score is a representation of how close a partition or the server is to optimal assignment of resources. The `optmem` HMC CLI command is used to actually initiate a DPO operation on the server. There are also partitions exclusion parameters and partition ordering parameters for these two commands.

Refer to the following URL for a detailed DPO documentation:
https://ibm.co/2RjPPYP

Network tuning in an MPP environment

When there are significant internode networking activities in an MPP environment, increasing the maximum transmission unit (MTU) size from 1500 (default) to 9000 (JumboFrames) may reduce the RX error count and improve network performance.

To change the MTU size, use the /usr/sbin/ifconfig command as follows:
ifconfig ${Interdace} mtu ${SIZE} up

For example:
```
ifconfig enp1s0f0 mtu 9000 up
```

Summary

SAS Viya is a complex set of AI and analytics solutions that require a properly planned infrastructure to meet user requirements. The underlying infrastructure matters to ensure performance expectations and SLA requirements are met. Key POWER9 configurations should be considered in order to ensure that an optimized infrastructure deployment is achieved. Contact your IBM and SAS sales representatives for any questions or assistance with selecting the right IBM POWER9 deployment and configuration for your needs.

Get more information

To learn more about SAS Viya on IBM Power Systems, contact your IBM representative or IBM Business Partner, or visit the following website: https://www.ibm.com/it-infrastructure/power/capabilities/sas-viya

About the authors

Beth L. Hoffman is an Executive IT Specialist and Solution Architect in the IBM Cognitive Systems ISV Ecosystem Technical Development organization. She leads the enablement of AI and analytics solutions on IBM Power Systems for the past 10 years. Beth has more than 15 years of experience consulting and collaborating with key software solution companies. You can reach Beth at bethvh@us.ibm.com or www.linkedin.com/in/bethhoffmanibm

Sandy Kao is an Open Source Ecosystem Enablement and Optimization Engineer in the IBM Cognitive Systems ISV Ecosystem Technical Development organization. Sandy has more than 15 years of experience specializing in software performance and optimization. Most recently she leads the optimization of key AI and open source data solutions on IBM Power Systems. You can reach Sandy at kao@us.ibm.com or www.linkedin.com/in/sandy-kao-373270123

Abhijit Mane is a Technical Consultant in the IBM Cognitive Systems ISV Ecosystem Technical Development organization. He leads the enablement of AI and analytics solutions on IBM Power Systems. You can reach Abhijit at abhijman@in.ibm.com or www.linkedin.com/in/abhijitmane

Deepak Narayana is a Technical Consultant in the IBM Cognitive Systems ISV Ecosystem Technical Development organization. He leads the enablement of AI, analytics, and open source data solutions on IBM Power Systems. You can reach Deepak at narayana@us.ibm.com or www.linkedin.com/in/deepakraj

Revision history

The revision history of this deployment guide is detailed in the following table.

Version	Date	Summary of changes
2.0	April 2020	• Additional content and new figures in the *Network configuration options* section • Adjustments to memory and storage in the *POWER9 hardware minimum configuration options* section • New content about file system options in the *Storage considerations* section • Additional content in the *Tuning best practices* section
1.0	November 2019	Initial version

IBM Systems
3039 Cornwallis Road
RTP, NC 27709

Produced in the United States of America

Please recycle