How to Run SAP on Amazon Web Services

Technical operation manual for solution engineers and architects written by Rackspace Application Services

While the deployment of an SAP® system on AWS can be done in a few hours, the skill of a network wizard, a security specialist, a server expert, a storage guru and a SAP Jedi is necessary for its preparation. Amazon provides extensive documentation on how to install and operate SAP systems for the those familiar with SAP. The intention of this document is to provide a concise overview enabling fast-drafting of infrastructures and pricing estimates for those who are not fluent in SAP.

This document is based on excerpts from the following pages:

SAP on AWS Overview and Planning

SAP on AWS Pricing and Optimization

SAP on AWS Implementation Guide

SAP HANA on AWS Implementation and Operations Guide

High Availability and Disaster Recovery Options for SAP HANA on AWS

Setting up AWS Resources and SLES for SAP HANA® Installation

Migrating SAP HANA Systems to X1 Instances on AWS

Additional information about SAP Solutions on AWS

SAP HANA on the AWS Cloud: Quick Start Reference Deployment

The following SAP Notes are related to the topic of SAP on AWS:

1588667 - SAP on AWS: Overview of related SAP Notes and Web-Links

1656099 - SAP on AWS: Supported DB/OS and AWS EC2 products

2539097 - SAP NetWeaver License on AWS

2591601 - SAP on AWS: Adaption of your SAP License

1656250 - SAP on AWS: Support prerequisites

1964437 - SAP HANA on AWS: Supported AWS EC2 products

2309342 - SUSE High Availability Extension on AWS for HANA

2198693 - Key Monitoring Metrics for SAP on AWS

2288345 - EIM Applications on AWS

2142455 - SAP Replication Server for AWS Cloud

2358420 - Oracle Database Support for AWS EC2

2251474 - SAP LVM configuration of AWS EBS Storage Manager and Provisioning Guide

Also read the SCN Wiki that contains all SAP Notes for Linux. To read the SAP notes and use the SAP Quicksizer, you need the credentials to access the SAP support network (aka S-user).

Table of contents

Introduction

Naming of SAP solutions can be somewhat confusing, especially because SAP has changed the name of some solutions several times over the last decades. For example, the core ERP solution changed from R/2 through R/3 and ECC to S/4. For more examples, see Figure 1 below.

For more details of the different SAP solutions, see the Appendix A.

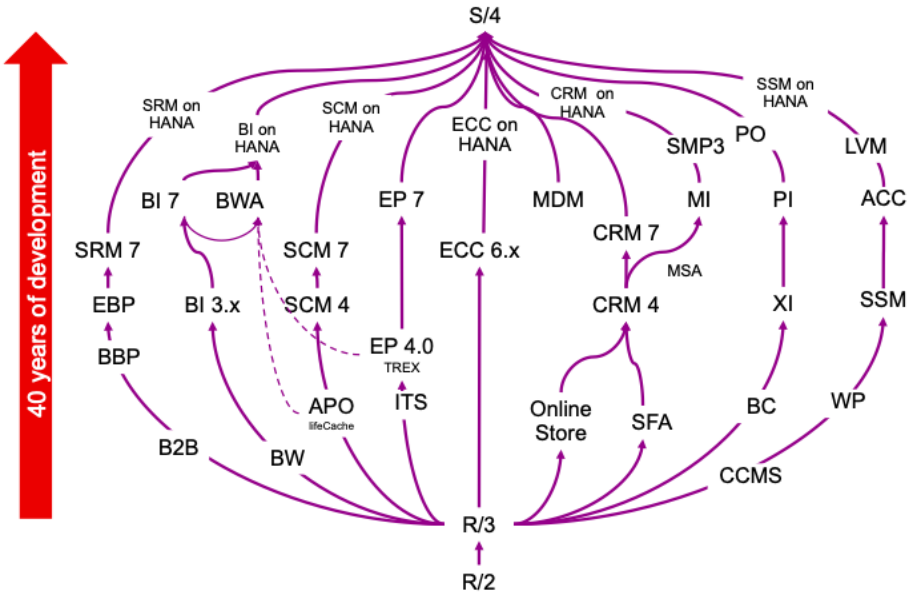


Figure 1: Historical development of SAP solutions and their naming

All of these “classic SAP” solutions literally have an identical architecture. The business processes written in ABAP (SAP’s own language based on COBOL) or JAVA are executed on the **SAP NetWeaver application server.** The business data as well as the SAP code is stored either on a traditional “anyDB” (Oracle, DB2, SQL, ASE, IQ or maxDB) or the new in-memory HANA[1].

The NetWeaver runtime acts as a **hardware abstraction layer,** enabling the business processes to be executed independently from the underlying platform (similar to .NET runtime). This is also true for cloud architectures such as AWS.

While there is generally only one database in a SAP system (some exceptions apply), there can be multiple NetWeaver application servers, but there must be always at least one. This basic infrastructure is accompanied by SAP WebDispatchers (which is just another NetWeaver App server) and other auxiliary components.

For details of the “classic” SAP NetWeaver Architecture, see Appendix B.

For a list of SAP solutions demanding either ABAP or Java application servers and some more technical terms not explained elsewhere, see Appendix C.

1 High-Performance Analytic Appliance, former SAP SE executive, Vishal Sikka, mentioned this architecture as "Hasso's New Architecture"

For an introduction of HANA technology, see Appendix D.

For a reference infrastructures and examples, see Appendix E.

In addition to the classic SAP solutions, there are also a wide range of solutions acquired by SAP over the years, including BusinessOne®, Hybris, Ariba®, SuccessFactors®, Fieldglass®, Concur® and Qualtrics®.

Each of these solutions introduce technologies, runtimes, etc., that are incompatible with classic SAP solutions. Architecture and sizing recommendations for these solutions are not included in this paper.

This paper and any AWS documentation for SAP describes only the setup of the classical SAP infrastructure. For anything else, the SAP documentation is only referenced.

What is supported by SAP on Amazon Web Services?

Operating systems (SAP Note 1656099):

- Red Hat® Enterprise Linux (RHEL) 6 or higher
- SUSE Linux Enterprise Server (SLES) 11 or higher
- Windows® Server® 2008 R2, 2012 (R2), 2016 and 2019

SAP products:

- All applications running on the NetWeaver 7 (SAP Kernel 7.21 PL #23 or higher)
- SAP liveCache 10.0 SP 25 released for EhP 2 for SAP SCM 7.0 and higher.
- SAP SCM Optimizer, version 12.0 or higher. See SAP note 1223407. And SAP note 1640509.
- TREX 7.10
- SAP BusinessObjects Financial Consolidation, version 10.0 or higher,
- SAP BusinessObjects
- SAP Business One (B1), according to SAP note 2058870.
- SAP Afaria 7 or higher on Windows 2008 R2 with m1.small, medium and large, m2.xlarge or cc2.8xlarge

Databases:

- SAP HANA certified configurations from the HANA Platform availability Matrix must be used. See SAP Note 1656099 for minimal Linux kernel versions.
- SAP ASE 15.7.0.051 or higher (need SAP NetWeaver 7.02 or higer)
- SAP IQ, version 16.0 SP08 PL20 or higher
- SAP MaxDB Version 7.8 or higher
- IBM DB2 LUW Version 9.7 or higher
- Microsoft SQL 2008 R2 or higher – please consider the new Microsoft SQL Server Use Terms as outlined in SAP Note 2139358. Amazon Relational Database Service (RDS) for MySQL and SQL Server.
- Oracle 11g R2 (11.2.0.4), 12c R1 (12.1.0.2), 12c R2 (12.2.0.1) or 18c only on Oracle Linux 6.4 or higher. For details and restrictions defined by Oracle see SAP Note 2358420:
    - Requires Oracle Linux 6.4 or later for the Amazon Machine Image (AMI) used for DB.
    - Oracle for SAP applications is supported on AWS EC2 only (not on AWS RDS).
    - Oracle Client for SAP application server is only supported on Oracle Linux 6 and 7.

    - Only single instance configurations of Oracle Database and Oracle Automatic Storage Management (ASM) are supported on AWS EC2.
    - Oracle Real Application Clusters (RAC) is **NOT** supported on AWS EC2.
- AWS Amazon Redshift database can be used by SAP PowerDesigner (PD) 16.6 SP05 or higher.
- Amazon Aurora is **NOT** currently supported by SAP Data Services SAP note 2739846

A few words on licences and licence keys

In public clouds, customer have to bring their own SAP license (aka BYOL).

- In general, SAP license fees are based on the number of named-users which may log on to the production system. With some exceptions, the number of cores doesn't matter.
- If customers extract data generated by SAP business processes from the database (e.g., to feed AWS analytics and AI tools), SAP consider this as "indirect access" and will insist that customers buy a "NetWeaver Foundation License for Third-Party Applications."
- While non-production systems also need a valid license key, they are free of charge[2].
- When buying licenses, customers can negotiate high discounts, but maintenance is always 22% of list price (18% in Germany, Switzerland and Austria).
- If a customer acquires SAP licenses excluding databases (as most do in America), then core based licensing of the database vendor applies!
- If a customer acquires SAP licenses including database run-time (as most in EMEA do), then the following charges apply:
    - Oracle, HANA — add 15 % to base price
    - MSSQL, DB2, ASE, IQ — add 8% to base price
    - maxDB (aka SAPDB) — add 5% to base price

Note

- SAP is terminating support contracts for run-time with Oracle, IBM and Microsoft for 2025.
- Oracle can only run on AWS on bare metal. This is not a technical issue but a rule imposed by Oracle.

Be aware that Microsoft closed the loophole of utilizing on-premises MSSQL licenses on "Dedicated hardware." You must license "Mobility" in order to deploy SAP on SQL Server in a shared environment. See Note 2139358.

In addition to the classic BYOL, there are several other options:

- SAP HANA Enterprise Cloud (HEC) where SAP owns and delivers the SAP HANA Enterprise Cloud managed service and supporting services on AWS infrastructure.
- SAP Cloud Platform (SCP) offers subscription and consumption-based licensing for the HANA service procured by SAP including the required AWS infrastructure.
- HANA express edition on AWS is a streamlined version for native HANA application development and data marts/analytics/big data with a free HANA license up to 32GB. Licences for 64GB, 96GB, and 128GB are available for purchase on the SAP Store.

2 For DR systems select "Backup System," provide DR system license key and generate the license.

After the installation of an SAP system, a licence key must be generated and installed within 3 months. This license keys are bound to MAC address of the first Network card, so they won't work on another server.

For systems running in VMs on AWS, SAP has changed this to the unique VM ID which is provided as part of the AWS. If this VM ID is not available within the AWS infrastructure, or cannot be read by the SAP kernel, no hardware key can be generated and no SAP license is created. See 2591601 - SAP on AWS: Adaption of your SAP License and 2539097 - SAP NetWeaver License on AWS.

Building a SAP Infrastructure on AWS

While every SAP system landscape is different in size of the key component size, they almost all follow the same pattern (with certain exceptions):

- Nearly every SAP customer runs more SAP solutions than just ERP (e.g., Business Warehouse (BW), Global Trade Management (GTM), Process Integration (PI) or the literally mandatory Solution Manager (SolMan))
- Every SAP solution landscape consists of at least of one development system (DEV), one Quality insurance system (QAS) and a production system (PRD), and sometimes many more
- Every SAP system consists of a database instance and at least one NetWeaver application server
- In nearly every case, you need some supporting components without a database such as the SAP NetWeaver Gateway

Let's start with the NetWeaver server; if no other information is available a best-practice size is:

- 4 vCores; 16GB if SAP gateway process run on dedicated server
- 8 vCores; 32GB if SAP gateway process run on this app server
- Add 1 vCore; 4GB if a SAP cloud connector run on the app server to use SAP Fiori
- For a stand-alone SAP gateway, 2 vCores and 8GB memory is sufficient
- in general SAP Fiori will add approx. 5% load on the application servers, which can be noticeable when you have a high number of users logged in

For the database, relatively modest VMs can be used with traditional disk-based databases like Oracle, DB2 or SQL. In many cases, customers even run a database and application server in the same VM, c alled a 2-tier Architecture. See the next chapter on sizing.

As an in-memory database, HANA needs a physical memory of the same size as the disk space for a compressed traditional DB. HANA needs anything between 64GB (the minimum for HANA to boot) and 25TB main memory. The sizing is entirely driven by the memory from where the numbers of vCores and demand on SSD disks are derived by formulas given by SAP.

Usually, the DEV systems need a HANA database of 128GB or 256GB, in rare cases 512GB. For the PRD and QAS system (recomended QAS has the same size as PRD) you will find any possible multiple of 128GB

For more details on classic SAP architecture see the appendix B.

Step 1: Sizing

Customers usually don't know how much compute power, memory, disk space and I/O their SAP systems will need. In the past they are used that the hardware vendor knows how to get that information.

The process of determining the necessary resources is called sizing, and consists of two parts:

1. The estimation of the maximum load to be expected
2. The determination of the minimum required hardware configuration

In principle there are three methods an experienced SAP infrastructure architect can use:

- SAP Quicksizer for greenfield SAP installations on AWS
- Analysis of Early Watch reports (EWA) for migrations of existing SAP systems to AWS
- SAP HANA sizing reports for migrations of existing SAP systems to HANA on AWS

SAP Quicksizer

The SAP Quicksizer[3] is an online tool available to every SAP customer and partner. Based on experience from previous projects, the tool calculates the minimal necessary SAPS, memory, disk space and IO throughput needed.

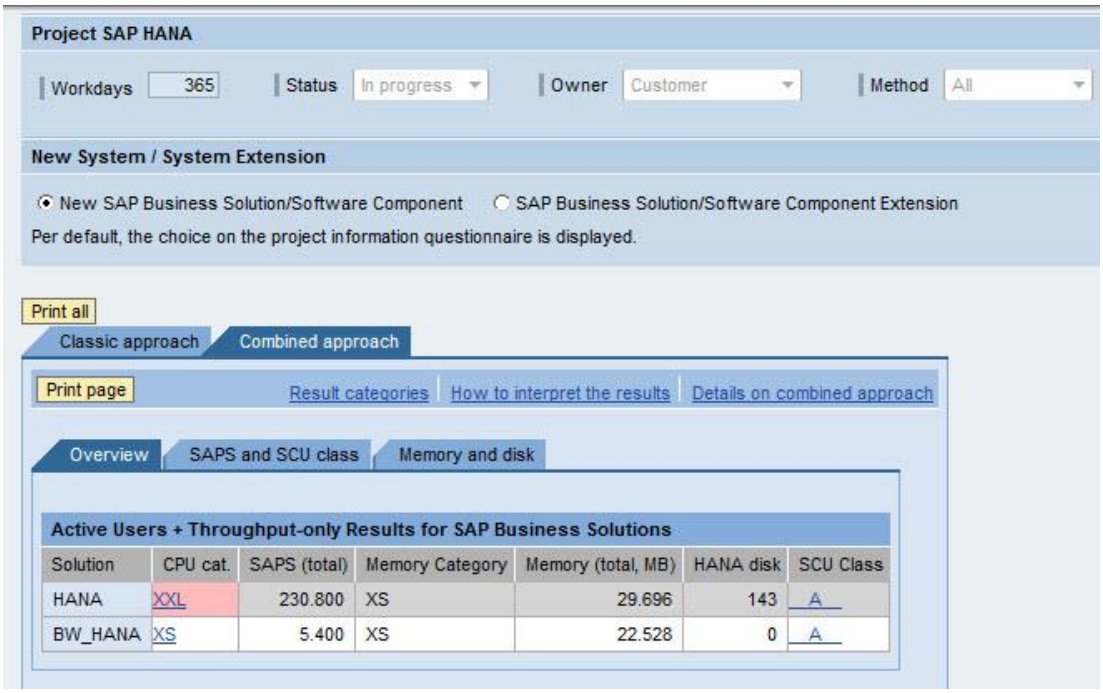


Figure 2: Output of SAP Quicksizer report

Note:

- Every SAP application has its own section in a Quicksizer project.

3 service.sap.com/quicksizing, SAP S-User credentials are necessary to access the tool

- Quicksizer does both user- and transaction-based sizing. The greater of the two is used to determine the hardware requirements.
- Quicksizer assumes a moderately customized system (less than 20% customer code only).
- Quicksizer includes a 40% security margin for “uncertainties.”
- Quicksizer considers productive systems only: no DEV, QAS or SBX.
- Quicksizer does not consider resource demand of OSs and hypervisors.
- Quicksizer follows the “garbage-in-garbage-out” concept. There is no check against nonsense entries.
- Quicksizer is continuously adapted to practical experience. To ensure that old versions are not used unintentionally, there is no offline version.

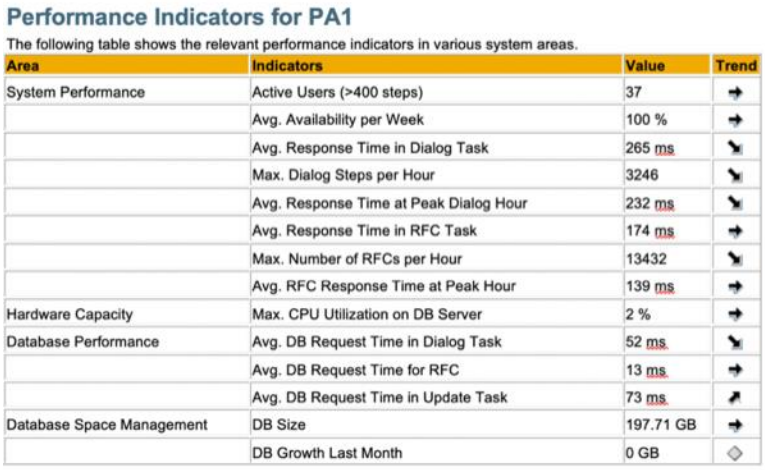
EarlyWatch Alert (EWA) Reports

Greenfield implementations with all their uncertainties are rare today. Most customers wanting to move SAP systems to the cloud have already used their SAP systems for many years on-premises or at a classic hoster. In this case, the individual resource demand can be derived from measurement rather than estimations.

As part of the maintenance contract, SAP customers receive EWA reports. These reports check the system parameters against default settings and recommend buying more SAP services. While the majority of the EWA reports focus on the status of the application, you can use the information about the platform utilization to determine the actual resource consumption of the SAP solution.

Figure 3: Example of EarlyWatch Alert title page

From the title page you can derive the type of SAP solution, the database and the SAP release.



Performance Indicators for PA1

The following table shows the relevant performance indicators in various system areas.

Area	Indicators	Value	Trend
System Performance	Active Users (>400 steps)	37	→
	Avg. Availability per Week	100 %	→
	Avg. Response Time in Dialog Task	265 ms	↘
	Max. Dialog Steps per Hour	3246	↘
	Avg. Response Time at Peak Dialog Hour	232 ms	↘
	Avg. Response Time in RFC Task	174 ms	→
	Max. Number of RFCs per Hour	13432	↘
	Avg. RFC Response Time at Peak Hour	139 ms	→
Hardware Capacity	Max. CPU Utilization on DB Server	2 %	→
Database Performance	Avg. DB Request Time in Dialog Task	52 ms	↘
	Avg. DB Request Time for RFC	13 ms	→
	Avg. DB Request Time in Update Task	73 ms	↗
Database Space Management	DB Size	197.71 GB	→
	DB Growth Last Month	0 GB	◇

Figure 4: Example of EarlyWatch Alert section performance indicators

From the section, “Performance Indicators,” (mostly on page 3) you can derive:

- Number of active users
- Maximum number of dialog steps per hour
- The actual DB size
- Last month DB growth
- Users measured in the system (unfortunately this has nothing to do with the “concurrent logged on” or “concurrent active” users necessary for sizing)

In the “Hardware Configuration” section is a “Hardware Capacity Check” where you can derive the:

- Hardware manufacturer
- Model of the database and application servers
- Number of cores and the amount of memory installed
- Maximum CPU load and memory utilization

Hardware Configuration

Host Overview

Host	Hardware Manufacturer	Model	CPU Type	CPU MHz	Virtualization	Operation System	CPUs	Cores	Memory in MB
de01c06n05	IBM	System x3850 5x [7143YH0]	Xeon E7 - 4870	Table info		SuSE Linux Enterprise Server 11 (x86_64)	80	40	1034130

hec01v017346	PC Vendor	[ ]	Xeon E7 – 4880 v2	Table info	Xen	SuSE Linux Enterprise Server 11 (x86_64)	16	16	65553
hec01v017347	PC Vendor	[ ]	Xeon E7 – 4880 v2	Table info	Xen	SuSE Linux Enterprise Server 11 (x86_64)	16	16	65553

Figure 5: Example of EarlyWatch Alert section on hardware type

CPU

If the average CPU load exceeds 75%, temporary CPU bottlenecks are likely to occur. An average CPUload of more than 90% is a strong indicator of a CPU bottleneck.

Memory

If your hardware cannot handle the maximum memory consumption, this causes a memory bottleneck in your SAP system that can impair performance. The paging rating depends on the ratio of paging activity to physical memory. A ratio exceeding 25% indicates high memory usage (if Java has been detected 0%) and values above 50% (Java 10%) demonstrate a main memory bottleneck.

Host	Max. CPU load [%]	Date	Rating	RAM [MB]	Max. Paging [% of RAM]	Date	Rating
hec01v017346	7	25.06.2018	✓	65.553	0		✓
hec01v017347	PC Vendor	25.06.2018	✓	65.553	0		✓

Note: For vitualization or IaaS scenarios (for example, IBM PowerVM, Vmware, Amazon Web Services, ...) it is possible that a CPU rating for some hosts is YELLOW or RED, even though the utilization value is quite low. In this case, the relevant host could not use maximum usable capacity due to a resource shortage within the virtualized infrastructure (for example, IBM PowerVM: Shared Pool CPU utilization).

Figure 6: Example of Early Watch Alert section on hardware utilization

Together with the SAPS number published at the SAP benchmark page, it's easy to calculate the maximum SAPS consumption on each server, and by adding the total number of SAPS the system "draws" under peak load. This is actually the number tried to guessing in a Greenfield sizing.

Without a doubt, orchestrating a cloud infrastructure based on EWA analysis is far more precise than the estimates of the Quicksizer.

Note:

- The averages for response times, users and transaction and time profiles in a EWA are derived from a full week (Monday to Sunday).
- The maximum CPU load is the highest average over a full hour. Relatively small load peaks, which usually cause the most trouble, are ironed out.
- The report covers only three weeks. If a high-load generating activity, such as for a year's end, you miss the real peak.
- CPU resource consumption is only measured as a percentage and not an absolute number independent of server and CPU type.
- Parameters like the IO throughput (which have become essential) are not measured at all.
- For each SAP system, you have to analyse another EWA report.

HANA Sizing Report

While there is a HANA section in the SAP Quicksizer, most customer aiming for a S/4HANA installation want to take over their historical data from their present system. SAP provides special reports to help predict the necessary memory considering the "compressibility" of the customer data as well as other factors.

The report for the SAP business suite (e.g., ECC, CRM, SRM, SCM, PI) has to be imported by the customers system operator (see SAPnote 1872170). For BW, the HANA sizing reports are already included in the actual BW release (SAPnote 2296290). Both reports are non-intrusive and need no downtime to implement. The reports provide all of the necessary information to size the database.

Example:

SIZING RESULTS IN GB	
Based on the selected table(s), the anticipated maximum requirement are	
for S/4HANA:	
- Memory requirement for the initial installation	2,188.1
- Net data size on disk for the initial installation	1,511.9
- Estimated memory requirement after data clean-up	1,517.9
- Estimated net data size on disk after data clean-up	1,536.6
Other possible additional memory requirement:	
- during the transition to S/4HANA (See FAQ)	18.9
Check the FAQ document attached to SAP Note 1872170 for explanations on how	
to interpret the sizing terms and	

Figure 7: Output of SAP HANA sizing report

Customers must be aware that according to SAP recommendations only approximately 60% of this memory can be used for data. The other 40% is consumed by internal processes and report calculations. Practical experience demonstrates that more than 60% of memory can be used for data, as long customer reports are relatively small. A report that can't be executed with the available memory will be aborted with an out-of-memory (OOM) message.

The size of the application servers doesn't change during a database migration and can therefore be derived from the EWA Report or you can use the best practice as given above.

A few words on Linux subscriptions

While you can use “plain vanilla” SLES or RHEL for application servers, SLES4SAP or RHEL4HANA is mandatory for HANA. These licences extend the support for a certain service pack from the default 6 to 18 months and provide additional essential HA libraries. RHEL4HANA for example adds 30-plus additional packages for HANA to standard RHEL.

Purchasing SLES4SAP and RHEL4HANA via the AWS Marketplace is also the only way to get in touch with the SAP experts within SuSE and Red Hat support organization. With standard SLES or RHEL, the customer is not flagged as an SAP customer in the joint support system. We have had cases where the default Linux support engineer mentioned to a customer that the kernel version or an OS setting is OK and fully supported while the SAP-related team stated correctly that for HANA this Kernel or OS setting doesn’t work and must be changed.

SAP uses SLES for SAP as a reference development platform. For example only SLES supports currently "live patching" for HANA omitting downtime for OS patching.

Step 2: Determine the necessary EC2 instance for compute

The result of a SAP sizing is expressed in SAPS, GB for memory and TB for disk. These numbers can be translated straight forward into AWS EC2 instance types using the following table after considering:

- For a 2-tier all SAPS and RAM used by database and SAP application server has to be in one VM.
- For a 3-tier System the rule of thumb is that 20% of the SAPS have to be on a single VM for the non-HANA databases while 80% can be distributed over multiple VMs for the application server part of SAP NetWeaver.

The following AWS EC2 instance types are supported in 2-tier or 3-tier configurations and can be used as application server(s) as well as a pure database server.

Instance Type	vCPU	Mem (GiB)	SAPS
m5.large	2	8	2,817
m5.xlarge	4	16	5,635
m5.2xlarge	8	32	11,269
m5.4xlarge	16	64	22,538
m5.12xlarge	48	192	67,615
m5.24xlarge	96	384	135,230
m5.metal	96	384	142,200

Figure 8: General-purpose (M) instances certified for SAP

Instance Type	vCPU	Mem (GiB)	SAPS
c5.large	2	4	2,650
c5.large	2	5,25	2,650
c5.xlarge	4	8	5,300
c5n.xlarge	4	10,5	5,300
c5.2xlarge	8	16	10,600
c5n.2xlarge	8	21	10,600
c5.4xlarge	16	32	21,200
c5n.4xlarge	16	42	21,200
c5.9xlarge	36	72	47,700
c5n.9xlarge	36	96	47,700
c5.18xlarge	72	144	95,400
c5n.18xlarge	72	192	95,400

Figure 9: Compute-optimized (C) instances certified for SAP

Instance Type	vCPU	Mem (GiB)	SAPS	HANA Scale-Up	HANA Scale-Out	Business One on HANA
r5.large	2	16	2,817			
r5.xlarge	4	32	5,535			
r5.2xlarge	8	64	11,269			✓
r5.4xlarge	16	128	22,538			✓
r5.12xlarge	48	384	67,615	✓		✓
r5.24xlarge	96	768	135,230	✓	✓	✓
r5.metal	96	768	142,200	✓		
x1e.xlarge	4	122	4,109			
x1e.2xlarge	8	244	8,219			
x1e.4xlarge	16	488	16,437			
x1e.8xlarge	32	976	32,875			
x1.16xlarge	64	976	65,750	✓	✓	✓
x1e.16xlarge	64	1,952	65,750			

x1.32xlarge	128	1,952	131,500	✓	✓	
x1e.32xlarge	128	3,904	131,500	✓	✓	

Figure 10: Memory-optimized (R and X) instances certified for SAP

Instance Type	vCPU	Mem (GiB)	SAPS	HANA Scale-Up	HANA Scale-Out
u-6tb1.metal	448*	6,144	480,600	✓	✓
u-9tb1.metal	448*	9,216	480,600	✓	
u-12tb1.metal	448*	12,288	480,600	✓	✓**
u-18tb1.metal	448*	18,432	444,330	✓	
u-24tb1.metal	448*	24,576	444,330	✓	

Figure 11: High Memory (U) instances certified for SAP

* Each logical processor is a hyperthread on 224 physical CPU cores

** Scale-out supported for SAP S/4HANA up to four nodes totaling 48TB of memory

High-Memory Instances and modern EC2 instances types such as m5, r5, and c5 leverage the **AWS Nitro System** (based on KVM instead the previous XEN virtualization).

This combination of AWS-built hardware and software components provide bare metal capabilities to eliminate virtualization overhead while offering and managing connectivity to Amazon VPC and EBS.

By offloading functions that have been traditionally supported through a hypervisor, Nitro instances enable applications to directly access underlying physical hardware.

You can use the same management tools whether the instance is virtualized or bare metal, and you can scale up or down – just like a virtual instance.

Amazon EC2 instance types that are certified and supported for SAP can be found be found at:

- Current generation instance types
- Previous generation instance types
- Instance type availability by Region

For additional information about EC2 instance-types for SAP see:

- Amazon EC2 instance types
- SAP Benchmarks Directory
- SAP Certified and Supported SAP HANA Hardware Directory

HANA specifics

The memory-optimized (R and X1) and Amazon EC2 high-memory (U) instance classes can each support SAP HANA deployments.

HANA scale-up certified Instance types as of Nov-2019.

Instance Type	vCPU	Mem (GiB)	SAPS	OLAP/OLTP	OLTP
r3.8xlarge		244		x	
cr1.8xlarge		244		x	
r4.8xlarge	32	244	38,200	x	
r5.12xlarge	48	384	67,615	x	
r4.16xlarge		488		x	
r5.24xlarge	96	768	135,230	x	
r5.metal	96	768	142,200	x	
r4.16xlarge	64	488	76,400	x	
X1.16.xlarge	64	976	65,750	x	
x1.32xlarge	128	1952	131,500	x	
x13.32xlarge	128	3904	131,500	x	
u-6tb1.metal	448	6,144	480,600	x	
u-9tb1.metal	448	9,216	480,600		x
u-12tb1.metal	448	12,288	480,600	x	
u-18tb1.metal	448	18,432	444,330	x	
u-24tb1.metal	448	24,576	444,330		x

Figure 12: HANA scale-up certified Instance types

Instance Type	vCPU	Mem (GiB)	SAPS	Max nodes
r5.24xlarge	96	768	135,230	16
x1.16xlarge	64	976	65,750	7
x1.32xlarge	128	1,952	131,500	25
x1e.32xlarge	128	3,904	131,500	25
u-12tb1.metal	448	12,288	480,600	4*

Figure 13: HANA scale-out certified Instance types as of Nov-2019

OLAP workloads like data marts, analytics, SAP BW/4HANA, SAP BW, and SAP BPC are supported on multi-node/scale-out configurations providing up to 100TB of memory when using the x1e.32xlarge instance type.

* Scale-out is supported for SAP S/4HANA (OLTP) with the u-12tb1.metal instance type, for up to 4 nodes, totaling 48TB of memory.

Instance Type	vCPU	Mem (GiB)	SAPS
c3.8xlarge		60	
r5.2xlarge	8	64	11,269
r5.4xlarge	16	128	22,538
m4.10xlarge		160	
R3.8xlarge		244	
M4.16xlarge		256	
r5.12xlarge	48	384	67,615
r5.24xlarge	96	768	135,230
x1.16xlarge	64	976	65,750

Figure 14: Business One on HANA certified Instance types as of Nov-2019

Purchasing Options

On-demand instances are paid by the hour, with no long-term commitments. Can be used for temporary SAP systems such as sandbox, PoC or additional app servers covering peak loads. Learn more

Reserved instances – a one- or three-year commitment to get a discount. They are suitable for any SAP workload. AWS options are standard, convertible, or scheduled instance types. Convertible reserved instance types provide flexibility to change instance families, OS and other attributes. This feature is particularly useful for SAP workloads. Learn more

When choosing the scope as either regional or zonal, consider that SAP application server and database must be in the same region. Learn more

Spot Instances take advantage of unused capacity in the AWS Cloud with up to a 90 percent discount. Don't use for SAP systems because Amazon terminates, stops, or hibernates your spot instance when the spot price exceeds your maximum price or capacity is no longer available. Learn more

Dedicated host is a physical EC2 server to be used with existing server-bound software licenses, such as a Windows Server, SQL Server, and SUSE Enterprise Server (subject to your license terms). Learn more

Step 3: Determine the necessary EBS storage

Amazon Elastic Block Store (EBS) is used for database, SAP software, data and log files and intermediate backups while Instance Store is used for OS swap only (because its gone when the instance is terminated) and Simple Storage Service (S3) for backup and archiving. EBS volumes are network-attached and exist independently from the instance. So EBS can be also used to hibernate a SAP system.

Amazon EBS options are General Purpose SSD (gp2) Volumes or Provisioned IOPS SSD (io1) Volumes. With gp2 the IOPS[4] depend on the volume size. If a particular SAP file system needs more IOPS than the gp2 volume provides, either change the volume type to io1 (provisioned IOPS) with more IOPS, or to increase the gp2 volume size.

SAP landscapes typically start on a gp2 volume and transition to io1 if more performance is needed. In some cases, it will be cheaper to allocate more gp2 space rather than changing the volume type to io1.

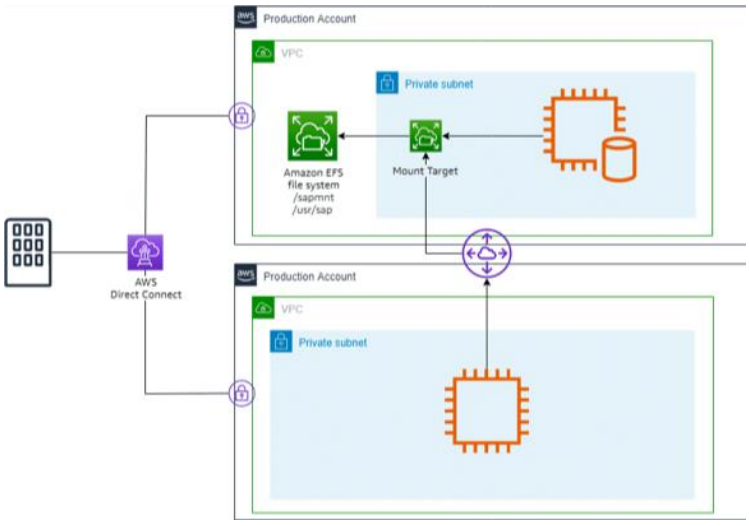
By attaching multiple Amazon EBS volumes to a single Amazon EC2 instance, you can partition the applications total IOPS load by allocating one volume for database log data, one or more volumes for database file storage, and other volumes for file system data.

SAP shared file system

Amazon EBS volumes can only be attached to one EC2 instance at a time. For volumes such as user/sap/ASCS/sapmnt and /usr/sap/trans or documents that have to be shared among many SAP systems, Amazon Elastic File System (EFS) must be used. See:

- EFS for Linux via the NFS protocol,
- FSx for Windows via the SMB protocol

Set up a AWS file system for SAP video



[4] Input / Output per Second

Figure 15: Example of Amazon EFS used for the SAP system mount and user directories across multible availability zones and regions

While it is necessary to use EFS for sharing global file system directories (like /sapmnt or /usr/sap/trans), EFS must not be used for DB data files or log files due to serious throughput and latency constraints. For details, please see the SAP on AWS Implementation and Operations Guide.

HANA specifics

When SAP introduced the HANA in-memory technology, their marketing claimed that “disk is dead.” However even with non-volaltile RAM there is still the need to “persist” customer data, logs, code and config files when power is down to avoid “Alzheimer” effects. In addition, this persistency layer (vulgo: disc) is also used for table joins, swapping of unused columns and keeping PSA and DSO tables not relevant for reporting. S/4HANA can be configured to swap transactions after a given time to disk to save precious memory.

The ABAP reports already mentioned in the section on sizing deliver in addition to the memory requirements also the so called “net data size on disk”.

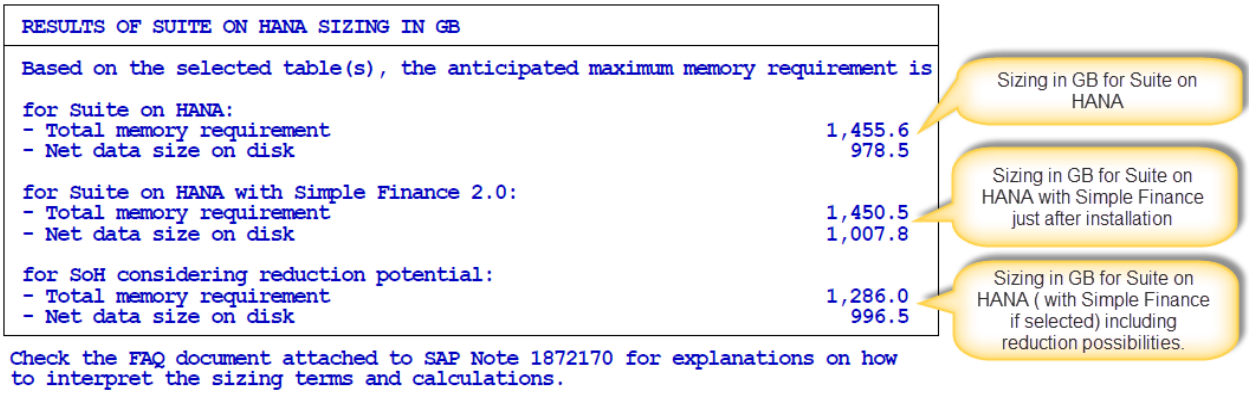


Figure 16: Typical output of HANA sizing script

With both numbers available the volumes can be determined:

- Data volume: 1, 2x net data size on disk
- Log volume: ½x memory up to 512GB, 512GB for larger systems
- Shared volume: 1x memory up to 1TB, 1TB for larger systems

A good rule of thumb for when the net data size on disks is not available: 3x RAM

To achieve necessary IO KPI’s for data and log volumes, multiple SSD’s must be deployed in parallel. AWS has worked with SAP to certify both EBS gp2 and io1 storage solutions for SAP HANA workloads with the configuration shown in the tables:

Instance Type	RAM (GB)	General Purpose SSD (gp2) storage (SAP HANA data and log volumes) – striped with LVM	Total max throughput (MiB/s)*	Total baseline IOPS	Total burst IOPS
X1e.32xlarge	3904	3x2048 GiB	480	18000	N/A
x1.32xlarge	1952	3x1024 GiB	480	9000	N/A
x1.16xlarge	979	3x700 GiB	480	6300	9000
x1e.4xlarge	488	3x400 GiB	480	3600	9000
r4.16xlarge	488	3x400 GiB	480	3600	9000
r4.8xlarge	244	3x300 GiB	480	2700	9000
r4.4xlarge	122				
r4.2xlarge	61				
r3.8xlarge	244	3x300 GiB	480	2700	9000
r3.4xlarge	122				
r3.2xlarge	61				

Figure 17: gp2-based storage configuration for SAP HANA data and logs

Instance Type	RAM (GB)	Provisioned IOPS (iol) storage (SAP HANA data and log volumes) – striped with LVM	Total max throughput (MiB/s)*	Total provisioned IOPS
x1e.32xlarge	3904	3x2048 GiB	960	15000
x1.32xlarge	1952	3x1024 GiB	960	15000
x1.16xlarge	979	2x1024 GiB	640	10000
x1e.4xlarge	488	2x600 GiB	640	9000
r4.16xlarge	488	2x600 GiB	640	9000
r4.8xlarge	244	2x450 GiB	640	9000
r4.4xlarge	122			
r4.2xlarge	61			
r3.8xlarge	244	2x450 GiB	640	9000
r3.4xlarge	122			
r3.2xlarge	61			

Figure 18: io1-based storage configuration for SAP HANA data and logs

In addition to the SAP HANA data and log volumes, all instances deployed by Quick Start will have the following storage configuration for root, SAP binaries, and SAP HANA shared and backup volumes:

Instance Type	RAM (GiB)	Root volume (gp2)	HANA binaries (gp2)	SAP HANA shared* (gp2)	SAP HANA backup** (st1)
x1e.32xlarge	3904	1x50 GiB	1x50 GiB	1x1024 GiB	1x8192 GiB
x1.32xlarge	1952	1x50 GiB	1x50 GiB	1x1024 GiB	1x4096 GiB
x1.16xlarge	976	1x50 GiB	1x50 GiB	1x1024 GiB	1x2048 GiB
x1e.4xlarge	488	1x50 GiB	1x50 GiB	1x512 GiB	1x1024 GiB
r4.16xlarge	488	1x50 GiB	1x50 GiB	1x512 GiB	1x1024 GiB
r4.8xlarge	244	1x50 GiB	1x50 GiB	1x300 GiB	1x1024 GiB
r4.4xlarge	122	1x50 GiB	1x50 GiB	1x300 GiB	1x512 GiB
r4.2xlarge	61				
r3.8xlarge	244	1x50 GiB	1x50 GiB	1x300 GiB	1x1024 GiB
r3.4xlarge	122	1x50 GiB	1x50 GiB	1x300 GiB	1x512 GiB
r3.2xlarge	61				

Figure 19: io1-based storage configuration for SAP HANA data and logs

* In a multi-node architecture, the SAP HANA shared volume is provisioned only once on the master node.
** see section on backup below

Note: Disk size, machine type memory, and network usage are calculated in gigabytes (GB), where 1GB is 230 bytes. This unit of measurement is also known as a gibibyte (GiB).

Storage for HANA dynamic tiering

To reduce the demand for valuable memory, SAP keeps infrequently accessed data on disk and call this feature HANA dynamic tiering. Managed by a dynamic tiering server the integrated component cannot be operated separately from the HANA database. The warm tier can keep up to five times more data than memory – backup storage must be large enough to hold both the in-memory data and the data that is managed on disk. For more, see HANA dynamic tiering on AWS and SAPnote 2555629.

As an alternative SAP introduced HANA native storage extension (NSE) with HANA 2.0 SPS 4. NSE allows you to define tables, columns or table partitions as warm to keep them on the HANA disc instead of using a additional VM for IQ. NSE loads pages of warm data into memory as required for queries.

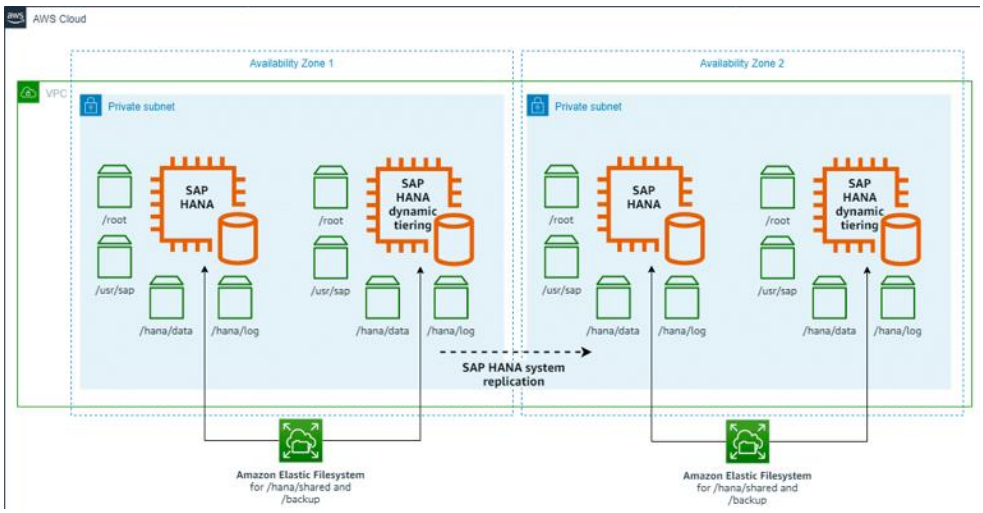


Figure 20: Multi-Availability Zone pattern for SAP HANA dynamic tiering

Step 4: Determine S3 capacity for backup

With all the HA features described in the next section, you may ask if Backups are necessary at all. Practical experience demonstrates that the likelihood that you need to recover because of physical hardware failure is nearly zero.

However practical experience also demonstrates that you will need to recover sooner or later because of a logical error – mostly because a user or admin deleted something which kills the consistency of the system. This is why you still need a backup. Since SAP depends on a database, we need to take database-aware backups of data and logs in order to achieve a point in time recoverability.

Alternatively, you can think about a shadow database approach such as HANA secondary time travel.

In regards of size of backup volume, HANA behaves like any other database – the size depends on the number of years collected data that has to be stored and the number of transactions. In case of a migration from an existing database to HANA you can just take the same number as for the old database, HANA will change nothing.

In case of a Greenfield installation, a best-guess rule of thumb is = 2x memory. Add more if you want several backup sets for audit reasons.

Best practice:

- For databases, Content Server, Adobe Document Server: Weekly full backup, 2-week retention (Wly-2W) = 2x storage
- For App Server, WebDispatcher: File System monthly full backup plus dayly incremental, 2-month retention (FS-Mly-Full+Dly-inc-2M) = 2.1x storage (assuming a 10% daily change rate)

For database backups, you can use third-party backup solutions that facilitate the database data and log backups required for conducting a point-in-time recovery. Typically, the only change is the backup destination, which will now point to Amazon S3.

Professional backup management solutions like Commvault and NetBackup integrated with the SAP-Backint API and AWS are highly recommended!

If you are not utilizing a third-party solution that directly backs up to Amazon S3, you can first back up to a local Amazon EBS volume attached to the instance, and then copy the backup data to Amazon S3. However, this is a risky and error-prone solution.

At least you should script the necessary tasks using the AWS Storage Gateway with Amazon SDKs, or the AWS Command Line Interface (AWS CLI).

Note: In HANA, it is not possible to delete just the log or incremental backup from backup catalog. If you try you will get an error message. The reason is obvious as deletion of any random log and incremental backup will make data backup useless for point in time recovery. See the SAP on AWS Backup and Recovery Guide for details on how to copy the catalog backup along with data and log backup.

HANA Studio offers management of data and log backups via “FILE” or “BACKINT” integration.

- With FILE, HANA studio manage writes to an NFS-mounted external resource.
- With the BACKINT API, backup tools like Linke’s Emory Backint for AWS S3 does backups directly from HANA to S3

This way HANA data and HANA transaction logs are secured. But what about configuration, executables, backup catalog and log files? The “FILE” option has the advantage that you can call the backup through a simple script, executed by cron for example on the master HANA node and include a copy of these flat files onto the same NFS resource (using “cp” or “tar”). This enables you to have more or less the same PIT (point in time) protected across all HANA data types and the data is residing together on the same NFS-mounted device.

To summarize the process for each data type:

1. DB logs should be backed up automatically and continuously by HANA towards an NFS directory so that logs are saved on a separate device immediately. This is determined by HANA configuration settings in regards to logging in the global.ini.
2. Data is backed up via the script execution as often as required towards the same NFS device, mostly once a day. By default, HANA uses the same names for each backup. If you want to keep your backups for one week for example, you need to re-name backup images accordingly, otherwise you overwrite your backup each day.
3. Flat files are included in each backup set. This makes recovery a lot easier and you don’t have to research when your configuration files may have changed in relationship to the backup image you want to use. You also don’t have to manage separate OS backups on the HANA system to capture changes to these files. Make sure to mount the external NFS device with the “nolock” option, otherwise your backup won’t start.

See the blogs HANA Backup Strategy on AWS about designs without any overhead of using third-party tool and HANA Housekeeping using HANACleaner for guidelines on automating the backup of housekeeping tasks.

HANA Restore

While Backup is a valid method to secure non-critical systems against physical disasters, recovery time is usually too long for production systems. In this case, HANA system replication should be used with backup as a second line of defense against disasters.

However, backups can be also used for other purposes, such as system copies to generate a new QAS system.

HANA offers 3 restore scenarios.

- Full restore of your last backup (changes and logs after the backup are lost).
- Full restore up to latest possible consistent point in time (backup and all available logs after the backup are used to recover the system).
- Point-in-time restore (backup and a subset of logs that were created after the backup are used to restore your system).

All three scenarios first include a complete restore of a full data backup and usually take several hours (depending on the size and the performance of your backup system). If you run into a log issue (missing or corrupted logs), you have to start the entire restore from the beginning.

No subsets of data, specific tables or view can be restored by backup. To also harden your system against logical and operator errors, consider implementing downtime-optimized HANA system-replication to enable secondary time travel as a shadow database.

Step 5: Availability / Disaster recovery for productive SAP systems

By definition, productive SAP systems are always among the most mission critical for every customer.

A Recovery Point Objective (RPO) = **0** (zero data loss) is **granted by default** in case of physical disasters as long the backup chain for data and logs are not interrupted and/or data is replicated synchronously to a second availability zone.

A Recovery Time Objective (RTO) = **0** (zero downtime) is **technically not achievable**. Just to minimize the RTO the different layers and components of an SAP system need different solutions. In reality HANA customers have to accept “worst case RTO” in the range of one hour with downtime optimized and two hours with cost optimized HA/DR configurations. A concept based on recovery from Backup has to accept at least eight hours RTO.

To secure the system against logical disaster (aka human error), you should consider a shadow database (aka secondary time travel in the case of HANA).

SAP single points of failure (SPoF) how to protect against them

Web layer: (SAP router, SAP web dispatcher, SAP cloud connector): The default HA/DR concept is redundancy. Just run multiple instances and configure a CloudWatch alarm to enable EC2 auto-recovery.

Application layer: Multiple instances are helpful, but not sufficient. In case an Additional Application Server (AAS) is failing and has to be rebooted, all users connected to this Application Server lose their open transactions

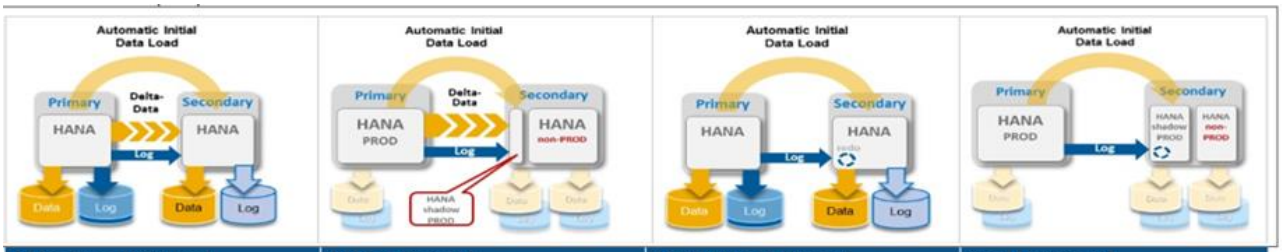
and have to log in again. Customers are likely to accept this because there is no mitigation and the reboot is quite fast.

In case the Central Services (ASCS) on the primary application server (PAS) is failing, the whole SAP solution has to be rebooted with all users losing their open transactions to avoid inconsistencies.

To mitigate this worst case, you can implement a “replicated enqueue server (ERS)” where a copy of the Enqueue table is kept in a second DC. This approach requires Disk sharing and special OS configuration.

Database layer: For Oracle, DB2, SAP ASE and SQL Server HA solutions have a look at the links. SAP HANA features its own HA solution called “HANA system replication” available in several options that can be combined:

- Synchronous for distances < 100km / Asynchronous for distances > 100km.
- Downtime optimized: Data is written into the memory of a hot standby system of same size as Production in a second DC. This hot-standby can be also used for read-only reports and secondary time travel (aka shadow DB) – for RTO < 1 hour.
- Cost optimized: Data is written to disc by a 64GB miniature HANA in second DC where a QAS system of the same size as Production can be shut down to donate the necessary resources in case the PRD system blow up – for RTO > 1 hour. Script the graceful shutdown of the QAS system and takeover of resources by the replication instance!
- Delta data shipping: Incremental data and log-transactions are transferred.
- Continuous log replay: Only log-transactions are transferred.



Downtime optimized Delta Data Shipping	Cost optimized Delta Data Shipping	Downtime optimized Continuous log replay	Cost optimized Continuous log replay
Incremental data and log-transaction data are transferred		Only log-transactions are transferred	
Data is preloaded into memory of hot stand-by in 2nd DC	Data is written to disk in 2nd DC	Data is preloaded into memory of hot stand-by in 2nd DC	Only data relevant for log replay are preloaded in 2nd DC
All resources of stand-by server reserved for possible take-over	Small fraction of resources in 2nd DC used for replication	All resources of stand-by server reserved for possible take-over	considerable fraction of resources in 2nd DC used for replication
Stand-by Server can be used in read-only mode	**Stand-by Server used for Quality insurance**	Stand-by Server can be used in read-only mode	Stand-by Server used for Development

Figure 21: Options for SAP HANA high-availability solutions

NOTE: Loading the row store is a main contributor of HANA startup times. Typically, 10 to 20GB of row store is loaded per minute. It is possible to keep the row store in memory while cleanly restarting HANA. Starting with HANA 2.0 SPS 04, the row store can be kept in memory, even in case of a dirty shutdown / crash – only dirty pages are reloaded from disk. For both cases the application has to be stopped or suspended cleanly beforehand. Details are described in Note 2159435.

SAP does not provide an automated fail-over for HANA solution landscapes so we also have to take care that the Application server can connect to the fail-over HANA instance. Such functionality is provided by Pacemaker and SUSE high-availability pattern together with SUSE SAPHanaSR resource agent package.

Auto-recovery for Amazon EC2 instance

When the connection to a particular instance is lost or the instance is unavailable, the CloudWatch auto-recovery alarm will detect the failure and trigger the migration of the instance onto physical hardware that is deemed healthy. This feature provides you with spare hardware, and you only pay for the CloudWatch alarm. This is sufficient for an application server and the web layer, but not to secure the database

Recommendations from the SAP on AWS High Availability Guide

Cause of failure	Recovery approach
Instance level hardware failure	Enable CloudWatch detailed monitoring, configure a CloudWatch alarm to enable EC2 auto-recovery
Logical volume corruption	Create new instance from AMI
Storage volume failure	Restore using S3 backups or point-in-time recovery using snapshots
Application and file system corruption, accdidental deletion of data, etc.	Recover the system to a point just before the issue happened

For PRD systems use architecture patterns designed to protect the SPoF of a SAP system by

Primary Availability Zone	Secondary Availability Zone	Failover mechanism
ABAP SAP Central Services (ASCS)	Enqueue Replication Server (ERS)	Enqueue logs are replicated from ASCS to ERS
Primary database	Secondary database	log shiping and DB replication
Additional application servers	Additional application servers	Manual or automated tools

AWS approach for component failover.

Automated alert with manual build new instances in the secondary Availability Zone from AMIs that are stored either in S3 or EFS. Recovery time to normal operations is less than an hour, but greater than just minutes.

Clustering solutions: below are some clustering solutions supported on AWS, depending on the OS/DB and SAP application combination:

- SLES HAE: SUSE Linux High Availability Extension
- RHEL HA: RedHat Enterprise Linux High Availability Add-on
- WSFS: Microsoft Windows Server Failover Clustering
- SIOS Protection Suite / SIOS DataKeeper SAP and HANA cluster software

While cluster seems to be the best option to improve availability by minimizing failover times, practical experience demonstrates that there is the risk that cluster worsens availability by initiating unnecessary fail-overs caused by their own bugs.

Manual trigger with automated build: write a “don’t panic” script that automates the build of new instances from AMIs that are stored either in S3 or EFS to be initiated after the admin is sure that fail-over is needed. Such scripts are also helpful for regular fire drills.

DR architecture patterns relevant for SAP systems are **passive**, **pilot light**, and **warm standby**.

Passive DR

During a failover, instances are launched based on their AMIs, backup and restore methods are used to restore operations. This approach is the most cost-effective but also has the longest recovery time because of the time required to provision the infrastructure.

Pilot light DR

Like Passive DR but using database replication. Quicker recovery time compared to backup/restore methods used in passive DR.

Warm-standby DR

Secondary Region hosts a scaled-down version of a fully functional SAP environment. During a failover, you scale the systems up quickly to handle the production load. Quicker recovery time compared to the pilot light pattern because some services are always running.

Step 6: Network

When predicting the traffic requirements to estimate the downstream traffic the SAPnote 2240690 - Front-end Network Bandwidth Sizing for SAP Fiori Apps and SAP S/4HANA does not really answer the question of how many KBs are needed per user, but gives at least a hint: “select a number between 10 and 25 KB….” we suggest 15 KB per user interaction step as the average network requirement.”

As a best practice for SAP, AWS VPN may be deployed in the initial stages of a project, and then as bandwidth requirements become known, customers may transition to AWS Direct Connect.

Note: in a Hybrid infrastructure the low keep alive time of AWS connection can force you to modify the RZ10 parameters to avoid getting kicked out of SAP.

In case of a migration to AWS flat files of several TB have to be transferred to AWS. Either dedicated connection to AWS, or temporary virtual private network (VPN), can be used to transfer the data.

Real-world experience shows that a VPN can support data transfer at rates of 100GB per hour. If this is not sufficient, customers can copy the data to an AWS Snowball device on-site and use a courier service to deliver it to the nearest AWS location for replication to the customer’s account.

Single VPC with only private subnets

Create a network that is only accessible from an existing, internal network, such as internally facing or back-office systems. This design is appropriate for customers who want to utilize their own internet service providers (ISPs) to control all internet-based traffic, and forgo public IP addressing capabilities from AWS.

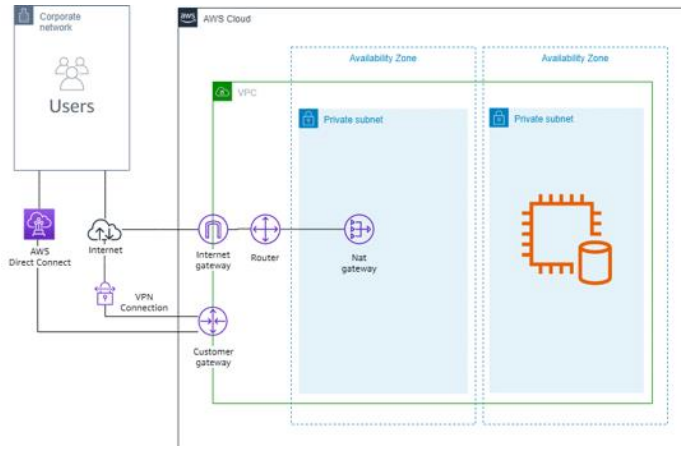


Figure 22: Single VPC with private subnets only

Single VPC with public and private subnets

Create a network that communicates with both internal (privately routed) and external (publicly routed) resources using a combination of public and private connections. This design is ideal for SAP workloads that need to accommodate a combination of public and private routing needs, such as all-in internet-facing, multi-tier web applications supported by databases or other privately routed backend systems.

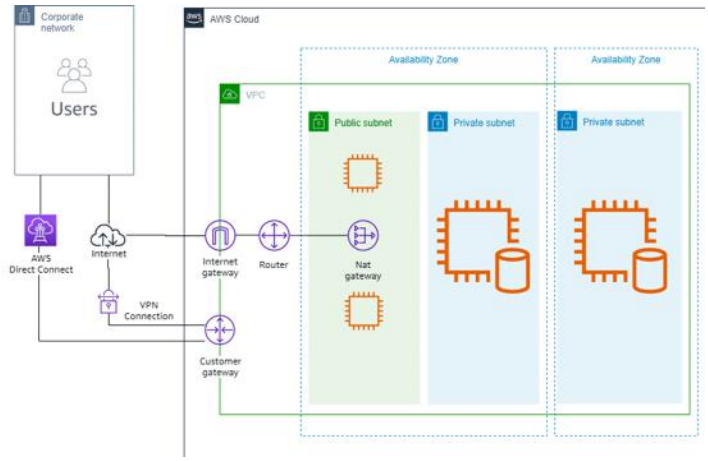


Figure 23: Single VPC with private and public subnets

Multiple VPCs

Are useful for to run an SAP production system in one VPC and an SAP non-production system in another VPC. These different VPCs need to communicate with one another for code transports.

Multiple accounts

Many SAP customers create multiple AWS accounts to streamline billing and restrict access to various environments, such as development, staging, and production, across different business and application teams.

VPC peering for multi-VPC deployments

A VPC peering connection is a networking connection between two VPCs that enables you to route traffic between them using private IP addresses. The VPCs can be in different Regions (also known as an inter-region VPC peering connection).

AWS Transit Gateway

A transit gateway acts as a regional virtual router for traffic flowing between your VPC and VPN connections. It comes with a default route table and can optionally have additional route tables.

Don't forget when setting up Amazon VPC for SAP

- Open ports 3200 or 3600 for SAP systems, or 30,000 in the case of SAP HANA.
- Don't separate SAP application and databases, you don't increase security but just add latency. In case they are separated across multiple subnets or VPCs: open the relevant ports and define security groups and network ACLs accordingly.
- In multi-account setups, explicitly share Snapshots/AMIs for SAP system builds, EC2 instance roles and S3 bucket access across account items.

Best practices for setting up VPC with SAP

- **IP range:** Make sure that the Classless Inter-Domain Routing (CIDR) range used on AWS is different than the on-premises range.
- **Elastic IP:** Use Amazon Elastic IP to ensure that the IP address is always associated with the respective instance, even if the instance is stopped and restarted.
- **Subnets:** It is recommended to have the SAP landscape split into at least two subnets.
- **Public subnet or demilitarized zone (DMZ**): SAProuter, webdispatcher, or any other public-facing applications will be hosted on your public subnet/DMZ.
- **Bastion host:** Use as a "jump box" in this subnet to connect to the SAP landscape.
- **Private subnet:** All SAP instances or, at the very minimum, an SAP database should be hosted on a private subnet. This subnet should be accessible via an on-premises network.

Click the links below for a more detailed discussion of Amazon VPC subnet zoning patterns for SAP.

- PART 1 Internal-Only Access
- PART 2 Network Zoning
- PART 3 Internal and External Access

AWS services for hybrid operation data transfer

- AWS VPN establish a secure tunnel from your network or device to the AWS global network.
- AWS Direct Connect provide more bandwidth using AWS Direct Connect locations.
- AWS Storage Gateway is a hybrid storage service providing storage for files, volumes, snapshots, and virtual tapes in AWS to on-premises applications.
- Amazon S3 Transfer Acceleration leverages edge locations useful for situations where global users are required to upload data to a centralized bucket, and where you need to transfer gigabytes to terabytes of data on a regular basis across continents.

Step 7: Auxiliary systems

Be aware of the necessary auxiliary system

- A internet gateway to allow access.
- A managed NAT gateway to allow outbound internet access for resources in the private subnet.
- A bastion host in the public subnet with an Elastic IP address to allow inbound SSH (secure Shell) access to the EC2 instance that host HANA.
- An optional remote desktop protocol (RDP) windows server to host SAP GUI and third-party front-end applications (like HANA studio) for managing SAP systems.
- A **jump server instance** is mandatory to allow the admins to access the virtualized systems.
- A **SAP router instance** is mandatory to allow users to access the virtualized systems.
- A **SAP Solution Manager** landscape with Dev and Production systems is highly recommended by SAP.

If your security policy requires truly internal VMs, you need to manually set up a NAT proxy on your network and a corresponding route so that VMs can reach the internet. It's important to note that you can't connect to a fully internal VM instance directly by using SSH. To do this, you must set up a bastion instance that has an external IP address and then tunnel through it.

When VMs don't have external IP addresses, they can only be reached by other VMs on the network or through a managed VPN gateway. You can provision VMs in your network to act as trusted relays for inbound connections, called bastion hosts, or network egress, called NAT gateways. For more transparent connectivity without setting up such connections, you can use a managed VPN gateway resource.

Step 8: Monitoring and Housekeeping

The AWS Data Provider for SAP is mandatory to monitor SAP workloads on AWS in SAP Solution Manager and to enable Amazon CloudWatch to gather infrastructure data send to SAP to get integrated support from SAP and AWS. The daemon starts automatically with your operating system and collects, aggregates, and exposes metrics to the SAP platform. SAP DB/OS Cockpit and SAP Support use the information gathered by the AWS Data Provider to analyze performance issues.

In order to set up the AWS Data Provider, you will need to grant the AWS Data Provider for SAP read-only access to Amazon CloudWatch, Amazon Simple Storage Service (Amazon S3), and Amazon EC2 so that you can use their APIs. You can do this by creating an AWS Identity and Access Management (IAM) role for your Amazon

EC2 instance and attaching a permissions policy. Watch the video below to see how to create an IAM policy and role to facilitate access to the relevant services.

You can also use the following AWS services:

- **AWS CloudTrail** to see who did a resize of an SAP system volume, or who stopped an Amazon EC2 instance that your SAP system is running on.
- **AWS Config** to assess, audit, and evaluate configurations of AWS resources that your SAP system is running on.
- **Amazon CloudWatch** performs monitoring of cloud services by default.
- **Session Manager** provides secure and auditable instance management, without the need to open inbound ports, maintain bastion hosts, or manage SSH keys. (Video)
- **Appstream 2.0** stream the SAP GUI to users omitting to install and update on every user's PC.

Automated SAP system management on AWS

Examples of common SAP management tasks that you can automate on AWS:

AWS CLI

- Add an Amazon EBS volume
- Take a snapshot of an Amazon EBS volume
- Take an image of an Amazon EC2 instance
- Stop or start your Amazon EC2 instance

AWS Systems Manager

- Back up HANA system
- Stop or start SAP system
- Automate and schedule patching jobs

AWS Lambda

- Refresh SAP systems with no production downtime
- Store SAP IDocs in Amazon S

CloudWatch Events

- Stop Amazon EC2 systems every day from 6 PM to 9 AM (hibernation)
- Back up SAP/DB every day at 10 PM using OS-level scripts

AWS Quick Start for SAP HANA

The AWS Quick Starts for SAP HANA provision the SAP-certified AWS configurations for HANA systems including all necessary auxiliary systems in less than an hour. The setup is based on AWS and SAP best practice and allow self-service and customization. Quick start also automates the HANA software installation (which takes more than an hour).

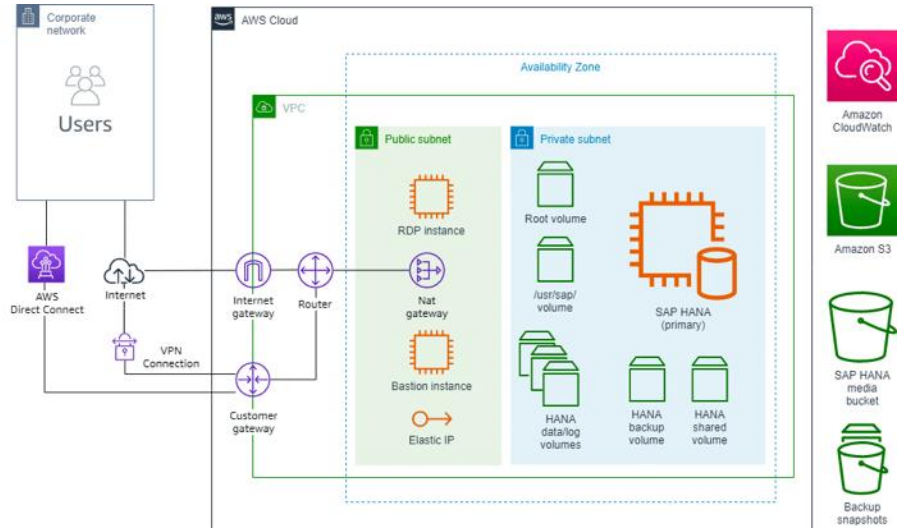


Figure 24: Single-AZ, single-node deployment built using Quick Start for HANA

The diagram below depicts a S/4HANA system built using the AWS Quick Start. Quick Start deploys the SAP application tier, the HANA database tier, Remote Desktop Protocol (RDP) and bastion hosts within a virtual private cloud (VPC) in your AWS account. The deployment includes an ABAP SAP Central Services (ASCS) server, a Primary Application Server (PAS) instance that provides SAP system utilities, an optional Additional Application Server (AAS) instance, and AWS core infrastructure services to scale out the SAP application tier.

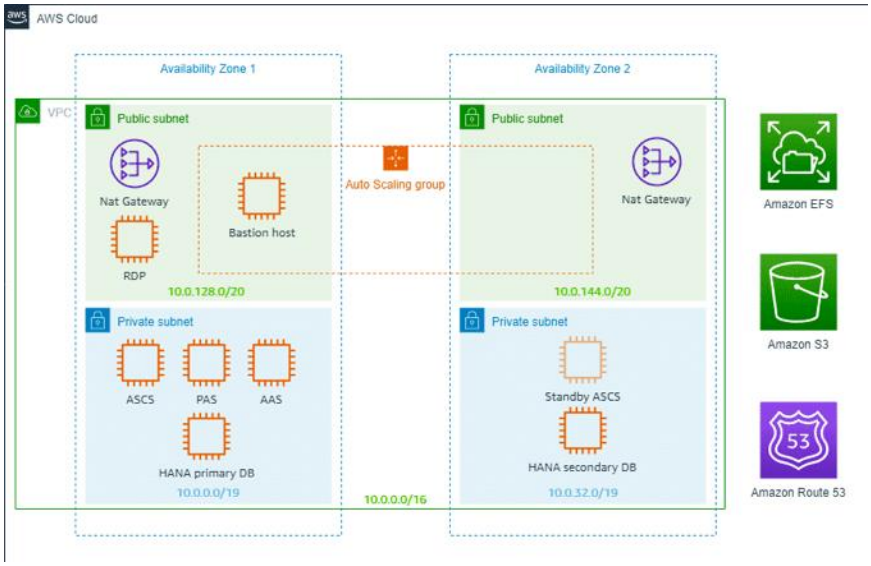


Figure 25: S/4HANA production system built using the AWS Quick Start

You can also leverage AWS CloudFormation or third-party tools such as Chef, Puppet, Ansible or Terraform.

Note

- **Parameterize inputs:** define unique attributes such as system ID, system number, file system sizes hostnames or operating systems as input parameters for AWS CloudFormation templates. Don't hardcode inputs.
- **Tagging:** leverage tagging to identify systems in the console that are billing to different business units. Tagging also informs the system of what needs to be stopped properly at certain times.
- **Build a repository of automations:** set up and configure multiple Amazon EC2 systems to provision an SAP landscape consisting of S/4, Fiori, ADS, ASCS, PAS, and AAS, ready for SAP software install.
- **Enable the AWS Cost and Usage Report:** to track associated costs.

Migrating SAP workloads to AWS

SAP distinguishes between homogenous migrations (same OS and/or DB on source and target) versus heterogenous (different OS and/or DB). This corresponds to AWS Rehost, also called “lift and shift” versus replatforming. A deeper change in business logic and data-model – like for SAP ECC to S/4 – is called **Transform** aka **refactoring** or **re-architecting** by AWS.

Rehost (homogeneous) migrations	Refactor/transform (heterogeneous) migrations
• Backup and restore • Replication (database or block-level) • HANA system replication	• SAP export/import migration process • SUM DMO with System Move option • Near-zero downtime migrations

For the necessary data transfer to AWS use Amazon VPC or Amazon Direct Connect. For large volumes, use AWS Snowball, a storage appliance that is shipped to the source data center and send back when data has been copied.

Backup/restore

- Build the target infrastructure in advance of your cutover date.
- Complete one full backup using established mechanisms and services.
- Continue to take incremental backups and archive log backups and copy to Amazon S3 over AWS Direct Connect or AWS VPN.
- Restore previous-day backups to Amazon EC2 until complete.
- Allow for several hours of downtime, which includes the validation process.

Replication

- HANA system replication (HSR) as described in the HA section. The HSR target can be initialized offline with a restore from a backup and then synchronized with the source system.
- Continuous data protection (CDP) tools track changes at the storage level and replicate them to the target. You can use such tools like AWS Server Migration Service (SMS) and CloudEndure also for lift and shift migrations. Be aware that AWS SMS is not supported by SAP

SAP export/import is the default process to change OS and/or database. To copy the exported files, you can use standard OS tools like SCP or AWS services like AWS Data Sync.

To keep the downtime as low as possible, you should consider allocating higher resources at source and target to run the export/import process in parallel

SAP Software Update Manager with Database Migration Option (SUM DMO) with system move option provides a one-step migration from other database systems to HANA. SUM offers near-zero downtime and zero downtime, and there are third-party solutions such as SNP to significantly reduce or eliminate technical downtime. However as lower the downtime as higher the price of the necessary infrastructure and consulting.

FAST SAP Migrations to AWS with the SAP Rapid Migration Test Program. SAP claim that customers can do a test with 2TB memory for just $1,000. Consulting costs are not included and $1,000 cover only 48 hours for the EC2 infrastructure, data transfer tool, network and storage cost. To give users a chance to try out the migrated system you have to add $2,800 per week.

For more see SAP Note 2377305 and HANA on AWS migration Tools and Methodologies

SAP system refresh on AWS

Quality assurance systems need to be refreshed on a regular base to grant that the same hiccups will happen during a test than in the production system. SAP system refreshes are considered to be time-consuming and error-prone tasks. There are several ways you can copy SAP production system data into QAS and test systems on AWS. A standard approach is to restore a backup from the production system, then copy it from Amazon S3 into the target system and start the recovery followed by **renaming the SAP system**, see SAPnote 1619720. Alternative options available only on AWS:

SAP system refresh using Amazon EBS Snapshots

AWS guarantees consistency of the snapshot when Amazon EC2 is shut down. File systems are created directly on the Amazon EBS volume. In this case, the outage for the QA system will be longer because you need to take a snapshot of the data volumes from the production system, bring them in, and create volumes before attaching them to the QA system.

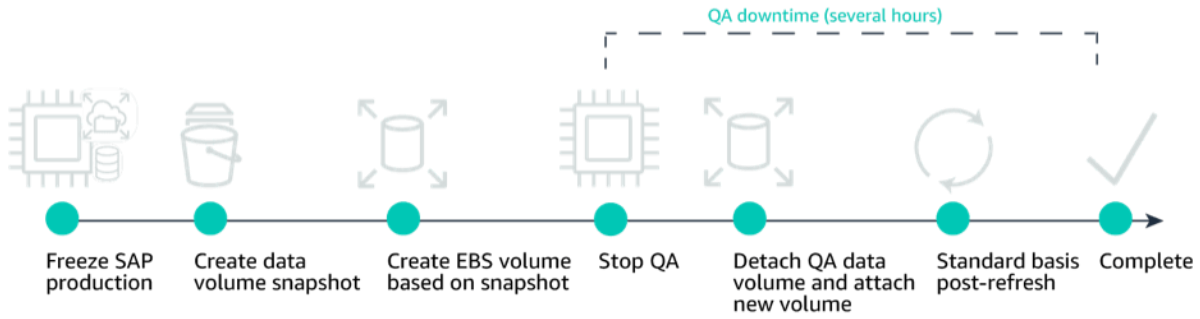


Figure 26: Serverless SAP system refresh

With AWS Step Functions and AWS Lambda, SAP system refreshes can be fully automated. See video

Tools to maintain HANA

Here are some helpful tools from SAP which are mostly unknown (maybe because they are free of charge and so there is no marketing).

HANACleaner is a tool that performs clean-up tasks like:

Cleanup of backup catalog entries	SAPnote 2096851
Cleanup of backups and backint.log	SAPnote 1642148
Cleanup of trace files	SAPnote 2380176
Cleanup of audit logs	SAPnote 2159014
Cleanup of HANA alerts and internal events	SAPnote 2147247
Cleanup of free log segments	SAPnote 2083715
Cleanup of multiple row store container	SAPnote 2222277
Cleanup of data file fragmentation	SAPnote 1870858
Creation of optimizer statistics	SAPnote 1872652
Optimize tables with columns	SAPnote 2112604
Cleanup of SAP HANACleaner logs	SAPnote 2399996

Implemented via Python script attached to SAPNote 1913302, designed by SAP support.

See the blog HANA Housekeeping using HANACleaner which provides guidelines on the necessary steps that need to be followed to automate housekeeping tasks using SAP HANACleaner script.

HANASitter pings the HANA system by default every hour. If the system is not available, HANASitter record traces for post mortem analysis. Otherwhise HANASitter check some critical features of HANA (see SAP Note 2399979).

Something you have to build yourself is a **HANA memory observatory** based on a SQL Script collection attached to SAPnote 1969700-Mini-checks. Monitoring long-term trends of HANA internal memory consumption is quite helpful to generate forecast and to avoid out of memory situations.

See also:

- FAQ: SAP HANA Performance Optimization
- FAQ: SAP HANA I/O Analysis
- How-To: Configuring SAP HANA Traces
- How-To: Troubleshooting SAP HANA Startup Times

Appendix A – Short description of SAP solutions

This section provides a highly condensed overview of the SAP solution portfolio and the underlying technologies as the basis for the discussion of how the individual components fit into the different cloud concepts.

If you can't find a three-letter acronym here, try Google. But be aware that during more than 40 years SAP and partners developed a nearly uncountable number of solutions, each of them praised by marketing as the next big thing. Many of them have been silently withdrawn from the market, so take a look at the date of entries. If the newest one is several years old, you can assume the product is obsolete.

The following list provides an overview of dependencies of SAP solutions on mandatory or optional components. For example, if a customer wants a S4/HANA system, the infrastructure always consists of a HANA database and an ABAP application server. However, a Fiori or archiving system is optional. While the HANA DB runs exclusively on Linux, you may also choose Windows for the app server (Linux is the recommended default)

Color code: mandatory; recommended; optional

SAP ECC, S/4HANA

SAP's classic solution for Enterprise Resource Planning (ERP[5]) was introduced as R/3 and later renamed Enterprise Central Component (ECC). S/4 is just the latest incarnation.

ERP deals with the fundamental business processes in every enterprise: financial accounting, production and human resources. In other words, among other functions SAP ECC ensures that orders can be accepted, fulfilled, tracked and paid for.

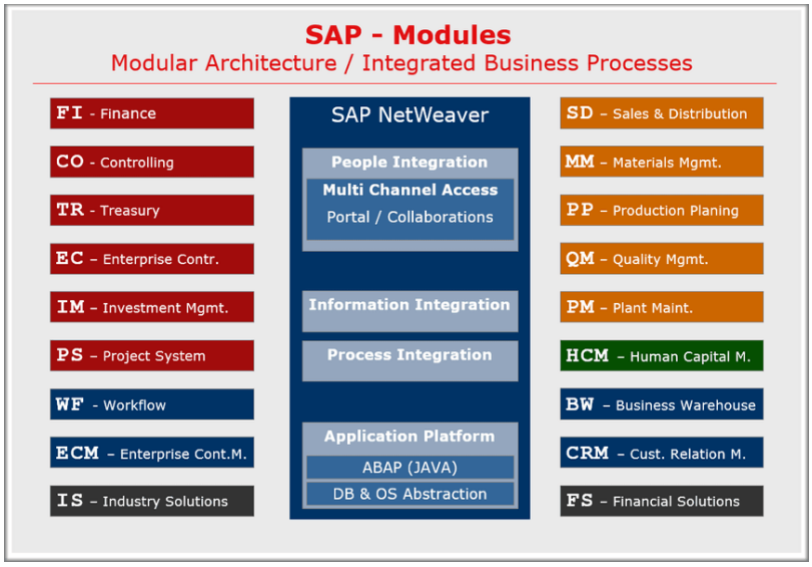


Figure 27: SAP modules within SAP ECC and S/4HANA

5 The term ERP was introduced by Gartner group and is not the name of an SAP product.

SAP ECC consists of several functional modules, each covering the demand of a specific line of business (LoB) in a company. The major ones are SAP Financials (FI), Sales and Distribution (SD), Production Planning (PP), Quality Management (QM), Warehouse Management (WM), Logistics Execution System (LES), Project Management (PS), and Plant Maintenance (PM). Human Resources (HR) or Human Capital Management (HCM) provide functions for payroll, incentives, statutory reporting and cost planning. SAP HCM also provides solutions for e-recruitment and e-learning.

Dependencies:

SAP ECC consists of several functional modules, each covering the demand of a specific line of business (LoB) in a company. The major ones are SAP Financials (FI), Sales and Distribution (SD), Production Planning (PP), Quality Management (QM), Warehouse Management (WM), Logistics Execution System (LES), Project Management (PS), and Plant Maintenance (PM). Human Resources (HR) or Human Capital Management (HCM) provide functions for payroll, incentives, statutory reporting and cost planning. SAP HCM also provides solutions for e-recruitment and e-learning.

Most of the SAP solutions described below are technically spin-offs from SAP ECC with enhanced functionality for individual LoBs. And yes, you are right if you think they are redundant to the modules!

- The ECC module SD (sales & distribution) and the SAP solution CRM both provide functionality for Sales;
- The ECC modules PP (production and planning), MM (material management), WM (warehouse management) and the SAP solution SCM (supply chain management) both provide functionality for production;

However the modules in ECC are focused on bookkeeping while the dedicated solution are focused on the functionality the individual LoB is asking for.

SAP also offers a wide range of industry-specific solutions (IS) for more than 25 industries, from IS-Apparel & Footwear, down to IS-Waste. These industry solutions consist of modified and extended ECC standard components. Some Industry solutions generate extraordinary loads. In IS-Retail for example, the typical business process is the analysis of the daily sales data collected by the Point-of-Sales (POS) systems in order to replenish the warehouse stock of the individual shops. As a result of an optimization run, the picking orders are placed and the delivery notes for the trucks are printed in accordance with the route plan. These processes are very CPU-intensive and also generate high IO loads. The same is true for monthly billing runs in IS-Telecommunication and IS-Utility.

Dependencies:

- Database: **HANA** or **ASE** or Oracle or **DB2** or **MSSQLSVR**
- CI/App Server: **ABAP**, JAVA in case you use Adobe
- Optional: Fiori (state of the art, tile-based user interface)

SAP Business Suite 4 HANA (S/4HANA) replace the classic SAP FI with Simplified Finance (sFIN) utilizing HANA to run trial Balance sheets, P&L's and Cash flow analysis any day of the quarter. sFIN merges the formerly separate datasets for external (FI) and internal (CO) transactions into a single dataset system to reduce the need for reconciliation and therefore time for financial closing significantly. It is also available as a add-on to “classic” ECC.

SAP provide a simplification list that documents application functionality changed, replaced or removed between ECC 6.0 and S/4. The SAP readiness Check, ABAP custom code analysis and the simplification item checks (SI-Check) are based on the list. See also the Blogs S/4 conversion – at a glance and The TOP simplification items. While the S/4 1909 simplification list describes over 600 simplification items on 1030 pages, typically 40 – to 80 are typical relevant for a customer.

- Database: **HANA**
- CI/App Server: **ABAP**, JAVA in case you use Adobe
- Optional: Fiori, Archiving (ILM retention management)

HR/HRM (Human Resource Management)

Many customers run the HR module of ECC on a dedicated system landscape to avoid a downtime of the complete productive ECC system for implementing “legal patches” necessary to adopt new legal rules only relevant for HR.

- Database: **HANA** or **ASE** or Oracle or **DB2** or **MSSQLSVR**
- CI/App Server: **ABAP**

Through the **SuccessFactors** acquisition, SAP offers cloud-based onboarding of employees, and performance and learning management, as well as HR analytics by HANA to allow the alignment to enterprise strategy. To grant proper payment and vacation management according to local laws, SuccessFactors needs to be tightly integrated into the ERP HCM.

SAP BW

As the first sibling of R/3, the SAP Business Warehouse (aka BIW) and Enterprise Warehouse Management (EWM) is still the most widely used SAP solution besides ECC. A dedicated data warehouse had become necessary because the performance of early platforms made it cumbersome to run transactions and reports at the same time on the same server.

In addition, the proliferation of SAP solutions dedicated to the different departments in an enterprise made it mandatory to consolidate the transactional data of all these individual systems for reporting and analysis. However, splitting off BW from ERP generates the need to extract, transform and load (ETL) data from the source system into the BW. This batch process puts such a high load on the source system that they run usually at night to avoid negative effects on user response times during daytime operation. So any reports derived from BW are always based on the “truth of yesterday.”

In contrast to an Online Transaction Processing system (OLTP) where all data records are stored in a normalized form for the sake of consistency, BW is an Online Analytical Processing system (OLAP), based on special data structures, known as multidimensional InfoCubes.

Roughly speaking, a dimension in an InfoCube corresponds to a classic report. In addition to the source data it also provides key figures like sums and averages. These figures are calculated directly after the data is loaded and is therefore immediately available when a user makes a query. This way a pre-defined report is displayed within seconds after a Line-of-Business manager calls it from his dashboard.

However, the design of multidimensional InfoCubes is a time-consuming task and needs specialized skills and experience. It can take weeks for a new type of report to become available to management. To solve this issue, SAP developed the in-memory-based Business Warehouse Accelerator (BWA) followed by HANA.

- Database: **HANA** or **ASE** or Oracle or **DB2** or **MSSQLSVR**
- CI/App Server: **ABAP**, JAVA Instance when BO Business Explorer (BEX) is used.

BW4/HANA

Just the newest, naturally HANA-only incarnation

- Database: **HANA**
- CI/App Server: **ABAP** (no need for Java because of BW/4HANA which simply does not support BEX
- Optional: Fiori, NLS (near line storage)

SAP Customer Relationship Management

Customer Relationship Management (CRM) is also found in many customer installations in spite of SAP's newer acquisitions of Hybris and Qualtrix customer experience (CX) solutions. CRM provides processes for interactions with customers such as marketing, sales, service and support transactions. Technically an offspring of R/3, the SAP CRM core is coded in ABAP. However, there are also some JAVA-based components.

SAP CRM can become quite complex. For example, the call center solution, Customer Interaction Center (CIC), requires a telephony gateway, a telephony server and a corresponding interface to the private branch exchange (PBX), the company's telephony system.

Even more complex are web shops. In addition to the basic SAP CRM system, the SAP Internet Sales scenario also consists of SAP Internet Pricing and Configurator (IPC), SAP Biller Direct, a catalog system, Knowledge Provider (KPro), permanent shopping basket, etc.

- Database: **HANA** or **ASE** or Oracle or DB2 or MSSQLSVR
- CI/App Server: **ABAP**

SAP Hybris

The acquisition of Hybris added multichannel and product content management software. The Hybris infrastructure consists of a web server, application server and database engine, all of which are business processes running on JVMs. By default, SAP Hybris uses Apache Tomcat as its application server

Hybris contains the following components:

- A web server for static content and directing dynamic requests, such as Apache or nginx.
- A Java application server for running the SAP Hybris core application.
- An optional caching mechanism for page and session caching, such as Redis.
- An optional search mechanism for product search, such as Apache Solr.
- Database: **HANA**, **ASE**, Oracle, **DB2** or **MSSQLSVR**, default HSQLDB — a light-weight SQL database that can run within a Java Virtual Machine.

The JVM of tomcat app servers need 4 vCPU and at least 6GB (but 10 to 14GB is better). In larger clusters, use 8 vCPU for each instance and cut the number of tomcat nodes in half.

The acquisition of Gigya added Customer Identity and Access Management (CIAM), enabling customers to fulfil the rules of DSGVO result in 3 new Hybris solutions:

Hybris Identity: Customer registration / authentication and multi-platform single sign-on.

Hybris Consent: Manage customer approval of to place cookies and receive marketing according to European DSGVO law.

Hybris profile: Store integrated customer profiles in a centralized customer database.

SAP C/4HANA

To counter the proliferation of salesforce.com, SAP introduced C/4HANA as CX-Suite (Customer eXperience) by combining existing Hybris modules and some acquisitions. Claimed to represent fourth generation of customer relationship management, C/4 consist of five components

1. SAP marketing cloud; formerly Hybris marketing
2. SAP commerce cloud; formerly Hybris commerce with components PCM, WCM, and DataHub
3. SAP sales cloud; former Hybris sales cloud consisting of SAP C4C sales (aka cloud for customer), including revenue cloud and the Callidus Cloud acquisition
4. SAP service cloud; consisting of C4C service (aka cloud for service) and the newly acquired coresystems
5. SAP customer data cloud; by acquisition of Gigya

They are all brought together via the SAP Cloud Platform (SCP).

SAP Data Hub (aka, on SAP Vora 2.0) is an attempt to occupy the big data market by running SAP analytical software on Hadoop.

SAP Upscale Commerce

Upscale

is a SaaS-based unified commerce platform that leverages AI-powered retail merchandising specifically designed to mid-market brands. Introduced in October 2018, Upscale Commerce seems still to be in incubator mode.

SAP SCM/APO

SAP supply chain management (SCM) also has a high number of scenarios, and SAP's first incarnation of an in-memory database is the APO LiveCache.

SAP Advanced Planner and Optimizer (APO) is the core component of SCM, covering forecasting future requirements on the basis of historical data by Demand Planning (DP), optimization of cross-plant distribution of orders onto the available transport, and production capacities by Supply Network Planning (SNP), Production Planning-Detailed Scheduling (PP-DS), Transportation Planning–Vehicle Scheduling (TP-VS), Vendor Managed Inventory (VMI) and Availability-to-Promise (ATP). For example, optimization can include providing a multilevel

availability check that can be carried out against material stocks, production, warehouse and transport capacities and costs across plants, etc.

All of these business processes demand complex optimization runs, which need extremely fast access to data that is impossible to achieve with hard disks. For this purpose, SAP developed LiveCache, an in-memory database based on MaxDB (previously called SAPDB).

Since SCM 7.12, APO liveCache can be migrated to HANA in an integrated SCM on HANA.

Other SCM components are SAP Event Management (EM), providing functions for managing deviations between planning and reality, and the SAP Inventory Collaboration Hub supporting cross-enterprise integration for Supplier Managed Inventories (SMI) or Vendor Managed Inventories (VMI).

The SAP Auto-ID Infrastructure (AII) provides connectivity of RFID scanners to SAP SCM and can generate extremely high IO loads. SAP Object Event Repository (OER) featuring PTA (Product Tracking & Authentication (PTA) record uniquely identified objects (like EPCs) detected by AII.

- Database: **HANA** or **ASE** or Oracle or **DB2** or **MSSQLSVR**
- CI/App Server: **ABAP**

SAP Supplier Relationship Management

Supplier Relationship Management (SRM) was the SAP component for purchasing and procurement departments, covering the complete process from placing the order to paying the invoice. The core component was SAP Enterprise Buyer Professional (EBP), enhanced by a catalog server. Optional components are SAP Content Integrator, SAP Bidding Engine for online auctions, SAP Supplier Self-Services (SUS) and SAP Live Auction Cockpit Web Presentation Server (LAC WPS) for online auctions (implemented as a Java applet).

- Database: **HANA** or **ASE** or Oracle or **DB2** or **MSSQLSVR**
- CI/App Server: **ABAP**

With the acquisition of Ariba, SAP added a SaaS solution for sourcing and payment processing. Many customers combine an on-premises SAP SRM system with Ariba in a hybrid model. In any case Ariba needs to be integrated into the SAP ERP system for bookkeeping.

The same is true for business travels booked via Concur and contingent workers leased and managed by Fieldglass — they need integration into the ERP bookkeeping system, so that bills can be paid and accounted correctly.

SAP PLM

SAP Product Lifecycle Management (PLM) is the SAP solution for product development, plant maintenance and quality assurance as well as hazardous substance management, industrial hygiene and safety and environmental protection. Although SAP PLM is a standalone solution, it is not an SAP system of its own. Instead it uses a combination of functions from SAP ERP, SAP CRM, SAP SCM and other components. Therefore, rather than requiring its own infrastructure, PLM is usually simply installed as an add-on to SAP ECC. The knowledge warehouse (KW) can be used to store and distribute large files like scans, CAD drawings, video files, etc.

SAP CPM

Corporate Performance Management (CPM) is part of SAP Financial Performance Management and has replaced Strategic Enterprise Management (SEM). It includes SAP Strategy Management through acquisition of Pilot Software, Business Planning and Consolidation (BPC) through acquisition of OutlookSoft and several other components from partners.

Due to the fact that these solutions are not based on the SAP standard web application server special care is necessary to implement them on public or private cloud infrastructures.

SAP GRC, CCM, GTS and EH&S

SAP Government Risk Compliance (GRC) is the solution for Sarbanes-Oxley Act (SOX) compliance. Continuous Controls Monitoring (CCM) helps ensure you're in full compliance with external and internal policies. The Global Trade System (GTS) ensures that companies don't export something that is on a black list for certain countries (in addition to do the custom declaration or even checking the financial status of the customer). Environment, Health & Safety (EH&S) manage the documents required for industrial hygiene and safety, and environmental protection, such as material safety data sheets, TremCards and waste manifests.

All components are based on the standard SAP web application server. Even Access Control (formerly Virsa Compliance Calibrator), a segregation of duties auditing software is coded in ABAP.

- Database: **MSSQLSVR**
- App Server: **ABAP**

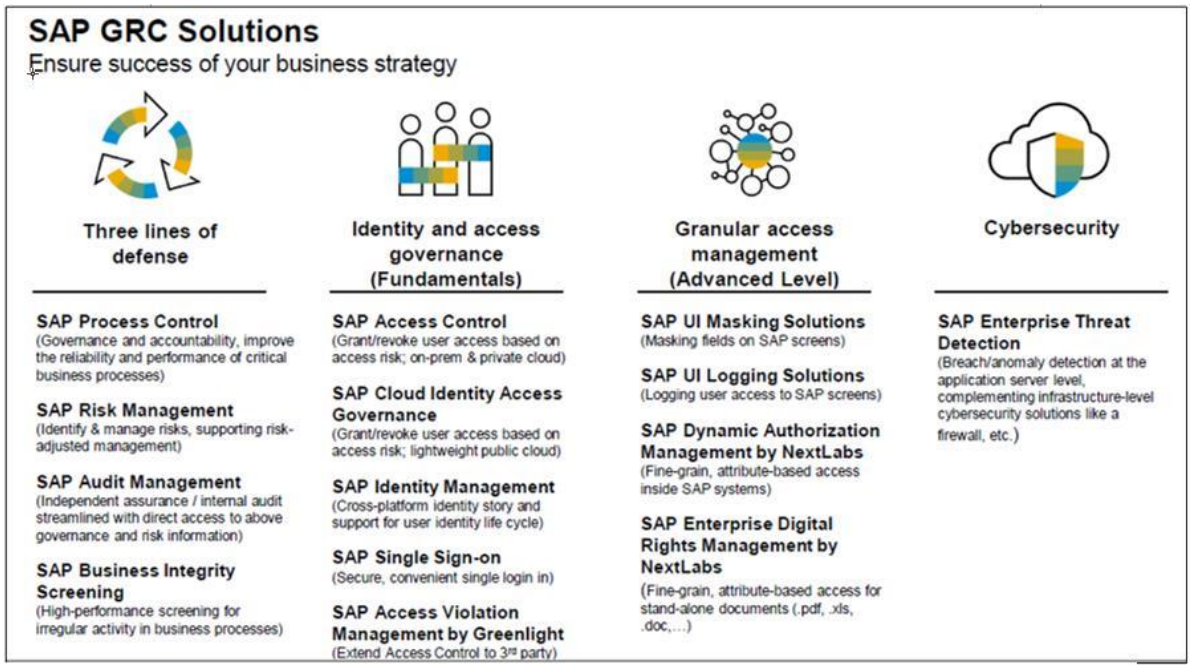


Figure 28: New SAP GRC portfolio

SAP SSO, IDM and DAM

SAP single sign-on (SSO), SAP Identity management (IDM) and SAP Dynamic Authorisation Management by **Nextlab** (don't confuse with Digital Asset Management by **Opentext**)

- Database: **HANA** or **ASE** or Oracle or **DB2** or **MSSQLSVR**
- App Server: **JAVA**

SAP ME/MII

SAP Manufacturing Integration and Intelligence (MII) and SAP Manufacturing Execution (ME) are the "classical" solutions to connect SPS and SCADA systems to SAP ERP and SCM. Together, with SAP Plant Connectivity (Pco) and a device management solution acquired from plat.one, they become part of the Leonardo IoT portfolio.

- Database: **HANA** or **ASE** or Oracle or **DB2** or **MSSQLSVR**
- App Server: **JAVA**

SAP Portal

The Enterprise Portal (EP) provides Single Sign-On (SSO) to SAP backend solutions with via browser. Implemented in Java, consumption of CPU resources can be quite high, especially for the connection to directory services like Microsoft Active Directory (AD). To enable Single Sign-On, the directory service must be sized adequately for hundreds of users logging into the SAP system at the same time.

With SSO functionality moved to Solution Manager, the SAP portal has become mostly obsolete. However, EP will stay but in conjunction with the Fiori frontend server

- Database: **HANA** or **ASE** or Oracle or **DB2** or **MSSQLSVR**
- App Server: **JAVA**

SAP TREX

In addition to Portal, SAP also developed its own search engine, called Text Retrieval and Information Extraction (TREX). Besides searching unstructured data, TREX can also index structured business data which made it a key component of BWA and later HANA.

Like LiveCache, TREX loads all its data into main memory. But TREX can be distributed horizontally over any number of blades necessary to provide the needed memory. For such a distributed system, the file system must be a clustered or shared file system to present all files to all nodes.

SAP Knowledge Warehouse

Knowledge Warehouse (KW) is used to create, manage and distribute documents, training materials and manuals, and so on. Large files can be replicated to local cache servers to improve access speed and to reduce network load over remote connections. This way a local server at a customer's premises can accelerate access times of files stored at a SAP KW hosted by a cloud provider.

SAP Mobile Platform

SAP has consolidated multiple acquisitions to the SAP Mobile Platform 3.0 (SMP3)[6]: Sybase Unwired Platform (SUP), Agentry (through the acquisition of SYCLO) and Mobiliser (through Sybase 365).

The enterprise mobility management suite (EMM), called SAP Mobile Secure includes mobile device management (MDM) by SAP Afaria, Mobile App Management by SAP Enterprise store and SAP Mobile App Protection for app wrapping, SAP Mobile Documents for Mobile Content Management, and Telecom expense Management built into SAP Afaria and offered by partners.

SAP MDM and MDG

As the result of mergers and acquisitions, enterprises are faced with the need to eliminate duplicate master records to grant consistent reporting. SAP Master Data Management (SAP MDM) is designed to tackle this problem. Master data from different sources can be compared semi-automatically to identify duplicate master data records. Within MDM, SAP Master Data Governance (SAP MDG) centrally creates, changes and distributes master data to multiple targets.

- Database: **HANA** or **ASE** or Oracle or **DB2** or **MSSQLSVR**
- CI/App Server: **ABAP**

SAP PO/PI/XI

The problem of Enterprise Application Integration (EAI) has existed since companies have used more than one computer. While direct database access seems to be a quick and easy solution it becomes a nightmare from a security and audit point of view. SAP Process Orchestration (PO, former Process Integration (PI), which was formerly exchange infrastructure (XI), acts as central data hub for interconnecting business processes across the boundaries of an enterprise.

Database: **HANA** or **ASE** or Oracle or **DB2** or **MSSQLSVR**

App Server: **JAVA**

6 global.sap.com/campaigns/digitalhub-mobile-platform/index.html

SAP Solution Manager (SolMan)

SAP Solution Manager (rarely called SSM) is the central monitoring system for a customer's SAP landscape. While SolMan provides deep drilling capabilities of SAP NetWeaver, only rudimentary HW- related capabilities are available based on the so-called SAPOSCOL agent, which has to be installed on all nodes.

As part of the SolMan, the Central User Administration (CUA) enabled Single-Sign-On (SSO). While CUA was a relatively quick and easy solution, it had some restrictions. Thus, SAP changed the approach to SAP NetWeaver Identity Management (IDM) as strategic recommendation for managing users across SAP systems. CUA is still supported, but it's a dead-end road.

SolMan is also instrumental in generating the installation keys necessary to install a SAP instance and enabling SAP support to access customer systems remotely. A SAP SolMan instance is mandatory for every SAP customer. Concepts where several customers shared a solution manager of a service provider are no longer tolerated by SAP.

- Database: **HANA** or **ASE** or Oracle or **DB2** or **MSSQLSVR**
- CI/App Server: **ABAP** and **JAVA**. Due to a bug in JAVA for Solution Manager 7.2, you need 256GB of HANA memory during the installation. After that the memory can be reduced to 128GB.

You don't need a Solution Manager anymore just to generate SAP license keys. In some cases, the SAP maintenance planner tool allows you to add systems by uploading the system information that XML generated for a technical system. However, we do not recommend it for the following reasons:

- It does not provide a holistic view of all the systems present in your landscape
- It does not include crucial SAP SolMan functions

So the recommended approach to upload and add systems in the maintenance planner tool is through SAP Solution Manager.

SAP LaMa

Landscape Manager — aka, Landscape virtualization manager (LVM), Adaptive Computing Controller (ACC) — enables start and stop of SAP and DB instances that still have to be installed in the "traditional way." Even though SAPnote2050537 talks about provisioning, this is restricted to cloning of PRD systems and the necessary post-processing to convert the clone into a QA system. You need an enterprise license to use this feature. LaMa capability depends on the installation of the SAPOSPREP agent on all servers and vendor-specific agents on FC-based storage devices. For HANA as a database the hdbclient must be installed on each HANA host.

SAP Business Objects (BOBJ)

With the Business Objects (BO) acquisition in 2007, SAP got another suite of business intelligence tools. User queries are executed on the BO server, which in turn fetch loads of data from the source databases. Despite the fact that BO uses a three-tier architecture, it's not based on classic NetWeaver runtime. BO is composed as such:

- Reporting: Crystal Reports
- Ad hoc queries: Web Intelligence (WebI)

- Data exploration: Explorer
- Dashboards: Xcelsius
- Data base-middleware: Universes XI

For more details, see the SAP BusinessObjects Enterprise Administrator's Guide

- Database: **MSSQLSVR**
- App Server: **Windows**, 8 vCores; 32GB RAM

SAP Fiori/NetWeaver Gateway:

- Database: **ASE**
- OS: **Linux**, 2 vCores; 8GB RAM

SAP Content Server (CS):

- Database: **ASE** (on same servers as application)
- OS: **Linux**, 4 vCores; 16GB RAM

ADS (Adobe Document server): be aware that you have to pay license fees if you deploy interactive forms

- Database: **ASE**
- App Server: **JAVA**, 2 vCores; 8GB RAM

SAP All-in-One

Basically a SAP ERP system with some functionality from BI, CRM and SCM delivered with more than 80 pre-configured, industry-specific solutions from various partners. An option for customerd wanting a cheaper solution than the full suite and accept that the solutions are reduced to the core processes of each particular industry.

SAP Business One

An ERP solution for small businesses was acquired in 2002 from TopManage. SAP Business One (B1) runs in a small VM, can use HANA and has no code or architecture common with the classical SAP applications.

SAP Business ByDesign

BBD is SAP's attempt to provide SaaS-based integrated SMB business management software. The state-of-the-art code was written from the ground up. Hosted exclusively by SAP, BBD is built on the principles of a service-

oriented architecture (SOA). The underlying technology stack is a multi-tenancy-enabled SAP NetWeaver stack, leveraging SAP's in-memory HANA database.

The User interface from SAPGUI & WebGUI to Fiori & Lumira

The classical user interface SAPGUI has to be installed at the end user's workstation. Alternatively the browser-based WebGUI can be deployed with a SAP internet transaction server (ITS) at the backend. To replace both altogether with BusinessObjects, SAP has introduced a new proprietary user interface toolkit called UI5. Based on this toolkit, SAP Fiori provides a collection of mini applications to access SAP software functions. Like its ancestor, the SAP portal, Fiori need a gateway server, a ABAP application server. Lumira is an analytic tool for users who yet don't know what they 're looking for yet, but also just want to slice and dice the bigger datasets.

Appendix B – NetWeaver architecture

The diagram below depicts the architecture that makes up a SAP System: Database Instance (bottom), Central Instance (middle), ABAP Application Server (left), Java Application server (right) and ICM (top).

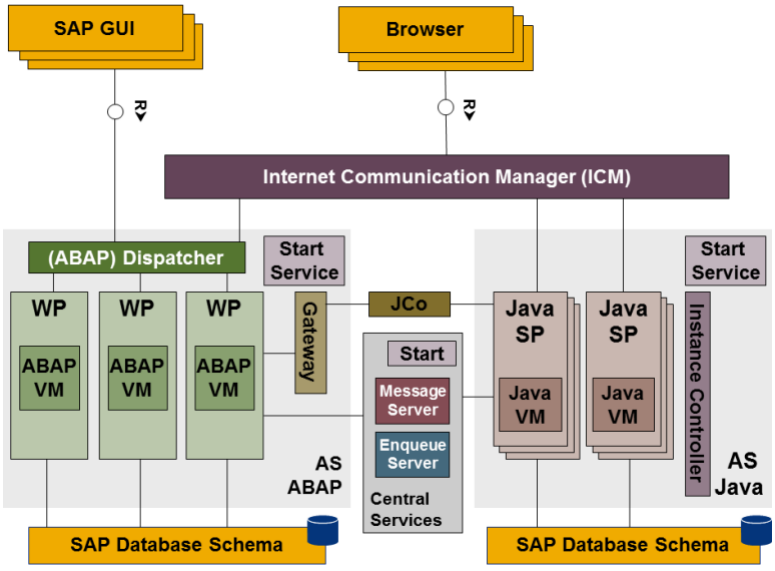


Figure 29: Components of a classical SAP system

The majority of SAP solutions only need the ABAP, and only a few need both (see Appendix C).

ABAP Central Instance, which is also referred to as Primary Application Server (PAS)

- **ABAP SAP Central Services (ASCS):** Made of the following two processes:
    - The message work process distributes users at logon to the app server with the lowest load at this time. This allows for the scale out of the SAP system.
    - The ENQUEUE work process maintains the lock table to secure the consistency of SAP business transactions stretching across multiple database operations. There is only one message and one ENQUEUE process needed per ABAP system. If you lose the message service, nobody can log in — all you have to do is to restart the message service. If you lose the ENQUEUE table, the whole system has to be rebooted to avoid inconsistencies.
- To mitigate the loss of an ENQUEUE table, HA options are available with a Master/Passive approach called “replicated ENQUEUE server (ERS).” This approach requires disk sharing and additional OS configuration.
- **Central Instance (CI):** The best practice is to run ASCS and PAS on the same host/VM.
- Gateway: The Gateway Process controls internal communication as well as communication with other SAP and non-SAP systems. Don’t mix up with SAP Gateway server.

- **ABAP Dispatcher:** The local instance “traffic cop” managing the worker threads (work processes) of the application server. Dispatchers from different application servers do not talk together (Hub Spoke),but talk with the message server of the ASCS to handle requests.
- **Work Processes:** Work Processes are the actual threads that perform the work requests. There are different types of work processes.
- **Dialog:** Processes the business transactions that end user requests. This could be recording a sales order or running a monthly roll up report.
- Update (1 and 2): Async Database changes (performs the DB commit request of Dialog/Background Work processes.
- **Batch:** Executes time-dependent or event-controlled background jobs
- **Spool:** Print formatting (to printer, file or database)

JAVA Central Instance:

Literally the same architecture as ABAP providing the same functionality, but in Java. A few SAP solutions like portal and PO run on Java (see Appendix C).

- **SAP Central Services** (SCS) with Message and ENQUEUE.
    - Also a best practice to run SCS on the same host/VM as the Primary Application Server.
- **SAP JVM:** Used to launch the Java Server Processors.
- **Java Controller:** This is the cluster manager process that is designed.
- **Java Server Nodes (processes):** These processes are the worker threads that load specific services. Typically, each node approaches 4GB without have negative performance results. Scale out design is approved by SAP.
- **SAP Deployment Manager (SDM) Process:** This is the process used to apply new code to the software components that make up the Java Server functionality.
- **Internet Communication Manager (ICM) Service:** HTTP Traffic handler.

Additional Application Server (AAS): Either ABAP or Java SAP systems can be scaled out with additional application servers which allow for handling the required processing load.

NetWeaver Java needs usual more resources than ABAP best practice: 8 vCores and 32GB for PRD, 2 vCores and 16GB for DEV and QAS.

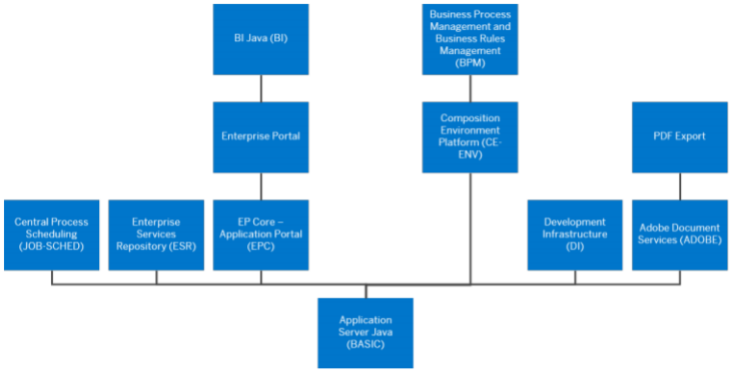


Figure 30: SAP solutions running on Java and their dependencies

Two-tier architecture

SAP classic architectures are based on three layers: Database, application server executing the business processes and End User front end. While the End User front end (either the locally installed SAPGUI or browser-based solutions) always runs at the customer (or a Citrix Farm), very small systems can run the database and the application server together on the same VM. This is called a 2-tier architecture or centralized deployment.

The VM needs five attached disk drives, each serving a specific role:

- **Root disk:** Contains the OS for the VM.
- **Swap disk:** Contains the OS's paging file.
- **SAP NetWeaver:** Contains the NetWeaver installation and the profile files.
- **Data volume:** Contains the database files.
- **Logs volume:** Contains the database-system logs used for maintaining data consistency, backup and recovery operations.

The required volumes for the database might be different depending on which database you're using

(for SAP HANA, see the HANA Planning Guide):

The SAP NetWeaver application server contain the HANA binaries and shared files.

You need an additional volume for storing database backups.

Please read the Installation Guides — Application Server ABAP with HANA Database on a Single Host (HANA_and_ABAP_on_One_Server.pdf and HANA_and_ABAP_on_One_Server_Replication.pdf).

Three-tier architecture

However, in most cases you don't want to increase an already large database VM further for the demand of application servers. Therefore, the most common architecture distributes the SAP system across multiple VMs with a dedicated VM for the database. All the NetWeaver AS nodes mount and access a shared file system that hosts the SAP NetWeaver profiles. This shared file system is contained on a persistent disk that is attached to VM 1, along with the SAP central services.

- Separate subnet for application tiers and database tiers
- When you have different subnets you can introduce different firewall rules
- None of these VMs should have IP addresses, only through internal communication

SAP application server and database must be in the same region and must not have a firewall in-between (you would not get extra security; as bad guys could capture the credentials of a SAP user and start a report to get data from the application server which would fetch them from the database evewn through a firewall).

Pro 2-tier — "Keep it as simple as possible"

- Complete system contained into a single VM
- Less effort for administration, hibernation and failover

- Less latency by omitting network traffic between DB and CI

Pro 3-tier — “Keep it as isolated as possible”

- Use of cheaper instance types for CI, especially with HANA DB
- Can scale to much higher user numbers
- Can dedicate app servers to usage types (batch vs. dialog), users group (per department or country), or connected solution (e.g., BO) to ease maintenance
- In two-tier config you can’t control the resource distribution in case either the DB or the CI starts to “eat the others breakfast” — for example, due to a memory leak

In summary, three-tier gives you more control, scalability, manageability and resilience

How many app servers?

- At least one (the CI aka PAS aka ASCS).
- As a rule of thumb a app server can support approximately 200 to 300 users and needs 100-150 MB of memory to keep their context.
- The internal dispatcher that can handle approximately 200 user requests per second. If overloaded, a dispatcher dialog queue will build up (check with transaction SM51). In this case, an additional application server becomes necessary to avoid prolonged response times.
- Thanks to faster CPU and multi-threading, you may reduce the number of app server instances when migrating from obsolete on-premises hardware to state of the art AWS instance types, as long as you increase the memory according to the higher user numbers.

How much resources per app server?

ABAP (CI, ASCS, AAS):

- 4 vCores; 16GB if NetWeaver gateway* run on dedicated server
- 8 vCores; 32GB if NetWeaver gateway* run on this instance
- Add 1 vCore; 4GB if a cloud connector run on the app server to use Fiori** out of SCP (SAP cloud platform)
  *SAP NetWeaver Gateway is used to setup a connection between SAP and target clients, platforms (like salesforce). using OData.
  **Fiori will add approximately 5% load on the application servers, which can be noticeable when you have many users are logged in.

JAVA (SCS, Java AAS):

- 8 vCores; 32GB (default)

How dynamic can the number of app servers be?

The distribution of users to app servers is controlled by the message service residing on the CI. When a user is logging onto the system, the message service sends him to the app server with the currently lowest CPU load. There is no rebalancing at a later time.

- You can add application servers at any time to cope with growing demand. As soon the CI becomes aware of the new AS, new users logging into the system will be sent to this AS. However already logged-in users stay on their overloaded app servers until they log out and log in again.
- However, when the load is going down you have to wait until the last user on an app server has logged out before you can shut the app server down. (In theory you can send a notice to the remaining ones and tell them to log out and wait until you have disabled the server before logging in again, but this is not practical in real life.)

SAP System Landscapes

Due to the fact that SAP systems execute mission-critical business transactions for most customers it would be unwise to implement patches, change customizing and develop customer code directly on a production system.

Therefore, every SAP implementation consists of systems dedicated to development (DEV) and test/quality assurance (QAS) in addition to the production system (PRD). Some customers deploy additional systems for training, sandbox, consolidation, pre-production, etc. All of these together are called the SAP System Landscape.

Code developed on the DEV system is moved first to QAS to be tested before being deployed on PRD. The SAP transport management system (TMS) controls the movement of code through the landscape. This has to be considered in hybrid models where DEV and QAS are deployed on AWS and PRD on private cloud. DEV needs another 100GB of diskspace for the ABAP code repository.

During migration downtime you export the customer data from the PRD system and the customer code from DEV to flat files and import to new build instances on the AWS. QAS and other non-production systems will be generated by a SAP system copy from the PRD afterwards. Also during operation, it is recommended to refresh the QAS system approximately every six months.

While the PRD system has to be up and running 24x7x365 in every case due to global business and nightly batch runs, in most cases the non-production systems can be hibernated to save costs. In principle they have to run only during business hours, however, exceptions need to be established with the customer for cases where developers need the systems at night and weekends for emergency patches.

Appendix C – Other terminology

Although not an exhaustive list of terminologies, the list below and in the other appendixes provide a good understanding of SAP sage and architecture patterns.

ABAP: SAP's own language based on COBOL to code the business processes independent from the underlying platform. The ABAP application server code is mostly written in C+. Customers can write their own code under the condition that the name starts with a Z, the so-called "Z-ABAPs."

SAPS: The horsepower of SAP. Basis of CPU sizing. Hardware vendors and hyperscalers have to run a benchmark with a certain set of SAP transactions to determine the SAPS rating of a certain server or instance type published at the SAP benchmark page.

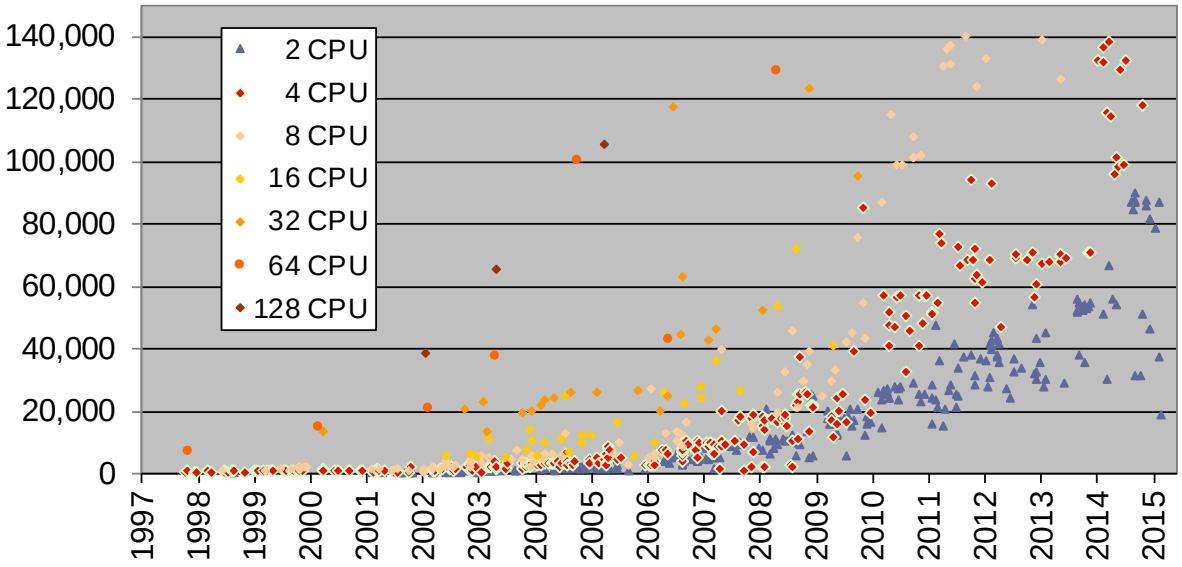


Figure 31: SAPS numbers achieved in benchmarks over time. Thanks to advances in CPU development, a single 2 socket server today can run a system that demanded a 64-socket superdome a decade ago.

In theory it's not possible to convert a SAPS value into a general value for CPU cores. In reality, you select a benchmark with a CPU you want to deploy and divide the SAPS rating of the server by the number of sockets and cores of the machine. See: www.sap.com/benchmark.

SAPS mathematics:

- SAPS of the application part can be divided over multiple VMs while the SAPS for the database have to be delivered by a single VM.
- 2-tier: In cases where a DB and app server run in the same instance type, the SAPS of DB and app server are simply added up.
  Example: A DB with 10k SAPS and an app server with 4k SAPS can share a single n1-highmem-16 with 14,670 SAPS.

- 3-tier: The total number of SAPS for application servers can be distributed over several instance types which have in total of least the same SAPS. The database just need a instance type with a SAPS rating large enough (exception: HANA scale out for BW).
  Example: A demand of 10k SAPS for app server can be distributed over three-times n1-highmem-4 with 10,740 SAPS in total.
- Rightsizing: If the EWA report reveals extremely low CPU use (below 20%), you may reduce the number of SAPS to reach an average of 40% to 60%. Make sure with the customer that the oversizing was not intentional to cover peak situations. In case the peak is only once a month or year you may add another instance just for the time of the peak.
  Example: An app server rated for 8k SAPS with a 15% utilization in the EWA report will fit easily into a single n1-highmem-4 for most of the time. During peak time you simply add another n1-highmem-4.

Platform Availability Matrix (PAM): hardware and platforms certified for HANA

SAP Basis: Mostly refers to the team responsible for the administration of the technical components of SAP solutions between the OS and the SAP GUI. This include monitoring, updating and patching, and tuning of the DB and app server, as well as OS/DB migration (for example, to AWS and/or HANA). However, no configuring business processes or code development is required.

SAP Development: Mostly refers to the team responsible for the adaption of the pre-packaged SAP business processes to the demand of the individual company by writing ABAP custom code (called Z-ABAP's because the names must start with an Z).

SAP router: A standalone program acting as:

- A proxy in connections between SAP systems, or between SAP and external networks.
- An extra firewall to the existing firewall (port filter).
- A gateway, through which connections to your firewall-protected system can be opened. It is usually installed directly on the firewall host.

SAP WebDispatcher: Entry point for HTTP(s) requests, load distributors and reverse proxies. Runs on a ABAP app server, but in the DMZ.

SAP Cloud Connector: Connect applications in the SAP HANA Cloud Platform with applications installed on AWS or on-premises. Due to reverse invoke support, no ports have to be opened in the on-premises firewall to allow external access from the cloud to internal systems. Run on Windows only, 2 vCPU, 4GB memory and 20GB disk.

SAP Landscape Transformation (SLT): SAP'd first ETL tool that loads and replicates data from a non-HANA Source System into the HANA Database, while migrating the data in SAP HANA format. SLT servers can be installed embedded on SAP ECC systems or on the separate ABAP application servers. See SLT server sizing guide. Don't mix up with HANA database replication, which is an HA configuration.

Fiori 2.0: SAPs new user interface replacing the classical SAPGUI. It's based on a predefined HTML5, CSS3 and JavaScript library. SAP offers more than 10 thousand apps in the Fiori reference libraries that run in any HTML5 capable browser on any device, even off-line (needs local data persistency). The SAP NetWeaver Gateway Service (aka front-end server, Fiori gateway) translates between internet standard protocol OData and SAP

Internal RFCs. The Gateway runs either embedded into ECC or on a dedicated ABAP app server serving several SAP solutions (central hub). It can run together with the Web Dispatcher on the same VM. A system landscape with Dev, Qas and Prd is recommended.

RFC: The ABAP communication method for SAP systems to talk to one another.

Jco: The Java communication method for SAP systems to talk to one another. Can also be used to integrate approved third-party applications.

RZ10: One of several thousand so called transaction codes that allow experienced SAP users to reach a certain page much faster than clicking through menus.

HANA Studio: Eclipse-based development and administration tool in the form of GUI tool. It needs a windows-based VM. See HANA Studio Installation and Update Guide.

HANA Cockpit: The successor of Studio, see HANA 2.0 Cockpit Installation and Configuration How to Guide. Comes as a separate SAP HANA system with at least 16GB RAM, running a special version of HANA, express edition, with an XS advanced runtime environment included, so there is at least no need for an app server. You can't deploy cockpit on an existing SAP HANA instance nor can you deploy XSA applications to HANA Cockpit. It is recommended for a production environment to install the SAP HANA Cockpit on a dedicated hardware. For non-production, you can deploy on an existing SAP HANA server as discussed in the SAP HANA Cockpit Installation and Update Guide.

Multi-tenant Database Container (MDC): A lightweight alternative to virtualization. While the individual containers are called tenants, the basic "container ship" is called the SystemDB. The SystemDB keeps the system-wide landscape information and provides configuration and monitoring system-wide. The tenant databases are isolated from each other in regard to application data and user management.

All tenants run with the same HANA version (revision number). HANA system replication works with an "all or nothing" approach — all tenant databases are subject to failover to another data center. Each tenant database can be backed up and recovered independently from one another. SAP made MDC mandatory even if you have only one tenant (SAPNote 2423367).

Scale-out: For SAP BW on HANA the database can be distributed over multiple nodes in a master-slave approach. This way smaller instance types can be used, however some performance penalties apply.

Scale-up: The only option for all other SAP solutions on HANA. According to practical experience SAP BW performance is better in scale-up.

I**Q:** Former Sybase Intelligent Query is a column-oriented disk-based OLAP DB.

ASE: Former Sybase Adaptive Server Enterprise is a row-oriented disk-based OLTP DB. Technically, the basis of the Microsoft SQL server. But while Microsoft invested in further development, Sybase didn't and SAP changed much on this. It will be offered by SAP as an alternative if customers complain about the cost of HANA.

MaxDB: Previously called SAPDB. Acquired 1997 from Software AG, who called it ADABAS-D.

LoB: Line of Business, the departments in a company dedicated to different tasks in a company like sales, finance, procurement, production, logistics. SAP sells its solution to top management and heads of LoB's rather to the IT department which is usually only aked afterwards to make it running.

Appendix D – HANA for Dummies

Based on the scientific research at the Hasso Plattner Institute[7] (HPI), SAP combined the technologies of TREX, MaxDB[8], and Ptime to expel Oracle, its arch-enemy from the SAP realms. Some sources claim that the name HANA stands for "Hasso's New Architecture."[9]

Part of the popularity of today's web search engines results from the fact that answers appear on the screen even before the user finishes entering the question. Why is such magic not available for enterprise business applications?

The answer is the difference in the necessary completeness of the result between enterprise applications and surfing the web. The response of a web search displayed on the first page represents only the hits rated most relevant for the query.

A legal business report, however, must reflect all relevant data in its result. Whereas a web search just has to sift through an indexed data set, the business application has to scan the complete dataset to guarantee completeness in addition to processing such complex aggregations.

Search engines like Google can be so astonishingly fast, because its searches just has to be "good enough" for the common user.

However, no tax authority will accept payment based on anything other than a complete scan through each and every accounting number. Therefore, some more advanced technologies than "just in-memory" are necessary to derive the complete answer from a business-grade system with the speed of thought.

In-memory: A no-brainer?

Given the fact that access to data in a computer's memory is orders of magnitude faster than to the data stored on disk, the concept of in-memory computing seems to be obvious even for the simplest minds. SAP has followed this approach already more than a decade ago with the APO LiveCache, literally, a maxDB running completely in main memory.

Thanks to advances in microchip technologies large amounts of main memory are now affordable. So simply enlarging the main memory until it can keep the complete dataset of an application seems to be a straightforward strategy.

This approach, however, will still not be sufficient to achieve the necessary performance for ad-hoc analysis. To enable business users to distill useful information from raw data within the blink of an eye, a deep understanding is necessary of how data is organized, not only in the main memory, but also in the CPU and intermediate caches.

Memory is slower than CPU cache

Even if main memory is several times faster than disk, it's still not as fast as the processor itself. Typical memory runs with clock speeds between 0.8 to 1.6 GHz. Whereas CPUs are rated for up to 4.0 GHz (peak). Therefore,

7 Donated to the University of Potsdam by Hasso Plattner

8 Acquired 1997 from Software AG, aka, ADABAS-D and SAPDB

9 Manager Magazine, September 2012

state of the art CPU designs deploy different levels of caching to decrease the latency for repetitive access to the same piece of data.

To load a value from main memory it has to be copied subsequently through intermediate caches until it reaches a register in the core. Accessing main memory can consume up to 80 times the number of CPU cycles compared to an access to the level 1 cache.

At lower cache levels the speed increases, but size decreases. Level 3 cache is measured in MB and runs with a little more than half of the CPU clock speed. Level 1 and 2 run at the same clock speed as the core itself, but they can only store KB of data.

Elaborate algorithms try to predict the next piece of data to be requested by the processor to have them available in cache. If this is not the case a so-called “cache misses” happens. A worst-case is a “full miss” when requested data has to be loaded from main memory. So even the fastest data transfer is futile if it delivers the wrong data. Therefore, data structures have to be optimized to maximize the likelihood that all of the data necessary for the next computing step are in the same cache line.

Row versus column orientation

All the facts described above result in the case of the optimal layout of the database tables. Whenever database structures are discussed, it is implicitly assumed that data is logically stored in two-dimensional tables, like a spreadsheet. In the physical world, however, all the bits and bytes representing the data are stored and transmitted in one single string.

Consequently, there are two ways to transform a table into a single string. You can either arrange one row of the table behind the other or you can queue each column after the other. The first option is called row oriented: the second column oriented.

The case for row-orientation: OLTP

For good reasons, most databases used for business applications store the data values in a row-oriented fashion. This way much of the data that belongs to the same business transaction like order numbers, customer numbers,who bought an item, the part number of the item ordered, the number of parts ordered, the price per piece and the total sum are stored in adjacent memory blocks.

This row-oriented organization of data increases the likelihood that all data belonging to a single business transaction are found in the same cache line, reducing the number of cache misses. The fact that for decades row-oriented databases have enabled sub-second response times even with disc-based storage demonstrates that this concept fits well with the OLTP systems.

The case for column-orientation: OLAP

Unfortunately, row-oriented storage is not well suited for reporting, when not the complete data set of a single business transaction is of interest, except when the part numbers (for example, how much of them are bought on average or the total sum per order?) are of interest.

In contrast to a typical business process, in a typical analysis, only a small number of attributes in a table are of interest for a particular query. Loading every row into a cache, when only a fraction of data is really used, is clearly not an optimal way for OLAP systems, even if they run completely in-memory.

Organizing the tables in a way that columns are stored in adjacent memory blocks make it possible that only the required columns have to be moved into cache lines while the rest of the table can be ignored.

This way, the cache has to keep only the data needed to process the request, reducing the data traffic from main memory to CPUs in-between CPUs and down through the whole cache hierarchy significantly. Maximizing the likelihood that necessary data can be found in the level 1 cache will obviously speed up the processing and minimize the response time.

Analysis of database access in enterprise warehouse applications as well as practical experience demonstrates that column-oriented solutions like Sybase IQ are an excellent choice for online analytical systems. The obvious disadvantage of these systems is that their performance with row-based transactions is poor. So what is good for business transactions is bad for reports and vice versa.

For many years the only answer to this dilemma was to deploy two sets of applications with databases optimized either for OLTP or for OLAP, doubling not only the amount of data to be stored and subsequently also the hardware and operation costs, but also stipulating the demand to synchronize the data between the different systems.

The secret sauce: Two engines under one hood

To combine the best of both worlds and support analytical as well as transactional workloads in one system, HANA combines the two different types of database architectures under one umbrella by means of two dedicated database engines:

- Column oriented for analytical operations
- Row oriented for transactional operations

For each individual data table, the best-suited engine has to be selected at the time of creation. Therefore, typical queries are analyzed with regards to their cache miss behaviour. Together with the weight of the query, this is used to determine the optimal layout — either row or column.

HANA also provides a library of business functions and allows the execution of the application logic directly on the database, thereby, avoiding unnecessarily moving of data to an external application server.

Volatile and persistent data storage

Main memory is fast but volatile by nature and loses all its content immediately in electrical power outages. To avoid this “Alzheimer effect,” HANA has to make all data persistent on non-volatile storage like SSD, flash or disk drives. For this purpose, SAP implemented a MaxDB shadow server providing a persistence layer shared by both database engines.

Changed memory pages are written asynchronously by default every five minutes as a savepoint to non-volatile storage. In addition, a database log captures all changes made by the transactions synchronously, ensuring that all committed transactions are permanent.

After a power failure or a maintenance shutdown the database can be restarted like any disk-based database. First, the database pages are restored from the last savepoint, and then the database logs are applied (rolled forward).

To guarantee high performance during a savepoint, SAP specified an IO throughput of 100.000 IOs per second for scale-up appliances with internal storage, and corresponding numbers for large scale-out HANA implementations with external storage. To reach this extraordinarily high throughput, SSDs are mandatory.

Appendix E – Reference infrastructures

There is always a cry for small, medium and large reference infrastructures — unfortunately, no two SAP landscapes are similar enough to define such references. The good news is that all classical SAP architectures follow the same pattern, independent from the size of the customer.

Even small customers with a single SAP ERP system fitting into a two-tier architecture need a development and a quality assurance system and some mandatory auxiliary systems.

But most small customers running ERP in three-tier landscapes, have more “auxiliary systems,” such as Fiori, WebDispatcher, Content server and Adobe Document server (ADS). -as in the example below.

Amount	System Name	Service	DB	Tier	RAM (GB)	vCPU	# Standby Nodes	Storage (GB) per Server	SLA	Backup Class
DEV Systems										
1	S/4HANA System (DB Server)	SAP S/4HANA On-Premises	HANA DB	DEV	256	32		768	95.0%	Backup-1GB-Wly-2W
1	S/4HANA System (App Server)	App Server for SAP S/4HANA		DEV	32	8		100	95.0%	Backup-1GB-FS-Mly-full+Dly-incr-2M
1	SAP Fiori DB+App	SAP Fiori	Other DB	DEV	64	16		150	95.0%	Backup-1GB-Wly-2W
1	SAP Webdispatcher	SAP Webdispatcher		DEV	8	2		50	95.0%	Backup-1GB-FS-Mly-full+Dly-incr-2M
1	SAP Content Server	Content Server	Other DB	DEV	16	4		300	95.0%	Backup-1GB-Wly-2W
1	ADS Server	Adobe Document Server	Other DB	DEV	64	16		150	95.0%	Backup-1GB-Wly-2W
QAS Systems										
1	S/4HANA System (DB Server)	SAP S/4HANA On-Premises	HANA DB	QAS	512	64	1,536		95.0%	Backup-1GB-Wly-2W
1	S/4HANA System (App Server)	App Server for SAP S/4HANA		QAS	32	8	100		95.0%	Backup-1GB-FS-Mly-full+Dly-incr-2M
1	SAP Fiori DB+App	SAP Fiori	Other DB	QAS	64	16	150		95.0%	Backup-1GB-Wly-2W
1	SAP Webdispatcher	SAP Webdispatcher		QAS	8	2	50		95.0%	Backup-1GB-FS-Mly-full+Dly-incr-2M
1	SAP Content Server	Content Server	Other DB	QAS	16	4	300		95.0%	Backup-1GB-Wly-2W
1	ADS Server	Adobe Document Server	Other DB	QAS	64	16	150		95.0%	Backup-1GB-Wly-2W
PRD Systems										
1	S/4HANA System (DB Server)	SAP S/4HANA On-Premises	HANA DB	PRD	512	64	1,536		99.5%	Backup-1GB-Wly-2W

1	S/4HANA System (App Server)	App Server for SAP S/4HANA		PRD	32	8	100		99.5%	Backup-1GB-FS-Mly-full+Dly-incr-2M
1	SAP Fiori DB+App	SAP Fiori	Other DB	PRD	64	16	150		99.5%	Backup-1GB-Wly-2W
1	SAP Webdispatcher	SAP Webdispatcher		PRD	8	2	50		99.5%	Backup-1GB-FS-Mly-full+Dly-incr-2M
1	SAP Webdispatcher HA	SAP Webdispatcher		PRD	8	2	50		99.5%	Backup-1GB-FS-Mly-full+Dly-incr-2M
1	SAP Content Server	Content Server	Other DB	PRD	16	4	300		99.5%	Backup-1GB-Wly-2W
1	ADS Server	Adobe Document Server	Other DB	PRD	64	16	150		99.5%	Backup-1GB-Wly-2W

Figure 32: Typical small system with S/4HANA, Fiori, WebDispatcher, content server and ADS

On the AWS, this will translate into the following:

Tier	Service	DB	Instance Type	vCore	RAM (GB)	Storage (GB)
Dev	S/4 DB	HANA	n1-highmem-32	**32**	**208**	1700
Dev	S/4 App		n1-highmem-8	8	52	100
Dev	Solman	ASE	n1-highmem-4	4	26	300
Dev	Fiori DB/AS	ASE	n1-highmem-16	16	104	150
Dev	WebDispatcher		n1-highmem-2	2	13	50
Dev	Content Server	ASE	n1-highmem-4	4	26	300
Dev	Adobe Document	ASE	n1-highmem-16	16	104	150
Qas	S/4 DB	HANA	n1-highmem-64	**64**	**416**	1700
Qas	S/4 App		n1-highmem-8	8	52	100
Qas	Fiori DB/AS	ASE	n1-highmem-16	16	104	150
Qas	WebDispatcher	ASE	n1-highmem-2	2	13	50
Qas	Content Server		n1-highmem-4	4	26	300
Qas	Adobe Document	ASE	n1-highmem-16	16	104	150
Prd	S/4 DB	HANA	n1-highmem-64	**64**	**416**	1700
Prd	S/4 App		n1-highmem-8	8	52	100
Prd	Solman	ASE	n1-highmem-4	4	26	300

Prd	Fiori DB/AS	ASE	n1-highmem-16	16	104	150
Prd	WebDispatcher		n1-highmem-2	2	13	50
Prd	Content Server	ASE	n1-highmem-2	2	13	50
Prd	Adobe Document	ASE	n1-highmem-4	4	26	300
HA	S/4 replica	HANA	n1-highmem-16	16	104	1700
HA	WebDispatcher		n1-highmem-2	2	13	50
Aux	SAP router/gateway		n1-highmem-2	2	13	50
Aux	Jump box		n1-highmem-2	2	13	50

Figue 33: Typical small system with S/4HANA, Fiori, WebDispatcher, content server and ADS on AWS

Be aware that SolMan Dev and Production instances have been added due to the fact that SolMan is mandatory for support and license key generation. SAP don't tolerate shared solution managers anymore.

If a customer doesn't accept the small reduction in HANA memory, you have to go to the next bigger instance type. Don't forget to use Red Hat Enterprise Linux for SAP or SLES for SAP

In comparison, a medium customer may need bigger instance types and more app servers but may not have more components — just different ones, as shown in the table below:

		Proposed Solution							
Application	**SAP Environment**	**FS**	**Instance**	**OS**	**DB**	**SAPS**	**vCPU**	**RAM**	**SAN GB**
SolMan	DEV	S4	DB/CI	AIX	DB2	1,100	1	10	556
	PRD	PRD	DB/CI	AIX	DB2	3,300	3	16	812
	SolManTotals					**4,400**	**4**	**26**	**1,368**
ECC	SBX	S4	DB/CI	AIX	DB2	1,650	2	12	1,068
	TRN	S5	DB/CI	AIX	DB2	1,650	2	12	1,196
	DV1	S6	DB/CI	AIX	DB2	1,650	2	12	940
	DV2	DEV	DB/CI	AIX	DB2	5,500	5	40	940
	QA1	QA	DB/CI	AIX	DB2				984
	QA2	S7	DB/CI	AIX	DB2	2,200	2	24	1,068
	QA3	S8	DB/CI	AIX	DB2	2,200	2	24	940
	FIX	S9	DB/CI	AIX	DB2	1,650	2	12	2,220
	PRD	PRD	AP1	AIX	DB2	4,400	4	32	50
		PRD	AP2	AIX	DB2	4,400	4	32	50

		PRD	AP3	AIX	DB2	4,400	4	32	50
		PRD	DB/CI	AIX	DB2	12,000	11	72	6,000
	DR	PRD	AP1	AIX	DB2	4,400	4	32	50
		PRD	AP2	AIX	DB2	4,400	4	32	50
		PRD	AP3	AIX	DB2	4,400	4	32	50
		PRD	DB/CI	AIX	DB2	12,000	11	72	6,000
	ECC Totals					**66,900**	**61**	**472**	**21,656**
GRC	DEV	DEV	DB/CI	AIX	AIX	4,950	5	38	300
	QAS	QA	DB/CI	AIX	AIX				450
	PRD	PRD	DB/CI	AIX	AIX	3,300	3	24	450
	GRC Totals					**8,250**	**8**	**62**	**1,200**
SAP Console	QAS	QA	MW	AIX	AIX	1,650	2	8	50
	PRD	PRD	MW	AIX	AIX	1,650	2	8	50
	DR	DR	MW	AIX	AIX	1,650	2	8	50
	SAP Console Totals					4,950	7	24	150
Grand Totals						**84,500**	**79**	**584**	**24,375**

Figure 34: Typical medium SAP system with ECC, GRC, SolMan and a SAP console

Ready-to-run test drive systems – the case for CAL

The SAP CAL (Cloud Appliance Library) provides an online repository of pre-configured SAP solutions including functional configuration and data for training and demo that can be instantly orchestrated on AWS.

There is an introductory YouTube video on the login page, to get a complete list of available solutions you have to click on the "show all solutions" button at the very bottom. At the time of this writing, the most appropriate S/4 demo is "SAP S/4HANA 1809 FPS02."

Click on a solution to get an overview of the components of the solution and the size of the AWS Cloud Platform instance types necessary for the installation. There is even a cost calculator (which obviously don't consider storage, network, backup and other nifty details).

To create the instances, you need a SAP S-user[10] and connect to your AWS Cloud Platform account (private key ID).

After the installation, read the guidance documentation for post-install configuration, the user and default password and also the demo guide for the scenarios available in the instance.

[10] Send a mail with name/email/phone/title/team to me to request a personal S-User.

Besides getting a idea about the sizing of VM for demos instances, CAL is proving “FrontEnd” VMs => the software that need to be installed locally to run the SAP Application. This FrontEnd comes pre-configured and you can RDP this VMs system to run the Demo Apps with no more effort to install the SAP client software

The CAL appliance comes with demo scenarios pre-configured including test data – so you don’t need a SAP consultant to configure your system, do the customizing and generate test data.

The cost of the AWS Cloud Platform infrastructure is billed directly to your account by AWS, SAP is not charging to use SAP Cal for 90 days, if you need to keep the system you can acquire SAP CAL licenses.