CHAPTER 2

SAP S/4HANA Systems on Public Cloud

SAP S/4HANA systems are based on a consistent evolution of SAP systems from R/3 versions to HANA-based SAP products. They support all important business processes of a company and thus represent the backbone of business operations.

Development of SAP S/4HANA

In 1979, the SAP R/2 system was introduced on mainframes and became one of the standard ERP systems worldwide. The architecture was kept stable for a long time, and only in 1992 SAP introduced a significant evolutionary step: the SAP R/3 system. For the first time, business transactions were performed in real time, and the system combined all major business processes within one large SAP system.

SAP then followed the path of ERP Central Components, or ECC for short. Here, customers could activate and use the most important functions that were important for the companies. There were also specific systems, such as Supplier Relationship Management, which could be installed in addition to the SAP ECC systems. From 2004 onward, there was a renewal in the system landscapes of the customers, and the older SAP R/3 systems were successively replaced by the SAP ECC systems. The duration of 12 years from R/3 to ECC shows the complexity and intricacy of the transformation of such ERP systems.

In 2010, the new database HANA was announced by SAP. The new platform HANA was placed by SAP not only as a new database but also as a development platform for OLAP (Online Analytical Processing) applications. Through this, many steps in evaluation of large amounts of data could simply be done in the HANA database, instead of within the application layer as before. As of 2013, SAP's most important applications

A. Bögelsack et al., *SAP S/4HANA Systems in Hyperscaler Clouds*, https://doi.org/10.1007/978-1-4842-8158-1_2

were then available on HANA – previously, only a few applications had been released for HANA. The "Suite on HANA" introduced additional features, such as access from mobile devices. This already worked before but was significantly simplified by the new platform.

In 2015, SAP introduced the S/4HANA system. This is a simplified (S) system for (4) HANA-based systems. It is at its core a completely redesigned system, still based on the ABAP (Advanced Business Application Programming) programming language, but completely rebuilt. With this announcement, SAP also made a shift away from supporting all common relational database systems to exclusively supporting SAP's own database HANA.

SAP S/4HANA in the Cloud

With the development of the new SAP S/4HANA platform, SAP also focused on the cloud. The new products should no longer be able to be operated only in the data centers of customers or partners but also in the cloud in particular. The basic idea of Software-as-a-Service came more and more to the fore. This was certainly due to the enormous growth of cloud services, but also due to the acquisitions of, for example, SAP Ariba or SuccessFactors, which were sold as pure SaaS solutions. SAP saw the future in the cloud business, and this is still one of the fundamental strategies in 2021.

As one of the first products based on the new S/4HANA architecture, S/4 Public Cloud Multi Tenant Edition (MTE) was launched in 2017. The offering is an SAP S/4HANA system delivered as SaaS. Customers could and can use the new S/4 functionalities based on it. In principle, the offer is aimed at customers who aim for a very high level of standardization and require little customizing. In return, customers then receive a system that undergoes regular upgrades per year and thus always contains the latest codebase. In 2019, the Multi Tenant Edition was renamed to Essentials.

Opposite the MTE, an S/4 Public Cloud Single Tenant Edition (STE) was also launched in 2018 – also as a pure SaaS solution. Through this, customers could obtain a new S/4HANA system and could perform more customizing here than with the MTE. In addition, STE also supported more industry-specific business processes. However, changes to the actual ABAP code were not possible here either. STE was renamed Extended in 2019.

In 2020, SAP launched the S/4 Private Cloud Edition in pilot operation, and from 2021, this edition was also available to all customers. This is an SAP S/4HANA system that is operated and supported by SAP for customers. The customers have the complete S/4HANA codebase at their disposal and can completely adapt the system (customizing). SAP performs an upgrade once a year. In contrast to the Essentials Edition, the Private Cloud Edition supports 25 industry-specific processes.

In addition to the cloud-based offerings, customers can still keep the traditional SAP system in the data center world (on premise). Here, the systems can be operated in the traditional data centers or also in the Hyperscaler clouds.

SAP S/4HANA Architecture

The architecture of SAP S/4HANA can be described from different perspectives, and the different use cases require different components. Thus, the structure of a BW/4HANA system is different from the architecture of an S/4HANA system. Nevertheless, there are basic components which are the same in all SAP systems.

Overview

The architecture of SAP S/4HANA systems is not fundamentally different from the architecture of SAP ECC systems. The same components still exist, which are summarized in an overview in Figure 2-1.

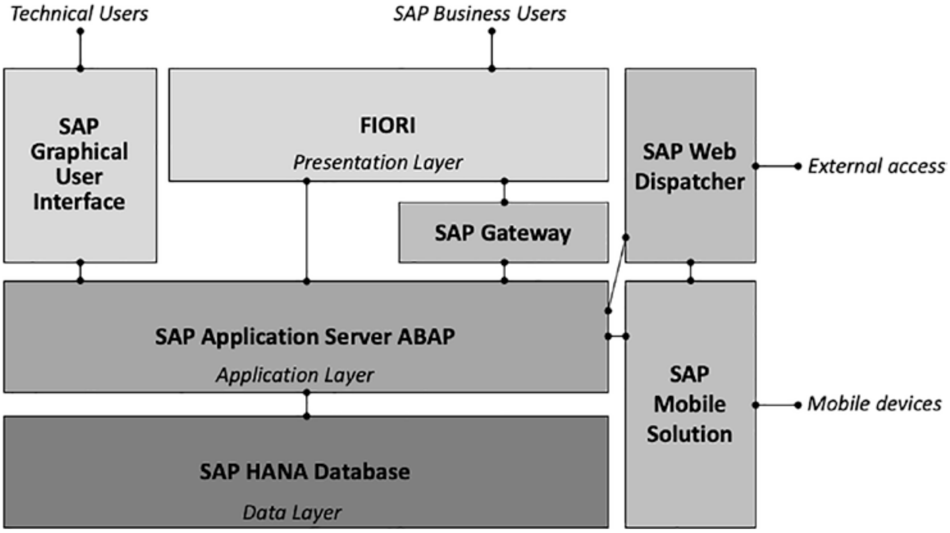

Figure 2-1. *SAP S/4HANA architecture*

The different layers of the SAP S/4HANA system are as follows:

- Data storage layer: This layer is fulfilled by the column-oriented HANA database.
- Application layer: This layer is formed by the application server ABAP.
- Presentation layer: This layer is formed by the SAP GUI and Fiori.
- Communication layer: This layer is formed by the SAP Gateway, SAP Web Dispatcher, and SAP Mobile Solution.

All the preceding layers are explained in a bit more detail in the following.

Data Layer

Data storage and processing take place in the database. In all SAP S/4HANA systems, it is a HANA database. The HANA database holds all data in relational tables in the main memory, which are connected to each other via key relationships. This basic functionality is the same in all SAP systems. The HANA database is a column-oriented database, which stores relations in a column-oriented manner. By storing data in main memory, the database gains very high speed compared to traditional databases, such as Microsoft SQL Server. However, this high speed also comes at a price: the HANA database requires very large and very powerful hardware.

Since the introduction of HANA, the functionality of the HANA layer has been successively extended. SAP's strategy is that certain steps and operations should take place in the database layer. This involves, for example, the preparation of data, which often used to happen in the application layer but should now be done in the database layer. This concept of "code pushdown" is lived in all S/4HANA systems and will also be incorporated into the new ABAP programs. Thus, there is a significant change compared to the previous SAP systems: the database is used as an intelligent component of the SAP S/4HANA system and is all the more important.

Application Layer

The application layer processes the data from the HANA database. In the SAP S/4HANA systems, this is the application server ABAP. In the previous SAP systems (such as ECC), there was still the application server Java, but this is no longer used strategically.

The ABAP programming language (Advanced Business Application Programming) is SAP's own language and has been used by SAP since time immemorial.

The ABAP application server not only processes the transactional data and programs of the S/4HANA system but also provides the functionalities for analytical tasks and search queries. This makes the application server one of the central components alongside the HANA database. The ABAP application server consists of some very important processes, which become very important in the chapters on implementations on the public clouds. These include the following processes:

- Dispatcher: The process distributes all incoming requests to the work processes.
- Work processes: The work processes process the user requests, run the ABAP programs, and access the data in the HANA database.
- Gateway: The gateway serves incoming and outgoing requests to the SAP S/4HANA system.
- Update: The processes update the results of transactions.
- Enqueue: The processes lock records to prevent parallel changes or changes in the background.
- Background: The processes are used for long-running programs (jobs) that need to change or evaluate a lot of data.
- Message server: The message server is used for communication between individual processes of the SAP system.
- Spool process: The spool processes serve print jobs, among other things, also for mass printing.

All processes can also be found in this form in older SAP systems and play the same important role there.

Presentation Layer

Over time, SAP has released a wide variety of programs that served as the central interface between users and SAP systems. Probably the best-known program is SAP GUI, which has been the central operating element of SAP for many years and many development steps. In addition, however, there were and are other products, such as for BusinessObjects.

For the SAP S/4HANA systems, there is a new component, Fiori, which is the central layer that takes over the tasks of representing data and providing functions. Fiori is no longer an optional component, as it was a few years ago, but is essential for the full use of SAP S/4HANA systems. Thus, Fiori also provides applications (apps), which are used for robotics or analytics, for example. Access via SAP GUI is still possible, but end users should ideally access via Fiori.

The three components Fiori plus the SAP HANA database plus the application server ABAP make up the SAP S/4HANA system. There are additional, optional components, which are needed to extend the functionalities or the integration.

Communication Layer

There are three important technical components that are installed in SAP system environments: the SAP Gateway, an SAP Web Dispatcher, and the SAP Mobile Platform.

The SAP Gateway should not be confused with the normal gateway of the ABAP application server. It is a new component, which can also be installed as a stand-alone component. The gateway offers SAP and non-SAP applications the connection to the SAP S/4HANA system. This allows non-SAP applications to access SAP system data via OData services (Open Data).

The SAP Web Dispatcher acts as a proxy for connections to the Internet in all system environments. The Web Dispatcher is not installed in parallel with the SAP system, but in a DMZ (demilitarized zone) where access to and from the Internet is secure. The SAP S/4HANA systems are never directly accessible from the Internet.

The SAP Mobile Platform is a third component that enables direct access to the SAP S/4HANA system from mobile devices. This way, users can access the SAP system natively (i.e., without having to use a web interface) from the mobile devices running Android or iOS. The SAP Mobile Platform is also a separate small SAP system, for which an appropriate architecture must be defined and implemented.

Although the elements of the communication layer are to be seen as optional, the components can be encountered in almost all environments.

Additional Components

Depending on the area of application of the SAP S/4HANA systems, additional applications/interfaces/infrastructure components can be provisioned in the customer environments. This depends very much on the requirements of the companies.

Very often, components for communication with other non-SAP systems, such as warehousing systems or archiving systems, can be found. These components must also be considered and taken into account when thinking about the architecture.

Use Cases of SAP S/4HANA on Public Clouds

This section will describe the important technical use cases for SAP S/4HANA systems on the public clouds. This is not yet about the concrete implementation, but about the importance of the use cases, such as correct sizing or the right selection of availability. These points must be addressed before provisioning the SAP S/4HANA systems.

Sizing

Importance of Right Sizing

Sizing provides the basis for subsequent stable system operation. An SAP S/4HANA system that is too small runs the risk of not being able to process user requests quickly enough. An SAP S/4HANA system that is too large may be able to process user requests well, but will incur high costs, which will be reflected in the public cloud. Therefore, it is important to find a good target value for the size of the new system/the system to be migrated. It is important to avoid unnecessary costs, but also to ensure stable system operation. In addition to the selection of the future size of the virtual machine(s), the sizing also includes the necessary memory requirements, as well as backup and other resources (such as load balancers). An SAP S/4HANA system that is sized too small can cause various problems. The key is to avoid them:

1. Stability problems: A sizing that is chosen too small can lead to significant problems in the stability of an SAP system. For example, a very high utilization of the main memory (greater than 98%) can lead to serious stability problems. In addition to the main memory, storage that is too small or an incorrect storage connection can lead to noticeable performance bottlenecks, which can have a major impact on users and business processes.

2. Low performance: The performance of SAP systems is critical in the execution of companies' business processes. For example, a lot of data needs to be processed in as little time as possible, and you can find industries, such as retail, where SAP systems need to process many small transactions in as little time as possible. Low performance can not only impact users but disrupt entire supply chains and cause delays that can cost customers a lot of money.
3. Frequent maintenance: If systems are not built to meet requirements, there is a risk that frequent maintenance will have to be performed. Such maintenance often requires downtime of SAP systems, which ultimately results in a severe impact on users and processes.
4. Downtime: Probably, the worst-case scenario, due to incorrect sizing, is the downtime of a system. This can be caused by an overload of the virtual machine and/or the memory and thus lead to a complete crash. Especially systems in high availability clusters show such patterns and should be sized with much care.

The data resulting from the sizing are the clues for the later sizing of the system and the design of the system. The next section describes how the process of sizing is performed for SAP S/4HANA systems.

Sizing Process

The process for sizing a system is usually based on the known framework parameters for a new system or for migrating a system. In both cases, however, the following steps are performed:

Step 1: Transparency to data volume. The size of a new system is first derived from the size of the current system or from the expected size. For this, it is important that existing systems are not taken over 1:1 in this way. For example, business warehouse systems often contain a lot of data that is no longer needed. The same applies to SAP ERP systems. Often, no archiving runs are carried out, and thus the systems grow day by day. It is therefore necessary to prepare the systems and perform an archiving/

selection of data to be deleted. In addition, a strategy should already be defined now on how to deal with future data growth. Nearline storage or data aging via SAP HANA can be used here, for example.

Step 2: Execute sizing. For the actual sizing, SAP offers sizing reports that can be executed within an SAP ERP system. For brownfield transformations/migrations, SAP OSS Note 18721170 applies to estimate the future size for a transformed system. After running the sizing reports in the system, a target size can be defined based on this. For greenfield implementations, the Quick Sizer should be used. Based on a few input factors, it can determine a first sizing for the future systems.

Step 3: Adjusting the results. After the first sizing is done, it is important to adjust the results. Two factors should be considered: future growth and specific usage scenarios.

- Growth: The general growth of SAP S/4HANA systems is assumed to be 10% per year. This growth primarily refers to the additional data that is created in a year. If a sizing of 3 TB is assumed for year 1, a target size of 3.3 TB must be expected in year 2. During the time when there were no HANA databases, growth played a role in storage sizing. After the introduction of HANA and in-memory technology, growth must be seen against the background of the main storage size of the virtual machines. Growth makes regular resizing of machines necessary.
- Special usage scenarios: Short performance peaks in the use of SAP S/4HANA systems can have a brief impact on users, but are usually tolerable. For longer performance peaks or even overloads due to long month-end/year-end closings, sizing should be adjusted based on the experience of the previous systems.

Step 4: Transfer sizing to hardware. For all new systems or systems to be migrated, an approximate new hardware size is specified with the number of CPUs and RAM, etc. The CPU-RAM ratio is fixed for SAP HANA-based systems in the public clouds. Thus, in the Hyperscaler Cloud, only the sizes of virtual machines are offered that correspond to the ratio and are therefore also approved by SAP. After an initial sizing has been created, a suitable virtual machine must be selected from the list of supported virtual machines. While mapping CPU and RAM to the available sizes is easy, there are further dependencies with the storage.

Step 5: Verification of the sizing. The final step in sizing is to verify the target size through a test installation or a sandbox installation including a migration of the data from the legacy system. After the system has been set up, it should always be subjected to a stress test or performance test. Only through a test can the actual performance be checked.

HANA Sizing

The sizing of HANA-based systems is based on the size of the main memory. This is done in Gigabytes or Terabytes. For all HANA-based systems, there are certified solutions from Hyperscalers as well as from traditional hardware manufacturers. A detailed list of certified solutions exists on the SAP website at the following link:

https://www.sap.com/dmc/exp/2014-09-02-hana-hardware/enEN/#/solutions?filters=v:deCertified

The sizing of HANA systems is determined by the ratio of CPU and main memory (RAM). This ratio is different for OLTP and OLAP workloads. OLTP stands for Online Transactional Processing, while OLAP stands for Online Analytical Processing. OLTP systems are the normal SAP S/4HANA systems, and BW/4HANA systems, for example, belong to the OLAP systems.

Many of the Hyperscalers use Intel-based hardware for SAP S/4HANA systems, which is effectively Intel Xeon processor based. For an Intel Xeon E7-8890 v4, the following ratios would apply to both workloads:

- OLTP: 1 TB per socket
- OLAP: 0.5 TB per socket

A lower memory ratio is therefore assumed for OLAP workloads. The background for this is the high use of memory in the OLAP environment for the intermediate storage of data, for example, during large load processes into SAP BW. These load processes can often require a lot of temporary memory. Therefore, a lower ratio is assumed when sizing.

It is important to note that sizing for OLTP and OLAP can only be indicative at the beginning and must be refined again later by other usage patterns (use cases). Nevertheless, the main memory remains the important criterion for HANA sizing in order to arrive at an initial sizing.

SAPS – SAP Application Performance Standard

The SAP Application Performance Standard Benchmark is an application benchmark that is executed within SAP systems. It therefore differs from synthetic benchmarks, such as iobench, which only test a specific aspect of performance. The SAPS benchmark was published by SAP and is still the standard benchmark used to measure and also certify the performance of hardware.

The SAPS primarily consists of simple transactions that are executed within the SAP system. These transactions are still available in the newer SAP S/4HANA systems and will continue to exist. These include, for example, the creation of a material master record via MM01 or the creation of bills of materials. The benchmark aims to allow as many parallel users as possible to process the predefined sequence of steps.

While the SAPS benchmark used to be primarily used as a means to measure the performance of hardware and certification of new platforms, it is still used today in sizing to achieve a simpler comparison. For projects migrating from a traditional, on-premise, data center to the cloud, the difficulty of comparing hardware performance arises. Therefore, the SAPS value of the old hardware is always considered a good guide for sizing the new virtual machine in the cloud.

Example: Replacing old hardware with the cloud

A customer was in the process of having its two older data centers on the company's premises demolished. The hardware inside had already depreciated and had not been renewed for several years. Among them were many SAP systems that had not yet been migrated to S/4HANA.

After an initial analysis of the situation, a first indicative schedule for a data center exit was created and sizing was started. Since the old hardware was already several years old, a number of SAPS was first defined for each server type based on historical data. There was very powerful hardware as well as less powerful hardware, some of which only reached 10,000 SAPS.

Then, based on the SAPS, the new target size of the virtual machines, such as DS3v2, was mapped, and the new VM was defined machine by machine. Since the hardware in the data centers was already very old, the customer was able to benefit from the very powerful hardware in the public cloud, and so only very small virtual machines had to be used. These small VMs have a very low price (even in the pay-as-you-go model). This meant that the customer was able to avoid having to renew the hardware and also benefited from a saving in operating costs. The return on investment (ROI) of the actual migration was achieved in less than a year.

The SAP benchmark provides added value as supplementary information, especially for older hardware and systems. Here, mapping based on CPU and main memory can lead to excessive sizing and thus additional costs. This can be prevented by using the SAPS value.

I/O Sizing

The SAP benchmark is also used as an indicator for calculating the storage throughput. The performance of the storage is indicated by IOPS. IOPS stands for “input/output operations per second” and indicates the maximum performance/current throughput of the I/O operations. While SAPS primarily shows the size of the CPU and main memory, the SAPS value cannot make any statement with regard to IO. However, this is very important because HANA databases have specific storage requirements.

In general, the SAPS value can be converted into IOPS as follows:

1 SAPS = 0.6 IOPS (example: 200,000 SAP = 120,000 IOPS).

Depending on the future SAP S/4HANA system and the usage of the system, the value for IOPS can still increase. There are S/4HANA systems that have been created with a ratio of 1 SAPS = 0.9 IOPS.

Now, the SAP HANA database is an in-memory database, but even with a HANA database, the data must be stored consistently on storage. For this purpose, different data areas are required, which should also be separated from each other:

- Redo Log Volumes: All changes to the HANA database and the data are logged in the Redo Log Volumes, so that even after a database failure, all changes can be restored (block size 4 KB up to 1 MB).
- Data Volumes: The data of the HANA database is stored on the Data Volumes. A change of the data occurs at the savepoints every 5 min by default (block size from 4 KB up to 16 MB, maximum 64 MB for super blocks).
- Backup Volume: All created backup data is stored in blocks (size up to 64 MB) on the Backup Volume.

The respective volumes differ in the requirements according to IO performance. The Redo Log Volumes have the highest requirements, since all transactions must be permanently written to the storage as quickly as possible. The Data Volumes have equally high requirements for performance, since no further changes are made to the data when a savepoint is created. The Backup Volume has the lowest requirements. Here, the data is stored asynchronously from the database.

The storage type within the cloud can also be selected according to the requirements:

- Redo Log Volume: Fastest storage
- Data Volume: Fastest storage
- Backup Volume: Normal Storage

The size of the respective volumes depends on many factors, but can be determined with the following “rules of thumb”:

- Redo Log Volume: The rule states that the size is selected as follows: Redo Log Volume = 0.5 x RAM.
- Data Volume: The simplest rule for sizing an SAP S/4HANA system is as follows: Size of Data Volume = 1x RAM.
- Backup Volume: The rule is to size the Backup Volume as follows: Backup Volume = 1x Data Volume + 1x Log Volume.

These calculations are only valid as a first reference point for the initial sizing and also only as a calculation of the size of the volumes, but not for their layout. The size of the volumes can be addressed very easily by the offered sizes of the different storage classes. However, the performance must not be neglected. The larger a volume becomes, the more IOPS are offered by the Hyperscalers. This can be easily illustrated with Microsoft Azure:

- The smallest Premium Disk P1 with 4 GiB can achieve up to 120 IOPS.
- The largest Premium Disk P80 with 32,767 GiB can achieve up to 20,000 IOPS.

It follows that to achieve a high IOPS number, several disks must be connected to each other in stripping.

Example Sizing I/O: SAP S/4HANA System in Azure

The following example is intended to show I/O sizing for a new SAP S/4HANA system in the Microsoft Azure Cloud.

Through the initial sizing, the following key data has been determined:

- RAM requirement: 3.5 TB
- SAPS requirement: 152,000
- IOPS requirement: 91,200

Based on the key data, the following instance type can be used in the Microsoft Azure Cloud:

- Target instance: M128ms with 128 vCPUs and 3892 GB RAM

This instance type is offered without further storage, so that the storage still has to be dimensioned for the Data Volumes and Redo Log Volumes.

- Size of Redo Log Volume: 2 TB
- Size of Data Volume: 4 TB
- Size of Backup Volume: 6 TB

However, the target size of the storage is only one parameter. The necessary IOPS are achieved by stripping the disks:

- Redo Log Volume: 2x P30 with 1024 GiB each and 5000 IOPS each = 10,000 IOPS
- Data Volume: 2x P40 with 2048 GiB each and 7500 IOPS each = 15,000 IOPS
- Backup Volume: 3x E40 with 2048 GiB each and 500 IOPS each = 1500 IOPS

This results in a target sizing with the following data:

Compute and RAM:

- 1x M128ms with 128 vCPUs and 3892 GB RAM

Storage:

- 2x P30
- 2x P40
- 3x E40

Costs

When using public cloud technologies, many companies pursue the goal of reducing the costs of the IT landscape in particular. Often, the migration of the complete IT landscape to the public cloud is aimed at very high cost savings, which are to be realized in the shortest possible period of time. The Hyperscaler providers of public cloud technologies offer a variety of cost reduction measures for this purpose. When applying these methods, business viability must still be ensured, and the impact on existing processes in the company must be taken into account. In practice, a step-by-step approach is therefore recommended, as the price optimization measures usually have an impact on the IT architecture and are accompanied by corresponding changes.

The costs of an IT infrastructure are made up of operating expenses (OpEx) and investment costs (capital expenditure, CapEx). Operating expenses are the ongoing costs of maintaining and ensuring the availability of the IT landscape. These are, for example, the electricity costs for operating a company's own on-premise data center, but also personnel costs incurred for maintaining and keeping the IT infrastructure up and running. Investment costs refer to one-off costs incurred as part of setting up an IT architecture. These are, for example, the procurement costs for physical components required in the infrastructure of the IT landscape, such as servers and data center premises. The total costs incurred for a company's IT environment are referred to as the "total cost of ownership" (TCO for short).

The advantage of using a public cloud is that the investment costs in particular are eliminated, as the Hyperscaler provider makes the physical components available. As a result, the responsibilities for the individual levels of the IT architecture are divided between the provider and the company. The public cloud offers the following models in which responsibilities are divided differently: Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS), Software-as-a-Service (SaaS). Depending on the model, an organization has the opportunity to reduce operating costs by reducing responsibilities. An overview of the distribution of elements in the different models is provided by the so-called "Shared Responsibility Matrix" shown in Table 2-1.

Table 2-1. *Shared Responsibilities*

Local (On-Premise)	Infrastructure-as-a-Service (IaaS)	Platform-as-a-Service (PaaS)	Software-as-a-Service (SaaS)
Data and Access - C	Data and Access - C	Data and Access - C	Data and Access - C
Application - C	Application - C	Application - C	Application - V
Runtime - C	Runtime - C	Runtime - V	Runtime - V
Operating System - C	Operating System - C	Operating System - V	Operating System - V
Virtual Hardware - C	Virtual Hardware - C	Virtual Hardware - V	Virtual Hardware - V
Compute - C	Compute - V	Compute - V	Compute - V
Network - C	Network - V	Network - V	Network - V
Storage - C	Storage - V	Storage - V	Storage - V
Responsibility of Customer - C		Responsibility of Vendor - V	

In addition to the model of distributed responsibilities for individual levels of an IT architecture, Hyperscaler providers offer pricing models that help companies to reduce costs. In connection with public cloud technologies, consumption-based pricing has become established. With this pricing model, the Hyperscaler provider only charges the company for the actual usage time of the resources and instances used. For example, if a virtual machine is only used during regular business hours and remains switched off during other times of the day, only the period of business hours during which the VM is switched on is charged. This model is also known as "pay-as-you-go" (PAYG for short). As soon as an instance is not needed continuously (24*7), there is the option to apply the usage-based pricing model.

Since all systems and applications are often continuously available in an on-premise environment, the decision on applicability with the PAYG model faces some challenges in practice during the migration of the IT architecture to the public cloud. Since the change from continuous availability to limited availability of selected applications is accompanied by a certain need for adaptation of internal processes, a selective analysis of usage is required before any adaptation takes place. One example of such a process adaptation is the way globally distributed development teams work. Due to the different time zones, the availability of a development system must also be ensured outside the business hours applicable to a specific country. Since the actual useful life of components of an IT architecture can only be determined during ongoing operation, it is suitable to establish the usage-based pricing model gradually after migration to the public cloud. All Hyperscaler providers provide various, fully integrable monitoring options for usage monitoring. In addition, the usage period of an instance can also be adjusted retrospectively for deployment in the public cloud.

In contrast to usage-based billing, Hyperscaler providers support a long-term commitment of a company to the public cloud resources offered. The billing model is offered as reservation of the cloud instances ("reserved instances"). Depending on the duration, the Hyperscaler provider grants a discount, which can be up to 70%. Through the commitment, the company makes a binding commitment to use the respective resource or service for a certain minimum duration, but in return receives a discount from the Hyperscaler provider.

In a practical enterprise environment, usage-based billing is usually not applicable to all components in an IT architecture. In particular, business-critical applications are usually subject to the requirement of 24*7 availability in order not to interrupt essential processes and thus possibly jeopardize the ability to do business. For productive workloads, the reserved instance model is therefore often suitable, since these are also usually used for longer periods.

For non-productive components, the decision is consequently made between the pay-as-you-go model and reserved instances. In order to make this decision as cost-effectively as possible, it is advisable to first determine the minimum required availability of each instance concerned. Using the pricing calculators provided by each Hyperscaler vendor, the monthly usage price for the usage-based model and for the commitment model can be calculated based on the instance type. The break-even amount of the respective instance type serves as the basis for deciding on the billing model. This can also be determined using the price calculators by comparing the two billing models and approximating the useful life. The amounts of the useful life are alternately approximated until the monthly price for both billing models is identical. The useful life that applies in this case represents the break-even for the particular instance type. The break-even represents the value of the useful life at which identical costs are incurred for both the usage-based pricing model and the commitment. From the break-even useful life onward, the reserved instance model is therefore suitable from an economic point of view, since the workloads are available for longer at the same price, in contrast to PAYG. The diagram in Figure 2-2 provides an example of the comparison between the ratio of the useful life and the monthly costs incurred in relation to the respective pricing model based on a virtual machine of type D4 v3 (4 vCPUs, 16 GB RAM, 100 GB temporary storage) in the Microsoft Azure Cloud.

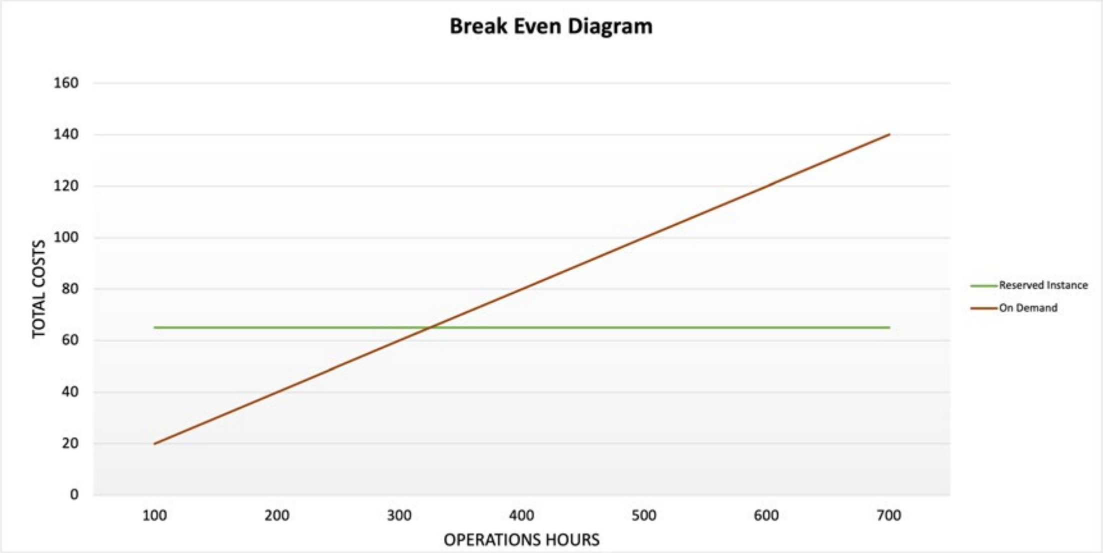

***Figure 2-2.** Break-even after 300 hours of operations*

To determine the break-even value, the following assumptions were used as the basis for the calculation in the Microsoft Azure price calculator:

- Region: West Europe
- Operating system: Linux (Ubuntu)
- Tariff: Standard
- Instance: D4 v3 (4 vCPUs, 16 GB RAM, 100 GB temporary storage)
- Virtual machines: 1
- For the comparison between usage-based payment and reserved instances, a useful life of three years was added for the long-term commitment.

Figure 2-2 shows that the costs for usage-based billing are higher than for reserved instances from an operational, monthly usage period of more than 320 hours. This value corresponds to the break-even.

With the help of the break-even of the usage time of an instance, the decision process is initiated. The determined break-even is compared with the actual usage time. This comparison distinguishes between the following cases, which serve as decision-making aids:

- Break-even is lower than the actual usage period: In this case, the reserved instance model is economically more efficient because the instance or service can be operated continuously (24*7) for a lower price in contrast to the usage-based billing model. In summary, the Hyperscaler provider's customer gets a higher utilization rate at a lower price.
- Break-even corresponds to the actual utilization period: The assessment of this scenario is analogous to the first scenario. If break-even matches the useful life, the user has a longer useful life of the instance at an identical price as with usage-based billing.
- Break-even is higher than the real useful life: If the break-even is significantly higher than the real useful life of the instance, the use of the usage-based pricing model is suitable. It is not possible to make a blanket decision recommendation for this case, as the size of the difference between the two amounts is particularly decisive.

> For example, if the break-even is only a few cents higher than the comparative value, the reserved instance model would be a more efficient option despite the minimally higher costs, since the instance would be continuously available in this case for a minimal increase in costs. Therefore, it is recommended to make an individual decision for each instance.

With reference to the exemplary comparison of the two billing models, the decision depends on the planned, monthly usage period of the virtual machine. If this should be higher than 320 hours, it is recommended to use the reserved instance model for economic reasons and cost efficiency. At this point, however, it must be taken into account that this is a long-term commitment.

In addition to the described pricing models to reduce costs, Amazon Web Services offers a volume discount for selected services. With an increase in usage and an associated, higher storage requirement, the costs decrease if more storage is consumed.

Another method for optimization is sizing the cloud infrastructure, which has already been described in more detail in the previous section. With the aid of resizing and right sizing, the components used in the architecture are checked for their actual utilization. The aim is to optimize the resources within the IT architecture, for example, to identify unused computing or storage capacities. Adjusting the resources used to the capacities actually used can support the reduction of costs in a company.

Example Calculation SAP S/4HANA System in Azure

The following example uses a productive SAP S/4HANA system in the Microsoft Azure public cloud to determine the break-even. The first step is to define the requirements and prerequisite needed to run an S/4HANA system in the Azure Cloud. The central business processes are carried out in the productive SAP S/4HANA system. Consequently, the system is continuously required to maintain the business capability. We use the following assumptions:

- No high availability or disaster recovery is considered in the first calculation approach.
- The second calculation approach provides for high availability via the provision of an additional availability zone.

- Initially, only the buildup of the resources required for operation is to be taken into account. Therefore, **no additional** Azure Native Services or other integration points are added in the calculation.
- The SAP S/4HANA system is composed of the ERP application with the associated **HANA database**, as well as the **Fiori environment**. To accommodate high performance and the principle of modularity, both environments are built in **separate VM instances**.

In the Azure price calculator, the following characteristics were assumed to calculate the total costs:

- Region: Western Europe.
- Licenses: The licenses for the operating systems are already in place, so the Azure hybrid benefit can be applied (bring-your-own-license).
- Reservation: The instances are reserved for three years.
- Storage: An SSD Premium Edition was assumed as the managed disk of the virtual machines for each instance.

Table 2-2 summarizes these components and their costs.

Table 2-2. *Components and Costs*

Environment	Workload	Type	Operating System	Storage	Instances	Costs
ERP	SAP	D16s v3	Windows	650	1	$384.63
	DB	M64ms	Linux	4096	1	$2945.36
Fiori	SAP	D8s v3	Windows	250	1	$164.89
	DB	M32ts	Linux	896	1	$648.87

The total monthly cost in this approach is $4143.75, and annually the total cost is $49,725.00. Due to the commitment for the reservation of the instances over three years, the complete sum for this period is $149,175.00.

Table 2-3 shows the cost for a high availability (HA) architecture.

Table 2-3. *Costs for Components for an HA Architecture*

Environment	Workload	Type	Operating System	Storage	Instances	Costs
ERP	SAP	D16s v3	Windows	650	1	$384.63
	DB	M64ms	Linux	4096	2	$5890.71
Fiori	SAP	D8s v3	Windows	250	1	$164.89
	DB	M32ts	Linux	896	2	$1369.73
Azure Site Recovery					2	$42.17

The total monthly cost in this approach is $7852.13, and annually the total cost is $94,225.56. Due to the commitment for the reservation of the instances over three years, the full amount for this period is $282,676.68.

High Availability

SAP S/4HANA systems must ideally be available 24 hours a day, 7 days a week. This also applies if there are problems in the data center or in the SAP systems. There are several ways to keep the systems highly available, which are described here.

Overview High Availability

SAP S/4HANA systems form the backbone of the companies' business and are an integral part of all business processes. Therefore, SAP S/4HANA systems are designed in the architecture to be as highly available as possible and to remain available and provide services even if a component fails.

The availability of an SAP system is measured in percent, and the architecture of a system is based on the importance/criticality of the system. For example, systems such as sandbox or test system are not very critical and therefore do not require high availability. However, production systems are very critical and are protected from component failure via high availability. Companies must evaluate the criticality of the systems in the architecture and align the technical architectures accordingly.

The availability of an SAP S/4HANA system is given as a percentage. The calculation basis is the maximum theoretical, calculated availability of the SAP system in the month in minutes. The maximum availability is 31 days * 24 hours * 60 minutes = 44,460 minutes.

Availability = (Max. availability - unavailability)/(Max. availability)*100

Based on the formula, an SAP system downtime of 700 minutes in a month, for example, can be said to have an availability of 98.43%:

Availability = (44,460-700)/44,460*100 = 98.43%

The overall availability of an SAP system results from the combination of the availabilities of the individual components. In the Hyperscaler Clouds, for example, customers are offered a possible availability of up to 99.99% for a virtual machine. However, it is important to understand that the virtual machine is only one component. It is important to consider the other components of an SAP system for overall availability. This includes the operating system, the network, the storage, and the file system, as well as the individual components of the SAP system. When these availabilities are all combined, the result is a different overall availability of a system. In the following, the total availability is calculated as a result of the availabilities of the other components:

$$\begin{aligned}Availability &= 99.99\%(VM) * 99.9\%(Storage) * 100\%(OS) * 99.9\%(Network)\\ &* 99.9\%(SAP) = \mathbf{99.69}\%\end{aligned}$$

The preceding calculation example shows how the overall availability of the SAP system falls, although the availability of the individual components is very high. This must be taken into account when discussing the availability of SAP systems. Even if only one small component in the overall S/4HANA system network is not available, it can directly affect availability. To rule out these possibilities, various high availability solutions can be used.

Availability Classes

The importance of S/4HANA systems for the companies can be derived from the criticality of the systems and a possible influence on the business processes. There are systems which are not important for the continuation of the companies. In addition, there are S/4HANA systems that are existential for the continuation of the business.

Companies categorize S/4HANA systems according to criticality and assign a corresponding necessary availability to each criticality. This availability should then be translated into a corresponding architecture in SAP systems. The architecture must be able to ensure the availability of the SAP systems and protect against failures. A categorization can be very different and can be named after, for example, metals:

- Gold class: Systems of the highest criticality level, which have a major impact on business processes due to a failure and as a result of which a company would no longer be able to operate. These are often the productive S/4HANA systems.
- Silver class: Systems of the medium criticality level, which have a limited impact on the company's business processes due to a failure (e.g., on only one area of the company). These are often quality assurance systems or productive business warehouse systems.
- Bronze class: Systems of the lowest criticality level, which have a very small impact due to a failure (e.g., a very small group of employees affected). These are often the development or even some quality assurance systems.

In addition to naming based on metals, it is also possible to simply work with numbers, which also reflect the criticality. Table 2-4 gives an overview.

Table 2-4. *Availability Classes*

Category	Gold	Silver	Bronze
Availability	99.5%	98%	95%
Max outage	4 hrs	14 hrs	37 hrs

The availability classes must be realized by a corresponding architecture. There are various options for this for the SAP S/4HANA systems in the Hyperscaler Clouds:

- Cloud native: Since the Hyperscaler Clouds all rely on virtualization, the S/4HANA systems also benefit from higher availability, which is given by the availability of the virtual machines in the cloud. For example, Microsoft Azure puts the target availability of the simplest virtual machines at up to 95%. If the availability turns out to be lower, service credits can already go back to the customer.

- Availability zones: In all large Hyperscalers, there is the option of additionally securing a virtual machine. This additional protection is provided by availability zones (availability zones, availability sets, etc.) and can increase availability up to 99.99%.
- High availability clusters: The availability of an SAP system is determined by the availability of its components. Since Hyperscaler Clouds only secure the virtual machines, high availability clusters are used to secure the other availabilities.

By using Hyperscaler Clouds, a basic availability of SAP S/4HANA systems can already be ensured.

High Availability of SAP S/4HANA Architecture

In recent years, SAP has worked continuously to eliminate the so-called single points of failure in SAP systems through redundancies. The architecture of an SAP system essentially consists of the following components, which can also be seen in Figure 2-3:

- SAP Central Services: SAP Central Services (SCS) provide the most important central services of the SAP system. These include the Enqueue Server and the Message Server. Without these two components, users can still work on the SAP system, but only with limitations.
- Application server: The application servers of an S/4HANA system handle user requests and hold the disp+work processes of an SAP system. Since an S/4 system can have multiple application servers, these servers can also fail without having an impact on the overall availability of an SAP system.
- HANA database: The HANA database stores the business data of the SAP system and can therefore be seen as a single point of failure. If the database is unavailable, the entire SAP system is unavailable.
- File shares: Typically, SAP S/4HANA systems need the files from the global mount point and the files from the transport directory accessible to all components of the SAP system and system line (i.e., development, quality assurance, and production system).

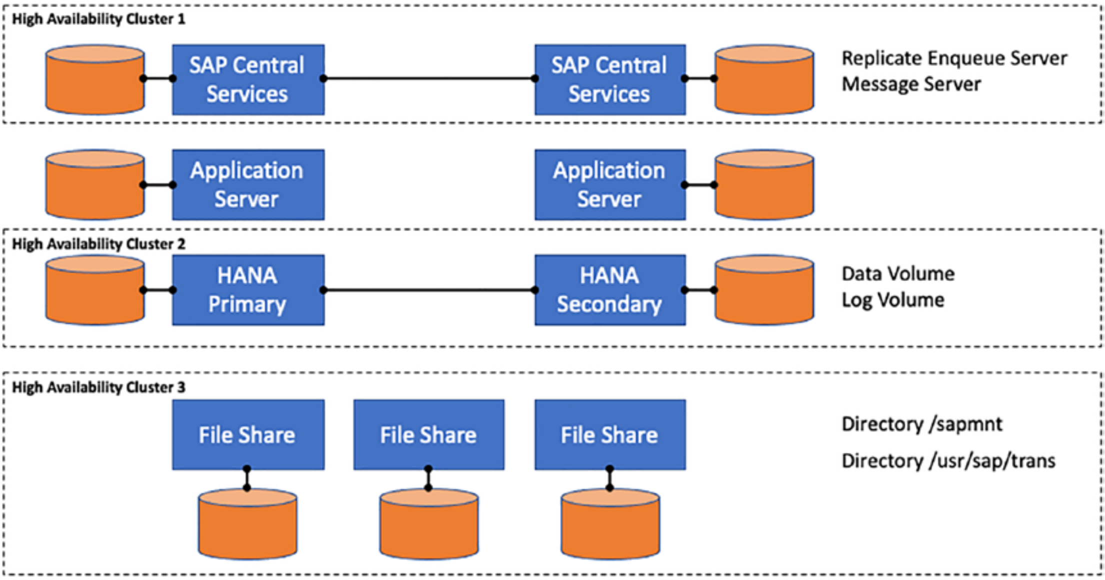

Figure 2-3. *Components of the SAP system*

Figure 2-3 shows the components of an S/4HANA system that should ideally be kept highly available. A virtual machine with connected storage is used for each component. The figure shows how the individual components are secured against failure:

- SAP Central Services: The central services of the S/4HANA system are secured by a high availability cluster. For this purpose, a Linux Pacemaker can be used, for example, which monitors the services and can restart the processes on the second node if necessary.
- Application servers: The application servers are normally not further secured against failure. Since there are usually several application servers in an S/4HANA system, the end users can continue to work even if one server fails.
- HANA database: The HANA databases of the S/4HANA systems are interconnected by HANA System Replication. Again, a high availability cluster is required to monitor the HANA services (such as the hdbindexserver).
- File share: The file shares are usually kept highly available as well. This can happen either through the clouds' native NFS means (e.g.,

Azure Share) or can be done through other clustering technologies (e.g., Gluster).

So, for very critical S/4HANA systems, the architecture required to achieve high availability can become very complex.

Example:

Looking at the exemplary architecture for a highly available SAP S/4HANA system, the question naturally arises as to the costs of such an architecture. It is therefore important that such architectures are only used for the critical systems, as otherwise the cost trap threatens and a very high invoice amount would be due from the cloud provider.

To ensure that this does not happen, an exemplary customer has subdivided its SAP system landscape as follows:

Highest criticality: Only two of the ten productive SAP systems were classified as so critical that they were given a highly available architecture. These were two ERP systems that were operated for two divisions of the customer. These two systems were built on the basis of the architecture shown before. These two systems were also equipped with active disaster recovery, which is described in the following subsection.

High criticality: All other productive SAP systems fell into this category. These included, for example, the business warehouse systems, the PI/PO systems, and other systems (such as Fiori Gateway, Solution Manager, and GRC). In the event of a failure of these systems, the customer would be impacted, but the customer's business would be able to continue. Therefore, the customer decided against a complex architecture for this.

Low criticality: All other systems (quality assurance, development, and sandbox systems) were given a very simple architecture with no further safeguards or complex architecture.

The customer worked with the business department to determine the criticality of the systems. This is not and cannot be a decision made by the IT department alone.

Disaster Recovery

Disaster Overall

While the high availability of S/4HANA systems protects against the failure of a component, it cannot protect against the failure of an entire data center and thus the entire system. The failure of a data center can be triggered by various factors:

- Airplane crash: The crash of an airplane on the data center can trigger the disaster and lead to a total failure of the data center. However, since plane crashes are very unlikely, this reason can be used more as a theoretical reason.
- Fire: In fact, there have been fires in data centers or even outside data centers in the past, which have led to an outage. The most recent example of a data center operator in France shows the possible extent of a fire. Huge amounts of customer data were lost.
- Power failure: Data centers are protected against power failures by uninterruptible power supplies (UPSs) or emergency power generators. Nevertheless, a failure can occur if, for example, maintenance work is carried out incorrectly or there is a problem with the UPSs.
- Cooling failure: A failure of the central cooling system is rarely possible, but can lead to a successive shutdown of the servers due to overheating.
- Connectivity failure: Data centers are connected to the outside world through multiple, redundant connections. Nevertheless, the famous example of the excavator (i.e., earthworks outside the data center) can lead to a loss of connectivity and thus affect SAP systems.
- Lightning strike: A lightning strike in a data center should not lead to an outage in any case for the large cloud providers, but it should for smaller cloud providers.
- Terrorist act: A bomb attack on a data center can lead to an outage, but must be classified as very unlikely. The data centers of the cloud providers are secured and monitored several times.

- Hacker attack: A hacker attack is a very likely event, but it cannot affect the entire data center.
- Emergency patching: Emergency patching cannot be seen as a disaster, but it can certainly be seen as a situation that can lead to a failure of the SAP system. This situation occurred, for example, after the discovery of the Heartbleed vulnerability.
- Earthquake: Although the probability of occurrence is rather low, earthquakes are considered a possible scenario for a data center failure.

In the event of a disaster, customers must either expect the SAP S/4HANA system to fail or there are precautions in place to keep the SAP system available. This is the goal of a disaster recovery.

Important Parameters in a Disaster Case

Although the failure of a data center is an unlikely event, it can nevertheless occur due to the reasons mentioned in the previous section. The disaster recovery mechanisms, which can be implemented in different forms, protect against the failure of a data center. The two most important parameters for implementing the mechanisms are RTO and RPO. RTO describes the Recovery Time Objective and indicates the time period in which a system should be available again after a disaster. The Recovery Point Objective (RPO) describes the maximum data loss after a disaster. Based on RTO and RPO, different mechanisms are implemented depending on the criticality of the SAP S/4HANA systems.

If a data center or an entire region fails (e.g., all data centers in Amsterdam), this is referred to as a disaster. Provided SAP systems are secured for this eventuality, there is an automatic switch to the second region. The time required to make the SAP system available to end users in the second region is defined as the RTO – Recovery Time Objective.

If a data center or an entire region fails (e.g., all data centers in Amsterdam), this is referred to as a disaster. SAP systems are basically protected against a disaster, but there are systems where no difference in data between the primary region and the secondary region (or primary data center and secondary data center) may occur. The RPO (Recovery Point Objective) refers to the maximum loss of data between the two regions and is expressed in minutes/hours.

Since not all SAP S/4HANA systems need to be immediately operational and available again after a disaster, the importance of the SAP systems is also taken into account to decide which mechanisms are implemented. Table 2-5 shows the different RTO and RPO of SAP systems.

Table 2-5. *Availability Classes*

Category	Gold	Silver	Bronze
Availability	99.5%	98%	95%
Max Outage	4 hrs	14 hrs	37 hrs
RTO	< 30 min	< 4 hrs	< 24 hrs
RPO	< 15 min	< 1 hr	< 4 hrs

For highly critical systems (gold category), a maximum data loss of 15 min is thus assumed and a maximum recovery time of 30 min. This means that critical production systems must be available again immediately after a disaster. Less critical systems (silver category) can also be restarted later, and systems in the bronze category can be available up to one day later. The following section shows how these goals can be achieved via technical mechanisms.

Possible Implementations

Various mechanisms can be used to protect SAP systems against the failure of a complete data center. For this purpose, a distinction is first made according to the type of mechanism. Each of the mechanisms has certain advantages and disadvantages:

- Hot standby: If, in the event of an SAP system failure, the DR (disaster recovery) data center is to have as short a period as possible until the system is available again, mechanisms are used for a hot standby. In this case, parts of the SAP system are synchronized (e.g., the database).
- Cold standby: Replication mechanisms are often used for a cold standby, which leads to an increased RTO.
- Backup and restore: For less critical systems, it is recommended to implement a pure backup and restore strategy in case of a disaster. This allows systems to be restored in a limited amount of time.

Each component of the SAP system can be secured against a disaster in different ways, and these are listed in Table 2-6. For SAP S/4HANA systems, for very critical systems, synchronizations are used via the onboard means of the HANA databases. This includes HANA System Replication, which brings the options of replication and full synchronization. Cloud onboard means can be used for the application server and SAP Central Services components. One example is Azure Site Replication, which replicates the Central Services virtual servers to a DR data center.

Table 2-6. *Comparison of DR Implementations*

Mechanism	Hot Standby	Cold Standby	Backup and Restore
Character	Synchronous – Primary and secondary sides are kept synchronous; there is no time offset between the primary and secondary sides	Asynchronous – There is a small time offset between the primary and secondary sides	Asynchronous – There is a large time offset between the primary and secondary sides
RTO	< 30 min	< 4 hrs	< 24 hrs
RPO	< 15 min	< 1 hr	< 4 hrs
Investment	High effort for initial setup	High effort for initial setup	Low effort
Operations	High costs due to active components in the secondary side	High costs due to active components in the secondary side	Very low costs due to storage of backups
Disaster recovery tests	Simple DR tests by simply switching to the secondary side	Simple DR tests by simply switching to the secondary side	High effort due to full restore tests

To secure S/4HANA systems against disaster, the mechanisms are considered and implemented at the beginning of the deployment. The implementation of an active DR mechanism, via HANA System Replication, for example, requires some effort and time. In addition, HANA systems must be operated on both the primary and secondary sides. This results in costs for the Hyperscalers, which can be significant.

The lower the RTO and RPO, the higher the costs for a DR mechanism during implementation and operation. In addition to the initial implementation, there are also high costs for the operation of synchronous and asynchronous solutions, since the components in the secondary side must be constantly available here.

Especially with HANA databases, the running costs can be very significant. For example, a virtual machine for a 3.8 TB HANA database in Western Europe (Amsterdam) in Microsoft Azure costs up to $22,000 (as of mid-2021) per month. These costs can be reduced by commitments, etc., but are incurred monthly for synchronous and asynchronous implementations. With a solution based on backup and restore, hardly any resources are required on the secondary side, and in principle only the costs for storing the backups are incurred.

In addition to the costs of implementation and operation, costs also arise from the actual DR tests. Here, testing is rather simple for synchronous and asynchronous solutions. There is a switch from the primary side to the secondary side and a final test:

1. Check of the synchronization between primary and secondary sides
2. Checking of the current data status by simple tests (e.g., number of currently existing users in the client)
3. Change from the primary to the secondary side
4. Switching off the primary side
5. Check of the secondary side and adjustment of connectivity
6. Check of the data status
7. Synchronization from the secondary to the primary side
8. Change from the secondary to the primary side and adjustment of connectivity

Performing a DR test for backup and restore involves more effort in this respect. For this, new virtual infrastructure must first be created, and then a restore must be performed. Performing these tests usually takes much longer and can include the following steps for a restore based on a database backup:

1. Checking the available backups in the secondary side.
2. Provisioning of the new virtual hardware (virtual machine, storage, network, resource groups, etc.).

3. Provisioning of the operating system and file system.
4. Provisioning of the empty S/4HANA system (without data).
5. Installation and configuration of the backup agent.
6. Execution of the restore of the database.
7. Testing of the S/4HANA system after successful restore.
8. Testing of the data status.
9. Deprovisioning of the virtual machine.
10. Regular testing of DR mechanisms is important for any organization and should be performed annually for all SAP systems.

Backup and Restore

In the previous section, the topic of disaster recovery was discussed, which is an elementary topic for modern SAP S/4HANA systems. Here, backup and restore was a way to restore the SAP systems in the event of a disaster. In general, however, backup and restore should not be missing in any environment, because only in this way can the data of the SAP systems be backed up, and the availability of the data in the event of an error can also be ensured.

Important Components for Backup and Restore

SAP S/4HANA systems consist of a number of components and important parts that must be included in a data backup. Depending on the criticality of the SAP systems, the data and configurations are backed up with different frequency. The following components are important for a backup:

1. Virtual machine: When backing up the virtual machine, the configuration of the VM is important. This includes the name of the VM, which virtual disks are attached to the VM, which VM template was used (e.g., in Azure M128), and which virtual network ports were configured. It is important that this information is backed up at regular intervals for a possible restore. However, it is not likely that the configuration will change too frequently.

2. Operating system: The operating system is the foundation of the SAP S/4HANA system. It contains the basic virtual machine configurations, file systems, and SAP system configuration files. Since it does not change frequently, operating systems are backed up at regular but not too frequent intervals.

3. File system: An SAP S/4HANA system usually does not have only one file system, but it includes quite a few file systems. It contains all important configuration files of the operating system, as well as the SAP system and the kernels of the database and the SAP system.

4. Database: The database contains all the important data of the SAP system, and regular backup is essential. For a backup of HANA databases, the backups of redo logs and data volumes are important. Backup can be done in different ways (full and incremental).

5. SAP-specific directories: SAP S/4HANA systems store various files in the file systems. These include the SAP kernel, database kernel, sapmnt and global mount file system directories, and the transport directory. All of these important directories are important for a backup.

In addition to the most important components listed earlier, there are often other directories and data that are important for a backup. These include, for example, interface and audit directories or directories for the exchange of files. However, since these are very individual, they are not considered further here.

Backup Classes

Not all SAP systems in a company are equally important or have the same criticality. As with high availability and disaster recovery, SAP systems are divided into different backup classes. Each class receives an individual backup plan, which corresponds to the importance of the data. This classification into backup classes is done similarly to how it is done for high availability, etc.

If the classification of backups is based on the role of the SAP systems, the components can be sorted as shown in Table 2-7.

Table 2-7. *Backup Classes*

Component	Gold	Silver	Bronze
System	Production	Quality	Sandbox, training, development
Virtual machine	Daily	Daily	Daily
Operating system	Every 4 hrs	Every 24 hrs	Every 24 hrs
File system	Every 4 hrs	Daily	Daily
Database	Data Volumes: Daily incremental backup Weekly full backup Log Volumes: Every 15 minutes	Data Volumes: Daily incremental backup Weekly full backup Log Volumes: Every 60 minutes	Data Volumes: Daily full backup Log Volumes: Every 4 hours
SAP-specific directories	Daily incremental backup Weekly full backup	Daily incremental backup Weekly full backup	Weekly full backup

The classification allows the different components to be backed up based on their criticality. Important components are backed up very frequently (such as log volumes) and less critical components are backed up only rarely.

Although storage space in the public cloud is very cheap, efficient retention of backups is important. The backed up data can otherwise lead to a significant cost contribution.

Retention

Today's S/4HANA systems can easily exceed the 10 TB limit and are now reaching a significant size that makes efficient data retention important. A simple calculation of the data to be backed up can illustrate the importance of the topic. For example, an S/4HANA system with a database size of 10 TB becomes a volume of 20 TB to be backed up per month.

How to calculate the needed target size for backup storage?

To calculate the total volume of data to be stored, a system of one size is assumed with the key data shown as an example in Table 2-8.

Table 2-8. *Exemplary System Size*

System Component	Size	Unit
Data Volumes	10,240	TB
Log Volumes	1024	TB
Operating system	256	GB
Change rate	10	%
Full backup	Weekly	
Incremental backup	Daily	

Based on the preceding example, the following volume results for the backup of the data:

- Complete backup: In total, 11.5 TB of backup volume must be calculated for the weekly complete backup.
- Daily backups: The rate of change also determines the volume to be backed up, which is up to 2.05 TB. These daily backups result in a weekly volume of 14.3 TB (7 x 2.05).
- Weekly volume: Based on the full system backup and the sum of the daily backups, the total backup volume is 25.8 TB.
- Monthly volume: The weekly backups of the data result in a total volume of 103.5 TB.

This backup volume must be held in addition to the normal data. In Hyperscaler clouds, the occupied storage must be paid for. Therefore, efficient storage of backup data is very important.

The preceding example shows the relevance of efficient storage of backup data, as the prices of the different storage classes in the Hyperscaler clouds vary greatly. This can be easily illustrated using the Microsoft Azure Cloud:

- Premium SSD storage: A premium disk (P30 with 1024 GiB) costs €125 per month.
- Standard SSD storage: A standard disk (E30 with 1024 GiB) costs €64 per month.
- Standard HDD storage: One standard disk (S30 with 1024 GiB) costs €35 per month.
- Standard BLOB storage: A capacity of 1024 GiB on the hot tier costs €17 per month.
- Standard BLOB storage: A capacity of 1024 GiB on the cold tier costs €8 per month.
- Standard BLOB storage: A capacity of 1024 GiB on the Archive Tier costs €3 per month.

The preceding price examples from mid-2021 show how serious the differences can be and how important it is to store data correctly on the respective storage classes. The Hyperscalers all offer respective storage classes that address different requirements of the SAP S/4HANA systems.

The different storage classes become relevant when it has to be determined for each S/4HANA system where the normal data should be stored and where the backup should be saved:

- Normal storage (SSD, HDD): The normal storage area of the S/4HANA systems should be stored on the normal disks of the Hyperscalers. The disks achieve the necessary performance for HANA-based systems, but are not suitable for storing backups.
- Hot tier: BLOB storage is assumed to be used primarily for backup of large amounts of data, which is not used often anymore. Nevertheless, hot tier classes offer a lot of storage space, and data can be accessed and read again within a short time.
- Cold tier: With BLOB storage of the cold tier classes, it is assumed that the stored data only needs to be read very rarely. Long memories are therefore used here, which keep the data retrievable but no longer offer high speeds.

- Archives: Archive memories are very slow storage media, but they are the cheapest option of memory. Here, however, it can take a very long time before data can be read again. The Archive class is more for long-term archiving.

The correct positioning of the backups depends on the duration for which the backups need to be stored (retention) and the way in which the data is accessed. Backup solutions for SAP systems also offer auto-tiering for this purpose, which distributes the data according to criteria to one of the available storage classes. In addition, these solutions also have other features that lead to a reduction in backup data. These include deduplication and compression, which are explained in the following section.

Backup Technologies

There are various procedures for backing up the components. All major Hyperscalers already offer native backup solutions for this, which also have integration with SAP's own Backint interface. This means that the backups can also be tracked within the SAP system.

Regardless of the solution selected, there are basically two different procedures that are used. These are snapshot-based backups and stream-based backups:

- Snapshot-based solutions: The backups are performed as snapshots of the virtual machines and the S/4HANA system within them. A snapshot records the current state of the data and the state of the S/4HANA system. All data in a VM is thus backed up and initially remains on the same storage as the S/4HANA system.

- Stream-based solutions: Backups are backed up from the virtual machine via an agent. The agent reads the data and backs up this data via pipes to a target environment. The target environment can be, for example, a virtual tape library (VTL) or a BLOB storage (BLOB = Binary Large Object). In any case, however, it is a second storage area that is independent of the storage of the S/4HANA system.

The important difference between both technologies is the whereabouts of the data. While with a stream-based backup, the data always remains on a different storage area than the S/4HANA system, with a snapshot-based backup, this backup must first be transported to a further secondary storage. This also fulfills the usual requirements for separating the backup areas.

When considering the costs of backups, the enormous price differences between the storage classes were highlighted. Since snapshots always reside on the same storage where the actual data resides, frequent snapshots can lead to an accumulation of data. This accumulation of data on expensive storage is not sensible and should be prevented.

Backup solutions are characterized by intelligent storage of data. Not all data is stored on the fast storage, so that there is no cost explosion. Auto-tiering is used for this purpose, which distributes the data based on access patterns. Data that is not used frequently is stored on a slow storage class. Data that is used very frequently is stored on a fast storage class. Backup solutions can automatically distribute the data and can also redistribute data after it has been written to the storage for the first time.

In addition to auto-tiering, deduplication and compression are among the capabilities of backup solutions that are also important in Hyperscaler clouds. Deduplication of data means the simple storage of redundant data. This is the case with operating systems, for example. Since the operating systems are backed up regularly, but the data does not change significantly and a large part of the data remains the same, the backups also only store the changed data. Redundant data (duplicates) are not stored many times. Deduplication can be expected to reduce backup data by up to 90%.

Compression enables a further reduction of the backup volume. Here, the backups are stored in compressed form. Due to the structure of the data, the backup solutions cannot always achieve the same results and compression rates. In the case of SAP systems, however, a not inconsiderable compression (often up to 50%) can be assumed.

A sound backup solution can help to back up the data properly and cost-effectively. This should always be considered before using the cloud-native solutions, even if a backup solution requires more effort in the initial setup.

Backup Solutions Out of the Cloud

For SAP S/4HANA systems, there are various solutions from the cloud and in the cloud. These include the established providers, such as Veritas or EMC, but also the Hyperscalers, which offer certified solutions.

As of May 2021, there are 42 certified backup solutions for SAP systems. These solutions do not all support the S/4HANA systems, but they have certain limitations. For example, there are providers who have focused strongly on Oracle. For all S/4HANA systems, there are 38 solutions that have achieved certification.

Regardless of which of the available solutions is to be used, certification of the solution by SAP is important. In this regard, a reference should be made to the Backint interface, which is a standardized interface for securing SAP systems.

All solutions that are certified for the Backint interface can also be used to secure S/4HANA systems. If the solution is not certified, an alternative should be considered. Backups are always an important point during audits and checks of SAP systems.

Hyperscalers offer a good starting point for backup in the cloud with native solutions. For example, they offer the following solutions:

- Microsoft: Azure Backup BackInt 1.0
- Google: Google Cloud Storage Backint Agent for SAP HANA
- Amazon: AWS Backint Agent
- Alibaba: Apsara HBR 1.6.0

Hyperscaler solutions are snapshot-based solutions that back up data to primary storage. They can be used to implement simple backup scenarios that meet the most important requirements. However, the solutions have a weakness: they do not allow efficient management of data and do not offer additional features such as deduplication and compression. The retention of the snapshots is not regulated/limited by the solutions, but the snapshots are stored until they are actively deleted again. Thus, the data remains on the primary storage, consuming storage space and causing high costs.

As an alternative to the backup solutions of the Hyperscalers, the established solution providers of backups offer their solutions. Two different classes can be distinguished here:

> Software-as-a-Service: Here, the providers offer the backup solution from the cloud. This means that the customer does not have to worry about configuring backup servers, etc., but can consume the service. The Metallic solution offered by Commvault can serve as an example here.
>
> Installation on the cloud: Here, the solutions are installed and configured as a separate installation on the cloud. This means that backup servers, agents, etc. have to be installed and configured. The customer has full control here and can make all the settings as required.

Table 2-9. *Some of the Available Backup Solutions*

Vendor	Solution	Backup-as-a-Service/ Installation on the Cloud
Actifio	**Actifio VDP 9.0**	Installation on the Cloud
AISHU Technology Corp.	**AnyBackup CDM 7**	Installation on the Cloud
Alibaba Cloud Computing Limited	**Apsara HBR 1.6.0**	Backup-as-a-Service
Arcserve	**Arcserve Backup 18.0**	Installation on the Cloud
Amazon Web Services, Inc.	**AWS Backint Agent**	Backup-as-a-Service
Microsoft Corporation	**Azure Backup BackInt 1.0**	Backup-as-a-Service
Bacula Systems	**Bacula Enterprise Edition 12.6**	Installation on the Cloud
Libelle AG	**BusinessShadow 6.5**	Installation on the Cloud
Catalogic Software, Inc.	**Catalogic Software DPX 4.6**	Installation on the Cloud
Cohesity, Inc.	**Cohesity DataProtect 6.0 for SAP HANA**	Installation on the Cloud
Cohesity, Inc.	**Cohesity DataProtect 6.5 for SAP HANA**	Installation on the Cloud
Commvault Systems, Inc.	**Commvault V11**	Installation on the Cloud
EMC Corporation	**Data Domain Boost for Enterprise Applications 4.5**	Installation on the Cloud
EMC Corporation	**Data Domain Boost for Enterprise Applications 4.6**	Installation on the Cloud
Dell Marketing LP	**Dell EMC NetWorker 19.3**	Installation on the Cloud
Dell Marketing LP	**Dell EMC PowerProtect Application Agent 19.5**	Installation on the Cloud
Linke	**Emory for SAP HANA 1.0**	Installation on the Cloud
Google Cloud	**Google Cloud Storage Backint agent for SAP HANA**	Backup-as-a-Service
Commvault Systems, Inc.	**Hitachi Data Protection Suite V11**	Installation on the Cloud

(*continued*)

Table 2-9. (*continued*)

Vendor	Solution	Backup-as-a-Service/ Installation on the Cloud
Hewlett Packard Enterprise	**HPE StoreOnce Catalyst Plug-in 2.2.0 for SAP HANA**	Installation on the Cloud
IBM – International Business Machines Corporation	**IBM InfoSphere Virtual Data Pipeline 8.1**	Installation on the Cloud
IBM – International Business Machines Corporation	**IBM Spectrum Protect for ERP 8.1**	Installation on the Cloud
Commvault Systems, Inc.	**Metallic 1.0**	Backup-as-a-Service
Micro Focus	**Micro Focus Data Protector 10**	Installation on the Cloud
Veritas Technologies LLC	**NetBackup**	Installation on the Cloud
QSFT India Pvt. Ltd.	**NetVault 12.3**	Installation on the Cloud
Rubrik	**Rubrik Cloud Data Management v5.0**	Installation on the Cloud
SEP AG	**SEP sesam 5**	Installation on the Cloud
Veeam Software Group GmbH	**Veeam Backup & Replication v10**	Installation on the Cloud

There are currently still very few providers offering a Backup-as-a-Service solution. It can be assumed that these services will increase in the coming years.

Integration and Network

Cloud computing is characterized by the fact that the services are generally available always and from everywhere. The providers of cloud computing also follow this credo, and thus all public clouds can be accessed via the Internet. However, communication of sensitive company data via the Internet is not acceptable, and so companies implement other ways to access the Hyperscaler clouds.

Access via Internet

All Hyperscaler clouds can be accessed via the Internet. This is done from the company's own network directly to the Hyperscalers. The portals of the public clouds can be accessed as normal via a browser, and so the first steps can be taken without any further costs for a network conversion.

Access via the Internet can be used as an initial option, but is certainly not a long-term solution. SAP S/4HANA systems that are provisioned and only available via the Internet are very vulnerable. In addition, all data exchanged between the SAP S/4HANA system and the company's other IT systems must pass through the Internet. This poses the risk of data manipulation, which is dangerous.

As a first step in using public clouds, access via the Internet is fine, but should be turned off as soon as possible. Access via the Internet can remain as a backup path but should then be supplemented by a virtual private network (VPN).

Azure ExpressRoutes/AWS Direct Connect/Google Cloud Interconnect

The usual way to connect an enterprise network to the public cloud is through direct links out of the network, via a wide area network provider (such as AT&T) to the public cloud.

The direct link creates a connection from the enterprise network to the WAN provider and then on to the public cloud. Here, enterprises often rely on the existing WAN providers, which then have to have the connection terminated in the appropriate regions of the Hyperscalers.

Such connections have significant advantages over a direct Internet connection:

- Higher transmission rates can be achieved, latency can be reduced, and the stability of the connection (fewer dropouts) can be increased.
- There is a higher level of security, as data is transferred point to point, and security can be implemented end to end by the companies.
- For some public cloud providers, direct connections can result in reduced costs for transferring data between/out of the cloud.

When creating the connection, it is important to note that WAN providers are not all present in all Hyperscaler regions. For example, if a customer selects the Amsterdam region and the WAN provider does not have a line there, there are two options:

- The customer can switch to another region, which happens very rarely because regions are chosen wisely and deliberately (e.g., because of regulatory requirements).
- The customer uses another region to terminate the connection and then uses the interregion connection of the public cloud providers. This results in higher hops in the network and thus higher latency, which is very important for SAP S/4HANA systems.

Since all public cloud providers have a connection between their own regions (backbone), customers often switch to the second option and connect the WAN link to another region. Before implementing a new public cloud solution, such points should be considered as they have an impact on the decision.

Creating the connection from the customer network to the public cloud requires a certain setup time. It is not uncommon for this to take more than eight weeks. Since the connection to the public cloud is usually the first step in projects, a lot of time can be lost at the beginning if the connection is ordered too late from the WAN provider.

Some companies pursue a hybrid cloud strategy. Here, the systems can be distributed across two different clouds, for example: SAP S/4HANA systems are run in the Microsoft Azure Cloud, and the non-SAP systems are run in the Google Cloud. In such a case, the company needs to create two connections (Azure and Google) and set up routing between the two connections. Such routing is often done via devices in the on-premise environment of customers. However, this involves a lot of overhead and a loss of performance due to many hops on the network. In addition, these network configurations must still be administered by customers. This can lead to a high level of complexity.

Hybrid Cloud via Cloud Connect Using the Example of Equinix Fabric

To reduce the complexity caused by many different WAN providers, manufacturers offer so-called Cloud Connects. Such Cloud Connects are to be understood as central entry points into the world of Hyperscalers. Here, customers benefit from a central entry point into all clouds and only have to keep a contract with one WAN provider, regardless of the number of connections.

In a Cloud Connect, the connection is established from the customer network via the WAN provider to a Cloud Connect provider. Connections are then made from the central Cloud Connect to the respective public clouds and the respective dial-in nodes of the public clouds. The customer does not need to use another WAN connection for this, but can use the backbone connection of the Cloud Connect providers.

Cloud Connects are a good way to offer multiple clouds via one provider. The offer from Equinix can serve as an example, which can offer a connection to the most important Hyperscalers/public clouds via the so-called Equinix Fabric:

- AWS via Direct Connect
- Azure via ExpressRoute
- Google Cloud via Carrier Peering
- SAP Cloud via SAP Cloud Peering
- IBM Cloud via Direct Link
- Oracle Cloud via Fast Connect

For customers with a maximally heterogeneous infrastructure spread across multiple clouds, a single, simple connection to the Equinix Fabric can be enough to connect all clouds together. This is a major advantage.

Comparison of Connection Types

All customers have the option to choose from the previously mentioned options – shown in Table 2-10 – for connecting to the public clouds. Thus, it is possible to start small at first and begin with the connection to the Internet. However, this cannot be recommended as a long-term connection. Then a direct connection to the public clouds should be created or a Cloud Connect should be used as an alternative.

Table 2-10. *Comparison of Connection Types*

Access	Extend Own Network	Security	Complexity
Internet	No	Very low	Very low
Direct connections	Yes	High	High
Cloud Connect	Yes	High	Medium

Although the Internet can score with very low complexity (customers can start using it immediately), the company's network area cannot extend the public cloud. In addition, the Internet does not offer reliable security for access, data transfer, and the actual protection of the public cloud.

Direct connections to the public clouds are the preferred path for many companies. Here, the network area can be extended, and the security of the SAP S/4HANA systems can be guaranteed in this case. However, the complexity can be very high due to a large number of WAN providers, direct connections with the cloud providers and routing.

Cloud Connects offer the advantages of direct connections and can reduce complexity by simplifying the network layout. The security of the SAP systems is guaranteed here in the same way as with the direct connections.

Example: Cloud Connect to Microsoft Azure in America and to Google in Europe

The company is a global corporation with subsidiaries and branch offices in America and Europe and began to gradually move its IT to the public clouds. A hybrid cloud approach was pursued in order to take advantage of the respective Hyperscalers and because there were already previous business relationships with the respective Hyperscalers. In America, for example, there were various local locations, all of which were connected to the WAN provider. From the WAN provider, there was a connection to Azure in America (ExpressRoute to Atlanta) and a connection to Google Cloud in Europe (Google Cloud Interconnect to Belgium). In addition, the company owned smaller co-location data centers, which were also connected:

– Americas: ExpressRoute to Azure to Atlanta with connection via WAN provider 1

– Americas: 15 smaller locations on WAN provider 1

– Europe: Cloud Interconnect to Google to Belgium via WAN provider 2

- Europe: Co-location data centers in Germany (e.g., Frankfurt and Munich) with connections via WAN provider 2
- Europe: Legacy data center with central switch and with routing infrastructure connecting to WAN provider 2

The company faced the challenge of having two different WAN providers in America and Europe, with which the company had to manage the connection to the clouds, as well as the routing between America and Europe (i.e., the two public clouds). To get out of this bottleneck, the company decided to rely on a central Cloud Connect and reduce the number of WAN providers to one.

The first step was to create the Cloud Connect in Europe. Here, the connection was created out of the existing WAN provider in Europe. The Cloud Connect was created in Belgium, since the Google Cloud also existed there. After that, the connections in America were rebuilt, and the ExpressRoute was dissolved. Instead, the backbone connection through the Cloud Connect provider came into use. All connections to the Microsoft Azure Cloud were handled through the Cloud Connect instead of the ExpressRoute. Further, the WAN provider in the Americas was replaced with the WAN provider from Europe. The smaller co-location data centers in Frankfurt and Munich were separated from the WAN provider in Europe and connected via the Cloud Connect provider's backbone connection. The provider also had a presence in the locations. This allowed the connections from the co-location data centers to the WAN provider to be disconnected. The company's old data center, which contained the central routing infrastructure, was also decommissioned, and routing was implemented via the Cloud Connect provider. All in all, the result was a significantly simplified picture:

- Americas: Different locations on the new WAN provider 2
- Europe: Cloud Connect via existing WAN provider 2
- Europe and America: Use of the Cloud Connect provider's backbone network

By consolidating and using Cloud Connect, the company was able to greatly reduce complexity and ultimately only had to manage one WAN provider.

Automation

The administration of SAP S/4HANA systems can be significantly simplified by automating the daily and recurring manual steps involved in administration. The same automation solutions can be used in the public cloud as in traditional data centers. Hyperscalers, however, also offer solutions for automating work on the public cloud, such as provisioning or start/stop. However, these automations are mostly limited to the services offered by the Hyperscalers. However, this can also automate some important functions and work steps. There are also tools offered by third parties that can be used in the cloud. SAP Landscape Manager (LAMA) is one example of this.

Automation Goals

The complete automation of all tasks of SAP Basis administrators is certainly not possible and is not the goal of automation. Especially when automation is started in the cloud, the following important goals are pursued:

- Process automation: The automation of repetitive work steps in the area of provisioning and operating SAP S/4HANA systems is one of the most important goals.
- Unification of the landscape: The system landscape is to be kept homogeneous. Through automation, systems are provisioned in the same way and thus unified.
- Configuration management: Many companies see the recording of the IT landscape and the maintenance of the configuration items (CIs) as challenging. The databases containing all configuration data (CMDBs) are often not always accurate and maintained. Leveraging Hyperscaler automation can address the issue.
- Keeping landscape up to date: Patching systems on a regular basis is becoming increasingly important due to the growing number of cyberattacks on businesses. For this, Hyperscalers bring the necessary tools to keep the SAP infrastructure up to date.
- Auditing: Audits require accurate data and often a lot of data from the companies. For this purpose, Hyperscalers offer simple options to generate important reports and pass them on to the auditors.

- Tagging: Tags are small pieces of information about SAP S/4HANA workloads that can be assigned to resources in the Hyperscaler portals. This can be used, for example, to identify which resources belong to which projects/departments/teams.
- Schedules: When SAP S/4HANA systems are switched on in the cloud, costs are generated. To minimize costs for systems that do not always need to be 100% available, schedules can be set up that turn resources on and off.

All of the preceding goals can be achieved in the rarest of cases, but rather companies focus on successive implementation. Furthermore, the implementation of complex automation is a very long-term undertaking.

Automation with Hyperscalers

All Hyperscalers offer customers the ability to automate using their own cloud-native tools. These automations can be very lightweight, but can also become increasingly complex.

Simple implementations involve standard operations in a cloud, for example:

- Starting and stopping virtual machines
- Creating, deleting, and modifying resources in the clouds (meaning all resources).

The simpler tasks can be performed using the tools offered by Hyperscalers, such as Azure Automation or even Google Composer.

The somewhat more complex activities, such as creating a new SAP S/4HANA system, are usually not possible via the cloud-native tools. For this, the following steps must be performed, for example:

1. Creating all infrastructure components (storage, virtual machine, network segments)
2. Creating/installing the operating system with the hostname and the desired IP configuration
3. Installing an empty SAP S/4HANA system
4. Filling the SAP S/4HANA system with data

Since the steps must be coordinated with each other, additional tools are required for automation. The automation tools that are already widely used in the public clouds, such as Ansible or Terraform, are ideal for this purpose.

Third-Party Vendors

Automation solutions exist and existed long before Hyperscalers entered the market. Repetitive administration steps were also automated in traditional data centers. At the latest since the introduction of virtualization (e.g., VMware), many tasks have been completely automated by the tools supplied by the virtualization manufacturers. Since Hyperscalers entered the market, there has been a proliferation of automation vendors. At the beginning of public clouds, Hyperscalers did not offer automation solutions, which allowed such vendors, such as Ansible or Terraform, to establish their position in the market.

One of the advantages in using third-party vendors is their strong focus on scripting and reusability. Most third-party solutions use a repository of scripts to get the job done. This repository can be extended with custom scripts, which are specifically customized.

By storing the scripts in a repository and using a third-party solution, Hyperscaler customers can take the existing scripts with them even if they switch platforms (i.e., from Azure to Google Cloud, for example) and use them in the new environment. However, if the customers only use the Hyperscalers' own automation, the work may not be able to be reused.

SAP's Own Automation

It makes sense for all customers to use SAP's own products for automation. Using SAP's solution seems logical, since SAP knows the S/4HANA systems best and can therefore build the appropriate software for them. SAP offers a solution for automating SAP Basis work (SAP LAMA) and a solution for automating repetitive work outside of technical topics (SAP RPA).

SAP Landscape Management is an evolution from SAP Landscape Virtualization Manager (SAP LVM) and SAP Adaptive Computing Controller (SAP ACC). Both products were initially used by SAP in its own data centers. SAP quickly recognized the benefits of the solution for, for example, mass operations in the data centers and quickly extended the solutions with some very interesting functions, such as copying SAP systems. SAP does not necessarily focus on the most common tasks that SAP Basis administrators have to do, but rather on the issues that require a lot of time and energy. These include the copying of SAP systems (copy, clone), as well as the so-called system refreshes (i.e., a copy from PRD to QAS with a renaming of the SID).

Due to its history (first SAP ACC, then SAP LVM), SAP LAMA offers very good support from other technologies, such as hardware manufacturers. For example, SAP LAMA can communicate with NetApp storage via a library. This support was mandatory when SAP ACC and SAP LVM implemented the first functionalities that also required addressing the storage in the data centers. Using the same principle, SAP LAMA also interacts with the public clouds, which all provide an interface for administration.

SAP LAMA exists in two different versions. On the one hand, customers can set up SAP LAMA directly in the data center or the cloud. On the other hand, customers can also use the SaaS version of SAP LAMA. This is provided by SAP, and therefore customers do not have to worry about the administration of the SAP LAMA system. This is also an SAP system that needs to be administered.

SAP Intelligent Robotic Process Automation

In addition to purely technical tasks, companies are also trying to automate other tasks. These are mostly recurring tasks and process steps that can also be performed by nonhuman resources. One way of automating these is SAP Intelligent Robotic Process Automation (SAP iRPA). Robots (bots) are small programs or scripts that perform exactly the same process steps over and over again. The process steps can also be refined via parameterization. Robotic solutions always have to contend with various challenges. Since the robots work on the user interfaces (UI), changes to the UI can result in major changes to the robots, which means that a large number of robots may have to be adapted. This is where SAP comes in and extends the solution. The intelligence in iRPA comes from SAP's use of more intelligent features: machine learning and artificial intelligence. In addition, the robots can now act using interfaces.

SAP has built iRPA as a pure cloud-based solution. Thus, new robots are created in a Cloud Studio via pure UI (without code). Currently, the solution has 200 preconfigured robots, which can be easily customized and deployed. The robots are used in many companies, but less in the technical environment of the administration of SAP S/4HANA systems.

Implementation Effort

Automation of daily work of SAP administrators is quite reasonable and very helpful. Only in this way companies manage to free the administrators' working time from daily tasks, and the administrators can take care of other tasks (such as the introduction of new features).

However, the implementation of automation solutions can become very complex. When using solutions provided by Hyperscalers, customers can get started very quickly and implement initial smaller tasks very quickly. However, as soon as the processes become more complex or several process steps need to be connected, more time needs to be invested for this. The complete implementation of a new solution, such as SAP LAMA, must be carried out as a separate project.

Horizontal and Vertical Scalability

The scaling of SAP S/4HANA systems refers to two directions and is differentiated into horizontal and vertical scaling. Horizontal scaling in SAP S/4HANA systems refers to the expansion of the existing system with more, additional HANA nodes or more application servers. Vertical scaling refers to the growth of the system and the increase in the sizing of the virtual machines and the SAP system due to increased requirements from the business.

Horizontal Scalability

Horizontal scaling describes the expansion of the SAP system through additional application servers or through additional HANA nodes. However, the expansion of the SAP system through additional application servers is not new to the SAP S/4HANA systems, but has existed since the beginnings of the SAP R/3 systems. Back then, too, further dialog instances could be added as additional application servers.

In principle, there are no limits to horizontal scaling, and up to eight or more application servers can be used for very large SAP S/4HANA systems. Often, the additions of application servers are not so much due to the performance limits of the individual servers, but rather certain application servers are only made available to certain groups of users. It is also often the case that application servers are only used for background jobs, since an influence on the end users by long-running jobs is to be avoided.

However, horizontal scaling is not only used for application servers but also for HANA scale-out systems. These HANA systems are used for large business warehouse systems and can thus hold very large amounts of data in the main memory, but distributed across multiple nodes. Especially for BW systems, the requirements for strong data growth exist. The background of the growth is the closings, where a lot of temporary data has to be stored, as well as large data load runs, which also lead to a lot of temporary data. Thus, BW systems show a significantly larger data footprint. This growth is addressed by provisioning through more HANA nodes.

The typical growth of the entire SAP S/4HANA systems is not addressed by horizontal scaling, but usually by vertical scaling.

Example: In many larger implementations of SAP S/4HANA systems, more than one application server is implemented. Very often, individual departments use the respective application servers.

A company from the pharmaceutical sector implemented four application servers for the most important S/4HANA system:

1. Application server: This server was intended for the finance department (FICO). Using the principle of logon groups, the server was integrated into only one group, which was also known only to the finance department.

2. Application server: This server was assigned to the majority of users. Also, here the principle of the logon groups was used, and the group DEFAULT was used for this.

3. Application server: This server was used for all incoming RFC connections. The background was security considerations on the one hand, but also the specific nature of RFC calls and the incoming transactions plus IDOCs.

4. Application server: This server was used for the long-running background jobs. Since there were a lot of background jobs scheduled in the SAP S/4HANA system (often with frequency of ten minutes), the customer decided to have a dedicated application server for this.

By separating the individual areas from each other, the company was able to better separate the load by users, increased security and the overall performance of the SAP S/4HANA system.

Vertical Scalability

Vertical scaling describes the growth of the SAP S/4HANA system through an increase in the size of the servers (HANA servers or application servers) or an increase in the resources of the components. This case occurs when more resources are needed than are available. This can happen when additional users gain access to the system or when new functionalities are to be provided.

Vertical scaling can therefore occur at different levels of the SAP S/4HANA system:

1. Changing the storage class: For some SAP S/4HANA systems, changing the storage from slower SSD storage to fast SSD-based storage may be necessary. This is a vertical scaling, which means more of a change effort, but can provide a huge speed increase.

2. Changing the VM template: For growing systems, other types of virtual machines can be used, which have more CPU and more main memory. These changes usually cannot be made on the fly, but require a reboot of the system. After a new size is used for the VM, the operating system and the SAP components (i.e., either the HANA database or the application servers) still need to be adapted.

3. Changing the HANA database: Changing the HANA database is one of the most common use cases. Here, for example, the allocation of main memory or the number of CPUs is changed to achieve higher performance.

4. Modification of the operating system: After increasing the size of the virtual machine or changing the storage, modifications to the operating system may also be necessary.

5. Modification of SAP application servers: Changing the buffers or even the number of disp+work processes is one of the most common adjustments to scale the SAP S/4HANA system.

Vertical scaling is often done not only for one component but is done for several of the preceding components. An increase in the size of the SAP HANA database can only take place if the virtual machine and the operating system have been adapted beforehand.

Vertical scaling works in both directions: resources can be added and systems can be enlarged. But resources can also be removed and systems downsized again. In the cloud environment, upsizing and downsizing is very important because it can save costs. In the course of effective capacity planning, upsizing and downsizing should be constantly considered. It helps here if the utilization of SAP S/4HANA systems and the respective components is considered over longer periods of time.

Consider Availabilities

Although public cloud providers always talk about unlimited availability of resources, there are limitations. Before planning capacity expansions, customers should ensure that the corresponding target VM templates are also available in the respective regions. It may well be that certain VM templates are not yet available in all regions, and customers will then have to switch to alternatives. Hyperscalers typically always have one/two regions where the latest technology is deployed earliest. Smaller regions for specific markets often cannot offer the full portfolio of services.

At the beginning of the COVID pandemic, there were also resource bottlenecks, sometimes significant, among public cloud providers. This was due to the enormous increase in the use of other Hyperscaler services. This high usage meant that other customers had no or only limited new resources available. The Hyperscalers use the public clouds not only for customers but also for their own services. At the time, it was not possible for customers to simply provision new virtual machines.

Very specifically, customers could no longer order large HANA systems (e.g., M-Class) in the Microsoft Azure Cloud. The underlying machines were used for other services. Microsoft struggled with a surge in Microsoft Teams users at the start of the pandemic. With many employees suddenly having to work from home, resources for MS Teams were steadily expanding. In addition, new requirements for fighting the pandemic emerged, with research institutions creating new and large systems on the cloud to perform complex calculations. All of this ultimately led to a situation where demand could no longer be met by other customers. Microsoft responded by introducing prioritization and requiring customers to register their needs. After such notification, Microsoft decided based on the criticality of the need and allocated (or not) resources to customers.

This example of the pandemic shows that there are limits to scaling in Hyperscalers. Even though such scenarios are very unlikely, bottlenecks can always occur, which nevertheless limit the seemingly unlimited resources.

Summary

This chapter has described the most important aspects of using public clouds for SAP S/4HANA systems. The right sizing of the systems is not only important from the point of view of users and performance but is particularly important because of the costs of such systems. An oversized system did not cause higher costs in the traditional data center world – but in the cloud, this is different and generates high costs. This must be avoided.

In the case of productive SAP S/4HANA systems, it is important to absorb the possible failures of the components of a public cloud (such as the virtual machines) and to continue to ensure the availability of the SAP S/4HANA systems. This was explained in the section on high availability, and an example architecture was also discussed, which is implemented in concrete terms in the sections on the respective Hyperscalers.

In the section on disaster recovery, it was shown how the SAP S/4HANA systems continue to be available through the onboard resources of the Hyperscalers even if a region fails. This can also be done by combining onboard means and the mechanisms of the SAP systems (such as HANA System Replication). The respective sections on Hyperscalers describe the implementation and guide you through the setup step by step.

The two important topics of backing up and restoring data were described, and the options for backing up data using Hyperscaler's own onboard resources, as well as third-party products, were discussed. These differ in how they are used and also in the knowledge required about the products.

The topics of integration and automation were described in this chapter, showing how easy it is to start using public cloud services. Skillful automation can reduce the initial effort required for cloud projects. Automation can save a lot of time and effort, particularly in day-to-day operations.

In order to always meet the requirements of the business departments and to meet the growing demands on the SAP S/4HANA systems, horizontal and vertical scaling were discussed. It is shown how this can be implemented in SAP S/4HANA systems.

The next chapter shows the deployment and migration of SAP S/4HANA systems to the public cloud.