Integration Strategy at SAP

CIO Guide: SAP Vision for Integrating SAP® Applications in Cloud and Hybrid Environments

Table of Contents

This CIO guide is concerned with SAP's **long-term approach to integration**. It emphasizes integration scenarios in the cloud and in hybrid environments. Special attention is given to the integration scenarios related to process and data-centric integration. The document also presents SAP® Data Hub, the data landscape management solution that enables agile data operations across the enterprise and supports data sharing, pipelining, and governance of all data in the connected landscape.

Executive Summary

SAP pursues the goal of making future integration in the cloud and hybrid landscapes easier and simpler. For process integration scenarios, we are establishing an integration approach that is based on two complementary pillars: public, fully documented application programming interfaces (APIs) and route-through middleware integration content. Both will be managed with our strategic middleware and core service SAP Cloud Platform Integration (which is the preferred option),[1] or a recent version of our on-premise SAP Process Orchestration software.[2] Future public APIs will be the elementary basis upon which SAP, its customers, and its partners will build integrations. SAP aims at creating these new public APIs based on aligned models for new process integration scenarios. The preferred protocols for these APIs will be representational state transfer (REST) and Open Data Protocol (OData). Aligned APIs will allow customers and partners to reduce the complexity of middleware content. This will translate in reduced total cost of operations (TCO) because it will minimize or even eliminate transformations in custom middleware content for SAP-to-SAP application integrations, particularly in the cloud. At the same time, extensive and more uniform integrations based on these APIs will result in a better user experience. Customers and partners will be able to build upon the same rich APIs that the SAP development organization will be using. Complementary to the aligned APIs, SAP will provide prepackaged, public route-through middleware content that customers can extend and modify. Since customers can switch to a middleware-based integration anytime, they can concentrate on their specific value-adding goals when building or extending this middleware content. For scenarios where SAP controls both ends of the integration and where SAP aligns consumers with public APIs, aspects such as upgradability and extensibility will also improve.

SAP will apply this future process integration approach in steps to new integration scenarios involving SAP applications. Complex scenarios, scenarios requiring customer extensions, integrations involving third-party or certain on-premise SAP applications, and integrations in the context of business-to-business (B2B) or business-to-government (B2G) communications will continue to require mediated or middleware-based integrations. For this reason, switching to mediated communication should be as simple and easy as possible. Even in these cases, public aligned APIs will make it easier for customers to build their own middleware content, so that they can integrate, for example, with third-party systems. This is precisely where customers and partners will take advantage of SAP Cloud Platform Integration content that SAP centrally provides in SAP API Business Hub. They will use this integration content as a reference template. After importing the public integration content from SAP API Business Hub into SAP Cloud Platform Integration design time instances, customers and partners may customize or extend this template content as needed. Afterward, they can execute the content in SAP Cloud Platform Integration runtime instances, which are hosted in major SAP data center locations for optimum performance. Customers will also be able to carry out selected configuration work in SAP Cloud Platform Integration design time, for example, to set up end points by themselves for integrations that do not require middleware.

1. The SAP Cloud Platform Integration service was previously known as SAP HANA Cloud Integration.
2. Since SAP Process Orchestration (version 7.5 and higher) supports the integration flow format of SAP Cloud Platform Integration. It can be used as middleware for hybrid scenarios too. Therefore, customers may choose to deploy integration content for SAP applications in the cloud into either SAP Process Orchestration or SAP Cloud Platform Integration.

API management solutions from SAP may well complement the above-mentioned integration options, as they provide orthogonal capabilities such as policies, security, and traffic management. Additionally, to help define the integration strategy that best suits a company's needs, SAP offers the integration solution advisor methodology (ISA-M).

In addition to dealing with process integration, this document discusses the common data-centric integration scenarios: extract, transform, load (ETL, both classic and real time); data replication; and data virtualization. Depending on whether your system landscape is centered on the SAP HANA® business data platform or is made up of heterogeneous systems, we provide different technology recommendations.

SAP Data Hub is a new category of software solution that provides a comprehensive answer to an emerging and painful challenge for enterprise customers: integrating data and establishing data-driven processes across an increasingly diverse data landscape. SAP Data Hub addresses data integration, data orchestration, and data governance capabilities across a complex landscape, as well as harnessing Big Data processing to create uniquely powerful data pipelines that are based on the serverless computing paradigm. Data and processes can be managed, shared, and distributed across the enterprise with seamless, unified, and enterprise-ready monitoring and landscape management capabilities. SAP Data Hub is an open solution that recognizes and reduces the complexity and diversity of today's enterprise data landscapes. It leverages other data-centric integration technologies mentioned in this document.

While SAP Cloud Platform Integration focuses on business process orchestration, SAP Data Hub clearly concentrates on data processing across landscapes. Therefore, these two technologies work together to get the most value out of data; for example, after SAP Data Hub manages processing of sensor data in a data lake, the resulting insights could be used as input for an ordering process.

Introduction

Integration has always been a challenge for software consumers and vendors alike. But it is a necessity and a fundamental prerequisite for streamlining business processes and for getting the most value out of data, which is often not readily accessible because it may exist in large quantities and in diverse formats, and can be spread over a complex system landscape.

Currently, SAP offers a variety of integration technologies and implementation patterns. We realized that some customers may find it difficult to have many different integration technologies to choose from. For this reason, we have developed a vision for a simpler integration approach in the cloud and in hybrid environments. For process integrations between SAP applications, SAP Cloud Platform Integration will be the preferred solution, since it will provide customers and partners with the flexibility they need to build integrations. This document explains how this will work for future integrations between new SAP applications. Customers already using our on-premise SAP Process Orchestration may continue to do so for hybrid scenarios.

Additionally, SAP recognized the need of modern enterprises to attain a centralized system view for data governance, monitoring, and orchestration. At the same time, mass data movements to a central, single store are often impractical, costly, and time-consuming. Today's complex distributed system landscapes demand new solutions for comprehensive and controllable data management. SAP Data Hub is SAP's answer to these challenges for enterprise customers. It builds upon many years of history in the integration realm, as well as on a vast experience in the areas of enterprise information management and data services.

Our overall goal is to follow a simpler and more uniform approach to integration across all our cloud applications. SAP is aware that much is still to be done to achieve this goal, and we are currently working on further simplifications and systematic unification of the respective approaches.

Although we intentionally recommend only a subset of the available integration technologies, customers are free to continue to purchase or use other technologies, depending on their specific needs.

This document covers the integration scenarios related to process and data-centric integration. Despite their current relevance, scenarios involving user interface integration, mobile integration, analytics, and integrations in the Internet of Things area go beyond this scope and therefore are not addressed in this guide.

This document is structured into the following sections:

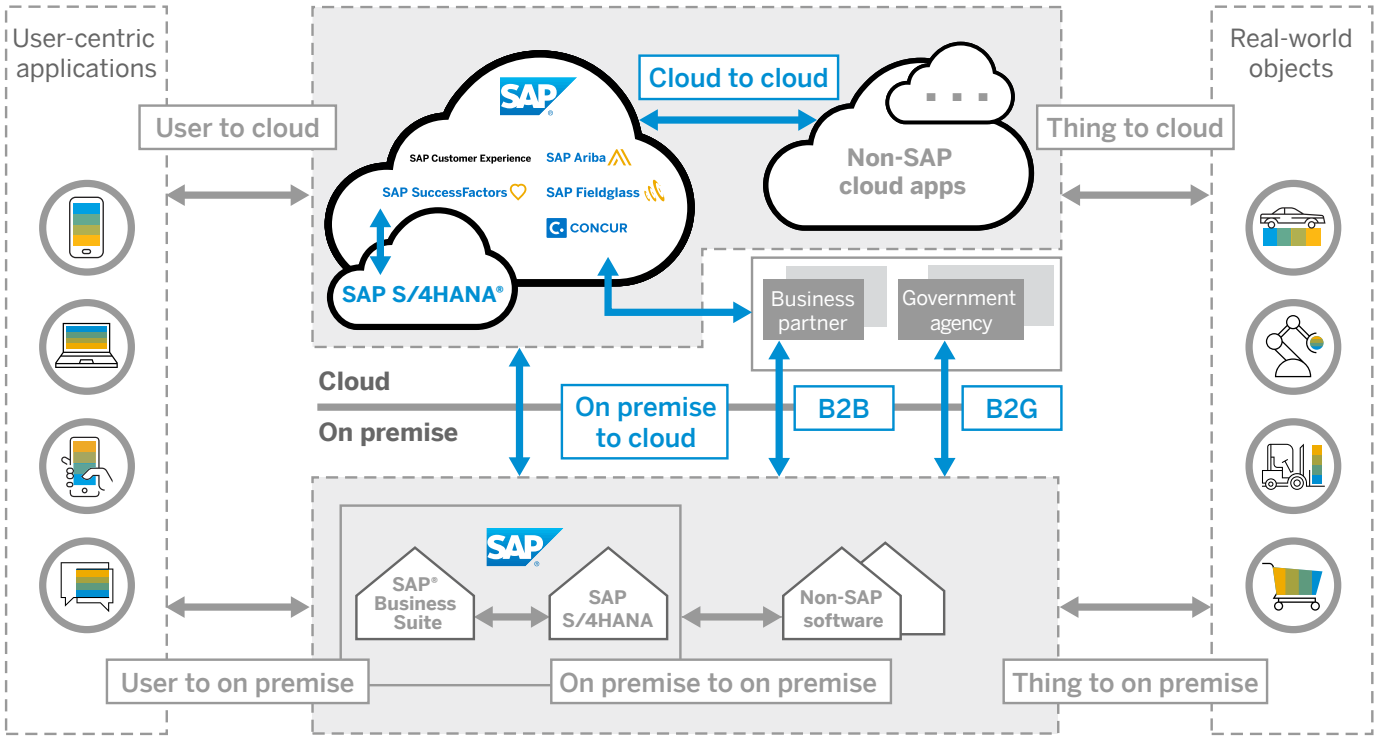
- "Integration Challenges" outlines the growing integration demands that arise when dealing with various integration domains. Many customers are dealing with hybrid landscapes spanning both cloud and on-premise deployments distributed around the globe. The associated requirements not only challenge the enterprise architecture but also offer opportunities for simplification and savings.
- "Key Integration Use Cases" describes the most important characteristics of the use cases that relate to process and data-centric integration. This background information is intended to facilitate your understanding of the future integration approach.

- “Future Process Integration Architecture” explains a high-level model for the future process integration architecture. This section also discusses the impact of application model design on this architecture. APIs are dealt with in the context of the new architecture, as they are an essential element in SAP’s vision for a consistent integration approach. While discussing the integration architecture, aspects related to design time, runtime, and operations are also considered. Finally, this section explains SAP Cloud Platform Integration in some detail and refers to further information.
- “Process Integration” examines how the future integration methods for application-to-application (A2A) and master data synchronization will be applied. It also describes the circumstances under which mediated communication and communication through aligned APIs work and the option to easily switch from one method to the other. This section ends with a summary of technology recommendations for the process integration–related use cases.
- “Data Integration and Virtualization” covers integration scenarios involving ETL and replication, as well as the related usage patterns. It concludes with a summary of the recommended technology for each of the commonly encountered use cases.
- “SAP Data Hub: Data Landscape Management Layer” explains SAP Data Hub, a new category of data landscape management solution that enables agile data operations across the enterprise. In a single solution, SAP Data Hub provides data governance, pipelining, and sharing capabilities for the connected landscape, making it possible for companies to accelerate and expand the business value of their data and reduce the complexity and diversity of their data landscapes.
- “Find Out More” provides links to further information.

Integration Challenges

Integration becomes a key enabler for the digital transformation of organizations. Figure 1 illustrates the integration domains and types of scenarios encountered when heterogeneous components and systems need to be integrated. The systems to be connected typically "speak" different communication protocols and store their data in different formats and structures. Figure 1 shows the main domains according to the integration solution advisor methodology (ISA-M) model. As cloud, mobile, and Internet of Things (IoT) scenarios are added to customer landscapes, the scope of integration continuously increases. Customers are challenged to extend their integration architecture to address these new domains of integration to exploit the richness of opportunities these areas offer.

Figure 1: Illustration of Integration Domains in Hybrid Environments



B2B = business to business, B2G = business to government

On their journey to adopt innovations in the cloud, customers increasingly extend and integrate their existing on-premise applications to the cloud. Customers already using cloud solutions expect on-premise and cloud solutions from SAP to integrate with each other, as well as with non-SAP applications. The integration of business partners and business networks (B2B) continues to be a key integration domain for customers. The same applies to integration between businesses and governmental agencies (B2G).

User-centric applications such as SAP Fiori® apps or mobile apps need to be integrated with applications in the cloud or on premise to provide an omnichannel user experience. With the emergence of the IoT, customers face new challenges: they need to integrate data from real-world objects such as sensors or assets with cloud or on-premise applications, either for analytical or transactional processing. To address these growing integration demands, customers and partners require a consistent and flexible cloud integration platform that covers process, data, user, and IoT-related integration scenarios. Such a platform is expected to lead to less implementation effort, better control and compatibility, and greater robustness, particularly in case of upgrades.

Enterprise architects defining the integration strategy in their company's system landscape usually try to find the best way to provide integration guidance across multiple teams, projects, and system integrators. For them it is important to look for the most suitable integration technology to approach new integration domains. The goal of the future SAP integration approach is to increase the straightforwardness of integration between SAP applications by aligning their processes and related data models, including publication of APIs. The purpose is to simplify new integration solutions, especially for line-of-business (LoB) cloud integration scenarios, and to further deepen integration between SAP applications over time.

In addition to the technology guidance given in this CIO guide, SAP provides the integration solution advisor methodology. This framework allows enterprise integration architects to blueprint their overall integration strategy based on a set of well-defined integration use case patterns. For more information on ISA-M, click **here**.[3]

As digitalization becomes an increasingly crucial topic, the accelerated growth in the digital space, characterized by exploding data volumes coming from various sources and having different types, poses additional challenges to today's enterprises. To exploit the value of these ever-growing data quantities, companies need to overcome silos and tackle the complexity of landscapes while getting better control of operations. SAP is committed to helping enterprises overcoming the challenges of growing, complex, and heterogeneous data landscapes. The purpose is to quickly drive results and business value from data distributed throughout this diverse data environment. Productization of complex data scenarios across data landscapes is difficult and costly due to limited tools that lack enterprise readiness. These new challenges require new technologies that can deal with data from systems in distributed landscapes comprising different data sources. Ultimately, it is about enabling comprehensive and controlled data management even in very complex system landscapes.

3. https://blogs.sap.com/2016/03/04/int203-integration-solution-advisor-methodology-isa-m-sap-teched-lecture-of-the-week.

Key Integration Use Cases

APPLICATION-TO-APPLICATION INTEGRATION

A2A integration covers the interlinking of business processes between one business's applications. For this kind of integration, it is particularly important to ensure process integrity, as well as the ability to monitor and trace on a single business-object level. Moreover, connectivity to a third-party software-as-a-service (SaaS) provider should also be considered. Orchestration capabilities are used to manage the message choreography between applications (for example, pull and push interactions). A further characteristic of this use case is that business users typically handle errors derived from failed messages in their application context (for example, processing of failed invoices in financial transactions). The integration of the SAP Cloud for Customer solution with SAP S/4HANA® for the purpose of facilitating the opportunity-to-order process is an example of A2A integration.

BUSINESS-TO-BUSINESS AND BUSINESS-TO-GOVERNMENT INTEGRATION

B2B integration refers to process interlinking between business partners. In addition to the requirements of the A2A integration, this use case requires support of a variety of industry standards – such as electronic data interchange (EDI) and Applicability Statement 2 (AS2) – and management of a potentially large number of business partners. As in the A2A use case, business users should be able to handle failed messages in their application context. Closely related to the B2B integration is the B2G use case, which covers the integration with governmental agencies and authorities. An example of this is the transmission of electronic documents to tax authorities, typically tailored to meet country-specific requirements.

MASTER DATA SYNCHRONIZATION

This use case refers to synchronization of master data (initial load or delta changes) between a system of records (source) and dependent applications (target). Master data synchronization enables execution of end-to-end business processes. Cost center replication between SAP S/4HANA and the SAP SuccessFactors® Employee Central solution is an example of this use case. Typically, master data is exchanged either in near-real time or batch. A further characteristic is the processing of mass or bulk data, optionally including data matching, cleansing, and consolidation.

DATA VIRTUALIZATION

Data virtualization provides users with real-time access (read and write) to data from an external system through a standardized interface (for example, SQL), regardless of its physical location (on premise, private cloud, or public cloud), how it is stored (database or file), or how it needs to be accessed – Java Database Connectivity (JDBC), Open Database Connectivity (ODBC), OData, REST, and so on.

DATA INTEGRATION

This use case is purely about data exchange between systems. In relatively simple system landscapes, this type of data exchange requires little or no orchestration at all and is not part of a well-defined business process. The decision to use data integration in addition to the required A2A capabilities mainly depends on criteria such as performance (mass or bulk data handling) and the development model on the data consumer side (for example, the consumer code may use SQL access instead of direct API calls). To load data, a consumer using a data integration technology can use a variety of technical APIs (JDBC, REST, remote function call [RFC], SOAP, OData, ODBC, and so on) that bypass the application and directly manipulate the data. The systems involved in the data exchange will not always have aligned data structures or data values (such as in the case of reference data). Therefore, it may be necessary to transform, cleanse, and enrich the data as part of the data exchange process.

USE CASES IN COMPLEX DATA LANDSCAPES

In addition to the integration use cases described above, special needs derive from complex data landscapes. Concrete tasks include data ingestion, data processing, orchestration of complex processes and workflows across system boundaries, control, operationalization and productization of complex data landscapes, metadata lifecycle with lineage and impact analysis, and data discovery.

Enterprise customers having such heterogeneous data landscapes need to manage data processing across data lakes, object stores, cloud and on-premise databases, as well as data coming from cloud and on-premise applications. A distributed data processing is required to perform data transformations, ensure data quality, and enable data preparation processes through a graphical user interface. To deal with the challenges that these complex data landscapes pose, customers require the ability to define data pipelines for native processing and orchestrate complex processes, including creation of workflows for operations and processes, with monitoring and analysis capabilities. Data processes must become transparent, from data ingestion into the landscape (and its processing) up to the integration of the resulting data into enterprise processes and applications. Data ingestion and processing must support both unstructured and structured data.

Ultimately, the use cases associated with complex data landscapes are about integrating the increasing number of (heterogeneous) systems, together with growing data quantities, not only to obtain a transparent view of the entire data processing but also to define, analyze, and monitor it to quickly drive results and business value from data distributed throughout the diverse data environment.

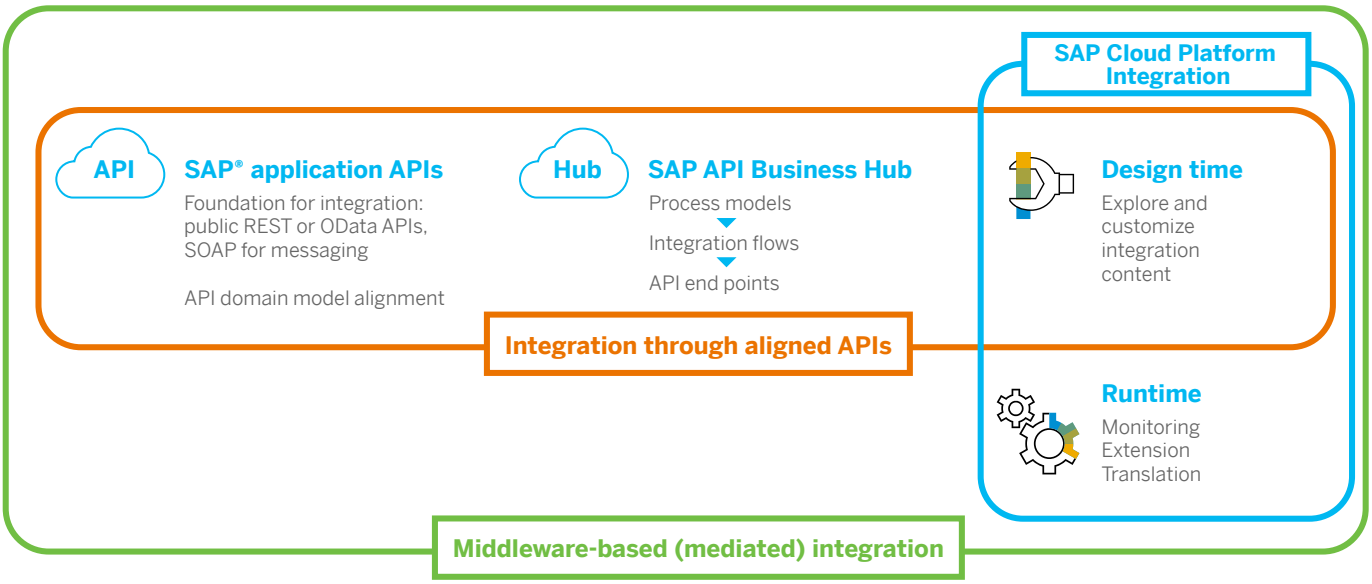
Future Process Integration Architecture

In this section, we will look at the architectural guidelines that SAP is establishing across its solution portfolio to support cloud-to-cloud and hybrid integration scenarios. The guidelines will apply to future (mainly A2A) integration scenarios. They will be adopted over time, starting with selected new applications.

First, let us consider the process integration scenarios involving SAP applications. In this setup, we distinguish three cornerstones (see Figure 2):

- Public APIs
- SAP API Business Hub
- SAP Cloud Platform Integration, our strategic middleware and core service of SAP Cloud Platform, which consists of a design time and a runtime.

Figure 2: Architecture Cornerstones for Future Application Integration in Cloud and Hybrid Scenarios



To implement our future integration approach, new applications will expose their business functionality as public APIs. These APIs not only will be the foundation for integration but also will be used for other purposes, such as facilitating extension applications. Since these APIs are public, customers and partners will also be able to take advantage of them. To help ensure completeness of functionality and usability, we will use REST or OData as the preferred protocols for these business APIs, and SOAP for asynchronous messaging.

SAP will continue to deliver further integration content for new hybrid and cloud integration scenarios, which can be used as a reference template for customizations. SAP API Business Hub is the place in the cloud where all integration content is being made accessible to consumers. Through this hub, we aim at achieving new levels of openness and transparency. The hub will combine a view of the existing business APIs with a view of the available integration content in one common platform. In the latter, customers can already find prepackaged integration content,

starting with high-level cross-application process models and data flow that describe the integration. From the process models, they can drill down to the integration flow integration content and from there to the actual API documentation (see Figure 2). This will be available for all public APIs and will include examples of API usage and more documentation. Customers can test each API against a sandbox or even against real end points in their landscape.

SAP Cloud Platform Integration is the strategic integration service within SAP Cloud Platform. It facilitates the integration of business processes that span different departments, organizations, or companies. It acts as a "broker" between the systems that can be connected with each other. The integration service then arranges for the technical communication between the components and the correct "translation" of data. Later in this section, there is more information about SAP Cloud Platform Integration.

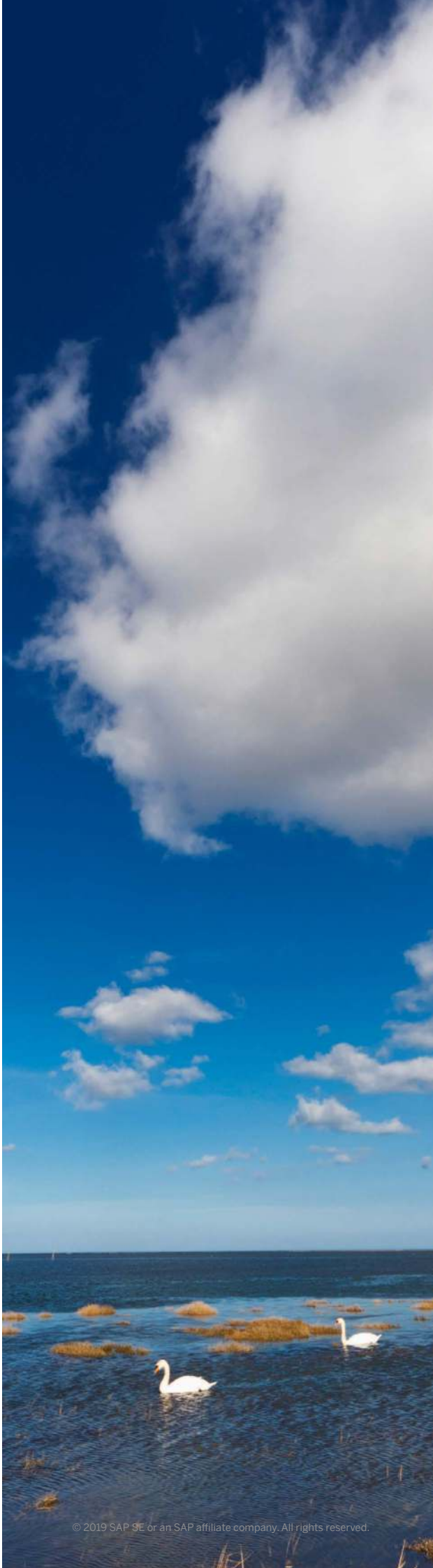
As stated previously, integration based on aligned APIs will become a fundamental option for integrating SAP applications in cloud and hybrid scenarios. Customers will be able to use dedicated configuration APIs to configure end points in SAP Cloud Platform Integration design time, where the end points can also be centrally managed (see Figure 3).

Another essential part of the future integration approach is giving customers the option to switch to a mediated integration, for example, by using SAP Cloud Platform Integration runtime, as required. Or, if they prefer the integration option based on aligned APIs for new SAP-to-SAP scenarios, they will definitely benefit from standardized protocols and simplified integration content that will be easy to customize and extend. Moreover, integration through aligned APIs will help to ensure high-quality standard integrations with a low TCO, as actual data and messages are sent using the most efficient protocol for any given scenario. The objective is to allow standard SAP-to-SAP integration scenarios to potentially run immediately without requiring configuration of middleware content. Customers with a company-wide hub strategy can, of course, deploy the route-through content that was created for integrations based on aligned APIs into their hub and customize it afterward. SAP Cloud Platform Integration design time will support the ability to manage and customize the integration scenarios based on aligned APIs. The service design time may be used to document all scenarios between systems, even when no actual translation is required and the communication does not take place through a middleware runtime (route-through content). At the same time, this capability will enable customers to benefit from a centralized view of the setup communication while using optimized network communication paths at runtime. A runtime can be seamlessly switched on anytime to execute mediated content derived from our standard integration content. For example, customers can add processing steps that involve additional systems in their landscape to the route-through content, and then enable message flow through the middleware runtime.

Complementary products such as SAP API Management technology can be optionally employed for both integration options (with or without middleware) to add capabilities such as traffic management (or throttling), access control, metering, or logging, if these features are needed for security purposes.

Our future integration approach will also focus on achieving so-called integration qualities, such as extensibility, master data integration, robustness, automation, integration monitoring, configuration alignment across SAP standard products, and identity support, among others. SAP is currently defining these and other qualities for bench-marking integrations.

The target architecture that corresponds to the integration approach relies on application domain model alignment and commonly acknowledged principles for exposing APIs. We will discuss these next.



APPLICATION DOMAIN MODELS AND INTEGRATION

A domain model defines an object model (data model, types, or values) as well as the set of possible actions and events.

Domain models are at the core of every SAP business application. For example, the applications that belong to the finance LoB define a domain model for accounting and controlling. Each domain model may be further subdivided into subdomains. The domain that describes the primary purpose of an application is called the core domain. To fulfill its purpose, an application may also use other supporting domain models. For example, travel itineraries and travel expense reports are the core domain of SAP Concur® solutions. For an expense report, the controlling domain is merely a supporting domain, as a cost object is needed to file an expense report. However, SAP has no primary interest in evolving these domain concepts. Therefore, our long-term goal is that applications follow the best-practice domain model that the leading application provides. Domain models manifest themselves not only in several internal models, such as application object models or persistency models, but also in external API models. For new applications, SAP seeks to align mainly the external API models in a first instance. An even farther-reaching alignment (down to the level of internal models) is only feasible and possible for newly developed applications and will be applied on a case-by-case basis.

Domain models play an important role in integrations. Prerequisite for integrating two applications without using a middleware is that the API layers of the applications are semantically aligned with regard to a leading core domain model. Therefore, at the time of application development, the leading application that will provide the core domain by means of an API should be determined. Consumers and nonleading providers should then aim at aligning their external APIs with this leading model. For both SAP and its customers, the key value of model alignment is a more uniform approach to concepts across SAP applications. This means that customers need to spend less effort managing data mappings between SAP applications. For SAP and its partners, this future approach will generate new SAP applications with deeper integrations and higher interoperability. In addition, customers will benefit from this alignment because it is expected to reduce the costs of maintaining customized middleware content. Moreover, lower costs can be expected to result from simpler system landscapes and reduced dependencies on middleware runtimes executing data translations. Upgrades of custom integration content, created based on future SAP standard route-through content, will be easier because customers can completely define the content and no content merging is required during the upgrade process.

APIS

As stated earlier, SAP envisions an integration architecture that is based on public, aligned APIs to ensure that our systems are open to integration with other applications and extension applications. By applying this integration approach, new applications will qualify for integration using fully documented public and aligned APIs, with REST and OData being the preferred integration protocols. We selected the REST protocols because of their pervasive use in the cloud today. Moreover, REST uses HTTP as the transport protocol and benefits from extensive standardization and tool support, such as API gateways, proxies, and others. REST also provides support for virtually any language.

OData is a special family of REST services. OData goes even further in terms of standardization and capabilities. OData gives consumers great flexibility in accessing exactly the data that is required for a given scenario. We will promote the use of OData under specific circumstances, most notably when OData providers are already available in an application environment – for example, an ABAP® programming language environment with SAP Gateway technology. Otherwise, plain REST is usually easier to implement and will be used as a baseline of what SAP will provide.

Public APIs will be designed for broad and intense reuse. This in turn will lower the barriers for new consumers to interact with a given application and reduce the costs for integrating with that application. All APIs will be versioned to balance the agility of API providers with the stability required by consumers. Public APIs will be carefully documented and will expose metadata in one of our allowed formats such as OData's Entity Data Model (EDM) or SOAP's Web Services Description Language (WSDL). Other REST services are described using Open API with additional support for RESTful API Modeling Language (RAML) by means of translations. With the introduction of SAP API Business Hub, we will have a central entry point for API documentation, where customers and partners will be able to browse through, discover, and try out APIs. In general, with this hub we will be following an open approach to enable more developers to easily integrate with our applications. Additionally, SAP API Business Hub will provide information on API compatibility, versioning, and deprecation.

To facilitate integration use cases (such as master data replication) and to enable sophisticated extensions, SAP will provide advanced API qualities on selected REST and OData APIs. For example, in the future, business and data change events will be exposed to subscribers to enable real-time integration capabilities. OData APIs may also support consistent initial loading of large data sets and delta load capabilities to support offline scenarios and data replication.

To provide a transition path to SAP S/4HANA in the cloud, business APIs (such as the BAPI® programming interface) and intermediate document (IDoc) interfaces will be used in hybrid scenarios to enable integration of legacy on-premise SAP Business Suite applications and SAP S/4HANA with the SAP S/4HANA Cloud suite.

INTEGRATION CONTENT

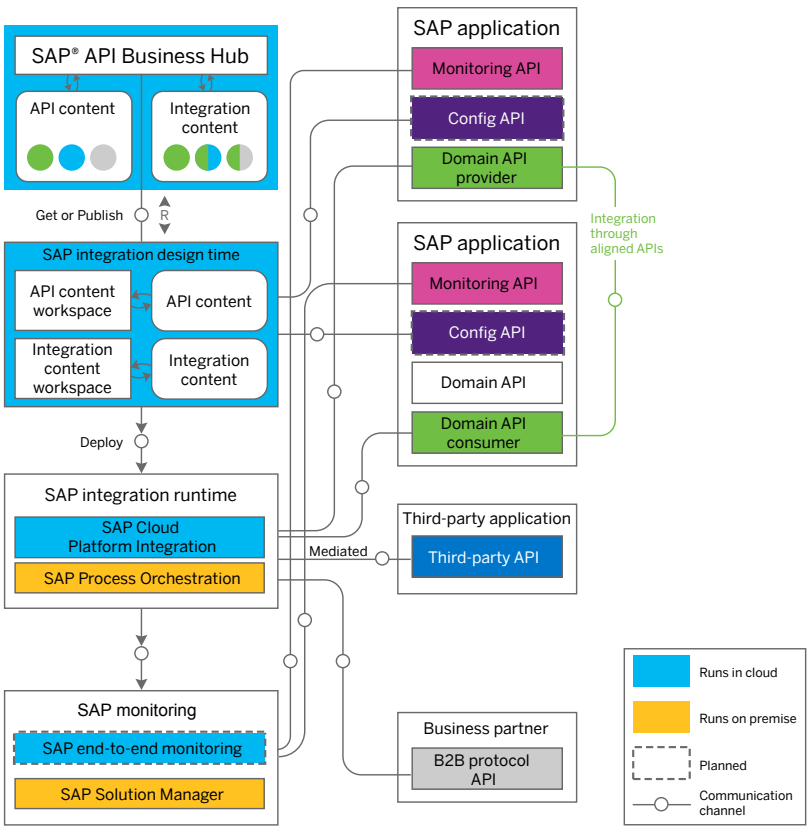
As far as integration content is concerned, SAP will follow a similar approach to transparency as for APIs. Every integration scenario will provide a corresponding business process model that describes the cross-application process, as well as a data flow model that indicates the objects that are being exchanged between applications. Finally, respective integration flows on SAP Cloud Platform will also be provided for individual interactions between applications. For integrations based on aligned APIs (not through a middleware), this will be the route-through content (see also the next section).

Integration content will be published in SAP API Business Hub so that it is accessible to the outside world, making clear which applications integrate as well as showing the involved scenarios. The goal is to have all relevant integration information on one single platform. This will enable customers to navigate through the different perspectives of integration scenarios and will make it easier for them to find a starting point for extending our standard integration content.

FUTURE RUNTIME, DESIGN TIME, AND OPERATIONS ARCHITECTURE

The following high-level architecture model outlines the future SAP process integration architecture with its components and their interactions (see Figure 3). The model attempts to cover not only cloud but also on-premise and third-party applications, as well as other aspects such as design time, runtime, and monitoring.

Figure 3: Future Process Integration Architecture Including Design Time, Runtime, and Operations



On the right side of the figure, you see examples of different applications that participate in integration scenarios. Within the applications, we modeled the different types of APIs that expose access to data and services of the respective applications. In case of an integration through aligned APIs, API consumers will be able to directly consume the services that API-provider applications expose. This integration option is highlighted in the diagram with a green communication channel.

The infrastructure components are depicted on the left side. SAP API Business Hub is where the available API types are being documented. Today, it already acts as a repository for integration content. The API content repository will contain documentation of the green domain API that the first application on the right-hand side exposes. As you can see, the standard integration content resides in the hub.

The green integration content in the storage of SAP API Business Hub indicates that the two SAP applications can be integrated without a middleware, meaning that no translation or mapping is required. This is called "route-through" content because it does not contain any mapping logic and can be used as-is. That is why this content has been colored completely in green (unlike the content that requires mappings, which is depicted with colored halves). Now, if a customer decides to switch to a mediated communication, this content will be loaded from SAP API Business Hub to SAP Cloud Platform Integration design time, where it will serve as a reference template for adaptations. More precisely, the customer copies the integration route-through content from the hub into his own workspace in the design time, configures the integration flow and end points, and finally deploys it from the design time to an SAP integration runtime. As shown in the model, both SAP Cloud Platform Integration and SAP Process Orchestration (version 7.5 and higher) support the format of SAP Cloud Platform Integration design time integration flow content. In this example, once a mediated communication has been established, the green consumer will no longer communicate with the green domain-aligned API. Instead, communication will be realized through the middleware runtime.

As mentioned before, SAP is offering prepackaged integration content as a reference template that allows customers to quickly implement new business scenarios. This reduces integration project lead times and lowers resource consumption significantly. In other words, SAP Cloud Platform Integration design time allows customizing of standard content from the content catalog (SAP API Business Hub). It is the central place and entry point for monitoring, configuration of integration end points, error resolution, and much more.

Since SAP has no control over APIs or consumer capabilities of third-party applications, mediated communication will be the usual method to integrate these applications. Similarly, integrations that use B2B protocols such as EDI or RosettaNet also require middleware.

The design time serves another purpose: in our future process integration architecture, it can be used to centrally configure the individual SAP applications, especially API providers and API consumers. For this purpose, in the future we plan for SAP applications to expose configuration APIs, which are colored purple in Figure 3. A similar concept is already being implemented for monitoring, which is a separate yet highly relevant topic.

On-premise and cloud applications expose monitoring APIs (shown in pink in Figure 3) for SAP Solution Manager to access a central monitoring. A planned end-to-end monitoring on SAP Cloud Platform will use these same APIs. Additionally, the middleware hubs also send monitoring information to the monitoring components. This will allow our customers to obtain a central overview of the entire integrated landscape.

SAP CLOUD PLATFORM INTEGRATION

While SAP API Business Hub is the future go-to place for discovering all integration content, SAP Cloud Platform Integration design time enables customers to actively customize integration scenarios using a Web application. The service supports two integration runtime environments for executing the integration content:

- SAP Cloud Platform Integration runtime (cloud-based)
- SAP Process Orchestration (on premise, as from release 7.5, support package 0)

For new integrations, content packages will always be created to achieve full transparency. We will offer seamless navigation from process models through integration flows down to APIs and message types used in the integration (see also Figure 2).

SAP Cloud Platform Integration enables communication in cloud and hybrid environments and is our preferred solution for dealing with process integration. Since SAP Cloud Platform Integration is offered as a service on demand, it provides the highest level of security features such as content encryption and signing, certificate-based authentication, encrypted data storage, and data isolation at runtime, as well as persistency. Furthermore, the service provides capabilities such as runtime for processing, data transformation, routing of messages, and diverse connectivity options.

An HTML5-based Web user interface is available for users to enable easy browsing and discovering of standard integration content that SAP delivers. As mentioned, SAP Cloud Platform Integration design time also will be used to configure end points that support remote configuration APIs, enabling customers to connect systems without requiring a runtime tenant for the service.

SAP Cloud Platform Integration makes mediated cloud integration reliable. Hence, it is the strategic integration service for customers with SAP applications in the cloud that require mediated communication. It provides ready-to-use connectivity across cloud and on-premise applications. For more information, click **here**.[4]

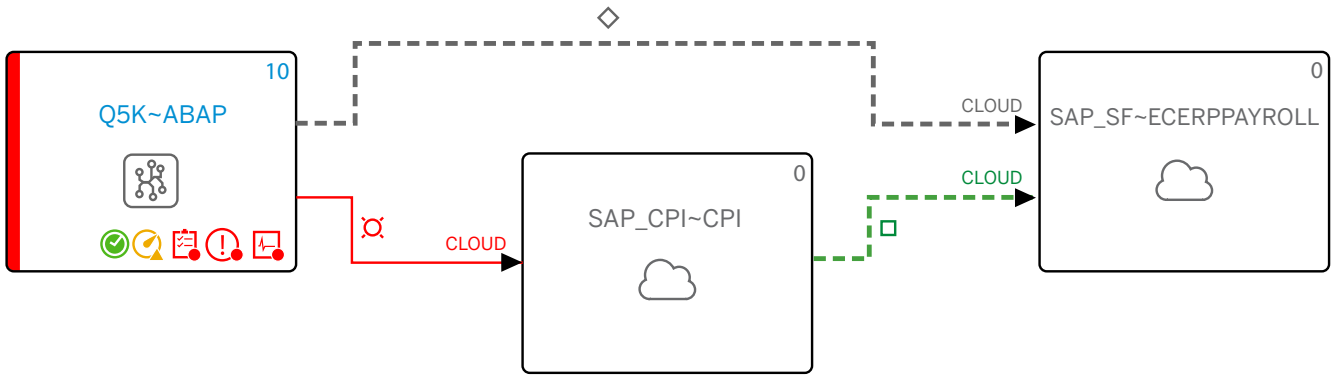
4. www.sap.com/products/hana-cloud-integration.html.

OPERATIONAL ASPECTS

SAP Solution Manager allows centralized monitoring of both on-premise systems and systems in hybrid landscapes. Additionally, SAP plans to offer lightweight cloud-based end-to-end monitoring as a cloud SaaS solution. In the future, all new SAP applications will provide the necessary APIs to monitor on-premise and cloud environments.

In SAP Solution Manager, integration scenarios are also modeled as operation views. Figure 4 shows an example view for a communication scenario between SAP S/4HANA and an SAP SuccessFactors solution through SAP Cloud Platform Integration middleware, beside an integration without middleware (gray path).

Figure 4: SAP® Solution Manager View of an SAP Cloud Platform Integration Scenario



In the future, SAP Cloud Platform, and specifically SAP Cloud Platform Integration, will closely integrate with end-to-end monitoring in the cloud to set up monitoring for integration scenarios based on the standard content that is contained in SAP API Business Hub. As far as operation-related activities are concerned, we are making efforts to reduce TCO by improving automation, monitoring, troubleshooting and support, and scalability.

For on-premise usage, the integration monitoring content for SAP Solution Manager will be continuously enhanced and made available.

For more information on SAP Solution Manager, click here.[5] Additionally, learn more about the SAP Solution Manager community.[6]

5. https://support.sap.com/solution-manager.html.
6. www.sap.com/community/topic/solution-manager.html.

Process Integration

According to the target integration architecture depicted in Figure 3, process integration will be based for both provider and consumer on aligned APIs and route-through content in SAP Cloud Platform Integration design time. As stated in that section, customers are free to choose the integration option that most suits their requirements. They can switch to a mediated integration anytime – for example, if monitoring or special auditing or security requirements have been defined, or if a customer pursues a general middleware-based integration approach.

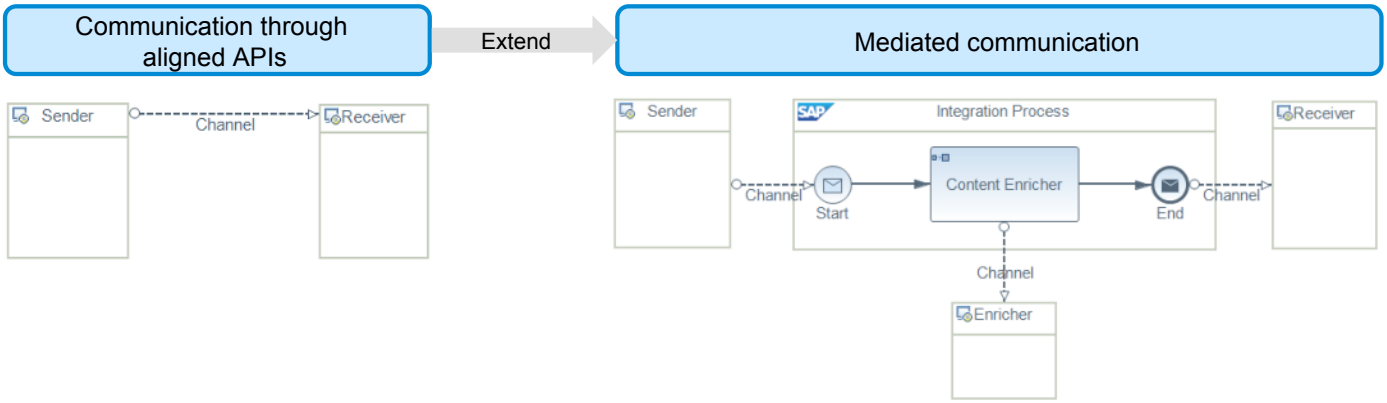
APPLICATION-TO-APPLICATION INTEGRATION

For A2A process integrations, a communication through API calls will be the preferred method for new SAP applications. This method presumes that three conditions are met:

- Aligned data domain model (for example, employee or cost center data entity). See also the "Application Domain Models and Integration" section.
- Aligned technical protocols (HTTPS REST or OData, and SOAP for asynchronous messaging)
- Aligned orchestration logic (for example, pull, or a synchronous request-reply pattern)

Even though communications based on aligned APIs do not need SAP Cloud Platform Integration runtime, we still require that new applications model the corresponding links in the service design time for documentation purposes, as well as to allow customers to extend them as required (see Figure 5). This route-through content is also used for managing end points and for governance purposes. In addition, even if aligned APIs are not available for integrations, SAP will deliver SAP Cloud Platform Integration content (including mapping) to enable ready-to-use integrations.

Figure 5: Extending a Straight Integration Flow with an Integration Process for Mediation



Since this domain model alignment is not likely to be given for integration scenarios involving third-party applications, an integration through aligned APIs will not be feasible in such cases. Therefore, a mediated integration – for example, through SAP Cloud Platform Integration runtime – is indicated. For this purpose, SAP Cloud Platform Integration offers special connectors for easily integrating third-party applications in the context of A2A integrations.

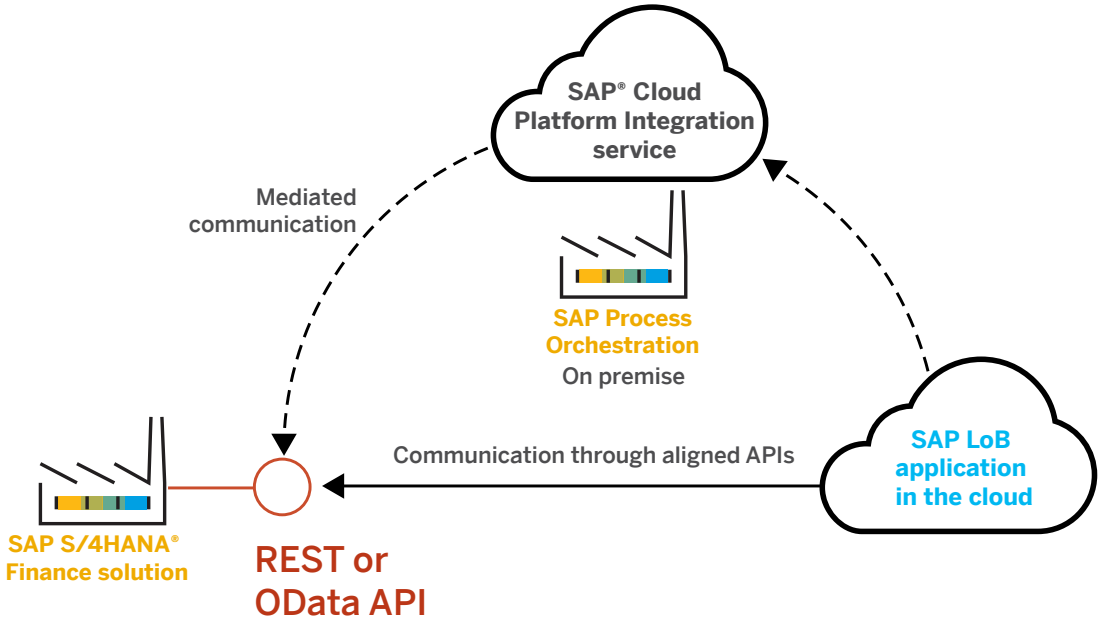
Example: SAP S/4HANA Finance Solution and SAP LoB Applications in the Cloud

Integration between SAP S/4HANA Finance (on premise) and an SAP LoB application in the cloud is a good example of an A2A integration between two SAP applications in a hybrid environment.

SAP S/4HANA Finance exposes either a REST or an OData API for the cost center master data, and the SAP LoB application in the cloud acts as a consumer that directly consumes this service. The preferred integration option for such a hybrid scenario is a communication through a corresponding REST or OData API. Prerequisite for this integration is that the respective external API models are aligned between the consumer and provider applications.

Suppose now that a customer decides to switch to a mediated communication, for example, because she wants to call an additional logging service each time master data is retrieved from SAP S/4HANA into the SAP LoB application in the cloud (see Figure 6).

Figure 6: Integration Options Between SAP S/4HANA Finance and SAP Concur® Solutions



To switch to a mediated communication, customers load the content that is provided on the content hub into their SAP Cloud Platform Integration design-time workspace and customize the integration flow. They sign up for an SAP Cloud Platform Integration runtime tenant and deploy the content into the instance. Alternatively, customers can use SAP Process Orchestration (from release 7.5) for this mediated communication, since it also supports the same integration flow format that SAP Cloud Platform Integration uses.

After reconfiguring the provider end point on the LoB SAP application side, the client will no longer communicate through aligned APIs with SAP S/4HANA but will use mediated communication instead.

This open and transparent approach combines the benefits from integration using our public APIs with the flexibility that our customers expect from our middleware solution.

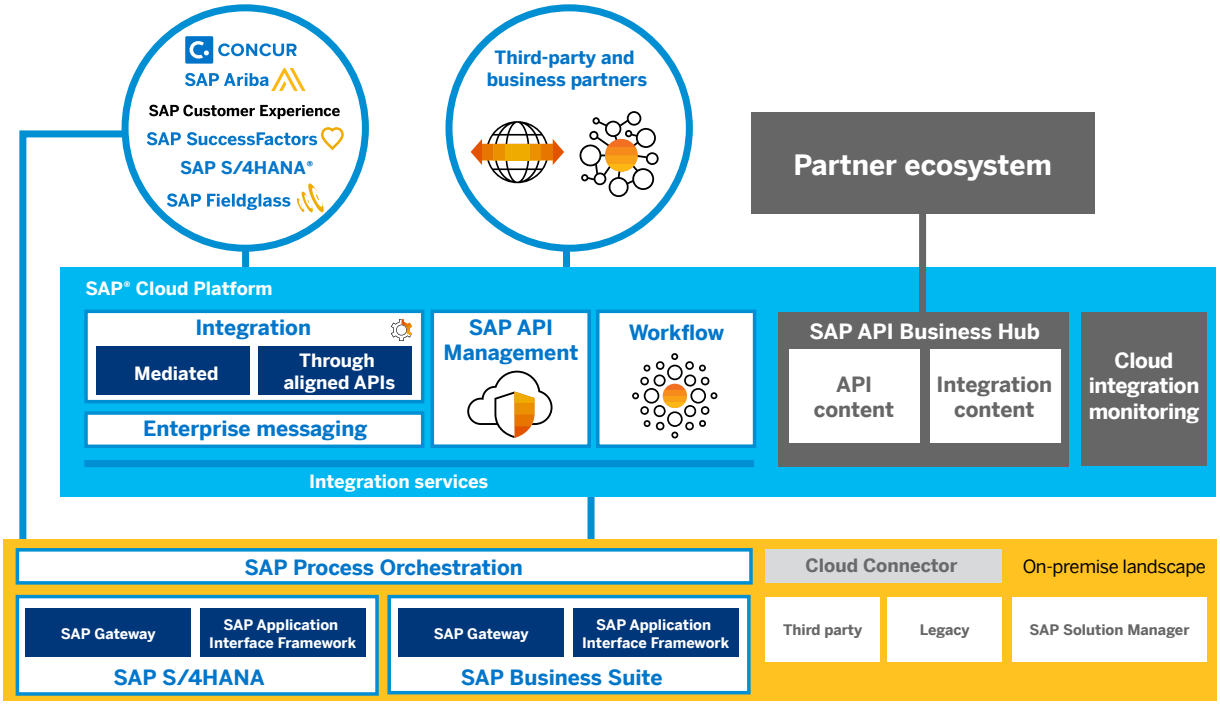
BUSINESS-TO-BUSINESS INTEGRATION

B2B communication is mediated communication. SAP Cloud Platform Integration will be used for all B2B communication scenarios. We advise against using application-specific implementations or third-party solutions to deal with these communication scenarios.

PROCESS INTEGRATION IN HYBRID SYSTEM LANDSCAPES

Figure 7 gives an overview of the SAP Cloud Platform Integration service and shows how it interplays with the related on-premise integration technologies.

Figure 7: Overview of Process Integration in Hybrid System Landscapes



As stated in the section "Future Process Integration Architecture," SAP Cloud Platform Integration is our strategic integration technology. It supports integration between SAP applications, third-party applications, and business partners through mediated communication scenarios. Additionally, it supports integration between SAP applications in the cloud, based on aligned APIs, including route-through content modeled in SAP Cloud Platform Integration service design time. In such a case, the runtime communication will not require middleware, and SAP Cloud Platform Integration will merely manage the end points of the communicating applications. Customers will be able to switch to mediated integration anytime as needed – for example, if custom extensions are required, or if the customer in general prefers a middleware-based integration approach.

In addition to this core integration service, SAP Cloud Platform provides complementary services. The SAP Cloud Platform Enterprise Messaging service, for example, provides high-performance and robust messaging capabilities that SAP Cloud Platform Integration uses. API management allows enrichment of APIs with policies, security, and traffic management capabilities. It enables customers to expose their digital assets, either within or outside their organizations, for example, by providing SAP-managed APIs to business partners (B2B) or consumers (business-to-consumer). More information on the SAP Cloud Platform API Management service can be found **here**.[7]

7. https://hcp.sap.com/capabilities/integration/api-management.html.

SAP API Business Hub is a catalog for standardized content lifecycle management, advertising, and distribution. Customers obtain a harmonized approach to learn about new SAP content (for example, APIs, integration or data flows, and mappings) and consume it in the same way, regardless of the type of content package or artifact. All APIs and integration content are published and consumed from the same hub.

Customers have the choice to deploy integration content based on integration flows for SAP applications in the cloud into either SAP Process Orchestration or SAP Cloud Platform Integration. Many of our customers currently use SAP Process Orchestration to integrate on-premise applications. If you are among them, you may also use SAP Process Orchestration (version 7.5 and higher) for integrations between on-premise and cloud applications from SAP. However, in a landscape where mediated cloud-to-cloud application integration is required, SAP Process Orchestration cannot be used. In contrast, SAP Cloud Platform Integration can be used for both types of integrations: between cloud applications and between cloud and on-premise applications. In general, while you may continue to take advantage of your SAP Process Orchestration investment, you should be aware that you may end up needing two middleware installations if you require additional mediated cloud-to-cloud integrations.

The SAP Application Interface Framework tool is an add-on to SAP Business Suite and an integral part of SAP S/4HANA to use for advanced interface monitoring, data validation, and error handling. It enables key business users to monitor interfaces they are responsible for and to handle business errors (for example, validating and reprocessing failed sales orders). Find more information on SAP Application Interface Framework **here**.[8]

SAP Cloud Platform Integration uses the SAP Cloud Platform Connectivity service to securely connect to on-premise applications.[9] It is the lightweight connection tunnel solution between SAP S/4HANA Cloud, other cloud solutions on SAP Cloud Platform, and existing on-premise systems. It provides fine-grained control over on-premise systems and resources that cloud applications should be able to access. SAP Cloud Platform Connectivity does not include any kind of orchestration or message mapping services. Find more information on SAP Cloud Platform Connectivity **here**.[10]

SAP Solution Manager facilitates centralized monitoring of both on-premise systems and systems in hybrid landscapes (see also the "Operational Aspects" section).

The API content in SAP API Business Hub provides a central catalog that includes APIs from different SAP applications in the cloud. The integration content in the hub complements the API content by providing rich content for prepackaged integration solutions (that is, integration packages that include integration flows).

8. https://help.sap.com/saphelp_aif10/helpdata/en/5a/3eacf824e74542abbd2271238dc70b/frameset.htm.
9. The SAP Cloud Platform Connectivity service includes connectors to connect to on-premise systems, to SAP Cloud Platform, and to interconnected cloud systems.
10. https://help.hana.ondemand.com/help/e6c7616abb5710148cfcf3e75d96d596.html.

Our long-term goal is to evolve SAP Cloud Platform Integration into an integration platform as a service (IPaaS) consisting of a rich set of integration services supporting various integration domains and use cases. One example of such an integration-related service is the SAP Cloud Platform Workflow service, which allows partners and customers to build, run, and manage workflows, from simple approvals to end-to-end processes that span different organizations and applications.

SUMMARY OF TECHNOLOGY RECOMMENDATIONS FOR PROCESS INTEGRATION

Technology recommendations can be derived from the SAP vision for a straightforward and consistent integration approach, shown in the following table.

Technology Recommendations for Process Integration Use Cases

	Integration use case		
Integration technology	Application-to-application integration	Master data synchronization	Business-to-business integration
APIs and SAP® Cloud Platform Integration route-through content[1]	●	●	
SAP Cloud Platform Integration with mappings[2]	●	●	●
Complementary technologies	• SAP API Business Hub • SAP API Management • SAP Cloud Platform Connectivity • SAP Application Interface Framework and SAP Gateway (SAP S/4HANA® and SAP Business Suite) • SAP Cloud Platform Enterprise Messaging • SAP Cloud Platform Identity Authentication		

● General recommendation

● Apps requiring mediated integration

1. This approach is the goal for **SAP-to-SAP** application scenarios. **SAP connectivity** is implemented through aligned APIs published in SAP API Business Hub. Customers may copy the corresponding SAP Cloud Platform Integration route-through content into a service instance and extend or modify it. **SAP Cloud Platform Integration runtime can be used optionally**.
2. General recommendation for third party (non-SAP)

For the SAP-to-SAP integration use cases, where A2A integration (and master data synchronization as part of it) is required, integration based on public (aligned) APIs and SAP Cloud Platform Integration route-through content will be the future method of choice. This is valid for both new on-premise-to-cloud and cloud-to-cloud integrations. Integration between applications through aligned APIs – that is, not requiring middleware – is possible only if the prerequisites regarding data model, orchestration pattern, and technical protocols are met. In this context, it should be stressed that SAP Cloud Platform Integration makes this connectivity both manageable and extensible. Route-through integration flows are provided to document scenarios that connect applications and the APIs they use for this purpose. Based on this content, customers can create their own extensions and apply centralized governance processes. The extended content can be optionally executed in SAP Cloud Platform Integration runtime. In the long run, SAP Cloud Platform Integration management capabilities will be able to automatically switch to a mediated communication, if so required. Customer-specific deviations from a communication solely through aligned APIs could be indicated, for example, when there is a need for integrating additional applications, such as archiving systems, or when value mapping in third-party systems must be looked up for data enrichment. In addition, customers may opt for a middleware hub-based integration approach as a general company decision.

SAP Cloud Platform Integration with mappings will be used if an integration through aligned APIs cannot be applied. For third-party and B2B integration, the integration service with mappings is the indicated option. For B2B purposes, SAP recommends using SAP Cloud Platform Integration as the general approach. Prepackaged integration content (for example, mappings and integration flows) developed by SAP or partners will be centrally published and managed in SAP API Business Hub.

Data Integration and Virtualization

Specifically, data integration refers to the following use cases: ETL (both classic and real time), data replication, and data virtualization. In this section, we will discuss these scenarios and provide technology recommendations to fulfill the requirements associated with each of them.

ETL is a process that consists of extracting data from a source system, transforming that data by applying rules to make it ready to be loaded into the target system, and loading the transformed data into the target system. Data replication is the process of copying data from one system to another in real time or near-real time. Real-time ETL combines the data replication scenario with the transform and load steps from the ETL scenario.

Finally, data virtualization provides real-time access to data (relational or nonrelational) from an external system through a standardized SQL interface, regardless of where the data is stored (for example, in the cloud or on premise), how it is stored (for example, in files, databases, and so on), or how it needs to be accessed (for example, using JDBC, REST, and others). Data virtualization is indicated when end users can reasonably consume the data (that is, when data amounts are manageable for the end user).

COMMON SCENARIOS FOR DATA INTEGRATION

The following scenarios are examples of cases where data integration technology may be indicated:

- Generally, when you want to access data of a remote system using local SQL. For example, say you develop an extension application on SAP Cloud Platform and you want to access data from SAP S/4HANA (on premise or in the cloud, such as on Amazon Web Services [AWS], among others). Data replication or data virtualization can be a good means in this specific case, since both allow for database-level joins of remote and local data with the option of database optimization. In addition, it allows for an "integrated experience" when developing highly data-centric applications.
- In case of a data migration from a legacy on-premise system to the cloud. Typically, in this scenario the legacy on-premise systems do not have adequate APIs. The requirement is to extract data directly from the database in a one-off manner and upload it to the cloud application. When assessing the feasibility of this scenario, data quality must be taken into consideration.
- If data replication from a legacy on-premise system to the cloud is required. Here, you need to transfer data from a legacy on-premise source system without adequate APIs directly from its database tables into a cloud application – for example, for analytical purposes. Data quality also plays an important role here.

AVAILABLE SOLUTIONS FOR HANDLING DATA-CENTRIC INTEGRATION SCENARIOS

Different technology recommendations apply, depending on whether your system landscape is SAP HANA–centric or rather comprises heterogeneous systems. Figure 8 depicts the current technologies used for dealing with data-centric integration and virtualization in the context of different communication parties.

The following sections will explain each of these technologies and provide recommendations on when to use them.

SAP Landscape Transformation Replication Server

SAP Landscape Transformation Replication Server is a trigger-based data replication solution that provides scheduled or real time data replication between solutions based on the SAP NetWeaver® technology platform and SAP NetWeaver–supported databases including SAP HANA.

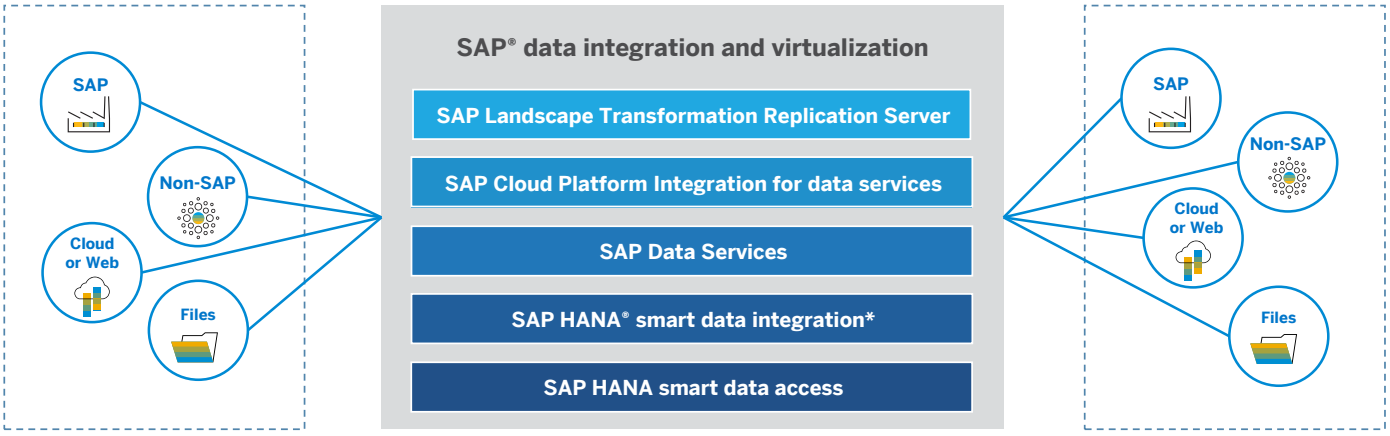
During replication, it provides lightweight transformations capabilities such as filtering and simple data manipulation.

SAP Landscape Transformation Replication Server is recommended for the following data replication scenarios:

- SAP NetWeaver to SAP NetWeaver
- SAP NetWeaver to SAP HANA
- SAP NetWeaver–supported databases to SAP NetWeaver–supported databases[11]

Customers may deploy SAP Landscape Transformation Replication Server on premise or in the cloud. This product can replicate data between on-premise systems or between on-premise and cloud systems. For replications into SAP HANA on SAP Cloud Platform, SAP Cloud Platform Connectivity must be used. For other on-premise-to-cloud integration scenarios, customers may set up any other virtual private network (VPN) tunnel to enable cloud and on-premise solutions to be in the same virtual network.

Figure 8: Available Technologies for Data Integration and Virtualization



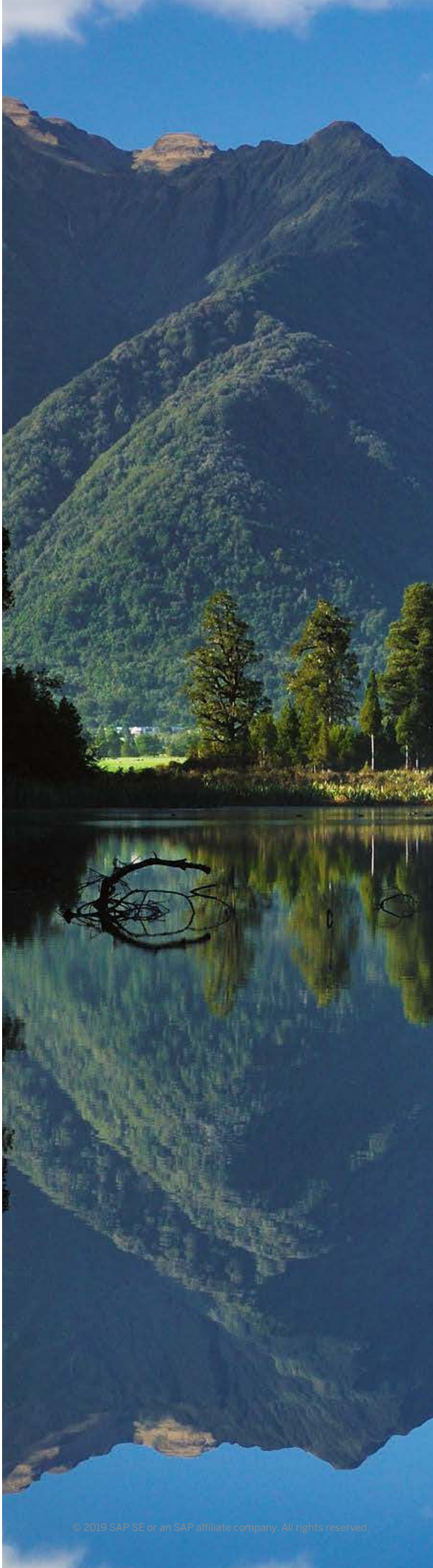
*When deployed on SAP Cloud Platform, this product is called SAP Cloud Platform Smart Data Integration.

11. Note: For customers who want to replicate from an SAP NetWeaver–supported database into SAP HANA but have never used the SAP NetWeaver Application Server component, we recommend using SAP HANA smart data integration instead.

SAP Landscape Transformation Replication Server in combination with SAP Data Services software is recommended for real-time replication from SAP systems when the data needs to be transformed before it is loaded into the target system. It is also recommended for real-time replication from SAP systems when the desired target system is supported by SAP Data Services and not supported by SAP Landscape Transformation Replication Server.

Note that customers who already use SAP Landscape Transformation Replication Server to replicate data into SAP HANA do not need to switch to SAP HANA smart data integration (or SAP Cloud Platform Smart Data Integration, when deployed on SAP Cloud Platform) unless implementation of real-time ETL scenarios is required.

More information on SAP Landscape Transformation Replication Server can be found **here**.[12] You can also learn more about the **SAP Landscape Transformation Replication Server community**.[13]



12. https://help.sap.com/viewer/p/SAP_LANDSCAPE_TRANSFORMATION_REPLICATION_SERVER.
13. www.sap.com/community/topic/lt-replication-server.html.

SAP Cloud Platform Integration for Data Services
SAP Cloud Platform Integration for data services is an ETL solution that can extract data from a variety of different on-premise SAP and non-SAP systems. It can transform that data using a very limited set of transformations and functions, and load it into cloud-based SAP applications (for example, the SAP Integrated Business Planning solution – see **here**[14] for more information). SAP Cloud Platform Integration for data services can also extract data from cloud-based SAP applications (for example, SAP SuccessFactors solutions and other cloud applications) and load it into a variety of different on-premise SAP and non-SAP systems. This ETL solution does not require a VPN (or other network solution) for loading data into the cloud.

The agent for SAP Cloud Platform Integration for data services must be deployed on premise and can be used to move data from on-premise to cloud-based systems and vice versa. It is the recommended solution for on-premise-to-cloud ETL scenarios when the target system is not running on SAP HANA, or when the target system is running on SAP HANA but for whatever reason SAP HANA smart data integration cannot be used. For more details on SAP Cloud Platform Integration for data services, consult the **Product Availability Matrix**.[15]

14. https://help.sap.com/viewer/da797ae2bf6246d58abd417f24915d55/1708/en-US/84c79657c82c6c10e10000000a441470.html.

15. https://support.sap.com/en/release-upgrade-maintenance/product-availability-matrix.html.

SAP Data Services

SAP Data Services is an ETL solution that can extract data from a variety of different SAP and non-SAP systems; transform that data using a rich set of transformation capabilities, functions, or custom-built functions; and load it into a variety of different SAP and non-SAP target systems. SAP Data Services has a rich set of connectors to third-party databases as well as to cloud and Big Data systems. Therefore, when compared to SAP Cloud Platform Integration for data services, SAP Data Services is a complete product with a lot more connectivity and a full set of transformation capabilities.

In addition to ETL scenarios, SAP Data Services also supports interactive data integration scenarios where applications send data to be processed in real time (for example, cleansing of address data). When combined with SAP Landscape Transformation Replication Server, SAP Data Services is also able to deal with real-time ETL scenarios.

SAP Data Services can be deployed on premise or in the cloud (for example, AWS, Microsoft Azure, and Google Cloud Platform) and can be used to move data between on-premise systems or between on-premise and cloud systems.

SAP Data Services is recommended for ETL and interactive integration scenarios in heterogeneous system landscapes.

SAP Landscape Transformation Replication Server in combination with SAP Data Services is recommended for real-time replication from SAP systems when the data needs to be transformed before it is loaded into the target system. It is also recommended for real-time replication from SAP systems when the desired target system is supported by SAP Data Services and not supported by SAP Landscape Transformation Replication Server.

Find more information on SAP Data Services **here**.[16]

SAP HANA Smart Data Integration

SAP HANA smart data integration is a comprehensive data integration solution that supports data replication, ETL, and real-time ETL scenarios from a variety of SAP and non-SAP systems. For data replication, it uses trigger-based or transaction log–based replication, depending on the source system.

SAP HANA smart data integration can be deployed on premise or in the cloud and can replicate data between on-premise systems or between on-premise and cloud-based systems. When deployed on the SAP Cloud Platform, it is known as the SAP Cloud Platform Smart Data Integration service and can replicate data between on-premise systems and SAP HANA running on SAP Cloud Platform.

SAP HANA smart data integration is the recommended solution for replicating data from SAP and non-SAP source systems into SAP HANA.

For data virtualization, SAP HANA smart data integration provides a host of adapters to connect to data sources, supporting JDBC, OData, and REST APIs.

16. www.sap.com/products/data-services.html.

SAP HANA Smart Data Access

SAP HANA smart data access is a data virtualization technology that is used to access data (through an SQL interface) from external data sources, in the same way as from local tables, without having to store that data in SAP HANA. The data in the external data sources may be stored in relational and nonrelational formats. A suitable adapter from a third-party vendor or from SAP HANA smart data integration can provide a relational interface to the stored data.

SAP HANA smart data access can be used when SAP HANA and the target data source can communicate through TCP/IP. In terms of deployment, SAP HANA and the target data source may be on premise or on premise and the cloud. When the target data source is on SAP Cloud Platform and SAP HANA is on premise, one can use an SAP Cloud Platform connector with WebSockets.

For cases where other platform-as-a-service (PaaS) cloud vendors such as AWS and Microsoft Azure are used, VPN can be used as long as the enterprise supports extending its network to include these cloud providers. If there is no VPN support, SAP HANA smart data integration provides adapters that can be used to create the connectivity.

SAP HANA smart data access can be configured to support adapters that run either in process or out of process. SAP HANA smart data access is recommended for running in-process adapters to get data from the following SAP databases (or components) to SAP HANA: SAP Adaptive Server® Enterprise (SAP ASE), SAP HANA, the SAP Vora™ engine, SAP IQ software, and the SAP MaxDB® database.

RECOMMENDATIONS FOR DATA INTEGRATION AND VIRTUALIZATION USE CASES

The following table summarizes the technology recommendations for the use cases related to data-centric integration and virtualization.

Technology Recommendations for Data-Centric Integration Use Cases

	Integration use cases			
Integration technology	Data virtualization	ETL	Data replication	Real-time ETL
SAP HANA® smart data access (with SAP HANA smart data integration adapters)[1]	●			
SAP HANA smart data integration[2]		●	●	●[3]
SAP® Cloud Platform for data services[4]		●		
SAP Data Services[5]		●		
SAP Landscape Transformation Replication Server[6]			●	●[7]

1. Recommendation for SAP databases and Apache Hadoop systems through the SAP Vora™ engine (for more details, see below)
2. Also known as SAP Cloud Platform Smart Data Integration when deployed on SAP Cloud Platform; recommendation for SAP HANA–centric landscapes
3. With SAP Landscape Transformation Replication Server
4. Recommendation for loading into SAP cloud-based applications running on SAP HANA
5. Recommendation for heterogeneous system landscapes
6. Recommendation for SAP systems
7. With SAP Data Services

The integration technology recommendations for data-centric integration use cases mainly depend on whether your system landscape is SAP HANA–centric or comprises heterogeneous systems. It also depends on whether license restrictions for accessing a specific source database exist.

SAP HANA smart data access should be used for data virtualization with any data sources. However, the following should be noted:

- For SAP databases, we use in-process adapters. ODBC adapters for SAP ASE and SAP IQ are available as a downloadable package.
- Access to SAP Vora is managed by an optimized ODBC driver for SAP Vora.
- Access to Apache Hadoop systems running Apache Spark is managed by a special connector called SAP HANA Spark controller.
- For all non-SAP data sources, there are two main options:
  - Use a third-party ODBC driver and configure it to run out of process
  - Handle any non-ODBC sources such as JDBC, OData, or REST-based interfaces with SAP HANA smart data integration adapters

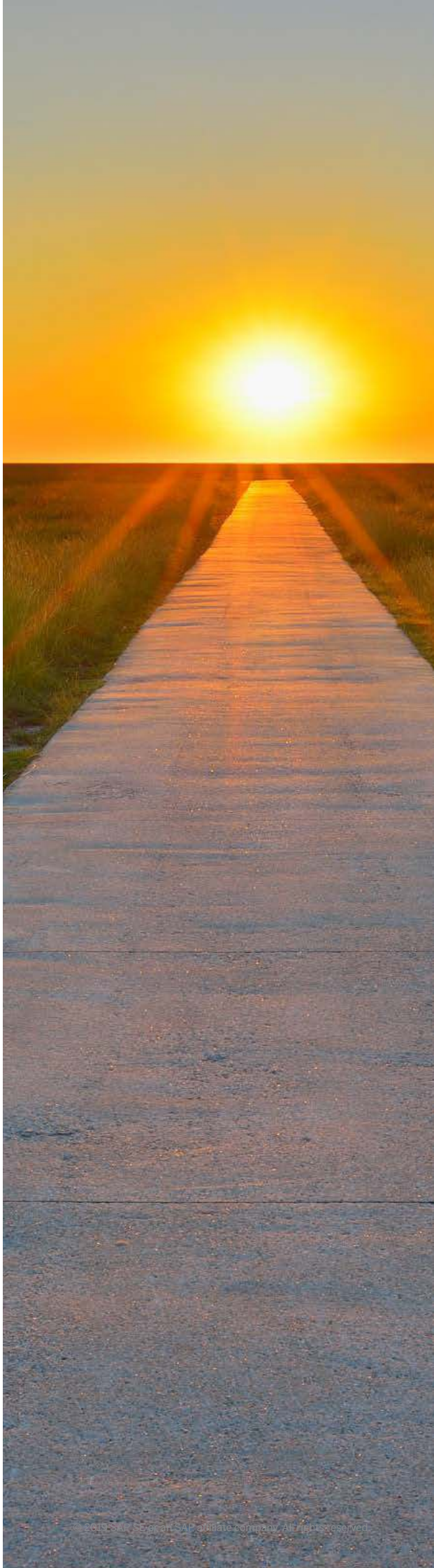
SAP HANA smart data integration is the technology of choice for SAP HANA–centric landscapes. SAP HANA smart data integration handles all data integration scenarios. It should be used for ETL and real-time ETL in SAP HANA–centric system landscapes, including SAP Cloud Platform Integration (in which case it is known as SAP Cloud Platform Smart Data Integration).

SAP Data Services should be used for ETL, while **SAP Landscape Transformation Replication Server** should be used for data replication. When combined, SAP Data Services and SAP Landscape Transformation Replication Server can also deal with real-time ETL. These solutions should be used in heterogeneous system landscapes that are not SAP HANA–centric.

SAP Cloud Platform for use with data services should be used for ETL from on-premise sources into SAP applications in the cloud running on SAP HANA (for example, SAP Integrated Business Planning) and cloud applications (for example, SAP SuccessFactors solutions).

SAP realized the necessity to integrate Big Data and enterprise data from systems deployed on premise and in the cloud. This requires a unified approach to modeling, orchestration, governance, and monitoring, all while ensuring that our customers' current investments are protected. SAP is working on these topics with the purpose of reducing the number of future technology recommendations for ETL (classic and real time), data replication, and data virtualization.

As a result of the efforts associated with the integration of Big Data and enterprise data, and to account for the fact that companies need to move from a world of centralized data to centralized data governance, SAP developed SAP Data Hub, which will be discussed in the next section.



SAP Data Hub: Data Landscape Management Layer

SAP Data Hub provides a comprehensive answer to an emerging and painful challenge for enterprise customers: integrating data and establishing data-driven processes across an increasingly diverse data landscape. The solution addresses data integration, data orchestration, and data governance capabilities across a complex landscape, as well as harnesses Big Data processing to create uniquely powerful data pipelines that are based on the serverless computing paradigm,[17] which allows radical scalability. Data and processes can be managed, shared, and distributed across the enterprise with seamless, unified, and enterprise-ready monitoring and landscape management capabilities. In doing so, SAP Data Hub functions as a catalyst for business transformation by addressing the following issues.

Data is kept in silos (files, Hadoop, data warehouses, data marts, cloud object storages) across and outside the enterprise. Typically, users cannot properly access the data and cannot work with it as they would require. Particularly, it is difficult, complex, time-consuming, and costly to connect Big Data into enterprise data processes.

Data technologies lack enterprise readiness, with no centralized governance for data, incompatible tool lifecycles, little automation in the landscape, and fragmented monitoring and tracing capabilities of individual technologies. No common security and access management is in place. Moreover, there is no transparency with regard to who changed the data, what has changed, or who is accessing it.

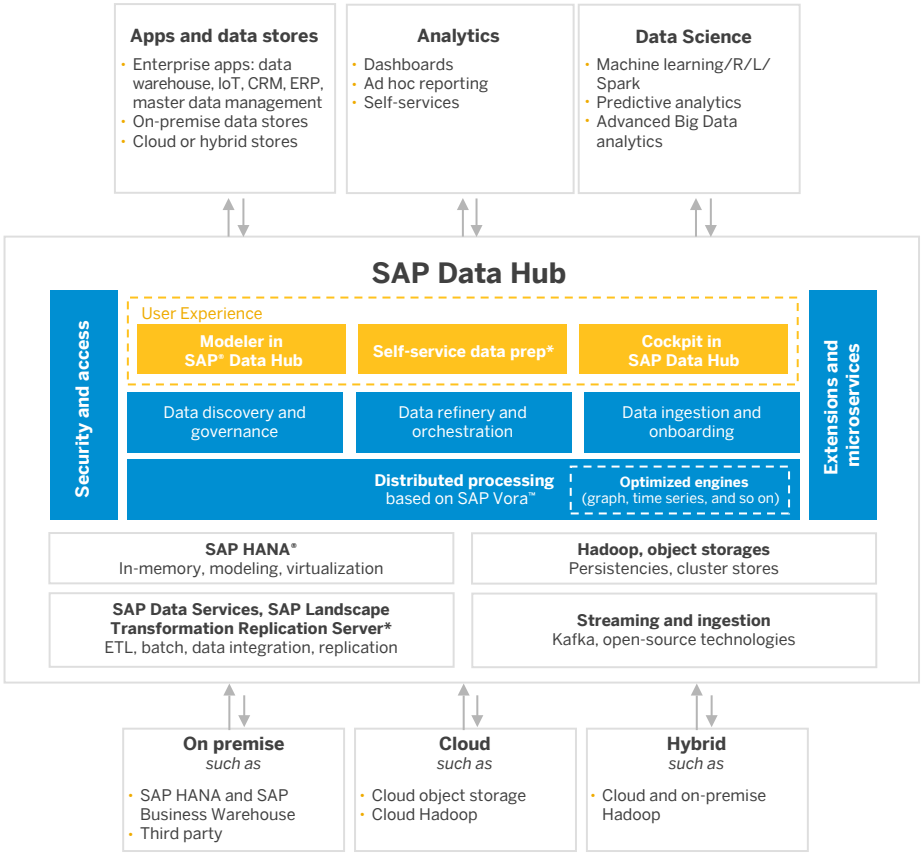
High effort is required to productize scenarios across the enterprise, leading to high investment in resources. Many nonintegrated (open-source) technologies coexist, such as those related to machine learning, Hadoop, Spark, Kafka, Cassandra, and others, which are needed to address pressing business needs. The great effort to integrate different technologies and systems prevents companies from creating business value. It is currently very hard to refine and enrich data across multiple systems so that it is usable by business users and data scientists.

Big Data initiatives typically require a specialized skill set to implement, scale, and create value out of the large data quantities.

SAP Data Hub can be regarded as a technology layer for unifying data sharing and pipelining (see Figure 9). It works for distributed environments and distributed execution of data logic, independent of the data integration technology that is used. Therefore, the data integration technology recommendations given in the previous sections continue to hold valid. SAP Data Hub leverages these technologies and acts as a data landscape management layer. These include SAP HANA smart data integration and SAP Data Services. In the near future, SAP Landscape Transformation Replication Server will also be integrated.

17. "Serverless" denotes a cloud execution model in which the cloud provider dynamically allocates machine resources. Of course, such an infrastructure does require servers, but consumers do not have to care about them.

Figure 9: SAP Data Hub As a Technology Layer Unifying Data Sharing and Pipelining



*Planned for future releases

SAP Data Hub offers a unified tool set for the design of data processing logic and landscape administration. The centralized governance allows a unified view, easily tracing how data has been used and by whom. It also enables you to assess the impact of future changes.

From the user's perspective (user experience), SAP Data Hub provides one overall landscape view on the usually heterogeneous landscape, from raw file storage (for example, Amazon S3), through Hadoop to SAP HANA or the SAP Business Warehouse (SAP BW) application. The cockpit for SAP Data Hub allows setup and provision of different zones, as well as the governance and lifecycle management between them. For example, production is under the control of the IT department, while data scientists work in laboratory environments. On the operations side, SAP Data Hub offers features that allow monitoring and analysis, connectivity management, data access and security, and data quality profiling.

"Data discovery and governance" offers a metadata-rich information catalog on top of the landscape with linage functions between the systems and data assets, including profiling and data preview on these data sets.

"Data refinery" is about transforming raw data into meaningful information for business users and data scientists. Today, you can obtain some of this functionality with open-source or third-party tools, but it is very point to point, expensive, and time-consuming. In contrast, with SAP Data Hub you can define complex data pipelines consisting of delivered operations for code execution and libraries. Moreover, you can create data flows consisting of complex data transformations (such as data quality functions or data masking) on your Big Data stores and execute these flows where the data resides, without having to extract the data first.

On the orchestration side, SAP Data Hub enables you to schedule data flows and complex pipelines across the landscape. You can also include tasks from "external" systems such as SAP Data Services, SAP BW, or SAP HANA smart data integration in one task workflow. Along with this, you obtain a holistic monitoring in a state-of-the-art user experience based on paradigms in SAP Fiori.

"Data ingestion and onboarding" provides predefined operators to ingest data from data sources such as Hadoop Distributed File System (HDFS), Amazon S3, SAP HANA, and others. It also has operators for streaming data such as Kafka and Message Queuing Telemetry Transport (MQTT).

SAP Data Hub is open because it takes advantage of open-source Big Data technologies, such as Python, Scala, Kafka, Hadoop, and Spark, and productizes them into end-to-end scenarios. The solution is also open on the enterprise side, allowing data to be shared no matter the source, through a tight integration with SAP Data Services. In other words, SAP Data Hub is not limited to connecting only SAP products: it explicitly focuses on the integration and interaction with open-source components (for example, Hadoop data lakes, Amazon S3, Kafka, and so on), as well as third-party databases and applications.

"Distributed processing" is based on SAP Vora, including optimized engines for handling graphs, time series, and other unstructured data. It processes data across data lakes (based on Hadoop), object stores, cloud and on-premise databases, and data warehouses. You can perform data transformations through data pipelines, ensure data quality, and manage data preparation processes through a graphical user interface.

The chosen architecture allows developing extensions and building microservices. The initial release of SAP Data Hub is offered as an on-premise or hosted application, whereby its runtime can be deployed in cloud environments. Therefore, its architecture is cloud ready by default. In future releases, PaaS and SaaS versions will follow.

Simplified Architecture View

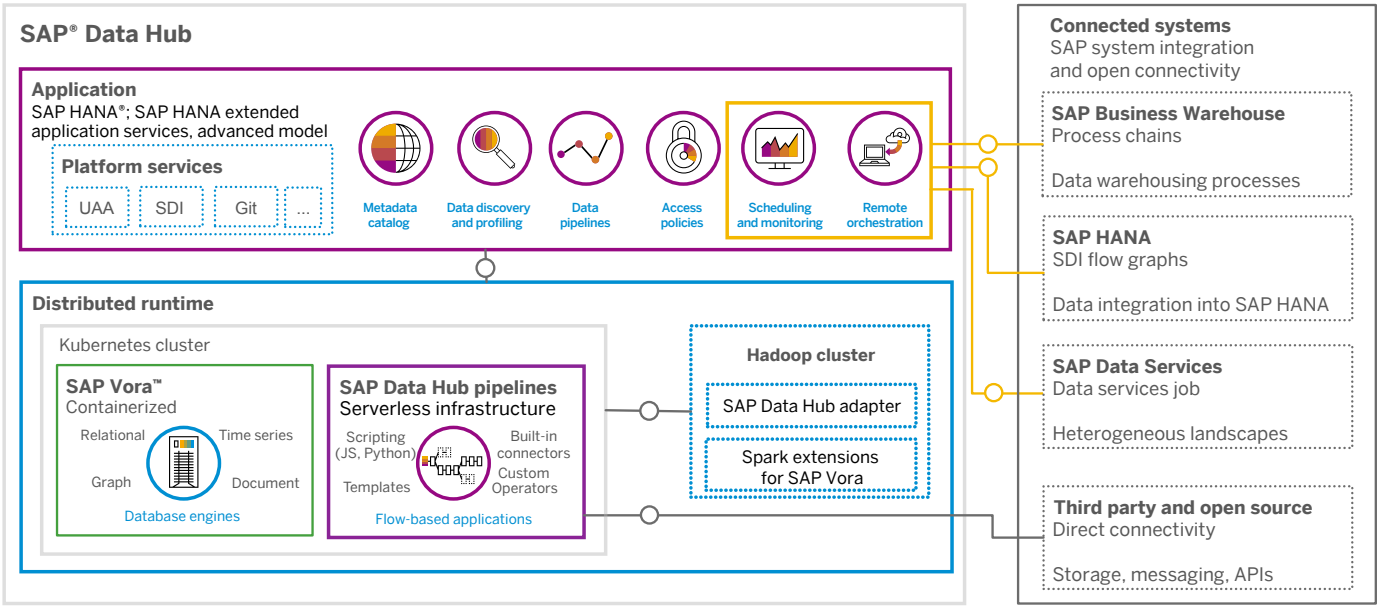
SAP Data Hub consists of two parts:

- An on-premise or hosted SAP HANA–based application that uses SAP HANA extended application services, advanced model[18]
- A runtime, which includes a Kubernetes and a Hadoop cluster

Figure 10 depicts the high-level architecture view of SAP Data Hub. The application can use platform components and services such as User Account and Authentication (UAA), SAP HANA smart data integration, and the version control system Git (Hub), among others. The metadata catalog is stored in SAP HANA. Data discovery and profiling can be carried out on all connected systems such as SAP Vora tables, as well as on the HDFS table in the Hadoop cluster. Data pipelines are designed and executed in the Kubernetes cluster. The link to SAP BW, SAP HANA, and SAP Data Services is only for remote orchestration and monitoring.

In Figure 10, you can observe two distributed runtimes: a Kubernetes cluster and a Hadoop cluster. The first consists of an embedded SAP Vora engine for data processing and an SAP Data Hub pipelines "serverless" infrastructure. The latter is a Hadoop cluster that is connected through an SAP Data Hub adapter and SAP Vora extensions for Spark. The SAP Data Hub pipelines within the Kubernetes cluster can access data from third-party and open-source data sources (for example, HDFS, Amazon S3, Kafka, and others).

Figure 10: Simplified Architecture View of SAP Data Hub



UAA = User Account and Authentication
SDI = Smart data integration

18. In the future, the application will also be available on SAP Cloud Platform.

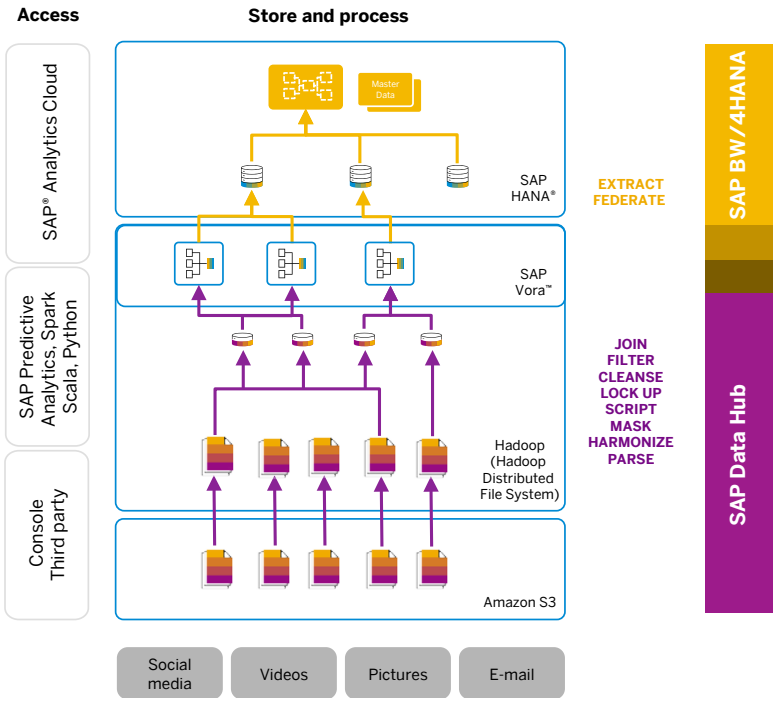
SAP Data Hub is not intended to store data, nor is it about moving data to a central location or "hub" to be processed there. It is a solution to orchestrate, monitor, and manage data between existing data stores. It allows data processing where the data resides.

Architecturally, the solution is based on a microservice design using public REST APIs for flexible consumption. It is prepared for multitenancy and cloud operations on SAP Cloud Platform in Cloud Foundry. Find more information on SAP Data Hub here.[19]

Example Scenario: Analyzing Web Log Data with a Data Warehouse

SAP Data Hub not only accelerates the creation and management of data pipelines that span various data sources, but it also provides fast execution of the pipeline activities themselves by distributing computational tasks to the native environments where the data resides. This federated "push-down" distributed processing helps ensure that the activities of the pipeline complete as rapidly as possible, delivering faster results to the business. In the example scenario outlined in Figure 11, the main tasks are to combine refined Big Data with enterprise data and corporate master data. To complete these tasks, SAP Data Hub manages data ingestion into Amazon S3, for example, as a landing zone for data coming from numerous sources, such as Google AdWords, YouTube, Twitter, and many more. SAP Data Hub orchestrates and schedules all related processes and operations on large data volumes, including joining, filtering, cleansing, and others. With SAP Data Hub, you can implement transformations and data pipelines. It also harmonizes data structures and looks up reference data. Through SAP Vora, data is extracted or federated into SAP BW/4HANA, where it is analyzed to get business-relevant insights.

Figure 11: SAP Data Hub – Analyzing Web Log Data with a Data Warehouse



19. www.sap.com/products/data-hub.html.

Relations Between SAP Data Hub and Other Technologies or Concepts

SAP Data Hub and SAP Cloud Platform Integration

SAP Data Hub deals with complex data processing involving multiple processes across heterogeneous landscapes and where Big Data is involved. This means that the focus is clearly on data orchestration. SAP Cloud Platform Integration, on the other hand, focuses on business process orchestration. Ultimately, data integration and process integration are two sides of the same medal: in many instances, they can be pragmatically combined. For example, after processing sensor data in a data lake, the resulting insights could be used as input for an ordering process. Therefore, the respective development teams at SAP are working closely together to achieve a tight integration (and reuse) of integration flows in SAP Data Hub, such as triggering integration flows through SAP Data Hub workflows. The idea is integration and reuse – not substitution.

SAP Data Hub and SAP Vora

SAP Vora capabilities are included in SAP Data Hub, but SAP Data Hub and SAP Vora target different use cases based on the specific customer need. SAP Vora is an enterprise-ready, intuitive, in-memory distributed computing engine to help organizations uncover actionable insights from Big Data typically stored in Hadoop and NoSQL solutions. It is positioned for both data scientists and as a part of a multitier data strategy with Hadoop.

SAP Data Hub simplifies the orchestration of complex data processes while providing governance across modern and diverse landscapes including Big Data stores, enterprise data stores (for example, SAP HANA, enterprise data warehouses, data marts, and databases), enterprise applications, and cloud solutions.

SAP Data Hub, SAP Data Services, SAP HANA Smart Data Integration, and SAP HANA Smart Data Quality

SAP Data Hub will leverage existing customer investments and execute SAP HANA smart data integration (and SAP HANA smart data quality) flow graphs running on SAP HANA boxes, as well as leverage SAP Data Services jobs that run on existing SAP Data Services job servers. It will not replace these technologies: SAP Data Hub will evolve into a central place to orchestrate, monitor, and later model all integration flows, where SAP Data Services jobs, SAP HANA smart data integration tasks, Big Data, and SAP Vora flows can be brought together.

SAP Data Hub and Data in Hub

SAP Data Hub is a product that enables data processing and management across complex landscapes (enterprise and Big Data) by integrating it into business applications.

"Data in hub" was an architectural concept mainly for data migration and integration from, for example, legacy or mainframe systems into SAP HANA. This concept and SAP Data Hub are not related and address two different use cases.

Find Out More

The materials listed here contain valuable information on SAP integration technologies and enabling services.

- SAP Cloud Platform documentation: https://help.sap.com/viewer/product/CP/Cloud/en-US
- SAP Cloud Platform Integration: www.sap.com/products/hana-cloud-integration.html
- Integration services of SAP Cloud Platform: https://hcp.sap.com/capabilities/integration.html
- J. Mutumba Bilay, P. Gutsche, and V. Stiehl, "SAP HANA Cloud Integration," 420 pages, 2016. You can order the book at www.sap-press.com/sap-hana-cloud-integration_3979.
- Data access and virtualization capabilities of SAP HANA: www.sap.com/product/technology-platform/hana/features/data-access.html
- SAP Data Hub: www.sap.com/products/data-hub.html
- SAP PowerDesigner®: Enterprise architecture tools for digital transformation success: www.sap.com/products/powerdesigner-data-modeling-tools.html
- Modeling guide for SAP HANA smart data integration and SAP HANA smart data quality: http://help.sap.com/saphelp_hana_options_eim/helpdata/en/b7/2a6833d8d54aa2be4c199ac4db6996/frameset.htm
- SAP Cloud Platform API Management: https://hcp.sap.com/capabilities/integration/api-management.html
- SAP Cloud Platform Connectivity: https://help.hana.ondemand.com/help/e6c7616abb5710148cfcf3e75d96d596.html
- SAP Application Interface Framework: http://help.sap.com/saphelp_aif10/helpdata/en/5a/3eacf824e74542abbd2271238dc70b/content.htm
- SAP Solution Manager: https://support.sap.com/solution-manager.html
- SAP Solution Manager community: www.sap.com/community/topic/solution-manager.html
- Integration solution advisor methodology (ISA-M): https://blogs.sap.com/2016/03/04/int203-integration-solution-advisor-methodology-isa-m-sap-teched-lecture-of-the-week

List of Abbreviations

A2A:	Application to application
API:	Application programming interface
AS2:	Applicability Statement 2
B2B:	Business to business
B2G:	Business to government
CIO:	Chief information officer
EDI:	Electronic data interchange
EDM:	Entity Data Model
ETL:	Extract, transform, load (applied to data)
HDFS:	Hadoop Distributed File System
HTML:	Hypertext Markup Language
HTTP:	Hypertext Transfer Protocol
IDoc:	Intermediate document
IPaaS:	Integration platform as a service
ISA-M:	Integration solution advisor methodology
IoT:	Internet of Things
JDBC:	Java Database Connectivity
L:	Low-level programming language of the SAP HANA database
LoB:	Line of business
MQTT:	Message Queuing Telemetry Transport
OData:	Open Data Protocol
ODBC:	Open Database Connectivity
PaaS:	Platform as a service
R:	Open-source programming language and software environment for statistical computing, data analysis, and graphics
RAML:	RESTful API Modeling Language
REST:	Representational state transfer
RFC:	Remote function call
SaaS:	Software as a service
SQL:	Structured Query Language
UAA:	User Account and Authentication
VPN:	Virtual private network
WSDL:	Web Services Description Language

Follow us

www.sap.com/contactsap

Studio SAP | 50200enUS (19/04)

THE BEST RUN SAP®