SAP DATA Intelligence Notes

1. **What is SAP Data Intelligence**

SAP Data Intelligence is a comprehensive data management solution that connects, discovers, enriches, and orchestrates disjointed data assets into actionable business insights at enterprise scale.

It enables the creation of data warehouses from heterogeneous enterprise data, simplifies the management of IoT data streams, and facilitates scalable machine learning. SAP Data Intelligence allows you to leverage your business applications to become an intelligent enterprise and provides a holistic, unified way to manage, integrate, and process all of your enterprise data.

With SAP Data Intelligence you can:

Discover and connect to any data, anywhere, anytime from a single enterprise data fabric

Transform and augment data across complex data types and curate a robust, searchable data catalog

Implement intelligent data processes by orchestrating complex data flows enriched with scalable, repeatable, production-grade machine learning pipelines.

**Why SAP Data Intelligence**

With SAP Data Intelligence, you can:

* Integrate any data, solving the data deluge problem
* Reuse any engine and orchestrate any SAP or third-party data processing engine
* Democratize intelligence by making machine learning easier to implement, scale, operationalize, and govern
* Ensure data quality by discovering, preparing, and governing data assets in the same tool
* Deploy on any mix of hyperscalers, whether hybrid or on premise
* AI/ML Capabilities
* Data Governance and Data Security

Q: What is Hyperscalers mean

companies now being called “hyperscalers” are those massive companies like Google, Facebook, and Amazon that are making efforts to not only dominate the public cloud and cloud services industries, but to expand their business into numerous related verticals, as well.

Q: What is Data Governance

Data governance helps ensures that data are reliable, valid, complete, timely, available to those with a legitimate need for and authority to access. Coordinated data governance also provides the opportunity to decrease data collection redundancies, standardize data-related processes and systems, and increase data system and

How SAP DI helps in Data Integration and Data Orchestration

* Connect and integrate everything: streaming, structured, unstructured
* Orchestrate across any cloud and on-prem deployment
* Leverage processing engines from SAP (for example, SAP HANA Machine Learning) and open source or 3rd party (like Python, R, NodeJS, TensorFlow, Spark etc.)

Q: what is Data Orchestration

Orchestration means Automating the tasks. For eg:

Cloud Orchestration is the process of automating the tasks needed to manage connections and operations of workloads on private and public clouds. For instance: Requesting an IP- you type an internet address or you do internet banking. From login to money transfer - all process has been automated. You have seen that whenever you try to refresh the page it won’t allow you as all process been automated.

Similarly new terms introduced in the market called Cloud Orchestration.

**Product Overview**

SAP Data Intelligence, trial edition is a pre-configured appliance for evaluating and testing SAP Data Intelligence, the successor to SAP Data Hub.

In addition to all the data sharing, pipelining and orchestration capabilities that were originally featured in SAP Data Hub, SAP Data Intelligence introduces a way to manage machine learning models with built-in tools for data governance, management, and transparency:

• Data Pipelines are data-driven applications consisting of reusable and configurable operators.

• Data Workflows orchestrate processes across a data landscape.

• Data Governance allows you to extract metadata from connected data stores. It supports discovery,

profiling and search.

• Automated Machine Learning (AutoML) lets you automate the creation of complex machine learning pipelines with intuitive tools for configuring machine learning artifacts.

**What is SAP Cloud Appliance Library**

SAP Cloud Appliance Library offers a quick and easy way to consume the latest SAP solutions in the cloud, such as SAP S/4HANA, SAP HANA Express Edition, Model Company Solutions, Industry Solutions etc.

It's an online library of latest, preconfigured, ready-to-use SAP solutions that can be instantly deployed into your own public cloud accounts (e.g. Amazon Web Services, Microsoft Azure and Google Cloud Platform) to kick-start your SAP projects - within few hours!

Architecture Overview

The SAP Data Intelligence, trial edition currently is available on Amazon Web Services, Microsoft Azure as well as Google Cloud Platform. While creating a solution instance of SAP Data Intelligence, trial edition, the SAP Cloud

Appliance library creates:

1. Jump Box to orchestrate the instances creation

2. Kubernetes cluster to run SAP Data Intelligence

The Jump Box is used during the creation of the solution instance to run the installer for SAP Data Intelligence. As a user of the SAP Data Intelligence, trial edition you typically do not need to access the Jump Box except for troubleshooting (in rare cases). The Jump Box is automatically terminated 4 hours after the installation has finished successfully (afterwards it cannot be accessed anymore).

SAP Data Intelligence leverages container technology, in particular Docker and Kubernetes, that is offered by the cloud hyperscaler vendor. It allows to query large amounts of data (in distributed storages like Amazon S3, Azure Data Lake v1 and v2 (ADL), Azure Storage Blobs (WASB), HDFS and Google Cloud Storage (GCS)) and to run highly scalable data flows and data-driven applications. As part of this pre-configured appliance, SAP Data Intelligence uses a Kubernetes cluster with three worker nodes.

SAP DI

Connection Management

Metadata Connections: Two folders External and Shared - External with read only access and shared one having Read and write access.

Go to connections --> select drill down as DI\_DATA\_LAKE connection.

This data lake connection is maintained by SAP - which means is Semantic Data lake which is maintained by SAP.

SAP Metadata Explorer

Terms and Tags in Profiling

Terms provides the definition/concepts of data and tags provides the classification/category of data

Rules provides