



GLOSSARY OF SAP DATA INTELLIGENCE

**A Quick Reference Guide to
SAP Terminologies**

Concept/Term	Definition
SAP Data Intelligence	An end-to-end data management solution by SAP to connect, discover, enrich, and orchestrate data. It integrates both SAP and third-party sources across hybrid landscapes.
Pipeline Modeler	A visual tool in SAP DI for designing data pipelines. Enables drag-and-drop configuration for data workflows.
Graph	A sequence of interconnected operators in SAP DI. Defines a complete data processing pipeline.
Operator	A building block of a graph that performs a specific task. Can be configured with inputs, outputs, and parameters.
Metadata Explorer	A central tool to explore, profile, and manage metadata in SAP DI. Helps track data lineage and quality.

Concept/Term	Definition
Data Lake	A storage repository that holds vast amounts of raw data. SAP DI can ingest and process data from lakes.
SAP HANA	An in-memory database used as a target or source in SAP DI. Supports high-speed analytics and processing.
SAP BTP	SAP Business Technology Platform, the foundation for SAP DI. Offers integration, extension, and data-to-value capabilities.
Connection Management	Tool to establish secure connectivity to data sources. Supports SAP and third-party data endpoints.
ML Scenario Manager	Manages the lifecycle of machine learning models in SAP DI. Enables versioning, deployment, and monitoring.

Concept/Term	Definition
Python Operator	Custom operator in SAP DI for executing Python scripts. Useful for complex logic and custom transformations.
Jupyter Notebook	Embedded tool within SAP DI for data exploration and scripting. Supports Python and R for data science tasks.
R Operator	Executes R scripts within SAP DI graphs. Useful for statistical analysis and visualization.
Docker	Used to package and deploy SAP DI operators in isolated containers. Ensures consistency and scalability.
Data Orchestration	Coordinating movement and transformation of data across sources. SAP DI handles complex orchestration workflows.

Concept/Term	Definition
Data Governance	Policies and processes to manage data quality, lineage, and compliance. Enforced via metadata and catalogs in SAP DI.
ML Operator	Operator in a graph for running machine learning models. Supports TensorFlow, Python, and ONNX formats.
SAP DI Modeler	A graphical interface for designing and running data pipelines. Streamlines pipeline creation and testing.
Data Provisioning	The process of making data available for use in SAP DI. Supports batch, real-time, and streaming provisioning.
Remote Source	Any non-local system that provides data to SAP DI. Can be SAP or third-party platforms.

Concept/Term	Definition
Data Wrangling	The act of cleaning, transforming, and preparing data. SAP DI provides visual tools and scripting for this.
SAP VORA	A distributed computing engine integrated with SAP DI. Used for big data processing over Hadoop or Spark.
HDFS	Hadoop Distributed File System, often used with SAP DI for storing large datasets. Common in big data scenarios.
ABAP Operator	Executes ABAP logic within SAP DI pipelines. Enables reuse of SAP ECC or S/4HANA business logic.
SAP BW/4HANA	A data warehousing solution connected via SAP DI. Enables data extraction and transformation from BW objects.

Concept/Term	Definition
Kafka Operator	Connects SAP DI with Apache Kafka streams. Used for real-time data ingestion and event processing.
Delta Lake	An open-source storage layer supported by SAP DI. Provides ACID transactions for data lakes.
OpenAPI	A specification for APIs, used in SAP DI to connect to external services. Supports REST-based interactions.
Data Catalog	Central repository for metadata and data assets. Helps in discovery, classification, and governance.
Data Lineage	Tracks the journey of data across systems. SAP DI provides visual lineage to ensure transparency.

Concept/Term	Definition
Tenant	A logical unit within SAP DI for isolating environments. Typically used for dev, test, and production segregation.
Machine Learning	Embedded in SAP DI to build and deploy predictive models. Integrated with pipelines for automation.
Inference Operator	Executes predictions using pre-trained models. Deployed as part of real-time or batch pipelines.
Version Control	Tracks changes to pipeline and operator definitions. Ensures traceability and rollback.
Data Integration	The core function of SAP DI to bring together disparate data sources. Supports both structured and unstructured data.

Concept/Term	Definition
Semantic Layer	Provides a business-friendly abstraction over technical data. Helps in consistent data interpretation.
Scheduling	Automates pipeline execution at defined intervals. Supports time-based or event-driven triggers.
Monitoring	Tracks performance and execution of graphs and jobs. Alerts for failures or anomalies in real-time.
Error Handling	Built-in mechanisms to manage failures in pipelines. Supports retries, alerts, and compensating actions.
Data Profiling	Evaluates the quality and structure of datasets. SAP DI provides visual and statistical profiling tools.

Concept/Term	Definition
SAP Cloud Connector	Securely connects SAP on-prem systems with SAP DI on BTP. Bridges cloud and on-premise securely.
Custom Operator	User-defined logic packaged into a reusable unit. Built using Python, R, or shell scripts.
Tenant Administrator	Manages users, roles, and settings within an SAP DI tenant. Ensures access and policy compliance.
Policy Management	Defines rules for data access, usage, and quality. Helps enforce governance across data assets.
External Metadata	Metadata imported from external systems into SAP DI. Helps unify visibility across the landscape.

Concept/Term	Definition
Object Store	A storage system for unstructured data used in SAP DI. Examples include AWS S3, Azure Blob.
Hybrid Landscape	Combination of on-premise and cloud environments. SAP DI is optimized to work across both.
Data Quality Rules	Checks applied to ensure correctness and completeness of data. Configured within the Metadata Explorer.
Data Tagging	Assigns labels to data assets for classification. Helps in search, filtering, and governance.
Pipeline Execution	The runtime process where SAP DI executes data graphs. Tracks progress, status, and outputs.