GXP DATA LAKE

INNOVATING PHARMA IN GXP DOMAINS WITH BIG DATA

WHAT IS IT?

The NTT DATA GxP Data Lake provides a highly scalable and compliant cloud-based data management platform for pharmaceutical companies that wish to gather various types of data from multiple IT systems and external data sources. It can be applied to a wide variety of business challenges, enabling pharma companies to become more data-driven and embrace Pharma 4.0.

THE BACKGROUND

The IT landscape of pharmaceutical companies has become a lot more complex in recent years and they will typically have a variety of systems for different business or functional areas, such as Manufacturing Execution Systems (MES), Laboratory Information Management Systems (LIMS), Quality Management Systems (QMS), customer management systems and specialized pharmacovigilance systems for detecting and prevention of adverse effects in pharmaceutical products.

These systems are designed and developed to achieve a particular outcome or automate a given set of business processes, and while they may be perfectly adequate for their designed purpose, the data they generate and manipulate is effectively locked inside each system.

With the rise of new technologies, hyperscale cloud-based platforms and predictive analytics, pharmaceutical companies want to extract the full potential of the data stored in their IT ecosystem by integrating all systems into a single data platform.

By doing so, they can achieve significant benefits by integrating data along the complete value chain, from early research to clinical stages, through to manufacturing, distribution and market surveillance. By moving from a chain into a value loop, they turn insights into business value.

MAKING SENSE OF THE DATA

The proliferation of information siloes, common to many large enterprises, makes it difficult to achieve integrated business intelligence, and manifests itself in different ways in a pharmaceutical company, such as a loss of competitive advantage, an inability to gain market insights, wide deviations from expected results and a higher risk to patient safety, product quality and efficacy.

Enterprises that have tried to break down these siloes using conventional data management solutions have often hit roadblocks due to the lack or scalability, not having the data in a central place for cross-references, the burden of performing error-prone manual activities for cleansing and harmonizing data, or problems validating solutions built on cloud infrastructure. These companies also have a huge challenge in maintaining data integrity, a key requirement that regulators focus on in audits.

The NTT DATA GxP Data Lake helps pharmaceutical companies overcome these data integration challenges.

NTT DATA GXP DATA LAKE

The most significant conceptual difference between a data lake and previous approaches to data integration is that a data lake does not force users into a common data model. That means that a data lake can act as a repository for large quantities and varieties of data, structured, semi-structured and unstructured.

As well as conventional text documents, reports and spreadsheets, this data might include web pages, emails, social media content, database query results, data visualizations, machine learning models, handwritten notes of practitioners, videos from production lines to analyze defects, and much more.

The NTT GxP Data Lake preserves all this data in its original form and capture changes to data and contextual semantics throughout the data lifecycle. This is particularly relevant in the pharmaceutical sector when it comes to ensuring compliance and performing audits, as the data lake keeps track of all transformations and updates performed on a given piece of data, so ensuring data integrity by design.

The data collected in the lake can then be made available to a wide variety of tools and analytical platforms, according to the desired use case.
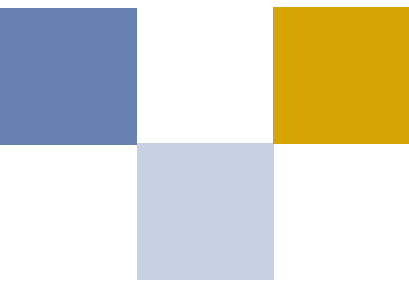
MAIN COMPONENTS

Data Lake

The Snowflake cloud-based platform provides standard data lake features including unlimited scalability, mixed data types, and support for different languages including SQL, Python and Java. Snowflake can be deployed on Microsoft Azure or AWS, according to customer preference.

Compliance Control Room

This acts as the central nervous system for the solution and can be deployed on top of the data ecosystem components to keep them all in check from a GAMP point of view. On a use case level, the criticality of a deployed use case can be determined, including alerting, authorizations and change control. The control room monitors aspects like data freshness, schema drift, value deviations and component health in order to ensure GxP-critical actions can be executed while keeping risks in control.

Graph Database

A graph database uses graph structures for semantic queries and is an important tool for uncovering relationships across diverse data.

Data Integration

Data for the data lake can be ingested directly from any SAP environment with the NTT SAP Connector, from Microsoft Azure Data Factory, Fivetran or using Matillon.

Data Catalog

The data catalog, MS Purview or Atlan, provides an inventory of all the data assets in an organization, and ensures that the most appropriate data for a particular application or can be quickly located. It is also leveraged to ensure governance across the data and manage the quality of that data.

Data Lineage

Data lineage tools keep track of the source systems from which the data originates, how the data was transformed along the way and if it is still trustworthy. It also allows impact analysis to be performed to update existing pipelines. The lineage can pinpoint what use cases consume data coming from the same pipelines and allow regression testing of the use cases before pipeline updates are brought live.
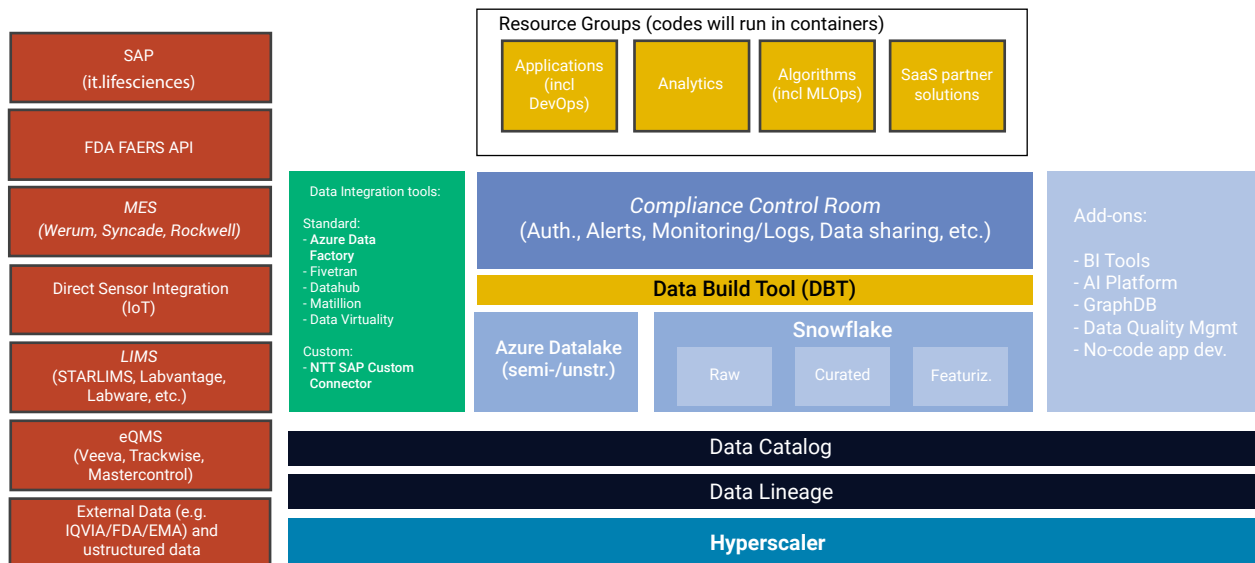
Data Transformation

The main tool to allow controlled data transformation is dbt, a powerful data transformation framework that makes it easy to build complex pipelines, collaborate with other data specialists, ensure version control through Github and bring new transformation and changes to existing transformations into production through a controlled methodology.

USE CASES

- **Data Sciences as a Service (DSaaS) / Direct Co-innovation** – NTT DATA can manage the machine language platform and data infrastructure for the customer, to facilitate co-innovation and provide the data science capacity that a pharmaceutical company may lack.

- **Predictive Quality Management** – Predict output quality parameters on the basis of at-line/on-line sensors. By integrating data from SAP, Data Historian, QMS and LIMS a solution can be built with dashboards to predict and do root cause analysis to understand line dynamics.

- **Integrated Pharmacovigilance** – Integrate external adverse effects data coming from the governing instances with regards to medicines. Provides disproportionality scores to easily detect anomalies and directly look back at production data of a batch to see if anomalies occurred during production.

- **Batch Record Automation** – Position all batch-record relevant information in the data lake before processing. Transforms the data to the right format, and allow direct batch record generation with data from multiple systems (SAP, MES, SCADA/Historian, QMS, etc.) In one real-world use case, this enabled review times to be reduced from two days to two hours and, eventually, just a couple of minutes minutes.

- **Personalized Medicine Platform** – Utilize a data lake setup to allow cell and gene therapies, including genomics data, privacy protection and full chain of custody (Multi-lake setup).

- **Archive/Paper Digitalization** – Use algorithms to digitize data currently captured on paper sitting in archives and desks. The documents can be classified and analysed to extract information from handwritten notes, for example. Once in digital form, this information provides the basis to create new insights and algorithms.

- **Change Assessment/Risk Modelling Framework**– Leverage historical change request data, including risk calculations, written text analytics and forwarding to stakeholders, significantly reducing time required in the change process by up to 70 percent, in one real-world use case.

- **Process Modelling –** Utilize algorithms to accurately model processes across different IT systems with data sourced from SAP, MES and SCADA systems. This allows for more predictive quality scenarios such as predictive quality management for final products and is particularly relevant to production line modelling, for example.

- **Predictive Maintenance –** Use datasets from various databases such SAP, SCADA and MES to discover cause-effect relations based on sensor (IoT) – and historical maintenance data, including researched relationships (e.g. pump vibration and breakdown models).

- **Cross-factory analytics –** Gain insights on production processes across sites to discover inefficiencies in the comparison, but also discover trends or relationships across sites. Ultimately this solutions allows production schedules to become reactive to factory outputs and produce output for optimal profits.

SAP (it.lifesciences)	
FDA FAERS API	
MES (Werum, Syncade, Rockwell)	
Direct Sensor Integration (IoT)	
LIMS (STARLIMS, Labvantage, Labware, etc.)	
eQMS (Veeva, Trackwise, Mastercontrol)	
External Data (e.g. IQVIA/FDA/EMA) and ustructured data	

Resource Groups (codes will run in containers)

- Applications (incl DevOps)
- Analytics
- Algorithms (incl MLOps)
- SaaS partner solutions

Data Integration tools:

Standard:
- **Azure Data Factory**
- Fivetran
- Datahub
- Matillion
- Data Virtuality

Custom:
- **NTT SAP Custom Connector**

Compliance Control Room (Auth., Alerts, Monitoring/Logs, Data sharing, etc.)

Data Build Tool (DBT)

Azure Datalake (semi-/unstr.)

Snowflake
- Raw
- Curated
- Featuriz.

Add-ons:
- BI Tools
- AI Platform
- GraphDB
- Data Quality Mgmt
- No-code app dev.

Data Catalog

Data Lineage

Hyperscaler

GxP Data Lake: components and services

BENEFITS

- Ensures data integrity by design

- Splits data lake into zones for optimal control

- Automates data cleansing and harmonization activities

- Enables big data analytics at scale

- Optimizes integration between data science, engineering, development and validation activities

- Allows self-service analytics for business users

- Generates in-depth insights of dynamics in processes, including influencing parameters and their dependencies

- Creates an effective starting point for continuous improvement processes

WHY CHOOSE NTT DATA FOR YOUR SAP LIFE SCIENCES PROJECT

WE ARE EXPERTS IN LIFE SCIENCES

NTT DATA has worked with a diverse range of enterprises in the life sciences industry for 25 years. We have a 20-year track record specifically helping manufacturers in this sector and have a full understanding of the end-to-end manufacturing and quality business processes used in this sector, which is characterized by increasingly complex systems and disaggregated value chains.

NTT DATA is an active member in the ISPE Pharma 4.0 community and we are passionate about utilizing data to improve the life sciences industry.

As well as a range of sector-specific accelerators, we can help life science companies unlock the true potential of their SAP S/4HANA environment with it.lifesciences, a collection of more than 300 industry best-practice processes already available and ready to adapt for your organization. These include processes to optimize R&D, compliant manufacturing, quality control, supply chain management, and more.

SAP Global Strategic Partner

NTT DATA is a formidable force in SAP services with more than 20,000 SAP professionals, 2,700 SAP clients worldwide, 21 delivery centers and operations in 41 countries. It is a truly global organization committed to seamlessly working together with its clients to deliver SAP-centric projects that can be delivered at scale and across multiple locations.

Our solutions leverage our business-specific expertise and skillsets in SAP technologies but also in other digital technologies, such as AI and IoT, for example. NTT DATA has achieved the highest accolade awarded by SAP, Global Strategic Partner, which recognizes its ability to offer a wide range of best-in-class business consulting and solution implementation services in support of SAP technology.

THE NTT DATA DIFFERENCE

NTT Data SAP Life Sciences Template

- Fully customized SAP S/4HANA system including all core areas such as finance, procurement, sales, logistics and production.

- Includes new SAP modules such as EWM, IBP, TM, VMS, etc.

- SAP Business Technology Platform and Cloud Integration.

SAP Global Strategic Partner

- **Internal Cold Chain** helps life sciences companies better manage their internal cold chain by ensuring the right freezing and thawing times for temperature-sensitive products are strictly enforced using scanners or RFID to identify and track batches.

- **CMO Cockpit & Integration Framework** has been developed to help pharmaceutical companies support the SAP processes for contract manufacturing in an integrated way. It includes a model CMO plant, data governance to simplify the onboarding of CMOs, and role-based apps to support manufacturing and supply chain processes.

- **Analytics Framework for Operational Excellence** allows pharmaceutical companies to easily monitor the KPIs necessary to achieve operational excellence. It includes a data model, a set of KPIs and extractors/views for data acquisition using S4/HANA.

- **EWM AP Integration** simplifies and shortens the time required to implement complex integrated serialization into warehouses processes used by life sciences, so ensuring compliance and improving supply chain visibility.

ABOUT NTT DATA

NTT DATA – a part of NTT Group – is a trusted global innovator of IT and business services headquartered in Tokyo. We help clients transform through consulting, industry solutions, business process services, IT modernization and managed services. NTT DATA enables clients, as well as society, to move confidently into the digital future. We are committed to our clients' long-term success and combine global reach with local client attention to serve them in over 50 countries. Visit us at nttdata.com.

JOIN US!