MindSphere Semantic Data Interconnect

Benefits
• Derive and understand correlation of data from disparate enterprise sources
• Unlock and maximize value of enterprise data with contextualization
• Enable comprehensive analytics of data across domains
• Reduce application development time by creating automatic schemas
• Improve productivity of application developers with semantic modeling
• Perform big data analytics and contextualize data

Features
• Connect external data from disparate sources
• Infer schemas from various data based on schema-on-read
• Maintain relationships across data sources using semantic modeling
• Perform big data semantic queries
• Provide workflows to register, extract, load, infer schema, build semantics and query

Correlating enterprise data from disparate systems to maximize value

Summary
The semantic modeling capabilities of MindSphere Semantic Data Interconnect enable the user to create relationships across multiple sources of data. This allows MindSphere to link and access data across disparate information sources, including manufacturing resource planning (MRP), enterprise resource planning (ERP), product lifecycle management (PLM) and Internet of Things (IoT) data lakes. Using Semantic Data Interconnect greatly simplifies the process with rich capabilities for schema inference, metadata catalogs, unique semantic correlation identification and built-in query capabilities. This significantly reduces the time required by enterprises to unlock and realize the value of data.

Additionally, the support of semantic correlation between digital and physical data in Semantic Data Interconnect enables the development of digital twins, digital threads and closed-loop feedback controls. Closed-loop and digital-thread analysis provides the ability to validate models by comparing the actual conditions with planned or modeled conditions to determine the reason behind a behavior.

Current situation
Today, every organization is struggling with the challenges of explosive data growth. Data is estimated to grow 10 times biannually with 90 percent of data created in the past two years. Addressing this challenge gets more complex when considering the variety and velocity along with the volume. Further, data integrity and related security makes this daunting. But even with these challenges, the potential value of this data to the enterprise is significant – it allows them to gain insights and create new business opportunities.

Most enterprises implement point-to-point integrations to collect and merge data, creating multiple copies and vague rules for data authority and a system of records. Enterprises try to harmonize the multiple data sources and dedicated business intelligence (BI) tools available, but most find it difficult to maintain consistency over even a short period of time. This often leads to cost overruns and failure to realize the true value of the data.
Challenges
To unlock the data’s potential, organizations need to create, maintain and support relationships with any combination of complex structures and data types across multiple domains; for example, original equipment manufacturers (OEMs), customers, products, assets, manufacturing, operations, financial data and location data. Additionally, the solution must enable easy linking of data with minimal or no manual effort, delivering the right amount of automation, governance, data control, relationships and access.

It should also be compatible with existing data storage mechanisms (such as enterprise systems, data historians, data lakes) and BI tools. Customers are also looking to minimize migration or duplication of data. All the new applications will need to have unified application and integration capabilities within a single platform while leveraging definition re-use for a quick, efficient and cost-effective methodology to add new data sources, types, attributes and associations.

Solution
Semantic Data Interconnect provides customers with the ability to correlate data from disparate systems and gain powerful insights, which is not possible when the information is in disjointed applications and storage locations. It enables an organization to build a graph composed of knowledge points (nodes) and a variety of semantic relations (edges) between them. This enables the user to get qualitative information by accessing the respective data and relations, as well as perform quantitative analysis.

Semantic Data Interconnect provides the following capabilities:
1. **Consumption of data** – Enables enterprises to connect external data from disparate sources to derive meaningful insights.
2. **Schema-on-read** – Offers the ability to infer schemas from data using a schema-on-read data handling strategy to increase versatility.
3. **Semantic modeling** – Maintains a relationship with data sources using semantic modeling capabilities.
4. **In-built queries** – Includes in-built big data query capabilities to enable retrieval of both explicitly and implicitly derived information.
5. **Data management** – Provides workflows to register, extract, load, infer schema, build semantics and query multiple data sources.

Semantic Data Interconnect includes a configurable extract, load, transform (ELT) integration process that supports the transformation and load of structured or semi-structured data without predefining or providing the schema.

The schema-on-read capabilities for external system data significantly reduces the need to have special skill sets to understand systems and their schemas, and in turn provides significant time, resources and investment savings.

The ELT engine facilitates scheduling automatic jobs to extract schema out of ingested files and store transformed data and metadata. Semantic Data Interconnect also provides versioning of the schema and related data, so you can understand the source and lifecycle of the data.

Using Semantic Data Interconnect enables you to leverage the generated schema to infer the correlation of data and recommend a semantic model. The ability to infer the semantic model significantly reduces the investment in skill set, time and resources and provides a value-added starting point for the data analyst. In cases in which the data analyst already knows the semantic model, Semantic Data Interconnect also provides the ability to import the semantic model as a starting point in the Owl file format or JSON input. Whether it is the inferred or imported semantic model, Semantic Data Interconnect enables this via a set of fully documented application programming interfaces (APIs), which can be leveraged with any third-party semantic modeler.
Finally, Semantic Data Interconnect provides users with the ability to define, build and save contextual/semantic queries on semantic models without requiring an understanding of how the data is represented in disparate systems. The related in-built query capabilities allow users to execute a query on huge volumes of data coming from multiple sources.

In summary, the Semantic Data Interconnect solution generates value from an organization’s data when following these six steps:

1. Register multiple data sources (for example, ingest files) in the source registry.
2. Ingest external data from identified sources.
3. Verify/import/create/adjust automatically built semantic model.
4. Map/validate semantic model properties to schemas extracted from ingested data using generated recommendations.
5. Create/update semantic queries.
6. Consume correlated data in the application or BI tools.

The capabilities of Semantic Data Interconnect allow customers to build domain-specific semantic models and access visualized data from the sources on dashboard applications (created by application developers using Semantic Data Interconnect) in a configurable way. Businesses gain powerful insights from the correlated data and MindSphere analytics to support faster and smarter business decisions.

MindSphere

MindSphere is the cloud-based, open IoT operating system from Siemens that connects real things to the digital world, and enables powerful industry applications and digital services to drive business success. MindSphere’s open Platform as a Service (PaaS) enables a rich partner ecosystem to develop and deliver new applications.