STARSCHEMA

SOLUTION BRIEF

Azure Data Lake

Full-stack data lake implementation

Common Data Lake Challenges

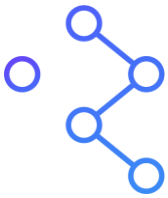
Data Engineering Challenges

- Large numbers of disparate data sources
- Huge volume of structured and unstructured data
- Very volatile data velocity

Operational Challenges

- Difficult to implement
- Lack of specialized skills
- Expensive operational costs
- Lack of integration between ETL, Store, Analyze and Reporting

Data Consumption Challenges



- Lack of sophisticated processing and automation of data collected
- Slow or poor quality data presented to users
- Difficult to use reports

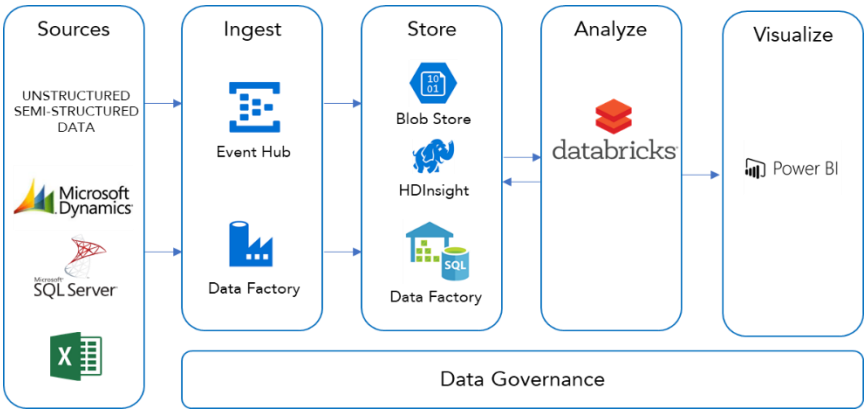
Highlights

- The flexibility of the Datalake architecture supports other Azure components, creating a custom designed architecture suited to your company's needs.
- HDInsights along with Event Hub store semi/non-structured data, and process them with Datakbricks jobs such as Spark or R.
- Microsoft Azure Datawarehouse, capable of storing and processing hundreds of terrabytes of data, is the central hub for all the structured and processed data coming from various sources eg. ERPs, flat files or HDInsights.
- PowerBI featuring stunning visualizations and dashboards offers responsive, interactive reporting for data consumers. PowerBI is also capable of connecting to Event Hub directly for real-time dahboards.
- A Centralized Data Governance Model (DGM) includes all the essential aspects that must be specified during the project. It is important not just as a one-time DGM setup but a set of rules and business processes to be maintained after the Data lake implementation.

Flexible, full-stack design

We design and implement data lakes to analyze data in distinct ways, gain insights and create value out of the data your organization generates and imports. Our standard BI Data lake solution implemented on Microsoft Azure platform is based on Lambda Architecture providing flexibility to process either structured data coming from traditional SQL databases or semi/non-structured data ingested from IoT devices, logs, documents.

Architecture



Data Lake Architecture

How We Work

Our implementation methodology involves a mixture of agile and waterfall process management. In the beginning of a Data Lake implementation, we use a waterfall-like method to gather all the required information and set-up the AsIs/ToBe and ToDo lists. With those inputs, Agile scrum project management is used to implement tasks in several sprints. We strongly believe that an efficient Data Lake implementation cannot be successful without involving the client early in the process.

About Starschema

At Starschema we believe that data has the power to change the world and data-driven organizations are leading the way. We help organizations use data to make better business decisions, build smarter products, and deliver more value for their customers, employees and investors. We dig into our customers toughest business problems, design solutions and build the technology needed to compete and profit in a data-driven world.

GET IN TOUCH
info@starschema.com
starschema.com
+1 415 944-5455