SNOWFLAKE

CENTRAL DATA CLOUD IN ENERGY TRADING

INDUSTRY PRESSURE

Energy trading has evolved from a simple value chain to a central element of the value chain. The underlying prices are dependent on availability, weather, times of day, and many other parameters. And it is precisely this data that needs to be centrally collected, processed, and made available to all the necessary processes, systems, data science and reports, thus creating the greatest challenge for trading. And even beyond that: If this data cannot be made transparent to the customer, there is a risk of losing them.

Our customer of the energy sector had to face the same challenges on a daily basis and was aiming for efficienty connecting new markets and thus data to it‘s IT landscape.

This data is to be made available to the systems and employees in order to increase the effective benefit. All types of new data should be made directly available, whether it is market, trading or master data. The project‘s goal was to dissolve the existing point-to-point connections of the operational systems and replace them with a central data platform.

By establishing such a platform, new data sources can be integrated faster, which reduces TCO costs, and by making data centrally available, employees are able to make faster decisions thus increasing revenue.

BUSINESS IMPACT

The energy trading sector is facing several challenges such as a rising complexity. With a strong competition and ever new power generation and usage options, such as public and private solar panels or private charging points, the pressure is on. There is no single point of truth anymore and companies must maintain this complex environment. To ensure energy input, transparency to the customer and managing the various data sources, a central data platform is key. In building this, a number of different challenges arise, such as the choice of architecture, the complexity of the technologies, provision of new data and, above all, the issue of time-to-market: How quickly can new requirements be implemented?

Improve data availability

New market data is directly available over the platform and results can be consumed in milliseconds. Additional sources are developed in weeks and not months to reduce time-to-market.

Transparency

Central data models are defined to improve the communication between the different departments. A data catalog helps to enable the transparency what data is available in the platform and how the data can be consumed. Also, data ownership and GDPR processes are centrally defined, to help understand the use of the data.

OUR SOLUTION

Creating the best solutions sometimes requires relying on multiple software vendors. Snowflake offers a variety of benefits such as the seamless integration into infrastructure landscapes and low administration overhead. Hyperscalers such as AWS, Azure or Google offer a reliable base.
With its more than 200 services, AWS for example provides the backbone of the infrastructure for the cloud data platform, which makes it possible to deploy the infrastructure very easily. Administration and maintenance is kept low, and is enhanced by the automation software WhereScape. This stores all metadata in a repository, as well as creates, maintains and generates Snowflake SQL structures through templates.

Snowflake

Snowflake, with its Data Cloud delivered as Software as a Service, is able to read and import JSON files stored in the S3 buckets. With the JSON files, it is possible to directly process the data using „schema-on-read“ without having to store the data further.
By integrating with the AWS landscape, Snowflake is able to leverage its full potential, for example, with External Functions, Lambda Python Connectors or S3 Bucket Integration. By using Secure Data Share, data can be provided via another standardized and secure interface.

AWS

As the central cloud infrastructure, a wide variety of services were enabled in AWS to provide the foundation for the platform. This involved the use of Lambdas – serverless instances that execute Python code, API gateways, SNS, S3 buckets and EC2 instances, among others.
Using Infrastructure-as-Code (IaC), templates could be defined that allowed new structures to be created consistently and deployed to the different environments.

WhereScape

WhereScape is a data automation tool that generates native Snowflake SQL code based on metadata and templates. The code is directly rolled out to the respective Snowflake database.
In addition, WhereScape provides a modeling component to build a Data Vault 2.0 warehouse. Furthermore, WhereScape generates a complete and always up-to-date documentation based on the metadata, provides a data lineage and impact analysis, and enables a structured deployment to further environments. Using the metadata and templates, it is possible to create new entities within Snowflake, manage them, and make the data available in just a few days.

CASE STUDY

About the company

The customer comes from the energy trade area and belongs to the leading ones on the German-speaking area. The company has been trading for more than 20 years and focuses on electricity, gas and fuels.

Challenges

The customer‘s challenge is to centralize the entire IT landscape in the cloud. This includes breaking point to point connections of operational and analytical systems with each other and building a central cloud data platform that is a single point of truth for all relevant data of the company.

API

The data is to be event-driven, loaded into the platform via a central API, and re-provisioned via a standardized API. This API is based on components from AWS such as the API Gateway and Snowflake SQL Interface and Secure Data Share, to implement all requirements.

Event driven

By using Snowpipe, new data in S3 can be loaded directly into Snowflake and queried from there via Lambda using Python Connector and SQL API. New data records in Snowflake are published to an SNS using External Functions and thus all subscribers are informed. This enables not only a data-driven integration to Snowflake but also a direct delivery of data to all data consumers.

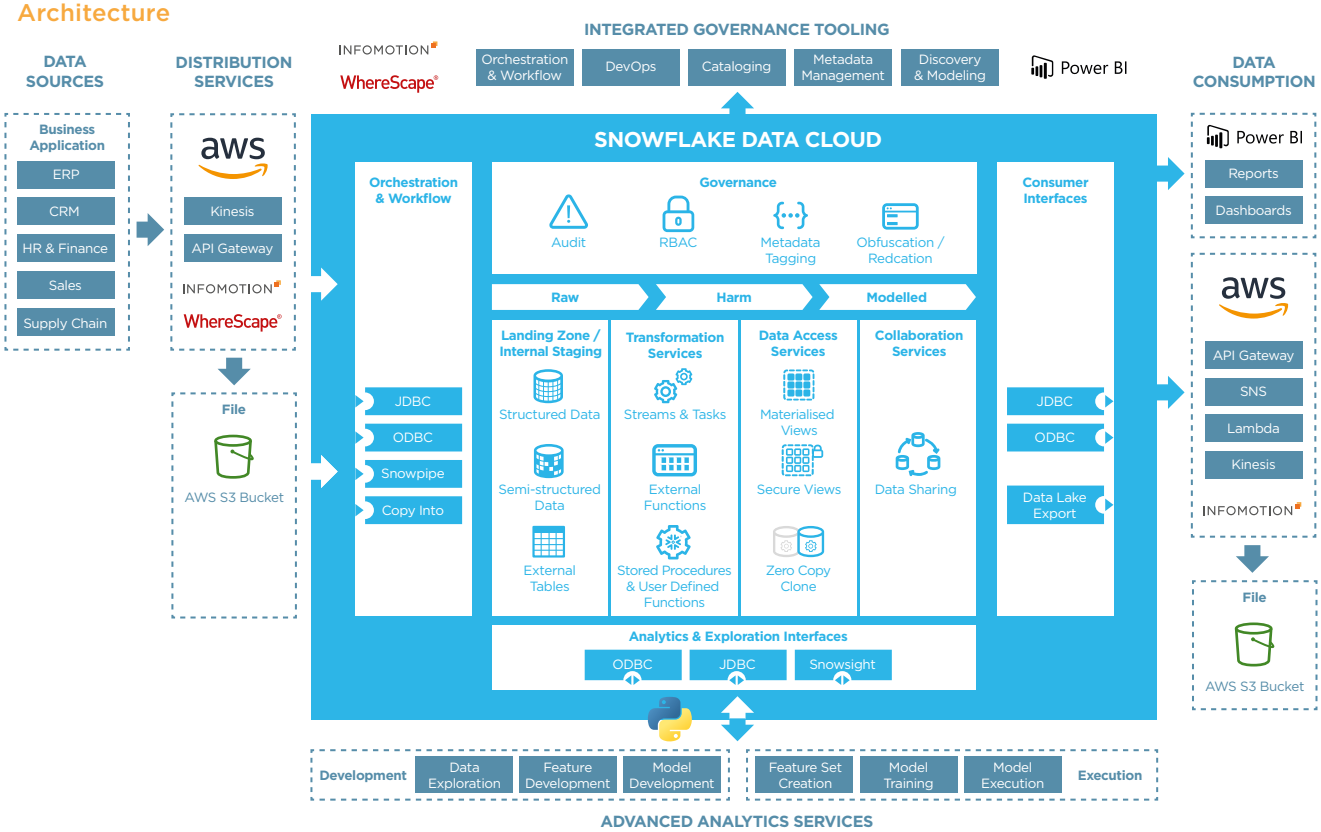
Time-to-market

The seamless integration of the three components into the AWS landscape allowed the first use cases to be implemented in just a few sprints. Using WhereScape and the open template base in combination with the metadata, a fast and standardized implementation of new requirements is possible. Bugs on the production side can be easily addressed by adapting a template or the metadata. New data objects and requirements can be implemented within a sprint, which shortens the deployment of new requirements enormously.

Performance

By using Snowflake, all types of data can be stored centrally and joined together. This includes master data, facts, key figures, market data and time series. Snowflake makes it possible to efficiently store billions of rows of time series on the one hand and to query and provide the data in milliseconds on the other.

Architecture



OUR PHILOSOPHY

We have been developing data warehouses, data lakehouses and cloud platforms for our customers with Snowflake since 2019. This expertise is demonstrated on the one hand by our best practices for analytics and data management, and on the other hand by a team of consultants, numbering 45 as of August 2021, who work intensively with Snowflake. As one of two Premier Partners in the German speaking region, we can accompany our customers in a targeted and tailored manner on their way to the cloud data platform and to a data-driven company.

With the help of our process model, we accompany clients from the cloud & data strategy to the architecture and development of MVPs to the takeover of operations. Together with the customer‘s requirements, we define the data, analytics and cloud roadmap for the company and find solutions tailored to the source systems. In doing so, we bring our expertise from countless analytics projects to enhance the Snowflake Data Cloud, to help clients deal with their data in an agile manner and minimize time-to-analytics. Analytics is an essential element in business decisions that can foster a continuous learning cycle. We train our customer’s employees and accompany them through this change process, to more data performance until the end.

www.infomotion.de

INFOMOTION
we love data

INFOMOTION
we love data