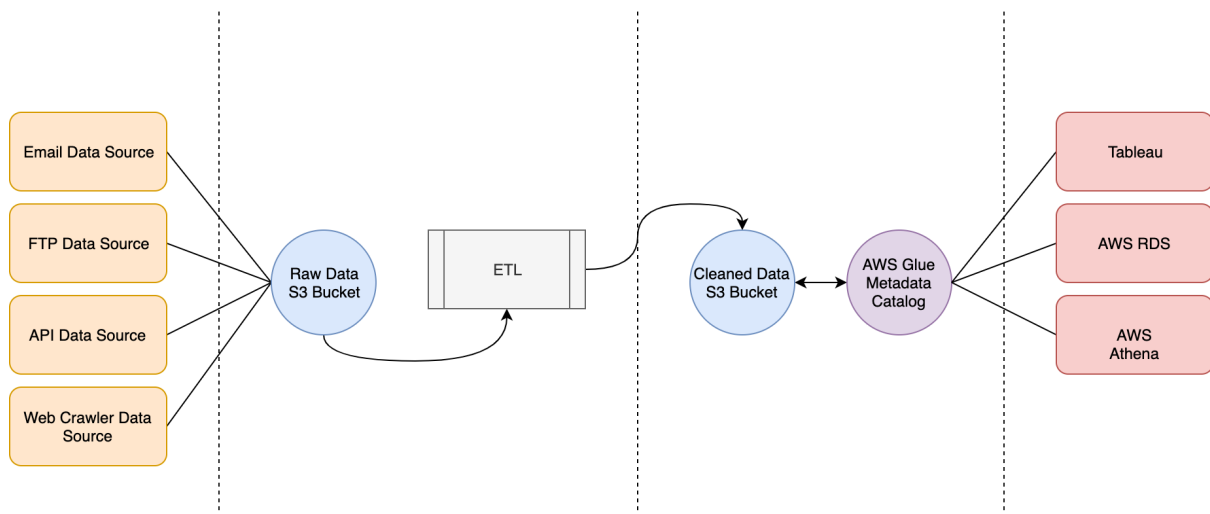


# Data Lake Case Study

The client is a german analytical company that collects and aggregates data to sell to industry players.

## Problem

The client has various data sources: databases, APIs, SFTPs, and data dumps from 3rd party vendors. These data sources must be unified into a single domain model, and access should be provided to BI teams using familiar tools such as RDS, Redshift, Tableau, and Athena.



## Solution

- korzh.cloud helped with technology selection.
- We designed a set of pluggable adapters to import specific data sources.
- Enabled adapters discovery and self-service via AWS Service Catalog.

Start a project?

[platon@korzh.cloud](mailto:platon@korzh.cloud)

<https://korzh.cloud>



- Conducted data lake workshop with customer employees.
- Created an automated workflow technology for improving data security and maintenance.
- Support infrastructure scales up and down depending on the current load.
- korzh.cloud set-up monitoring and alerting through logging and custom metrics implementation.

## Methodology

We used scrum - periodically contacted the clients, gave them an account of what was happening, shared the project results, and made the necessary adjustments based on the feedback.

## Deliverables

- Documentation / Wiki
- Code
- Workshop artifacts

Start a project?

[platon@korzh.cloud](mailto:platon@korzh.cloud)

<https://korzh.cloud>

