Data Governance
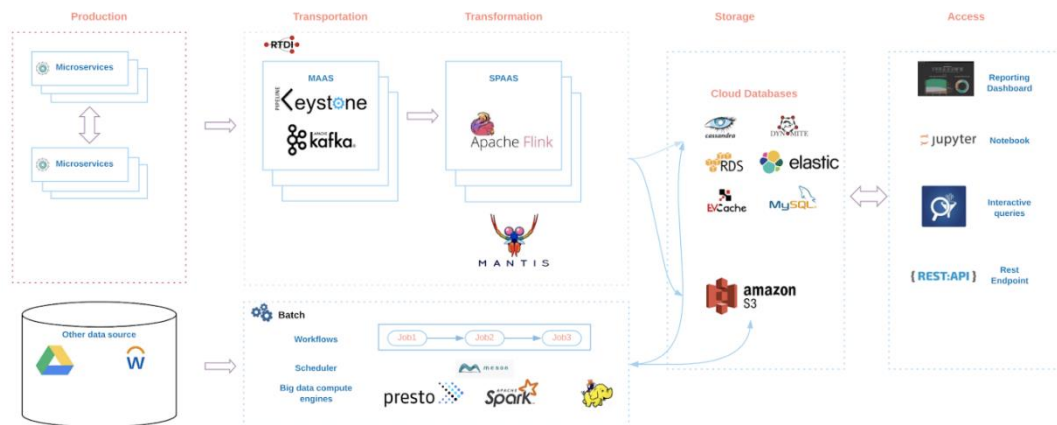
Netflix's data system governance is shared by multiple teams due to the complexity of the Netflix data landscape.To map out all the data-artifacts, a comprehensive and precise data lineage system must be built. This is a challenging task that calls for a scalable architecture, a robust design, a strong technical team, and cross-functional cooperation.

(Netflix's data landscape) (source: Netflix)

Data Governance Implementation:

Establish a robust data stewardship program→ Build a complete data lineage system

- ❿ Create a flexible data model which includes various types of data artifacts and their relationships to support different business use cases.
- ❿ Create the system in a way that it can easily be integrated with an increasing number of data tools and platforms.
- ❿ Create a system level instrumentation approach for Spark. (Since Spark is the company's main big-data compute engine and every time Spark was upgraded, the Spark plan also changed.)