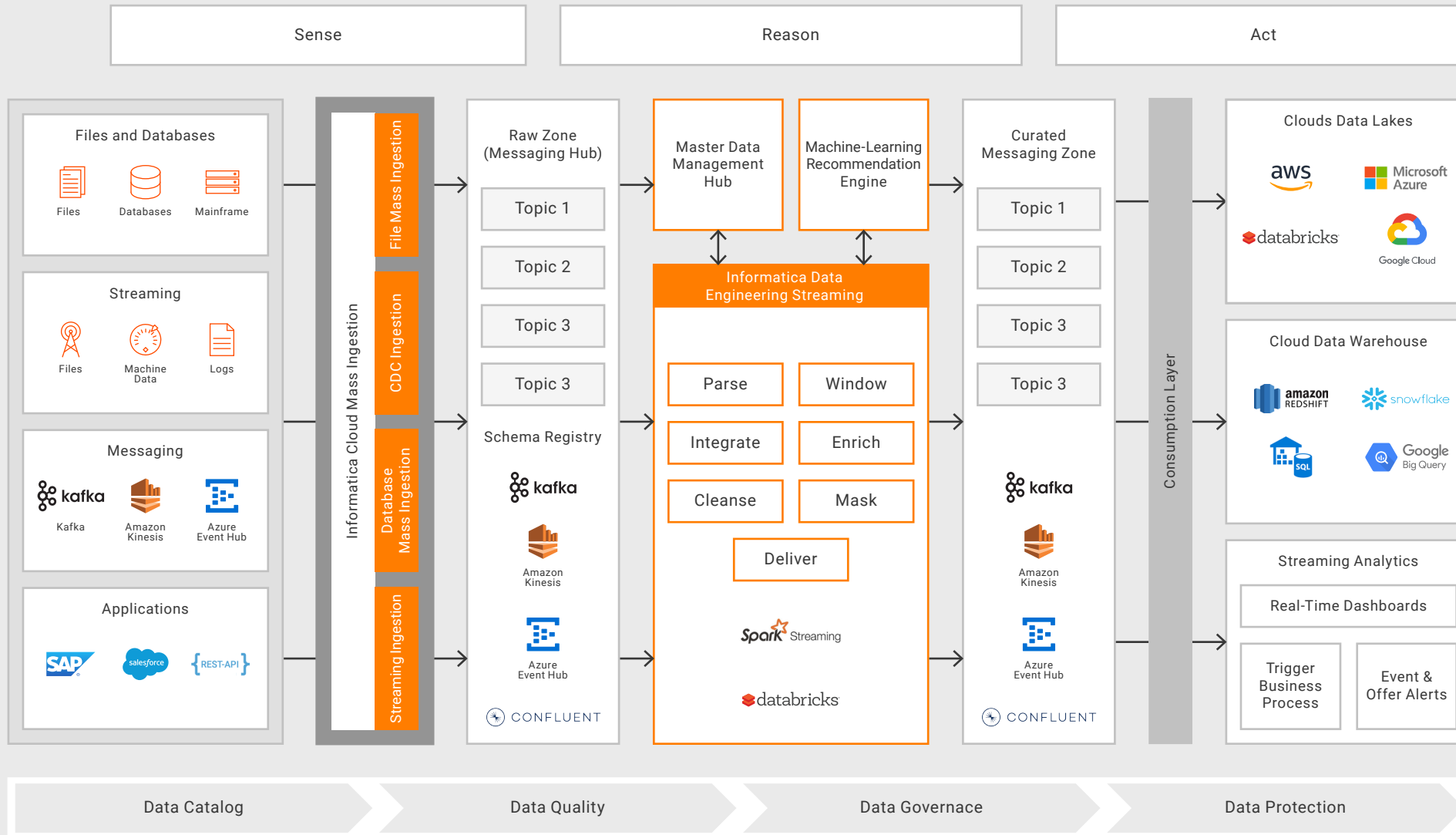




End-to-End Streaming Data Management Solution for Kappa Architecture

Kappa architecture is a streaming-first architecture deployment pattern – where data coming from streaming, IoT, batch or near-real time (such as change data capture), is ingested into a messaging system like Apache Kafka. A stream processing engine (like Apache Spark, Apache Flink, etc.) reads data from the messaging system, transforms it, and publishes the enriched data back to the messaging system, making it available for real-time analytics. Additionally, the data is distributed to the serving layer such as a cloud data lake, cloud data warehouse, operational intelligence or alerting systems for self-service analytics and machine learning (ML), reporting, dashboarding, predictive and preventive maintenance as well as alerting use cases.



5 key features of Informatica's streaming solution to support a Kappa Architecture

- Ingest data from a variety of sources using Informatica's Cloud Mass Ingestion (Sense):** Informatica offers the industry's first cloud-native, schema-agnostic [mass ingestion solution](#) for ingesting data from variety of sources, including files, databases, CDC, IoT, HTTP, REST API, logs, clickstreams, and other streaming sources onto messaging systems like Apache Kafka, Confluent Kafka, Azure EventHub, and Amazon Kinesis. It offers edge transformations so data engineers can apply simple edge processing (for example, filtering, splitting records, etc.) before loading the data into cloud data warehouses or data lakes.
- Process streaming data by applying complex transformations (Reason):** Informatica Data Engineering Streaming (DES) provides an AI-driven scalable streaming data processing solution that leverages the power of open source Apache Spark Streaming for horizontal scaling. It also offers complex transformations on the streaming data to enrich, process, cleanse, and aggregate streaming data with an easy-to-use experience. It includes support for parsing complex unstructured data using [Informatica Intelligent Structure Discovery](#) (ISD) and Confluent Schema registry – which also addresses schema drift use challenges. Automatic addressing of schema drift helps customers to keep their streaming jobs running without frequent restarts, hence ensuring uninterrupted streaming data pipeline. It helps customers reuse the transformation logic and apply data quality rules as well as perform cached and un-cached lookup into various sources – all in real time as part of the streaming pipeline.
- Operationalizing actions on streaming data (Act):** Informatica Data Engineering Streaming (DES) provides capabilities to operationalize complex business rules as well as ML models as Python code. Additionally, [Informatica Cloud Application Integration](#) (CAI) helps customers trigger business processes on events so that the integration with downstream business applications can be completely automated.
- Metadata-driven approach for end-to-end data management:** Informatica provides a metadata-driven platform for end-to-end data management, which helps customers to make the switch from Lambda to Kappa architecture (or vice versa) as the organization's needs evolve. This helps customers to be agile in terms of handling ever-changing business requirements without spending a lot of time and money on re-engineering the solution.
- Catalog and govern streaming data management pipeline:** Informatica Enterprise Data Catalog (EDC) and [Informatica Axon Data Governance](#) offers the ability to extract metadata from a variety of sources and provides end-to-end lineage for the Kappa architecture pipeline while enforcing policy rules, providing secure access, dynamic masking, authentication and role based user access. This helps customers manage and govern the streaming pipeline with ease.