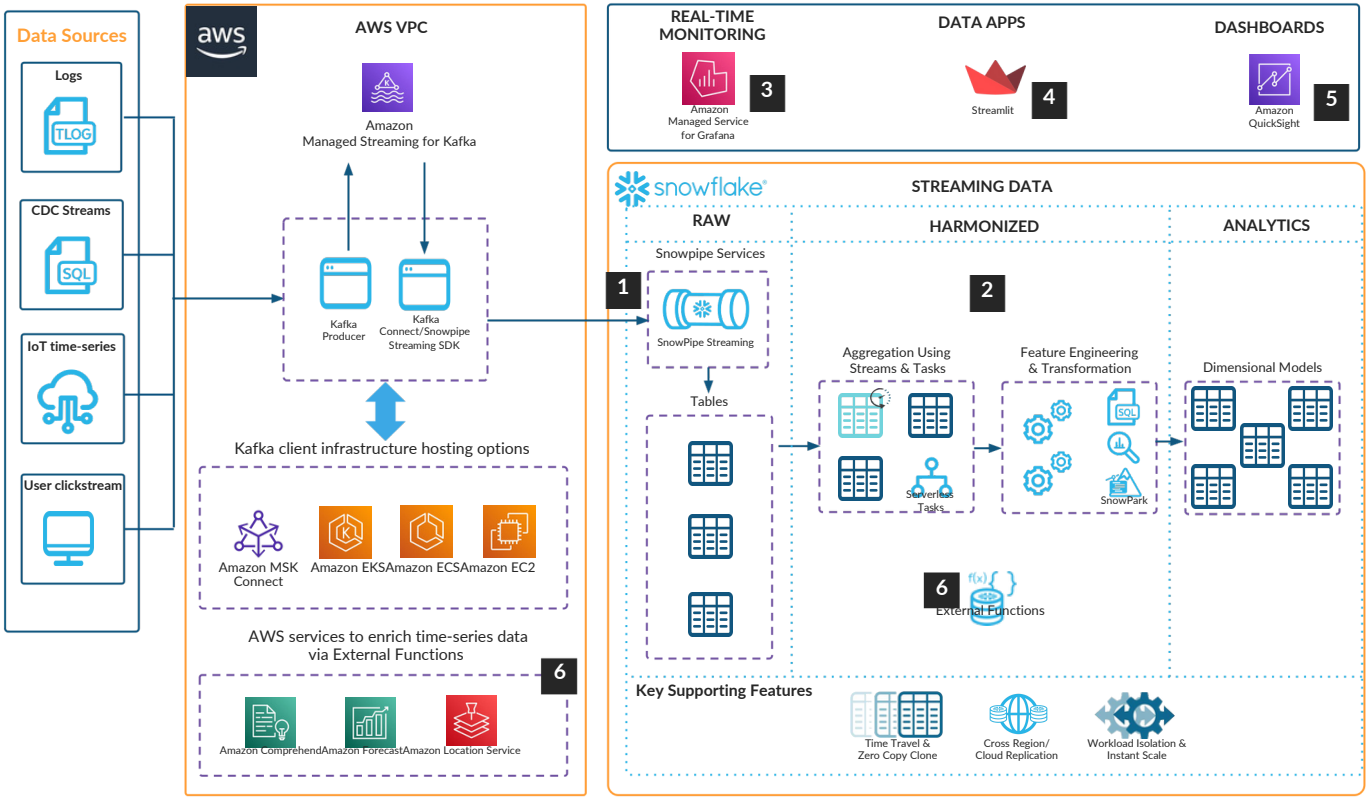
AWS - SNOWPIPE STREAMING WITH AMAZON MSK
REFERENCE ARCHITECTURE



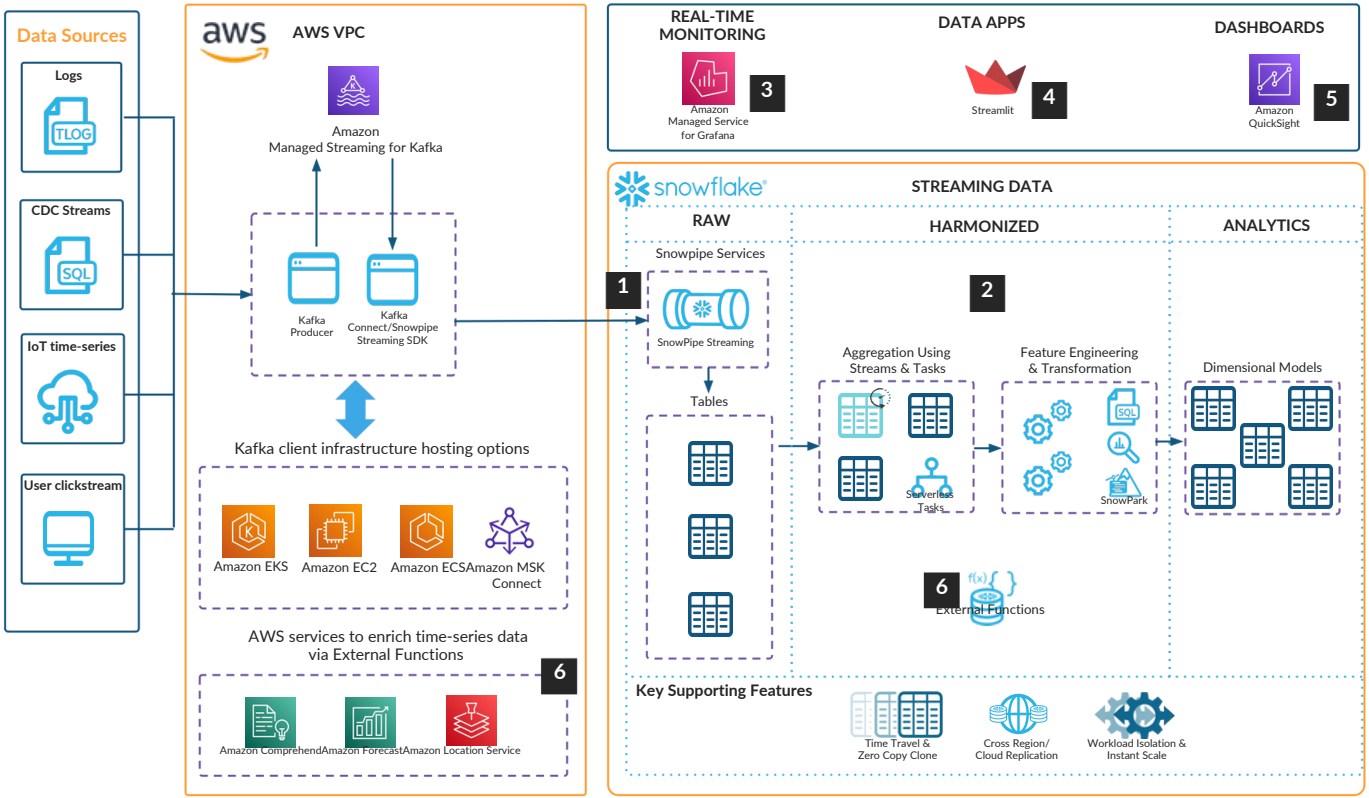
OBJECTIVE

Snowflake enables organizations to be data-driven by offering an expansive set of features for creating performant, scalable, and reliable data pipelines that feed dashboards, machine learning models, and applications. But before data can be transformed and served or shared, it must be ingested from source systems. Snowpipe Streaming is designed as an ingestion method for our Snowflake Connector for Kafka to address the issue. Whether you use Snowpipe Streaming as a standalone client or as part of your Kafka architecture, you can create scalable and reliable data pipelines with a fully managed underlying infrastructure with built-in observability.

DESCRIPTION

1. Raw data lands in Snowflake using Snowpipe Streaming for real-time streaming data. Kafka connector for Snowpipe Streaming can be hosted on Amazon MSK Connect, Amazon EKS, Amazon ECS or Amazon EC2
2. Data gets transformed using Snowpark to leverage the power of Java or Python for semi-structured Streaming data types. Data can then be aggregated and prepared using SQL for analytics
3. Real time dashboards on streaming data driven by separate compute and integration with Grafana
4. Streamlit Apps can be used to display real time dashboards from the raw data zone, as well as historical analysis from the Analytics zone
5. Data gets analyzed and visualized using separate compute and BI tools like Amazon QuickSight
6. AWS services can be called via External Functions to further enrich time-series data in Snowflake

AWS - SNOWPIPE STREAMING WITH AMAZON MSK
REFERENCE ARCHITECTURE



OBJECTIVE

Snowflake enables organizations to be data-driven by offering an expansive set of features for creating performant, scalable, and reliable data pipelines that feed dashboards, machine learning models, and applications. But before data can be transformed and served or shared, it must be ingested from source systems. Snowpipe Streaming is designed as an ingestion method for our Snowflake Connector for Kafka to address the issue. Whether you use Snowpipe Streaming as a standalone client or as part of your Kafka architecture, you can create scalable and reliable data pipelines with a fully managed underlying infrastructure with built-in observability.

DESCRIPTION

1. Raw data lands in Snowflake using Snowpipe Streaming for real-time streaming data
2. Data gets transformed using Snowpark to leverage the power of Java or Python for semi-structured Streaming data types. Data can then be aggregated and prepared using SQL for analytics
3. Real time dashboards on streaming data driven by separate compute and integration with Grafana
4. Streamlit Apps can be used to display real time dashboards from the raw data zone, as well as historical analysis from the Analytics zone
5. Data gets analyzed and visualized using separate compute and BI tools like QuickSight
6. AWS services can be called via External Functions to further enrich time-series data in Snowflake