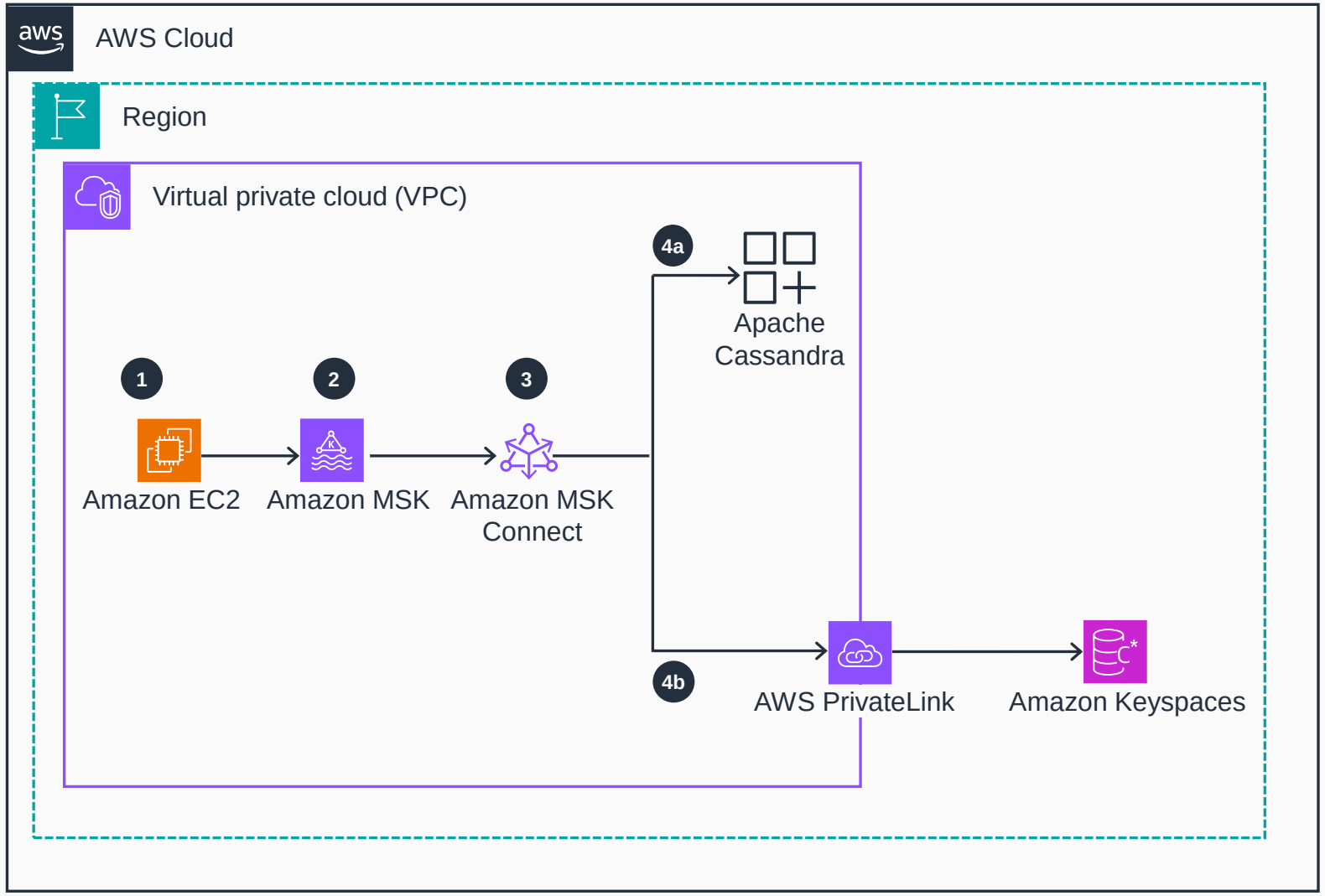
Guidance for Continuous Data Migration from Apache Cassandra to Amazon Keyspaces

This architecture diagram shows how to perform concurrent dual writes to Apache Cassandra and Amazon Keyspaces for Apache Cassandra. It uses Amazon Managed Streaming for Apache Kafka (Amazon MSK) during data migration operations.



1. The Apache Kafka console producer command line interface (CLI), hosted within an **Amazon Elastic Compute Cloud** (Amazon EC2) instance and functioning as a Kafka client, directly produces and publishes messages to a Kafka topic within the **Amazon Managed Streaming for Kafka** (Amazon MSK) service.

2. The Kafka client sends the messages to **Amazon MSK**, which processes these messages, distributes, and replicates data across its brokers for reliable processing and fault tolerance.

3. The **Amazon MSK** Connect feature, which is a component of the **Amazon MSK** service, ingests the data from the Kafka topic and writes it to both the Apache Cassandra and **Amazon Keyspaces** data stores.

4a. The Apache Cassandra Sink connector, which is integrated within the **Amazon MSK** Connect feature, selects the routed messages and inserts them into the designated Apache Cassandra table.

4b. Simultaneously, the **Amazon MSK** Connect feature routes the messages and inserts them into the **Amazon Keyspaces** data store. This is done through an **Amazon Virtual Private Cloud** (Amazon VPC) endpoint, which is powered by the **AWS PrivateLink** service.

Reviewed for technical accuracy September 13, 2024

AWS Reference Architecture