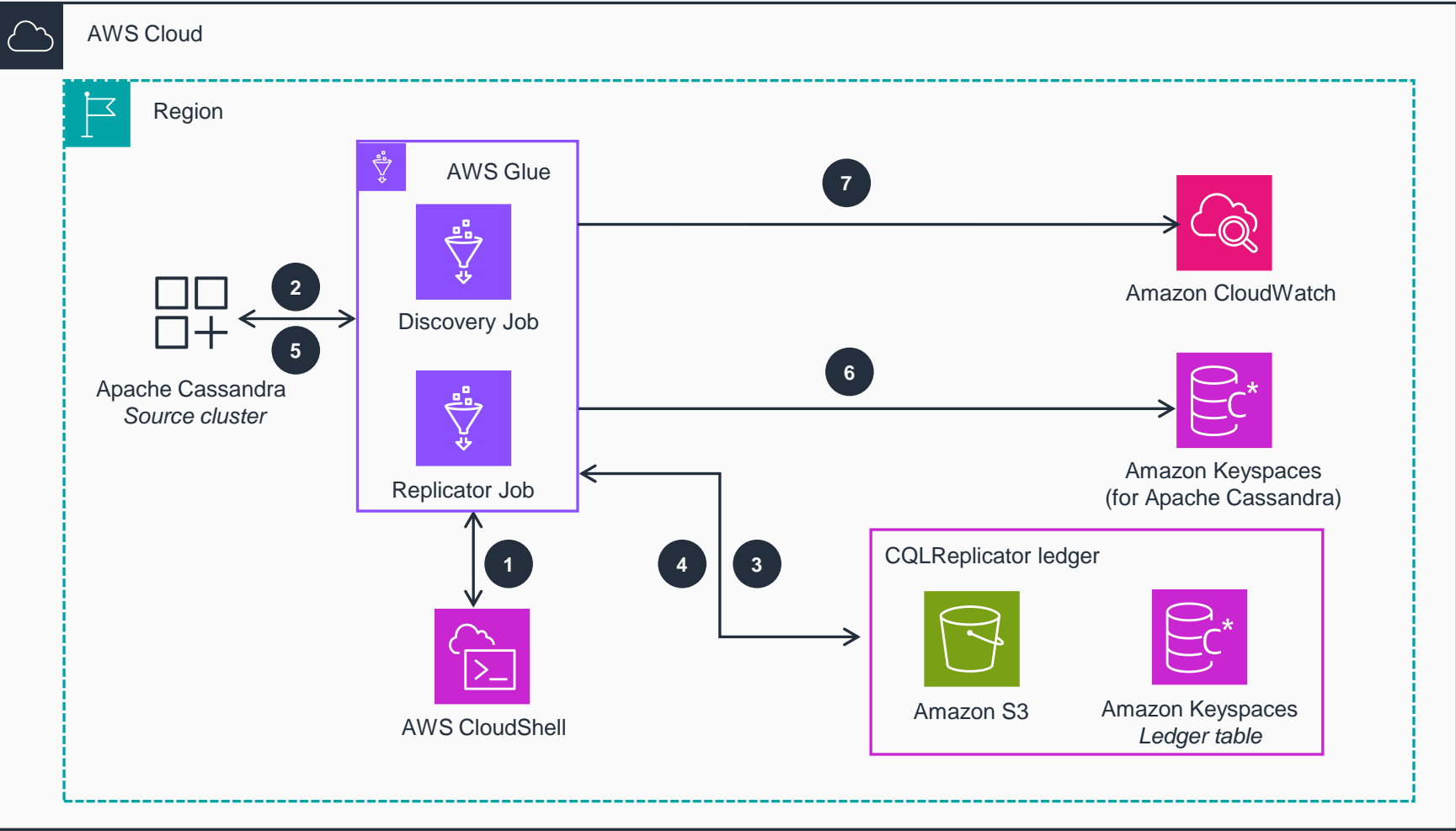


# Guidance for Near Real-Time Data Migration from Apache Cassandra to Amazon Keyspaces

This architecture diagram shows how to migrate self-managed Apache Cassandra clusters to Amazon Keyspaces using a tool called the CQLreplicator.



- 1 Initiate the CQLReplicator in **AWS CloudShell**, which creates two **AWS Glue** jobs called **Discovery** and **Replicator**.
- 2 The **AWS Glue** Discovery job connects to the Apache Cassandra source cluster and collects all the primary keys with the latest timestamp.
- 3 The **Discovery** job then persists the primary keys in an **Amazon Simple Storage Service (Amazon S3)** bucket. If the Discovery job has been executed previously, it compares the old and new primary keys to determine the latest changes. It then stores the changes in **Amazon S3** and records the location of the bucket key in a ledger table stored in **Amazon Keyspaces**.
- 4 The **Replicator** job scans the ledger table in **Amazon Keyspaces** for new change sets.
- 5 When the **Replicator** job finds a change set, it reads the keys and queries the Apache Cassandra source cluster for the latest row data.
- 6 The destination **Amazon Keyspaces** table will receive the data, including the complete primary key and the most recent row values.
- 7 The **AWS Glue** jobs push all the logs to **Amazon CloudWatch**.