# WvChalange Tasks

Architecture Diagram



Data Stores

Crawler

Data Catalog

AWS Management Console

Schedule or Event

CSV

S3 Bucket / RAW
S3 Bucket / Parquet

Glue

Athena Query

**1**   **2**   **3**   **4**
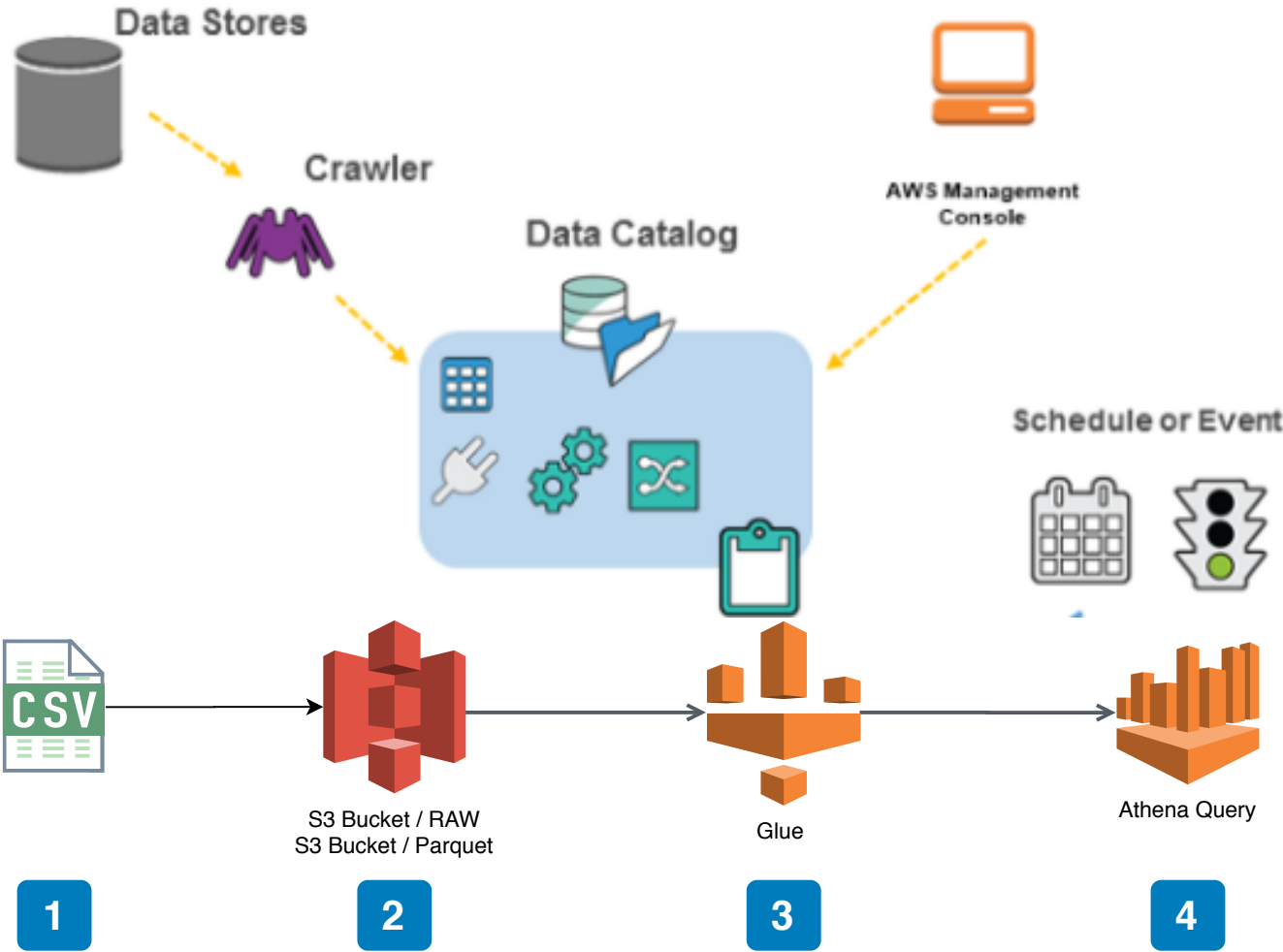
**1** Ingestion of CSV flat file. Storage was set up and a Iac template created.

**2** Files land on a folder named Raw on an S3 Bucket

**3** Created Crawler on Glue to acomodate raw data to right metadata on Data Catalog. as well as a Glue job running with pa pyspark transformation to aggregate data.

**4** 2 table were created named raw for original files and Parquet for aggregated output data as well as an View with UI names.