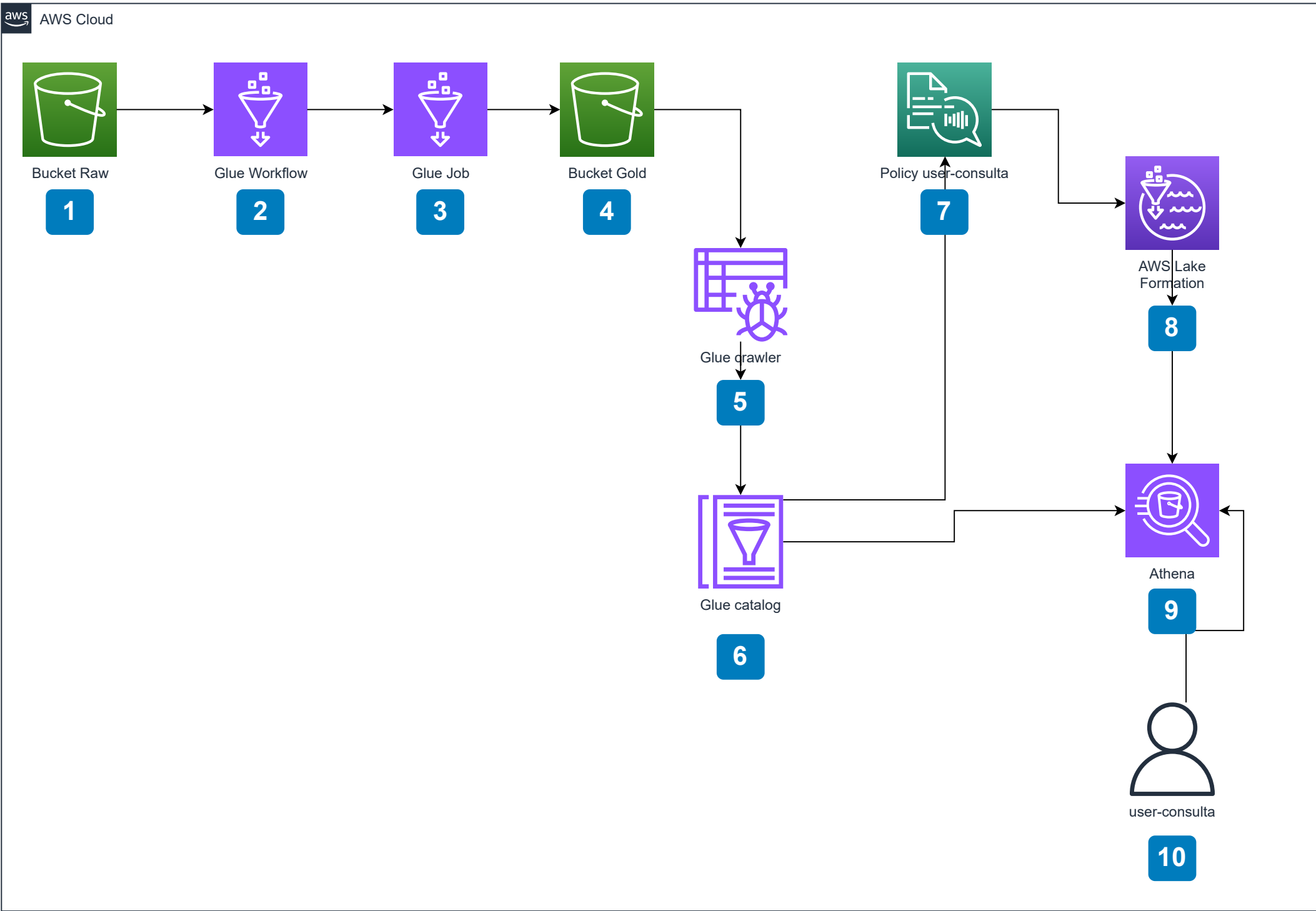


Desafio Técnico A3Data



- 1 Bucket camada Raw que armazena as tabelas csv em formato bruto
- 2 Glue Workflow responsável pela orquestração do Job ETL
- 3 Glue Job ETL em pyspark, responsável por fazer as transformações e converter os arquivos em parquet(formato colunar)
- 4 Bucket camada Gold, contem a pasta com os dados finalizados pelo Job ETL, disponiveis para consumo
- 5 Glue Crawler que escreve os metadados no catálogo de dados
- 6 Glue data catalog, responsável por disponibilizar os dados para o consumo no athena
- 7 Policy limitadora de acesso ao bucket para evitar modificações em pastas e deleção de arquivos, esta policy também dá acesso ao KMS para decriptação/encrptação dos dados e também limita algumas atividades do athena e do glue.
- 8 AWS LakeFormation responsável por limitar o acesso aos dados do catálogo
- 9 AWS Athena, responsável por executar as queries ad-hoc
- 10 Usuário consumidor com acesso a apenas consulta aos dados