



Kamran Ali
@aTechGuide

How to build Data Pipelines in AWS using 7 Simple services [2023 Edition]



Let's look into



Buckle Up, to build **end to end** Data Pipeline



First Question ...



How to Load Data in AWS?



1. AWS Database Migration Service

1. It can be used for **Snapshot Import** of Data or Setting Up **Continuous Replication**
2. Support both Homogeneous and Heterogeneous Data Migration
3. It offers features such as Schema Conversion, Data Validation, error Handling etc



We have a
mechanism to Pull
Data.

But ...



Where to Save data?



2. Amazon S3

1. S3 stands for Simple Storage Service
2. It is an Distributed Object store
3. It is highly available, Scalable and Secure.
4. We can build Data lake, do data back ups at Low cost



Can we build Tables
on top of Raw Data?



A Schema on Read
will be helpful.

How to do that?



3. AWS Glue Data Catalog

1. Glue Catalogue is scalable
Collection of Tables organised into databases
2. Its uniform repository where disparate systems can store and find metadata to keep track of data in data silos



How to Query the Tables?



4. Amazon Athena

1. Amazon Athena is an **Interactive Query** service to analyse data in S3
2. We can use standard SQL to run ad-hoc queries and get results in seconds.



Now, we need to

1. Clean the data
2. Join it with other tables
3. Aggregate data
4. ...



How to do compute?



5. Amazon EMR

1. Amazon EMR is a **Managed Cluster** platform that simplifies running big data frameworks, such as Apache **Hadoop** and Apache **Spark** etc.
2. We can use above frameworks and other open-source projects (e.g. Zeppelin), to process data for analytics purposes and business intelligence workloads.



We have all the
pieces but how to
run everything
Automatically and in
Sequence



How to Automate the Workflow?



6. Amazon MWAA

1. Amazon MWAA stands for Managed Workflows for Apache Airflow
3. It is a **managed** service for **Apache Airflow** that we can use to build and manage our workflows in the cloud



Also, we need to
build **Graph** on top
of aggregated
Metrics
And
Dashboard
containing those
Graphs



How to build visualisations?



7. Amazon QuickSight

1. It is a fast business analytics service to build visualisations
2. It can discover AWS data sources and enables organisations to scale to hundreds of thousands of users
3. It uses Super-fast, Parallel, In-Memory, Calculation Engine (SPICE) to deliver fast and responsive query performance



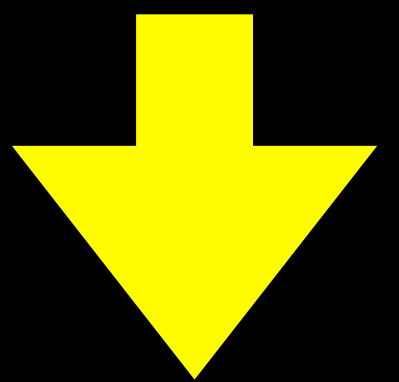
We have an End to
End Pipeline 🎉



Over to You,

If we are using
Tableau, How shall
our pipeline
change?

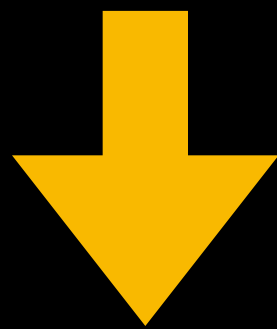
Let me know in the
Comments





Want to be a part of
Community?

Join the Discord



<https://discord.gg/Pc9ed8krYK>



Kamran Ali
@aTechGuide

That's a Wrap

1. Follow Me on LinkedIn for more valuable content - I'm Kamran Ali

2. Turn on the bell Notification in my profile

