Orchestrate Redshift ETL using AWS Glue and Step Functions

Aditya Chouhan, Khushi Soni, Nikhil Soni, Prof. Manoj Kumar Gupta

Department of CSIT

Acropolis Institute of Technology & Research, Indore, M.P., India

Abstract*: Data Analytics and Machine Learning work-streams rely on ETL for their basis. ETL cleanses and organizes data using a set of business rules to meet particular business intelligence requirements, such as monthly reporting. Still, it may also handle complex analytics to enhance back-end operations or end-user experiences. An organization's ETL is frequently used to Retrieve data from legacy systems, To improve data accuracy and reliability, clean the data. Update a target database with data*

Keywords: ETL tools, Data integration, Amazon redshift

REFERENCES

[1]. Amazon Redshift and the Case for Simpler Data Warehouses Anurag Gupta, Deepak Agarwal, Derek Tan, Jakub Kulesza, Rahul Pathak, Stefano Stefani, Vidhya Srinivasan.

[2] Daniel J. Abadi, Samuel R. Madden, and Miguel Ferreira. Integrating compression and execution in column-oriented database systems. In Proceedings of the ACM SIGMOD Conference on Management of Data, pages 671–682, 2006.

[3] https://aws.amazon.com/blogs/big-data/orchestrate-amazon-redshift-based-etl-workflows-with-aws-step-functions-and-aws-glue/

[4] https://docs.aws.amazon.com/redshift/latest/mgmt/overview.html