

KEERTHANA ARUL

Senior Data Engineer

LWD- 31st Oct

+91 9047487520 | keerthanaarul1512@gmail.com

LinkedIn URL: <https://www.linkedin.com/in/keerthana-arul-k151296>

About

Overall Experience with 4 years in Big Data Professional which includes Apache spark, Hadoop, Hive, and have an exposure in cloud technology like Azure and AWS.

Eager to expertise in latest technologies and contribute to the team success through Hard work and excellent Organizational skills.

Education

- B. TECH-EIE | MAY-2018 | SRM UNIVERSITY|8.5 CGPA
- HIGHER SECONDARY | MAR-2014|NATIONAL HIGHER SECONDARY SCHOOL|92%
- SSLC| MAR-2012|GOVT GIRLS HIGHER SECONDARY SCHOOL|91%

Skills Set

- ETL Developer
- Cloud services-AZURE, AWS
- Trained in basic Python
- Big Data concepts includes Hadoop, Hive, Spark
- Hands on Experience in Azure Databricks, Azure Data Factory, Azure datalake storage gen2, PowerBI, AWS lambda functions, AWS step functions, EMR.
- Good knowledge about Datawarehouse concepts.

Experience

L&T Infotech, Dec 2021 – current

Project: 1

Client: Marsh McLennan (Jan 2022 – current)

Role: Pyspark developer with AWS

Description: It involves extracting the sources, applying transformation and storing it to the AWS s3 after that query and performance checks. The data will be extracted from the oracle data warehouse followed by flattening of the json's, applying the transformations

finally loaded into the s3 and also stored in hive tables. The downstream will access this data from the respective hive source views accordingly.

Responsibility:

- Developed a json flattening scripts, prephysical and physical stage modules.
- Developed a dynamic DDL creation script, which will create the json query table automatically based on the columns present in the json.
- Production monitoring and bug fixing.
- Source view creation for new sources.
- Worked on the AWS lambda and step functions for creating each layer of the job work flow.

Infosys, Nov 2018 – Dec 2021

Project :1

Client: Microsoft - Workload Tagging for Microsoft Customers (May 2020 – Dec 2021)

Role: Azure data bricks, ADF Developer

Description: Extracting the customer support data to derive and analyze the total workload distributed for each agent based on the different region and incident severity.

Responsibilities:

- Developed an efficient code to tag the work done by each agent irrespective of the time constraint and certain acceptance criteria using Pyspark.
- Developed an automation script to load this incremental data into ADLS path and Archival.
- Created a final report for all the SBU workload tagged data using PowerBI tool.
- Performed all the Business validation Tests for all the sources and created the KPY queries for the same scenarios.

Project: 2

Client: Citizen's bank Account (June 2019- April 2020)

Role: ETL Developer

Description: Extracting the data from several sources applying transformation and loading this data after the query tuning and performance checks. This processed data further

moved from On-Prem to AWS S3 using the Talend tool. The final data is Published to downstream as a view.

Responsibilities:

- Created a metadata for all the sources which contains information for the Preprocessing and Processing techniques.
- Load Historical data from On-Prem to AWS S3 as a one-time load.
- Automated the ingestion Process by using Autosys tool.
- Performed the data validation between source data and target data.

Certifications:

- Infosys Global Agile Developer Certifications
- Internal Certification (Infosys) in Pyspark
- Microsoft certified on Azure Fundamentals-AZ900
- Microsoft certified on Azure Data Fundamentals-DP900

I hereby declare that the information I provided above are true to my knowledge.

Regards,

Keerthana Arul