

Indhuja N

Data Engineer

(BigData,PySpark,Cloud,ETL,RDBMS)

Email: indhujarajan96@gmail.com

Address: Bangalore, India

Phone: +91 6381181377



JOB SKILLS

- Having 6+ years of overall IT experience with Tata Consultancy Services in Application Development in **Big Data, Spark and Python**.
- 5 years of extensive working experience on Data Pipeline project for batch processing using components like **Apache Spark(Pyspark), AWS Glue, EMR, RDS, S3, Redshift, DAG, Appflow, Kinesis stream, Firehose, Lambda, Cloud Watch, IAM, CloudFormation, Athena, Cognito, DynamoDB, EC2, API Gateway, SNS, SES and Python**
- Experienced in Full LifeCycle (design, development, testing and implementation) of applications.
- Exposure in working with Spark Data Frames and optimized the SLA's.
- Proficient in developing SQL with relational databases like Postgres.
- Involved in Agile Scrum methodology that leverages the client big data platform and used version control tool Git, Bitbucket, project tracking software Jira.
- Experienced in preparing and executing Unit Test Plan and Unit Test Cases after software development.
- Having 1 year of experience as a Full Stack Developer - Java, Spring Boot and Angular5, Implemented AWS Cognito for authentication and authorization.
- Having 1 year of experience as ASP.NET Developer for Cyber Security Framework application using other tools like SQL and d3js
- Implemented Real Time Data Platform (RTDP) which aims at improving parameters like Speed, Accuracy and Availability of Data and related Reports generated within the organization using DMS,Kinesis Data Stream, Firehose, S3, Aurora, Lambda, Glue, SQS, DataSync and Microsoft Power BI for Visualization

TECH. SKILLS

- **Programming & ETL:** Python , Pyspark
- **DB & Orchestration:** PostgreSQL, MySQL, MSSQL AWS DAG (Airflow)
- **Big data & Cloud Computing:** Cloud (AWS), PySpark
- **Machine Learning & Framework:** Pandas, Python Dedupe
- **Secondary Skill Set:** Angular5, Java, Spring boot, Dot Net, HTML, JS & CSS

EXPERIENCE

Nov 2021–till

Data Engineer

PwC

Insurance and Health Domain

- Experience in building the ETL orchestration using DAG, Appflow, S3, Glue to load the curated data to Redshift from salesforce and to export the curated data into S3
- Implemented the MDIF in the orchestration to control the flow using Metadata control tables.
- Experience in using the Glue spark to curated the data from different sources after data cleansing, filtering and data transformation.
- Experience in building Data pipeline to insert the insurance related information from MS SQL to Redshift using DMS,lambda and EMR (ETL).
- Developed the lambda with Event bus trigger to insert the bulk data on daily basis to Redshift for Cost optimization

Sep 2017- Oct 2021

Data Engineer

Tata Consultancy Services

Citizen Data Bpost

- Involved in Extracting, Transforming and Loading of large sets of data and analyzed them by using Spark.
- Experienced in performing file validation using lambda , cleansing , matching the records based on the individuals from different surveys and historical data loads to get the unique individuals ids using Pyspark in AWS EMR.
- Experience in using the python dedupe (Machine learning) library for training and matching the individuals based on certain specific rules.
- Experienced in performance tuning of Spark Applications by setting correct level of Parallelism, driver and executor memory tuning.
- Experience in creating CloudFormation script for AWS Services – Kinesis Stream, Kinesis Firehose, S3, Lambda, DMS, SQS and SNS.
- Implemented Azure authentication for Power Bi dataset to perform insert and delete data in real time.
- Implemented Cognito Authentication for Kibana (Elastic Search).

- B.Tech – Electronics and Instrumentation Engineering

EDUCATION

2013-2017

PERSONAL SKILLS

- Empathic listener, influencing, **leading**, & efficient.
- Passionate to learn new skills, **problem solving**, open to new challenges.
- Interact and Provide updates to top management and **co-ordinate** across the teams.