

PUJA PRIYA

DATA ENGINEER

ABOUT :

- Working as Data Engineer to design, implement, and configure different Big Data projects by using Apache Spark, Hive, Impala, Sqoop, Oozie, Nifi and Autosys.
- Having experience in Apache Spark to implement Big Data applications.
- Involved in developing framework using Pyspark, Python, JSON
- Having experience in data ingestion using YAML through AWS services like AWS Lambda, S3.
- Having experience in Azure DevOps
- Having experience in all stages of the project including requirements gathering, designing, documenting, architecture, and development.
- Experience in using source control tools like Jira, GIT.

SKILLS :

- Big Data
- Apache Spark
- Apache Nifi
- Pyspark
- Schedulers - oozie, autosys, airflow
- Sqoop
- HIVE
- IMPALA
- DREMIO
- Python
- AWS- LAMBDA, S3, CLOUDWATCH.
- AZURE DEVOPS
- JENKINS
- GIT, JIRA

EDUCATION

BACHELOR OF TECHNOLOGY

WBUT 2017

WORK EXPERIENCE

DATA ENGINEER

TRAVELOPIA | JULY 2022 - PRESENT

Making Business with Travel Brands
1. Yachts Hotel Booking : Capturing data from stored procedure database and reading it through AWS lambda scheduled by event bridge and sending current state of hotel booking to sqs handled by DLQ and sending it to Hubspot

DATA ENGINEER

IDFC FIRST BANK | FEB 2021- JULY 2022

- CREDIT CARD COLLECTIONS /Sfdc collections
- Responsible to migrate credit card data from heterogeneous sources to Object store Datalake using Apache Nifi · Responsible to write various Pyspark applications to process data present in object store and to generate various reports. · Responsible to generate credit card reports as per the guideline provided by RBI · Involve in standardizing data present on Nutanix object store and migrate that data to cloud
- NIFI MIGRATION
- Responsible to migrate datalake flow from old DI to new DI to balance the load. Create and Configure the flow
- Tested and deployed to cloud

BIG DATA CLOUD DEVELOPER

COGNIZANT TECHNOLOGY SOLUTIONS | MARCH 2018-FEB 2021

- Data Lake development in Wildcats for Nationwide Mutual Insurance Company
- Data available from different sources like spreadsheets or csv format from Nationwide are preprocessed using python scripts, ingestion of data in GDH using YAML in S3 bucket. Validation of S3 bucket data using Python in AWS lambda.
- AbbVie Health Care Professional Commercial Data Lake Development
- Data is available from Teradata source, Data is sqooped from database to hive tables. Tables are used to make a framework using python/pyspark script via configuration files.
- State Auto an Insurance Company
- State Auto has embarked on an enterprise data lake platform to support the decision support and data provisioning to various internal and external downstream systems
- PFIZER DAVID DASHBOARD DEVELOPMENT
- Developing David Finance Dashboard and analyzing the data by different types of visualization and narratives using TIBCO Spotfire tool, oracle database, Iron Python and Java Script.