Prashant Kumar Rana

Data Engineer

My Contact

✉ pkr.prashant12@gmail.com

☎ +91-9667361369, +91-7008751017

📍 Bhubaneswar, Odisha

in https://www.linkedin.com/in/prashant-kumar-rana-875707123/

Skill

- Azure Data Factory
- Databricks,
- pyspark
- Delta Lake
- ETL Pipeline
- Azure Data Lake
- Microsft SQL Server

Languages

- English
- Hindi
- Odia

Education Background

- Trident Academy of Technology, Bhubaneswar

  Mechanical Engineering,
  Completed in 2016

Career Objective

To obtain a dynamic and challenging position to prove my skills that can provide a steady growth and Learning opportunity to accomplish the organization's goals and also expand my own knowledge for future exposure.

Extensive experience in Microsoft Azure Data Factory, AzureDatabricks,Pyspark,ETLpipeline creation,Dataflow creation, DataLake usage & Microsoft SQL server.

Experience in Azure Data Integration from different source systems.

Professional Experience

Senior Executive
Xerago, Chennai, Oct 2022-30th June 2023

- ETL pipeline creation using Azure Data Factory,
- Azure Databricks, Pyspark. Data integration from different servers to Azure
- DataLake.Work flow creation in Databricks for delta record upserting. End to end design building for Data migration in Azure.
- Designing & Development of Extraction
- Transformation Loading (ETL) methodology for supporting data transformation and processing using Azure Data Factory &SQL Server.
- Usage of pyspark in Databricks Notebook.
- Communicating with client for better understanding the requirements and make the process errorless.
- Working with Banking Sectorfor Singapore Market.

Associate Data Analyst
Vidooly media tech pvt. ltd, Aug 2018-July 2022

- End to end design building for data migration in Azure.
- ETL pipeline creation using Azure Data Factory,Azure Databricks, Pyspark & Delta Lake.Data integration from different servers to Azure DataLake. Involving into ETL Process fordata loading into DB using SQL server

- Preparing reports as per the client's requirement and create insights of the data report.
- Worked to provide different Competitor analysis of client and their growth in weekly, monthly and Yearly basis. To understand type of content around which the audience engagement is more.

Project :

Domain: - Insurance
Client: - Singapore based client
Role and Responsibility: Azure Data Engineer
Work Duration – Oct 2022 to June 2023

Project Scope:

Internal applications send the data on CSV format on daily basis in the cloud storage location. There are some data validations needed to be performed on this incoming data. Once everything is passed data should be persisted as delta table in Databricks for downstream system.
Internal Application sends CSV file in Azure data lake storage. Validation needed to apply on this follows:-
- Check for duplicate rows. If it contains duplicate rows, file need to be rejected.
- Need to validate the date format for all the data fields. If validation fails file will be rejected.
- Move all the rejected files to Reject folder.
- Move all the passed files to Passed folder. Store as Delta table.

Responsibilities:

- ETL **pipeline** creation using **Azure Data Factory, Data Transformation, Azure Databricks**, **Azure Datalake & Blob storage**, **SQL Server, Pyspark**.

- Get the schema form **SQL** server and connect it into the Databricks.

- Connect the Databricks with Landing folder**.**

- Do the validation of the file with the schema we are getting from SQL server.

- Incorporate Azure key vault to the Databricks for all the credentials.

- Based upon the file is passed or failed we move the file in to Passed or Reject folder (Using pyspark in Databricks Notebook).

- Communicating with client for better understanding the requirements and make
  the process errorless.

Project :

Domain: - Automobile
Role and Responsibility: Azure Data Engineer
Work Duration – March 2020 to July 2022

Project Scope:

Load data from different sources with different type of data (csv,excel etc). Dump
the data in to ADLS Gen2 Clean the data and create cleansed delta tables.
Create fact and dimension based on business logic and save it.

Responsibilities:
- Validate the Data between source and sink.

- Send alerts when data is wrong or Junk (Data quality check).

- Send alerts when pipeline gets failed.

- Create pipelines using **Azure Data Factory, Azure Databricks, Azure Datalake & Blob storage, SQL Server, pyspark.**