Hathija Bibi R

Data Engineer

Data Engineer with 5 years of experience in Pyspark, GCP, Microsoft Azure in developing successful ETL tools requiring high performance, scalability and short response time. Key participant and a team player in all phases of development life cycle of the projects with good interpersonal skills.

SKILLS

Programming Language:

Python

Framework:

PySpark,Hadoop,

Spark, Hive, Oozie,

Netezza

Database:

Oracle, PL/SQL

Cloud:

GCP, Azure

Versioning: Git

Education:

B.E: Computer Science & Engineering (2014-2017)

College: K.Ramakrishna Engineering College. Trichy

CGPA: 7.51

Diploma: Computer Science & Engineering (2011-2014)

College: MIET Polytechnic College. Trichy

Percentage: 89.71%

SSC (2009 - 2011)

School: St. Anne's Girls Higher Secondary schools

Percentage: 66.4%

EXPERIENCE

TATA Consultancy Service as System Engineer (June 2017 – July 2021)
The Nielsen Company (US) LLC --Ecommerce – *Rakuten is an e-commerce and retail company looking for ways to manage the large volume of sales data for data analysis. We have developed ETL tool to store and maintain the large volume of data in Azure cloud*

- Actively participated in data gathering and data processing from different data sources (Amazon, wine.com, jet.com, Rakuten)
- Ingesting data from various online sources(S3 client, Sqooping from MySql, file processing) in Big data platform
- Processed raw data from different sources based on the various business logics and algorithms.
- Enhanced the data cleanup process in HDFS and improved the overall performance of the reports/queries
- Deployed the codes in the production environment and involved in the post production activities.
- Developed various migrating scripts from OnPrem to cloud

Capgemini (July 2021– Present)

HEMA — *HEMA is a large e-commerce retail store. Here, we have developed scripts to migrate data from API platform channel to Cloud and process the data as per the business needs.*

- Developed scripts to extract the raw data from various sources.
- Involved in data cleaning and data processing to convert the raw data to desired format.
- Transformed the processed data into further processing.
- Involved in weekly and monthly job run activity.
- Actively involved in testing the developments before production release.
- Participated in the production deployments using Jenkins CI/CD pipeline.
- Monitored post production performance of the system.
- Worked on resolving bugs fixes.

Declaration: I hereby declare that the above furnished information is true to the best of my knowledge.