PRIYANKA B

priyabigdataspecialist@outlook.com
+91 84310 95319

Career Objective

A desire to work with large, complex datasets to use my skills in big data engineering, analytics, data visualization and communication to make data-driven decisions and drive business success. And also where I can learn new skills, that add value to the organization and my career.

Profile Summary

I have over 4 years of experience with big data technologies including Python, SQL, and Scala, as well as Data Bricks, Microsoft Azure, and the Hadoop ecosystem with a focus on core frameworks.

- Able to data analysis, cleaning, wrangling, performance tuning, deployment, visualization and reporting.
- Experienced in Spark ecosystem components like Spark batch Processing, Spark queries.
- Comfortable to work with Azure Resources (Databrick, Azure blob storage, Azure Data Factory, Azure Key Vault, AAD etc.).
- Good Experienced in working data warehouse system, developing data pipeline, implementing complex business logic and optimizing business logic.
- Good knowledge of AWS (Athena, Glue, S3 and Air flow).
- Experienced with Hadoop Developer, have a strong background with file distribution systems in a big-data.
- Excellent communication, organizational skills, attention to detail & the ability to meet tight deadlines.
- Source code management and Version Control system using Git and GitHub.
- Strong communication, presentation & interpersonal skills. Ability to interact with ease and professionalism.
- Good knowledge on Spark Streaming and Kafka, design and develop real-time data pipelines and streaming applications to process and analyze large volumes of data in near real-time.
- Proficient in SQL-based technologies (MySQL, Oracle DB, etc.).
- Proficient in Python and SQL, with experience using PySpark to perform data analysis and manipulation on large datasets.
- Using JIRA for Tickets, Tasks & Reports.

Programming Languages: Python, SQL, Pyspark

Hadoop Ecosystems**,** HDFS, YARN, MapReduce, Spark, Hive, Azure Databricks, Azure blob storage G2, Azure Data Factory

Key Skills and Knowledge

BigData Technologies

- Databricks
- Azure
- Hadoop Ecosystems
- Pyspark

Languages

- SQL
- Pyspark
- Python

Databases:

 SQL-based technologies (MySQL, Oracle DB, etc.).

OS: Windows

JIRA: Tickets, Tasks, Reports

GitHub: Git, GitHub, Git Bash

Work Experience 1

Company: GMS Solutions Pvt Ltd (Sep 2018 - Dec 2020)
Client: TORONTO DOMINION BANK
Role : Software Engg
Team: 12

Client offers various financial products of loans, credit cards, savings, also car, home, travel, life & health insurance which they provide online. On top of that collected data, we perform batch processing, extract data from various sources, transform it, and load it into a central data store. And then aggregating and analyzing, client use our analyzed data to improve business and marketing and advertising more relevant to the customers.

Responsibilities:

- Extracting data from Oracle to Azure SQL using ADF. Second Source Json files available at ADLS.
- Loading data from AzureSQL and Adls using Pyspark Dataframe.
- Using PySpark & Databricks For processing and Applying Analytics on data, as per client Requirement.
- Developed code and Testing & implementation will done in Data Bricks.
- Tuning & Optimizing Involved in developing the Databricks.
- Setting up and maintaining batch processing infrastructure, including scheduling and orchestration tools
- Designing and implementing data pipelines to move and transform large volumes of data.
- Ensuring data quality and integrity throughout the ETL process with optimized techniques.
- Investigatand optimize model performance. Monitoring, debugging and continuous improvement.
- Communicates findings and obstacles to team members to achieve best approach.
- Collaborating with data scientists and analysts to understand data requirements and develop data solutions.

Worked Tools/Technologies: PySpark, Oracl, DataBricks, CICD Pipe line, Github, ADLS, ADF JIRA.

PERSONAL DETAILS

DOB: 10/02/1998

Languages:
- English,
- Kannada,
- Telugu

Hobbies:

- Upgrading with new technologies,
- Listening to music,
- Chess,
- Carrom board .

Marital Status: Unmarried

PresentLocation: Bengaluru

COMPUTER SKILLS

- Power BI,
- working with various Databases,
- Big data tools and Technologies,
- MS Office Package.

Highest Education

Bsc- Tumkur University-2018

Work Experience 2

Company: Resource Pro (March 2021 -Jan 2023)
Client: Cansure Underwriting
Role : Software Engg
Team: 10

Client offers various insurance coverages to the people in Canada. Insurance contains Various coverages including Personal, Property, Cargo, Vehicle, Home Insurances, Event Liability. Extracting data from various sources, transform it, and load it into a central data store. And then aggregating and analyzing, client use our analyzed data to improve business and Renewal and Claim, Analyzing various coverages.

Responsibilities:

- Setting up and configuring a Sqoop connection to an Oracle database.
- Using Sqoop to extract data from the Oracle database and load it into Hive tables.
- Reading data from the Hive tables using PySpark and performing data processing and analytics on the data as needed.
- Configuring and scheduling tasks using Autosys to automate the execution of the data pipelines and analytics processes.
- Debugging and troubleshooting issues that arise during unit testing, including identifying and fixing any issues with the data or the code
- Tuning the performance of the PySpark code and the data pipelines to ensure that they are running efficiently.
- Developing PySpark code to meet business requirements, including implementing data processing and analytics logic as needed.

Worked Tools/Technologies: Spark and Python, AutoSys, Hadoop, Hive, Sqoop, Yarn, Git Hub, PyCharm, Notebook, Windows, Jenkins, RLM, CICD Pipe line, Agile JIRA.

`