S MD NAZEER HUSSAIN

hussainsmdnazeer@gmail.com

+91 91108 80565

# Career Objective

**Key Skills and Knowledge**

**BigData Technologies**











Databricks

Azure

Hadoop Ecosystems AWS

Pyspark

**Languages**







SQL

Pyspark

Python

**Databases:**

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

**OS**: Windows

**JIRA**: Tickets, Tasks, Reports

**GitHub:** Git, GitHub, Git

Bash

A desire to work in 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

**Impact Automation Services Private Limited**

**5.6 Yrs | Jan 2018 - Aug 2023**

I have 3.6 years of experience in Pyspark, SQL, Scala, Data Bricks, AWS, Microsoft Azure, Hadoop ecosystem and its core frameworks.

* 2 years worked as a system engineer.
* Able to data analysis, cleaning, wrangling, performance tuning, deployment, visualization and reporting.
* Experienced in Spark ecosystem components like Spark batch Processing, Spark SQL queries.
* Comfortable in work with Azure Resources (Databrick, Azure blob storage, Azure Data

Factory, Azure Key Vault, AAD etc.).

* Experienced in Spark ecosystem components like Spark batch Processing, Spark SQL queries.
* 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 withease and professionalism.

**Programming Languages:** Python, Scala, SQL

**AWS:** S3, Glue, Athena, RDS, Cloudwatch.

**Microsoft Azure:** Azure Databricks, Azure blob storage G2, Azure Data Factory, Azure Key Vault, AAD, Azure SQL.

**Hadoop Ecosystems:** HDFS, YARN, MapReduce, Spark, Hive, etc.

# Project 2

**PERSONAL DETAILS**

**Languages:**

* English,
* Hindi
* Telugu
* Urdu

**Hobbies:**

* Upgrading with new technologies,
* Listening to music,
* Carrom board .

**Marital Status:** married

**PreferredLocation:**

Bengaluru

**COMPUTER SKILLS**

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

**Client:** Banking: TD Bank **Role** : Software Engg **Team:** 10

Client offers financial products of loans, credit cards, savings also car, home, travel,

li.fe & 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 various sources such as databases, ADLS and so on and transforming that data as needed using Spark SQL, Python, or Scala code.
* Loading the transformed data into the desired destination, such as a data warehouse or a database.
* Using Databricks For processing and Applying Analytics on data, as per client Requirement.
* Developed code and Testing cum implementation will done in Data Bricks.
* Tuning & Optimizing Involved in developing the Databricks.
* Setting up and maintaining batch processing infrastructure, including scheduling and orchestration tool.
* Designing and implementing data pipelines to move and transform large volumes of data using ADF.
* Ensuring data quality and integrity throughout the ETL process with optimized techniques.
* Investigate and 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.

**Highest Education**

• B.COM, 2014, 60.6%

**Worked Tools/Technologies:** PySpark, Kafka, DataBricks, CICD Pipe line, Github, Azure Blob Storage, Power BI. JIRA.

# Project 1

**PERSONAL ABILITIES**

* Ambitious, Smart working, energetic and well Disciplined.
* Positive thinking, self- motivated and flexible.
* Team player and strong leadership skills.
* Good communication skills.
* problem solving, quick learning ability.

**Client**: Retail: WALMART Bank **Role** : Software Engg **Team:** 12

Client collects huge amount of usage information about all their products, enterprise customers and a huge global Products, data analytics is a critical component of client business. 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 marketing and advertising more relevant to the customers.

* 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
* 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, Hadoop, Hive, Sqoop, Yarn, Git Hub, PyCharm, Notebook, Windows, CICD Pipe line, Agile JIRA.

 Certifi

ICSI | CNSS Permanent Certified Network Security Specialist from ICSI, UK (International CyberSecurity Institute)

[https://www.credential.net/deac23b5-f9e8-4ac0-aa51-28facad67c95](http://www.credential.net/deac23b5-f9e8-4ac0-aa51-28facad67c95)

Q1. How conversant are you with Python/Scala on a scale 1-5

----> PYTHON(3/5),PYSPAK,SQL

Q2. Would you be interested to take a data structures and Algorithm test in any programming language of your choice

--->Yes

Q3. What Development Techniques do you adopt for your data project?

---> pyspark,data brick,agile developement

Q4. Explain in brief what measures were taken to assure data quality.

--> unit test,data typemismatch.regression test

Q5. What data Modelling approach did u adopt for enabling Business Intelligence If an

--->No

Q6. What technical gaps do you want to work on once you join our organisation?

---->Ok with any techology

cation