

**Rohit Bukan**

**Phone: (+91)-9637338798**

[**rhtbukan@gmail.com**](mailto:rohitbukan@gmail.com)

**PROFESSIONAL SUMMARY**

* Having 4 years of total experience in IT industry, out of which around 3.1 years of extensive experience in **Big Data** technologies like **Hadoop** and **Spark** Eco System.
* **C-Dac** [ **Center of Diploma in advanced computing** ] Certified Full Stack Software developer.
* Hands on knowledge in design and development of solutions for Big Data technologies such as **Hadoop**, **Hive, Sqoop, Spark and AWS**
* Good understanding of Hadoop architecture and its components such as **HDFS, MapReduce, Partitioning, Bucketing, Controlling**
* Hands-on Experience in performing analytics on structured data in the **Hive with Hive**
* **queries, Partitioning, and Bucketing.**
* Experienced in Amazon Web Services (AWS) such as **EMR, EC2, S3 and RDS.**
* Experience in developing Spark applications using RDD transformations (**Spark core**) and **DataFrame** (Spark SQL) with Spark DSL & SQL Functions.
* Extensive knowledge on serialization- using **AVRO, PARQUET ,JSON,XML& ORC** File formats.
* Involving in handling large sets of **Structured, Semi-structured and Streaming data.**
* Created Hive tables to store structured data into HDFS and processed it using **HiveQL**.
* Good knowledge **in incremental imports, partitioning and bucketing** concepts in Hive and Spark SQL needed for optimization
* Plan, develop and implement successful large-scale projects from conception to completion
* Extensive experience in working with the projects using **Agile** and **Waterfall** Software Development Life Cycles.
* Ability to be a splendid team player and communicate/interact effectively with global clients.
* Strong analytical skills, customer focused, takes responsibility for projects and drive results through execution.

**TECHNICAL SKILLS**

**Data Eco System** : Hadoop, HDFS, Hive, Sqoop, NoSQL (HBase), Spark, Amazon AWS.

**Distribution** : Cloudera

**Databases** : MS SQL Server, MySQL Hbase

**Languages** : Scala, HiveQL ,Java, Angular, C

**IDE and Build Tools** : Eclipse, IntelliJ, JIRA and Putty

**Version Control** : GITHUB,GITLAB, SVN

**Operating Systems** : Linux, Unix and Windows

**Tools**  : Eclipse,Visual Studio Code,Notepad++,Intelli

**PROFESSIONAL EXPERIENCE**

***Organization: Carelon Global Solution Dec 2022 - Till Date***

**Role: Software Engineer II.**

Carelon Global Solutions, formerly Legato, combines expertise, data, and digital capabilities to provide predictive and proactive healthcare products and healthcare services company solving the challenges to achieving full and complete care.

**Project Role:** **Data Engineer**

**Responsibilities:**

* Developed **spark** code and deployed it in EMR which will take data from **Postgres** table

apply business logic on that data as per business requirement and store that data in target table.

* Handled multiple reads and writes with Postgres database.
* Developed Scala code using **GitHub/Bitbucket** that bitbucket develop branch is integrated with **S3 Bucket** using **Bamboo pipeline**.
* Running spark-submit in **EMR** who takes the files from S3 bucket and processes the data from source table and load that processed **data** to target table.
* Reading CSV, **Parquet** files applying business logic on that data and storing data to Postgres database.
* Reading the data from Postgres table performing analytical operation on that data and storing it into S3 bucket as **fixed-width** file.
* All this pipeline **triggered** through **automation** (Time-Based trigger & Event-Based Trigger) in which we are using some AWS services like **S3,EMR,Lambda,EC2,Step Functions.**

***Organization: Quantiphi Analytics Solution PVT.LTD March 2021–Dec 2022***

**Role: Data Engineer.**

Quantiphi provides Applied AI and Big Data solutions to global Fortune 500 clients across sectors viz. CPG, Insurance, Healthcare, Retail, Media and Entertainment, Ed-tech, and Technology. Quantiphi’s solutions have helped these enterprises create tangible business value by building smarter products, achieving frictionless customer experiences, automating complex processes, and managing risks.

**Project Role:** **Spark Developer**

**Responsibilities:**

* Performed **Import and Export** of remote data from **AWS s3**
* Developed **spark** code and deployed it in EMR.
* Importing data from multiple sources such as **AWS Simple Storage Service(S3)** into Spark RDD
* Involved in delivering the resultant data to RDS and snowflake
* Scheduled necessary scripts in **EC2** to integration with other servers.
* Handled multiple reads and writes with with **Azure Datalake and Blob storage**.
* Involved in handling team for a soften project delivery.
* Wrote Spark applications for data validation, cleansing, transformation, custom aggregation
* Developed custom aggregate functions using Spark SQL and performed interactive querying.
* Implemented Spark using Scala and SparkSQL for faster testing and processing of data.
* Created and worked on Sqoop jobs with incremental load to populate **Hive** External tables.
* Created Hive tables as per requirement, internal and external tables are defined with appropriate static and **dynamic partitions**, intended for efficiency.
* Work with client/customer to create technical strategies and frameworks
* Working with other Delivery Leads to perform impact analysis on key initiatives
* Understand and execute Change and Incident management
* Work with the Project Manager in the production of Project Work Package for Production related Products

**Technologies:**  **AWS S3, IAM role,EC2 ,Scala and Spark**

-------------------------------------------------------------------------------------------------------------------

**Role: Data Engineer** .

**Technologies/Tool**: **Java,Angular,GCP,Analytics,Hadoop,Glue,RDS,S3,EC2,IAM,Warehouse, HDFS, Hive, Spark, Sqoop, Scala, AWS, Eclipse, MySQL, XML, Json**

**Responsibilities:**

* Creating S3 Buckets and Storing Complex data and performing analytical operations.
* Retrieving snowflake warehouse data and storing it in to **AWS**.
* Developed Scala code to read the data from the **Snowflake and GCP**.
* Developed Spark code in **Scala** using **Spark SQL** & Data Frames for aggregation.
* Experience in working with Spark SQL for **Azure data Processing**.
* Worked On Java & Angular Full stack development project.
* Handson Experience wit version controlling Tool like **GitHub** and **GitLab**.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**EDUCATION**

* **Center of Diploma in Advanced Computing** **[C-Dac]** February2020–February 2021
* Bachelor of Engineering (Computer Engineering) from Pune University in 2019.
* Diploma in Computer Engineering, from MSBTE in 2015.

**PERSONAL DETAILS**

* Marital status: Un-Married
* Nationality: Indian
* Date of Birth: 29 November 1995
* Permanent address: At-Post Bukanwadi Tal&Dist-Osmanabad,Maharaadhtra -413509