

# VAIJANATH MADLIKAR

## DATA ENGINEER

+91 9036679793

vaijanath.m79793@gmail.com

Bangalore



### ABOUT ME

Overall, around 4+ years of professional experience in Information Technology and around 3 years of expertise in BIGDATA using Hadoop and Spark framework and Analysis, Design, Development, Testing, Documentation, Deployment, and Integration using SQL and Big Data technologies.



### SUMMARY

- Expertise in using major components of Hadoop ecosystem components like HDFS, YARN, MapReduce, Hive, Impala, Sqoop, HBase, Spark, Spark SQL, Kafka, Spark Streaming, Hue.
- Good understanding of distributed systems, HDFS architecture, Internal working details of MapReduce and Spark processing frameworks
- Having good knowledge in writing MapReduce jobs through Hive and Sqoop.
- Hands on Experience in developing Spark applications using PySpark Data Frame, RDD, Spark SQL
- Extensive knowledge in writing Hadoop jobs for data analysis as per the business requirement
- Worked on structured and semi structured data storage formats such as Parquet, ORC, CSV, JSON
- Data extraction, join operations, writing custom UDF's as required and having good experience in optimizing Hive Queries.
- Proficient with Spark Core, Spark SQL for processing and transforming complex data using in-memory computing capabilities written in Scala. Worked with Spark to improve efficiency of existing algorithms using Spark Context, Spark SQL, Data Frame, Pair RDD's and Spark YARN.
- Experienced in building and deploying Spark applications on Hortonworks Data Platform and AWS EMR
- Active team player with excellent interpersonal skills, keen learner with self-commitment & innovation.
- Proven ability to manage all stages of project development Strong Problem Solving and Analytical skills and abilities to make Balanced and Independent Decisions.

### PROJECT 1:

**TOOLS:** Sqoop, Spark, MySQL, SQL, Python, AWS S3

Description: The project was to create enterprise application which ingest the data from RDBMS system to on premise and cloud big data warehouses which includes snowflake, hive using spark and Sqoop. Insurance related data was stored in the MySQL data warehouse which was used for reporting purpose.



### Technical Skills

**Programming:** Python, Scala, Java, PySpark

**Big Data Technologies:** Hadoop ecosystem, HDFS, Hive, Spark, Sqoop.

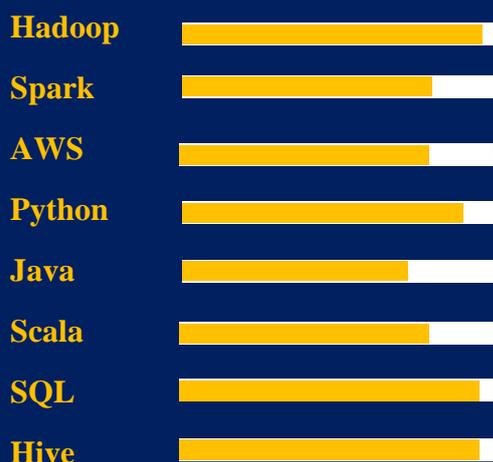
**Cloud Platforms /Services:** AWS Lambda, EMR, EC2, S3, Glue, Athena, IAM.

**Frameworks and Tools:** Postman, VScode, IntelliJ, Eclipse, JIRA, GIT, Jenkins, Putty, WinSCP, SQL Developer, MSOffice (Word, Excel, PowerPoint, Visio).

**Database:** Oracle, SQL Server, MySQL.



### Skills



### Work Experience

**Working from Accenture as Software Developer from November 2018 to till date**

**TASK HANDLED:**

- Used data ingestion SQOOP for loading and incremental updating of data from MySQL to hive
- Created big data application using Pyspark which loads the data from AWS S3 and ingestion into snowflake.
- Data cleansing with the use of spark framework (Spark SQL) and storing the results into hive for analysis.
- Imported data from AWS S3 into spark RDD and performed transformations and actions on RDDs.
- Experienced in performance tuning of Spark Applications for setting right Batch Interval time, correct level of Parallelism and memory tuning.
- Created schema in Hive with performance optimization using bucketing & partitioning
- Involved in the end-to-end deployment process.
- Written Hive queries for transformation and stored result back to hive table.
- Experience with different data formats like JSON, AVRO, PARQUET, and ORC and compressions like snappy.

**PROJECT 2:**

**TOOLS: Sqoop, Spark, SQL, Java, HBase**

**TASK HANDLED:**

- Took part in several discussion with Onshore SMEs for logic understanding and sharing the same with other team members
- Worked with the Informatica team to understand the data loading logic and implemented created the data pipeline in Spark using Java.
- Created a macro for DDL creation of all the HIVE tables that were used in Staging, L1 and other layers.
- Developed the logic for extraction of data from Staging to L1 layer by using Spark and Java for both fresh and incremental records.
- Took care of most of the transformation of data in L1 layer as per the business requirements.
- Used property file to handle SQL, to improve the readability of the code.
- Used YAML file for Data quality checks.
- Developed script on Scala using Spark for handling incremental data.
- Created Hive table on top of HBase by using HBase storage handlers and appropriate column mappings.
- Used file formats like ORC in HIVE and parquet with Spark.
- Took part in several agile cadences like: Scrum call, retro sessions, backlog grooming, and Sprint planning and unity hours.
- Prepared design documents in Visio
- Have worked with SIT and UAT teams for helping them understand business logics and scenarios.