**SALMAN SAYYAD**

**+91-751-756-4738|python831547@gmail.com|**[**linkedin.com/in/salman-sayyad-88b016a**](https://www.linkedin.com/in/salman-sayyad-0b88b016a)

Big Data Engineer with 3+ year of experience leveraging Big Data Transformations, Aggregations, Statistical Modelling, Data Mining, Data Warehousing, Machine Learning, deep learning algorithms to solve challenging business problems on time series , computer vision and Natural language processing .

**Professional Experience**

**Ample SofTech -** Data Engineer.

**Client -** Swiggy : Apr 2018 - Present

**Project -** E-commerce

Swiggy is India's largest and highest-valued online food ordering and delivery platform. Logs generated by millions of users through Swiggy website and mobile application saved in AWS S3 bucket through Kafka server. Project involves transformation, aggregation of log data using AWS EMR, AWS Glue, AWS Athena, Apache Spark, Scala, PySpark and implementation of advance data analytics and data warehousing, machine learning, deep learning after processing data packets in big data pipeline.

**Technology**

**Spark 2.7.x, AWS S3, EMR, Glue, Athena, Scala 2.12.x, Kafka, Sqoop, SQL, Hive, HBase, RedShift, PySpark, Python 3.7, Pandas, Numpy, Intellij, Jupyter Notebooks, Hadoop 2.x .**

* Batch processing of data from AWS S3/Hadoop using Apache Spark 3.0.x., Scala 2.12.0 on AWS EMR cluster.
* Stored meta data in AWS Glue catalogue from S3 using Glue crawlers and used AWS Athena to create tables from Glue metadata.
* Used Athena to create Hive tables from S3 and Glue metadata, created reports in QuickSight.
* Created hive schemas using performance techniques like partitioning and bucketing.
* Used Linux shell scripting to create and execute hive tables and queries.
* Performed ETL, Data wrangling, data analysis using Scala, PySpark, AWS Glue, AWS Athena.
* Experience with Kafka, Flume and Sqoop.
* Developed Spark code using Scala and Spark-SQL for faster processing of data.
* Developed Scala scripts, UDF's using both Data frames/SQL and RDD/MapReduce in Spark 3.0.x for Data Aggregation, queries and writing data back into RDBMS through Sqoop.
* Involved in complete end to end code deployment process in Production.
* Worked on back-end using Scala 2.12.0 and Spark 2.7.0 to perform several aggregation logics.
* Experience in working with NoSQL databases like Hive 2.1.1, HBase and Redshift.
* Excellent analytical, Interpersonal and Communication skills, fast learner, hardworking and good team player
* Very good understanding of Machine Learning and Deep learning algorithms like Linear regression, Logistic regression, SVM, KNN, Decision Tree, Random Forest, Backword propagation etc.

**Infosys BPM -** Process Executive /Sales Reporting Executive

**Client -** Chobani : May 2016 – Mar 2018

**Project -** E-commerce

Chobani is an American food company specializing in strained yogurt. Project involves analysis and reporting of daily sells transaction data using SQL Server and Power BI.

**Technology**

**Power BI, Sql, Microsoft SQL Server, SSIS**

* Imported data from SQL Server DB, Azure SQL Data Warehouse to Power BI to generate reports.
* Created Dax Queries to generated computed columns in Power BI.
* Generated computed tables in Power BI by using DAX.
* Involved in creating new stored procedures and optimizing existing queries and stored procedures.

**Academic Qualification:**

**Dr. Babasaheb Ambedkar Marathwada University, Aurangabad. *B.Sc. Computer Science* Aug-2014**