

VIJAY SANJAY NALE

Big Data Developer

Email : vsnale.04@gmail.com
Mob No: 8208078704
[LinkedIn](#)
[Github](#)

CAREER OBJECTIVE

3.6+ years of overall IT experience, which includes relevant experience of 2 years in Big Data in designing, building, integrating, Testing and Maintenance end-to-end data pipelines.

EDUCATION

Master of Computer Application(C.B.S.E Board) (July '16 - Apr '19)
Shivaji University, Kolhapur : 8.74 CGPA

KEY SKILLS

Programming Languages	Python, Scala
Hadoop Platform	Apache Hadoop, Cloudera Distributed Environment
Hadoop Ecosystem/ Framework	Sqoop, Hive, Kafka, Airflow, HBase, Apache Spark (Spark Streaming, Spark SQL)
Database	SQL, PostgreSQL, Mysql
Tools	Pycharm, Scala IDE, WinSCP, Putty

PROFESSIONAL EXPERIENCE

- A) Associate - Bajaj Finserv Ltd., Pune (Sep '21 – Present)
 - Experienced with the tools in **Hadoop Ecosystem included Hive, HDFS, Map Reduce, Sqoop, Spark.**
 - Excellent with the tools in Hadoop Ecosystem including **Resource Manager, Node Manager, Name Node, Data Node** and Map Reduce paradigm.
 - Worked on **Hive Queries** for creating and querying Hive tables to retrieve useful **analytical information.**
 - Used **Sqoop to import large data** from traditional **RDMS to HDFS** and also handled incremental Sqoop.
 - Created **partitions, buckets** in **Hive**, Handled the **integration of Spark with Hive.**
 - Optimized the existing pipelines** which resulted in the reduction of execution time **by 10 times.**
 - Experience of AWS component like **S3, EC2, EMR, AWS Athena, Crawler, Redshift, Glue.**
 - Developed **Complex SQL queries** to retrieve data from the database as well as for performance tuning.
 - Used **Airflow** to automate jobs.
 - Experience in developing spark application using **spark-SQL** in **databricks** for data extraction, transformation and aggregation from multiple file format for analyzing & transforming the data.
 - Used **Spark-SQL** to process the data and to run on **Spark engine.**
- B) Data Science Intern - iNeuron.ai (Dec '20 – Mar '21)
 - Involved in requirement gathering and architecture design.
 - Developing statistical models for various predictive methods such as forecasting, classification, clustering and regression.
 - Involved in parameter tuning process for optimal model hyperparameters.
- C) Data Analytics - Acceline Digital Media Pvt Ltd (Oct '19 – Sep '20)
 - Excellent SQL skills for daily testing of data using various SQL queries.
 - Reviewed analyze and implement necessary changes in appropriate areas to enhance and improve existing system.
 - Perform DAX queries using Power BI tools.
 - Designed, developed and tested various Power BI visualizations for dashboard and ad-hoc reporting solutions by connecting from different data sources and databases.
- D) Data Science - Innodatatics (Mar '19 – Sep '19)
 - Involved in Data Pre-processing Techniques such as data cleaning, visualizing, EDA, finding outliers.
 - Machine Learning Algorithms Evaluation, Model Deployment process once signoff is in place.

PROJECT

❖ Analyze Insurance Data.

The theme of project is to **perform the analysis on huge dataset** which are generated by websites, app and other sources like employee portal etc. In this Application-Quotes level data get generated on daily basis. **Business look for number** so here we are **analyzing multiple things** like **how many payment success and their gross premium** are made by current month and also by each Category and from which UTM source.

- Daily **3 million data** records are processed, data is received multiple data producers.
- Importing data using **Sqoop into hive** and HDFS from existing SQL server (like Application-Quote Level, Order data, Customer data, Product data etc.).
- Write Script for processing data through **Spark Core and Spark SQL** to generate business report.
- Technology Involved: Python, PySpark, Hive, Sqoop, Airflow.

❖ Credit Card Fraud Detection System.

With the increasing digitalization and online transactions, it becomes ever so important for the **credit card to be able to recognize "genuine" and "fraudulent" transactions** in order to provide their customers with a more secure and a seamless experience. **Detect fraudulent transactions at the shortest possible time** (Since the transactions are happening in real time, timing constraints play a very important role).

- Efficiently **transfer data** (card_member, card_transactions, member_score etc.) from **databases to HDFS**.
- Resultant data get stored into **Hive-Hbase** table performs **fast querying** and displays records which is helpful for the support team to handle end customer queries on a daily basis.
- Technology Used: Python, Spark, Hive, Hbase, AWS RDS, Sqoop, Airflow.

Certificate

- **Machine learning Master Certification** from iNeuron.ai
- **Big Data Engineering** from Trendytech.