




ANUGRAH VISHAL

Big Data Developer

Pune, India 
+91 – 7905687340 / 8960422302 
anugrah.vishal1@gmail.com 



PROFESSIONAL SUMMARY

- 3 years of experience as a Spark/Hadoop Developer using data handling and transformation in Hadoop and Spark Ecosystem.
- In-depth understanding of Spark Architecture including Spark Core, Spark SQL, Dataframes including spark streaming.
- Implemented Spark using Scala and utilizing Dataframes and Spark SQL API for faster processing of data.
- Experience in migrating large sets of data using Sqoop, from Oracle databases to HDFS and vice-versa.
- Performed Data Analysis on large datasets using Hive Queries as per the client's requirement.
- Motivated team player with excellent communication, interpersonal, analytical and problem-solving skills.



SKILLS

- Data Integration: SQOOP, FLUME.
- Data Storage: HDFS, ORACLE, MS SQL.
- Data processing: PIG, Hive, Spark over Scala and Pyspark(RDD, DATAFRAME).
- Workflow automation: OOZIE, Airflow.
- Platform/OS: Linux, Windows.
- Software Packages: Eclipse, WinSCP, PuTTY.



EXPERIENCE

Senior System Engineer | Infosys Limited 03/2017 – Present

Technologies: Hadoop-Hive | Spark-RDD, Data Frame

Domain: Retail

Cluster Configuration:

- No. of Nodes: 10 nodes (64GB RAM each)
- Cloud: On-Premise
- Storage - HDFS (17 TB)

Roles and Responsibilities:

- Import data using Sqoop into HDFS(Datalake) from Oracle database.
- Support code/design analysis, strategy development and project planning.
- Create Schema RDD and load it into Hive Tables and handled Structured data using SparkSQL.
- Gather the business requirement from client and interpret their requirement to analyze the problem and get to a solution in terms of data engineering.
- Involve in creating Hive Partitioned tables into bucket, loading data into those tables including writing Hive queries.
- Performed advanced procedures like text analytics and processing, using the in-memory computing capacities of Spark using Scala and pyspark.
- Implemented Spark using Scala/PySpark and SparkSQL for faster processing of data.
- Performed Transformation and Action on RDDs as per the client requirement.
- Create reports for the BI team after querying data using Hive and Spark.
- Written various functions in Scala and Pyspark , which have automated the daily processes.



EDUCATION

- **Bachelors of technology**

Mechanical Engineering from Pranveer Singh Institute of Technology, Kanpur
72%

08/2012 – 06/2016

- **Intermediate**

St. Joseph's Convent School, Kanpur
79.2%

04/2011 – 05/2012

- **High School**

St. Joseph's Convent School, Kanpur
8.3 CGPA

04/2009 – 05/2010



INTERESTS

- Hive
- Hadoop
- Spark
- Scala
- Python

- PIG



LANGUAGES

- English
- Hindi



DECLARATION

I hereby affirm that the information in this document is accurate and true to the best of my knowledge.

(Anugrah Vishal)