
Professional Summary:

- Having 4.1 years of professional experience in SDLC(SoftwareDevelopmentLifeCycle), analysis, design and development of **Hadoop, Spark, ETL, Datawarehouse** and Web applications and 2 years work experience with **GCP**.
- Bigdata developer with 4 years of working experience on Bigdata components including **Hadoop, PySpark, SparkSQL, GCP, Hadoop Eco-System HDFS, MapReduce, Hive, Sqoop, Impala, Python, Scala, Kafka, Airflow**.
- Proficient in designing, developing and implementing scalable and efficient data processing pipelines using tools such as Hive, Spark, Pyspark and Sparksql
- Extensively worked on ETL design and development using PySpark and SQL development using Oracle.
- Worked on end-to-end application-level performance tuning re-designing data models, choosing proper distribution keys re-write Green plum code.
- Experience on designing and implementing complete end-to-end **Big Data** projects.
- Expertise in handling **Airflow** for job deployment and scheduling.
- Expertise in **ETL**: Extract, Transform and Load from various data sources to target.
- Experience in Datawarehouse/datavault2.0 concepts and implementation.
- Good exposure on **DB2, NoSQL**databaseslikeHBase, DynamoDB.
- Experience in automating the project processes using **Unix shell/Batch/Python** scripting.
- Experience with front end and backend web development technologies.
- Experience in Agile methodology.
- Proficiency at various levels of application development i.e., implementation and documentation.
- Effective team player with excellent problem-solving abilities, a willing and quick learner with strong technical and communication skills.

Education:

- Completed Bachelor of Engineering from Anna University, Chennai, Tamil Nadu (2020).

Technical Summary:

Languages	Python, SQL, scala
Applications	Sales, Finance, Banking
Server Scripting	Unix, Shell, Windows Batch, Python
Frameworks	jQuery, Hadoop, Spark

Web Technologies	HTML5, XML, JSON
Database	Oracle, MySQL
Reporting and Testing	Qlik sense
Versioning	SVN, GitHub
Operating Systems	Windows, Linux
Methodology	Agile, Object-Oriented Analysis/ Design, ETL

Professional Experience:

- Worked as a Consultant in **Capgemini Technology Services India Limited, Bangalore** from mar'2020 to Apr'2024.

Experience Summary:

Data Engineer | Capgemini | Sales

Dec 2022 - Apr 2024

Responsibilities:

- Designed and implemented a large-scale sales data processing and analytics solution for a multinational sales company.
- Collaborated with stakeholders to gather requirements and understand the business objectives for sales data analysis.
- Utilized Hadoop ecosystem tools such as Hive, MapReduce, Yarn and Spark to process and analyze large volumes of sales data.
- Developed and optimized data pipelines using Pyspark and SparkSql for efficient data ingestion, transformation and analysis.
- Integrated Kafka for real-time data streaming, enabling the capture and processing of sales data as it arrives.
- Leveraged Airflow for scheduling and orchestration of data workflows, ensuring timely execution and monitoring of sales data processing jobs.
- Worked with Oracle databases for storing and retrieving the sales data, ensuring data integrity and reliability.
- Utilized GCP's BigQuery, cloud storage and data flow for scalable storage and processing of sales data.
- Developed sales data models and implemented advanced analytics algorithms using Spark's Machine learning capabilities.
- Collaborated with the data visualization team to create interactive dashboards and reports, providing insights into sales performance, trends and forecasts.
- Conducted performance tuning and optimization of data processing jobs, reducing processing time and improving overall system efficiency.
- Implemented data quality checks and monitoring mechanisms to ensure the accuracy and consistency of sales data.
- Worked closely with the sales team to provide ad-hoc data analysis and reporting support for decision making.

Environment: Hadoop, HDFS, Hive, PySpark, SparkSQL, Kafka, Oracle, Airflow, GCP, BigQuery, Qlik.

Consultant | Capgemini | Oil & Gas

Mar 2021 - Nov 2022

Responsibilities:

- Working in Agile Methodology, involved in designing/ architecture and picking the stories. Then discussing the requirements directly with the product owners.
- Involved in understanding the requirements and estimating the efforts required.

- Importing and exporting data into HDFS from Relational databases and vice versa using Sqoop.
- Created partitions, bucketing across state in Hive to handle structured data.
- Involved in moving all log files generated from various sources to HDFS for further processing through Kafka, Spark.
- Exported the analyzed data to the relational databases using Sqoop for visualization and to generate reports for the BI team.
- Involved in converting Hive/SQL queries into Spark transformations using SparkRDDs, PySpark and have a good experience in using Spark-Shell.
- Evaluated Oozie for work flow orchestration in the automation of MapReduce jobs, Pig and Hive jobs.
- Captured the data logs from webserver into HDFS using Flume for analysis.
- Designed and developed various ETL and Frame work jobs to crunch and load the history data of Oil& Gas and Finance data.
- Hands on experience in the coding automation scripts using Unix-shell/python for downloading files for ingestion set-up and archival process.
- Designed complexSQL scripts. Played a key role in data modelling datawarehouse for reporting needs.
- Improve query performance by following performance enhancement tips and database best practices.
- Lead and managed a team of 6members technically throughout the development and redesign process.

Environment: Hadoop, HDFS, Hive, PySpark, Scala, SparkSQL, Kafka, BatchScripting, Qlik.

Software Engineer | Capgemini | Multi National Bank

mar2020-Feb2021

Responsibilities:

- Responsible for loading data from Oracle to HDFS.
- Responsible for developing the UDF's.
- Understanding the client requirements and developing the code and enhancements to existing code to meet the requirement.
- Responsible in migration of project from On-Prem to Cloud, This included end to end changes to runto existing pipeline in cloud without changing functional aspects.
- Created Big Query tables and hive tables for all the important stages to storedata.
- Implemented Ad hoc jobs for exporting final transformed data to SQL server.

Environment: Hadoop,HDFS,Hive,Sqoop, Spark,Scala,ETL,Oracle,Talend6.1/Alteryx, Shellscripting.

Personal Details:

Father's Name : N. VenkataRamana
Date Of Birth : 01/01/1999
Nationality : Indian
Gender : Male
Languages Known : English, Telugu and Tamil

This is to declare that all the above entries are true to the best of my knowledge.

Bangalore

(N V Phaneendra)