
Professional Summary:

- Having 5.6 years of professional experience in SDLC(SoftwareDevelopmentLifeCycle), analysis, design and development of **Hadoop, Spark,ETL,Datawarehouse** and Web applications and 4 years work experience with **GCP**.
- Bigdata developer with 5 years of working experience on Bigdata components including **Hadoop, PySpark, SparkSQL,GCP,HadoopEco-SystemHDFS,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 in designing and implementing complete end-to-end **BigData** projects.
- Experience working in **Binary Search, Quicksort, Merge Sort and Radix sort**.
- Experience working in data structures such as **Array, Struct, LinkedList, Hash Map**
- 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/datavault 2.0 concepts and implementation.
- Good exposure on **DB2, NoSQL** databases like HBase, DynamoDB.
- Experience in automating the project processes using **Unix shell/Batch/Python** scripting.
- Experience with frontend 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 Degree from SKU University, Anantapur (2018).

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
Cloud	GCP

Professional Experience:

- Worked as a Data Engineer in Likedin service private Limited ,**Bangalore** from March 2023 to till date.
- Worked as a Quality Engineer in LTIMindtree Limited, Hyderabad from Jan 2022 to Feb 2023.
- Worked as a Associate Engineer in Adrola software technologies India Private Limited, Banglore from Nov 2019 to Dec 2021

Experience Summary:

Data Engineer | Sales

Mar 2023-Till Now

Responsibilities:

Designed and Implemented a Scalable Data Architecture: Spearheaded the design and implementation of a large-scale, high-performance sales data processing and analytics solution, tailored for a global sales organization. Ensured the system could efficiently handle vast datasets and provide insights on sales trends, performance, and forecasting.

🔗 **Stakeholder Collaboration & Requirement Gathering:** Collaborated closely with business stakeholders and sales leadership to gather detailed requirements and align project objectives with business goals, ensuring that the data solution met organizational needs and provided actionable insights for decision-making.

🔗 **Big Data Technologies Implementation:** Led the adoption and integration of the Hadoop ecosystem (Hive, MapReduce, Yarn, Spark) to process and analyze large volumes of sales data in a distributed computing environment, enhancing the scalability and efficiency of the data processing pipeline.

🔗 **Development of Data Pipelines Using Spark & PySpark:** Designed and developed efficient, high-throughput data pipelines using PySpark and SparkSQL, optimizing data ingestion, transformation, and analysis processes to support fast-paced decision-making and reporting requirements.

🔗 **Real-Time Data Streaming Integration with Kafka:** Integrated Kafka for real-time data streaming, allowing sales data to be captured and processed as it arrived, significantly improving the responsiveness and timeliness of reporting and analytics.

🔗 **Data Workflow Orchestration with Airflow:** Implemented Apache Airflow for scheduling, orchestrating, and monitoring complex data workflows, ensuring that sales data processing tasks were executed reliably and on time, minimizing delays and bottlenecks in reporting.

🔗 **Database Management & Data Integrity:** Worked extensively with Oracle databases to store and retrieve sales data, implementing robust solutions for ensuring data integrity, availability, and reliability in a high-transaction environment.

☐ **Cloud Solutions for Scalable Storage & Processing (GCP):** Utilized Google Cloud Platform (GCP) services, including BigQuery, Cloud Storage, and Dataflow, to facilitate scalable and cost-effective storage and processing,

and querying of large datasets, enabling faster insights and data processing at scale.

- 🔗 **Advanced Analytics & Machine Learning Models with Spark:** Designed and implemented advanced sales data models leveraging Spark's machine learning capabilities, creating algorithms that predicted sales trends, identified key performance drivers, and provided predictive insights for strategic decision-making.
- 🔗 **Collaborative Reporting & Data Visualization:** Partnered with the data visualization team to develop interactive dashboards and reports using tools like Tableau and Looker, providing the sales leadership with real-time access to performance metrics, trend analysis, and forecasting data.
- 🔗 **Performance Tuning & Job Optimization:** Conducted performance tuning and optimization of Spark jobs and data processing workflows, achieving significant reductions in data processing times and improving the overall efficiency of the data pipeline.
- 🔗 **Data Quality Assurance & Monitoring:** Implemented automated data quality checks and monitoring mechanisms to ensure the accuracy, consistency, and integrity of sales data throughout the ETL pipeline. Proactively identified and addressed data anomalies, preventing potential issues in downstream reporting and analysis.
- 🔗 **Ad-Hoc Analysis & Decision Support:** Worked closely with the sales team to provide on-demand data analysis, reporting, and insights to support critical business decisions, ensuring that the team had accurate, timely, and actionable data at their fingertips.
- 🔗 **Continuous Improvement & Innovation:** Regularly assessed emerging technologies and methodologies to enhance the sales data processing infrastructure, ensuring that the solution remained cutting-edge and capable of scaling to meet future business needs.

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

Quality Engineer | Oil & Gas

Jan 2022-Feb 2023

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 Spark RDDs, PySpark and have a good experience in using Spark-Shell.
- Evaluated Oozie for workflow orchestration in the automation of MapReduce jobs, Pig and Hive jobs.
- Captured the data logs from web server into HDFS using Flume for analysis.
- Designed and developed various ETL and Framework 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 complex SQL scripts. Played a key role in data modelling data warehouse for reporting needs.
- Improve query performance by following performance enhancement tips and database best practices.

- Lead and managed a team of 6 members technically throughout the development and redesign process.

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

Associate Engineer | Multi National Bank Nov 2019-Dec 2021

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 existing pipeline in cloud without changing functional aspects.
- Created BigQuery tables and Hive tables for all the important stages to store data.
- Implemented Adhoc jobs for exporting final transformed data to SQL server.

Environment: Hadoop, HDFS, Hive, Sqoop, Spark, Scala, ETL, Oracle, Talend 6.1/Alteryx, Shell scripting.

Personal Details:

Father's Name : Nagireddy S
Date Of Birth : 12/06/1997
Nationality : Indian
Gender : Male
Languages Known : English, Telugu

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

Bangalore

(Yathesh Kumar)