

Sai Kumar Gunda

940.331.4745 gundasaikumar95@gmail.com Frisco, TX

PROFESSIONAL SUMMARY:

- Big Data Developer with 5 years of IT experience in a variety of industries as well as practical knowledge of Teradata, Hadoop, Hive, Spark, Python, SQL, and HQL.
- Experience working on Oracle Database and Oracle EBS and cloud Applications.
- Hands-on Experience on the ETL tool named K2 VIEW.
- Hands-on expertise with a variety of data science programming languages, as well as excellent proficiency with structured, semi-structured, and unstructured data.
- Experience on Python code base related to Apache spark performing the actions, transformations on RDDs, Data Frames & Datasets using SparkSQL and Spark streaming contexts.
- Develops logical and physical data flow models for ETL applications.
- Provide System and Application Administration for Informatica PowerCenter and Power Exchange
- Designs, develops, automates, and supports complex applications to extract, transform, and load data.
- Identifies and manages interfaces, service levels, standards and configurations.
- Created scalable, effective data pipeline processes to manage data intake, cleansing, transformation, integration, and validation necessary to give analysts access to ready-made data sets.
- Analyzed how to leverage Spark Context, PySpark, SparkSQL, DataFrames, Spark YARN, Resource Manager, Memory Tuning, Data Serialization, and Memory management to improve the efficiency of the current Hadoop techniques.
- Familiar with Data Science project life cycle, including Data Acquisition, Management, Data Cleansing, Data Manipulation, Standardization, Visualization, Feature Engineering, Modeling, Evaluation, Optimization, Testing and Deployment.

TECHNICAL SKILLS:

Big Data Technologies	Oracle SQL, HDFS, Hadoop, Hive, Sqoop, PySpark, AWS, EC2, EMR, Athena, Airflow, SQL, YARN
Database	Oracle 11g/12c, SQL Server
Datawarehouse Technologies	Teradata, Redshift, Snowflake
NoSQL Databases	HBase, MongoDB, Cassandra
Programming Languages	Python, Java, C, C++, JavaScript
Data Science	Machine Learning, Neural Networks, NLP, Tensor flow, Pandas, Sklearn
Operating System	Windows, LINUX

PROFESSIONAL EXPERIENCE:

Data Engineer | Jan 2023- May 2023 | Irving, Texas

Client: Verizon

Role: Data Engineer

Environment: Hive, PySpark, Cloudera, Jira, Hive, Cloudera, Python, Azure

Responsibilities:

- Requirement gathering, business logic alignment/development and analysis.
- Build possible solutions and get signing off from Business stakeholders.
- Ingesting the needed data from Mainframe and Microsoft SQL Server 2008 in to HDFS using Sqoop and importing various formats of flat files in to HDFS.
- Developed various shell/Python/PySpark/Hive scripts to perform ETL, stale table cleanup, Table compaction to tackle small file issues, Purge older log files.
- Enhanced and Automated data pipeline monitoring and code deployment process.
- Generated final reporting data using Power BI for testing by connecting to the corresponding Hive tables using Hive ODBC connector.
- Worked on hive and Spark Performance techniques to improve the run time and compression techniques.
- Implemented data cleansing and custom groupings as required by the analytics team.
- Created data flow and design documents.
- Provide ongoing support for BI, Analytics data pipeline.

Capgemini | Data Engineer | Dec 2018 – Dec 2021 | Mumbai, India

Client: Heathrow Airport

Role: Data Engineer

Environment: Oracle Database, Azure, Cloudera, Bit Bucket, Hadoop, Hive, PySpark , Autosys, Jenkins, Sqoop, HQL, Unix, Python, Shell scripting, SQL, Git.

Responsibilities:

- Using Hadoop, Hive, and PySpark, created an end-to-end ETL pipeline for data processing for downstream consumers.
- Participated in the Teradata to Hadoop ETL process migration.
- Developed solid skills in performance tuning and debugging for Spark applications and Hive scripts to achieve optimum performance.
- Knowledge of the methodologies for partitioning, bucketing, and performing joins on hive tables, as well as experience establishing hive external and controlled tables.
- Utilized caching, effective, and efficient joins, broadcast variables, and other data processing tools in Spark.
- Used Sqoop to import data from Teradata to Hive tables to perform data validation.
- Experience in Monitoring and Optimizing the Hive and Spark jobs to improve the run time of the data pipelines.

- Migrated the code from Dev to Test and Test to Prod. Wrote the migration documentation in detail for system compatibility, object and parameter files for smooth transfer of code into different environments.
- To load data from staging to the target tables, bteq scripts were created.
- Participated in Bit Bucket and Jenkins-based continuous integration and application deployment.
- Created a framework for rejection to filter the data in accordance with the business standards.
- Data modelling and table design expertise.
- By generating the Jil Files, Autosys was used to schedule the Data Pipelines.
mostly in charge of optimizing long-running spark applications, creating unique spark udfs, resolving issues, etc.
- Primarily responsible for fine-tuning long running spark applications, writing custom spark udfs, troubleshooting failures etc.,
- Utilized caching, effective, and efficient joins, broadcast variables, and other data processing tools in Spark.
- Participate actively in meetings, code reviews, and technical problem-solving sessions.
- Participated in sprint planning and worked in an agile setting.

WSD | Junior Data Engineer | December 2017 - Nov 2018 | Hyderabad, Telangana, India.

- Participated in creating a standardized data structure now used by over 50 telecommunication companies.
- Designed slowly changing dimensions type1 and type2 data loads, increasing speed by 20%.
- Extracted, transformed, and loaded data from various sources.
- Cleaned and normalised data ensured data quality and ran daily data checks.
- Design, develop and maintain ETL data pipelines and architectures.
- Troubleshoot data-related issues and ensure the accuracy and integrity of data.
- Develop data-driven recommendations and create reports and visualisations to communicate insights.
- Analyse data, develop data models and algorithms and identify patterns and trends in data.

EDUCATION

M.S in Computer Science	Southern Arkansas University (3.7 GPA)	Jan2022- May 2023
B.Tech in Computer Science and Engineering	JNTU Hyderabad, INDIA.	Oct 2013 - June 2017