



VARUN GANGASANI

varunkumar.gdsk@gmail.com | +1-972-261-4466 | [LinkedIn](#)

Objective:

An accomplished and detail-oriented Senior Data Engineer with a strong background in building scalable data solutions on leading cloud and Big Data platforms to support robust data architectures, integration frameworks, data processing/ETL pipelines. Seeking a challenging role to leverage expertise in designing and implementing innovative Data Engineering solutions on Cloud platforms, foster team collaboration and create a measurable impact in the success of a dynamic organization.

Certifications:

Databricks: [Data Engineer Associate](#) | May 2024 – May 2026

Microsoft Certified: [Azure Data Engineer Associate](#) | Apr 2024 – Apr 2025

Skills:

Programming Languages: Python, Scala and Java

Scripting: Bash and PowerShell

Cloud Platforms: Azure and AWS

Databases: Oracle, SQL Server, HBase, Snowflake, Cassandra, Cosmos DB

Devops: Azure Devops, Jenkins, Terraform, GitHub, Chef, Puppet and SCM

BI Tools: Tableau, Qlik View, Qlik Sense, Power BI, OBIEE and Hydra, Grafana

Tech Suite: Cloudera Hadoop, Apache Spark, , PySpark, SQL, T-SQL, HDFS, MapReduce and YARN, Hive, Impala, Kafka, Sqoop, Oozie, Zookeeper, Azure, ADF, Azure Datalake Storage Gen2, Blob Storage, Unity Catalog, Delta Lake, SCD Types, Azure SQL DB, Synapse Analytics, Synapse Studio, Synapse DB, Databricks, Query Editor, HD Insight, Nifi, Azure IoT Hub, Event Hub, Stream Analytics, Stream Lit, Spark Streaming, Structured Streaming, Flink, Azure Resource Manager, Auto Loader, Serverless Functions, Logic apps, Key Vault, Docker, Kubernetes, Snowsql, Snow(Dimension and Fact) Tables, Star Schema, Teradata, Mongo DB, Apache Airflow, Splunk, Control-M, IntelliJ, PyCharm, Visual Studio, Jupyter Notebooks, SVN Tortoise, Zen Hub, Radar, Pager Duty and DBeaver, Trino.

Experience:

Data Engineer | Client: Apple | Innovative Information Tech. | June 2023 – Present

Role: Data Engineer at Apple Media Products Analytics & Data Engineering (AMP ADE)

- As part of AMP ADE team, worked on creating, maintaining ETL pipelines in CDP/Hadoop environment that are crucial in building day-to-day media product analysis reports.
- Designed a series of data validation and notification jobs to ensure the data from upstream sources is available to process as per the agreed SLAs.
- Worked closely with a team of data engineers from various upstream source systems to design efficient pipelines in Spark using SQL and Python to process data into HDFS to help users analyze customer behavior, purchase patterns.
- Designed data pipelines using Azure Databricks, ADF and Synapse Analytics to ingest and transform various Apple media products data from different sources.
- Creating jobs using Databricks/Spark and IoT Hub based on user requests, tune PySpark jobs for better performance and efficient use of resources.
- Designed case specific Dimension and Fact tables in Snowflake to identify and establish relations between various offers initiated by different Apple Media services.
- Created a new system to update delta lake tables with adjusted data and alert users upon report refresh jobs completion by eliminating the intermediate status tables.
- Creating custom SQL scripts to validate daily ingested data and identify anomalies.
- Proactively monitor, maintain, identify resource crunches, bottleneck issues and tune jobs for better performance that are consuming huge resources and running long.
- Working closely with BI Teams to create efficient Tableau/Qlik/Grafana reports by creating report specific tables and building views on top of transformed fact data.

Data Engineer | Client: Credit Suisse | Innovative Information Tech. | Oct 2017 – May 2023

Role: Data Engineer in designing and building data integration, curation, modeling and processing frameworks on Cloudera Hadoop/Big Data and Azure Cloud platforms.

- Working closely with stakeholders, business analysts and project manager and product architects to discuss, define technical requirements, create project documentation.
- Responsible for creating a framework on Cloudera Hadoop using SparkSQL and Python to efficiently ingest, transform and process large data sets into intermediate Hive tables.
- Hands-on experience with validating data in internal stage files before loading into staging tables using a series of pre-determined data quality checks.
- Involved in building an end-to-end automated distributed scalable data processing framework handling different file formats such as Parquet, JSON, CSV, Avro and XML.
- Used Azure services like ADF, Synapse Analytics, Azure Databricks and ADLS to revamp existing ETL pipelines and automate them to work on Azure.
- Setting up jobs that can leverage Azure Databricks, Apache Kafka and Azure DevOps tools to process, filter and update adjusted data into fact tables in ADLS.
- Creating data validation jobs using Spark and Python which is leveraged to create a series of Tableau dashboards to further analyze the integrity of data.

- Experienced in handling large datasets using Partitions, Spark In-Memory capabilities, Distributed processing and Broadcasts in Spark, Effective & efficient Joins during ingestion and transformation process.
- Experienced in designing Hive tables using partitioning, bucketing, incremental stats strategies to improve performance.
- Designed Sqoop jobs to load data into HDFS from Oracle systems and leverage Impala to derive insights for different finance applications.
- Experienced in Using Terraform and Jenkins CI/CD to automate pipeline deployments.
- Worked on creating pipelines/jobs based on user requests to ingest, transform, and write data to Hive fact tables that are used in report creation using Tableau, Qlik.
- Experienced in orchestrating workflows using Airflow, Control-M and to run, maintain and monitor large number of jobs in batch mode on daily basis.
- Performed duties of a scrum master to ensure that project deliverables are completed within agreed timelines.
- Worked as a tester for different applications within finance platform under the Carbon program initiated by Credit Suisse to develop new skills and effectively perform in all technical roles involved in the complete lifecycle of a project.

Big Data Engineer | NFC Solutions | Jan 2017 – Oct 2017

Role: Big Data Engineer in building ETL solutions

- Worked on CDH platform to build transformations using PySpark and SparkSQL.
- Developed a python module to read excel sheets, convert them into data frames using pandas and store them as parquet files.
- Performed sourcing of data from source systems to HDFS and imposing Hive Schema on the data to establish relations.
- Implementing a series of mandatory validation checks using SQL to analyze, detect and log defects in the imported data.
- Responsible for developing and maintaining the visualization reports using Tableau.
- Working with multiple individuals from different levels of SDLC to coordinate, prioritize, schedule and track project goals.

Education:

Master's in Engineering from University of South Florida, Tampa, FL (2015-2016).

Bachelor of Technology in Electronics and Communications Engineering from Mahaveer Institute of Science and Technology, Hyderabad (2011-2015).