

# NAGA PRADEEP DHANENKULA

## Cloud Data Engineer

(469) 653-0898 | [pradeep.dn27@gmail.com](mailto:pradeep.dn27@gmail.com) | [LinkedIn](#) | [GitHub](#) |

### Professional Summary

As a skilled Cloud Data Engineer with 5+ years of experience, I have a proven track record of solving critical business problems related to data. My expertise lies in creating efficient, parameterized data pipelines using both On-Prem and Cloud ETL tools for various use cases, including Data Migration of business applications and loading data into Data Warehouses on-premises and on the cloud. Additionally, I possess extensive knowledge in creating Ad Hoc automation scripts using Python to tackle complex challenges.

### Technical Skills

<b>Programming Language:</b>	SQL, Hive-SQL, Python, PySpark
<b>Big Data Ecosystem:</b>	Hadoop, MapReduce, Hive, Apache Spark, Databricks, Kafka, Airflow
<b>Cloud Platforms:</b>	Amazon Web Services, Microsoft Azure, Snowflake
<b>Data Concepts:</b>	Data Mining, Data Warehousing, Data Modeling, ML Algorithms, A/B Testing
<b>ETL/ELT Tools:</b>	DBT (Data Build Tool), SSIS, Fivetran, Airbyte
<b>Database:</b>	MySQL, PostgreSQL, HBase, DynamoDB, MongoDB
<b>IDE Tools:</b>	Visual Studio Code, PyCharm, Jupyter Notebook, Sublime Text
<b>Alerting Tools:</b>	DataDog, PagerDuty
<b>Methodology</b>	Agile, Scrum, Kanban
<b>Version Control:</b>	Git, GitHub
<b>DevOps Tools:</b>	Docker, Kubernetes, Terraform, CI/CD
<b>Soft Skills:</b>	Problem Solving, Change Management, Incident Management

### Professional Experience

#### Azure Data Engineer

Jul 2023 - Present

##### McKesson, TX

- Implemented solutions for Ingested and processed **over 10 TB** of data from diverse sources using Hadoop, MapReduce, HBase, Hive, and Sqoop, supporting scalable data processing across multiple projects.
- Optimized data processing by converting Hive/SQL queries into Spark transformations, resulting in a 40% increase in processing speed and efficiency.
- Implemented **DBT** transformations leading to a 40% improvement in query performance for analytics and reporting purposes in Snowflake, enhancing overall data processing efficiency.
- Integrated Apache Storm with Kafka to perform web analytics and to process clickstream data from Kafka to HDFS, enabling real-time analytics on user interactions and website traffic.
- Conducted performance tuning of Spark jobs in **Databricks**, optimizing resource allocation and configuration settings, resulting in a 50% improvement in job throughput and cost optimization on AWS.
- Developed functions in AWS Lambda that aggregated data from incoming events and then stored the result data in Amazon DynamoDB, enabling real-time data processing and storage.
- Developed and maintained 15+ dashboards on Tableau and AWS QuickSight for sales data reporting, providing actionable insights that contributed to a 10% increase in sales performance.
- Followed a rigorous change management process, **worked closely with the business for User Acceptance Testing (UAT) and signoff**, and assisted operations teams during deployments. Provided thorough handover of the developed solution with proper documentation.
- Developed **Data-Lineage diagram** using python to visualize the data flow from source system to the Data Warehouse layer, resulting in a 25% reduction in debugging time.
- Worked in Agile/Scrum environment, collaborating with cross-functional teams to deliver high-quality data engineering solutions iteratively and adaptively.

#### AWS Data Engineer

Mar 2021 – Dec 2021

##### Trigent, India

- Developed scalable data pipelines using AWS Glue for automated data ingestion and transformation, resulting in a 50% reduction in manual data processing efforts and improved data quality for downstream analytics.
- Enhanced data analysis capabilities by implementing advanced statistical models and complex data transformations using Python and SQL, resulting in a 20% improvement in predictive accuracy for business forecasts.
- Utilized AWS EMR (Elastic MapReduce) for distributed data processing and analytics, leveraging Hadoop and Spark clusters to handle large-scale data workloads and performing complex transformations.
- Successfully implemented POC (Proof of Concept) in a Development Databases to validate the requirements and benchmarking the ETL loads.
- Implemented Type 2 **slowly changing dimensions** (SCD) to capture historical changes in dimensional data for accurate trend analysis and reporting in data warehousing projects.
- Implemented partitioning and bucketing in Hive, developing queries to process data and generate data cubes for visualization, enabling faster query performance and efficient data organization.
- Loaded the aggregated data into MongoDB for reporting on the dashboard. Worked on MongoDB schema/document modeling, querying, indexing, and tuning.

Jan 2020 - Feb 2021

## Azure Data Engineer

### Tolaram Group, LAGOS, Nigeria

- Designed and developed a Centralized Data Lake platform for Analytics and Reporting needs of multiple countries spread across 2 geographical regions, as an **Enterprise Data Platform (EDP)** resulting in a 30% reduction in reporting time and a 25% increase in overall data accuracy.
- Designed and developed a Centralized Data Lake platform, resulting in a 30% reduction in reporting time and a 25% increase in data accuracy across multiple countries.
- Analyzed client's business requirements, estimated costs and timeframes, and prioritized product backlog items. Proposed technical solutions for new items in the backlog and adept at sprint planning.
- Successfully onboarded 2 new countries into the EDP platform using parameterized ADF Pipelines to move data from source system into the staging layer of enterprise data lake in Snowflake resulting in a 20% reduction in data ingestion time.
- Developed Spark Applications using PySpark and Spark-SQL for data extraction, transformation, and aggregation from multiple data file formats.
- Performed POCs** on various topics, such as usage of snowpark in Snowflake, Change Data Capture (CDC) analysis of source systems, Incremental load Data pipeline design in ADF, Data-Lineage visualization using Python, Masking policy Admin script automation using Python, Web scraping using python deployed in Azure Functions.

Jun 2018 - Dec 2019

## Associate Data Engineer

### Tolaram Group, LAGOS, Nigeria

- Successfully implemented Data and Documents migration of 10+ Business applications from legacy systems to new enterprise applications for 2 countries in Africa region.
- Developed an optimized Python script in **SSIS** Script task to load data from CSV files to SQL Database, resulting in a 30% reduction in time taken for data load.
- Orchestrated 30+ extraction, data quality, validation and reconciliation reports using Microsoft Power BI for business validation of the migrated data.
- Setup Data gateway to auto refresh the reports in Power BI Service. Reduced Teams effort in manual refresh by 5 hours per week.
- Utilized Azure services to build and automate data pipelines, effectively handling large volumes of data and reducing pipeline execution times by 40%.
- Developed and optimized Spark jobs within Azure Databricks environment, leveraging Azure Blob Storage for data storage, leading to a 40% improvement in data processing performance and scalability.
- Designed and implemented **CI/CD** pipelines Terraform, automating the deployment of Docker containers to Kubernetes clusters and reducing manual intervention by 60% in the release process.

Jul 2017 – May 2018

## Data Analyst

### Espire, India

- Utilized Python and SQL to analyze datasets from diverse sources, contributing to the successful completion of 10+ data analysis projects that supported business objectives.
- Designed and executed 15+ data visualization solutions using Tableau and Microsoft Excel, effectively presenting insights and trends to stakeholders
- Conducted data cleaning and data wrangling processes to ensure data quality and consistency, resulting in more reliable analysis outcomes.
- Managed MySQL databases, including data querying, optimization, and maintenance tasks to support data analysis and reporting requirements.
- Leveraged Azure cloud technologies to store, process, and analyze large datasets, optimizing performance and scalability of data solutions.
- Applied **A/B** testing on various business strategies, which provided insights that led to a 10% increase in campaign effectiveness.

## Certifications

- Big Data with PySpark by DataCamp - [Certificate](#)
- Data Manipulation with Pandas by DataCamp - [Certificate](#)
- Intermediate with Python by DataCamp - [Certificate](#)
- Intermediate with SQL by DataCamp - [Certificate](#)
- Lean Six Sigma Black Belt - [Certificate](#)

## Leadership & Volunteering

2016 - 2017

### Vice President, Health Club, Vellore Institute of Technology

- Managed a core team consisting of more than fifty members.
- Delivered 15+ public address sessions on substance abuse to spread awareness among students on campus.
- Guided and mentored students in the orphanage to help them grow and improve their academics.

## Education

### Master of Science in Data Science

University of North Texas

2022 - 2023

TX, United States

### Bachelor of Technology in Electrical and Electronics

Vellore Institute of Technology

2014 – 2018

Vellore, India