**BONDILI NISHITHA BAI**

469-659-7554

nishithab16@gmail.com

Denton,

Texas

76205 [Medium](https://medium.com/@nishithab16) [GitHub](https://github.com/nishitha15) [LinkedIn](https://www.linkedin.com/in/nishitha-bai-b40115352/)

* Data Engineer with over 6 years of experience designing, developing, and optimizing scalable data pipelines using **Hadoop, Spark, Kafka, Airflow, and Hive** for large-scale data processing.
* Proficient in **Python, PySpark, and Scala**, building robust ETL workflows and implementing real-time data processing solutions using **Spark Streaming and Kafka**.
* Expertise in cloud platforms including **AWS, Azure, and GCP**, leveraging services like **S3, Redshift, Glue, ADF, Synapse, Snowflake, BigQuery, and Dataflow** to architect scalable and cost-effective data solutions.
* Strong background in **SQL and NoSQL databases** such as **PostgreSQL, Oracle, DynamoDB, and Cassandra**, with advanced query optimization and data modeling for efficient storage and retrieval.
* Hands-on experience with **CI/CD pipelines, infrastructure automation using Terraform**, and workflow orchestration with **Apache Airflow and Cloud Composer**, ensuring reliable and automated data processing.
* Designed and optimized data architectures on **Google Cloud (BigQuery, Dataflow, Pub/Sub, Cloud Data Fusion, Dataproc)** to enable high-performance analytics and real-time data processing.
* Expertise in **dimensional modeling (star and snowflake schemas)**, **BigQuery performance optimization (partitioning, clustering, materialized views)**, and **serverless architectures using AWS Lambda and Google Cloud Functions**.
* Implemented automated monitoring and alerting for data pipelines using **AWS CloudWatch, Azure Monitor, and GCP Stackdriver**, ensuring real-time issue detection and minimizing downtime.
* Extensive experience optimizing **big data frameworks (Hadoop, Spark, Dataproc)** for improved performance, scalability, and cost efficiency.
* Skilled in designing and optimizing data warehouses and lakes using **Snowflake, Redshift, and BigQuery**, leveraging partitioning, clustering, and materialized views for high-performance analytics.
* Implemented data integration and streaming solutions using **Apache Kafka, AWS Kinesis, and Google Pub/Sub**, ensuring seamless data flow and real-time analytics.
* Developed and automated **data quality validation** checks within ETL pipelines using **BigQuery, Cloud Dataflow, and AWS Lambda**, ensuring data integrity and compliance.
* Strong expertise in data governance, security, and compliance, implementing IAM policies, encryption techniques, metadata management, and data lineage tracking using tools like **Apache Atlas, Glue Data Catalog, and AWS Lake Formation**.
* Built interactive dashboards using **Power BI and Tableau**, delivering real-time business intelligence insights and supporting data-driven decision-making.
* Strong analytical and problem-solving skills with a track record of optimizing **ETL performance**, reducing query execution times, and improving data processing efficiency across cloud and on-prem environments.
* Hands-on experience with CI/CD pipelines and infrastructure automation using **Terraform, Jenkins, and Git,** streamlining deployments and improving collaboration across teams.

# Skills

|  |  |
| --- | --- |
| * Programming: Python, SQL, Bash, Shell scripting * Frameworks: Databricks, Snowflake, Airflow * Database: PostgreSQL, Amazon Redshift, Oracle, MySQL, MongoDB * API: RESTful API Design, API Gateway, MuleSoft API. * Data Integration: AWS Glue, Apache NiFi, Informatica, Azure Data Factory, Cloud Data Fusion | * ⁠Cloud Platforms: AWS (S3, Redshift, Glue, Lambda, Athena, EMR, Kinesis), Azure (ADF, Synapse, Data Lake), GCP (BigQuery, Dataflow, Cloud Composer, Pub/Sub) * ⁠Data Modeling: Normalization, Star Schema, Snowflake Schema * ⁠Tools: Git, Jenkins, Docker, Kubernetes, Jira, * ⁠Testing: Unit Testing, Integration Testing, Performance Tuning |

**Experience**

Data Engineer II

**Lowes, Charlotte**

*January 2024- Present*

* **Developed and implemented automated testing frameworks** using **Test-Driven Development (TDD) practices**, ensuring high-quality **GCP-based data solutions** met business requirements.
* **Designed and optimized data architectures on Google Cloud Storage and BigQuery**, enabling efficient storage, retrieval, and processing of large-scale datasets for analytics.
* **Built and managed data pipelines using Apache Beam with Dataflow**, automating data ingestion, transformation, and processing for real-time and batch workflows.
* **Implemented streaming data ingestion solutions using Pub/Sub and Dataflow**, reducing data latency and enabling real-time analytics for faster decision-making.
* **Developed scalable ETL workflows using Cloud Data Fusion and Dataflow**, ensuring seamless extraction, transformation, and loading of structured and unstructured data.
* Developed a metadata-driven **ETL framework** to dynamically process multiple datasets, reducing manual effort and increasing pipeline scalability.
* **Optimized query performance in BigQuery**, leveraging **partitioning, clustering, and materialized views**, reducing query execution time and cost.
* **Ensured data security and governance compliance** by implementing **IAM policies, encryption techniques, and VPC Service Controls**, protecting sensitive information across GCP environments.
* **Designed and optimized data warehouses in BigQuery**, improving storage efficiency and analytical performance for business intelligence and reporting.
* **Automated ETL workflows using Apache Airflow with Cloud Composer**, ensuring efficient orchestration of data pipelines and timely data availability.
* **Utilized dbt for data transformations in BigQuery**, enhancing data consistency, integrity, and governance for reliable analytics and reporting.
* **Designed and implemented dimensional modeling** using **star and snowflake schemas** in BigQuery, improving analytical query performance and business intelligence insights.
* **Built robust error handling and logging mechanisms** within **Cloud Composer workflows**, ensuring pipeline resilience and quick troubleshooting of failed processes.
* **Integrated BigQuery with Google Cloud Functions and Cloud Run**, enabling event-driven processing and serverless compute scalability for high-volume data workloads.
* Designed and developed custom data transformation logic **using dbt (data build tool) in BigQuery**, ensuring efficient and reusable data models for analytics.
* **Optimized Spark jobs on Dataproc**, improving the efficiency of large-scale data transformations by leveraging **autoscaling, parallel processing, and optimized partitioning**.
* **Developed custom data quality validation checks** within **ETL pipelines using Cloud Dataflow and BigQuery**, ensuring high-quality data for downstream analytics.
* **Implemented version control and CI/CD pipelines for data workflows** using **Cloud Build and Terraform**, streamlining deployments and improving collaboration across teams.
* Implemented BigQuery BI Engine to accelerate dashboard performance in **Looker and Power BI**, improving query response times for business users.

Data Engineer II

**Fannie Mae, Reston, VA**

*September 2022- Dec 2023*

* Optimized complex SQL queries in **Amazon Redshift and RDS**, improving **query execution speed** and **reducing computational costs** through **indexing, partitioning, and query tuning**.
* **Built and maintained ETL workflows** using **AWS Glue**, enabling **seamless data integration, transformation**, and automation across AWS services.
* **Developed automated data pipelines** leveraging **AWS Data Pipeline** and **AWS Lambda**, streamlining **data movement** and reducing **manual intervention**.
* **Engineered and optimized data warehouse solutions** using **Amazon Redshift**, enhancing **data storage**, retrieval, and **analytical performance**.
* **Created dynamic dashboards in Power BI**, visualizing **business KPIs** and **performance metrics** to support **data-driven decision-making**.
* **Optimized data partitioning and indexing strategies** in **Amazon Redshift**, improving **query performance** and minimizing **data retrieval times**.
* **Automated data validation processes** using **AWS Lambda**, ensuring **data accuracy**, **compliance**, and reducing errors in reporting.
* **Integrated real-time data processing** using **AWS Kinesis** and **AWS Lambda**, enabling **fast, scalable data ingestion** and **analytics**.
* **Built and deployed serverless data applications** with **AWS Lambda and AWS API Gateway**, enhancing **data accessibility** and **system scalability**.
* **Monitored and managed data pipelines** with **AWS CloudWatch**, ensuring **reliable performance**, proactive issue detection, and system health.
* **Ensured data governance and compliance** by implementing **metadata management** and **quality controls** using **AWS Glue Catalog** and **AWS Lake Formation**.
* Developed custom DAX measures and calculated columns in **Power BI**, enhancing **data modeling and analytical capabilities** for complex business reporting.

Data Engineer I

**Disney, Florida**

*October 2020- July 2022*

* **Developed Python scripts** to automate **data processing** and **validation**, ensuring **accurate data workflows** and optimizing **operational efficiency**.
* **Designed and optimized SQL queries** for efficient **data extraction, transformation, and analysis**, significantly improving **query performance** and **execution time**.
* **Implemented SQL-based data validation techniques**, ensuring **consistent and reliable data** across diverse systems and applications.
* **Built and maintained ETL workflows** using **Azure Data Factory**, enabling seamless **data integration, transformation, and automation** across platforms.
* **Developed automated data pipelines** leveraging **Azure Data Factory and Azure DevOps**, streamlining **data movement** and enhancing **process efficiency**.
* **Engineered and optimized data warehouse solutions** with **Azure Synapse Analytics**, improving **data storage, retrieval**, and **analytical performance**.
* **Partitioned and indexed data** in **Azure SQL Database**, improving **query speed** and reducing **data retrieval times** for operational and analytical tasks.
* **Established and enforced data governance** protocols to ensure **data quality, integrity**, and **compliance** across all Azure-based data platforms.
* **Created interactive dashboards in Power BI**, visualizing **key business metrics** to enable **real-time, data-driven decision-making**.
* **Implemented row-level security (RLS)** in **Power BI**, ensuring **role-based access control and compliance with organizational security policies.**

Data Analyst

**Wipro, Hyderabad**

*January 2019 - September 2020*

* Cleansed, transformed, and validated datasets using **SQL, Python, and Excel**, ensuring **data accuracy and consistency** for analytics and reporting.
* Designed dynamic dashboards in **Power BI and Tableau**, presenting key insights to drive strategic decision-making.
* Enforced data security through **row-level security (RLS) in Power BI**, restricting access based **on user roles and compliance policies.**
* Automated dataset refreshes and report updates in **Power BI Service,** minimizing manual effort and ensuring real-time insights.
* Built and optimized **ETL pipelines** to process large datasets efficiently, enabling seamless data transformation and integration.
* Developed scalable data architectures leveraging **AWS services like Redshift, Glue, and Lambda** to enhance data processing performance.
* Streamlined deployment of data workflows using **Jenkins and AWS CodePipeline**, accelerating development cycles and improving operational efficiency.
* Enhanced **ETL performance** through **tuning techniques**, optimizing data transformation speed and reducing computational costs.
* Implemented **proactive monitoring and fault-tolerant processes**, reducing system downtime and ensuring business continuity.
* Established **metadata management and data quality controls**, improving data discoverability, governance, and consistency across platforms.

IT Analyst

**Wipro**

**Hyderabad**

*June 2018 - December 2018*

* Optimized automated data pipelines using UiPath and Automation Anywhere to enhance workflow efficiency.
* Analyzed business processes to identify and implement RPA automation solutions for improved operations.
* Designed, tested, and debugged RPA bots, ensuring smooth workflow execution and troubleshooting issues.
* Documented process flows, configurations, and user guides to support scalability and future maintenance.
* Leveraged data analytics to extract insights and drive process optimization through automation.
* Deployed and monitored RPA bots, ensuring seamless execution and proactive issue resolution.
* Collaborated with analysts, developers, and engineers to align automation strategies with business needs.
* Automated repetitive data tasks, reducing manual effort and boosting system performance.
* Mapped business workflows for automation, contributing to process redesign and efficiency gains.
* Enhanced RPA expertise by engaging in skill development and automation best practices.

# Education

Master of Science (M.S.) in Computer Science

**University of North Texas, Denton, TX, US**