Biniyam Kebede

📍 Stone MTN, GA   ✉ findbene@gmail.com   ☐ 404 603 0407   in linkedin.com/in/byka
 github.com/findbene   🌐 https://findbene.github.io/tech-with-biniyam-portfolio/

Cloud-certified Data Engineer building scalable, automated data pipelines using Python, SQL, & AWS.

Technical Skills

- **Cloud Platforms & Data Warehouses: AWS** (S3, EMR, Lambda, Kinesis, API Gateway, Redshift, Step-Function, SNS, SQS, IAM), Snowflake, Azure (Fundamentals).

- **Languages & Databases:** Python (Pandas, PySpark, FastAPI, Boto3), SQL (Advanced), PostgreSQL, HTML/CSS.

- **Big Data & Streaming:** Apache Spark, Apache Kafka, ETL/ELT, Batch & Real-Time Data Pipelines.

- **Data Engineering Concepts:** ETL/ELT Pipelines, Data Modeling, Data Warehousing, Data Quality & Validation, Code Versioning (Git), OOP Principles

- **Orchestration & Transformation:** Apache Airflow, dbt (Data Build Tool), Docker

- **BI & Data Visualization: Power BI** (DAX, Power Query), Microsoft Power Platform

- **Data Concepts:** Data Modeling (Star Schema), Data Quality & Monitoring, API Development, Git, CI/CD

DATA ENGINEERING PROJECTS

1. Event-Driven Data Ingestion and Processing Pipeline (AWS)

- **Technologies:** AWS Lambda, API Gateway, S3, SQS, Python (Boto3)

- Architected a scalable, serverless pipeline to ingest real-time data streams via an AWS API Gateway endpoint.

- Engineered an AWS Lambda function to process and validate incoming data, placing valid records onto an SQS queue to ensure data durability and decouple ingestion from processing.

- Developed a second Lambda function, triggered by the SQS queue, to perform final data transformation and cleansing before storing the structured data in a Snowflake data warehouse via its Python connector.

- This event-driven architecture ensures high availability and low-latency processing for real-time analytics needs.

2. ETL Modernization & Data Quality Enhancement with dbt and Snowflake

- **Technologies:** dbt, Snowflake, SQL, Git

- Developed a modular data transformation project using dbt to build analytics-ready data models from raw data stored in Snowflake.

- Wrote custom data quality tests in dbt (e.g., uniqueness, not null, accepted values) to automate data validation, leading to a 95% reduction in manual validation efforts and ensuring high data integrity.

- Implemented dbt models to transform complex, multi-source data into a clean, star-schema format, improving query performance for BI tools by an estimated 40%.

- Managed the project with Git for version control, following best practices for branching, commits, and pull requests.

3. Interactive BI Dashboard for Sales Analytics

- **Technologies:** Power BI, DAX, Power Query, Snowflake

- Connected Power BI to a Snowflake data warehouse to build a comprehensive sales analytics solution.

- Utilized Power Query to clean, transform, and model complex datasets. Wrote advanced DAX calculations to create critical KPIs like YTD revenue, customer acquisition cost, and regional performance.

- Designed a multi-page interactive dashboard that allowed business stakeholders to perform self-service analytics, reducing requests for ad-hoc reports by 30%.

4. Real-Time E-commerce Analytics & Demand Forecasting Platform

- Architected an end-to-end data platform to process **millions of daily e-commerce events** from **Kafka** streams, REST **APIs**, and batch files. This platform directly solved key business challenges in **demand forecasting** and **fraud detection**.

- Engineered a data transformation process with **Apache Spark** that **reduced order processing time from 4 hours to 30 minutes**, enabling near real-time inventory analytics and personalized customer marketing campaigns.

- **Collaborated with marketing and data science teams** to build clean, reliable data models in **Snowflake** using **dbt**, which served as the foundation for a new customer segmentation model, leading to a **10% increase in conversion rates**.

- Implemented a comprehensive data quality framework with dbt tests, **improving data accuracy by 20%** and building trust across analytical teams. The entire workflow was orchestrated using **Apache Airflow**.

- **Technologies:** *Kafka, Spark, Airflow, dbt, Snowflake, AWS S3, Python (PySpark), SQL.*

5. Scalable Clinical Data Warehouse & Patient Monitoring System (Healthcare)

- Designed and deployed a HIPAA-compliant data platform on **AWS** to centralize and process **terabytes of patient electronic health records (EHR)**, directly addressing the business need for improved patient outcomes and operational efficiency.

- Built and maintained critical batch data pipelines that **achieved 99.9% uptime**, ensuring clinical staff had reliable and timely access to patient data for critical care decisions.

- Implemented a cost-effective data warehousing strategy using **Snowflake**, which **reduced annual data storage costs by 15%** while simultaneously improving data accessibility for clinical research teams.

- Partnered with clinical staff and hospital administrators to translate their operational needs into technical solutions, delivering automated monitoring dashboards in **Power BI** that helped **optimize hospital operations and lower patient care costs**.

- **Technologies:** *AWS (S3, Lambda), Snowflake, Power BI, Python, SQL, dbt.*

6. Serverless API for Real-Time Data Access

- Developed a secure, high-performance RESTful API using **Python (FastAPI)** to provide internal teams with programmatic access to customer and patient data models stored in Snowflake.

- Deployed the API as a serverless application using **AWS Lambda** and **API Gateway**, ensuring automatic scaling to handle thousands of requests per minute with zero infrastructure management. This directly supported real-time applications built by other teams.

- **Technologies:** *Python (FastAPI), AWS Lambda, API Gateway, Snowflake, SQL, Docker.*

PROFESSIONAL EXPERIENCE

Online Data Analyst (Remote) | TELUS International | Jan 2025 - Present

- Executed data validation and quality checks on diverse datasets to enforce data cleanliness and support ML model training.

- Annotated and classified data according to strict engineering guidelines to ensure high-quality inputs for production AI systems.

Quality Control Inspector & Data Analyst | Ricoh Inc. | Nov 2019 - Dec 2021

- Managed high-volume data quality within MES/SAP systems, ensuring the accuracy of critical production data for downstream analysis.

- Developed performance dashboards in Excel by transforming raw system exports into actionable visualizations for management.

High Tech Machine Operator | Alcon Pharmaceuticals | May 2017 - Nov 2019

- Leveraged the Manufacturing Execution System (MES) to manage and ensure the integrity of high-volume production data streams at the source.

- Contributed to implementing an automated MES data encoding process, reducing manual data entry errors by 95% and improving downstream data quality.

- Provided on-the-floor technical support for data-generating production systems, troubleshooting both hardware and software faults to minimize data loss and system downtime.

IT Support Specialist | Cita Technology | Jan 2016 - Feb 2017

- Developed and assisted with basic automation scripts using PowerShell and Bash to streamline repetitive system tasks and configurations.

- Managed and resolved technical incidents using Jira, adhering to ticket prioritization to ensure operational uptime for end-users and systems.

CERTIFICATIONS

- AWS Certified Cloud Practitioner (CLF-C02)

- Microsoft Certified: Azure Fundamentals (AZ-900)

- Microsoft Certified: Power Platform Fundamentals (PL-900)

EDUCATION

Bachelor of Science in Applied Physics, Minor in Mathematics
Addis Ababa University, Addis Ababa, Ethiopia