

# Semeon Dawera



571-234-1672



semeon.dawera@gmail.com



<https://www.semeon.work>

---

## Summary

As a Senior Data Engineer at Google, I specialize in creating robust data pipelines. My experience extends to cloud services, programming, databases, data modeling and Reporting. With industry experience in Energy (Utility), Education, Retail and Cybersecurity, I bring a wealth of knowledge and skills to drive data-driven solutions

---

## Technical Skills

- Databases: MS SQL Server, PostgreSQL, MySQL, MongoDB, Snowflake, Redshift, BigQuery
  - ETL & Data Pipelines: Apache Airflow, AWS Glue, DBT, Python, Apache Kafka, Apache Spark, SSIS, Databricks
  - Cloud Platforms & Infrastructure :
    - AWS – S3, Lambda, RDS, EC2, Databrew
    - Google Cloud Platform – BigQuery, Dataflow
    - Azure - Data Factory, SQL Database
  - Programming & Scripting: Python, SQL, Java, C#, PySpark, JavaScript
  - DevOps, Version Control & CI/CD: GitHub, Jenkins, Docker, Kubernetes
  - Business Intelligence & Reporting: Power BI, Tableau, Looker
  - Big Data & Analytics: Hadoop, Hive, Spark
  - Data Modeling & Architecture: Erwin, DBSchema
- 

## Certifications

- AWS Certified Data Engineer
  - AWS Certified Solutions Architect
  - GCP Professional Data Engineer
  - GCP Professional Cloud Architect
  - Microsoft Certified: Querying Microsoft SQL Server 2012
  - Java Programming Certification – Northern Virginia Community College
- 

## Education

Bachelor of Science in Civil Engineering  
Addis Ababa Institute of Technology (AAiT)

---

# Professional Experience

## **Mandiant Cybersecurity (Acquired by Google)**

Senior Data Engineer

January 2024 – Present

- Built ETL pipelines using Apache Airflow, AWS Glue, and PySpark to ingest and transform massive volumes of cybersecurity telemetry for threat intelligence reporting
  - Developed data quality reports using Amazon Web Services (AWS) Databrew for Mandiant client onboarding, reducing onboarding time by 40%..
  - Collaborated with DevOps for deployment and troubleshooting of Google Cloud Platform (GCP) data systems.
  - Automated threat log imports from Amazon S3 to Snowflake tables using Snowpipes.
  - Extracted data from GCP BigQuery to generate IR billing reports.
  - Analyzed and identified anomalies in Incident Response (IR) logs using Machine Learning (ML).
  - Created Pyspark framework for data transformations according to Mandiant's specifications.
  - Automated data transformation deployments with Jenkins and Docker.
  - Integrated Managed Defense data using Google's internal ETL tool.
- 

## **Kohls**

January 2024 - September 2024

Data Engineer Consulting

- Engineering the migration of Apache airflow 2 jobs to Apache Airflow 2
  - CI/CD automation via Jenkins to help with migration to Airflow 2
  - Created Airflow 2 operators and modules to be shared across 60+ engineers
  - Led a new project to stream retail sales transactional data using MySQL with CDC (Change Data Capture), Kafka and BigQuery
  - Helped troubleshoot failing Airflow Dags and worked with DevOps on release steps
- 

## **Trilogy Education (Acquired by 2U)**

September 2018 - January 2024

Senior Data Engineer/ BI Developer

- Wrote and automated multiple reporting SQL queries against snowflake database to generate student progress reports
  - Designed Tableau dashboards to track core KPI's like Attendance and Assignment
  - Wrote Salesforce SOQL queries then created data pipeline to run those queries via simple\_salesforce python package to incrementally pull 2U's Sales data to Snowflake
  - Created and automated data pipeline to pull data from 2U's Student Portal using Apache Airflow
  - Created a data model to store student engagement and performance like Quiz and Test Scores, Course Completion Rate and GPA
-

**Atsite (moved to InSite)**

Senior Data Engineer / BI Developer

*March 2017 – August 2018*

- Automated data retrieval from Weather Underground REST APIs using Python
  - Built ETL pipelines using Airflow, Python, and Amazon Redshift to handle multi-tenant IoT sensor data from 50+ commercial buildings
  - Managed data integration using Apache Airflow for diverse sources into Amazon Redshift.
  - Automated IoT device data flow and optimized energy bill data handling.
  - Designed fault-tolerant Airflow ETL workflows for frequent data updates.
  - Engineered Python-based data pipelines for meter readings.
  - Conducted rigorous SQL testing for data quality.
  - Led the Unified Data Store project for data modeling.
  - Developed efficient Tableau and Qlik Sense reports for analysis.
  - Created SQL reporting data layers with the help of views and stored procs.
  - Managed data backups, AWS server setup, and ELK stack for dashboards.
  - Conducted Machine Learning analysis in Kibana for energy consumption predictions.
- 

**U.S. Department of Labor**

Data Analyst

*November 2014 – February 2017*

- Transformed VBA reports into SSRS reports with SQL.
- Designed ETL processes for IBM DB2 to SQL Server data transfer.
- Scripted DB2 table definitions for migration.
- Led report development and data warehousing.
- Adapted DB2 scripts to SQL Server.
- Improved SQL Server database performance and queries.
- Converted MS Access reports to SQL Server Reporting Services.
- Translated MS Access Visual Basic code into T-SQL.
- Explored NoSQL databases for large datasets.
- Migrated SQL Server tables to AWS via Azure Migration Wizard.
- Contributed to ICD10 modification in Java.
- Analyzed ECS data and maintained data quality.
- Led data quality project for ECS business processes.
- Resolved database anomalies and facilitated changes.
- Organized developer-analyst meetings for data quality assurance
- 
-