


# Michael Lubinsky

---

 [mlubinsky@hotmail.com](mailto:mlubinsky@hotmail.com) |  408-775-4518 |  USA Citizen |  San Jose, CA  
Home page: <https://michael-lubinsky.github.io/>

## Summary

Results-driven Data Engineer with over 10 years of experience designing scalable big data solutions.

Skilled in Python, Apache Spark, SQL, Snowflake, and Databricks.

Expertise in ETL pipeline development, data modeling, cloud computing (AWS, GCP, Azure), and machine learning applications.

Adept at transforming complex data into actionable insights through data visualization and advanced analytics.

## Technical Skills

- Big Data: Apache Spark, Hadoop, Hive, Kafka, Databricks, Snowflake
- Real-time Processing: Apache Flink, Apache Spark Structured Streaming
- Databases: PostgreSQL, MySQL, Amazon Redshift, Presto, Druid
- Programming Languages: SQL, Python, Java, Scala, C/C++, Bash
- Data Engineering Tools: Apache Airflow, REST API development
- Cloud Platforms: AWS (S3, EMR, RDS), Azure, Google Cloud Platform
- Data Visualization: Looker, Tableau, Superset, Grafana

## Work Experience

### Samsung — San Jose, CA

#### Senior Data Engineer, contractor (Jan 2023 – Present)

- Designed and deployed ETL pipelines using Apache Airflow, Databricks, and AWS.
- Created the logical and physical data models for Data warehouse and Data Lakes.
- Optimized complex SQL queries and improved query performance by analyzing and tuning execution plans.
- Developed anomaly detection algorithms for time-series data using NumPy and Pandas.
- Collaborated with cross-functional teams on project designs.
- Built dashboards in Grafana and Tableau.

### Roku — San Jose, CA

#### Senior Data Engineer (Apr 2020 – Dec 2022)

- Built and maintained data pipelines using Apache Airflow, Apache Spark and SQL to track and analyze user behavior (click stream) across Roku devices.
- Delivered A/B testing framework and deployed machine learning models.
- Designed star schema models for data warehouse.
- Optimized data partitioning strategies resulting in a 35% improvement in query performance.

- Produced interactive dashboards in Looker for internal and external customers.
- Implemented multi-format data ingestion (JSON, Parquet, Avro), processing over 10 TB of data daily.

## **Arm — San Jose, CA**

### **Senior Software Engineer (Apr 2018 – Apr 2020)**

- Architected IoT data acquisition pipelines using MQTT and Kafka, handling >10K device messages daily.
- Built SQL-based materialized views and stored procedures for PostgreSQL.
- Implemented scalable anomaly detection algorithms to identify faults in IoT device behavior.
- Managed Docker environments and automated deployments on AWS infrastructure.

## **Apple — Cupertino, CA**

### **Senior Software Engineer (Nov 2013 – Mar 2018)**

- Led the development of a distributed, high-performance data pipeline processing over 5 TB/day.
- Implemented supervised and unsupervised machine learning pipelines (clustering, regression models, feature selection, etc).
- Developed REST APIs serving internal and external applications, supporting thousands of daily users.
- Built database models and schema designs leveraging PostgreSQL and NoSQL technologies for dynamic scaling.
- Collaborated across teams to deploy data analysis tools using RabbitMQ, Hadoop, Tableau, and Python.

## **Education & Certifications**

- Master of Science in Electrical Engineering (with Honors) — Moscow Institute of Electronic Engineering
- Machine Learning — Coursera (2011)
- Introduction to Artificial Intelligence — Coursera (2011)
- Apache Spark — Udemy (2015)