

# Michael Lubinsky

---

[mlubinsky@hotmail.com](mailto:mlubinsky@hotmail.com) mobile: 408 775 4518 Nationality: USA

## Summary

---

More than 10 years experience designing, implementing, and operating highly scalable and reliable services.

## Technical Skills

---

- Big Data: Apache Spark, Apache Flink, Hadoop, Snowflake
- Databases: PostgreSQL, MySQL, MariaDB, Oracle, Redshift, Presto/Trino, Druid
- Languages: Python, Java, Scala, C, C++, bash
- Apache Airflow, Kafka
- Cloud platforms: AWS, GCP, Databricks
- Data analysis and machine learning: A/B testing, ML ops, Pandas, scikit-learn, PyTorch, Tensorflow
- Reporting tools: Looker, Tableau, Superset

## Working Experience

---

### **Samsung. San Jose, CA 2023 - now. Contractor**

- Building pipelines for Maps Location Services using big data ecosystem
- Using Hadoop, Apache Spark, SQL, Java, Python, cleaning and preparing data for ML learning
- Using visualization tools: Tableau and Looker

### **Roku. San Jose, CA Apr 2020 – Dec 2022. Senior Data Engineer**

- Developed the data pipelines using Apache Airflow, Apache Spark, Redshift, Snowflake and Kafka
- Data acquisition, database design and implementation, dashboard/reporting using Looker and Tableau
- Worked with different data formats: JSON, Parquet, Avro
- Build A/B testing framework, deploying ML models to production
- Worked with AWS cloud services (S3, EC2, RDS)

### **ARM. San Jose, CA Apr 2018 – 2020. Senior Software Engineer**

- Developed the data pipelines for IoT (Internet of Things): data acquisition using MQTT protocol
- Created database schemas, tables, stored procedures, materialized views
- Implemented time series data analysis and anomaly detection algorithms
- Environment and tools: AWS S3, Hadoop, Hive, Python, Java, SQL, Docker

## **Apple. Cupertino, CA Nov 2013 – Mar 2018. Senior Software Engineer**

- Developed the high performance, scalable and distributed pipeline: data acquisition, parsing, uploading, analysis and reporting.
- Implemented data clustering, supervised and non-supervised learning algorithms.
- Developed database models and REST API, which is used by external applications.
- Environment and tools: Java, C++, Python, PostgreSQL and no-SQL databases, RabbitMQ, Hadoop, Tableau.

## **SAP. Palo Alto, CA Nov 2011 – Oct 2013. Senior Software Engineer**

- Created the logical and physical data modeling for HANA database; data partitioning and replication.
- Database performance profiling and tuning: # of threads, memory allocation.
- Run TPC-H database benchmarking
- Environment and tools: Linux, C++; Java, Python and SQL.

## **Yahoo Inc. Sunnyvale, CA Nov 2010 – Nov 2011. Senior Software Consultant in User Data and Analytics Group**

- I wrote map-reduce jobs for processing the hundreds of GB of data on Hadoop using Java and Pig Latin.
- I implemented the sampling algorithm, which dynamically allocates the number of Hadoop reducers per key proportionally to the data distribution
- Applied statistical methods (regression, classification) for Apache log data analysis.
- Environment and tools: Linux, Java, Hadoop.

## **Education & Qualifications**

---

- Apache Spark, Udemy 2015
- Introduction to Artificial Intelligence, Coursera 2011
- Machine Learning, Coursera 2011
- MS in EE (with honor) - Moscow Institute of Electronic Engineering