

BRENDAN COHEN

New York, NY | 407.454.0817 | brendanjcohen@icloud.com | [linkedin.com/in/brendan-cohen/](https://www.linkedin.com/in/brendan-cohen/)

EXPERIENCE

Capital One | Senior Data Engineer

July 2024 - Current

- Architected and led migration of event-driven **Spark** data pipelines from **EMR** to **AWS Glue**, improving performance by 65% and cutting **AWS** costs by \$60K annually.
- Designed and deployed a failover automation framework for **AWS Glue**, enabling 200+ data engineering teams to fully automate their quarterly disaster recovery exercises.
- Led **Spark** pipeline integration between **AWS** and **GCP (BigQuery, Google Cloud Storage)** to enable cross-cloud data platform evaluation and drive cost optimization of more than \$1M annually.
- Engineered batch **Spark** data pipelines processing 2B+ records daily as part of an enterprise-wide data platform modernization, establishing the target for marketing analytics.

Capital One | Data Engineer

March 2022 - July 2024

- Implemented an event-driven **Spark** pipeline processing over **150 gigabytes** of clickstream data daily, replacing a legacy batch job and improving data availability for 500+ internal consumers.
- Partnered with data analysts and product teams to design custom **Spark** data pipelines that supported evolving use cases for analytics on the Capital One Credit Card site.
- Redesigned event-driven marketing analytics pipeline orchestration pattern in **Apache Airflow**, introducing advanced resiliency options and reducing deployment lead time by 80%.

Perficient | Associate Data Engineer

March 2020 - March 2022

- Led a **Spark** development team in migrating 1850 Scala applications from **Cloudera Hadoop** to **Databricks, Delta Lake**, and **Amazon S3**.
- Restored archival functionality of over 350 million user event logs through open source ETL integration of Perficient software, improving storage efficiency and query performance.
- Developed modules for **Spark** structured streaming in **Microsoft Azure** environment, utilizing **medallion architecture** and **Delta Lake** on **Azure Databricks**.

EDUCATION

University of Florida | Bachelor of Science in Computer Science

December 2019

SKILLS

- Languages: Python • Java • Scala • SQL
- Data Engineering: Apache Spark • Databricks • Apache Airflow • Delta Lake
- Cloud Platforms: AWS (S3, Glue, EMR, Lambda, SNS, CloudWatch) • GCP • Azure
- Databases: Snowflake • Amazon RDS • MySQL
- DevOps/Other: Git • Jira • Jenkins • Terraform