Meeta Pandit

meeta.pandit890@gmail.com | 206-829-0633 | linkedin.com/in/meetapandit | github.com/meetapandit | Seattle, WA

Summary

Data Engineer with 6+ years of experience developing large-scale and high-performance data pipelines, data models, and fine-tuning queries to optimize performance of databases. Self-driven and motivated Data Engineering Boot camp graduate, having delivered 3 capstone projects using Python, SQL, Spark, Kafka, Docker, and Google Cloud Platform.

Technical Skills

Languages: SQL, Python, Shell Scripting

Database Systems: MySQL, Redshift, PostgreSQL, MS-SQL Server, SSIS, Cassandra

Distributed & Cloud Technologies: Apache Kafka, Apache Spark, Google Cloud Platform, Azure Databricks, Parquet, S3, HDFS

Analytics & DevOps Tools: Tableau, Google Analytics, Git, Docker, Apache Airflow

Education

Washington University in St. Louis Aug 2023 - Jan 2024
Data Engineer Boot Camp

University of Washington Sept 2015 - June 2017
Master of Science in Information Management GPA: 3.83

University of Mumbai Aug 2008 - July 2012
Bachelor of Science in Computer Engineering

Bootcamp Capstone Projects

Fraud Detector System using Kafka (github.com/meetapandit/kafka_fraud_detector_system) Jan 2024

- Ingested customer order transactions using kafka-python client to Apache Kafka installed on Docker
- Built Apache Kafka Consumer that classified the raw transactions as fraudulent vs legit which increased customer reliability

Equity Market ETL using PySpark on GCP (github.com/meetapandit/equity_market_analysis) Nov 2023

- Processed equity market data using Apache Spark dataframe API and stored partitioned datasets as parquet files in GCP
- Computed latest trade price, prior day closing price, and 30-min moving avg using SparkSQL and scheduled data pipeline in Airflow

Stock Market Analysis (github.com/meetapandit/stock_market_trends) Aug 2023 - Sep 2023

- Designed an application using Python to extract, transform, and store NYSE stocks data to Google Cloud storage
- Deployed application to GCP using cloud run, scheduled it using Airflow, and created interactive dashboard using Plotly Dash (Python)

Work Experience

Digital Analyst, Seattle, WA Oct 2020 – Aug 2023
Biossance.com

- Implemented data model and ETL to unify retail sales across 10+ countries, reducing dashboard consolidation time by 4 hours
- Built ETL for promo analysis in MySQL for analyzing customer segments, increasing return customer rate by 30% and new customer acquisition by 25% by creating 3 categories of data based on buying patterns
- Developed analysis for VP of E-Commerce, proving that Gift-with-Purchase increased repeat purchasers by 30% and lifetime value by 100% by analyzing first-time vs repeat purchase orders
- Launched a dashboard for monitoring campaign performance by integrating Google Analytics and Salesforce data, reducing update time from 2 hours to 10 minutes
- Redesigned orders ETL, making dashboard near-real-time by directly using Shopify API, eliminating upstream dependencies
- Created ETL to analyze frequently bought together products, aiding in personalized campaign creation by analyzing buying patterns
- Received org-wide catalyst award for proactively identifying a broken business process that directly impacted weekly leadership updates and for continuously updating stakeholders about progress which increased customer trust

Business Intelligence Engineer, Seattle, WA Oct 2018 – May 2020
Amazon.com

- Automated weekly business review dashboard, reducing update time from 2 hours to 15 minutes by eliminating manual updates
- Built workflow in Redshift by computing scores for features for 3P sellers and presented findings on high attrition rate to leadership
- Developed ETL using PostgreSQL tables from Tableau, increasing business teams' productivity by 20% by tracking dashboard adoption

Business Intelligence Engineer, Boise, ID June 2017 – Oct 2018
Micron Technology Inc

- Constructed parameterized SQL stored procedures to improve the execution time of SSRS reports by 60%
- Designed ETL to track operations health metrics which reduced the expiry of raw materials at warehouse facility using SQL and SSIS