

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 the performance of ETLs. Completed 2 hands-on data engineering bootcamps using cutting-edge technologies and winner of best capstone at DataExpert.io bootcamp.

Technical Skills

Languages: SQL, Python, Shell Scripting

Databases: MySQL, Redshift, PostgreSQL, Trino, Snowflake, Cassandra

Distributed & Cloud Technologies: Apache Kafka, Apache Spark, DBT, Databricks, Google Cloud Platform, Apache Iceberg, Parquet, S3

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

Education

DataExpert.io led by Zach Wilson Data Engineer Boot Camp	May 2024 - July 2024
Washington University in St. Louis Data Engineer Boot Camp	Aug 2023 - Jan 2024
University of Washington Master of Science in Information Management	Sept 2015 - June 2017 GPA: 3.83
University of Mumbai Bachelor of Science in Computer Engineering	Aug 2008 - July 2012

Bootcamp Capstone Projects

NYC Citibike data pipeline (https://github.com/meetapandit/nyc-citibike-data-pipeline)	Jan 2024
<ul style="list-style-type: none">Built a pipeline to track bikes' status at stations, enabling real-time updates to users using Kafka and Spark Streaming on DatabricksDeveloped an ETL in Snowflake using DBT for processing 10+ years (2.2 GB) of historical data for seasonality and trend analysisCreated interactive visualization using Streamlit to share insights about locations where bikes are faulty and most popular stationsPublished a Medium article (https://bit.ly/3W49Mml) outlining project execution and to share my learnings with data enthusiastsWon the best capstone for delivering high-quality data pipeline which also demonstrated a positive environmental impact	
Fraud Detector System using Kafka (github.com/meetapandit/kafka_fraud_detector_system)	Jan 2024
<ul style="list-style-type: none">Ingested customer order transactions using kafka-python client to Apache Kafka installed on DockerBuilt 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
<ul style="list-style-type: none">Processed equity market data using Apache Spark dataframe API and stored partitioned datasets as parquet files in GCPComputed latest trade price, prior day closing price, and 30-min moving average using SparkSQL and scheduled the job in Airflow	

Work Experience

Digital Analyst, Seattle, WA Biossance.com	Oct 2020 – Aug 2023
<ul style="list-style-type: none">Implemented data model and ETL to unify retail sales across 10+ countries, reducing dashboard consolidation time by 4 hoursBuilt 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 patternsDeveloped 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 ordersLaunched a dashboard for monitoring campaign performance by integrating Google Analytics and Salesforce data, reducing update time from 2 hours to 10 minutesReceived org-wide catalyst award for proactively identifying a broken business process that directly impacted weekly leadership updates and for keeping stakeholders updated about the progress which increased customer trust	
Business Intelligence Engineer, Seattle, WA Amazon.com	Oct 2018 – May 2020
<ul style="list-style-type: none">Automated weekly business review dashboard, reducing update time from 2 hours to 15 minutes by eliminating manual updatesBuilt workflow in Redshift by computing scores for features for 3P sellers which improved survey results of customer satisfaction rateDeveloped ETL using PostgreSQL tables from Tableau, increasing business teams' productivity by 20% by tracking dashboard adoptionDeveloped dashboard using Tableau and Redshift to track 3P sellers' KPIs, and presented findings on high attrition rates to leadership	