Meeta Pandit

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

Summary

Data Engineer with 6+ years' experience building scalable pipelines and data models that drive business impact. Graduated with AI and 2 data engineering bootcamps and participated in the Seattle Data and AI Hackathon. Actively participating in Toastmasters club activities.

Technical Skills

Languages: SQL, Python, Shell Scripting

Databases & Cloud Technologies: dbt Core, Trino, Apache Iceberg, Snowflake, Databricks, Apache Kafka, Apache Spark, GCP (CloudRun, CloudStorage, BigQuery, IAM), Parquet, MySQL, PostgreSQL, Redshift

Orchestration & Analytics Tools: Git, Apache Airflow, Docker, Tableau, Streamlit, DOMO, Google Analytics, Salesforce Analytics

AI Tools: RAG, Milvus vector database, Neo4j, LangChain, MCP

Work Experience

Data Engineer | Bill.com   Apr 2025 - Sep 2025
- Migrated dbt models from Snowflake to medallion architecture in Apache Iceberg, saving ingestion costs and reducing time to insights
- Optimized data models by removing inefficient joins and table scans by adding query pruning, predicate pushdown, and clustering
- Increased productivity and data credibility by creating reusable unit and integration test scripts, cutting manual work by 25%

Analytics Engineer | BigPanda   Nov 2024 - Mar 2025
- Identified and resolved slow joins and full table scans by query profiling in Snowflake, increasing dashboard refresh time by 15%
- Redesigned dbt models using materialized views, reducing full table scans and union all, and accelerated query runtime by 30%
- Built a unified account health data model in dbt and Snowflake by integrating multi-source metrics, enabling consistent insights that improved subscription growth and reduced churn

Data Engineer | Biossance.com   Oct 2020 – Aug 2023
- Developed a data model using MySQL and ETL to unify sales from disparate data sources, reducing reporting time by 4 hours
- Implemented data model and pipeline to unify retail sales across 10+ countries, reducing dashboard consolidation time by 4 hours
- Developed and deployed data pipeline for promo analysis in MySQL for analyzing customer segments, increasing returning rate by 30% and new customer acquisition by 25% by creating 3 categories of customers 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
- Built end-to-end data pipeline for analyzing campaign performance, reducing update time by 90% by integrating multiple data sources
- Received Catalyst Award for proactively identifying and resolving a critical business process that impacted weekly leadership reports

Business Intelligence Engineer | Amazon.com   Oct 2018 – May 2020
- Automated weekly dashboard with disparate data sources using Redshift and Tableau, reducing dashboard update time by 85%
- Built a data pipeline in Redshift by computing feature scores for 3P sellers, which improved customer satisfaction and retention rates

Business Intelligence Engineer | Micron Technology   Jun 2017 – Oct 2018
- Built data pipeline and monitoring dashboard to track storage and self life which increased efficiency of manufacturing and operations

Projects

AI-Powered Communication Tool | Claude Sonnet 4.5, OpenAI text embeddings small, OpenAI TTS   Oct 2025
- Built AI-powered tool using OpenAI embeddings, extracting contextually relevant words from images for empowering users with speech impairments (https://github.com/meetapandit/Oct-4-Hackathon-2025-)

NYC Citibike data pipeline | Kafka, Databricks, Streamlit (https://github.com/meetapandit/nyc-citibike-data-pipeline)   Jul 2024

Data Engineer Boot Camp at DataExpert.io   May 2024 - Jul 2024
- Delivered real-time bike status updates by building a Kafka and Spark Streaming pipeline on Databricks and won best capstone award

Equity Market ETL | GCP, PySpark and Airflow (https://github.com/meetapandit/equity_market_analysis)   Nov 2023

Data Engineer Boot Camp at Washington University in St. Louis   Aug 2023 - Jan 2024
- Built Airflow pipeline using Pyspark, CloudStorage and BigQuery to analyze billions of multi-exchange trading records, automating KPIs

Certifications & Professional Development

Google Cloud Certified Professional Data Engineer   Dec 2024 - Dec2026

Databricks Certified Data Engineer Associate   Oct 2024 - Oct2026

Education

University of Washington |M. Science in Information Management | 2017

University of Mumbai |B. Science in Computer Engineering | 2012