Aakash Deepak Bartakke

Binghamton, NY | (607) 338-8761 | aakashbartakke21@gmail.com| linkedin.com/in/aakash-bartakke | github.com/adb-21

EDUCATION

Binghamton University, SUNY | Master of Science in Computer Science (AI Track) | *Expected May 2026*
Dr. Babasaheb Ambedkar Technological University | Bachelor of Technology in Computer Science | *Jun 2020*

PROFESSIONAL EXPERIENCE

LTIMindtree | Senior Data Engineer
Pune, India | Dec 2020 – Jun 2024

- Architected cross-factory inventory intelligence data warehousing platform processing 100M+ daily records from 40+ data sources, **replacing 8-10 second single-factory QAD queries with sub-2 second Power BI dashboards** for 200+ supply chain users, **unlocking €100K profit** in year one through product phase-out decisions and overstocking prevention
- Engineered waste calculation ESG reporting system processing 500K+ daily records through SQL, SSIS and multi-level BOM-to-emissions logic, helping **reduce absolute waste from 35,744 tons (2023) to 31,599 tons (2025) and lowering tax liability by 14% (€2M+ annually)**
- Redesigned a critical ETL architecture by introducing delta load mechanism, cutting runtime **60% (4hr to 90min)** through **SQL query tuning**, index creation, and partition design, ensuring **on-time daily data delivery** for 90+ business users
- Eliminated **80% manual monitoring** (2hr daily per person across 10 engineers) by building **automated pipeline monitoring** dashboards replacing manual query analysis, while developing 3hr-saving parameter-based historical load mechanism using Azure data factory pipelines
- Accelerated pipeline performance across 10+ applications through **database optimization** (indexing, partitioning, query tuning) and SSIS dataflow materialization, reducing runtimes from 30min to 10min, 2.5hr to 1.5hr, maintaining 100% SLA in time sensitive environment

Binghamton University | Research Assistant
Binghamton, NY | Dec 2024 – Present

- Architected real-time IoT data infrastructure integrating 100+ thermal, pressure, and flow sensors across server racks using **Apache Kafka streaming pipeline** for thermal cooling research, implementing **event-driven data validation framework** managing DHCP compliance and MAC verification across 30+ hardware devices

PROJECTS

NYC Taxi Analytics Platform | Databricks, Delta Lake, Snowflake, AWS S3, Power BI | *Dec 2025 – Feb 2026*

- Engineered production-grade **Medallion Architecture** (Bronze→Silver→Gold) processing **200M+ historical records plus 4M daily NYC taxi trips using Databricks, PySpark and Delta Lake on AWS S3**, implementing idempotent date-partitioned ingestion with automatic backfill and sub-6-minute end to end data load
- Built analytics-ready gold layer with 10+ business metrics (daily stats, hourly trends, location analytics, driver behavior) **powering demand forecasting, surge pricing, and zone-based fleet optimization dashboards** projected to deliver 8-15% revenue lift and 5-10% idle time reduction
- **Automated Snowflake warehouse ingestion** using Snowpipe, Streams, and Tasks for exactly once processing, eliminating manual data loads while implementing outbound snapshot pattern cleanly decoupling Databricks from Snowflake

Real-Time Product Sales Pipeline | Kafka, AWS EC2, S3, DynamoDB, Airflow | *Jun 2025 – Jul 2025*

- Developed end-to-end **streaming pipeline** using Apache Kafka, **Python multiprocessing for parallel country-based sales ingestion,** and **Kafka consumer batch processing** on AWS EC2, optimizing cloud infrastructure costs to less than $4 per day through efficient DynamoDB storage and Airflow orchestration

TECHNICAL SKILLS

Data Engineering: Apache Spark (PySpark), Databricks, Apache Kafka, Apache Airflow, Delta Lake, SSIS
Cloud & Warehousing: AWS (S3, EC2, DynamoDB), Azure (Data Factory, Data Lake, SQL), Snowflake (Snowpipe, Streams, Tasks)
Languages & Tools: Python (Pandas, NumPy, Boto3), SQL (T-SQL, Stored Procedures), Java, Git, Docker, Linux
Analytics & Visualization: Power BI (DAX, M Query), Data Modeling (Star Schema, Snowflake Schema), ETL/ELT Patterns
Methodologies: Medallion Architecture, Incremental Processing, CDC, Data Quality, Agile/Scrum

CERTIFICATIONS

Microsoft Azure Fundamentals (AZ-900)