# Naveed Mohammad Bhat

Data Engineer | Python, Apache Spark, SQL, Databricks & Azure Cloud Infrastructure

+91 91496 21219 | naveedbhat085@gmail.com | Handwara, J&K, India | linkedin.com/in/naveedbhat085 | github.com/NaveedBhat | naveedbhat.in

## PROFESSIONAL SUMMARY

Data Engineer with hands-on experience building data pipelines on Databricks using Python, SQL, and PySpark, with exposure to machine learning and predictive modeling, including a deep learning classifier that reached 99.58% accuracy. Has practical experience applying GenAI and LLMs in automation workflows and is comfortable working with business stakeholders to translate analysis into clear, actionable recommendations. Brings 2 years of production software engineering experience with a focus on reliability and testing.

## TECHNICAL SKILLS

**Core Skills:** Python, Apache Spark, SQL, Cloud Infrastructure, Databricks, Azure, Android Jetpack, Flutter, Java, Dart, Kotlin

**Data Engineering:** SQL (Window Functions, Query Optimization), Data Modeling, Data Warehousing, ETL/ELT Pipelines, Databricks, PySpark, Apache Spark, Delta Lake, Unity Catalog, Medallion Architecture, Azure Databricks, Apache Airflow, Snowflake, dbt, Star Schema (Dimensional Modeling), SCD Type 2, Incremental Models (MERGE Upsert), dbt Snapshots & Macros, Jinja Templating, Data Quality Testing, Snowflake External Stages

**AI / ML:** TensorFlow, Keras, CNN, Transfer Learning, Pandas, NumPy, Matplotlib, Jupyter

**Cloud Infrastructure & Automation:** Azure Cloud Infrastructure (Databricks), AWS Cloud Infrastructure (EC2, S3), GCP, n8n, LLM Orchestration (Gemini Flash, Groq), Firebase, Power BI, Tableau, Streamlit, Git, GitHub, uv

**AWS Concepts (via AWS Cloud Solutions Architect certification):** Redshift, Lambda, IAM Roles & Permissions, Hive

**Databases:** PostgreSQL, Firebase Firestore, MySQL, SQLite, Hive NoSQL, MMKV, Schema Design

**Backend & APIs:** REST APIs, Firebase Cloud Functions (Node.js), Razorpay Payment Integration, GitHub REST API v3, Gmail/Drive/Sheets API, Telegram Bot API, Webhooks, JSON

## WORK EXPERIENCE

### Data Engineering

*Jun 2026 – Present*

**Insights Factory Pvt Ltd**, *Remote*

- Built and orchestrated production batch pipelines on Databricks (PySpark, Delta Lake, Unity Catalog, Workflows) using Medallion Architecture, data quality gates, audit logging, and email alerting.
- Developed ELT on Snowflake with dbt Core (incremental MERGE, SCD Type 2 snapshots, star schema) and a metadata-driven Apache Airflow framework (Docker) loading 10 vendor sources.
- Wrote and optimized SQL using window functions and CTEs against large relational datasets.

***Tools & Technologies:*** *Azure, Databricks, PySpark, Delta Lake, Unity Catalog, Apache Airflow, Snowflake, dbt Core, Docker, SQL (Window Functions), Python*

### Mobile Application Developer

*Dec 2024 – Nov 2026*

**Acmo Network Pvt Ltd**, *Kashmir, India*

- Shipped production Flutter/Android apps and modular SDKs, integrating REST APIs with structured error handling for reliable, secure data flow across 3+ enterprise clients.
- Architected Tyrads SDK, a multi-platform ad SDK adopted by 3+ production clients; served as primary technical contact for international clients.

***Tools & Technologies:*** *Flutter, Kotlin, REST APIs, Firebase, Android JetPack Compose, Android Development, Dart*

## PROJECTS

### Production ELT Pipeline: dbt & Snowflake (Medallion Architecture + Star Schema) | GitHub

- Built a production ELT pipeline on dbt Core and Snowflake: S3 external stage to Bronze/Silver/Gold (Medallion) with incremental MERGE upserts and watermark macro.
- Engineered metadata-driven Jinja models: a 34-column OBT from a 3-way join, split into SCD Type 2 snapshot dimensions and a fact table (Star Schema) for 5,000 bookings.
- Enforced data quality with 40 dbt tests (30 generic, 10 singular) and 5 reusable macros across every layer, with full ref() lineage and zero hardcoded SQL paths.
- Used ephemeral models to feed snapshots with no extra Snowflake objects, plus custom schema routing (BRONZE/SILVER/GOLD) and env_var credential handling.

***Tools & Technologies:*** *dbt Core, Snowflake, AWS S3, Python, SQL, Jinja, Git*

### Metadata-Driven Airflow & Snowflake ETL Framework | GitHub

- Architected a scalable, metadata-driven ETL pipeline using Apache Airflow (Dockerized) to ingest, validate, and load diverse vendor datasets into a Snowflake data warehouse.
- Designed a Master-Worker DAG architecture controlled by a centralized Snowflake config table, enabling the dynamic generation of concurrent pipelines for 10+ data models.
- Engineered robust Python (Pandas) data quality gates to handle dirty vendor data, strip rogue columns, and ensure 100% type-safe loads via PyArrow, alongside a centralized ETL logging system.

***Tools & Technologies:*** *Apache Airflow, Snowflake, Python (Pandas), Docker, SQL*

### Customer & Orders Analytics ETL Pipeline | GitHub

- Designed the data model and engineered a governed, production-grade batch ETL pipeline on Databricks using SQL and PySpark (Medallion Architecture: Bronze/Silver/Gold data warehouse) with Unity Catalog governance, automated Workflows, data quality gates, and email alerting.

***Tools & Technologies:*** *Databricks, PySpark, Delta Lake, Unity Catalog*

### AI-Powered Medicinal Plant Identification System

- Built an end-to-end deep learning pipeline classifying 41 medicinal plant species from 10,000+ botanist-verified images; achieved 99.58% accuracy, outperforming VGG16/ResNet50 baselines.

***Tools & Technologies:*** *Python, TensorFlow, Keras, Streamlit*

### ARIA - Autonomous AI News Intelligence System | GitHub

- Designed a 42-node autonomous pipeline processing 1,900 articles daily from 14 RSS feeds with LLM-based relevance scoring, dual-AI failover, and structured weekly reports.

***Tools & Technologies:*** *n8n, Gemini Flash, Groq LLaMA, Google Sheets API*

## EDUCATION

**M.Tech - Computer Science & Engineering**, Central University of Kashmir — *2023 – 2025, CGPA: 8.0/10*

**B.Tech - Computer Science & Engineering**, University of Kashmir — *2018 – 2022, CGPA: 7.0/10*

## CERTIFICATIONS & ACHIEVEMENTS

- Data, Data, Everywhere - Google Data Analytics Certificate, Coursera, 2023; AWS Cloud Solutions Architect - Coursera / AWS, 2023
- Microsoft Certified: SQL AI Developer Associate - Microsoft, Aug 2026