



Aaditya Bhilegaonkar

☎ [6028042664](tel:6028042664) ✉ [Email](#) 🐙 [Github](#) in [LinkedIn](#) 🌐 [Open Source Contributions](#)

SUMMARY

Data Engineer with 2+ YOE specializing in **next-generation AI infrastructure** and open-source data systems. Combines foundational data engineering at scale with hands-on execution in multi-agent orchestration.

WORK EXPERIENCE

[AI Data Intern](#)

University of Phoenix

May 2025 - August 2025

- Accelerated analytics-ready data delivery by 30% across 10+ sources by optimizing **Python/SQL** pipeline execution logic and eliminating processing bottlenecks.
- Decoupled data quality checks from the transformation layer into dedicated ingestion and loading stages, improving pipeline performance by 20%.
- Restructured S3 object storage layouts using high-cardinality keys to maximize partition pruning, slashing **Redshift Spectrum** query costs by 70% on 100–200GB weekly datasets.
- Automated Redshift cluster storage maintenance by orchestrating **Airflow DAGs** to execute VACUUM and ANALYZE operations, reclaiming 20+ engineering hours monthly.

[Software Developer - Data](#)

Amdocs

June 2022 - July 2024

- Restored billing data integrity for **200M+** subscribers across carriers like Vodafone and Bell Canada, preventing **\$100K+/month** in revenue leakage by remediating critical pipeline logic defects.
- Triaged an active **10M+ row** Oracle data corruption event within 7 hours using system log forensics and targeted SQL impact assessment queries to execute live production recovery.
- Optimized legacy SQL execution paths and refactored shell-based batch logic, yielding an **18%** runtime reduction on **500–800GB** telecom datasets.
- Deployed automated schema-conformance validation layers across **5+** ingestion pipelines processing **100GB+/day**, neutralizing downstream integration errors for analytics teams.

PROJECTS

[Agentic Data Engineering Platform | Google ADK, BigQuery](#)

- Architected an end-to-end multi-agent system utilizing the Google ADK to automate data preparation, job scheduling, and BigQuery warehouse loading, truncating ETL pipeline development timelines from 5+ hours to minutes.

[Data Lineage Analysis Claude Skill | Python, dbt, Airflow](#)

- Reduced data pipeline impact assessment from hours to seconds across 10+ file types by building an open-source Claude Code skill identifying all affected downstream consumers and auto-generating staged impact reports per change.

[LLM-Augmented Metadata Pipeline | PySpark, AWS Glue, LLaMA](#)

- Optimized ingestion time by **20%** and analyst query turnaround by **30%** by building a **PySpark + AWS Glue** Medallion ETL pipeline with **LLaMA** trained on table metadata to auto-generate SQL.

[RAG Retrieval Optimization | Milvus, Python](#)

- Improved **Recall@5** by **14%** and search latency by **35%** by benchmarking hybrid chunking strategies across **3** retrieval modalities using **Milvus** with nDCG@10/Recall@5 evaluation pipelines.

TECHNICAL SKILLS

Programming Languages : Python, Unix/Linux, SQL.
Agentic AI & LLM Systems : LangChain, Agent Orchestration, Claude Skills, Agent MCP, Google ADK, Milvus, RAG.
Cloud & Infrastructure : AWS, Docker, Kubernetes.
Distributed Systems : Spark, Kafka, Flink.
Data Warehousing : BigQuery, Snowflake, Redshift, OLAP/OLTP.
Orchestration & Tooling : Airflow, dbt, Fivetran.
LLMs : Claude, Gemini, ChatGPT.

EDUCATION

Arizona State University, Tempe, USA

Master of Science in Computer Science

May 2026

Pune Institute of Computer Technology, Pune, India

Bachelor of Engineering in Electronics and Telecommunications

May 2022