

# Omkar Agarwal

☎ +91-9511236581    ✉ agarwalomkar03@gmail.com    🔗 linkedin.com/in/omkar-agarwal/

## SUMMARY

Data Engineer with 3+ years of experience in building and optimizing scalable data pipelines for both batch and real-time processing. Skilled in working across multi-cloud environments with a proven ability to solve complex data problems in fast-paced settings.

## PROFESSIONAL EXPERIENCE

### Senior Software Engineer

*Persistent Systems Limited*

Jan 2024 – Present

*Pune, India*

- Built scalable batch ETL pipelines using **Azure Data Factory Studio** to move on-premise **SQL Server** data to **ADLS Gen2**, leveraging **Medallion Architecture** for structured layering.
- Processed **terabytes of data** with **Apache Spark** on **Azure Synapse Analytics**, boosting pipeline throughput by **35%**.
- Collaborated with cross-functional teams to translate business needs into scalable data solutions across cloud platforms.
- Developed a custom **Azure-based data pipeline** to monitor WhatsApp policy consent logs. Enabled conversion trend insights and helped shape digital adoption strategies.

### Software Engineer

*Persistent Systems Limited*

July 2022 – Jan 2024

*Pune, India*

- Processed **terabytes of data** using **Apache Spark** on **Dataproc** (GCP) for batch ETL workflows, implementing optimizations like column pruning, partitioning, and broadcast joins to boost performance by **35%**.
- Modeled and maintained analytical datasets in **BigQuery** by building partitioned and clustered tables, reducing query time and enabling self-service analytics that cut analysis time by **60%**.
- Enforced schema validation and data quality checks across **multi-cloud systems**, ensuring consistency and integrity in analytical workflows.
- Contributed to core insurance platform development using **.NET**, **C#**, and **SQL Server**, supporting key features like policy booking and quote generation.

## SKILLS

**Technical Skill:** SQL, NoSQL, Python, Statistics, ETL, ELT, Data Visualization, Apache Spark, Data Lakes, Data Modeling, Docker(Basics)

**Tools:** GCP (GCS, Big Query, DataProc, Pub/Sub, Dataflow), AWS (S3, Athena, Glue), Power BI, Excel, Azure (ADLS Gen2, Data Factory, Synapse Analytics, Bus Service, Functions), MongoDB

## PROJECTS

### eCommerce Analytics System (Python | Azure | Power BI)

Personal Project

- Simulated real-time user activity logs using a custom Python script and ingested existing batch product data; stored both in **ADLS Gen2**.
- Processed batch data using **PySpark** to derive insights like top-selling products and customer behavior trends, visualized through **Power BI**.
- Streamed live logs using **Azure Stream Analytics** directly into **Power BI** dashboards and triggered alerts for suspicious activity using **Azure Functions** to detect potential DDoS attacks.

### Anuvad – AI Voice Cloning Translator (OpenAI | Django | AI/ML)

Corporate Team Hackathon

- Built multilingual AI-based text-to-speech translator with voice cloning using **OpenAI APIs**. Backend in **Django**, frontend in **ReactJS**.

## EDUCATION

### Dr. D.Y. Patil SOET (SPPU)

*Bachelor of Engineering in Computer Engineering*

2019 – 2022

*Pune, India*

## CERTIFICATIONS & ACHIEVEMENTS

---

**Certifications:** Google **Professional Data Engineer** (08/2024 – 08/2026), Google **Associate Cloud Engineer** (11/2023 – 11/2026), AWS **Certified Cloud Practitioner** (06/2023 – 06/2026), Microsoft **Azure Fundamentals** (07/2022 – Present)

**Achievements:** Two-time **Bravo Award Winner** for outstanding performance and recipient of the **Top Talent Award 2023**