

---

---

# Kaustubh Upadhyaya

Bengaluru, KA • +91 84319 88388 • kaustubhupadhyaya@gmail.com • linkedin.com/in/kaustubhupadhyaya

---

---

## PROFESSIONAL SUMMARY

Detail-oriented Data Engineer with 2 years of experience building scalable ETL pipelines and automating validation with Airflow, SQL, and Great Expectations. Skilled in AWS, API development, and cross-functional collaboration. Proven success in reducing data errors, optimizing performance, and delivering reliable cloud-integrated solutions. Passionate about data quality and driving business impact through engineering.

---

---

## WORK EXPERIENCE

### Data Engineer

*Eli Lilly and Company - Bengaluru, India | July 2023 – Present*

#### Pipeline Development & Automation

- Spearheaded the migration of 50+ DAGs to Kubernetes-based Airflow, decreasing data error rates by 15% and improving overall pipeline reliability by 20% with robust data validation.
- Built automated Airflow error detection, cutting incident response time by 25%.
- Optimized 100+ Jenkins CI/CD jobs for SQL validation, boosting runtime efficiency by 20%.

#### Data Validation & ETL Quality Assurance

- Designed 30+ Great Expectations suites to enforce schema, null, and uniqueness checks across ETL pipelines handling 10 million+ records/day in a Data Lake.
- Embedded validation in Airflow DAGs, reducing manual QA by 70% and data issues by ~40%, with monitoring via AWS CloudWatch.

#### API & Database Engineering

- Developed REST APIs for real-time data ingestion and transformation, exceeding SLAs for uptime and latency.
- Designed performant PostgreSQL schemas for multi-occurrence fields, improving query speed by 40% via indexing and partitioning.
- Automated SQL execution integrated with AWS S3 for ingesting structured/semi-structured data; DBT integration planned.

#### Infrastructure & Cross-functional Collaboration

- Automated S3-to-SFTP folder sync using Shell scripts and AWS Lambda, reducing setup effort by 60% and enabling scalable data warehousing.
  - Led secret management across Kubernetes environments using AWS Secrets Manager, coordinating with 5+ teams to manage 50+ secured variables.
  - Configured modular ETL scheduling in Airflow with cron triggers and custom mail alerts, reducing misconfigurations by 25%.
  - Enhanced data integrity by implementing data validation rules and constraints in collaboration with DBAs, ensuring data consistency across 10+ tables; this improved data accuracy.
  - Delivered case studies leveraging S3-hosted data lakes and transformation logic, accelerating clinical decision-making by 20%.
- 
- 

## EDUCATION

Bachelor of Engineering, Information Science & Engineering  
BMS College of Engineering, Bengaluru | Aug 2023 Graduate | CGPA 8.34/10

---

---

## SKILLS & CERTIFICATIONS

- **Languages & Querying:** Python, SQL, Bash.
  - **Data Engineering Tools:** Apache Airflow, Jenkins, Great Expectations.
  - **Databases:** PostgreSQL, MySQL, Oracle DB, RDBMS.
  - **Cloud & Storage:** AWS (S3, EC2), GCP, HPC systems.
  - **Data Integration:** REST APIs, JSON, XML.
  - **Testing & Automation:** Unit Testing, Data Validation, CI/CD Pipelines.
  - **Other:** ETL Development, Secret Management, Schema Design, Monitoring & Alerting Systems, Cross-functional Collaboration.
- 
-