Rohit Yadav
Data Engineer

+353 899 706295
Dublin, Ireland
rohit.yadav.analytics@gmail.com

Data Engineer with about 2 years of hands-on experience using SQL, Python and cloud tools to build ETL/ELT pipelines and turn messy data into clean, reliable datasets for the business. I enjoy solving data problems, improving pipeline performance and data quality, and working closely with analysts and data scientists so they have the right data for dashboards and decisions.

SKILLS

Programming & SQL	SQL (PostgreSQL, MySQL, SQL Server – joins, CTEs, window functions), Python (Pandas, ETL scripts, automation)
Data Engineering	ETL/ELT pipelines, batch processing, basic streaming, data modeling (star/snowflake schemas)
Big Data & Streaming	Apache Spark (batch jobs, transformations, aggregations), Kafka or similar (basic producers/consumers)
Databases & Storage	Relational databases, cloud data warehouses (Snowflake / BigQuery / Redshift or similar)
Cloud & DevOps	AWS / Azure / GCP (storage, compute, basic IAM), Linux, Bash scripting, Git, GitHub, basic CI/CD
Data Quality & Docs	Data validation checks, monitoring of pipelines, documentation and data lineage basics
Collaboration & Soft	Working with analysts and data scientists, problem-solving, explaining technical topics to non-technical stakeholders

TECHNICAL EXPERIENCE

Data Engineer   **Aug 2022 — Sept 2024**
Ipsos Pvt. Ltd.   *Dublin, Ireland*

- Designed and productionised ETL/ELT pipelines in SQL and Python to consolidate data from APIs, files and databases into a central warehouse, cutting manual reporting effort by around 30%.
- Modelled star and snowflake schemas and optimised queries (indexes, partitions) to speed up key dashboards and reports, reducing average query times for analysts and data scientists.
- Implemented data validation checks, alerts and basic monitoring on critical pipelines, reducing broken refreshes and increasing stakeholders' trust in the data.
- Contributed to migrating and scheduling batch jobs with Spark and Airflow (or similar tools), simplifying daily operations and reducing manual intervention in routine workflows.
- Built and maintained datasets in the data warehouse to serve BI tools (Power BI / Tableau / Looker), enabling self-service reporting for business users.
- Worked with relational databases (PostgreSQL, MySQL, SQL Server) to design schemas, create indexes and troubleshoot performance issues in production queries.
- Used cloud services (AWS/Azure/GCP) such as object storage and compute instances to run data workloads, manage data lakes and support scalable pipelines.
- Applied Git/GitHub and basic CI/CD practices to version-control pipeline code, review changes and automate deployments across environments.
- Automated recurring data tasks with Linux and Bash scripting, reducing manual steps and improving reliability of nightly and hourly jobs.
- Documented pipeline logic, data models and runbooks, and shared knowledge with analysts and data scientists to make it easier to understand and use engineering outputs.

PROJECTS

Sales Analytics ETL Pipeline (End-to-End)   **2026**
Personal Project   *Dublin, Ireland*

- Built an end-to-end ETL pipeline in Python and SQL to ingest daily sales data from CSV/public datasets into a cloud data warehouse, standardising schemas and handling missing or dirty records.
- Designed a star schema with fact and dimension tables (customers, products, dates) and implemented transformation logic in SQL to make sales analysis faster and easier.
- Orchestrated daily batch loads using Apache Airflow so the sales data refreshes automatically, reducing manual report preparation.
- Added data quality checks (row counts, null checks, duplicate detection) and basic logging/alerts to quickly detect and fix issues in the pipeline.
- Used Linux and Bash scripts to automate supporting tasks (file moves, job triggers), improving reliability of the overall workflow.
- Built a simple Power BI/Tableau dashboard on top of the curated tables to visualise revenue trends, top products and key customer segments for stakeholders.

EDUCATION

Master of Science in Data Analytics, *Dublin Business School*   Jan 2026
Bachelor of Science in Computer Science, *S.I.E.S*   June 2022