

# David Nady

## Data Engineer

+20 10 124 79330 | [david.naddie@gmail.com](mailto:david.naddie@gmail.com) | [linkedin.com/in/david-nady](https://linkedin.com/in/david-nady) | [github.com/david11133](https://github.com/david11133)

### EDUCATION

#### Minia University

Bachelor of Science in Computer Science

Egypt, Minia

Aug. 2021 – Jun. 2025

**Relevant Coursework:** Database Design, Information Retrieval, Data Analysis, Data Mining, Artificial Intelligence, Algorithms, Data Structures, Operating Systems, File Processing, Software Engineering

### TECHNICAL SKILLS

**Languages:** Python, Java, C#/C++, SQL, Javascript

**Databases:** PostgreSQL, MS SQL Server, MongoDB, Cassandra

**Cloud & Tools:** AWS (S3, EC2), Snowflake, BigQuery, Redshift, Docker, Git, GitHub, VS Code

**Data Engineering:** Apache Airflow, Spark, dbt, Pandas, NumPy, Apache Kafka

**Others:** CI/CD, REST APIs, Linux, Power BI, Looker, Agile, Selenium, Beautiful Soup

### EXPERIENCE

#### Data Engineer - Google Summer of Code (GSoC) Contributor

May 2025 – Sep 2025

Integrated Ocean Observing System (IOOS)

Remote

- Developed data pipelines for 3 oceanographic datasets in NetCDF and CSV, standardizing them into a unified columnar storage Parquet schema to be used by researchers studying ocean and weather (see [final-report](#)).
- Reduced memory overflow achieving 78% gains by implementing chunking strategy for large NetCDF files.
- Scaled workflow by integrating Dask for parallel processing, used diagnostics and memory profiling to measure scheduling overhead and startup latency prior to implementation, resulting in cutting run time by 50%.
- Implemented an automated data quality testing framework using pytest, validating data integrity, sensor fusion accuracy, depth assignments, and scientific profile conversions through unit and functional tests.
- Built a custom multi-threaded downloader in Python for the Oleander dataset, automating retrieval most recent up-to-date data and unpacking of +250 of historical files from [ERDDAP](#) servers.

#### Python Developer (Professional Training)

Jul 2024 – Aug 2024

Information Technology Institute (ITI)

Egypt, Minia

- Built end-to-end analytics-ready data pipelines 3M rows in Python 3.11 + dbt Core 1.8. as a final project.
- Authored 30+ dbt models (staging → marts) with 80% test coverage (unique, not-null, accepted-values).
- Wrote complex SQL (CTEs, window functions, upserts) to cleanse and enrich 30+ disparate sources.
- Designed a star-schema dimensional model (1 fact, 5 dim tables) simulating business OKRs.
- Earned the highest accolade with a perfect 100% score by consistently maintaining exemplary discipline, and demonstrating outstanding dedication and learning aptitude.

### PROJECTS

#### NASA Data Ingestion Pipeline [🔗](#) | Python, Airflow, Kafka, Spark, Cassandra, Docker

May 2024

- Built a real-time incremental pipeline that ingests data using NASA's NEOWS APIs, processing up to 10,000+ near-Earth object records daily for analysis and decision-making.
- Wrote Python (PySpark, boto3, confluent-kafka) to orchestrate micro-batch ETL: Kafka → Spark Structured Streaming → Cassandra, and built 7 DAGs in Airflow AWS MWAA that cut manual intervention.
- Optimized data processing with Spark Streaming, boosting throughput by 30% and reducing latency by 25%.
- Deployed the solution in Docker containers on AWS, enhancing scalability and ensuring environment consistency for continuous integration and deployment.

#### Uber ETL Data Pipeline [🔗](#) | Snowflake, PostgreSQL, Python, Mage, Looker Studio

Jan 2024

- Designed a Snowflake centric star-schema data model that increased query performance 42% (avg. 3.1s to 1.8s) and clustering on high-cardinality keys and materializing 7 aggregate tables.
- Built a Python-based Mage ETL pipeline (Airflow-like) that ingests 1.2M ride records from CSV files into Snowflake with 99.9% SLA, and anomaly-detection module (Python scipy) auto-quarantines 2% outlier rows.
- Orchestrated the consolidation and migration of disparate data sources from multiple CSV files, effectively reducing data redundancy by 20% and streamlining the data architecture.
- Engineered a comprehensive and dynamic dashboard with Looker Studio, enabling in-depth data analysis and providing actionable insights through interactive visualizations and real-time data updates.

### CERTIFICATES

Data Engineer Associate certificate Datacamp