

Ashwin Pradeep

Bengaluru, Karnataka, India

ashwinpradeep.k@proton.me | [+91 6361859795](tel:+916361859795) | [Portfolio](#) | [Linkedin](#)

Data Engineer specializing in building scalable, containerized data platforms and end-to-end cloud data warehouses. Experienced in provisioning and orchestrating distributed environments using AWS (EKS, ECR, S3), and PySpark to support production-grade data workloads. Proficient in designing robust data modeling pipelines using Medallion architecture and relational engines to transform raw data into highly optimized, analytics-ready infrastructure.

Education

<i>MSc Data Science</i>	August 2022 - June 2024
Christ (Deemed to be University), Bangalore, India	CGPA: 8.3/10
<i>BSc Computer Science with Data Analytics</i>	June 2019 - May 2022
PSG College of Arts and Science, Coimbatore, India	CGPA: 8.5/10

Experience

Software Engineer | Openturf Technologies, Bangalore April 2025–June 2026

- **Data Science Education Platform:** Spearheaded the design and deployment of an end-to-end, multi-tenant data science platform on AWS EKS for a premier educational institute, empowering 3,000+ students to interactively build and execute production-grade ML projects.
- **Scalable Data Architecture:** Integrated Kafka, PySpark, and PostgreSQL/MySQL to handle massive data workloads and improve platform reliability. Innovated a hybrid container runtime (local sandboxes vs. cloud pods) to significantly reduce cloud compute costs.
- **Telecom CDR Analytics:** Engineered backend Node.js services and alert management portals for a tier-one Indian telecom giant. Processed millions of hourly Call Data Records (CDRs) and partnered with stakeholders to design reporting dashboards that track critical, real-time network performance metrics.
- **Generative AI Workflow Product:** Contributed to the development of an internal AI system that enables enterprise clients to seamlessly construct Retrieval-Augmented Generation (RAG) and LLM-driven workflows for automated document processing.

Data Analyst | Namaste IT Products and Solutions, Ernakulam May 2024–March 2025

- Devised ML engineering for a **Machine learning system for Image Captioning**. Developed an end-to-end machine learning pipeline for an image captioning system, managing the entire lifecycle from unstructured image/text data preprocessing and feature scaling to model training and deployment.
- **Engineered SQL and Python data pipelines** to ingest, clean, and model multi-regional order fulfillment and supply chain datasets across 450+ distinct business zones — implementing data quality rules including null checks, referential integrity constraints, and deduplication logic designing optimized dimensional data models and Power BI dashboards to streamline operational KPI tracking.
- **Served as a Power BI Specialist**, leading advanced data modeling, schema optimization, and DAX reporting workshops to upskill enterprise clients on analytics best practices

Data Analytics Intern | Thriftizer Solutions LLP, Bangalore Feb 2024–Apr 2024

- **Automated Data Pipeline:** Developed Python scripts to automate the extraction and aggregation of metrics from Google Analytics, significantly reducing manual data collection time.
- **Behavioral Modeling & Strategy:** Analyzed and modeled user behavior data to uncover insights that directly informed product marketing campaigns, ultimately improving client platform discoverability.

Skills and Interests

- **Languages & Scripting:** Python, SQL, C++, Node.js, Linux Shell Scripting
- **Data Processing & Architecture:** PySpark, Apache Kafka, dbt, Snowflake, Databricks, Medallion Architecture, Dimensional Modeling
- **Cloud, DevOps & Orchestration:** AWS (EKS, ECR, S3), Docker, Apache Airflow, Git
- **Databases:** PostgreSQL, MySQL
- **Analytics, Visualization & AI:** Power BI, Tableau, scikit-learn, Generative AI

Projects

SQL-Based Data Warehouse (Medallion Architecture) [In Progress]

Technologies Used: SQL, PostgreSQL, Medallion Architecture, Data Modeling, Query Optimization

- **Designed and implemented a relational data warehouse** from scratch within a **PostgreSQL** environment, systematically organizing data into structured Bronze (Staging), Silver (Cleansed), and Gold (Analytical) layers.
- **Authored complex SQL scripts, views, and transformation logic** to execute the ELT process, managing data type-casting, deduplication, and schema enforcement during layer-to-layer transitions.
- **Engineered optimized Star Schema dimensional models** using SQL DDL, building high-performance Fact and Dimension tables tailored for fast, seamless downstream business intelligence query execution.

IPL Data Analysis | *Python, PySpark, Databricks, AWS (S3), Matplotlib*

- **Engineered an end-to-end data pipeline** using Apache Spark to process and analyze 16 years of Indian Premier League (IPL) data (2008-2024), orchestrating data ingestion from Amazon S3, transformation, and complex querying via Spark SQL.
- **Leveraged distributed processing capabilities** in PySpark to efficiently handle and transform large-scale datasets, conducting in-depth analysis of team performance, historical player statistics, and match outcomes.
- **Delivered actionable business insights** by designing interactive visualizations in Matplotlib, uncovering key trends and enabling data-driven decision-making for IPL performance metrics.

Achievements & Certifications

- Data Engineering Professional Certificate by Snowflake
- Data Engineering with dbt - LinkedIn Learning
- Complete Guide to Analytics Engineering - LinkedIn Learning
- Data Engineering Foundations – LinkedIn Learning
- Learning Docker & Docker Compose – LinkedIn Learning
- Introduction to Databricks – 365DataScience
- Runners-up with a cash prize in “Gateways ‘22 “ Hackathon for building an IoT- Integrated AI-Assistant.