

Dhairya Pratap Singh Gaur

+91 7454948411 | dhairyagaur98@gmail.com

LinkedIn | **GitHub**

PROFILE

Data Engineer with 3+ years of experience leveraging big data technologies (**Python, SQL, PySpark, Databricks**) to build scalable **ETL** pipelines, process large datasets, and deliver actionable insights. Proven ability to streamline workflows and enhance data quality across cloud-based platforms (**GCP**).

TECHNICAL SKILLS

Programming Languages:	Python, SQL, C++
Big Data Technologies:	Spark, PySpark, Spark SQL
Cloud Computing:	Databricks, Google Cloud Platform (GCP)
Data Engineering Concepts:	Data Warehouse, Data Lake, Data Lakehouse, Medallion Architecture
Visualisation Tools:	Amazon QuickSight

EXPERIENCE

KPMG <i>Data Engineer</i>	November 2021 - Present <i>Bangalore, India</i>
-------------------------------------	--

- Ported and optimised SQL scripts from Teradata for catalog tables to ensure efficient and effective migration.
- Built metadata scripts and constructed **ETL** data pipelines within the **Medallion Architecture** framework.
- Implemented data processing and transformation logic using **PySpark** and optimised queries with **Spark SQL**, reducing runtime by 50%.
- Deployed and managed scalable data solutions leveraging **Databricks** for collaborative development and performance tuning.
- Designed and implemented detailed reports and data analysis pipelines.
- Developed interactive, data-driven dashboards utilising **Amazon QuickSight** for visualisation.
- Managed database operations and optimised queries using **pgAdmin** with **PostgreSQL**.

SmartBridge <i>Machine Learning Engineer (Intern)</i>	May 2020 - June 2020 <i>Hyderabad, India</i>
---	---

- Developed an automated AI system to predict individual life expectancy, achieving a predictive accuracy of 93%.
- Incorporated various factors such as demographic, health, lifestyle, and socioeconomic data.
- Utilized IBM frameworks, including :
- IBM Cloud** for infrastructure and scalability.
- IBM Watson** for data processing and machine learning model development.
- IBM Machine Learning Instance** for model training, deployment, and real-time predictions.

CERTIFICATIONS

- Databricks Fundamentals
- Databricks Lakehouse Fundamentals
- AWS Databricks Platform Architect
- GCP Databricks Platform Architect

EDUCATION

Guru Gobind Singh Indraprastha University <i>Bachelor of Technology in Electronic and Communication Engineering</i>	CGPA - 7.1 <i>August 2017 - July 2021</i>
Kendriya Vidyalaya Sangathan <i>Physics, Chemistry, Mathematics, English and Computer Science</i>	12th - 77.4% <i>April 2015 - March 2016</i>