

# Dhairya Pratap Singh Gaur

+91 7454948411 | dhairyagaur98@gmail.com

LinkedIn | GitHub

## PROFILE

Data Engineer with over 4 years of experience building scalable **ETL** pipelines, big data processing, and cloud-based solutions. Proficient in **Python**, **SQL**, **PySpark**, **Databricks**, **Snowflake**, and **GCP**, with expertise in optimizing workflows and improving data quality. Passionate about delivering high-performance data solutions that drive business insights.

## TECHNICAL SKILLS

<b>Programming Languages:</b>	Python, SQL
<b>Big Data Technologies:</b>	Spark, PySpark, Spark SQL, HDFS
<b>Cloud Platforms:</b>	Databricks, Snowflake, Google Cloud Platform (GCP)
<b>Data Engineering Concepts:</b>	ETL, Data Warehouse, Data Lake, Data Lakehouse, Data Modelling
<b>Visualization Tools:</b>	Amazon QuickSight
<b>Other Tools:</b>	pgAdmin, Git/Github, CI/CD

## EXPERIENCE

### KPMG

*Data Engineer*

November 2021 - Present

Bangalore, India

- **Client Engagement 1 : Data Platform Modernization using PySpark**
- Optimized 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 optimized queries with **Spark SQL**, reduced query runtime by **60%**.
- Deployed and managed scalable data solutions leveraging **Databricks** for collaborative development and performance tuning.
- **Client Engagement 2 : Data Pipeline Implementation and Analytics Development**
- Designed and implemented detailed reports and data analysis pipelines, increased data scalability by **30%**.
- Developed interactive, data-driven dashboards utilizing **Amazon QuickSight** for visualization.
- Managed database operations and optimized queries using **pgAdmin** and **PostgreSQL**.

### SmartBridge

May 2020 - June 2020

*Machine Learning Engineer (Intern)*

Hyderabad, India

- Developed an automated AI system to predict individual life expectancy, achieved a predictive accuracy of **94%**.
- 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 Lakehouse Fundamentals
- Databricks AWS Platform Architect
- Databricks Azure Platform Architect
- Databricks GCP Platform Architect
- Databricks Platform Administrator

## EDUCATION

### Guru Gobind Singh Indraprastha University

CGPA - 7.1

*Bachelor of Technology in Electronic and Communication Engineering*

August 2017 - July 2021

### Kendriya Vidyalaya Sangathan

12th - 77.4%

*Physics, Chemistry, Mathematics, English and Computer Science*

April 2015 - March 2016