



ABHISHEK TOMAR

SnowPro Certified Data Engineer

Phone - +91 7302213176

Email - abhiyou@gmail.com

Location - Noida



CAREER OBJECTIVE

Data engineer with expertise in Snowflake, Airflow, SQL, Python, Pyspark, and PowerBI. Passionate about problem solving and implementing new skills and technologies.

SOCIAL

LinkedIn - [linkedin.com/in/abhishek-tomar](https://www.linkedin.com/in/abhishek-tomar)

Kaggle - [kaggle.com/tomarat](https://www.kaggle.com/tomarat)

GIT Hub - github.com/Abhishek1310AT

TECHNICAL SKILLS

Snowflake	AWS	Azure
SQL	Python	Pyspark
Airflow	Hadoop	Hive
Power BI		

CERTIFICATION AND TRAININGS

Data Science Bootcamp

Azure Data Factory

Databricks Distributed Computing

Odin School

Udemy

Coursera

6 Months

-

ACADEMICS

- Sharda University**

Bachelor of Technology (EC)

CGPA: 7.2

- Intermediate**

Science Background

PROJECTS

Project 1

[Tools: **Airflow + S3 + Snowflake**]

- Used Python code to extract data from the chess.com API - [**Web scraping**]
- Loaded the extracted data into multiple JSON files containing streamers, game_statistics, and player_info, into an Amazon S3 bucket using **boto3**.
- Created a **Snowpipe** to automatically retrieve data from S3 bucket.
- Utilized **Amazon SQS** with Snowpipe to automate the loading of any new files present in the bucket to the external stage in Snowflake.
- Developed a Python DAG to scrap the data and a Snowflake DAG in **Airflow** to automate these processes on an hourly basis.

Project 2

[Tools: **Snowflake + Pyspark**]

- In this project you are going to analyze the bank loan dataset.
- Import the source file in **Pyspark** which is stored in the **Snowflake** database.
- The various insights like number of bankruptcies, highest credit score and average monthly debt was found out.

Project 3

[Tools: **EMR + Hive + MongoDB**]

- Using **AWS EMR cluster**, load the data from **S3** bucket to **Hive** via **Putty**.
- Perform analysis using SQL queries in **Pyspark**, and
- Store the result into MongoDB using **Pymongo**.

Project 4

[Tools: **Python+ EDA + Power BI**]

- The aim is to identify the item category has highest growth in which region.
- Libraries like **Pandas, Numpy, Seaborn** and **Matplotlib** are used while performing EDA and to visualize the data.
- Created a dynamic visualizing dashboard using PowerBI tool to analyze each parameter.

ACHIEVEMENTS

- As a HackerRank enthusiast, I earned the Five-star Golden badge for SQL, showcasing my passion for technology and problem-solving.

INTERESTS

- I placed 2nd in the National Air-Rifle Shooting Championship, proving my commitment to excellence and the sport.

