

# Rohan Kumar

## Computer Scientist

✉ rohanvaswani16@gmail.com ☎ +923362591060 📍 Gulf residency near teen talwar Clifton block 8 Karachi  
in <https://www.linkedin.com/in/rohanvaswani16>

### PROFILE

---

Highly motivated and skilled Data Engineer with more than one year of experience in designing, developing, and maintaining data pipelines and databases. Proficient in ETL/ELT processes, data warehousing, and data modeling. Adept at leveraging programming languages, such as Python, for data mining and analysis. Strong problem-solving and analytical abilities with a proven track record of optimizing data processing workflows. Committed to delivering data-driven solutions that drive business growth and efficiency.

### PROFESSIONAL EXPERIENCE

---

#### Data Engineer, Freelancing

01/2022 – present

As a freelance Data Engineer, I have successfully executed diverse projects, showcasing expertise in web scraping, Apache Airflow, Apache PySpark, and ETL tooling like Pentaho. My experience extends to cloud platforms, including Google Cloud Platform, AWS S3, and Azure Databricks, as well as data warehousing, Azure Cosmos DB, and Data Lake solutions.

#### Jr. Data Analyst, Data Check Limited

02/2023 – 05/2023  
Karachi, Pakistan

Design and implement data pipelines using ETL/ELT tools, ensuring seamless data extraction, transformation, and loading processes. Utilize Oracle SQL Developer to manage and optimize databases for efficient data storage and retrieval. Apply Python scripting for data mining and analysis, extracting valuable insights to drive data-driven decision-making. Employ Apache Airflow to create and manage complex data workflows and pipelines, enhancing overall data processing efficiency.

#### Jr. Business Intelligence Analyst, Sindh microfinance Bank Ltd

12/2022 – 01/2023  
Karachi, Pakistan

Spearheaded the design and implementation of data pipelines using Oracle SQL Developer and ETL/ELT tools, ensuring seamless data extraction, transformation, and loading from diverse data sources. Develop and maintain Key Performance Indicators (KPIs) for datasets, enabling stakeholders to make data-informed decisions and measure performance metrics effectively. Conduct thorough data quality checks and implement data validation procedures to ensure the accuracy, reliability, and consistency of data.

#### Summer Internship, State Bank of Pakistan

08/2021 – 09/2021  
Karachi, Pakistan

During my internship, I gained valuable experience in Relational Database Management Systems (RDBMS) and specifically honed my skills in Oracle. Additionally, I expanded my knowledge in network and information security, further enhancing my skill set in data management and protection.

### PROJECTS

---

#### Pipeline Data with Airflow, Consumer Financial Protection Bureau

we already have an API available for consumption which has parameters. We have frequent requests from our internal stakeholders to be provided with an up-to-date dashboard for their monitoring purpose. of State, From Date To Date, you can sample fetch data from it.

- 1) we used this list for Recursion on API to get all states' Data.
- 2) API Hit URL
- 3) Scrapping of Data for Washington State for Toda
- 4) Dumping Data into Google Sheets

#### ETL Data warehousing using Pyspark, SCD2 implementation

The task involves creating MySQL procedures and scripts to generate the schema and migrate data from an OLTP (Online Transaction Processing) system to an OLAP (Online Analytical Processing) system using a star schema for all the dimensions, including date, and fact. Additionally, the task requires implementing a PySpark Slowly Changing Dimension (SCD) on the employees' dimension using either SparkSQL or Spark Native API.

To validate the SCD2 implementation, the task requires demonstrating test cases on the employee dimension, including a scenario where a new record is inserted and another scenario where a record is updated. The SCD before and after the test cases should be shown to ensure that the implementation is correct.

### **Madhav ecommerce sales Dashboard, Power BI**

Analyze e-commerce sales data and created an interactive dashboard using Power BI. Used complex parameters to drill down in worksheet and customization using filter and slicers. Created Connections, join new tables, and enable user-driven parameters for visualizations.

### **Display with Data with Streamlit, Financial Consumer Complaints across the state.**

We were made to take it to the next level and create a Streamlit dashboard to showcase financial consumer complaints across the state. The dashboard will enable users to understand state-wise complaints and actionable items.

### **Store Furniture Data Prediction, Data Science**

We predicted previous data of store using python Jupiter library on Anaconda software.

## **SKILLS**

---

**MySQL / Oracle SQL database / SQL Microsoft Server**

**Apache-Airflow**

**Quality Assurance**

**ETL/ELT Tools**

Pentaho and Talend

**Google Cloud Platform**

**Data Science**

Machine Learning, Predictive analyzing, Deep learning

**Apache Spark**

**NoSQL**

MongoDB, CouchDB, CouchBase , Amazon SimpleDB, Riak, HBase, Redis

**Amazon web server**

AWS S3, AWS glue and lambda

**Python (Web Scraping and Analysis)**

**Data Warehousing**


**Power BI**

BI & data visualization tool that convert data from various data source into dashboard.

**Microsoft Azure**

## **EDUCATION**

---

**Bachelor of Science (B.S) - Computer Science** 

Shaheed Zulfikar Ali Bhutto Institute of Science and Technology, Karachi

09/2018 – 07/2022

Karachi, Pakistan

## **ACHIVEMENTS**

---

**Guest Speaker session, SZABIST**

Delivered a captivating lecture on Big Data and data warehousing concepts to MS Computer Science students at SZABIST.

Discussed the latest advancements and innovations in Big Data technologies, emphasizing their significance in handling large-scale datasets and extracting valuable insights.

Presented case studies of successful data warehousing projects, showcasing how industries leverage data warehousing to drive business intelligence and analytics initiatives.

Shared insights into industry trends, outlining how businesses are adapting to the dynamic landscape of Big Data and data warehousing solutions.

**Runner Up Prize, Final Year Project**

**Badminton Champion, silver medalist**

## **CERTIFICATES**

---

**Certified Data Engineering** 


By KarachiDotAI

**Microsoft Azure for Data Engineering** 

By Microsoft

**Microsoft PowerBI**

by Microsoft

**Introduction to Databases** 

by Meta

**Data Engineering Essentials using SQL, Python, and PySpark**

by Udemy (Durga Viswanatha Raju Gadiraju)

**Foundations: Data, Data, Everywhere** 

by Google

**Python for Data Science, AI & Development** 

by IBM