

**Sumit Kumar**

**E-mail:** [sumit.datascientist@gmail.com](mailto:sumit.datascientist@gmail.com)

**Mobile:** +91 9873788393

<https://sumitdatascientist.github.io/>

<https://www.linkedin.com/in/sumit-datascientist/>



---

### Profile Objective:

To have expertise with a progressive organization where I can employ **Data Engineer** knowledge and skills to contribute to the growth of the organization and personal growth in an effective manner.

### Summary

- ⌘ Offering **3.5 + years** of experience in a lead-level position across any functional sector within an IT organization of repute.
- ⌘ Solid programming skills with **Python and its libraries, Python, Spark, and Spark SQL.**
- ⌘ Experience in developing Spark applications using **Azure Spark and SQL in Databricks** for data extraction, transformation, and aggregation from multiple file formats for analyzing and transforming the data to uncover insights into customer usage patterns.
- ⌘ A good understanding of **Azure Spark architecture**, including **Spark Core, Spark SQL, Data Frames, Spark Streaming, Driver Nodes, Worker Nodes, Stages, Executors, and Tasks.**
- ⌘ An Excellent communication skills, an excellent work ethic, and a proactive team player with a positive attitude.
- ⌘ Domain knowledge of **retail, telecom, IVR, online food ordering, design, logistics, travel, and health.**
- ⌘ Strong skills in visualization tools **Power BI, Confidential Excel- formulas, pivot tables, charts, and DAX commands.**
- ⌘ Expertise in various phases of project life cycles (**design, analysis, implementation, and testing**)

### Technical Skills

- ⌘ **Programming Languages:** Python, SparkPython, SparkSQL, SQL, Scala
- ⌘ **PythonLibraries:** Pandas, NumPy, Scikit-Learn
- ⌘ **Azure Architecture:** Data LakeStorage, Data Factory, Databricks, SQL Database
- ⌘ **Databases:** SQL Server 2014/2012/2008/2005, MySQL, PostgreSQL
- ⌘ **Cloud Platform:** Azure
- ⌘ **Other Tools:** MS Excel (VBA, Macro), Power BI(DAX), Jupyter Notebook, PyCharm, Visual Studio
- ⌘ **Other Skills:** Machine Learning Algorithms, Data Structure, Algorithm, OOPs

### Academic Details

- ⌘ Bachelor of Technology (ME) from IEC College of Engineering and Technology, (AKTU University) with **69.80%** in 2017.
- ⌘ HSC from BSEB Board with **74.60 %** in 2013.
- ⌘ SSC from BSEB Board with **76.40 %** in 2010.

## Professional Experience

**Webhelp [Jan 2021 – Present]**

**DLF Centre Court, Gurgaon**

**Data Engineer**

### Responsibilities:

- ℵ Analyze, design and build modern data solutions using the **Azure Service** to support the visualization of data. Understand the current production state of the application and determine the impact of a new implementation on existing business processes.
- ℵ Extract, **transform, and load** data from source systems to **Azure Data Storage** services using a combination of **Azure Data Factory, Spark SQL, and Azure Databricks. Data ingestion** to one or more Azure Services (**Azure Data Lake, Azure Storage, Azure SQL**) and processing the data in **Azure Databricks**
- ℵ Created pipelines for multiple clients in **ADF** using **linked services, datasets, and pipelines** to extract, transform, and load data from different sources like **Azure SQL, blob storage, Azure SQL Data Warehouse**, write-back tools, and backwards.
- ℵ Developed **logical KPIs** as per client requirements using **PySpark and Spark-SQL** for data extraction, transformation, and aggregation from multiple file formats for analyzing and transforming the data to uncover insights into customer usage patterns.
- ℵ Responsible for estimating the **cluster size, monitoring and troubleshooting the Spark Databricks cluster.**
- ℵ Developed **JSON scripts** for deploying the pipeline in **Azure Data Factory (ADF)** that process the data using the SQL Activity.
- ℵ Created builds and releases for multiple projects (modules) in a **production environment using Azure Services.**
- ℵ Hands-on experience in developing **SQL scripts** for automation purposes

**Extension Technology [Aug 2019 – Sep 2020]**

**Faridabad, Haryana**

**Junior Software Developer**

### Responsibilities:

- ℵ Working on **SQL Queries** for **Data Visualization** and **Creation Dashboard.**
- ℵ Made different types of **reports** in **Zoho Analyst** as clients needed.
- ℵ Got Recommended for Many Projects by Senior Managers.
- ℵ Worked on Zoho tools like Creator, Crm, Books, Inventory backend in **Python and Deluge Script.**
- ℵ API Integrated 3rd party applications like **India Mart, Flipkart, Websites**, etc. to Zoho tools.
- ℵ Also Worked on **ERP Next Software** Backend and **ERP Next Implementation.**

## Certificate

Python 101 for Data Science: **Cognitiveclass** [6 Weeks]

Python for Data Science: **IBM** [4 Weeks]

5 Days Data Science Workshop: **GreyAtom** [5 Days]

Machine Learning: **Udemy** [4 Weeks]

## Endorsement

I hereby declare that the information furnished above is complete and true to the best of my knowledge.

**Place: India**

**Sumit Kumar**