*SUMIT KUMAR*

*DATA ENGINEER*

*E-mail*: sukumar8393@gmail.com
*Mobile: +91 9873788393, +91 9315607406*
*https://sumitdatascientist.github.io/*
*https://www.linkedin.com/in/sukumar8393/*

## *Profile Summary*

- ⋉ Offering **4+ years** of experience in a lead-level position across any functional sector within an IT organization.
- ⋉ Solid programming skills with **PySpark, Spark SQL, Python** and **its libraries**.
- ⋉ Experience in developing **ETL** Pipeline processes using **Azure Data Factory** and **Azure Databricks** for transforming the data to uncover insights into customer usage patterns and storing in **Azure Data Lake Storage.**
- ⋉ Excellent communication skills and a proactive team player with a **positive attitude**.
- ⋉ Domain knowledge of **retail, telecom IVR, online food ordering, travel, and health**.
- ⋉ Basics skills in visualization tools **Power BI and Confidential Excel.**

## *Technical Skills*

- ⋉ **Programming Languages**: Python, PySpark, SparkSQL, SQL
- ⋉ **Azure Architecture**: AzureData Lake Storage, Azure Data Factory, Azure Databricks, SQL Database
- ⋉ **Databases:** SQL Server 2014, MySQL, PostgreSQL
- ⋉ **Cloud Platform:** Azure , Google
- ⋉ **Other Tools:** MS Excel (VBA, Macro), Power BI(DAX), Jupyter Notebook, PyCharm, VisualStudio
- ⋉ **Other Skills:** Pandas, NumPy, Data Structure, Algorithm, OOPs

## *Professional Experience*

*Data Engineer*
*Webhelp, Gurgaon*
*[Jan 2021 – Present]*
*Responsibilities:*

- ⋉ Designing and **implementing data storage solutions, data pipelines, and data integration** patterns based on business requirements. This involves selecting appropriate Azure services, such as **Azure Data Factory, Azure Databricks** and **Azure Data Lake Storage.**
- ⋉ Creating and maintaining big data **pipelines** for multiple clients through **ADF** to ingest data from various sources including **databases, APIs, streaming data into the Azure data ecosystem** and doing some transformation using **PySpark** - **SparkSQL** in **Azure Databricks**
- ⋉ Developed **logical KPIs** as per client requirements using **PySpark** and **Spark-SQL** in multiple file formats for analyzing and transforming the data to uncover insights into customer usage patterns.
- ⋉ Responsible for estimating the cluster size, monitoring and troubleshooting the **Spark Databricks cluster.**
- ⋉ Worked on billion rows of data using **PySpark and Spark SQL**.
- ⋉ Created builds and releases for multiple projects (modules) in a production environment using **Azure Services**.
- ⋉ Hands-on experience in developing **Python and SQL** scripts for automation purposes.

*Junior Software Developer*
*Extension Technology*
*Faridabad, Haryana*
*[Aug 2019 – Sep 2020]*
*Responsibilities:*

- ℵ Worked on **Python** and **SQL Queries** for Data Visualization and Creation Dashboard.
- ℵ Worked on different types of **data formats** and **transform** the data into useful insights for clients needed.
- 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 ERPNext Implementation using **Python and SQL script**.

## *Academic Details*

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

## *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**