

PRAVEEN KUMAR SHARMA

📞: +91-9717830866 ✉ praveensharma2261988@gmail.com



DATA PROFESSIONAL – SOFTWARE ENGINEER

Targeting Assignments In: **Software Development Lifecycle (SDLC)**, **Data Engineering**, functions to **deliver robust solutions for complex technical requirements**

Profile Summary

MCA offering around 8 years of rich experience in Data Engineering / Analysis:

- Innovative developer having strong technical expertise in **Python Programming, PySpark** with knowledge of OOPS Fundamentals Concepts.
- Experienced in **Extract-Transform-Load (ETL), Hadoop Big Data & AWS Cloud Architecture**; worked as an Integral part of team for data gathering, data analysis framework entailing data cleaning, processing, analysis.
- Experienced in **ETL Development** using **SPARK** platform through **PySpark RDD and Dataframes**.
- Successfully executed key projects like **Honda, HP, Turner Analytics**.
- **Customer-centric professional & Effective Leader** with proven talent in **motivating & mentoring team members** and enabling knowledge-sharing amongst them; effective communicator with strong analytical, multi-tasking & problem-solving skills.

Technical Skills

Programming Languages: Python, PySpark

Frameworks: Customized Data Driven Framework

Databases: PostgreSQL, MySQL, SQL Server | **Languages:** SQL

AWS Components: EMR, EC2, S3, RedShift, RDS

Azure Components: DataBricks, DataLake, BlobStorage, FileShare

ETL Tool: DataIKU, Anaconda

Tools & Technologies: PyCharm, Eclipse, JIRA, Putty, VDI Instances, Git, SQL Workbench, PostMan, Tableau, Citrix

Operating Systems: UNIX, Linux, Windows

Work Experience

Since Jan'19: **LiquidHub Analytics Pvt. Ltd. Gurgaon, India as Senior Consultant**

Feb'15 - Dec'18: **Aricent Group, Gurgaon, Haryana, India as Senior Software Engineer**

Jun'14 – Jan'15: **Chhabra Info Solutions Pvt. Ltd., New Delhi, India as Software Engineer**

May'13 – Mar'14: **Inventxo Technologies Pvt. Ltd., New Delhi, India as Software Engineer**

Projects with LiquidHub Analytics Pvt. Ltd. - Gurgaon, Haryana

Title: Unilever

Period: Mar'21 – May'21

Role: Data Engineer

Team Size: 3

Tools: PySpark, Azure Cloud(DataBricks), SQL Server, Git

Description:

The idea is all about to automate the existing manual workflow of generating reports manually by processing/transforming data from SAP data as per business needs and eventually deploying to output point.

Responsibilities:

- Using PyCharm IDE, written PySpark dataframes on top of SAP tables to perform transformation/processing of data and executed same code remotely(Using DataBricks Connect library) over Azure Cloud DataBricks Cluster.
- Written Notebooks/Jobs inside Azure Cloud DataBricks environment to schedule the workflow.
- Managed SQL Server DB as per changing business needs.

Title: Friesland Campina (FC)**Period: Nov'20 – Feb'21****Role: Data Engineer****Team Size: 5****Tools:** Python (Pandas, Numpy), SQL, DataIKU, AWS**Description:**

Friesland Campina is a Dutch multinational dairy co-operative which is based in Amersfoort, Netherlands.

Goal of this project was to identify associations between products sold in F&B by identifying if specific SKUs are usually bought together or in short time intervals by customers. That will allow F&B to offer associated products to customers that are not already buying these combinations.

Responsibilities:

- Written SQL views to extract data from schemas created specifically to store Food & Beverages (F&B) related data and further created Python recipes using Jupyter notebooks in DataIKU to manipulate same as per Data Scientists needs.

Title: HP Project**Period: Nov'19 – Oct'20****Role: Data Engineer****Team Size: 3****Tools:** Python (Pandas, Numpy), SQL, JIRA, Putty, Tableau, PySpark, Unix Instances**Description:**

HP is specialized in developing and manufacturing computing, data storage, and networking hardware, designing software and delivering services.

In this Project, HP devices manufacturing and testing data is collected from different locations and further used for analysis and in development applications.

Responsibilities:

- Written Data Pipeline for end-to-end processing of data involving data collection, extraction, validation and transformation/processing and deploying to AWS Cloud Architecture leveraging Amazon Cloud Components (AWS) (such as AWS EMR, AWS EC2 instances, Amazon S3, AWS RDS).
- Developed Tableau Dashboards and Workbooks as per client requirements.
- Automated Tableau Workflow by first pulling input data from AWS S3 bucket and generate tableau extract (.hyper) file dynamically from same and publish same file to Tableau Server.

Title: Honda Project**Period: Mar'19 – Oct'19****Role: Data Engineer****Team Size: 8****Tools:** Python (Pandas, Numpy, BeautifulSoup, JSON), SQL, JIRA, Putty, PyCharm, Unix Instances**Description:**

Honda Motor Company, Ltd. is a Japanese public multinational conglomerate corporation primarily known as a manufacturer of automobiles, motorcycles, and power equipment and has been the world's largest motorcycle manufacturer since 1959.

This project is for enhancement of Honda Sales by developing Machine Learning Models trained and tested over Honda's Data. The Key idea for sales enhancement is to optimum utilization of incentives/promotions data.

Responsibilities:

- Extracted and integrated data from different data sources through web scraping using RESTful API's, setting DB Connections, through remote servers and primarily provided same to Data Science team to develop Machine Learning Model for client's sales enhancement through optimum utilization of incentives/promotions data.

Education

Master of Computer Applications (MCA) from ABES Engineering College, Ghaziabad, Gautam Buddh Technical University (GBTU), Lucknow in 2012; *secured 72.16%*

Bachelor of Computer Applications (BCA) from Institute of Management Education (IME) Sahibabad, Chaudhary Charan Singh University (CCSU), Uttar Pradesh in 2009; *secured 69.78%*

Personal Details

Date of Birth : 22nd June 1989
Address : J-1320, Sector-9, Vijay Nagar, Ghaziabad - 201009, UP
Languages Known : English& Hindi

ANNEXURE

Key Projects Executed

With Aricent - Gurgaon, Haryana

Title: Turner Analytics Project	Period: Mar'15 – Dec'18
Role: Senior Software Engineer	Team Size: 2
Tools: Customized Framework, AWS Components, PostgreSQL, Python, SQL, JIRA, Putty, Junos Pulse, VDI Instances, Agile Methodology, Git, Robot Framework, SQL Workbench	

Description: Turner Broadcasting System hosts major brands like TBS, TNT, CN, CNN & PGA. The project includes ETL processing of the data received from multiple vendors like Omniture, Freewheel, DFP and loading into AWS Cloud.

Responsibilities:

- Managed the design, development and implementation of End-to-End Migration Process from the existing Hadoop Big Data Architecture to AWS Cloud Architecture
- Developed Custom Framework and writing Python Modules for development and support of ETL Jobs
- Created S3 Buckets, managing policies for S3 Buckets, Utilized S3 Bucket and Glacier for storage & backup on AWS
- Written & executed Hive Queries using Qubole Data Service for providing GUI for all Hadoop Schemas/ Databases
- **Written Python Modules for:**
 - Making HTTP API Calls, Pulling & Parsing JSON / XML Data and normalizing the same using Pandas Module
 - Creating Database Connection with Oracle, PostgreSQL, AWS RedShift and performing CRUD operations with them
 - Preparing reports for tracking of pending/completed/failed tasks/jobs
 - Automating day- to-day routine jobs as per client requirement and specifications

With Chhabra Info Solutions Pvt. Ltd.

Title: SPARJ Project	Period: Jun'14 – Jan'15
Role: Software Engineer	Team Size: 4
Tools & Technologies: Python, JIRA, Spark, WebServices, HTML, CSS	

Description:

1. A web application 'SPARJ' is being developed which will provide an integration between SPARK (cisco's chat application) and JIRA (Bug/Task reporting tool).
2. It will create a chat room in SPARK as soon any JIRA ticket/issue is raised and will add the required members in the chat room (It was a manual process earlier)

Responsibilities:

- Involved in the Web Service and Front-End Development.
- Worked on Development of creating client-server architecture using Python/Django.

With Inventxo Technologies Pvt. Ltd., New Delhi, India

Title: AD - SCHEDULER Project

Period: May'13 – Mar'14

Role: Software Engineer

Team Size: 3

Tools & Technologies: Java, WebServices, HTML, CSS, JavaScript

Description:

1. This Desktop application helps to schedule the ad on WEB SITES, MOBILE WEB and APP.
2. It schedules the ad on WEB page with time frame, target goals, priorities, platform and geographies.
3. It has the capability to differentiate the ad on Test and live environment so the that the ad could be tested on real site before going to production.

Responsibilities:

- Involved in the API development of the Server side.
- Involved in Application development which was for Web Application supporting mobile app, mobile web and Desktop.