

# Sumit Pawar

**Big-Data Analytics, Big-Data Engineer**

H.NO. 886, Sector 14,  
Old DLF Colony,  
Gurgaon. Mobile: +91-  
+91-8930050050  
+91-8851589616  
Email: [sumitpwr7@gmail.com](mailto:sumitpwr7@gmail.com)  
[hadooper786@gmail.com](mailto:hadooper786@gmail.com)

I am a **Big-Data Analytics** and **Big-data Engineer** consultant having 7.2 years of IT industry experience. I am looking for a challenging and rewarding opportunity with an organization of repute that recognizes & utilizes my true potential while nurturing my analytical skills in the field of **Big-Data Analytics** and the **Big-data Domain**.

## **Technical Expertise:**

Category	Entities
<b>Big Data Hadoop Ecosystem</b>	Hadoop-2.7.2(stable), Hive, Pig, Sqoop, HA cluster with QJM, PySpark, Flume, Tableau, Cloudera 5, MapR 6.1.0
<b>Programming Languages</b>	Shell script, Python, Hive scripting, Pig Latin Scripting
<b>Database Tools</b>	MS-SQL, MySQL, Oracle, DB-Visualizer, DBBeaver.
<b>Operating System</b>	Windows(7/8/10), Linux(Ubuntu, Cent-OS)
<b>Other Data Tools</b>	MS-Excel, MS-Access, MS-Word

## **BIGDATA SKILLS**

- **Big Data Hadoop 2.7.x**
  - Experience in installation, configuration, supporting, and monitoring Hadoop clusters using **Apache, Cloudera distributions, and AWS**.
  - Experience in working with **MapReduce** programs using Apache Hadoop for working with Big Data.
  - 7.2 years of comprehensive experience as Big-Data engineering and data analytics.
  - Familiar with data architecture including data ingestion pipeline design, Hadoop information architecture, data modeling, and data mining, and advanced data processing.
  - **Written Shell Scripts in Linux enviromnet** for ETL process and job scheduling on Bigdata platform.
  - Setting permissions, Commissioning, Decommissioning, Balancing, and Managing Nodes.
  - Experience in importing and exporting data into Hadoop File system using **Sqoop** and **FLUME**.

- Worked on Big data tools –**Hadoop, Zookeeper, Hive, Pig, Sqoop, and Flume.**

#### **Horton-works Sandbox(Sandbox HDP 2.2 VirtualBox)**

- A hand on exposure on **Horton works Sandbox** using Oracle virtual box.
- Analysis of the huge data file by **using the Horton works cluster (HCatLog, PIG, HIVE, and Flume).**

#### **Technical Achievements**

- **Big Data Analytics with Hadoop and Apache Spark certificate** from “Lynda.com”
- **Big Data and Hadoop Essentials certificate** from **Udemy.**
- **Certification in Basic python** from **Skillsoft.**

#### **Extracurricular Activities**

- Certificate of **NCC (National cadet core)** with two years training completion in 2008

#### **Career History :**

**April 2021**

##### **Organization Name: Mindtree**

##### **Project : SOR VGP**

Description: SOR SOR is complete great solution for marketing analysis about product which relate to similar Group. This solution has different kind of data exists where giving great accuracy about the sales and prestaton of product in store. This project had great analysis over grained data using different visualization Method to end user to designed further business plans

**MAY 2017 -Feb 2021**

##### **Organization Name: Optum Global Solutions(United Health Group) (Permanent)**

##### **Project: DATA LIBERATION ANALYTICS(ITA DLA):**

Description: Data liberation Analytics is a comprehensive solution for all internal applications of the united health group. In which Data comes in Data Lake from multiple sources which include legacy RDBMS system, application server, Datameer, SharePoint, etc. We used Sqoop, Python API, and shell scripting for auto data ingestion and derived data from source data. For End-user data access, we provide a solution using Datameer, Hive warehouse access, or other data visualization tools like Talend, Tableau, PowerBI,

Excel workbook, etc.

**Job Responsibilities:**

- As a Data operation support engineer, I had to observe Operation running WebUI where we can observe each Job running status in one view which updates every minute.
- I had also need to follow up on job status from the TWS window which has all logs and job details with running machines and timing.
- Working on service-now incidents and recover the failed jobs by analyzing the configuration files.
- I had to find out the change impact with the application server during job running timing and inform to admin level person to reduce the possible impact.
- Working on the Datameer incidents and working with the customer to resolve the issue.
- I had open different types of incident tickets in ServiceNow according to application behavior and assigned them to correct workgroups to find the appropriate solutions for that.
- I had post job status to all major assets whose status is very important for other depending jobs.
- Working with different user stories related to the next release as per customer requirements.
- Creating new Data assets in the HIVE data warehouse on the BDPAAS side.
- Need to work with the onshore person to discuss Operational activity before day end.

**Projects: SmartSave (Data Park):**

Description: SmartSave is a very impressive data solution where we provide smart data decommission solutions to different active applications in UHG. In which we analyze the application behavior and its data challenges issues facing existing data models due to different data storage like data amount, variety of data, data consistency, etc. In the SmartSave program as a Big-data engineer, I worked with multiple applications under DATA PARK solution like NDB, FDS, ODXC, UHCCS, BASICS, COSMOS, CAMS, CHWY, etc. Working under the DATA PARK solution I took end-to-end responsibility during data decommissions.

**Job Responsibilities:**

- Working with the application team and analyze application data according to the client's requirement
- Recommend the correct data solution according to the client's requirement after data analysis.
- Procure the different environments with the help application team, used in decommission like source access request, BDPaaS tenant volume, access using different groups according to its hierarchy.
- Creating scripts according to data variation at application side data.

- Automate the process using **shell script** where numbers of tables are huge and application access window time is less.
- Write Java programs to solve the data compatibility issue between source and destination.
- Writes the automated scripts for data compression at the BDPaaS side for smartly using volume and increases the data query access response at the application level.
- Implement the Object Storage data model as an application data backup solution to avoid data loss situations in bed circumstances.
- Write a data validation script to make the data validation process faster.
- Provide the knowledge-sharing session to the application team to feel comfortable with the solution.

### **Projects: Zorro(EMC to BDPaaS)**

Description: Zorro(EMC to BDPaaS) mainly the project in which I work with the team as a Big-data developer for moving data from the EMC repository to BDPaaS(Big-Data Platform as a Service) to reduce the data cost used in the EMC server for the application. In Zorro, I have to mainly deal with XML format from XDB server using XQuery. Application data stored in XML format on the EMC server and our application internally worked with XQuery to collect the data from the XDB server.

### **Job Responsibilities:**

- Work with the team to analyze how data will move from EMC to BDPaaS.
- Decision-making to decide the tenant volume for application data on BDPaaS.
- Procure the Volume on Object storage as well on BDPaaS.
- Fetching the data from the EMC server using XQuery.
- Develop the solution for data abstraction from XML format using PIG Latin.
- Develop the Script to convert the XML format of data into CSV format.
- Create the Hive Script for creating the tables from extracted data.
- Data validation from the EMC server to Hive tables.
- Prepare Job Aid Documentation, Job process for User.

### **Projects: Online Archival :**

Description: Online Archival Project is based on decommissioning the application data from its relational database server to the BDPaaS environment to reduce the data storage cost as well as improving the analysis power. In Online Archival I focused on Database servers like the mainframe, Oracle, DB2, etc. In which I have to Archive the data of currently running applications in which working lives. In the online archival, I have to deal with different data formats like XML, CSV, JSON, etc.

### **Job Responsibilities:**

- Analyze the source side data formats and prepare strategy according to data complexities.
- Working with the application team to procure different access regarding the application database.
- Procure the tenant volume according to application data volume.
- Procure the volume on Object Storage as well according to data importance.
- Creating SGOOP script and handling the different forms of data during import from the source side.
- Creating hive tables for imported data in Hive Warehouse.
- Data validation from source (RDBMS) to destination (BDPaaS) about volume as well as in data format.
- Create a Job Aid Document as well as KT for the application team.

### **Awards and recognition :**

- Awarded by **Aquamarine Recognition** for dedicated work and timely delivery in the COSMOS application.
- Awarded by **Topaz Recognition** for the best performance in the BASICS application.

### **Project – POE Media data Analysis**

**Ogilvy & Mather, US. (Permanent)**

**Description:** POE Media Data analysis projects about campaign data based on the data analysis of media data in huge amounts and to develop the best throughput method to analyze the data and get the required result which is expected.

### **Job Responsibilities:**

- Good exposure on the Cloudera Hadoop cluster.
- Implemented Proofs of Concept on Hadoop stack and different big data analytic tools, migration from different databases (i.e. Teradata, Oracle, MYSQL) to Hadoop.
- Developed MapReduce programs to parse the raw data, populate staging tables, and store the refined data in partitioned tables in the EDW.
- Created reports for the BI team using Sqoop to export data into HDFS and Hive.
- Writing MySQL scripts according to the client requirements for business analytics.
- Created Hive queries that helped market analysts spot emerging trends by comparing fresh data with EDW reference tables and historical metrics.
- Administrator for Pig, Hive installing updates, patches, and upgrades.

Analyzed the data by performing **Hive queries** and running **Pig scripts** to know user behavior.

**DEC 2013 – MAY 2016**

**Hadoop-2.7.0,HDFS,HIVE, SQOOP, PIG,**

**FLUME**

**Project Name: Online Shopping**

**Primo-logic System Pvt. Ltd. (Permanent)**

**Description:**

Tapiture is one of the leading online shopping in the US. I am involved as a Hadoop Developer with a team size of 6. Here we got TB's of data which is generated from an E-Commerce portal. We are using Hadoop to store and analyze that data into a single place like HDFS and process that data by using Hive to get the Price and Product Info. Once we got the Info we need to share that Info for further Analysis by using Tableau.

**Responsibilities:**

- Moved all crawl data flat files generated from various retailers to HDFS for further processing.
- Involved in importing and exporting data from MySQL to HDFS using SQOOP.
- Writing Hive complex queries to create the table and process data in the Hadoop File System.
- Creating Hive tables partitions, loading with data, and writing hive queries which will internally in Map-Reduce way.
- Written the Apache PIG complex scripts to process the HDFS data.
- Created External Hive Table on top of parsed data increasing the performance of HiveQL.
- Created Hive tables to store the processed results in a tabular format using Excel.

### **Educational Background**

Year	Degree	Name of the University
2009-13	B. Tech (Bachelor of Technology)	Maharishi Dayanand University, Rohtak, Haryana
2008	Higher Secondary	Board of School Education Haryana
2005	Secondary	Board of School Education Haryana

### **Personal Details**

Category	Details
Sex	Male
Notice period	30 Days
Nationality	Indian
Language known	Hindi, English