# Vinaypal Waghmare

**Associate Consultant | Centtra Solutions Pvt. Ltd.**
Phone: +91 8407957956
Email: waghmarevinaypal@gmail.com

## SUMMARY

- **3.3** years of experience in IT industry including technical proficiency in big data environment with extensive expertise on Hadoop, Hive, Sqoop, spark ecosystem and Python.
- Experience with working on cloud infrastructure Amazon Web Services (AWS).
- Experience in launching EMR cluster, RDS, EC2 instances, S3 Bucket.
- Good understanding and knowledge of Hadoop architecture and Hands on experience with Hadoop components such as Name Node, Data Node and Map Reduce concepts, Spark Execution Concepts and HDFS Framework.
- In-depth understanding of Spark Architecture including SparkCore, SparkSql, and Data frames.
- Experience in working with Hive data warehouse tool -creating tables, data distribution by implementing Partitioning and Bucketing, writing and optimizing the HiveQL queries.
- Knowledge on Map Reduce Programs and using Apache Hadoop Java API for analyzing the data.
- Worked on data processing and transformations and actions in spark by using Python (Pyspark) language.
- Hands-on Core Python, NumPy, Pandas and Object-Oriented Programming (OOPS) using python.
- Experience on analysis of data using Python. Preprocessing, cleaning, aggregating, visualizing the data using different python libraries.

**Technical Skills:**

| | |
|---|---|
| **Big Data Platform** | Hadoop, Hive, Sqoop, Spark |
| **Programming** | Python, SQL, Linux Shell script, Java, C |
| **AWS Services** | EC2, S3, EMR, RDS, Databricks |
| **Operating Systems** | Linux (CentOS), Windows |
| **RDBMS** | MySQL Oracle, |
| **Data Analysis** | Python: Pandas, Matplotlib, Seaborn |

## PROFESSIONAL EXPERIENCE:

**Company: Centtra Solutions Pvt. Ltd.**                                    **May 2020 – Present**
**Role: Associate Consultant**

**Description:** Worked on a Financial service data analysis project. They were having huge volume of data and they had required fast and upscaled technology to process the data. The traditional technology stack was expensive to scale and not flexible to support this work.

**Technologies:** MySQL 5.1, Hive 1.1, Spark 1.6 (PySpark), Amazon S3, EMR, Linux

**Responsibilities:**
- Attend the daily scrum meetings and get the requirements for the project
- Get the data from S3 bucket for processing.
- Involved in loading data from Amazon S3 to HBASE and Hive environment i.e., HDFS using Spark.
- For that like we have built a data ingestion framework, our data ingestion framework support three ingestion pattern
  - Database ingestion,
  - File-based Ingestion,
  - Incremental data load.
- Implement Data Validation, Quality Checks.
- Involved in creating tables, loading data into hive tables from HBase using Spark.
- Involved into importing data from HBase to Hive staging table and clean the data and eventually load it to the Hive Final Table using Pyspark.
- Load the processed final data to Amazon S3.
- Perform UAT of implementation.
- Engage with BA to understand the requirement clearly.
- Monitoring logs and troubleshooting the issues occurred during the process.


**Company: Centtra Solutions Pvt. Ltd.**                                    **July 2019 – May-2020**
**Role: Associate Consultant**

**Description:** Worked on Healthcare and Insurance Domain project. In this project we have developed a robust pipeline health and insurance company data.

**Technologies:** HDFS 2.7.2, Sqoop 1.4.5, Hive 1.1, Airflow, Linux

**Responsibilities:**
- Import Data into the Hive from various Relational Databases (Oracle, MySQL) using Sqoop.
- Write Hive DDL to Create Hive Table for Optimize Query Performance.
- Ingest Flat Files like Delimited, etc. into Hive Warehouse.
- Engage with BA to understand the requirements clearly.
- Comes up with missing scenarios and get them clarified with BA.
- Implement Data Validation, Quality Checks.
- Perform UAT of Big Data implementation.
- Involved in Collecting Business Requirements from Business Users, Translate into Technical Design (Data Pipelines and ETL workflows).
- Involved in import data from various RDBMS into HDFS using Sqoop which includes Incremental

Load to populate Hive External Table and vice-versa.
- Involved in import data from various RDBMS into Hive Tables which includes Queries using Sqoop.
- Designed both Managed and External Hive Tables and Defined static and dynamic partitions as per requirement for optimized performance on production datasets.
- Worked with various File Formats like Text File, ORC Files, Avro Files and various Compression Formats like Snappy.
- Written Hive Queries for Data Analysis to meet the business requirements.
- Involved in developing shell scripts and automated data management from end-to-end integration work.

**Qualification:**

| Education | University | Year |
| --- | --- | --- |
| Bachelor of Engineering | Savitribai Phule Pune University, Pune | 2018 |
| Diploma | MPGI's Vishwabharti Polytechnic, Nanded | 2014 |
| SSC | Oxford English Medium School, Nanded | 2010 |

**Languages:**
Hindi, English, Marathi

**Strengths:**
Logical Thinking, Quick Learner, Team player

**Hobbies:**
Cooking, Cycling, Gardening.