

## JAYRAJ PANDURANG KAKDE

**Phone:** +91-9158046087, **Email:** [jayrajkakde12@gmail.com](mailto:jayrajkakde12@gmail.com)

---

### CAREER OBJECTIVE

---

To work in a challenging environment demanding all my skills and efforts to explore and adapt myself in different fields and realize my potentials where I get the opportunity for continuous learning technologies.

---

### SOFTWARE SKILLS

---

- ♦ **Hadoop Platform:** Map-Reduce, Hive, Spark, Spark-SQL, Pyspark.
- ♦ **Desktop Operating Systems:** Linux(Ubuntu), Windows Family.
- ♦ **Languages:** Python, SQL.
- ♦ **Databases:** MYSQL, SQLServer.
- ♦ **AWS Hadoop Services:** S3, EMR, RDS, EC2.

---

### PROFILE SUMMARY

---

- ♦ Overall industrial experience of 3.2 years and 2 years of extensive experience in IT industry mainly in Hadoop Ecosystem Developing and PySpark and Amazon Cloud Service (AWS).
- ♦ Hands on experience with Big Data core components and Eco System including Data Ingestion and Data Processing (PySpark, Hive, HDFS)
- ♦ High level knowledge of Apache Spark & Spark.
- ♦ Capable of processing large sets of structured, semi-structured data.
- ♦ Perform tasks such as writing scripts, SQL queries.
- ♦ Have a hands-on experience on Amazon web services mainly S3, EC2, AWS EMR, RDS.
- ♦ Develop batch processing, integration solution and process structured and Semi- structured data.
- ♦ Experience in transferring data from RDBMS to HDFS and Hive.
- ♦ In-depth understanding of Spark Architecture including SparkCore, SparkSql, Dataframes.
- ♦ Gather and process raw data at scale and analyze processed data.
- ♦ Very Good understanding of Partitioning, Bucketing concepts in Hive and designed both Internal and External tables in Hive to optimize performance.
- ♦ Strong problem solving & technical skills coupled with clear decision-making.

---

## WORK EXPERIENCE

---

**Company:** Brain Vision Technology

**Designation:** Associate Consultant.

**Period:** Sept 2019 – Present.

**Location:** Pune

---

## PROJECT

---

**Project : Telecom Business Intelligence Unit**

**Project Domain - Telecommunication**

**Duration:** April 2021 – Till Date

**Role:** Associate Consultant

**Description :** The Purpose of The Project is to do the ingestion of the Telecom data that is present on cloud storage as well as Clients Database to Data pipeline which was modeled as per the client. It is Cloud to HBase ingestion Project. Data Migration Utility to migrate existing system into Data hub which was built on top of Hadoop and uses spark for operations.

**Tools & Technologies:** Hadoop Eco-system, Pyspark, Python, S3, RDS(MySQL), EMR, HBase, Hive.

### Roles & Responsibilities:

- Understand business process and draft business Requirement document based on functional design documents.
- Worked on design and implementation of PySpark based framework.
- We have to load data and perform multiple operations in order to get meaningful insights that will help the client in decision making and analysis.
- Maintain existing ETL workflows, data management and data query components
- Develop automation and data collection frameworks
- Develops innovative solutions to Big Data issues and challenges within the team
- Involve in creating Hive tables, loading with data and writing Hive queries.

**Project : Data Analytics for Retail Domain**

**Project Domain- Retails**

**Duration:** Sept 2020 to March 2021

**Role:** Associate Consultant

**Description:** The Purpose of the project is to analyze the data provided using big data tools. To maximize their benefits, retailer needs to start changing their views regarding transactional data and start looking at it more from retailer's perspective.

**Tools & Technologies:** Spark, Python, SQL, S3 storage, AWS, Databricks.

**Roles & Responsibilities:**

- Experience in developing spark applications using spark-SQL in databricks for data extraction, transformation, aggregation from multiple file formats for analyzing and transforming the data.
- Extract Transform and Load data from sources system to AWS Data Storage Services.
- Responsible for estimating Cluster size and monitoring of the spark databricks cluster.
- Ability to apply the spark DataFrame API to complete the data manipulation within spark session.
- We used for processing as Jupiter-Notebook, spark-sql and for storage AWS S3.
- At the end uploading the output data to output directory on S3.

---

**EDUCATION**

- BE from University of Aurangabad in Mechanical Engineering with the percentages of 68.20 %.
- Diploma from MSBTE, Pune in Mechanical Engineering, Shreeyash Polytechnic, Aurangabad.

---

**PERSONAL DETAILS**

**Name** : Mr. Jayraj Pandurang Kakde  
**Date of Birth** : 12th April, 1996  
**Gender** : Male  
**Languages Known** : English, Hindi and Marathi.

---

**DECLARATION**

I hereby declare that the above-mentioned information is correct up to my knowledge and I bear the responsibility for the correctness of the above-mentioned particulars.

**Place:** Pune

**Date:** / /

Jayraj Pandurang Kakde