

## Candidate Summary

<b>PROFILE ID</b>	50834
<b>SUBMISSION ID</b>	2405S22293
<b>LEAD TIME TO START</b>	1 Month
<b>TOTAL EXP (Years)</b>	3+
<b>RELEVANT EXP (Years)</b>	3+
<b>SKILLS</b>	Big Data, Data Ingestion, Apache Kafka, Operating Systems, Ecosystem, Analytics, Linux, JIRA, Cleaned Data, Python (Programming Language), MapReduce, Amazon S3, Transformation (Genetics), Loans, Query Performance, SQL (Programming Language), Data Extraction, Sqoop, Revenue Growth, Pyspark, Extract Transform Load (ETL), AWS Elastic MapReduce (EMR), File Format, Databricks, Hadoop Distributed File System (HDFS), Data Processing, Execution Time
<b>PROFILE SUMMARY</b>	<ul style="list-style-type: none"> <li>■ Big Data Engineer with 3+ years of excellent track record of designing and implementing efficient ETL processes.</li> <li>■ Cloud Technologies as AWS (S3, EC2, IAM, VPC, CloudWatch, EMR, AWS Glue, Redshift, Athena, Lambda), Azure (Azure Databricks, ADLS Gen2, BLOB, Azure Data Factory, Azure key vault)</li> </ul>

# Data Engineer

## SUMMARY

---

- Skilled Big Data Engineer with 3+ years of excellent track record of designing and implementing efficient ETL processes.
- Proficient in Python, SQL, Hadoop, PySpark and cloud-based data platforms.
- Masters in Collaborating with cross-functional teams to deliver high-quality solutions saving project costs by around 20%.

## SKILLS

---

### TECHNICAL SKILLS:

- Big Data/Hadoop ecosystem - Hadoop, Spark, PySpark, HDFS, HIVE, Sqoop, Oozie, YARN, Cloudera, MapReduce, kafka, ELT and ETL
- Cloud Technologies - AWS (S3, EC2, IAM, VPC, CloudWatch, EMR, AWS Glue, Redshift, Athena, Lambda), Azure (Azure Databricks, ADLS Gen2, BLOB, Azure Data Factory, Azure key vault)
- Programming Language - Python and SQL.
- Operating System - Linux (RHEL 7.3), Windows.
- Other Skills – Ticketing Tool (Jira), Git, Agile, CI/CD
- Big Data 101 By IBM Languages
- English • Hindi • Marathi

## QUALIFICATIONS

---

- 2018 - 2021 Bachelor of Technology ( Maharashtra Institute of Technology )
- 2015 – 2018 Diploma ( Marathwada Institute of Technology Engineering College )

## PROFESSIONAL CERTIFICATIONS

---

- Programming Language Python and SQL

## ORGANIZATIONAL EXPERIENCE

---

- May 2022 – Present Big Data Engineer - ( Bravezone Softech ) ( Pune, Maharashtra, India )
  - Designed and managed data pipelines in Azure Data Factory for Azure Data Lake Storage Gen2, resulting in a 40% increase in data processing effectiveness.
  - Leveraged Azure Databricks to optimize data processing and analytics, reducing processing time by 50%.
  - Improved query execution time by 30% through PySpark optimization techniques like partitioning and bucketing.
  - Engineered Databricks optimization strategies, resulting in a 40% enhancement in query performance.
  - Utilize broadcast join optimizations in Databricks, achieving a 16% improvement in query performance by minimizing data shuffling and optimizing data distribution.
  - Executed data storage optimizations using the vacuum command in Databricks, reducing storage overhead by 30%.
- March 2021– April 2022 Big Data Engineer - ( Bravezone Softech ) ( Pune, Maharashtra, India )
  - Designed and implemented effective database solution (Azure Blob Storage) to store and retrieve data.
  - Deployed Azure Data Factory for creating Pipeline to orchestrate the data in SQL Database.
  - Evolved Spark Application using Pyspark and Spark SQL for data extraction, transformation, and aggregation from multiple file format for analyzing & transforming data to uncover insight into the customer usage patterns.
  - Proficient in working with file formats like CSV, Parquet, Avro, and ORC, reducing storage costs by 22%.
  - Enhanced performance by loading processed data into Spark DataFrames, implementing partitioning and bucketing techniques, and selecting efficient file formats such as ORC and Parquet, resulting in an 18% speedup in data processing.

## LANGUAGES

---

- Hindi
- English
- Marathi