

# Ajay K

📍 : Bengaluru, 560029, INDIA, IN,

☎ : +91-9542380365, ✉ : kadivalaajay8@gmail.com,

🌐 <https://www.linkedin.com/in/ajay026/>

🔗 : <https://github.com/Ajay026>.

## Profile Summary:

- Over 6+ Years of overall IT experience in Application Development.
- Working experience in **Hadoop** ecosystem (Gen-1 and Gen-2) and its various components such as **HDFS**, Job Tracker, Task Tracker, Name Node, Data Node, Resource Manager (YARN).
- Experience with components such as **Cloudera** distribution encompassing components like MapReduce, **Spark**, **SQL**, **Hive**, **HBase**, **Sqoop**, **Pyspark**.
- Good skills on **NoSQL** Database- **HBase**.
- Knowledge in **Data Warehousing** Concepts in **OLTP/OLAP** System Analysis and developing DatabaseSchemas like **Star Schema** and **Snowflake Schema** for Relational and Dimensional **Modelling**.
- Good hands on in creating custom **UDF's** in **Hive**.
- Load and transform large sets of structured, semi-structured and unstructured data from Relational **Database** Systems to HDFS and vice-versa using **Sqoop** tool.
- Good Experience on architecture and components of Spark, and efficient in working with **Spark Core**, **Data Frames/Data Sets/RDD API/Spark SQL**, **Spark streaming** and expertise in building PySpark and Spark-Scala applications for interactive analysis, batch processing and stream processing.
- Hands-on experience in Spark, **Scala**, **SparkSQL**, Hive Context for Data Processing.
- Experience on **Azure** cloud i.e., **ADF**, **ADLS**, **Blob Storage**, **Databricks**, **Synapse**, **CI/CD pipelines** etc.
- Extensive working experience in an Agile development Methodology & Working knowledge on Linux.
- Expertise in working with big data distributions like **Cloudera** and **Hortonworks**.
- Experience in working with **Hive optimization** techniques such as **partitioning**, **bucketing**.
- Experience in tuning and debugging Spark application and using **Spark** optimization techniques.
- Expertise in developing batch data processing applications using **Spark**, **Hive** and **Sqoop**.
- Experience in working with **CSV**, **JSON**, **XML**, **ORC**, **Avro** and **Parquet** file formats.
- Limited experience in creating and designing data ingest pipelines using technologies such as **Kafka**.
- Good knowledge in working with **ETL** methods for data **extraction**, transformation and loading incorporate-wide ETL Solutions and Data Warehouse tools for **reporting** and **data analysis**.
- Basic Experience in implementing **Snowflake** Data Warehouse.

## TECHNICAL SKILLS:

- **Big Data Technologies:** Hadoop, Spark, Hive, Sqoop, Kafka, PySpark, HBase
- **Cloud Technologies:** Azure (Azure Storage, Azure Synapse, ADF, Azure Data Bricks).
- **Languages:** Scala, Python, SQL
- **Databases and Tools:** Oracle, MySQL, SQL, NoSQL.

- **Platforms:** Windows, Linux.
- **IDEs:** Eclipse, Cloudera, Hortonworks.
- **Scheduling:** Airflow.
- **Project Management Tools:** Jira, GitHub.

## **Certifications:**

- Completed Microsoft Azure-Fundamentals (AZ-900).
- Completed Microsoft Azure Data Fundamentals (DP-900)
- Completed Microsoft Azure Data Scientist Associate (D100).
- Completed Microsoft Azure Power BI Data Analyst Associate (PL-300).
- Completed Microsoft Azure Data Engineer Associate (DP-203).
- Completed Microsoft Azure Developer Associate (AZ-204)
- Snowflake Pro Core Certification (COF-C02)

## **PROFESSIONAL EXPERIENCE:**

**Price Waterhouse Coopers Pvt Ltd:** Oct 2021 – Till date

**Client:** EDP Migration

**Remote**

**Role:** Big data Consultant

## **Responsibilities:**

- Responsible for building scalable distributed data solutions using Spark.
- Ingested log files from source servers into HDFS data lakes using Sqoop.
- Developed Sqoop Jobs to ingest customer and product data into HDFS data lakes.
- Developed Spark streaming applications to ingest transactional data from Kafka topics into Cassandra tables in near real time.
- Developed an spark application to flatten the transactional data coming from using various dimensional tables and persist on Cassandra tables.
- Involved in developing framework for metadata management on HDFS data lakes.
- Worked on various hive optimizations like partitioning, bucketing, vectorization, indexing and using right type of hive joins like Bucket Map Join and SMB join.
- Worked with various files format like CSV, JSON, ORC, AVRO and Parquet.
- Developed HQL scripts to create external tables and analyze incoming and intermediate data for analytics applications in Hive.
- Optimized spark jobs using various optimization techniques like broadcasting, executor tuning, persisting etc.
- Responsible for developing custom UDFs, UDAFs and UDTFs in Hive.
- Analyze the tweets json data using hive SerDe API to deserialize and convert into readable format.
- Orchestrating Hadoop and Spark jobs using Oozie workflow to create dependency of jobs and run multiple Jobs in sequence for processing data.

**Tech stack:** Scala, Hadoop, Spark, Spark SQL, Spark Streaming, Hive, Cassandra, MySQL, HDFS, Apache Kafka.

**Client:** Trimble

**Project:** Migration

**Remote**

**April 2022 – June 2022**

**Role: Data Engineer**

**Responsibilities:**

- Working with Structured data that is being ingested into Azure File storage explorer.
- Create ETL pipeline in Snap logic tool to bring the data into azure Databricks workspace.
- Applied transformation logic including, Spark sql, pyspark operations on data.
- Applied optimizations logics i.e., Partitioning, broadcast joins etc.
- Create ETL pipeline on Databricks transformed data to dump the target directory called Salesforce.
- Analyze the resultant data with data bricks tool.
- **Tech Stack**: ETL, Databricks, azure, Snap logic, Salesforce.

**Accenture Solutions Pvt Ltd:** Feb 2019 – Oct 2021

**D-VOIS Communications Pvt Ltd:** Nov 2017 – Feb 2019

**Side Projects:**

- **Link to Repository:** <https://github.com/Ajay026>

**Big Data Technologies:** Hadoop, Spark, Hive, Sqoop, Kafka, PySpark, HBase.