



# AJAY KUMAR

## DATA ENGINEER

 vajay8679@gmail.com

 +91 7247593467

 Indore, Madhya Pradesh

### EDUCATION

#### IPS ACADEMY INDORE

JUL 2014 - JUN 2018  
BACHELOR OF ENGINEERING  
(MECHANICAL) - CGPA 6.9/10

#### JNV CHHINDWARA

CLASS XII - JUNE 2013  
81% (PCM)

### TECHNICAL SKILLS

#### PROGRAMMING LANGUAGES

Python (Pandas, Numpy), Php

#### BACKEND FRAMEWORKS

Codigniter, Laravel, Django

#### WEB DEVELOPMENT

HTML, CSS,  
BootStrap, JavaScript

#### DATABASES

SQL, DBT, MongoDB

#### BIG DATA FRAMEWORKS

Spark, Hadoop, Hive, Kafka,

#### SCHEDULERS & AWS

Airflow, S3, EMR, EC2, Glue,  
Lambda, Cloudwatch, Athena

#### CONTROL SYSTEMS & DOCUMENTATIONS

Git, BitBucket, Slack, Zoho,  
Notion

### WORK EXPERIENCE

#### Ineuron - DATA ENGINEER INTERN AUG 2022 - PRESENT

##### Data Pipeline-Scheduled-Based

- Created scheduled based Data Pipeline on AWS with the help of EventBridge connecting with S3, Glue Crawler and ETL-Job using Workflow.
- We can schedule the Job through Event Bridge and Workflow will run accordingly and we can see output through Athena where we can run different query.

##### Data Pipeline based on Messaging Using PySpark and Airflow

- The speed of getting data pipelines up and running, Getting new engineers up to speed, Complexities
- Spending less time worrying about the service management, Providing business value, Overall engineering cost

##### Hadoop-Pyspark-Movie analytics-Project

- The main goal of the project is to Analyze the data from various data sources and extract meaningful insights from the data and to know the audience interests. the data
- will be present in the hdfs storage layer in csv format with delimiter(::).

We will be reading data from Hdfs layer and process the data using pyspark.

##### ETL Data Pipeline on AWS EMR Cluster

- Created EMR Cluster, Created S3 Bucket to store raw and transformed data.
- Inside EMR in hive created tables and performed casting and filtration and send data to final table and then we can visualize transformed data with the help of tableau.

##### Pyspark Basic Operations

- Different Pysparks operations performed with the help of Hadoop HDFS File System. All Crud operations like Modification, Drop, Create, Select with Different operations With Different Joins.
- Work with Different conditional operations With PySpark.

#### DigiPrima Technologies - SOFTWARE ENGINEER MARCH 2021 - PRESENT

##### Health Care Project - ID CARE

- ID Care is NJ's largest network of board-certified infectious disease specialists. With over 50 certified physicians onboard, ID care delivers customized solutions to each patient.
- Clinical leaders, researchers, and expert physicians focus on identifying, treating, and preventing infectious diseases.

##### SurveyJS

- SurveyJS is a set of JavaScript components that allow you and your users to build surveys, quizzes, polls, and other web forms, store them in your database.
- Visualize survey results in custom dashboards.

#### Vinti Web Solutions - WEB DEVELOPER JAN 2020 - FEB 2021

- Worked as Web Designer as a beginner and work on few projects.
- Worked on Different minor projects and helped in some major projects as Web Developer. Work with HTML, CSS, JS, BOOTSTRAP & PHP