

# PAVAN YADAV

Maa Durga PG, 4th Street, Vpg Ave Extn. Rd, Mettukuppam, Chennai, TN  
Ph: 9411060791 | Email: raj.pavan034@gmail.com  
LinkedIn: www.linkedin.com/in/pavan034

## Professional Summary:

---

- Experience in IT industry for implementing, developing and maintenance of various applications by using Spark, Scala, Hadoop Ecosystem, Kafka, Elasticsearch and Kibana.
- Good Knowledge on Hadoop Architecture and Daemons such as HDFS, Hive and Spark.
- Having good experience in configure Hadoop, Kafka, Spark, Elastic search clusters.
- Having experience in developing a data pipeline using Kafka to store data into Elasticsearch, mysql, HDFS.
- Strong experience on setting up Mongo DB Shard Cluster, Hadoop Cluster, Cassandra cluster with Docker on Linux Environment using own Dockerfile and docker-compose.yml file.
- Strong experience on build the docker for any kind of application and deploy.
- Hands on experience to capture the Google Analytics data.
- Hands on experience to create pipeline from Google BigQuery and MySQL to MSSQL Server.
- Used Scala SBT to develop Scala coded spark projects and executed using spark-submit.
- Good experience in writing Spark applications using Python and Scala.
- Having good experience in collecting data from external sources using Confluent Kafka
- Good Analytical skills with good communication, presentation and goal-oriented team player. Knack for thriving in high pressure environments managing multiple task while meeting the tight deadlines.
- Great motivation to learn new skills/technologies, excellent analytical/problem-solving skills, fast-learner, resourceful, committed, hard-worker, and self-initiative

## Professional Experience:

---

### Scala Developer

Aaum Research and Analytics Private Limited, Chennai (Aug 2018 – Present)

#### Responsibilities:

- Create and maintain optimal data pipeline architecture.
- Drive continuous improvements of products and processes.
- Automating manual processes, optimizing data delivery, re-designing infrastructure for greater scalability.
- SparkSql to aggregate the data.
- Setup and use MongoDB and MongoDB Shard Cluster.
- Responsible for creating, managing and monitoring the Hadoop Cluster.
- Responsible for setting up and maintaining Docker containers over Linux environment.

## Technical Skill:

---

**Languages:** C#, Scala, Java, Python, ASP. Net Core, Android, PHP, HTML, CSS

**Database:** MongoDB, MySQL, MsSql Server, Oracle, Postgres

**Hadoop Components:** HDFS, Hive, Oozie, Kafka, Spark, Elasticsearch, Kibana, Sqoop

**Softwares:** Visual Studio, Docker, NetBeans, Eclipse, Andriod Studio, R Studio, Gitlab

**OSs:** Windows 7-10, Cent OS, Linux(Ubuntu , Fedora, RedHat & Linux Mint)

## Education:

---

Vellore Institute of Technology, Chennai, Tamil Nadu  
Master of Computer Application (CGPA – 8.67)

Jul, 2016 –May, 2018

Dr. BhimRao Ambedkar University, Agra, Uttar Pradesh  
Bachelor of Computer Application (Percentage – 73.6%)

Jul, 2010 –May, 2013

## Projects:

---

➤ Project Name: **GPI Flash**

Technologies: Scala, Spark, Postgres DB

Responsibilities:

- Developed code using Spark Scala to read the bulk data (CSV files) from Amazon S3 bucket.
- Performed aggregation using join columns and stored the result into Postgres DB.

➤ Project Name: **IOT Analytics**

Technologies: Java, Scala, Kafka, Elasticsearch, Kibana

Responsibilities:

- Developed code using Spark Scala to generate and send data to Kafka.
- Used Connector to send the live stream data from Kafka to Elasticsearch.
- Used Kibana to see the graphical view of the live stream data.

➤ Project Name: **Project Monitoring Review Committee**

Technologies: C#, ASP.Net, MVC, Oracle DB

Responsibilities:

- Web Development for monitoring & taking reports of the DRDO Project.
- Worked on Oracle Database to save the data.
- Developed Web service for backend using C#.
- 

➤ Project Name: **B.TECH (an Ecommerce site) BI**

Technologies: Python, SQL, MSSQL Server, MySQL, BigQuery, MS Analysis service, Tabular model, Powerbi

Responsibilities:

- Capture the data from the external sources (like: Google Analytics, funnal, adjust, compartra, RMS) and push into Google BigQuery using python scripts and connectors.
- Create a connector to pull the BigQuery and MySql data to SQL Server using linked server.
- Create a tabular model using Visual Studio, import the data sources to tabular model and deploy to SAAS
- Write the SQL refresh script into SQL Agent to refresh the data in tabular model.
- Create a on-primsis Gateway and configure in Powerbi Portal to use the data sources in powerbi reports.

## Awards & Recognition:

---

- Rising Star Award for the month April 2019 in the Project.
- Appreciation & Cash Award of 15K for the DRDO Assignment.

**Hobbies:** Singing, Playing Indoor games

**Languages:** English, Hindi

I hereby declare that the information furnished above is complete and true to the best of my Knowledge.

Date: 16th June, 2019  
Place: Chennai

PAVAN YADAV