

Amit Sharma

Data Engineer

From: Kurukshetra, Haryana
Mobile: +91 9416042840
Mail: amitsharmajir@gmail.com
LinkedIn: <https://www.linkedin.com/in/amitsharma-kkr>

Summary

Professional Data Engineer with 5+ years of industry experience including around 4+ years of experience in Big Data technologies. Expertise in Spark/Hadoop development experience, automation tools and E2E life cycle of software design process. Good communication skills, dedicated to maintain up-to-date IT skills and industry knowledge.

Technical Skills

- Languages: Scala, Python, Java
- Frameworks: Apache Spark
- OS : Windows, Linux
- Softwares: IntelliJ, Apache Zeppelin, Jupyter
- Databases: Cassandra, MySQL, MongoDB, Solr
- HDFS (Apache Hadoop)
- Amazon EC2, EMR, S3 (AWS)

Certifications

- Microsoft Certified Professional
- Microsoft Technology Associate
- IBM DB2 Academic Associate

Career History

Data Scientist

Spineor Webservices Pvt Ltd

March 2017 - Present

Mohali, Punjab

1. Project: Auto Parts Catalogue (Big Data Processing)

Description: Created an Auto Parts catalogue from raw unstructured data. It involved ETL process using Apache Spark and Scala to clean and preprocess data so that it be prepared for analysis. The catalogue thus formed, was also used in constructing data feeds and sitemaps for the auto parts e commerce portal.

Technologies used: Apache Spark, Python, Scala.

2. Project: Data feed and sitemaps

Description: Created Google and Bing data feeds and their corresponding sitemaps for the auto parts sale. Also handled the Google Content API which involved the updation, modification, addition and deletion of the millions of auto parts products using batch processes.

Technologies used: Apache Spark

3. Project: A clustering approach to analyze auto parts sales

Description: Established a clustering paradigm to boost the auto parts products sale based on their spatial co-ordinates. The Euclidean distance was used as the distance norm for DBSCAN. The arbitrary clusters thus formed presented a well-defined division of sales in order to analyze the sale pattern of autoparts in various regions.

Technologies used: Java, Apache Spark

4. Project: Building Automation System

Description: Created backend for building automation system which consists of various APIs including management, real time streams and query engine(using websockets). The management api data is stored in mongodb using reactivemongo driver. The stream comes from IOT devices which are dumped into Kafka and then stored into elasticsearch using kafka connect api. The Http server is built using akka http.

Technologies used: Scala, Kafka, Akka http

Education

• **B.Tech** (Computer Science & Engineering) with **66.5%** from Kurukshetra University.

• **12th** with **aggregate 76%** from S.M.B. Gita Sr. Sec. School (H.B.S.E) in 2012.

• **10th** with **aggregate 88.2%** from S.M.B. Gita Sr. Sec. School (H.B.S.E) in 2010.