Ankit Jhunjhunwala

Gurgaon, Haryana
8513804966
Akijhunjhunwala63@gmail.com

Summary

Big Data Engineer with a demonstrated history of working in the Information Technology industry. Skilled in Technologies like Spark, Python, SQL, Hadoop, Hive. Possessing strong knowledge of ETL & Data Warehousing concepts & experienced in applying data driven solutions to increase accuracy and efficiency of data processing. Skilled in extracting meaningful data, analysis & visualization of data, creating predictive data models using machine learning algorithms to deliver insights and implement action-oriented solutions to complex business problems.

Skills

Languages: Python, Java, Scala, SQL

Databases: Microsoft SQL Server, IBM Netezza, Oracle Exadata

Operating Systems: Linux, Windows, Unix

Big Data Technologies: Hadoop, Hive, Spark, Sqoop, Pentaho, SSIS, Zookeeper, Airflow

Cloud Service: GCP

Version Management: GitLab, BitBucket

Experience

Senior Data Engineer | Dunnhumby | Nov'20 - PRESENT

Clients: Monoprix, Coop-Norway, SOK-Finland | Domain: Retail

- Build, enhance and operate efficient and compliant data platforms to deliver high quality production data.
- Import client data from Axway to GCP and loading the data into datalake in the form of dimension and fact tables containing raw data.
- Transformation on datalake tables by using PySpark and loading them into mapped staging layer.
- Publish the data into final MART layer by adding surrogate keys for each entity.
- Load the data into time-series DB to visualize the quality of data using Grafana & Influx db.
- Validation of data before it gets loaded into each layer.
- Add segmentation on generated MART data as per the rules specified by the client.
- Export Segmentation & MART data into Oracle Exadata Warehouse, transform the data & load into final warehouse table.
- Establish connection between the final warehouse table & dhShop product from where customers can see their data.
- Automate the entire data workflow and orchestration through Apache Airflow.
- Apply data governance & retention policies on historical data.
- Write optimized queries using SQL, Pyspark to analyze the datasets & get valuable insights from data keeping track of effective cluster utilization for cost reduction.
- Process enhancement & automation of some existing functionalities.

System Engineer | Tata Consultancy Services | July'18 – Nov'20

Data Engineer (Feb'20-Nov'20) | Client: Bank of Montreal
Data Pipeline Creation & Building Machine Learning Models using Spark

- In-depth understanding of Spark architecture and its components such as Spark-SQL and Spark-MLlib.
- Created data pipelines and generated the data for the input of Machine Learning models incorporating the required transformations on data. Used Alation for data governance.
- Used Spark-SQL to load CSV, JSON, TXT, PARQUET data and created schema RDD's and loaded it into Hive Tables in HDFS and handled structured data.
- Written optimized queries on the Data Frames using Spark-SQL to analyze the dataset and to get valuable insights from it.
- Used classical machine learning algorithms like regression, clustering, and random forest from Spark's MLlib library to predict valuable business insights.
- Created reports on the predicted values.

ETL Developer (Apr'19-Jan'20) | Client: Bank of Montreal
Data Migration from legacy RDBMS to IBM Netezza data warehouse

- Involved in requirement gathering, data modelling & creation of mapping documents.
- Responsible for transforming and loading data from legacy RDBMS to Netezza Data Warehouse.
- Created optimized Netezza Scripts to load data from flat files to Netezza data warehouse.
- SSIS package creation & execution for data integration.
- Validation & Testing of data through SSMS.

Hadoop Developer (Oct'18-Mar'19) | Client: Bank of Montreal
Lift & Shift Migration of Legacy Data Warehouse to a Concentrated Data Lake Environment Using Hortonworks HDP Cluster

- Responsible for transforming data from Netezza Data Warehouse to HDFS System using Hive warehouse.
- Created Hive schemas using partitioning, bucketing & compression techniques.
- Built Hive tables, loaded data into it and developed Hive queries that runs internally in map reduce way.
- Batch processing TB of data using Spark & Scala.
- Created Pentaho jobs, UNIX shell scripts and used it as a wrapper to call Hadoop components.
- Framework Execution and Validation of data through SSMS.
- Continuous Integration & Deployment through Bit bucket, Bamboo, Artifactory and Ansible tool.

Quality Assurance Intern | Tata Consultancy Services | Jul'17 – Aug'17
Android Automation Testing Using Xamarin UITest

- Worked within a team to develop different test cases to check the functionalities of the hybrid applications and to find the defects or vulnerabilities present in the application.
- Automation of all the test cases using Xamarin Test Recorder.

Education

Year	Degree/Exam	Institute	CGPA/Marks
2018	B.Tech	Institute Of Engineering & Management, Kolkata	8.76
2013	Higher-Secondary	Doon Public School, Dhanbad	82.4%
2011	Secondary	Gyan Bharati School, Raniganj	10.0 CGPA

Academic Projects

Facial Recognition System | July' 17 – May'18

Developed a facial recognition system using HOG (Histogram of Gradients) Algorithm & SVM (Support Vector Machine) classifier using our self-prepared datasets. It identifies all the faces in a live video & matches it with the faces that are already there in its trained datasets and provide results with an accuracy of more than 95%.

Grant Application Submission & Tracking System | Ministry of Ayush | Aug'17 – Mar'18

Worked within a team to develop a website for the submission of grant application and their tracking.
This application will act as a connecting bridge between an individual and the Ministry of Ayush in which one can upload their research protocol and apply for grant hassle-free and at the same time can track the real time status of submitted application. The official administrators can make changes to the application status as and when required. (Hosted by **Central Government of India**: https://aiia.gov.in/)

Awards & Achievements

- Got **'On the spot'** award for the outstanding performance in the team.
- Selected for Grand Finale of **Smart India Hackathon 2017**, organized by **MHRD** and secured **5**th rank and won the **Delloite Innovation Award** for developing Grant Application Portal & Tracking System for **Ministry of Ayush**.
- Ranked 143 out of 1 lakh+ candidates all over India in TCS Codevita - 2016.