Ankit Agrawal

Lead Consultant -
Data Engineering

Personal Info

Phone
+917906417818

E-mail
anki.ak27@gmail.com

Location
Bengaluru, India

Hard Skills

PySpark , MySQL, ETL, Python,

AI ChatBots, SparkSQL, Hive

PostgreSQL, Sqoop, Tableau

Scala, Informatica, Shell Scripting

Databricks, Delta Lake, ADF

Airflow, GIT, REST APIs

Automation Anywhere 360

Education

2011 – 2015: B.Tech
HCST, Farah, Mathura, UP – 73.7%

2010 – 2011: Intermediate
DPS, Aligarh, UP – 77.8%

2009 – 2010: High School
JMPS, Agra, UP – 83.4%

Hobbies & Interests

Cricket, Badminton, Travel & Trekking, Music

Languages Known

English

Hindi

An undeniably inquisitive and highly motivated data enthusiast. To utilize my technical experience as an Experienced Data Engineer with over 8+ years of exposure in Big Data Domain and functional knowledge across multiple formulations for the competitive aspirations of the organization. I look forward with the same interest, to apply my skills and knowledge in a progressive organization where growth prospects are unlimited, which would enhance and give ample exposure. Good communication and interpersonal skills with confidence to learn new technologies and undertake challenges.

Experience

2021- 09 - Present **Genpact Headstrong Capital Markets**

Lead Consultant

- ✓ Requirement analysis and design of the AI bots associated data processing pipelines.
- ✓ Ingesting data from multiple sources like **MySQL**, shared drives, salesforce to staging layer **HDFS**.
- ✓ Cleanse, transform data using **PySpark** to gain insights, visualize using Tableau dashboards.
- ✓ Optimizing other automation bots **SQL** queries into python pandas.
- ✓ Designed and developed approach to implement new features to improve chat forum.
- ✓ Performance tuning of code to enhance/improve the chats processing.

2020- 02 - 2021- 09 **Danske IT**

BigData Developer

- ✓ Fetching Data from mainframes to daily **hive** tables using bash script (having **sqoop** commands and other hive commands to create schema and other necessary details).
- ✓ Then run **pyspark** model of consolidated flow on airflow to consolidate data.
- ✓ It scheduled on **airflow** on daily basis and load the other control table to have proper statistics with proper data load.

2016- 09 - 2020- 02 **DXC Technology**

BigData Developer

- ✓ Used SQOOP to import vast data from traditional RDMS to HDFS. Involved in writing import query for the incremental data on a scheduled basis.
- ✓ Used Oozie to automate the Sqoop jobs.
- ✓ Worked on Hive queries for creating and querying HIVE tables to retrieve useful analytical information.
- ✓ Developing Scripts and Batch Job to schedule various Hadoop Program.
- ✓ Monitoring of Data Pipelines to ensure the thorough transfer of data.
- ✓ Developing Scripts and Batch Job to schedule various Hadoop Program.
- ✓ Written documentation to describe program development, logic, coding, testing, changes, and corrections.

2015- 11 - 2016- 09 **Hewlett Packard Enterprise**

Informatica Developer

- ✓ Extracted data from oracle database and spreadsheets and staged into a single place and applied business logic to load them in the central oracle database.
- ✓ Used Informatica Power Center 10.1.1 for extraction, transformation, and load (ETL) of data in the data warehouse.
- ✓ Extensively used Transformations like Router, Aggregator, Joiner, Expression and Lookup, Update strategy and Sequence generator and Stored Procedure.
- ✓ Developed complex mappings in Informatica to load the data from various sources. Implemented performance tuning logic on targets, sources, mappings, sessions to provide maximum efficiency and performance.
- ✓ Parameterized the mappings and increased the re-usability. Used Informatica Power Center Workflow manager to create sessions, workflows and batches to run with the logic embedded in the mappings.
- ✓ Used the PL/SQL procedures for Informatica mappings for process control in incremental load. Created the ETL exception reports and validation reports after the data is loaded into the warehouse database. Written documentation to describe program development, logic, coding, testing, changes and corrections.