Anurag Ambuja
Data Engineer

CCA cloudera
DATA ANALYST

Professional with around 10+ years of experience in Software Development, Automation & Bug Fixing in Retail, Networking and Banking Sector.
Currently working as Big Data Engineer with 5+ years of extensive experience in Big Data Ecosystem.

• <u>Linkedin</u> • <u>Medium</u> • <u>Github</u> • *Gurgaon, India* • *+91-8800442797* • *anurag.ambuja@gmail.com*

PROFESSIONAL EXPERIENCE

Lead Data Engineer, IHS Markit
09/2020 –   *Gurugram, Haryana*

Role & Responsibilities
- Requirement Gathering, Analysis and designing technical Solutions. Do effort estimation and planning
- Implement the business requirement on **BigQuery**

Lead Data Engineer, Dunnhumby
01/2019 – 09/2020   *Gurugram, Haryana*

Role & Responsibilities
- Analyse problem statement and Design/Modify Data Lake Solution for Transactional and Dimensional Data.
- Migration of existing structured data from RDBMS database, flat files or SAS dataset to **Hadoop/Hive** using Sqoop or Spark.
- Writing spark job to compute aggregated data for faster processing of data using **Pyspark**.
- Use Spark API over Cloudera Hadoop **YARN** to perform analytics data on Hive.
- Involved in converting Hive/SQL queries into Spark Transformation using Spark dataframes and **Python**.
- Parsing different file formats like xml, json and converting it into structured format using python.
- To develop Hive programs to cleanse and parse data in HDFS obtained from various data sources and load it in Hive tables.
- Creation of tables, data distribution by implementing partitioning and bucketing, writing and optimizing the HQL queries on Hive data warehouse tool.
- Automate the developed solution using Airflow
- Create **Grafana** Dashboard using InfluxDB for visualization
- Perform the POC; develop the code and test the code in UAT env & local env using **docker** container and promote and maintain the code using **Git** in Production.
- Thoroughly document release procedures and provide training to support team
- Write tests and verify product capabilities
- Provide training to new joiners

Senior Engineer (Bigdata, Spark, Python)
Dunnhumby
02/2017 – 12/2018   *Gurugram, Haryana*

IT Analyst (Bigdata, Python, Shell)
Tata Consultancy Services.
08/2014 – 01/2017   *Gurugram, Haryana*

System Engineer (Shell, SQL, Python)
Tata Consultancy Services.
08/2012 – 08/2014   *Gurugram, Haryana*

Associate System Engineer (Shell, SQL)
Tata Consultancy Services.
08/2010 – 08/2012   *Bangalore, Karnataka*

SKILLS, TOOLS & TECHNOLOGIES

Hadoop, **Python / Shell Scripting**, **Impala**, **SQL, Hive**, **Google Cloud/ Azure**, Sqoop, **Spark**, Docker, **Airflow**/Oozie/ControlM, **Grafana**, **Redis/InfluxDB**, Oracle, Make/JAM, **Git**/ACME

EDUCATION

B.Tech. (Computer Science and Engineering)
Cochin University (CUSAT), Kochi
08/2006 – 07/2010   *70 %*

Schooling
C.B.S.E, Delhi
1993 – 2005   *Bokaro, India*

CERTIFICATIONS

- *Microsoft Certified: Azure Fundamentals*
- *Cloudera Data Analyst, CCA 159 (license ID # 100-023-599)*
- *ITIL Foundation Exam, APMG In.*
  (Credential # 02999923-01-HQYQ)
- *Other Learning Certifications*

Honors

- *On the Spot Award – Clients (Deutsche Bank & American Ex)*
- *Appreciation Certificate – Client (American Ex.)*