

# Arnab Mondal



<https://github.com/arnabm14> | <https://www.linkedin.com/in/arnab1408/> | [arnavm014@gmail.com](mailto:arnavm014@gmail.com) | +91-6290729774

**Career Objective:** I want to build pipelines and create infrastructure which can ingest and transform Petabytes of Data per day. I want to revolutionize Big Data and how we perceive it so that it becomes a trivial solution for everyone.

## WORK EXPERIENCE

**phData**, Bangalore India, Remote Work

**June 2022 – Present**

### Data Engineer

- Created a migration plan and pipeline along with 3 of my colleagues to migrate 200 Terabytes of data.
- Exceeded the expectations and delivered within the time frame with proper documentation to the client.
- Analyzed and created a migration validation tool to ensure proper data transfer and prevent data loss.
- Designed the architecture and laid out the plan for the migration process and presented it to stakeholders.

**Skills Learnt:** Migration, Snowflake, DBT, Data Validation, Airflow, ETL, ELT

**AFour Technologies**, Pune India

**December 2021 – June 2022**

### Senior SDE: Data Engineering

- Developed and Created software that worked on Openstack and maintained production-level code.
- Worked with a team of 10 people and led a team of 4 to resolve bugs and push new features to production.
- Edited and created proper documentation for my code and KT videos to help future developers Leveraged.
- Created Glue Job on AWS using PySpark to ingest data into the RAW layer and follow the 3 stages of ingestion.

**Skills Learnt:** AWS Glue, S3, Athena, Airflow, Crawlers, Cross-Account, Openstack Ecosystem, ETL

**Tata Consultancy Services (TCS)**, Kolkata India

**July 2019 – December 2021**

### SDE: Data Engineering

- Worked intensively with Big data technologies like Impala, Spark, Kafka, Hive and HDFS.
- Created ETL Data Ingestion Pipelines to upload processed data to the Data Warehouse.
- Data Visualisation and Daily Report Generation using a combination of various ETL tools and scripts.
- Created an ETL tool from scratch and scheduled it using Apache Airflow to ingest 1TB of data per day.

**Skills Learnt:** PySpark, MySQL Server, Hive, PowerBI, Airflow, Kafka, HDFS, Impala, YARN, Cloudera

## EDUCATION

**Institute of Engineering and Management, Kolkata**

**2015-2019**

Bachelor Of Technology | Major: Computer Science and Engineering (CSE) | CGPA: 8.45/10

- Related Coursework: Computer Architecture, Network System, Digital Communication, Operating Systems.

## TECHNICAL PROJECTS

**E-Commerce App with Smart Recommendation System**

**2021**

- Chose a specific niche of Pet food products and used collaborative filtering to recommend the right product to the user and built a home to the checkout web application hosted on an AWS EC2 instance.

## ADDITIONAL EXPERIENCE AND AWARDS

- **Rising Star@phData:** Awarded for outstanding performance and immediate project-ready skillset.
- **500+** Leetcode Questions solved
- **Appreciation Certificate@TCS:** Awarded to the Data Platform team for delivering the project on time
- **HackerRank:** 6 stars in Algorithm and Problem-solving section of HackerRank.

## SKILLS AND TECHNOLOGIES

Snowflake, TRAM, PySpark, AWS, Azure, GCP, Hadoop, Kafka, Impala, Python, MySQL, Airflow, GLUE, S3, Athena, HIVE, HDFS, HBase, C++, Java, Git, Agile, CI/CD, Jenkins, Docker, Cloudera, Databricks, Distributed Systems, Hadoop, OOP, REST API, JIRA, Sentry, Kubernetes, Concurrency, DBT, Exadata, Qubole, Kerberos, Openstack, Nova, Neutron, keystone