

Mayur Surkar

Data Engineer

+918856840298

mayursurkar@gmail.com

[linkedin.com/in/mayur-surkar/](https://www.linkedin.com/in/mayur-surkar/)

www.credly.com/users/mayur-surkar.60194bd7

TECHNICAL SKILLS

Big Data

- **Data Ingestion** – Sqoop
- **Data Processing** - Apache Spark (Pyspark & Scala), Hive, Hadoop
- **Database** – MySQL, IBM DB2, HBase
- **Scheduler** – Airflow
- **ETL Tool** - DataStage
- **Cloud** - AWS (S3, EC2, RDS, Athena, EMR), IBM Cloud

Language

Python, C, Scala, SQL

Operating System

Unix, Windows

CERTIFICATION

- **Microsoft Certified: Azure Fundamentals**
Microsoft, April 2021
- **Cloud Core IBM**, Nov 2021
- **Data Engineer – Data Integration IBM**, May 2021

SUMMARY

Having 2 years of comprehensive experience working as a Data Engineer with persistent analytical and inquisitive nature to drive improvements. Like to face the challenges in the ahead to enhance my skills for professional growth, while being resourceful and innovative. Key competencies include designing and implementing efficient ETL pipeline, refining data for business reporting and using design to solve complex problems. Proficient in range of technologies including Spark, Hadoop, Sqoop, Cloud, HBase, Scala, Python and SQL.

PROFESSIONAL EXPERIENCE

Data Engineer

IBM India Pvt. Ltd. Pune | March 2021 – Present

Master Data 360:

Provides a holistic 360-degree view of an organization's customers and product by harnessing and integrating data from internal and external sources into a single cognitive data platform, improves sales and revenue.

Skills/Technologies Used: Spark, Airflow, SQL, Python, IBM Cloud, Hadoop

Roles and Responsibilities:

- Incorporated data from more than **45 sources** into cloud object storage using **Sqoop**.
- Implemented new pipelines to onboard new sources and enhanced existing pipelines to handle increased data volumes.
- Created **spark** jobs to meet the requirements of end users for data transformation and aggregation.
- Deployed pipelines in a Kubernetes cluster hosting 18 worker nodes on **IBM Cloud**.
- Resolved production issues and assisted the support team to fix issues to operate pipelines 24/7.
- Performed execution, monitoring, integration and loading of data into the database and cloud via DataStage and **Airflow** pipelines.
- Achieved data extraction, data load and quality validation by executing **SQL** queries in **IBM DB2**.
- Monitored daily, weekly, and monthly load jobs and provides reports on a complete run.
- Debugged and troubleshooted data integration solutions and performed testing and modification on them.
- Used **Jira**, **Service Desk**, **ServiceNow**, and other documentation and software development tools to follow Agile methodology.

EDUCATION

- **International Institute of Information Technology Pune** (2016-20)
B.E. Electronics & Telecommunications (CGPA: 8.36/10)
- **Jawahar Navodaya Vidyalaya Wardha** (2014-15)
HSC (87.20%)
- **Jawahar Navodaya Vidyalaya Wardha** (2012-13)
SSC (CGPA: 9.8/10)