

# ANKUR DIXIT

## Sr. Data Engineer & Product Owner

### CONTACT

-  New Sangavi, Pune, India
-  +91 9561090906
-  ADIXIT201@gmail.com
-  <https://www.linkedin.com/in/ankur-dixit-8a459445/>

### TECHNICAL SKILLS

#### Programming Languages

Python, Scala, SQL, Java, C/C++,  
Shell Scripting

#### Python Libraries

Pandas, Scikit-Learn, TensorFlow,  
Pytorch, Matplotlib

#### Hadoop Technologies

Hive, Spark, HDFS, Map Reduce,  
YARN, Kafka, Pig, Hive, Sqoop,  
HBase, Oozie, Zookeeper,  
Cloudera Manager and Horton  
Works

#### Cloud Platforms

Azure Cloud, AWS Cloud  
architecture, AWS Big Data  
computing, Azure logic app

#### Machine Learning

Decision Tree, SVM, Naïve Bayes,  
Regression, Classification, Neural  
networks

#### Databases

Oracle, DB2, MySQL, Sybase,  
PostgreSQL, MS SQL Server

#### NoSQL Databases

MongoDB, Cassandra, HBase

### PROFILE

- Experienced Data Engineer with expertise in the entire project life cycle, including data infrastructure building, ETL processes, and data modeling.
- Proficient in statistical programming languages like Python, SAS, and Apache Spark, and experienced in analyzing data using various tools for data mining, data cleansing, and machine learning applications.
- Skilled in utilizing cloud platforms such as Azure and AWS, and familiar with Microsoft tools including Power BI, Azure Data Factory, and Synapse analytics suite.
- Proven track record in developing predictive models, integrating data lake storage solutions, and implementing automation for deployments. Strong background in ETL processes, data warehousing concepts, and version control systems, and adept at building interactive dashboards and performing ad-hoc analysis using Tableau and Power BI.
- Proficient in working within Agile/Scrum methodologies and possesses strong mathematical and analytical skills.

### PROFESSIONAL EXPERIENCE

#### BARCLAYS Pune, India

Nov 2018 - Present

##### Lead Data Engineer

- Participated in requirement grooming meetings, provided estimates, and managed the end-to-end development, delivery, and validation of code for data pipelines.
- Utilized a variety of technologies including Pandas, Scikit-Learn, TensorFlow, Pytorch, Matplotlib, Apache Spark, Scala, IntelliJ, NoSQL databases, Jenkins, Docker, GitHub, Kubernetes, HDFS, Hive, Kafka, Azure Data Factory, T-SQL, U-SQL, Azure Data Lake Analytics, Azure Databricks, Azure Event Grid, Azure Logic Apps, Azure Functions, Azure Storage, Azure Service Bus, Snowflake, and more for various data processing and analytics tasks.
- Developed and deployed machine learning predictive ETL pipelines, and performed complex data operations, data ingestion, data migration, and data storage and processing across various cloud platforms.
- Worked on PySpark and Spark SQL transformations for complex data operations, and utilized Azure Data Factory Triggers, Azure Event Grid, and Apache Airflow for workflow orchestration.
- Tested databases using complex SQL scripts and collaborated with business users to define data requirements and visualize data using Tableau and Power BI tools in real-time.

##### Senior Data Engineer

- Built machine learning predictive ETL pipelines accessed using RESTAPI for geological data interpretation and optimization of oil exploration processes.
- Developed Spark applications using PySpark and Spark-SQL for data extraction, transformation, and aggregation from multiple file formats to uncover customer usage patterns.
- Implemented environment and deployment automation, infrastructure-as-code, and deployment data pipeline specifications using utilities like GitHub

## CONTACT

-  New Sangavi, Pune, India
-  +91 9561090906
-  ADIXIT201@gmail.com
-  <https://www.linkedin.com/in/ankur-dixit-8a459445/>

## TECHNICAL SKILLS

### Visualization Tools

Tableau, Power BI, Advanced Excel

### Version Control Systems

Git, GitHub, CVS, SVN

### IDEs

PyCharm, IntelliJ IDEA, Jupyter Notebooks, Eclipse

### Operating Systems

Unix, Linux, Windows

## EDUCATION

### Bachelor of Technology-

Computer Science

Uttar Pradesh Technical university,

Lucknow, India

2007 – 2011

## AWARDS

Emerging Leader – Barclays, 2023

Service Excellence – Barclays, 2020 & 2021

Team Star – Harman, 2016

Actions, Terraform, Azure DevOps, and CI/CD automated pipelines for release cycles.

- Optimized data partitioning strategies within Azure Data Factory and configured auto-scaling options to handle varying workloads efficiently. Implemented Azure Logic Apps, Azure Functions, Azure Storage, and Service Bus queries for large enterprise-level ERP integration systems.
- Developed automation systems using Logic apps and JSON templates to remediate Azure services and utilized AWS Data Pipeline for scheduling Amazon EMR cluster to clean and process web server logs. Additionally, proficiency in scripting languages like Python, PowerShell, or Azure Data Lake Analytics U-SQL was utilized for data transformation and automation tasks.
- Migrated data warehouses to Snowflake Data warehouse, extracted data from Data Lakes, EDW to relational databases for analysis, and worked on Spark for performance optimization within Hadoop using Spark Context, Spark-SQL, Data Frames, and Pair RDDs.
- Leveraged Spark-SQL & PySpark based models on Azure Databricks ML workspace to build & deploy ML models at scale for various use cases such as suspicious phishing, covid-19 impact, customer activity, ecosystem, and retention/stickiness.

### IBM Pune, India

May 2017 – Oct 2018

#### Data Engineer

- Built enterprise-level Data warehouses and data lakes, automating data ingestion and scheduling jobs at various frequencies on AWS cloud, modernizing the data warehouse environment using AWS cloud-based Hadoop system, and developing ETL Processes in AWS Glue for data migration into AWS Redshift.
- Utilized Python, Spark on AWS Elastic Search, Sage maker for Analytics & Machine learning models, and integrated AWS Lambda functions with IAM roles using Cloud Watch Triggers.
- Designed logical and physical data models for snowflake, developed large-scale data intelligence solutions around data warehouse, and created data pipelines in Azure Data Factory for efficient data movement, transformation, and orchestration.
- Optimized code and queries within Azure Databricks notebooks for efficient data processing and analysis, integrating Azure Databricks with other Azure services for end-to-end data analytics solutions, and providing training on Azure Cloud and ETL best practices.
- Ingested data through cleansing and transformations using AWS Lambda, AWS Glue, and Step Functions, developed MapReduce programs in Hive, Pig, and Python for Data Analysis, and implemented data quality, integrity, and reliability solutions in data pipelines through data governance and enrichment.
- Created complex Reports in Power BI for data visualization and was involved in building unit tests and integration tests for automated data validation quality checks before deployment.

### HARMAN Pune, India

May 2015 – Apr 2017

#### Big Data Engineer (HADOOP)

- Managed and monitored Hadoop clusters through Hortonworks (HDP) distribution and Cloudera Manager to ensure continuous efficiency.
- Developed Spark Applications using Python to handle data from various RDBMS and Streaming sources, focusing on performance improvement and algorithm optimization within Hadoop.

## CONTACT



New Sangavi, Pune, India



+91 9561090906



ADIXIT201@gmail.com



<https://www.linkedin.com/in/ankur-dixit-8a459445/>

## REFERENCES

Prabhat Kumar

Senior Executive, Rabo Bank  
Netherlands

+31 638157750

Prabhat.kumar@rabobank.com

Pratikshya Samantray

Director of analytics, Deloitte, India  
+91 9886700988

Hitesh Menghani

Vice President in Risk department,  
Barclays, India  
+91 9822068653

- Utilized Azure Data Factory for data transformation, cleansing, and enrichment to meet SLAs and support marketing campaigns and reporting.
- Designed and implemented Cloud Data and Analytical architecture solutions using Azure, including platforms such as Azure SQL Database, Azure Data Lake, HDInsight, and Data Factory.
- Extracted and loaded data into the MS Azure Data Lake using Sqoop and designed data models for client transactional logs, analyzing data from Cassandra.
- Wrote Map Reduce Jobs using PySpark and NumPy and conducted cluster testing of HDFS, Hive, Pig, and MapReduce to enable access for new users.
- Maintained Tableau functional reports based on user requirements

## SEARCE Pune, India

Jun 2013 – Apr 2015

### Big Data Engineer

- Proficiently worked with various Hadoop Ecosystem components including HBase, Sqoop, Zookeeper, Oozie, Hive, and Pig within the Cloudera Hadoop distribution.
- Managed the entire lifecycle of Hadoop jobs, leveraging technologies such as Sqoop, Pig, Hive, MapReduce, Spark, and Shell scripts.
- Supported enterprise Data Warehouse operations and spearheaded Big Data application development using Cloudera and Hortonworks HDP.
- Developed Pig scripts to transform raw data into actionable insights as specified by business users and installed Hadoop, MapReduce, and HDFS to execute multiple MapReduce jobs in Pig and Hive for data cleansing and pre-processing.
- Utilized Spark API over Hortonworks Hadoop YARN to perform analytics on data in Hive and improved the performance and optimization of existing algorithms in Hadoop using Spark Context, Spark-SQL, Data Frames, Pair RDDs, and Spark YARN.
- Developed Spark code using Scala and Spark-SQL/Streaming for efficient testing and processing of data and created a Java-based Spark job to index data into Elastic Search from external Hive tables within HDFS.
- Utilized Spark streaming to process real-time data from Kafka, storing stream data to HDFS using Scala and NoSQL databases such as HBase and Cassandra, and produced various data visualizations using Python and Tableau.