

Priya Sharma

Junior Hadoop Developer

SUMMARY

Junior data engineer with 1.8 years of internship and contract experience across Hadoop-based ETL workflows. Built ingestion pipelines with Sqoop and Spark, modeled Hive datasets for operations analytics, and supported HDFS reliability checks in a 6-node cluster handling 300 GB daily logs.

[Recruiter's Note - Hannah]: This summary works because it gives a time frame, stack context, and scale in one block. Junior candidates who use this pattern get screened faster.

EXPERIENCE

Junior Hadoop Developer

DataRiver Analytics

2025 - Present, Pune

- Maintained 42 scheduled Airflow jobs for daily ingestion from CRM and billing systems into HDFS, improving data freshness from T+1 day to T+3 hours
- Optimized 14 Hive queries using partition pruning and file format conversion to ORC, reducing median report runtime by 48%
- Built Spark validation checks for null spikes and duplicate keys, catching 27 data-quality breaks before BI publication

Data Engineering Intern

StreamGrid Labs

2024 - 2025, Bengaluru

- Developed Sqoop imports for 18 MySQL tables into HDFS and managed incremental loads with watermark columns
- Created MapReduce job to aggregate telecom usage events at 95M records per day, cutting batch completion by 31%
- Documented recovery runbooks for HDFS under-replicated blocks and supported weekend on-call rotation

[Recruiter's Note - Hannah]: Notice the verbs: maintained, optimized, built, developed, created, documented. They signal hands-on contribution, not passive exposure.

PROJECTS

- Capstone: Fraud Signal Pipeline on Hadoop and Spark. Ingested 150 GB transaction logs, engineered 24 risk features, and generated hourly fraud-risk tables consumed by a Tableau dashboard
- Mini Project: Log Search on HBase. Indexed 12M web records and implemented query filters for status_code and URL path analysis

[Recruiter's Note - Hannah]: Projects are written in production language with data volumes and consumers. That is exactly how a developer resume earns trust.

priya.sharma@email.com

000 - 00 - 000 - 00

Pune, India

linkedin.com/in/priyasharma-data

github.com/priya-Hadoop

TECHNICAL SKILLS

Hadoop Ecosystem:

HDFS, YARN, MapReduce, Hive, HBase

Processing:

Apache Spark, PySpark, Kafka basics

Ingestion:

Sqoop, Flume, REST API connectors

Orchestration:

Apache Airflow, cron-based scheduling

Ops:

Linux shell, Ambari monitoring, log analysis

Languages:

SQL, Python, Java basics

[Recruiter's Note - Hannah]: The categories are practical. Recruiters do not need long tool inventories; they need a clear map of what you can run in production.

EDUCATION

B.Tech, Computer Science, Pune Institute of Technology, 2024

CERTIFICATIONS

- IBM Data Engineering Professional Certificate, 2024
- Cloudera Hadoop Fundamentals, 2024
- Databricks Lakehouse Fundamentals, 2025

[Recruiter's Note - Hannah]: Certifications support the profile, but they do not replace project evidence. Keep them concise and current.

LANGUAGES

English Proficient ●●●●●