

# Arjun Mehta

## Senior Hadoop Developer / Big Data Team Lead

### SUMMARY

Senior data engineer with 8 years of experience designing, deploying, and tuning distributed data platforms built on the ecosystem. Managed a 200 node HDFS cluster processing 3.4 TB daily for a European fintech company, reducing batch ETL runtime by 62 percent through Spark migration and partition optimization. Led a cross functional team of six engineers and mentored three junior developers to mid level promotions within 18 months. Certified Cloudera Data Engineer and IBM Big Data Architect.

*[Recruiter's Note - Hannah]: The summary opens with years of experience plus a concrete cluster size and data volume. Hiring managers can instantly gauge seniority. The 62 percent reduction and team size prove leadership, not just technical skill.*

### EXPERIENCE

#### Senior Hadoop Developer / Team Lead

FinData GmbH

2019 - Present, Berlin

- Architected a multi tenant HDFS namespace strategy supporting four business units on a shared 200 node cluster, reducing storage costs by 28 percent through erasure coding and tiered storage policies.
- Migrated 53 legacy MapReduce jobs to Spark SQL on YARN, cutting average execution time from 42 minutes to 11 minutes while processing 3.4 TB of daily transaction data.
- Designed a Kafka to HDFS ingestion pipeline handling 2.8 million events per minute from payment microservices, integrated with Schema Registry to enforce Avro schema evolution across 14 topics.
- Led weekly architecture reviews for a team of six engineers, introduced code review gates in Git that reduced production incidents by 35 percent over two quarters.

*[Recruiter's Note - Hannah]: Every bullet here follows the verb plus tool plus scale plus result formula. I can see cluster size, data volume, event throughput, and percentage improvements without reading twice. This is exactly what a strong Hadoop experienced resume looks like.*

#### Senior Hadoop Developer / Team Lead

FinData GmbH

2016 - 2019, Berlin

- Built and maintained an 80 node HDFS cluster on Cloudera CDH, implementing Kerberos authentication and Ranger policies that passed two consecutive SOC 2 Type II audits without findings.
- Developed Oozie coordinator workflows scheduling 72 daily Hive ETL jobs, with SLA alerting that kept 99.5 percent of downstream dashboards updated before the 08:00 CET business deadline.
- Automated Sqoop incremental imports from four Oracle and two PostgreSQL sources into HDFS, replacing manual scripts and saving the operations team 15 hours per week.

arjun.mehta@email.com

+49 170 555 2288

Berlin, Germany

linkedin.com/in/arjunm

github.com/arjunm

### TECHNICAL SKILLS

#### Core Ecosystem:

HDFS, YARN, MapReduce, ZooKeeper

#### Query and Analytics:

Hive, HBase, Pig, Presto, Impala

#### Processing and Streaming:

Apache Spark (Core, SQL, Streaming),

Kafka, Flink

#### Orchestration:

Oozie, Apache Airflow, Azkaban

#### Cloud Platforms:

AWS EMR, S3, Azure HDInsight, GCP

Dataproc

#### Languages:

Java, Scala, Python, Bash, SQL

*[Recruiter's Note - Hannah]: Grouping tools by function shows architectural thinking. A flat comma list would force me to mentally categorize everything myself, and at 200 resumes a day I will not do that.*

### CERTIFICATIONS

- Cloudera Certified Data Engineer (CDE)
- IBM Certified Big Data Architect
- Databricks Certified Associate Developer for Apache Spark

*[Recruiter's Note - Hannah]: Three recognized certifications from different providers show continuous learning. Cloudera plus IBM plus Databricks covers on premises, architecture, and modern Spark, which is a strong combination for a senior candidate.*

- Optimized Hive queries on a 900 GB partitioned dataset by introducing ORC columnar format and predicate pushdown, reducing average query time from 14 minutes to 2.8 minutes.

*[Recruiter's Note - Hannah]: Notice the progression: the earlier role shows strong individual contributions, while the current role emphasizes team leadership and architecture ownership. That trajectory is what separates a senior resume from a mid level one.*

#### EDUCATION

**M.Sc. Computer Science, Technical University of Munich, 2016**

**B.Sc. Information Technology, University of Pune, India, 2014**

#### LANGUAGES

**English Proficient** ●●●●●

**German Proficient** ●●●●●