

# Basit Ali - Principal Big Data Engineer

+966503972756 | Basitali4545@yahoo.com | [LinkedIn](#)

## SUMMARY

---

Principal Data Engineer with 9+ years of experience designing and leading large-scale data platforms across healthcare, telecom, government, retail, and IoT sectors. Skilled in building secure, scalable data architectures, real-time processing systems, and cloud-based solutions that improve performance and reduce costs. Recognized for leading teams, improving engineering standards, and ensuring data solutions deliver clear value across the organization.

## WORK EXPERIENCE

---

### Lean Business Services, Riyadh Saudi Arabia

February 2025 – Present

#### Senior Big Data Engineer

- Architected and implemented a large-scale Patient Profile Platform using Apache Spark, consolidating and centralizing healthcare data from hospitals and medical facilities across the Kingdom into a unified enterprise DataLake establishing a single source of truth for patient records at national scale.
- Designed and developed patient linkage profiling algorithms to match and merge patient identities across multiple disparate hospital systems, eliminating data duplication and enabling a complete 360-degree longitudinal patient view across facilities.
- Built and managed end-to-end data pipelines responsible for ingesting, transforming, and loading structured and unstructured clinical data from various source systems including lab results, radiology, pharmacy, and outpatient records into the centralized Data Lake.
- Developed advanced analytics and insight-generation layers on top of the centralized Data Lake, enabling stakeholders to derive actionable population health insights, clinical trends, and operational performance metrics from unified patient data.
- Engineered a metadata-driven ingestion framework using Streamlit and Python to automate data onboarding, enabling configuration of SQL, MySQL, Oracle, and NoSQL sources without manual coding — significantly accelerating data delivery and reducing onboarding effort across new hospital integrations.
- Architected real-time data processing pipelines using Spark Streaming, Kafka, and Apache NiFi to ingest high-volume healthcare data streams, ensuring low-latency availability for critical operational and clinical decision-making.
- Optimized long-running Oracle extraction jobs and batch processing workflows, reducing processing windows and improving reliability and predictability of daily data loads across all integrated facilities.
- Implemented strong data governance, quality control frameworks, and privacy safeguards across all pipelines, ensuring compliance with healthcare data protection regulations and national data security standards.
- Reduced infrastructure costs by optimizing event-driven architectures, eliminating redundant pipelines, and decommissioning inefficient legacy workflows — improving overall system performance and scalability.

### Ai-Elements, Riyadh Saudi Arabia

September 2023 – January 2025

#### Senior Big Data Engineer

- Designed and implemented robust data pipelines that ingest data from multiple sources and transform it into actionable insights, ensuring availability for analytics and decision-making.
- Collaborated with cross-functional teams to convert business requirements into efficient data workflows, improving operational efficiency and reducing costs.
- Led development and maintenance of end-to-end pipelines using Airflow, NiFi, and Spark, enabling real-time ETL from diverse systems to target platforms.
- Translated business needs into scalable pipeline solutions while coordinating with stakeholders to gather, refine, and document requirements.
- Managed and optimized on-premises data workflows, ensuring security, reliability, and compliance with organizational standards.
- Implemented best practices across on-premises infrastructure, delivering a resilient and scalable architecture.
- Applied Clean Code practices and refactored legacy codebases, reducing technical debt by 30% and improving maintainability.
- Built a Cloudera-based PoC for AI-driven insights using Spark and Hadoop to process large datasets, train ML models, and enhance performance.

## **Systems Limited, Islamabad Pakistan**

July 2019 – September 2023

### *Senior Big Data Engineer*

- Designed and delivered Microsoft Fabric Medallion Architecture and scalable data pipelines using Apache Spark, Hadoop, and Big Data ecosystems, improving ingestion, transformation, and analytics performance.
- Built distributed ETL frameworks with Spark and Python on AWS EMR, optimizing multi-terabyte processing, improving data accuracy, and enabling actionable insights.
- Implemented strong data governance, privacy controls, and end-to-end documentation, ensuring compliance and operational consistency.
- Engineered AWS-based data platforms using EMR, Glue, Lambda, Kinesis, Step Functions, CloudFormation, RDS, S3, and Redshift, enabling both batch and real-time processing at scale.
- Developed real-time ingestion pipelines using Kinesis Firehose, incremental processing with Apache Hudi, and automated orchestration using AWS Lambda and Step Functions.
- Built multi-cloud pipelines integrating Azure and AWS, including ADF, Azure Functions, Queues, Redis, and Databricks for near real-time ETL and cross-platform data movement.
- Applied Clean Code, SOLID principles, and TDD, refactoring legacy systems and reducing technical debt by 30% while improving code quality and maintainability.
- Implemented CI/CD automation using Jenkins, integrating build pipelines with AWS to orchestrate Spark jobs, data workflows, and infrastructure deployments.
- Designed and optimized SQL-based data validation, creating stored procedures and automating test routines to support QA teams during data verification.

## **Metis International Private Limited, Islamabad Pakistan**

February 2018 – June 2019

### *Big Data Developer*

- Developed graphical representations of data using Neo4j, enhancing data visualization capabilities.
- Designed and implemented a comprehensive web-based Big Data portal using Python and Django, facilitating efficient data management and analysis.

### *Big Data Developer (Dubai Future Accelerators - Onsite)*

- Conducted social media analysis to gain insights into crime perspectives, hate speech, and sentiment analysis using Python.
- Utilized Python for Social Media Analysis on Drugs, identifying the location of users posting content related to the purchase or sale of drugs on social media platforms, supporting the Anti-Narcotics Department.

### *Front End Developer*

- Employed Angular 2 and Angular 5 technologies to develop an interactive front-end application.
- Created real-time alert systems, live vehicle tracking features, geofencing functionality, and analytical dashboards for monitoring fuel consumption, temperature changes, speed, and driving behavior.

## **CORE ARCHITECTURE COMPETENCIES**

---

- Data Lake, Lakehouse, Data Mesh, Data Fabric & Medallion Architecture
- Cloud Data Architecture, Modeling, Governance & Metadata Management
- ETL/ELT Architecture, Automation & Real-Time Streaming
- Data Integration, Federation, Virtualization & Master Data Management
- AI/ML & Generative AI-Enabled Data Platform Architecture

## **CERTIFICATIONS AND TRAININGS**

---

- Microsoft Certified Azure Data Engineer Associate
- Cloudera Technical Professional (CTP) Accreditation
- Cloudera Technical Specialist (CTS) Accreditation
- Cloudera Technical Expert (CTE) Accreditation
- Future Leadership Program offered by Lahore University of Management Sciences (LUMS)

## TECHNICAL SKILLS

---

1. **Big Data & Analytics:** Cloudera Data Platform | Apache Spark | Kafka | Airflow | Databricks | Apache Hudi | Hadoop | HDFS | MapReduce | Neo4j | Hive | Power Apps
2. **Cloud Platforms:** Azure (ADF, Synapse Analytics, Databricks, Functions, DevOps) | AWS (EMR, Glue, Lambda, Kinesis, Redshift, S3, Step Functions) | GCP | Microsoft Fabric
3. **Programming:** Python | Java | SQL | Node.js | Angular | Spark
4. **Databases:** HP Vertica | Teradata | Redshift | Snowflake | BigQuery | Azure Synapse | MongoDB | Elasticsearch | Redis | PostgreSQL
5. **DevOps & Tools:** GitHub | Azure DevOps | Jenkins | TFS | Trello | Jira | MS Office | MS Visio
6. **Artificial Intelligence:** Generative AI | RAG Applications | Agentic AI | Deep Learning | Machine Learning

## EDUCATION

---

**International Islamic University, Islamabad Pakistan**  
Bachelor of Science in Software Engineering

February 2018

## ACHIEVEMENTS

---

- Team of the Year (2021) and Star Performer (2020, 2021) – Systems Limited.
- Lead and mentor a high-performing Big Data engineering team.
- Completed the Future Leadership Program – LUMS (REDC).