

Anis Sayed

Principal Data Architect & DevOps Lead

Dubai, UAE | Portfolio: anissayed.dev | LinkedIn: linkedin.com/in/anis-sayed-4039aa202

SUMMARY

Principal Data Architect & DevOps Lead with 10+ years building, migrating and automating large data platforms for telecom, banking and government. Hands-on across Cloudera and Hadoop, Kafka, Spark, containers and CI/CD, and leads client delivery from sizing and design through build to handover.

Currently leads technical delivery of Cloudera-based data platforms for government and enterprise clients across the GCC at BBI.ai. Previously brought containers and gated CI/CD to a bank's analytics platform, and contributed to migrating the Jio Big Data Lake from Hortonworks to Cloudera CDP.

SKILLS

- Data platforms: Cloudera CDP, Hadoop (HDFS, YARN, Hive, Ozone)
- Streaming & ingestion: Kafka, NiFi
- Processing & workflows: Airflow, Oozie, Spark & PySpark
- Containers & orchestration: Docker & Podman, Kubernetes, Red Hat OpenShift
- CI/CD & automation: Ansible, Azure DevOps, GitHub Actions, Jenkins, Python & Bash, SonarQube, Checkmarx & Nexus
- Cloud & infrastructure: Azure data services, RHEL, VMware & Nutanix, Terraform (demo/experiment)
- Security & governance: Kerberos, Ranger, Knox & TLS
- Observability & reliability: Grafana, Prometheus & alerting, Runbooks, readiness & rollback
- AI/ML enablement: Cloudera Machine Learning, Containerised ML workflows, Governed AI access (exploration)

EXPERIENCE

Principal Data Architect & DevOps Lead

BBI.ai, Dubai, UAE | Apr 2025 – Present

- Technical and delivery lead for a government security organisation's data platform programme, which took three evidence-processing use cases into production.
- Administer Cloudera CDP and Cloudera Machine Learning through Cloudera Manager: cluster health, configuration, security (Kerberos, TLS, Ranger) and upgrades.
- Build platform automation with Ansible and GitHub Actions, including pre-installation readiness checks, remediation and certificate tooling.
- Lead pre-sales architecture: platform sizing, growth models and RFP compliance reviews.
- Validate upgrade and migration runbooks, and lead escalations on client Cloudera and data-warehouse estates.

Client assignments (client names withheld):

- Government security organisation (Client delivery): Technical and delivery lead for a multi-phase data platform programme, from platform install to three production use cases and handover.
- Government agency (Proof of concept): Built a raw-to-gold lakehouse proof of concept on Cloudera CDP with PySpark and Hive.
- Government agency (Client delivery): Wrote the deployment manual and configuration for an open-source Hadoop environment with Kerberos.

- Regulated financial-services client (Client delivery): Automated pre-installation readiness checks for a Cloudera cluster.
- Telecom operators (Client delivery and proof of concept): Cross-hypervisor migration proof of concept; validated upgrade runbooks and led a support escalation.
- Public-sector, banking and telecom prospects (Pre-sales): Platform sizing and RFP compliance reviews for proposals.

Senior Technology Engineer, Big Data & DevOps

Emirates NBD, Dubai, UAE | Oct 2022 – Mar 2025

- Built a framework for running Spark and big-data jobs in Docker/Podman containers on YARN in CDP, and moved server-based workflows onto it.
- Orchestrated containerised big-data jobs through YARN and Kubernetes.
- Designed GitHub Actions CI/CD for the big-data estate with SonarQube quality gates, Checkmarx security scans and Nexus artefacts.
- Implemented on-demand, self-hosted GitHub Actions runners with Podman (container in container).
- Deployed containerised AI/ML model workflows on the CDP Private Cloud cluster.
- Led the advanced-analytics team's side of the Hortonworks to Cloudera CDP migration.

Big Data DevOps Administrator (Assistant Manager)

Reliance Jio, Mumbai, India | May 2018 – Sep 2022

- Helped administer the Jio Big Data Lake. The platform, run by the wider team, had approximately 800 servers and 50 PB of HDFS capacity.
- Lead contributor to the Hortonworks (HDP) to Cloudera CDP migration, including NameNode federation, Ozone, erasure coding and end-to-end pipeline testing. Recognised with the RIL Samman award.
- Designed and configured a Kafka setup of around 150 nodes in a single cluster, using config groups and ZooKeeper znodes.
- Designed the big-data CI/CD toolchain: Git repositories, Jenkins pipelines for Maven/SBT artefacts, Rundeck for deployments, Airflow for workflow management and Nexus for artefact versioning.
- Implemented CI with Azure DevOps and Jenkins (agents, distributed builds, plug-ins) and CD with Jenkins and Azure DevOps release automation.
- Supported production Kubernetes clusters and ran Kubernetes proofs of concept.
- Worked on migrating a three-tier application from on-premises Hadoop to Azure (Synapse, Blob Storage, Azure SQL, Event Hubs).
- Hardened the platform with Kerberos, LDAP/SSSD, TLS, Ranger and Knox, and monitored it with Check-MK, Grafana, Prometheus and Dr. Elephant.

Big Data / Hadoop Administrator

Pune-based data technology firm, Pune, India | Jun 2015 – May 2018

- Installed and administered Hortonworks Hadoop clusters.
- Set up monitoring and alerting for cluster services and documented operational procedures.

PROJECTS

- Turning evidence files into governed, queryable data (Client assignment via BBI.ai · Government security organisation, 2025 – 2026): Technical and delivery lead for a government security organisation's data platform, from platform install to three evidence-processing use cases in production and handover.

- Migrating one of India's largest data lakes to Cloudera CDP (Employer · Reliance Jio, 2018 – 2022): Lead contributor to the Hortonworks-to-Cloudera migration of the Jio Big Data Lake (about 800 servers and 50 PB of capacity), while helping run it day to day.
- Containers and CI/CD for a bank's analytics platform (Employer · Emirates NBD, 2022 – 2025): Introduced containerised Spark workloads and gated GitHub Actions CI/CD to a bank's Cloudera-based advanced-analytics platform. ML model workflows used the same path.
- Re-platforming a Hadoop application onto Azure (Employer · Reliance Jio, 2018 – 2022): Worked on migrating a three-tier application from on-premises Hadoop to Azure Synapse, Blob Storage, Azure SQL and Event Hubs. Related work included NiFi ingestion into ADLS and Azure DevOps CI/CD.
- Automated readiness checks for Cloudera installs (Own tooling · built during BBI.ai client engagements, 2025): Reusable Ansible automation, triggered from GitHub Actions, that validates and remediates Cloudera CDP prerequisites across a fleet and writes per-host CSV evidence.
- A raw-to-gold lakehouse proof of concept on Cloudera (Proof of concept via BBI.ai · Government agency, 2026): Designed and built a config-driven proof of concept that ingests unstructured and semi-structured events into HDFS and transforms them with PySpark into partitioned Parquet tables in Hive.

CERTIFICATIONS

- CDP Certified Administrator – Private Cloud Base, Cloudera (Issued Aug 2022)
- Certified Kubernetes Administrator (CKA), The Linux Foundation / CNCF (Issued Dec 2020 · expired Dec 2023), Credential ID LF-r8am1pfpc5
- Microsoft Azure Administrator, Microsoft (training in progress)
- Red Hat OpenShift Administration, Red Hat (training in progress)

Awards: A* performance rating, three consecutive years (Annual performance reviews); RIL Samman Award (Reliance Industries)

EDUCATION

- Bachelor of Technology, Pune University, 2015
- Diploma in Engineering, Zeal Education Society, Pune, 2012

LANGUAGES

English, Hindi, Marathi, Arabic (intermediate)