

DATA PLATFORM MODERNISATION · CAPITAL MARKETS

Migrating 2.2 petabytes without stopping the market

A zero-downtime Cloudera Data Platform sidecar migration for one of the oldest stock exchanges in Asia, moving years of live production data with no data loss while trading workloads ran throughout.

2.2 PB

MIGRATED, ZERO DATA LOSS

100%

UPTIME DURING MIGRATION

32

NODE CDP PRODUCTION CLUSTER

4 mo

END TO END; 12-WEEK ACTIVE MIGRATION

THE CHALLENGE

Modernise a live exchange, with zero room for error

The client is **one of the oldest stock exchanges in Asia**, running a mission-critical data platform behind market surveillance, regulatory reporting, and analytics across millions of trades each day.

The platform had reached the end of its supported life on an end-of-life operating system, raising security, compliance, and vendor-support risk. Yet the institution could tolerate neither data loss nor unplanned downtime during any transition.

THE SOLUTION

A sidecar migration, proven before cutover

The scale and criticality ruled out an in-place upgrade. Translab built a parallel **Cloudera Data Platform (CDP)** cluster alongside the live environment, seeded and validated it in full, then cut over only once it had earned production trust.

Built on prior engagements with the exchange, covering Spark, Ranger and Atlas, and Kafka and StreamSets, the team arrived with deep knowledge of the estate and delivered the engagement end to end, with no sub-contractors.

HOW THE PLATFORM MOVED, SAFELY

Legacy Hadoop cluster

END-OF-LIFE · UNSUPPORTED OS

HDFS

Hive

Kafka

Kudu

Impala

Sqoop

Spark

Cloudera BDR replication

Dual run, then cutover

New CDP Private Cloud Base

32 NODES · FULLY SUPPORTED

HDFS

Hive (ACID/ORC)

Kafka

Kudu

Impala ×16

YARN pools

Security parity preserved: Kerberos authentication, Ranger authorisation, and Knox gateway migrated with full fidelity.

Bulk seed

Full HDFS, Hive & Kudu replication

Live delta

Continuous change replication

Validate

Row-count & format reconciliation

Cut over

Controlled maintenance window

THE SAFEGUARD THAT MADE IT ZERO-RISK

Before any cutover, the new and legacy clusters were run in parallel as a **dual run**, with both production systems live at once. This preserved a continuously available fallback and let the team confirm parity under genuine production load before legacy was retired, turning a high-stakes migration into a controlled, reversible sequence rather than an act of faith.

OUTCOMES & IMPACT

A modern platform, delivered without disruption

- ✓ **Zero data loss, fully verified.** All 2.2 PB reconciled by row count across every Hive database, with HDFS integrity confirmed and lineage preserved for regulatory assurance.
- ✓ **Zero unplanned downtime.** Live ingestion, batch, and interactive query workloads continued uninterrupted across the twelve-week active migration.
- ✓ **Security posture strengthened.** Kerberos, Ranger, and Knox configured on the new cluster, with encryption standards aligned to regulatory expectations for a market institution.
- ✓ **Performance stabilised.** Impala admission control re-tuned to true physical capacity, removing the queue build-up a misconfiguration would otherwise have caused.
- ✓ **A foundation for what comes next.** A fully supported CDP platform now anchors a follow-on Disaster Recovery programme and readiness for new analytics and AI workloads.

The cost of waiting

End-of-support dates are fixed, and Cloudera's move to a subscription-based CDP has changed the commercial picture too. Each quarter on an unsupported distribution compounds security exposure and technical debt, making the eventual move harder and more expensive.

What it unlocked

Beyond removing risk, the platform gives headroom for growing data volumes, faster regulatory reporting, stronger operational visibility, and a base for real-time and AI-led workloads.

A REPEATABLE, THREE-PHASE METHOD

1 · Evaluate

Plan and analyse every workload and dataset before any build.

**2 · Develop & test**

Build CDP, **migrate each workload**, then test, secure, and certify against a non-production cluster.

**3 · Productionise**

Repeat on the live sidecar, **dual-run** old and new, then cut over and retire legacy.

Modernise on your own terms, not under pressure

This is not a challenge unique to one exchange. Cloudera has consolidated the older CDH and HDP lineages into the unified CDP and set firm end-of-support timelines against the legacy distributions. For any enterprise still running CDH, legacy HDP, or an early CDP release, and especially for platforms above several hundred terabytes carrying live production workloads, or regulated institutions where downtime and data loss are not options, the question is no longer whether to migrate, but how to do so without disrupting the workloads the platform exists to serve.

The institutions that act early, on a proven sidecar pattern, modernise on their own terms rather than under the pressure of a looming deadline or a security incident. Whether adopting CDP Private Cloud Base, planning disaster recovery, or retiring end-of-support risk, the same approach that protected a live exchange at petabyte scale is ready to protect yours.