Achieving Real Success with Hadoop and NoSQL

Presented by Amir Halfon, CTO Financial Services, MarkLogic Corporation

MarkLogic

Typical Enterprise Data Management

- Hadoop; RDBMS; NoSQL; search
  - Data duplication, inconsistency, Integration nightmare, multiple indexes, limited cross-functionality…
- Hadoop + NoSQL: no enterprise features, no Integrated search
  - Still hard to deploy and manage
- Traditional RDBMS: enterprise features galore, little flexibility or agility
- **What's the alternative?**

MarkLogic

Hadoop's Advantages

- Hadoop and HDFS tackle the Volume issue
  - HDFS provides scale and economies of scale
- File-based nature allows for greater Variety
  - Raw data is fine and any shape will do
- Map-reduce enables massive parallel scaling, providing answers in minutes/ hours that previously took days/weeks
  - ..but it's back to writing programs that read a bunch of files…

MarkLogic

Hadoop's Limitations

- Hadoop was designed for batch processing
  - Does not support real-time applications on its own
  - Limited support for SQL-based analytics
- Requires expertise to configure, deploy and manage
- Until recently has not supported mainstream security authentication technologies
- **You need a database**

MarkLogic

30 YEARS OF SUCCESS

ACID IS INDISPENSABLE

ENTERPRISE HARDENED

RDBMS: THE GRAND DADDY OF DATABASES

MarkLogic

30 YEARS OF SUCCESS

ACID IS INDISPENSABLE

ENTERPRISE HARDENED

RDBMS: THE GRAND DADDY OF DATABASES

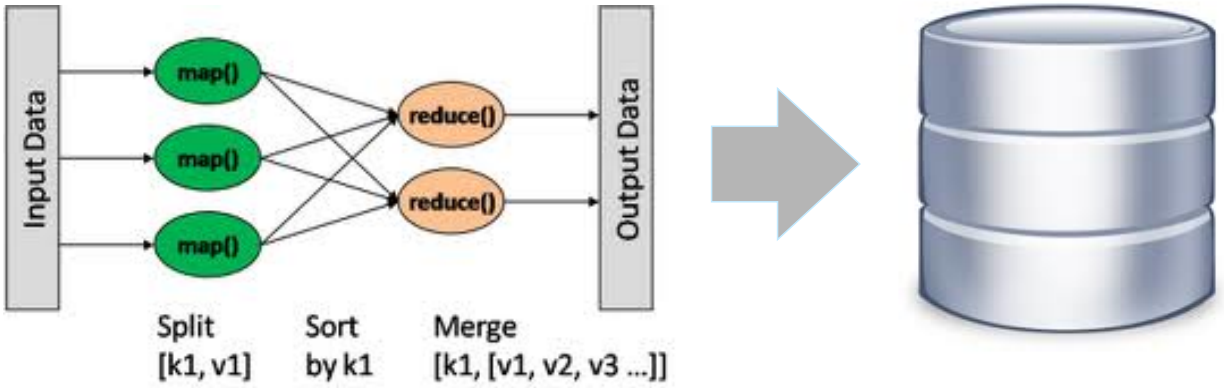
TODAY'S DATA DOESN'T FIT

DATA MODELING IS SLOW

SCALING IS SLOW & EXPENSIVE

MarkLogic

Hadoop + RDBMS



MarkLogic

SCHEMALESS

FAST & AGILE

RUNS ON SCALE-OUT CLUSTERS

NOSQL: THE YOUNG UPSTART

MarkLogic

SCHEMALESS

FAST & AGILE

RUNS ON SCALE-OUT CLUSTERS

NOSQL: THE YOUNG UPSTART

DATA IS MORE THAN KEYS & VALUES

BUSINESS REQUIRES ACID TRANSACTIONS

IT REQUIRES HA/DR/SECURITY

MarkLogic

12 YEARS OF SUCCESS

ACID IS INDISPENSABLE

ENTERPRISE HARDENED

ENTERPRISE NOSQL: THE BEST OF BOTH

SCHEMA AGNOSTIC

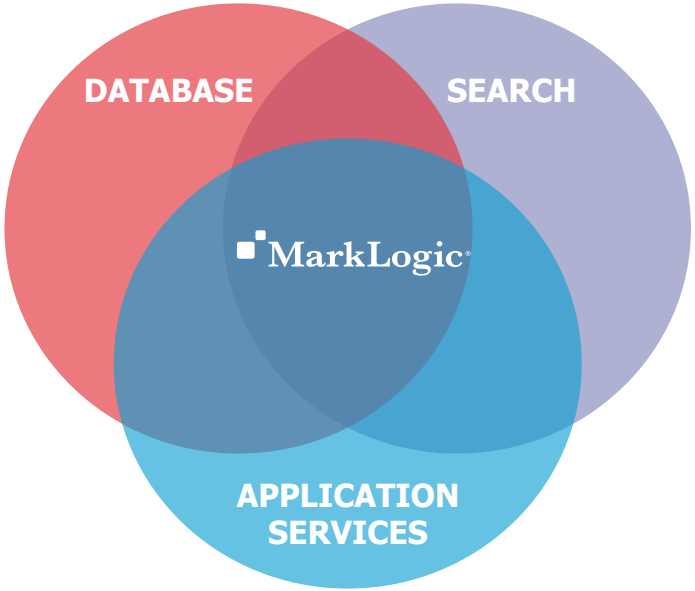
FAST, AGILE & POWERFUL

RUNS ON SCALE-OUT CLUSTERS

MarkLogic

MarkLogic Enterprise NoSQL

- Search & Query
- ACID Transactions
- High Availability / Disaster Recovery
- Replication
- Government-grade Security
- Elasticity
- Cloud Deployment
- Hadoop for Storage & Compute
- Semantics



MarkLogic

Why Integrated Search?

- Interactive access to unstructured text
- Best if integrated at the DB layer:
  - Quicker time to information
  - Greater than the sum of its parts
  - No separate indexes to maintain
  - One less friction point
- Search as a text processing engine
  - Turn text into numbers, create new dimensions
  - Infer new information from text search and enrich

MarkLogic

MarkLogic – the Best Database for Hadoop

- Drop-in enterprise database for real time querying, search and analytics
- Mission critical applications directly on HDFS
- Leveraging existing investment without compromising enterprise capabilities
- Simplified architecture – minimizing duplication and reducing risk
- Save money by leveraging common infrastructure – no more ETL
- Leverage MarkLogic's sophisticated security model for sensitive data

MarkLogic

Example: Investment Bank

MarkLogic

Post Trade Operations

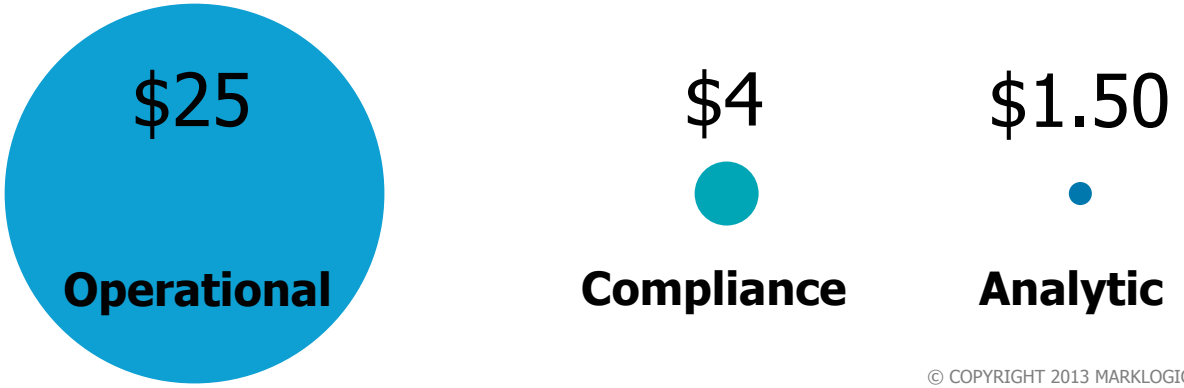
MarkLogic

Post Trade Operations v2

Total Size of Trade Data (TB)

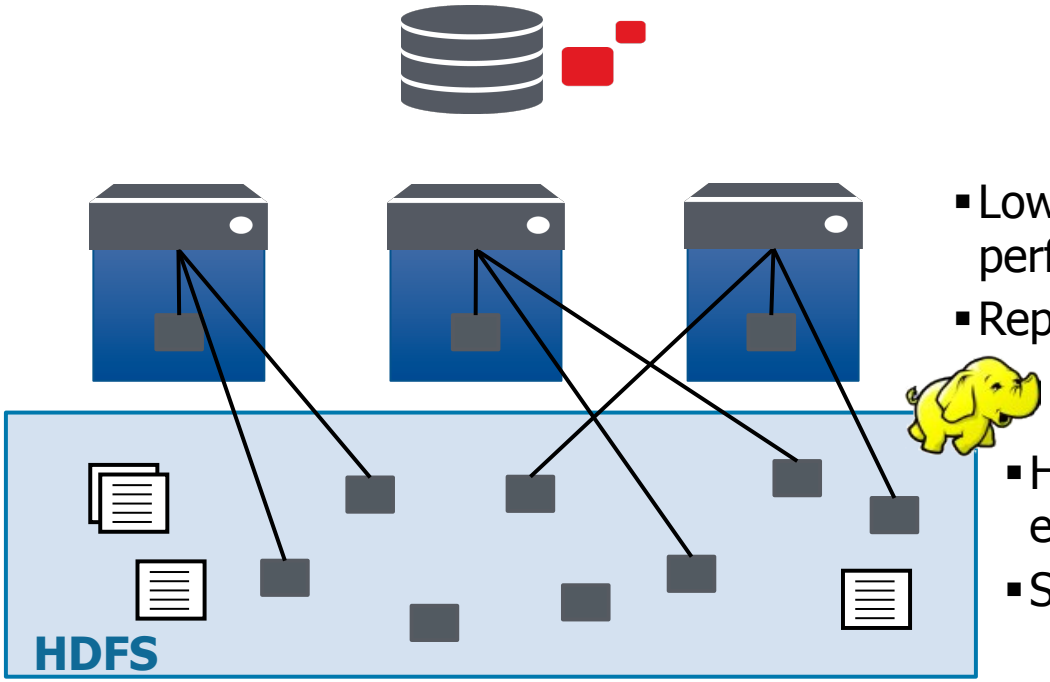
Total Cost ($000)

Effective Unit Cost



MarkLogic

Tiered Storage Using Hadoop



~$25/GB

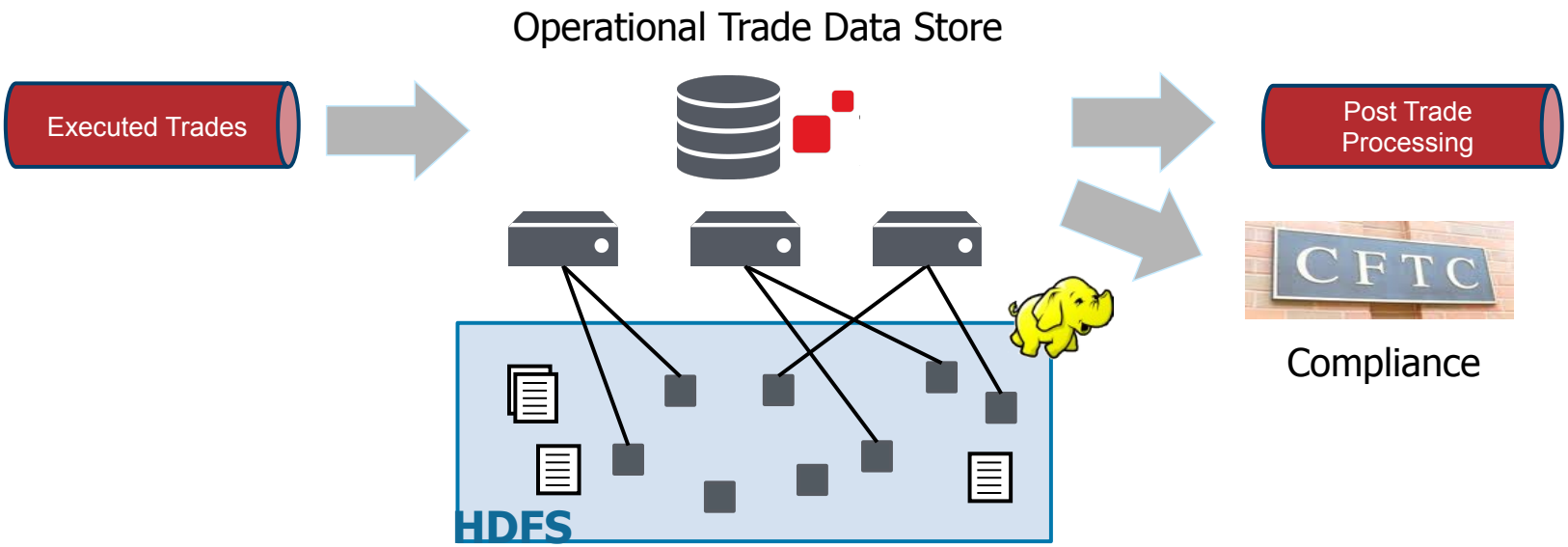
- Low density for ingest performance
- Replication for HA

~$1/GB

- High density for efficiency
- Shared-disk failover

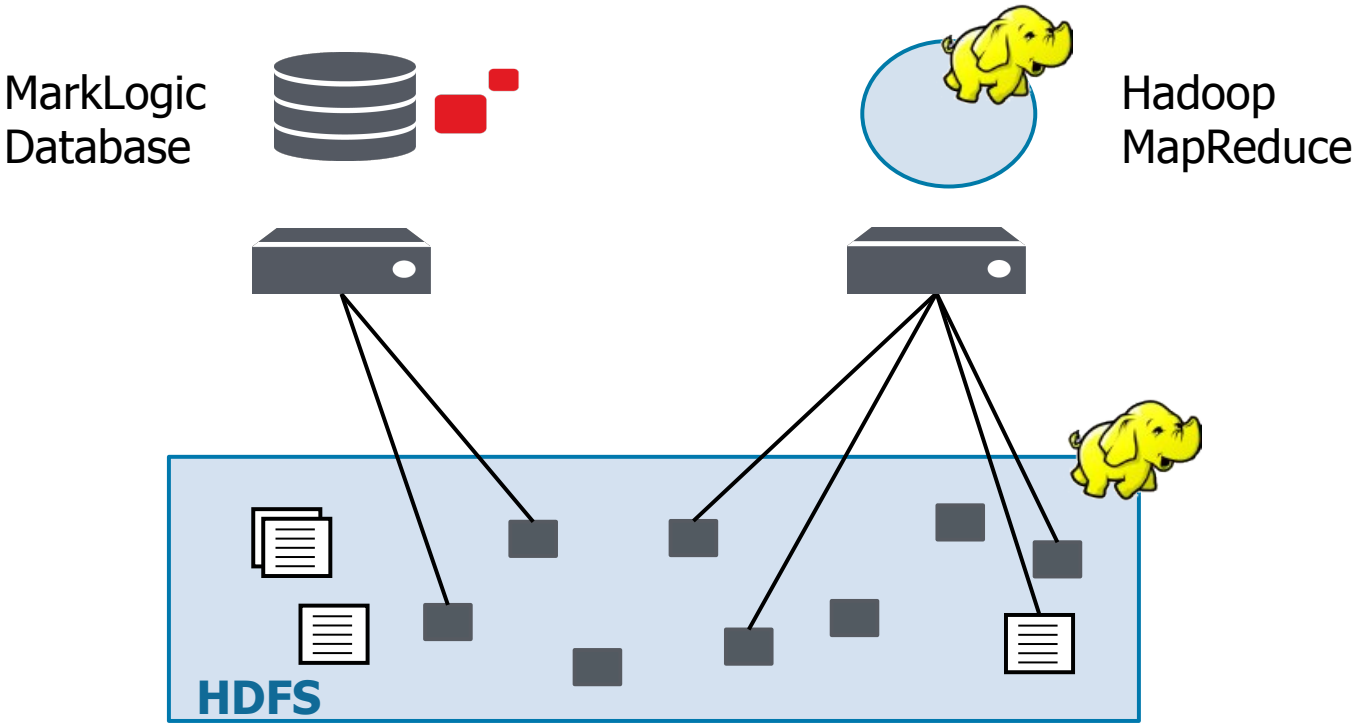
MarkLogic

Post Trade Operations v3



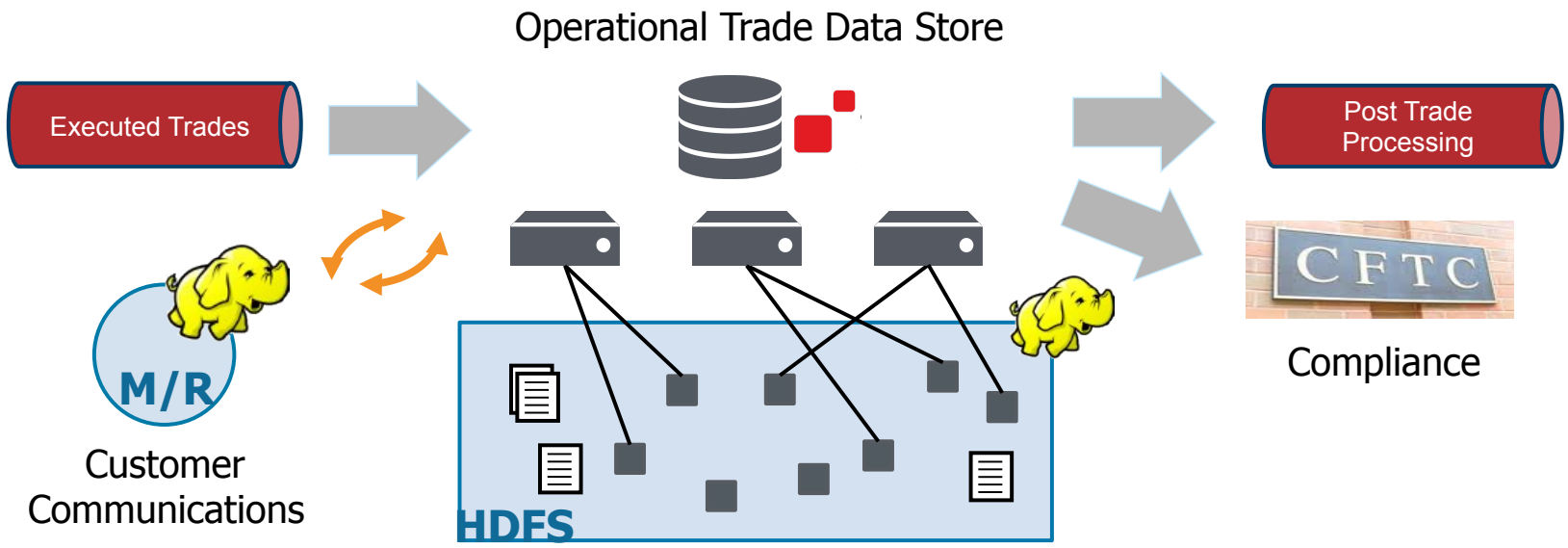
MarkLogic

Mixed Workloads with Hadoop



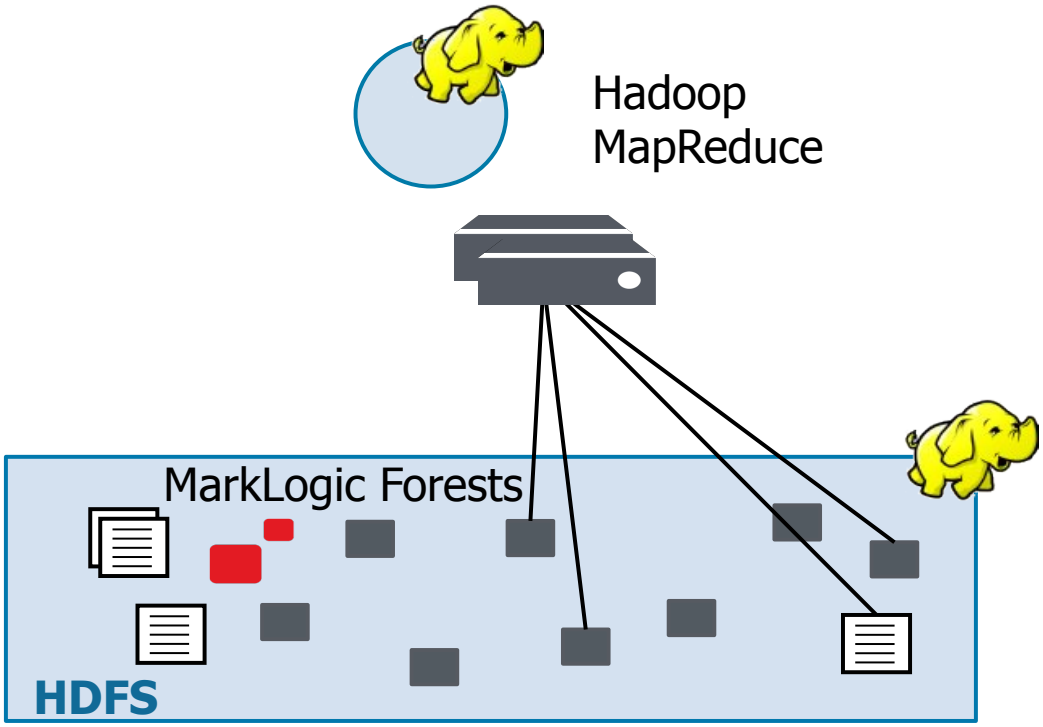
MarkLogic

Post Trade Operations v4



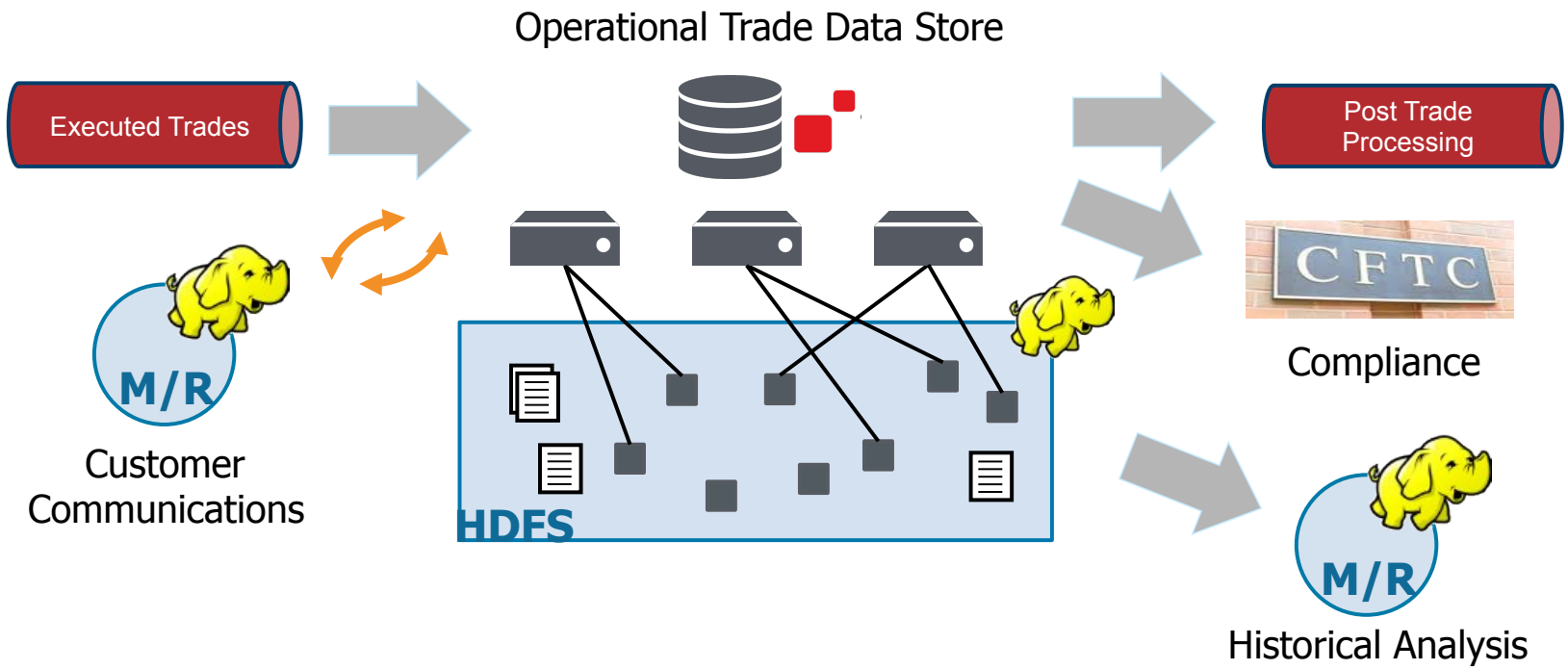
MarkLogic

Direct Access

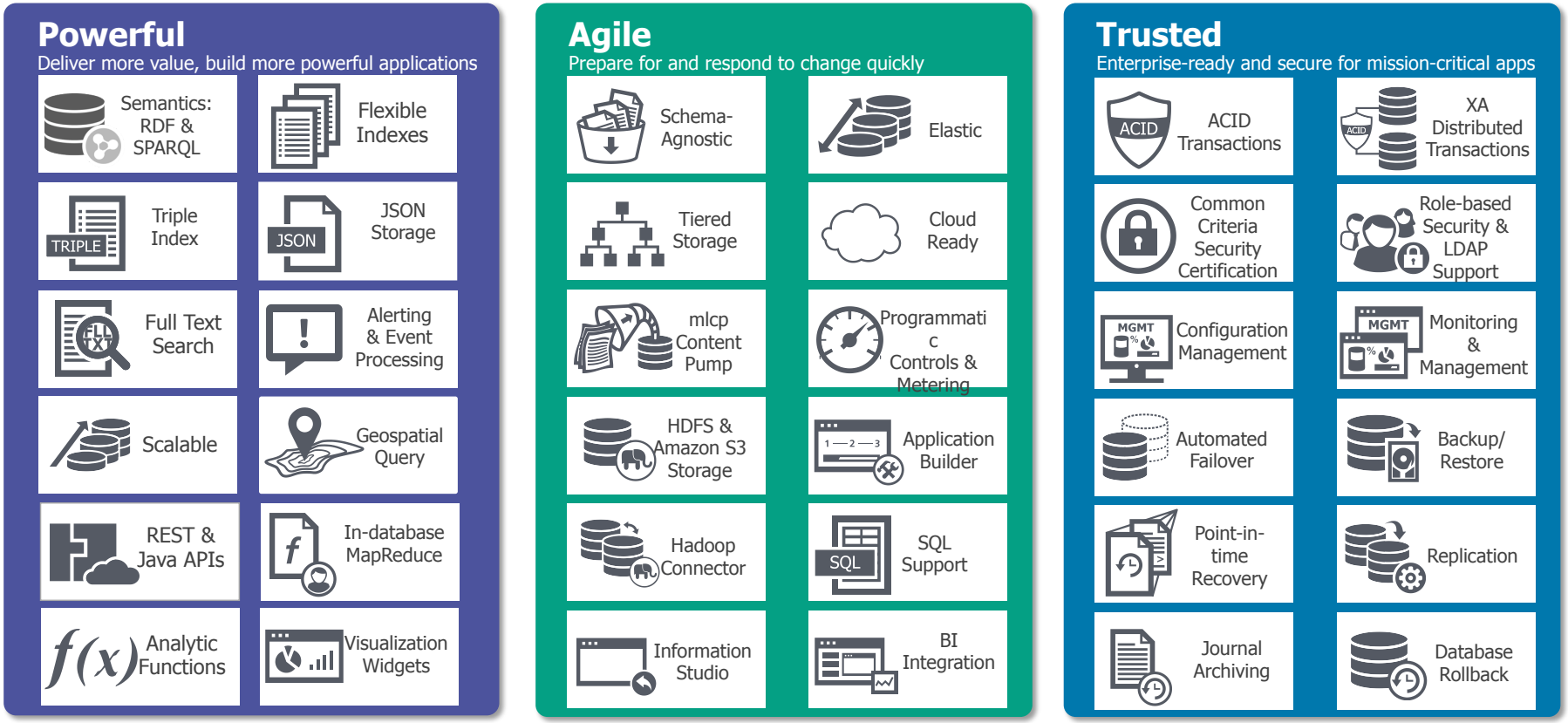


MarkLogic

Post Trade Operations v5



The Rest of the Story



MarkLogic

Summary

- Hadoop and NoSQL are becoming the new generation EDM platform
- It's not about SQL, it's schema on read that's different
- NoSQL does not mean no enterprise features
- Firms are already using this platform to increase agility and reduce costs

MarkLogic

Check us out

developer.marklogic.com

Data Governance for Regulated Industries session

Booth 2A