



EZMERAL DATA FABRIC 資料平台

Danny Ho

HPE EZMERAL DATA FABRIC



Ezmeral Data Fabric ecosystem pack
開源軟體

商業第三方產品

JDBC/ODBC/NFS/各種API etc.

基礎軟體

HPE Ezmeral Data Fabric
filesystem
分散式文件系統

HPE Ezmeral Data Fabric
Database
資料庫

HPE Ezmeral Data Fabric
Event Store
事件Stream處理

HPE Ezmeral Data Fabric Object
Store
物件存儲

多通訊協定、開源套件企業支持、支援各種基礎設施部署、水平擴展、結合HPE Green Lake提供私有雲體驗

資料就是

生命力

對於啟動**新一代**營運模式至關重要

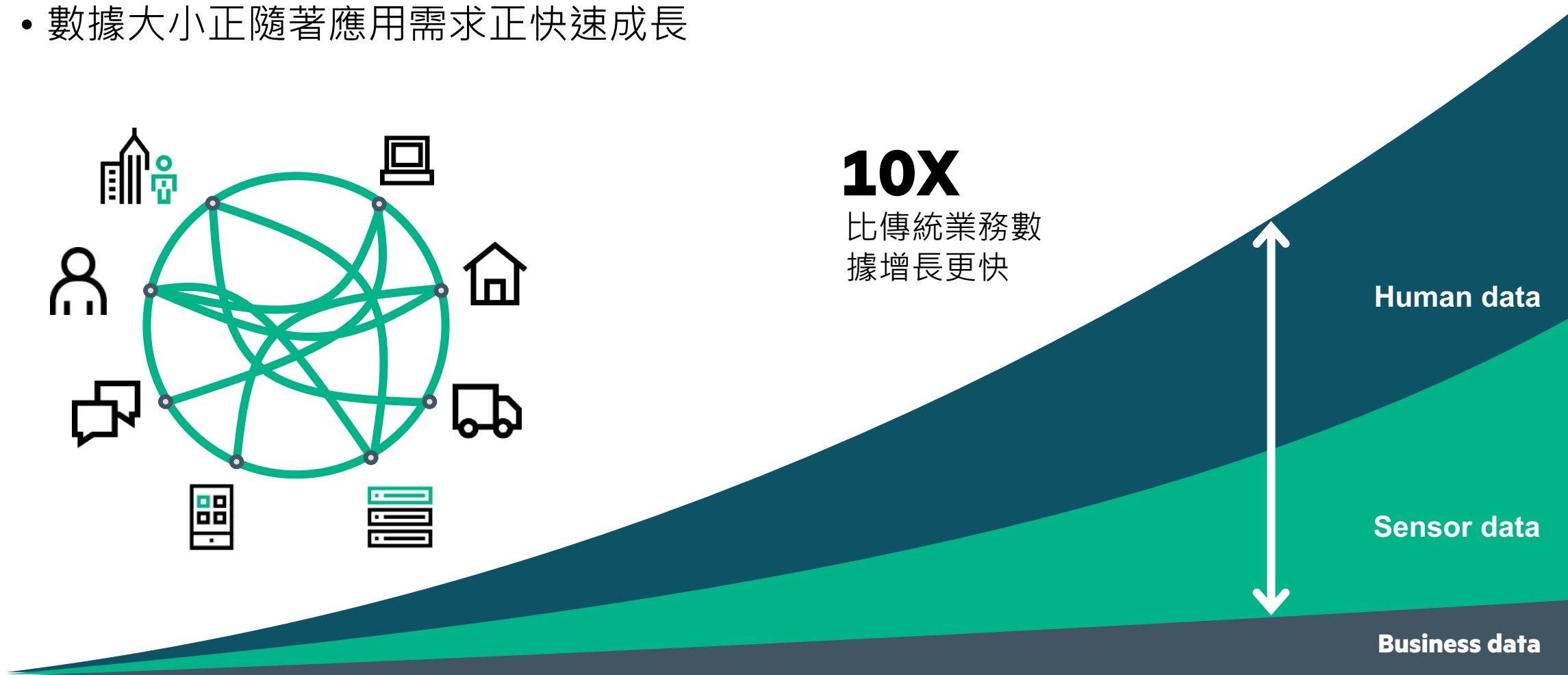
提供豐富的體驗

加快決策速度

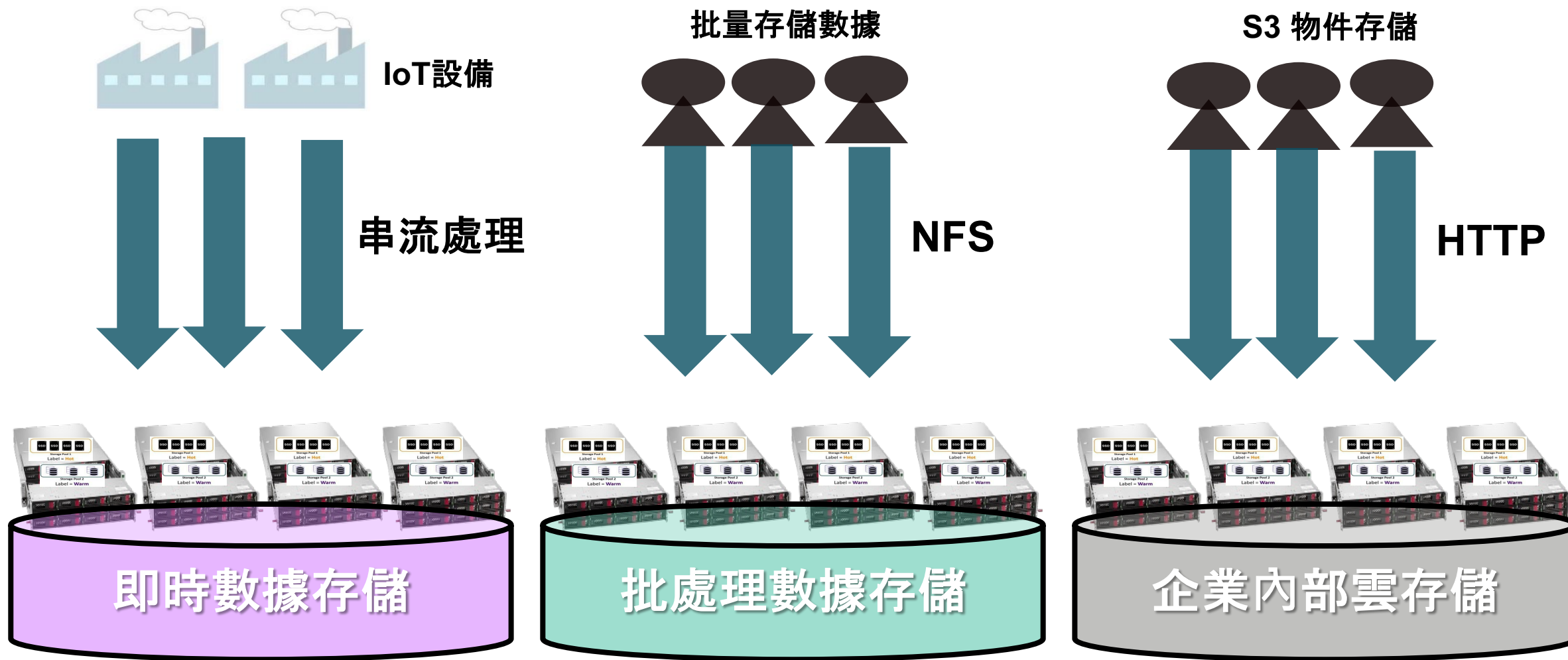
創新應用活力

數據推動向數位化企業轉型

- 數據大小正隨著應用需求正快速成長



企業中各種資料儲存的情境



企業常碰到的資料問題

如果沒有
資料優先現代化，就無法
從**Edge**端到資料中心再到雲端，實現完整的數位轉型

資料存取方式多樣

缺乏針對分析和 **AI**
進行優化並可擴展、
安全的數據平臺。

預處理開源軟體維運困難

不斷變化、不斷擴
大的開源生態系統。

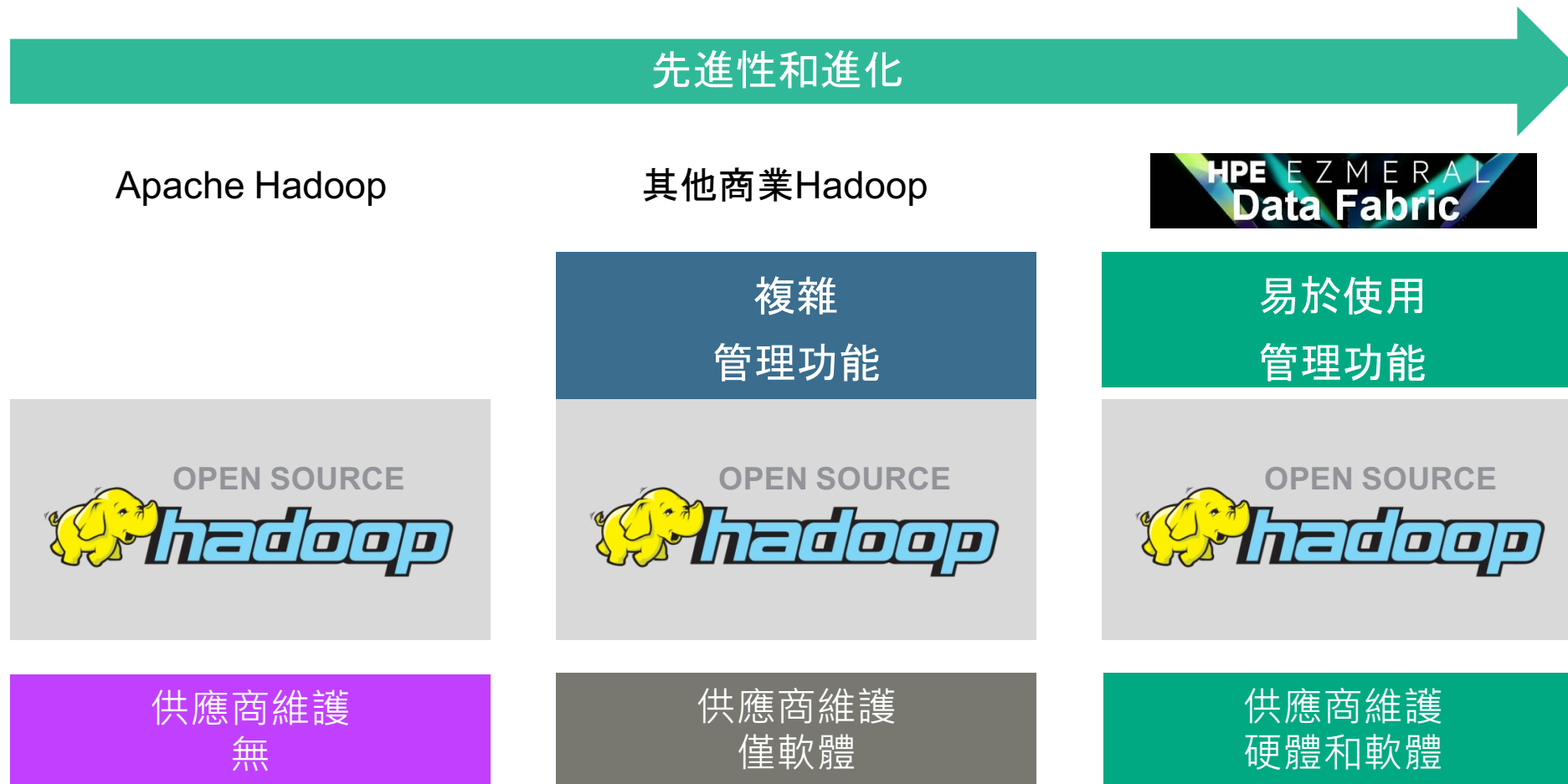
Edge端資料量增長

Edge-to-Cloud 各種
混合部署模型

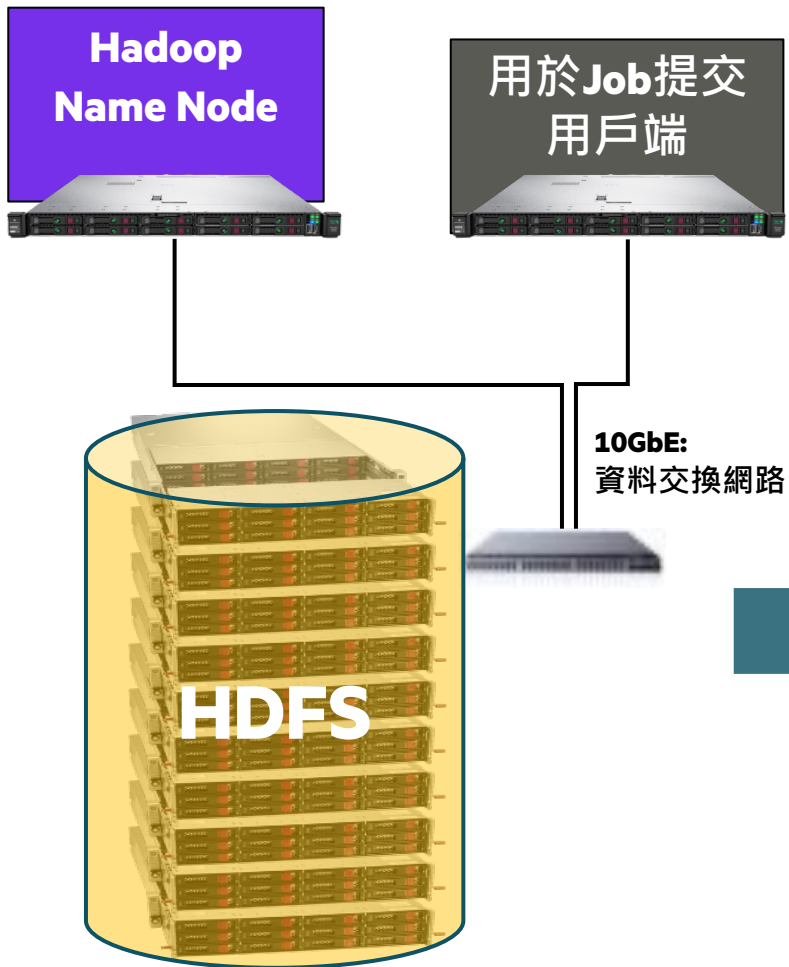
資料成本難以預計

不可預測的成本
和容量需求

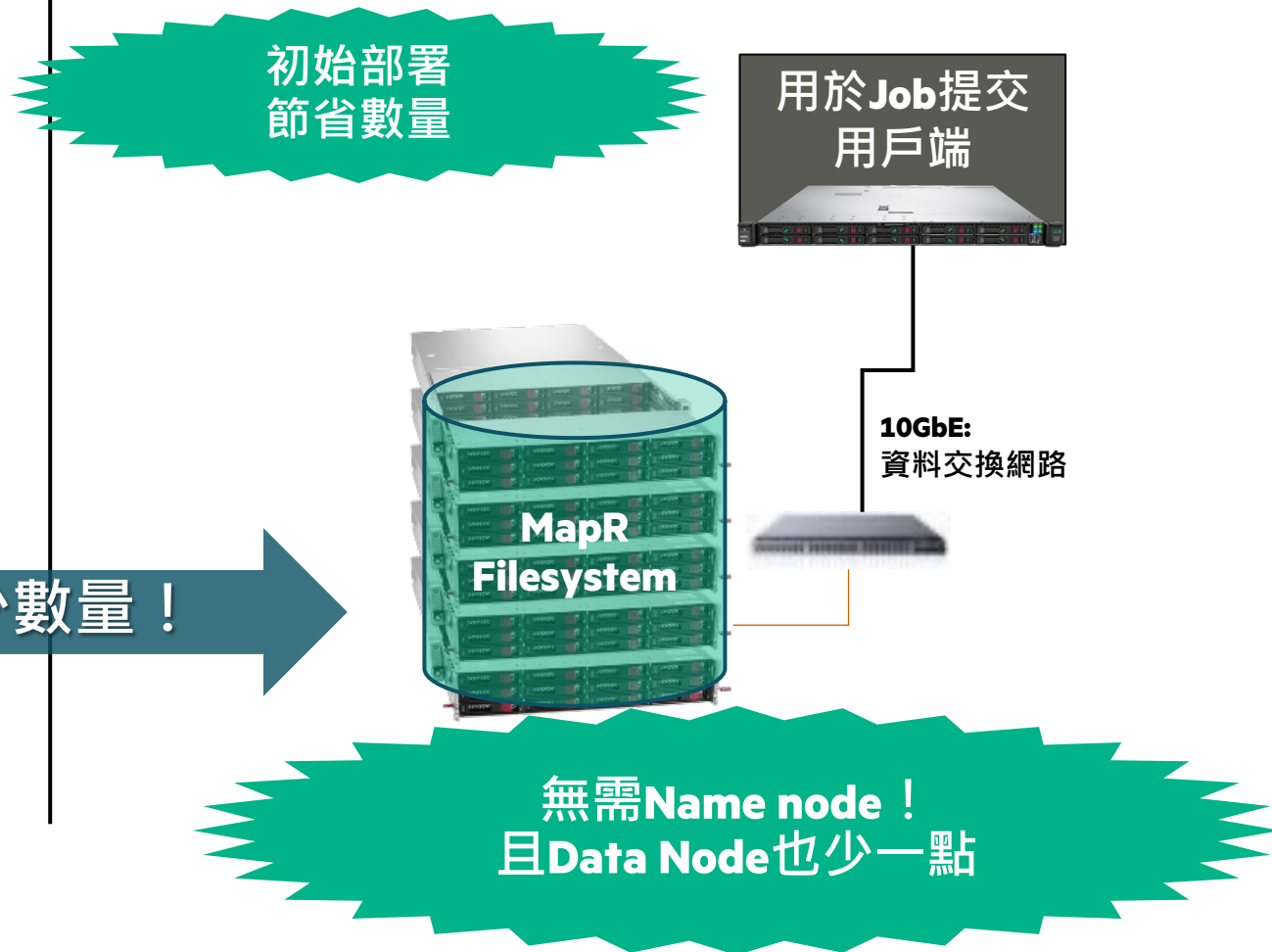
DATA LAKE 解決方案選擇



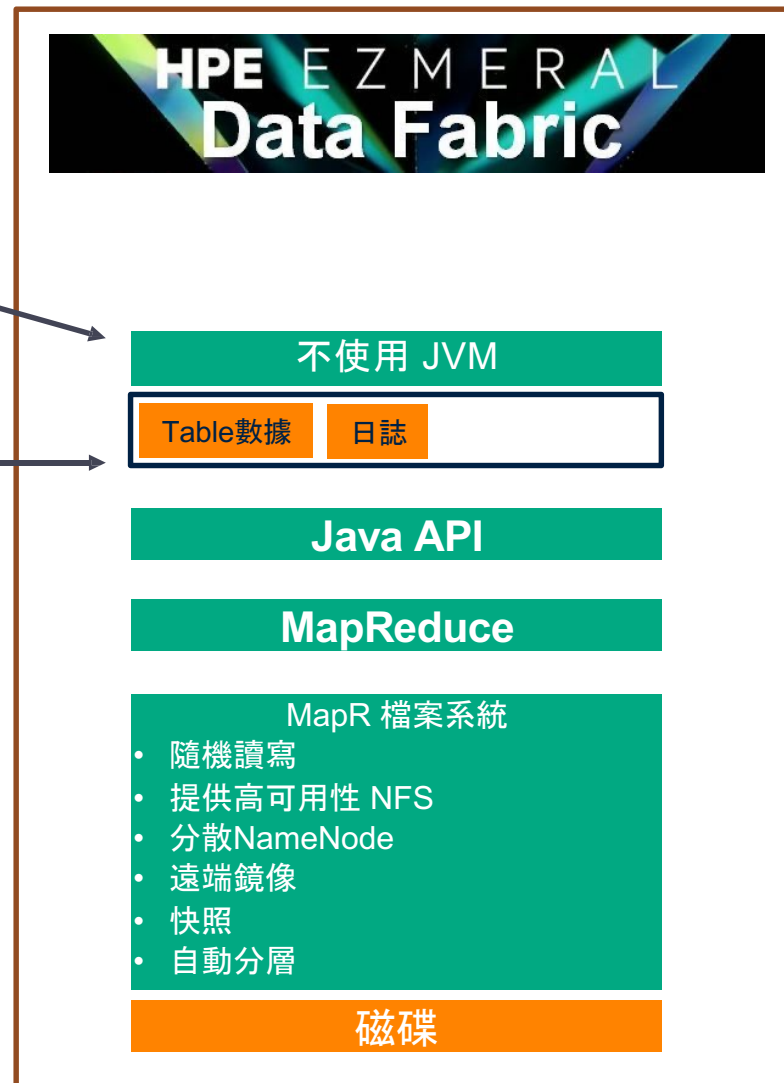
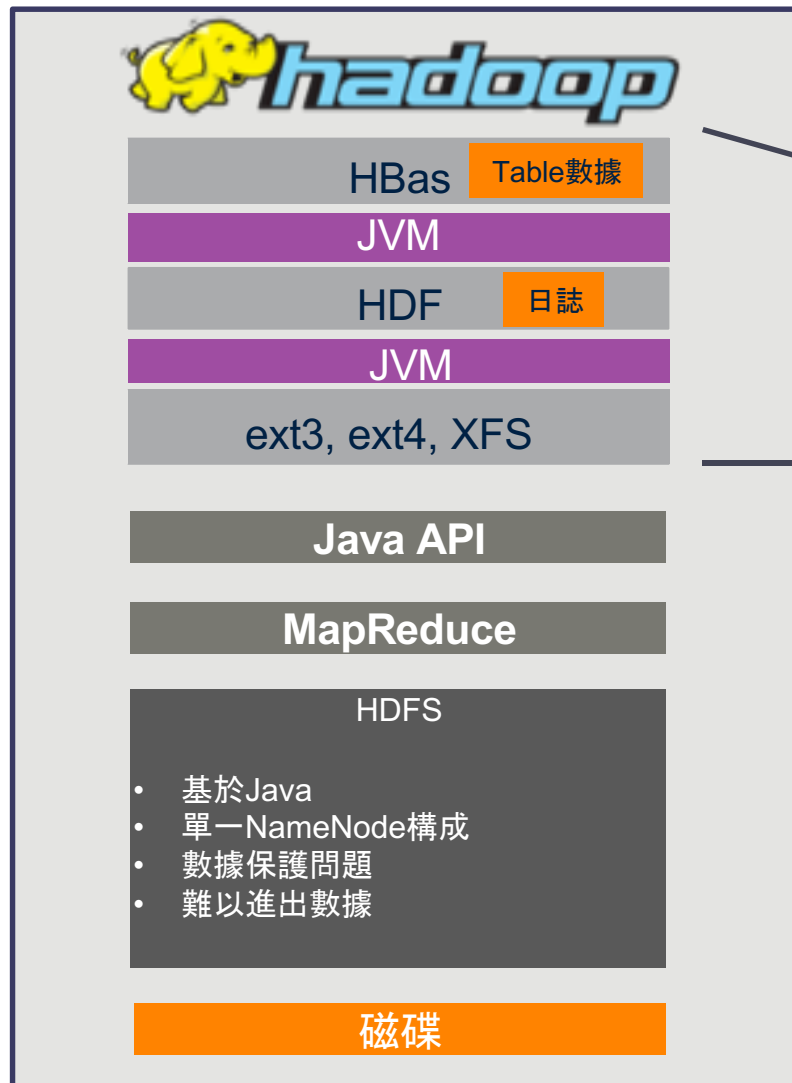
Apache Hadoop 和 Cloudera 和 Hortonworks



HPE Ezmeral Data Fabric



DATA FABRIC 解決 HADOOP 效能問題

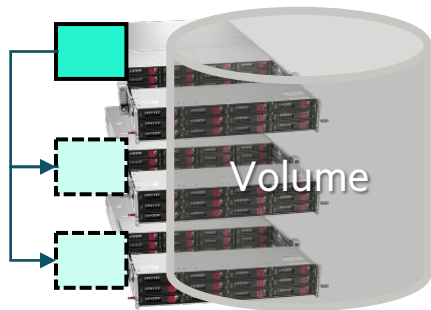


HPE EZMERAL DATA FABRIC FILESYSTEM

可靠的資料備援保護

高可用性

將數據的副本保存在三個不同的節點上
即使兩個故障，也可以訪問數據



作為 NFS 伺服器的高可用性

在發生故障時繼續 NFS 服務



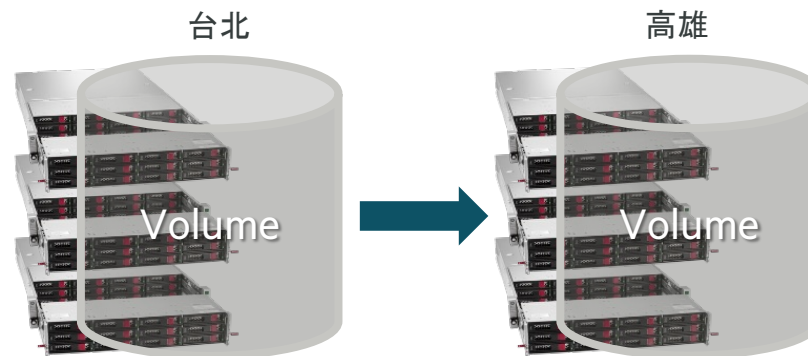
快照

切換到多個過去狀態



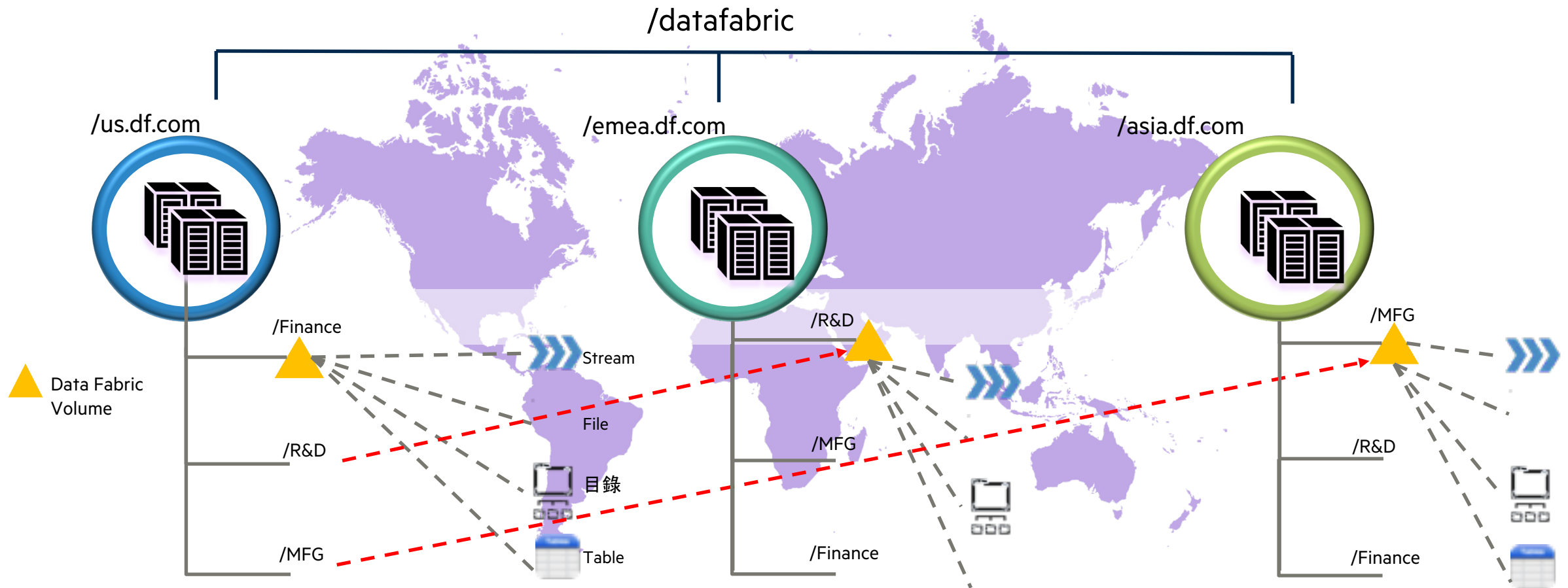
複製

在資料中心之間複製數據

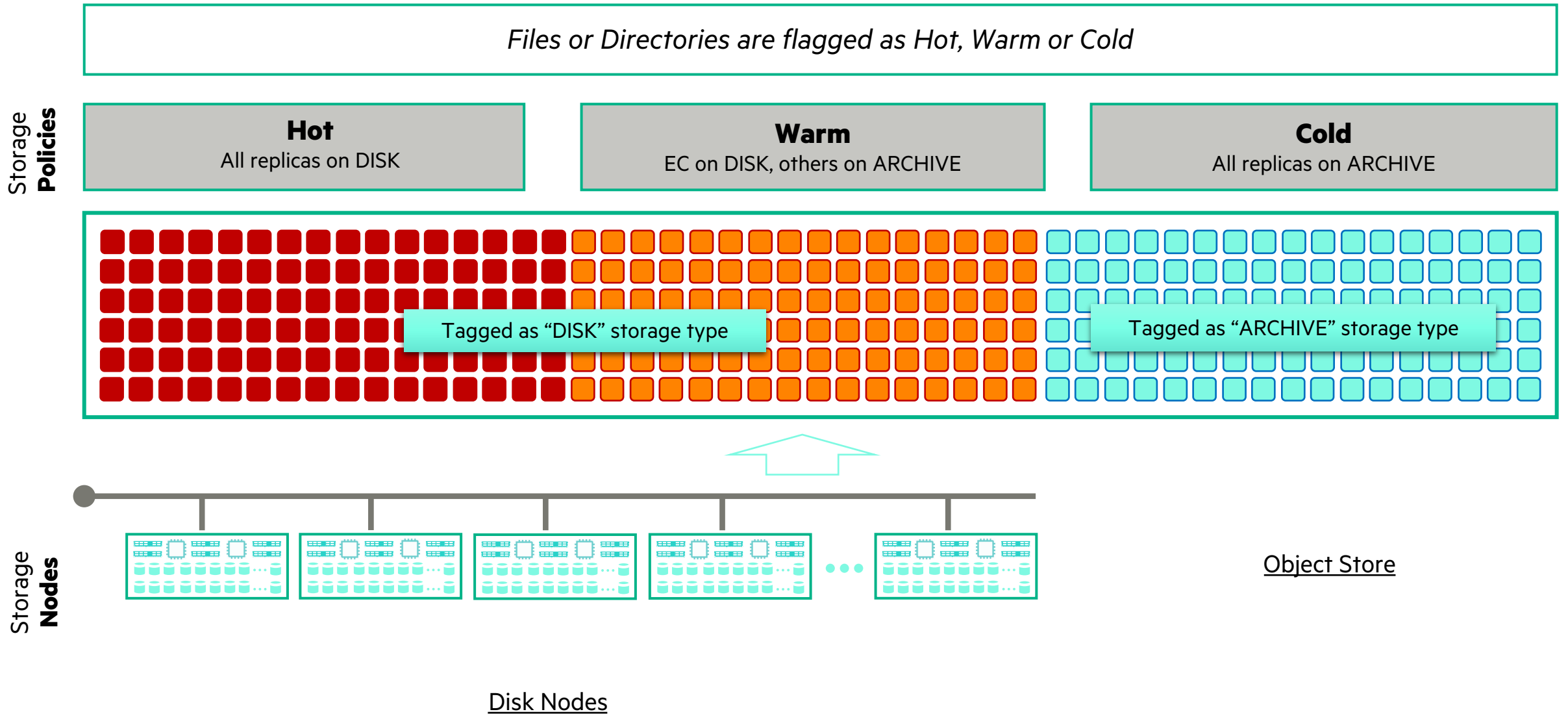


輕鬆訪問所需的數據：GLOBAL NAMESPACES

通過相互信任遠端叢集的存儲空間來訪問任何遠端叢集上的任何數據集，就像本地群集的一部分一樣



資料分層存放



HPE EZMERAL DATA FABRIC OBJECT STORE

HPE EZMERAL DATA FABRIC OBJECT STORE (S3) 雲端體驗落地化

Object Store 可以保存大量非結構化數據，可以從任何位置輕鬆訪問這些數據

數據儲存

- 用於構建本地應用程式或遷移到雲原生應用程式的持久性數據存儲。

數據分析

- 查詢數據以執行分析和深度學習，從而獲得對客戶、運營或市場的寶貴見解。

多樣化檔案格式儲存

降低儲存豐富的多樣化格式（音樂、視頻和圖像）的成本。

串流媒體影像儲存



Backup Archive



原始資料的數據備存



混合雲平台數據取用

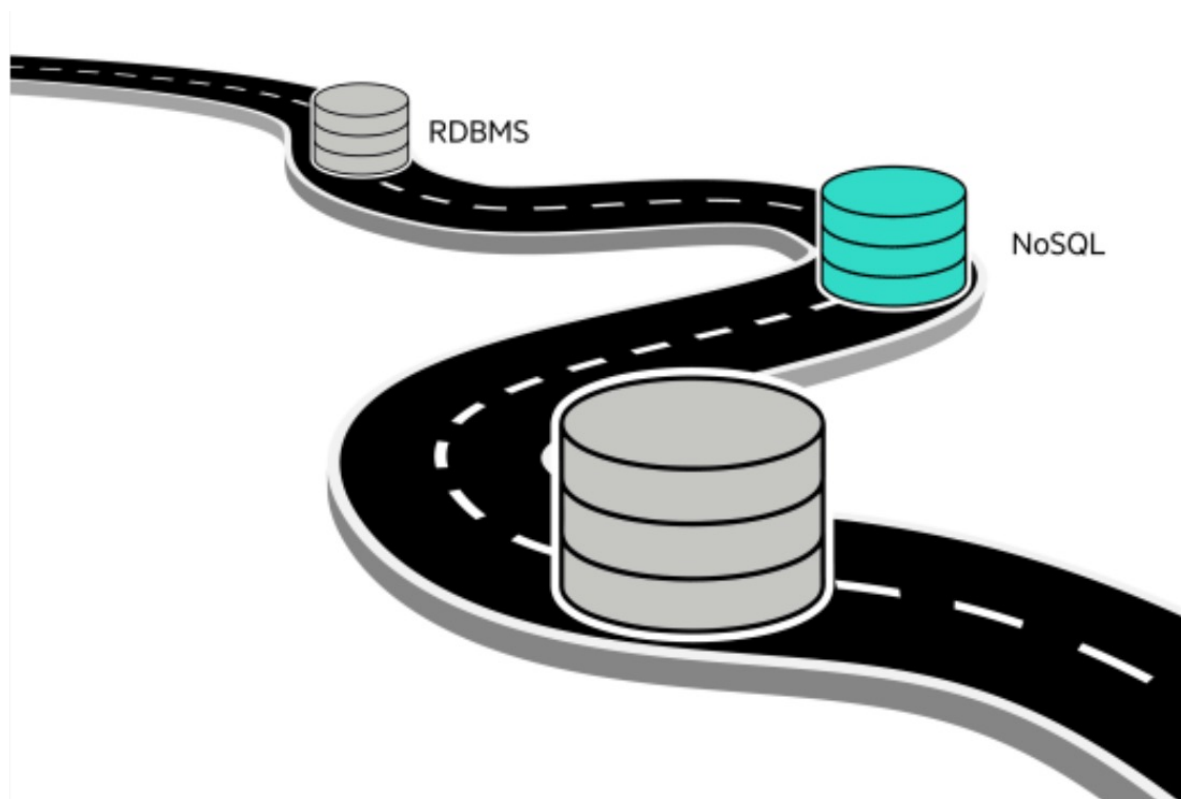


遠端辦公室資料儲存



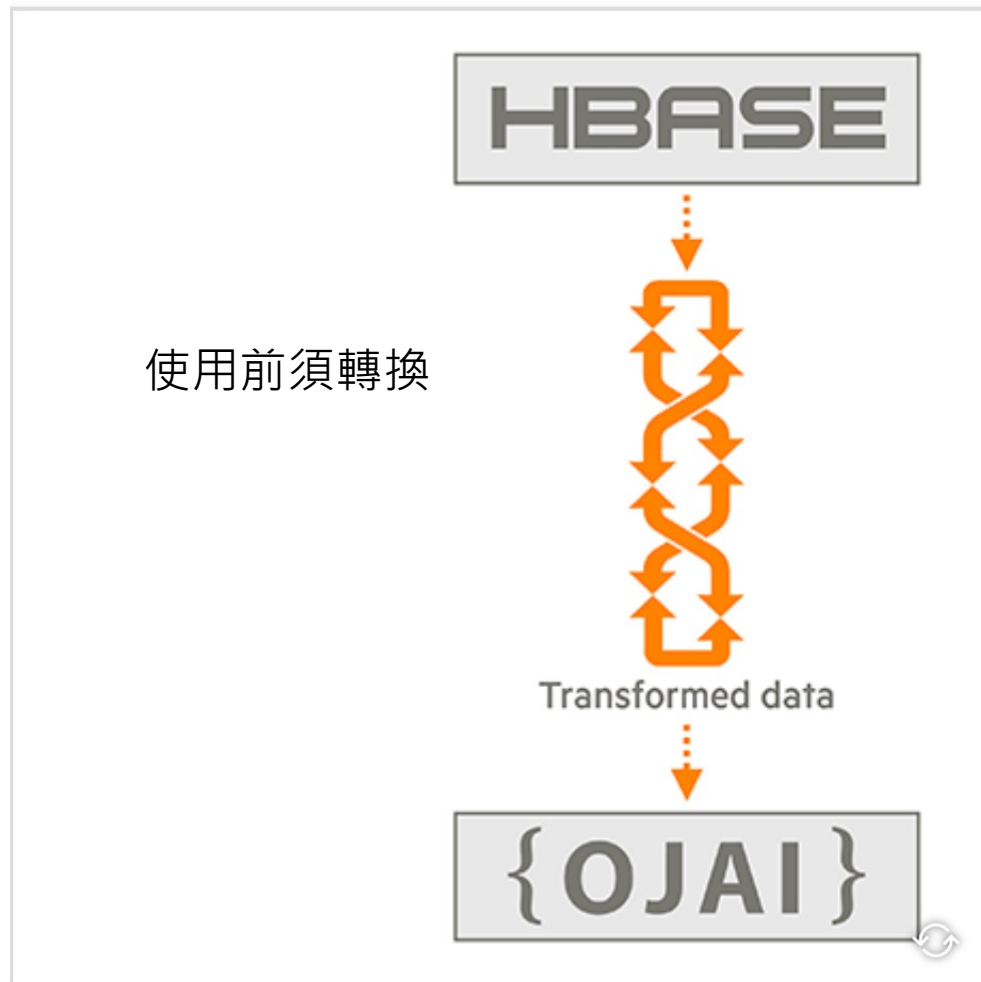
HPE EZMERAL DATA FABRIC DATABASE

EVOLUTION OF DATABASES SOLUTION: NOSQL

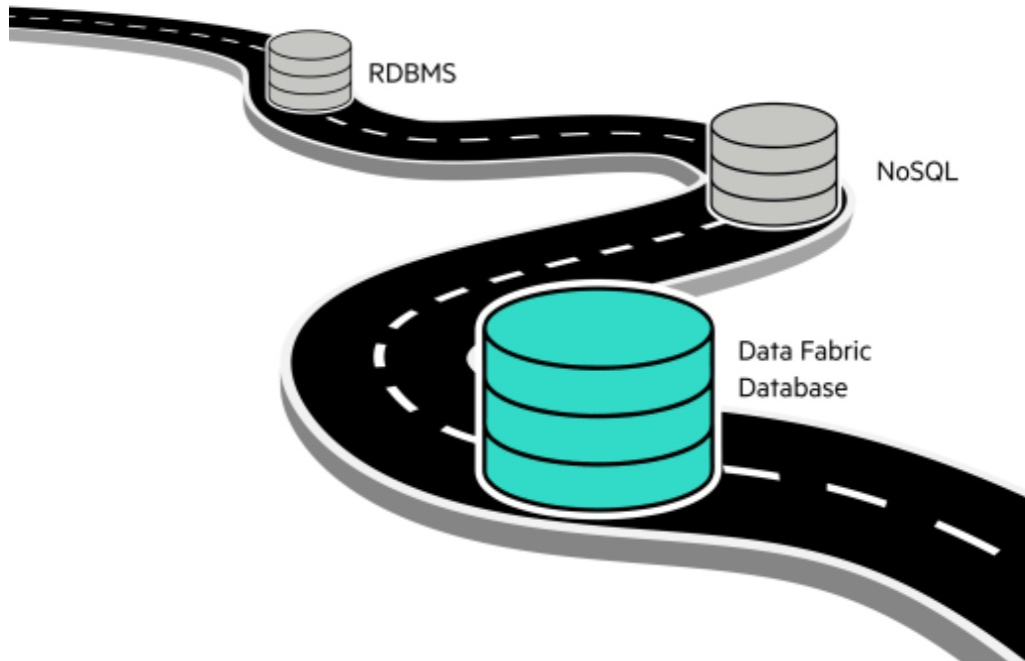


- NoSQL
 - 可更改DataSet
 - 版本控制
 - 適用於多種數據類型，包含非結構化數據

NOSQL PROBLEMS



EZMERAL DATA FABRIC DATABASE



- Ezmeral Data Fabric
 - Transparent, binary format
 - 可以儲存舊有的HBase tables
 - 適用於所有數據類型

為什麼使用**HPE EZMERAL DATA FABRIC DATABASE**

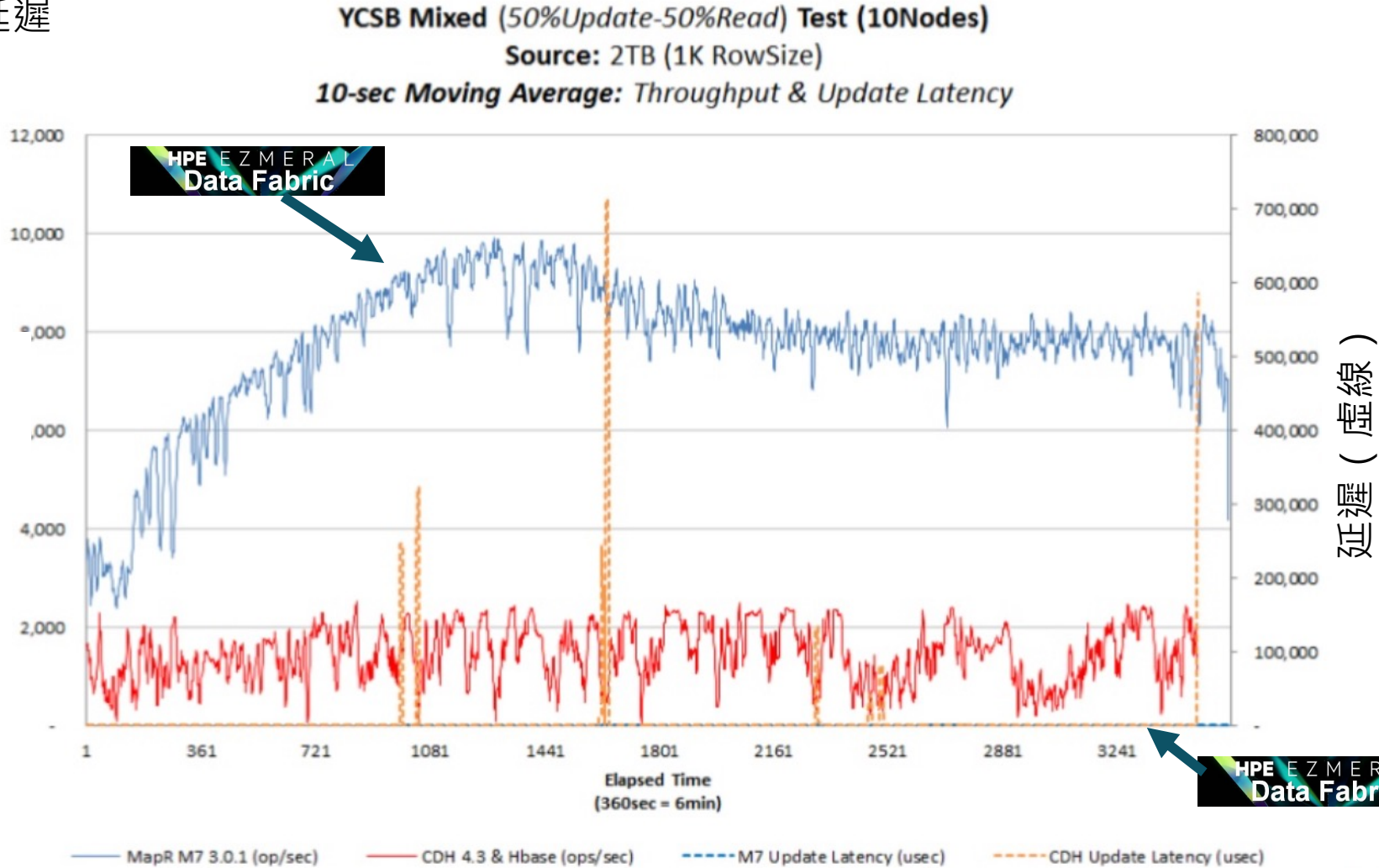
- 使用 **SQL** 進行集成分析：HPE Ezmeral Data Fabric 提供了低延遲、分佈式 SQL 查詢引擎。
- 彈性且集中化分析：Hadoop 和 Spark 於相同的叢集中從而提供更高效率的數據架構。
- 應用程式的全域訪問：資料表的應用程式訪問可分佈於異地端。
- 靈活的數據模型：支援多種數據形態的存取。
- 更高的存取效能：相較於NoSQL有額外的堆疊消耗，直接將資料寫入硬碟優化數據I/O。

Ezmeral Data Fabric 傳輸量和延遲都優於HBase

傳輸量和更新延遲

↑
傳輸量越高越好 (實線)

傳輸量



↓
延遲 (虛線)

延遲是越低越好

DATA FABRIC DATABASE ADVANTAGES

Feature	Data Fabric Database	HBase
Transparent Document Database Storage	✓	✗
High Availability at Mission-Critical Levels	✓	✗
Instant Recovery from Failures	✓	✗
Continuous Low Latency (No Compaction Delays)	✓	✗
Less Administration (No Processes to Manage, Self-tuning)	✓	✗
Online Data Protection (Snapshots, Mirroring)	✓	✗
Scalability (Number of Rows Supported)	Trillions	Billions

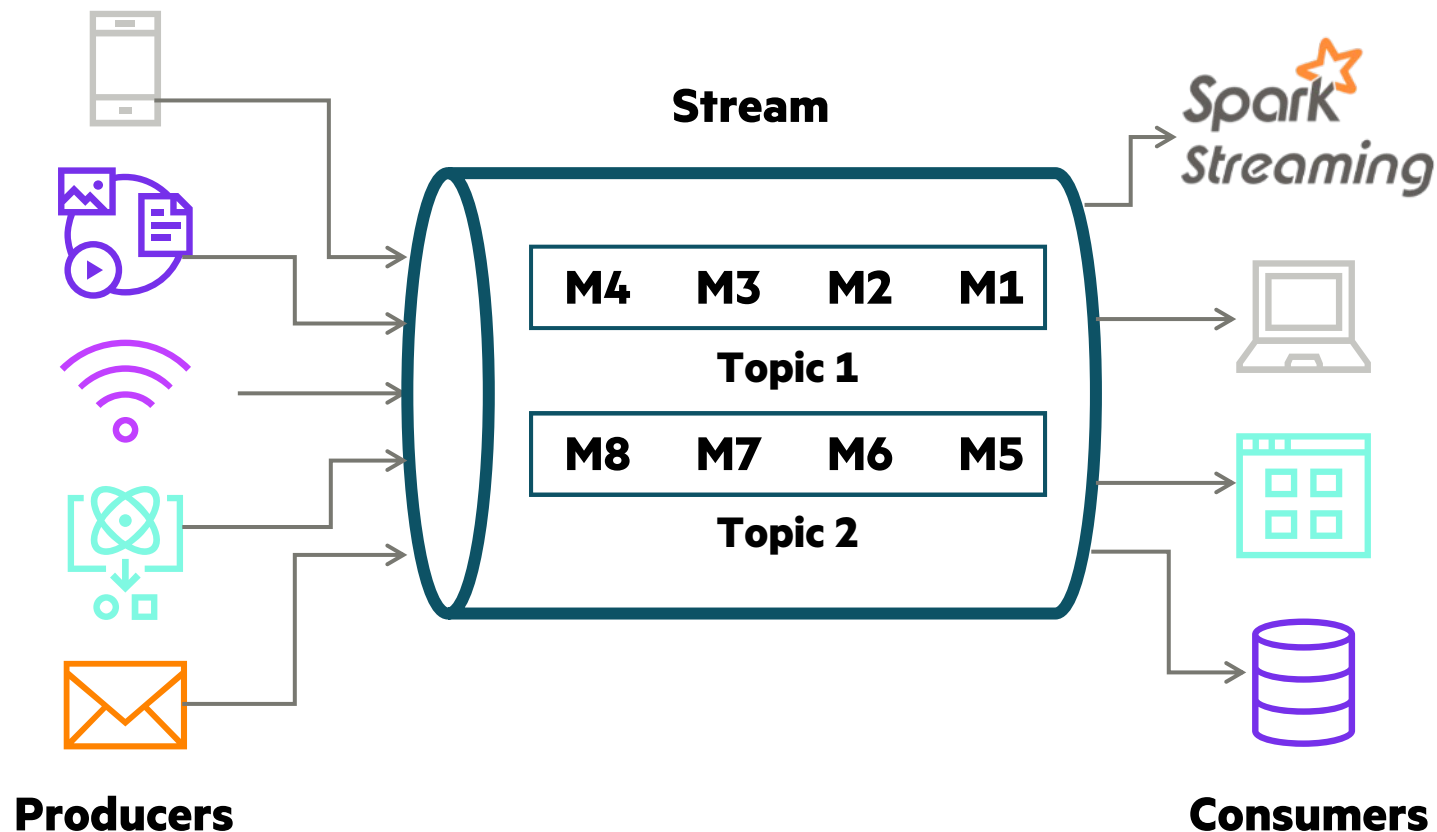
HPE EZMERAL DATA FABRIC EVENT STORE

HPE EZMERAL DATA FABRIC EVENT STORE 串流資料存放

Streams, topics, and partitions:

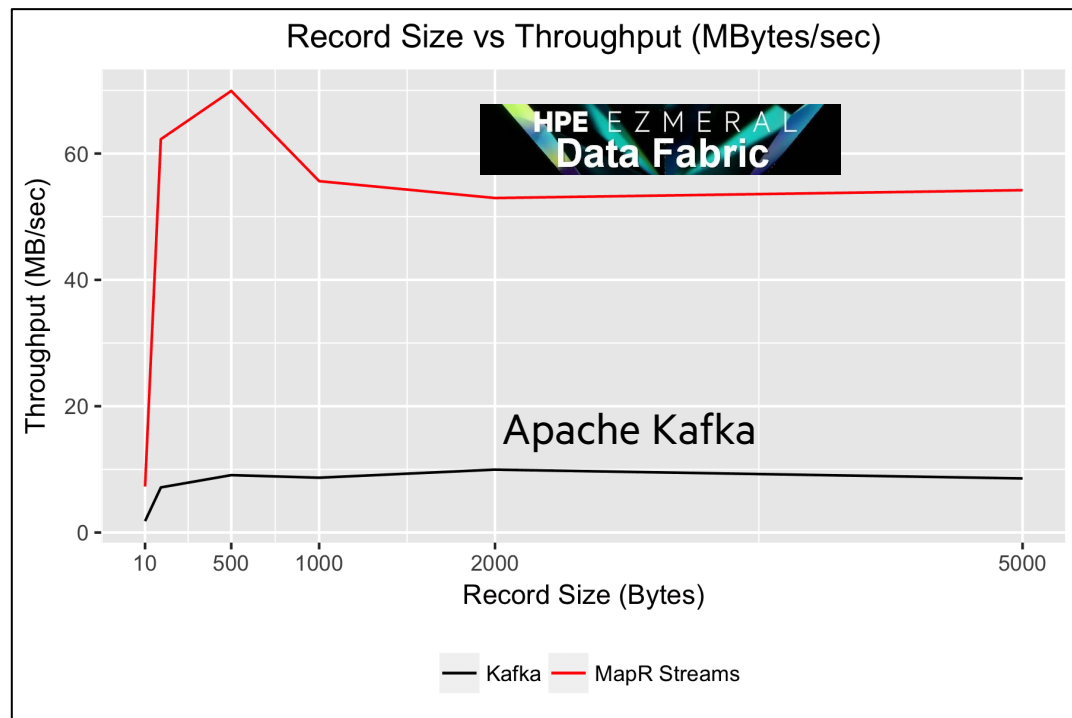
- Stream是一個邏輯和管理單元，其中包含許多主題。
- Topic是從一組Producers傳遞到一組consumers的事件Queue。
- 存在於Topic中的Partition是並行、有序、不可變的Message Queue。

Producers and Consumers 共用主題架構。

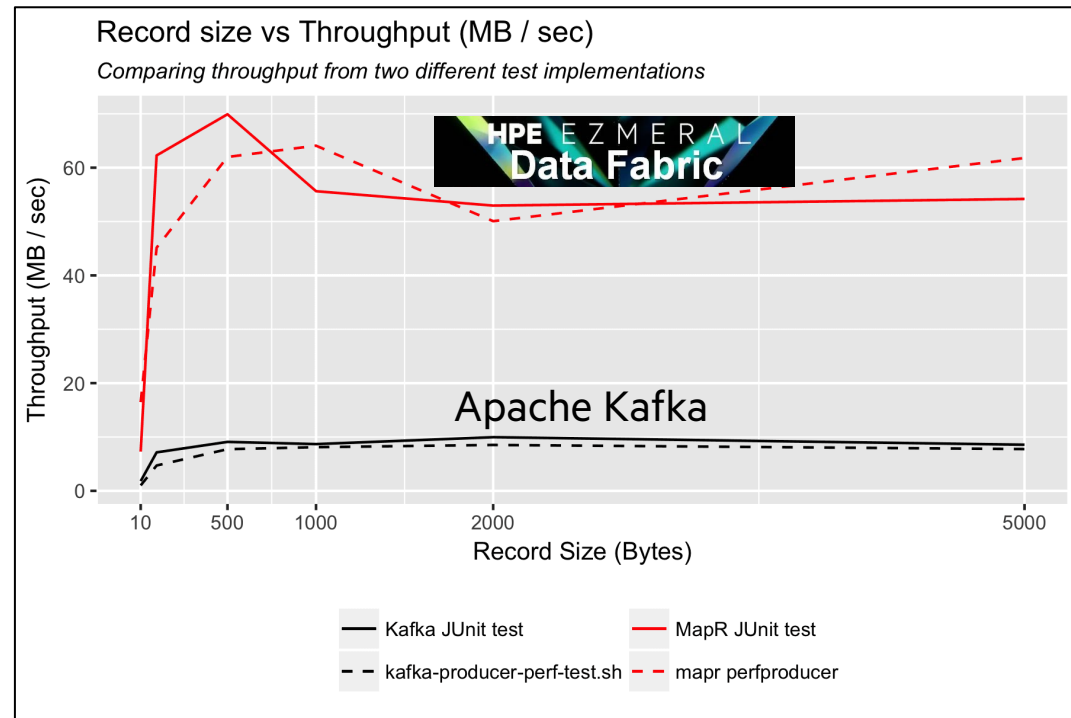


EZMERAL DATA FABRIC EVENT STREAMS VS. APACHE KAFKA

- Intel Xeon E5-2660 16核心, 2.2GHz
- 記憶體：112GB
- 3節點MapR v5.2



Data Fabric的輸送量比**Kafka**高得多



Data Fabric 傳輸的數據量和頻率高於**Kafka**

<https://developer.hpe.com/blog/vyvg31gkBKH6DoBpk6Ov/kafka-vs-mapr-event-store-why-mapr>

EZMERAL DATA FABRIC ECOSYSTEM PACK

EZMERAL DATA FABRIC ECOSYSTEM PACK

數據使用可以立即開始 快速供給各種應用情境



Analytics with Hadoop

批處理

Apache MapReduce
Apache Hive, Apache Pig
Apache Sqoop and Sqoop2
Apache Flume, Apache Hue



Analytics with Spark

即時處理

Apache Spark Core
Apache Spark SQL
Apache Spark Streaming
Apache Spark MLlib
GraphX, SparkR
Spark Operator for k8s



NoSQL DB

Wide-Column Database using HBase API
JSON document database using OJAI API
Multi-master table replication
Secondary indexes



串流消息傳送

Support for Kafka API
Kafka Connect
Kafka REST Proxy
KSQL, Kafka Streams
Kafka Schema Registry



互動式 SQL 查詢

Apache Drill : Schema-less
ANSI-compliant
distributed SQL query engine,
Queries on File

分散式文件系統、資料庫和串流

伺服器

伺服器

伺服器

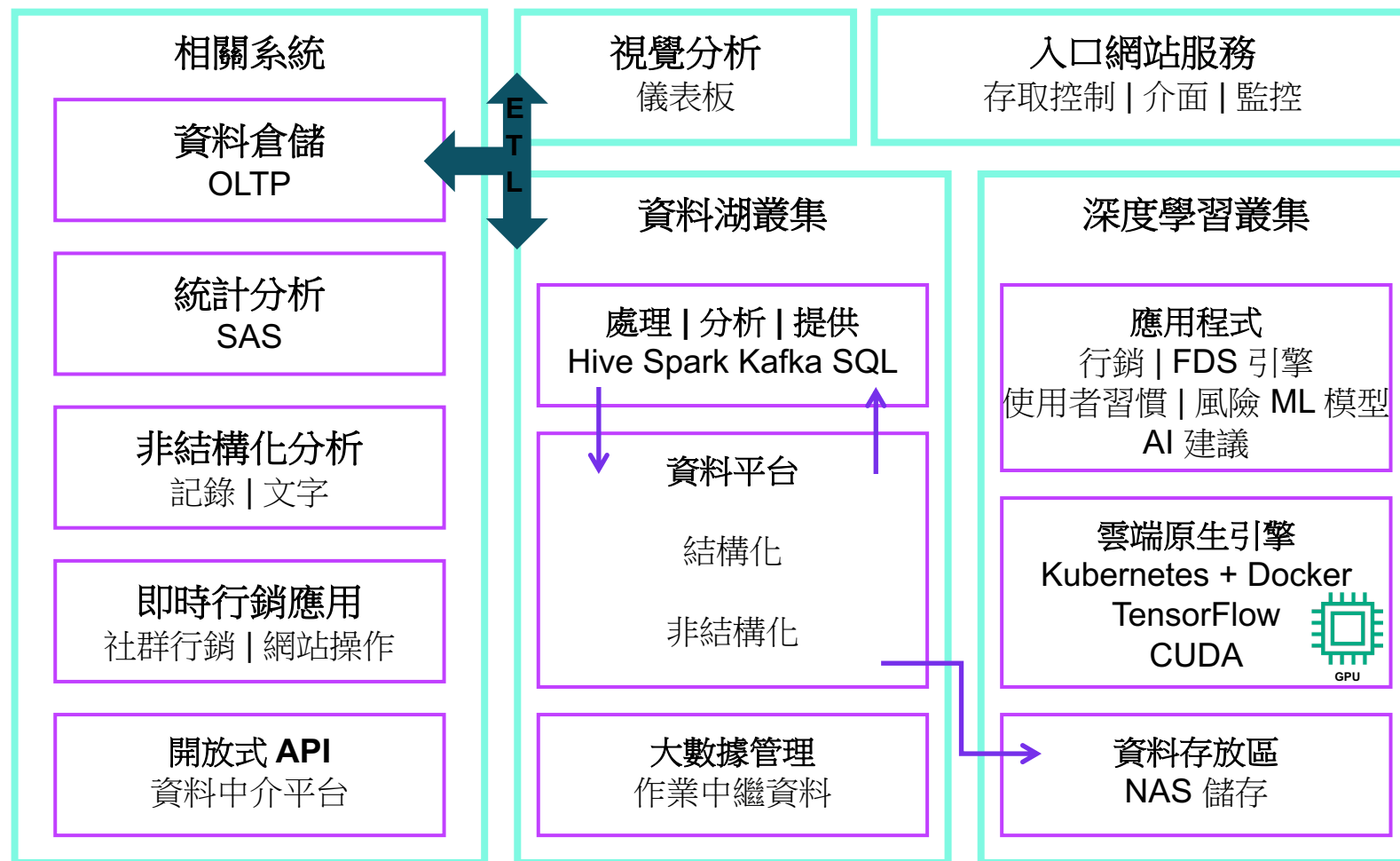
伺服器

- Ezmeral Data Fabric 生態系統，我們提供各種分析工具，使Data Lake不僅存儲數據。

- Data Fabric核心功能
- 提供分散式數據存儲
Software Defined Storage

資料優先現代化案例分享

- 合併目前孤島分散的資料
- One Platform減少維運人力
- 設計新服務改善業務流程
- 自助式資料取用，使團隊進行使用案例的創新



透過 HPE EZMERAL DATA FABRIC 幫助企業資料存儲面臨的挑戰



缺乏針對分析和 AI 進行優化
並可擴展、安全的數據平臺。

- *Open Interface (HDFS, NFS)*
- *CSI (Container Storage Interface)*
- *ETL (Extract, Transform, Load)*
- *FIPS Level 1 Compliance*

不可預測的成本和容量需求

- *HPE Green Lake Service*
- *最佳化保存成本(Data Tiering)*



不斷變化、不斷擴大的開
源生態系統。

- *Auditing 資料取用與治理*
- *Native S3 落地雲體驗*
- *Stream 串流資料*
- *Database 資料庫*

Edge-to-Cloud 各種混合部署模型

- *資料池管理(異地備援)*
- *Scale Out (依據儲存量擴展)*
- *Global Namespace(跨環境相同存取體驗)*

THANK YOU
