

HADOOP HBASE

國立臺北科技大學資訊工程系
郭忠義

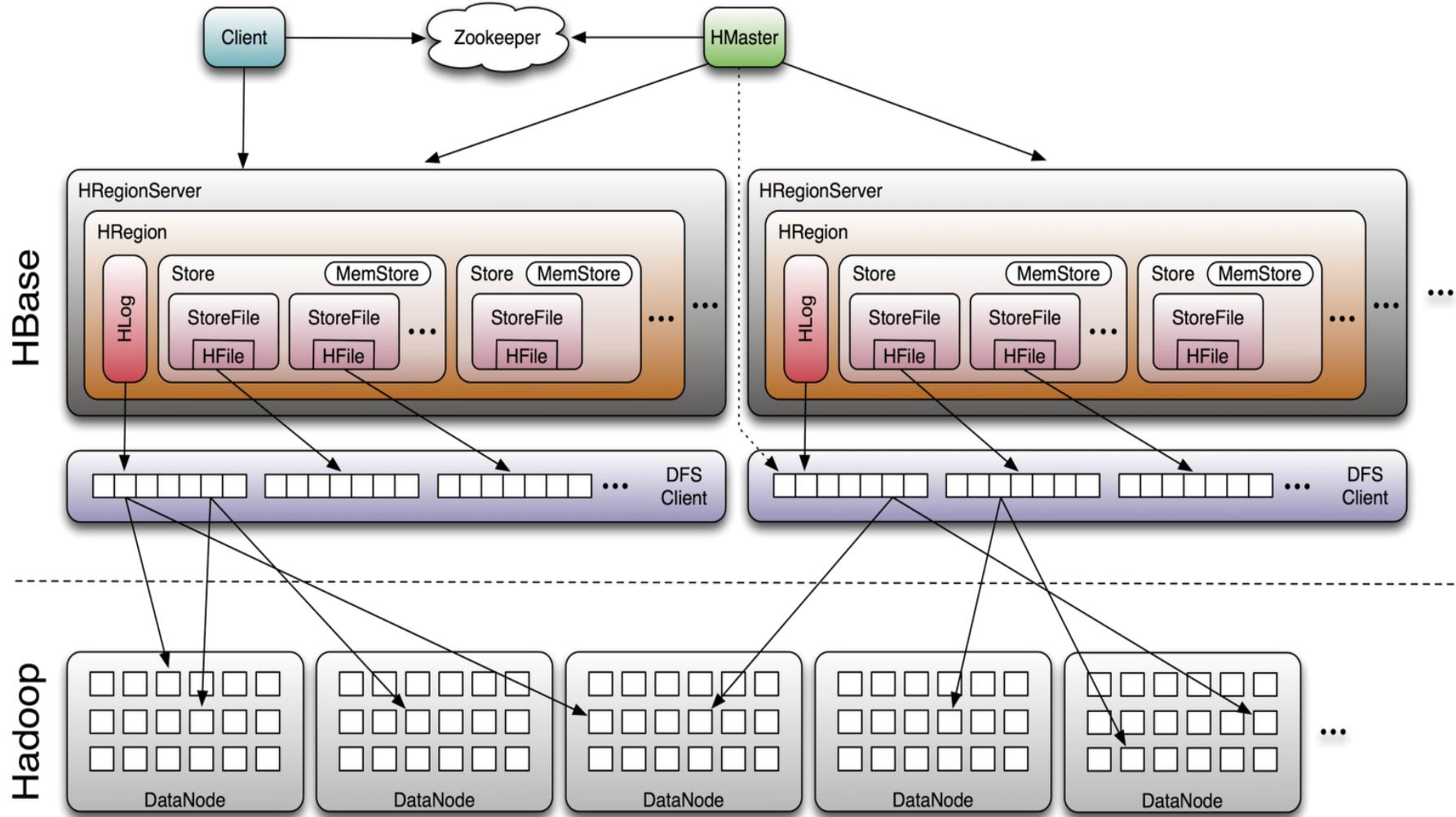
HBase 1/4

- 架構在HDFS分散式資料庫，與關聯式資料庫不同。
- 使用列 (row) 和行 (column) 為索引存取資料值，查詢時較像在使用 map 容器 (container)；
- 每一筆資料都有一個時間戳記 (timestamp)，同一個欄位可依不同時間存在多筆資料。
- 資料表 (table) 由許多 row 及數個 column family 組成，每個列都有一個 row key 做為索引；
- column family 是 column label 集合，內含多組 label，label 可視需要隨時新增，不用重新設定整個資料表。
- 存取資料表時，通常使用 ('row key', 'family:label') 或 ('row key', 'family:label', 'timestamp') 的組合取出需要的欄位。

HBase 2/4

name	timestamp	score	phone
ricky	T2	Math: 65	
	T5	Eng: 30	
james	T1		61234
	T3	Math: 90	
	T6	Math: 95	61235
jason	T4		64321

HBase 3/4



HBase 4/4

- 為方便分散資料和運算工作，將整個資料表分為許多region，一個region由一到數個列組成，分別存放在不同HBase主機，這些存放region的主機是region server，
- master server紀錄每一個region對應的region server；master server會自動將不能提供服務的region server上的region重新分配到其他region server上。
- HBase可供MapReduce程式當作資料來源或儲存媒介，提供TableMapper及TableReducer類別讓程式的Mapper及Reducer類別繼承，可把MapReuce中的(key, value)方便從HBase中取出和存入。