

●HBase

- 차드 월터스와 짐 캘러만에 의해 시작 (2006년)
- 구글의 ‘빅테이블: 분산스토리지 시스템의 논문 (2006년)
- 하둡 0.15.0의 하위 프로젝트로 등록 (2007년 10월)
- 아파치 최고 수준의 프로젝트로 승급 (2010년 5월)
- HBase의 실 서비스 : 어도비, 스템블 어폰, 트위터, 야후 등 (현재)



●특징

- Region 서버들간의 자동 복구
- 자동화된 Table Sharding
- 확장성(Scalability)
- Client 접속을 위한 쉬운 Java API 제공
- 강력한 읽기 / 쓰기 일관성

● 구조이해

●HBaseMaster

- RegionServer 관리 기능 (ON, OFF)
- RegionServer 오류 검사
- Region 설정 (ROOT, META)

●HRegionServer

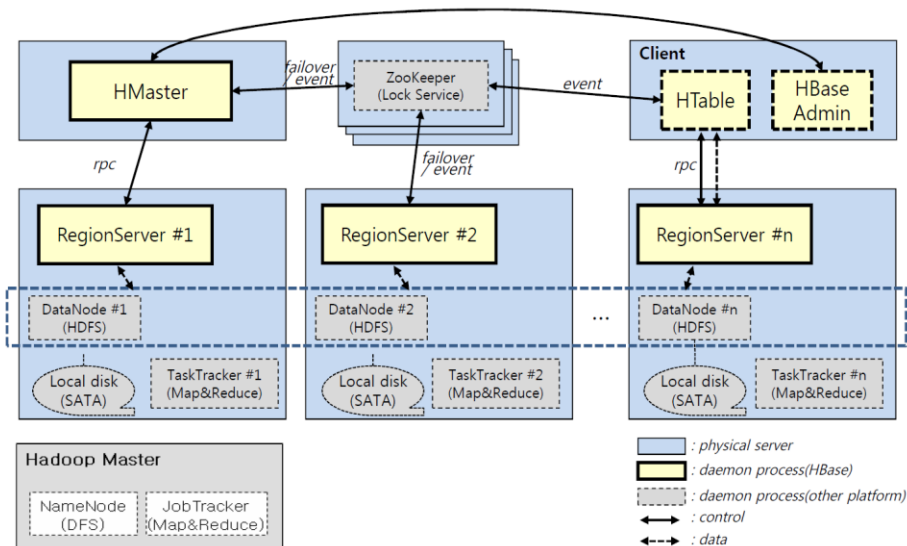
- Data Read / Write
- Cache 관리
- 데이터 서비스

●HBaseAdmin (Client)

- 테이블 추가/삭제
- 인스턴스 생성
- 스키마 변경

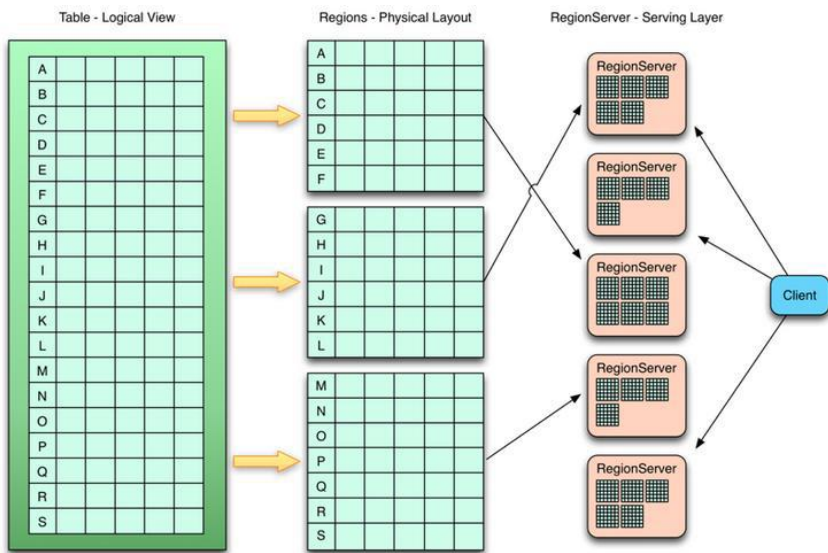
●Htable (Client)

- 액세스
- RegionServer에 대한 리스트 기록



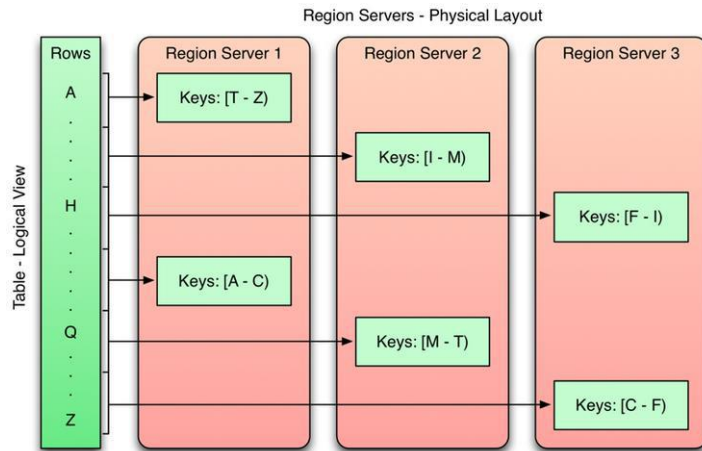
●오토샤딩

- Serving Layer : Hlog + Hfile + StoreFile + MemStore
- Physical Layout : DataNode + HDFS Client
- Logicla View : HBaseAdmin + HTable



●분산저장

- Physical Layout : HDFS + DataNode
- Logical View : ResionServer
- HBase의 확장 단위는 “Region”
- Load balancing과 failover를 제공
- 증가하는 데이터와 scale 확장인 경우에는 자동으로 분산됨



● 저장구조

< Logical Layer >

Row Key	Time Stamp	ColumnFamily contents	ColumnFamily anchor
"com.cnn.www"	t9		anchor:cnnsi.com = "CNN"
"com.cnn.www"	t8		anchor:my.look.ca = "CNN.com"
"com.cnn.www"	t6	contents:html = "<html>..."	
"com.cnn.www"	t5	contents:html = "<html>..."	
"com.cnn.www"	t3	contents:html = "<html>..."	

< Physical Layer >

물리 파일_1 (HDFS)

Row Key	Time Stamp	Column Family anchor
"com.cnn.www"	t9	anchor:cnnsi.com = "CNN"
"com.cnn.www"	t8	anchor:my.look.ca = "CNN.com"

물리 파일_2 (HDFS)

Row Key	Time Stamp	ColumnFamily "contents:"
"com.cnn.www"	t6	contents:html = "<html>..."
"com.cnn.www"	t5	contents:html = "<html>..."
"com.cnn.www"	t3	contents:html = "<html>..."

● 데이터 모델

- 셀 : row + column + timestamp
- 컬럼 : 키와 값으로 이루어진 데이터 구조
- 컬럼패밀리 : 컬럼들의 그룹
- Customer:name와 Customer:City는 Customer라는 컬럼패밀리의 멤버컬럼임.

The diagram illustrates a data model table with the following structure and annotations:

Row Key	Customer		Sales	
customer id	Name	City	Product	Amount
101	Jonh White	Los Angeles	Chairs	400.00
102	Jane Brown	Atlanta	Lamps	200.00
103	Bill Green	Pittsburgh	Desk	500.00
104	Jack Black	St. Louis	Bed	1600.00

Annotations:

- cells**: Points to the cell containing 'Jonh White'.
- Column Family**: Points to the 'Sales' column family.
- Column**: Points to the 'Amount' column.
- ROW**: Points to the row containing '102'.