

HBase

Agenda

1. Wprowadzenie do baz NoSQL na przykładzie HBase
2. Architektura
3. Model danych
4. Korzystanie za pomocą CLI
5. Metody dostępu do danych

NoSQL

- Czy to naprawdę oznacza NoSQL ?
- “Not only SQL”
- Nie spełnia zasad ACID
- Skalowalny
- Rozproszony
- “Flexible data model”

NoSQL in minus

- Dojrzałość systemów
- Brak profesjonalnego wsparcia
- Analityka nie zawsze jest łatwa

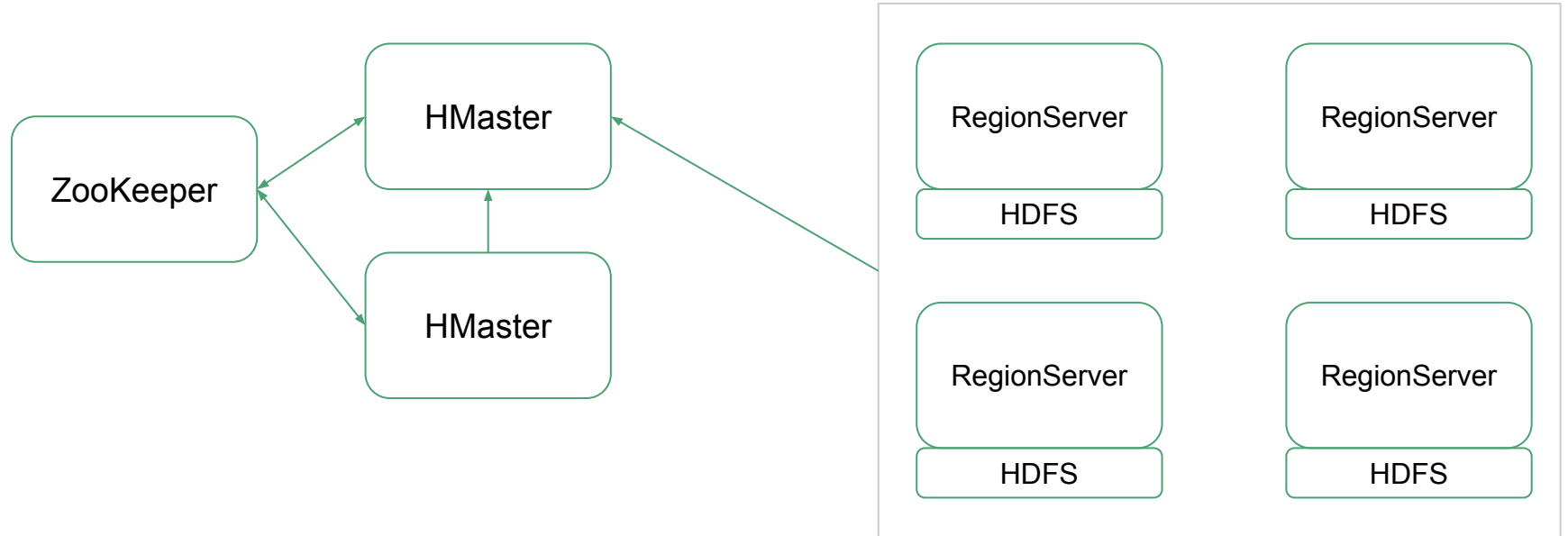
Klasyfikacja

Klucz - wartość	Apache Ignite, ArangoDB, BerkeleyDB, Couchbase, MemcacheDB, Project Voldemort, Redis, ZooKeeper, Geode
Dokument	Apache CouchDB, Couchbase, MongoDB, ElasticSearch, Solr
Rodzina kolumn	Amazon SimpleDB, Accumulo, Cassandra, Druid, HBase
Graf	Apache Giraph, Neo4J

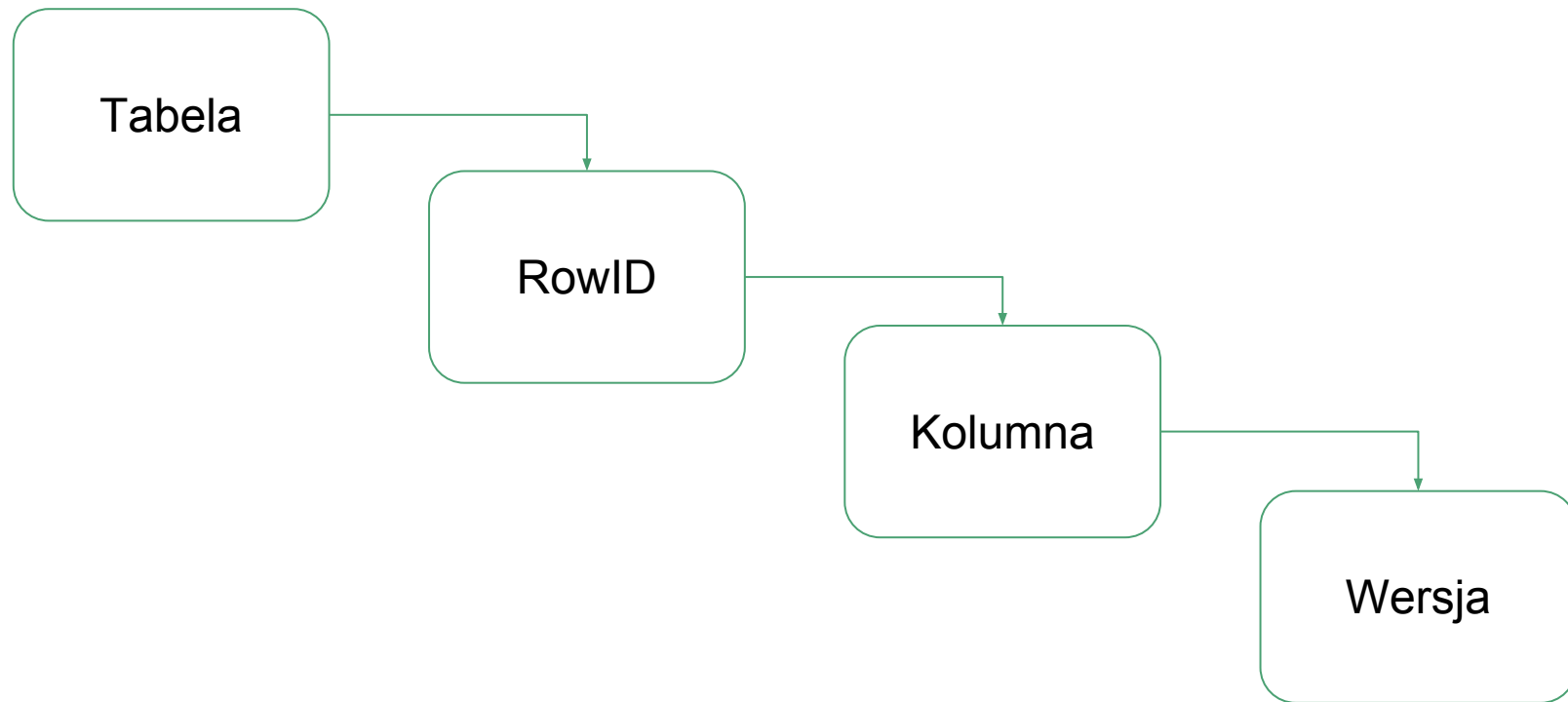
HBase - przykłady użycia

- Mozilla uses HBase to store all crash data in HBase
- Facebook - to store real-time messages
- Infolinks - to process advertisement selection and user events for the In-Text ad network, Infolinks uses HBase. It is an In-Text ad provider company. Moreover, to optimize ad selection, they use the reports which HBase generates as feedback for their production system.
- Twitter - A company like Twitter also runs HBase across its entire Hadoop cluster. For them, HBase offers a distributed, read/write the backup of all MySQL tables in their production backend. That helps engineers to run MapReduce jobs over the data while maintaining the ability to apply periodic row updates.
- Yahoo! - There HBase helps to store document fingerprint in order to detect near-duplicates

HBase Architektura



Model danych



Dostęp do danych

- HBase shell
- Spark
- MapReduce
- HBase API