ETH zürich

Week 5 - Column Stores

Aljaz Medic

Plan for today

1. Exercise Sheet Questions?
2. Quiz
3. HBase Recap (API, Physical Architecture, Optimizations)
4. BloomFilters
5. Old Exam Questions

QUIZ

HBase was inspired by Google's Bigtable.

True.

Give all of the four main design improvements in wide column stores compared to key-value stores.

Column families, versioning, flexible schema, multidimensional key model.

Column families in HBase group together data that is accessed together.

True.

QUIZ

HBase guarantees ACID properties at the column level.

False. (HBase guarantees ACID properties at the row level.)

Each row in HBase must contain the same set of columns across all rows in a table.

False. (Tables can be very sparse, rows may differ in their columns.)

HBase provides a SQL-like query language similar to relational databases.

False. (It provides a low-level API, similar to key-value stores.)

QUIZ

RDBMS are optimized for sparse access and scalable writes; HBase is for complex joins and strong consistency

False. (Other way around.)

Both key-value stores and wide column stores rely on a flat key model without data grouping.

False. (Wide column stores introduce an extra grouping layer with column families.)

In HBase, columns can be added dynamically as long as they belong to a pre-defined column family.

True.

QUIZ

HBase relies on a single master node (HMaster) to coordinate operations across the cluster.

True.

List all primary operations supported by HBase.

Get, Put, Scan, Delete

The 'Get' operation in HBase always returns exactly one value.

False. (If versioning is enabled, multiple versions can be returned.)

QUIZ

A region in HBase includes rows with IDs in the closed interval [min, max].

False. (Regions are of the form of left-open interval [min, max).)

Explain what happens when a region becomes too large.

RegionServer automatically splits the region into two.

Why is HBase fast even though HDFS has high latency?

Caching, Buffering, Bloom filters, Short-circuiting.

QUIZ

Each RegionServer is responsible for all the column families of a specific region.

True.

The HFile is a sorted list of key–value pairs ordered primarily by the row ID.

True.

Name all logical levels of storage in HBase.

Table → Region → Store → HFile → HBlock → KeyValue

QUIZ

HBlocks are another name for HDFS blocks.

False. (HFiles, can have many HBlocks (~64kB or more),

and are persisted using HDFS Blocks. (64-128 MB).)

HBase data is always stored directly on each RegionServer's local disk.

False. (RegionServers store data on HDFS.)

LSM trees make writes sequential and efficient while allowing fast reads through merges and indexing.

True.

QUIZ

HBase writes are always slower than reads because they must get sorted and hit the disk.

False. (Writes can be faster as they are buffered in memory (MemStore + HLog).)

What is the purpose of the Write-Ahead Log (HLog) in HBase?

Ensures durability of writes by logging changes before they are written to disk.

Describe what happens during an HBase ‘flush’.

MemStore contents are sorted and written to a new HFile on HDFS.

QUIZ

How does HBase maintain data locality when compacting HFiles?

By performing compactions on the same node to preserve short-circuit access to data blocks.

What fields make up an HBase key?

Row ID, Column Family, Column Qualifier, Version, Type

If a Bloom filter says a key is present in a file, it might still be absent.

True.

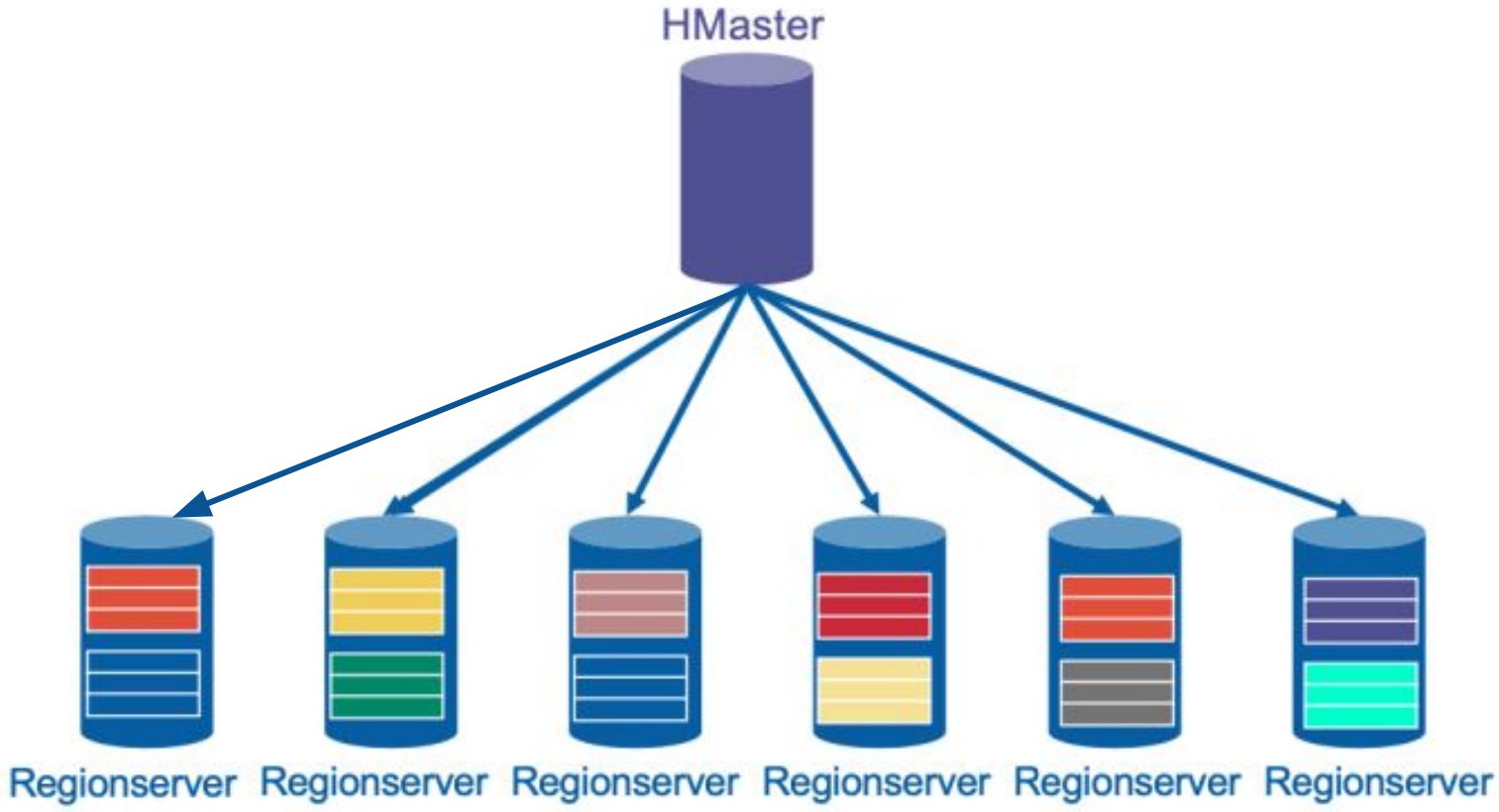
HBase - Logical Model

1. **Row ID**s are sorted
2. **Column families** are predefined, **Columns** added during runtime

Row ID	A	B	1	2	I
000					
002					
0A1					
1E0					
22A					
4A2					

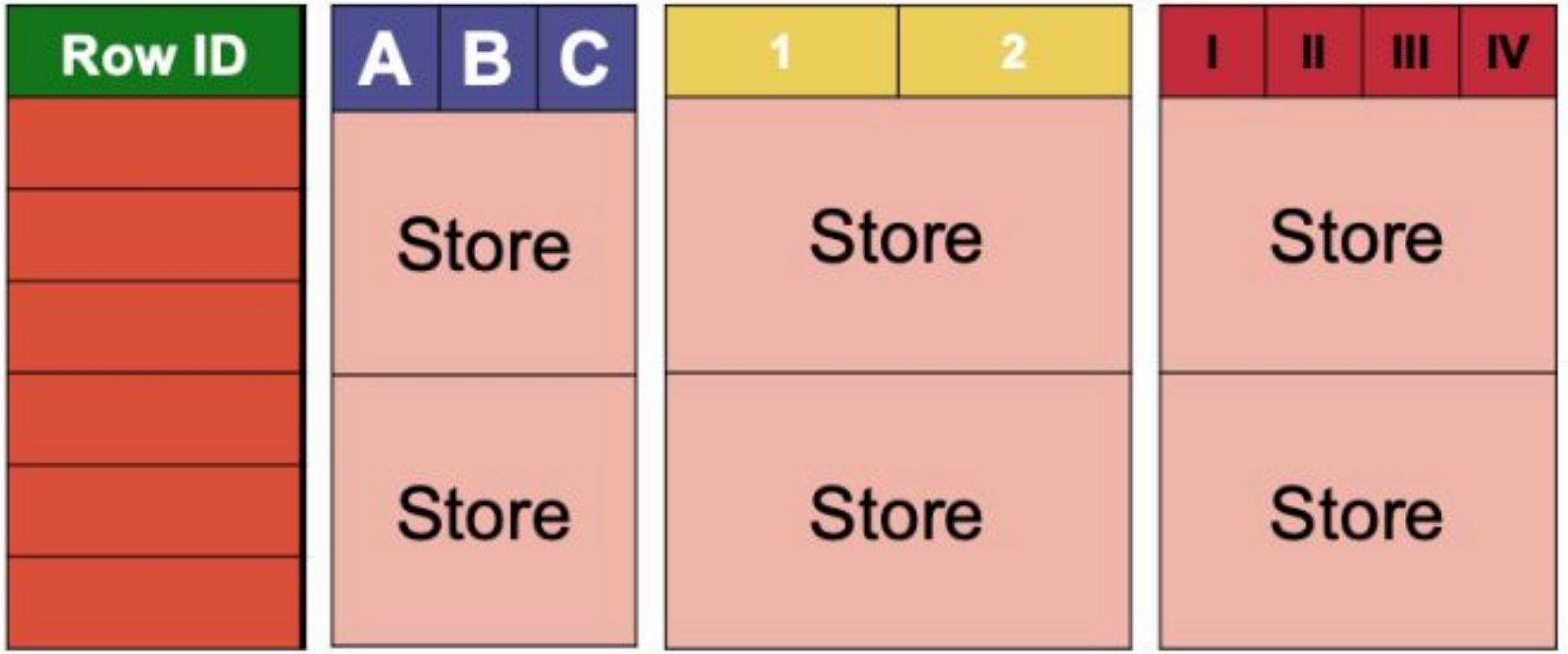
HBase - System Architecture

1. **HMaster** + **RegionServer**
2. Usually ran on the same machines as HDFS



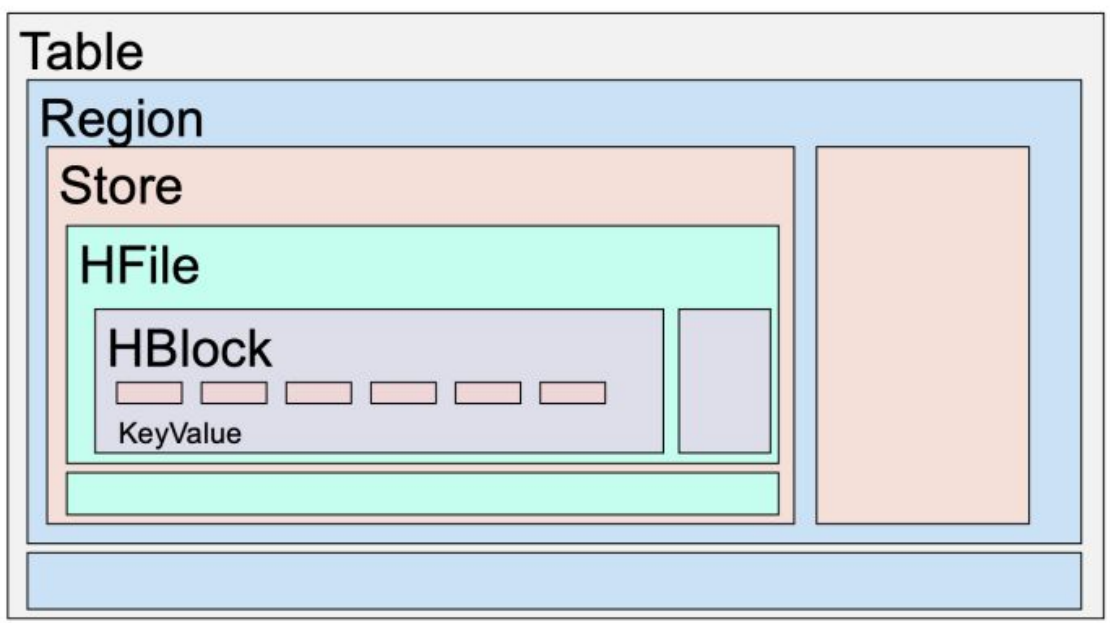
HBase - Physical architecture

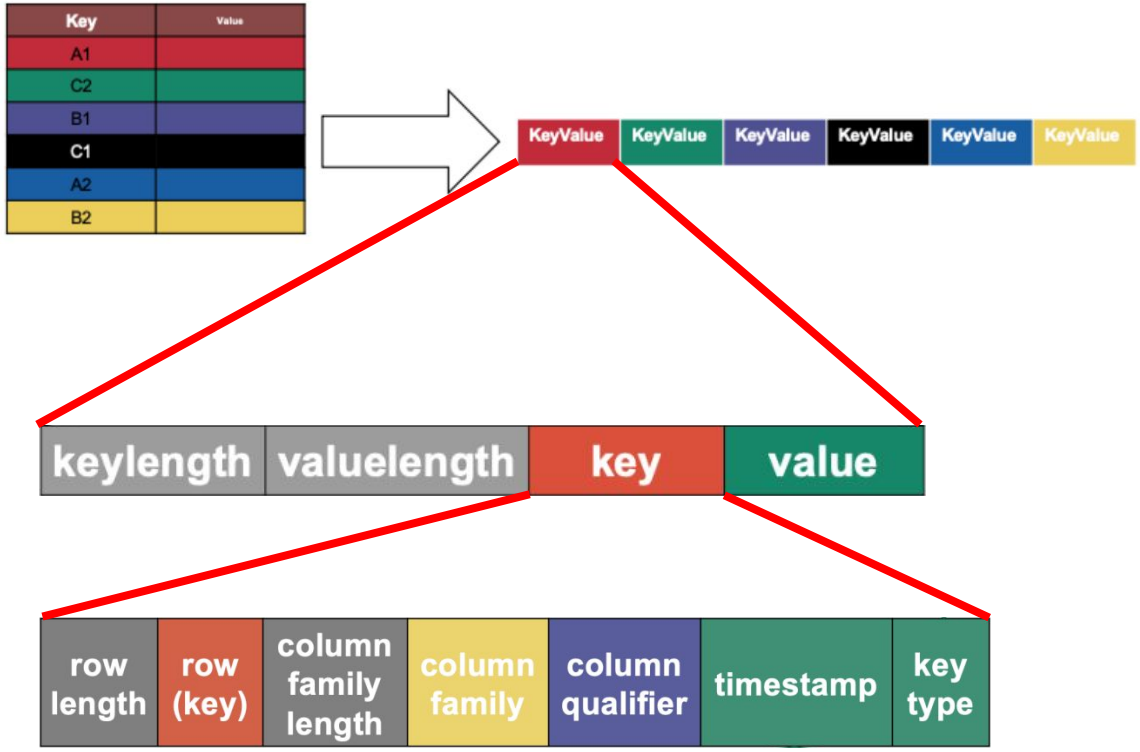
1. **Regions** of Row IDs [min, max)
2. Partition on **Region** x **Column family** = **Store**



HBase - Physical architecture

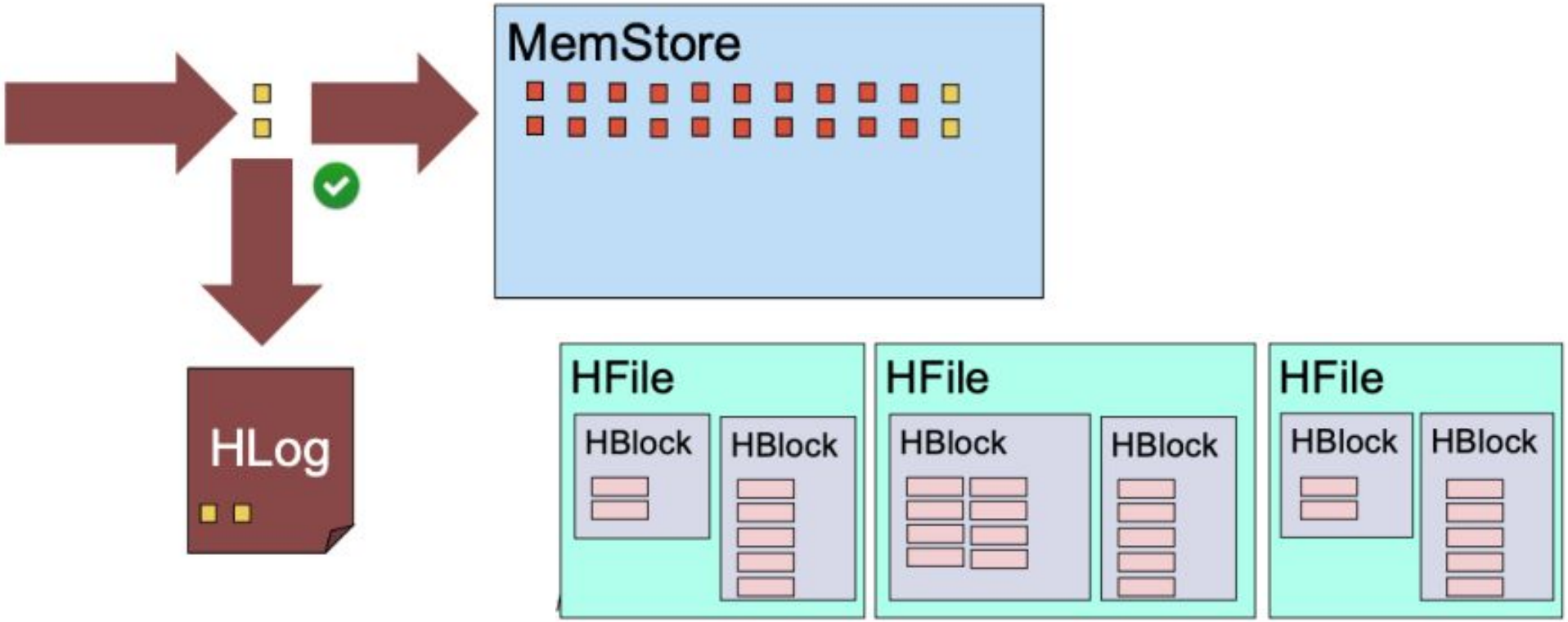
1. **HFile** contains index of **HBlocks**
2. **HBlocks** are sorted and contain a sequence of **KeyValue** objects





HBase - Read and Write

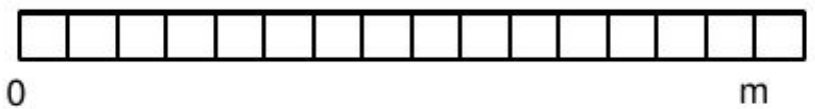
1. **Read:** From MemStores + HFiles (Bloom Filters)
2. **Write:** Through HLog to MemStore, sort there, flush sorted to HFile. Later compaction.



Bloom Filters: insertion

- Bit vector of size m
- k hash functions h_i that map to $\{0 \ldots m\}$

Fresh Bloom filter: empty array



Insert element a:

$h_0(a) = 2, h_1(a) = 7, h_2(a) = 3$

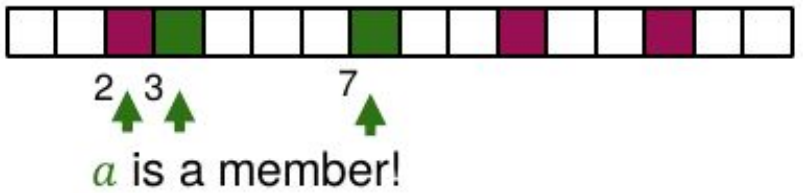
Insert element b:

$h_0(b) = 10, h_1(b) = 2, h_2(b) = 13$

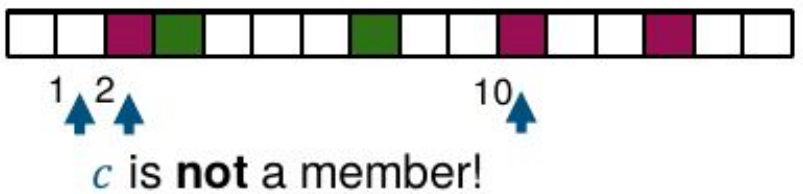
Source: Piet De Vaere, 2024

Bloom Filters: membership testing

Test membership of a:
$h_0(a) = 2, h_1(a) = 7, h_2(a) = 3$



Test membership of c:
$h_0(c) = 1, h_1(c) = 2, h_2(c) = 10$



Problem: False positives (FPs) on membership tests

Source: Piet De Vaere, 2024

HS23 Exam Q23: Bloom Filters

For each of the following statements on Bloom filters, indicate whether it is True or False.

True	False	
○	○	A hash function mapping an input string to its length is a good choice for the hash function in the Bloom filter.
○	○	A hash function mapping each input string to a unique value could be a good choice for the hash function in the Bloom filter.
○	○	Using more hash functions in a Bloom filter can decrease the likelihood of false negative errors.
○	○	The primary use case of Bloom filters is to speed up updating existing cells in HBase. Filters help the RegionServer to find the HFile with the current cell value which should be updated.

False. (Hotspots, collision resistance, …)

True. (Discussion.)

False. (No FN, only FP)

False. (Bloom filters are used for lookup, not updating)

HS24 Exam Q19: HBase

True	False		
○	○	It is possible to restrict a scan to an interval of rows in HBase.	True.
○	○	The total number of columns per table is limited to 256 in HBase.	False.
○	○	A row ID, column family and version always uniquely identify a value stored in HBase.	False. (Also column qual.)
○	○	Data stored in HBase is often denormalized.	True.

HS23 Exam Q24: HBase

○	○	In case of a server failure, HBase will lose all the data stored on the failed server.	False.
○	○	With log-structured merge trees, we more often merge smaller files than larger ones.	True.
○	○	A region is assigned redundantly to multiple RegionServers, the default being 3.	False.
○	○	A RegionServer may be responsible for multiple regions at the same time.	True.

ETH zürich

See you next week!

Aljaž Medič
amedic@ethz.ch

Slides

Suggestions