

Apache HBase At Netease

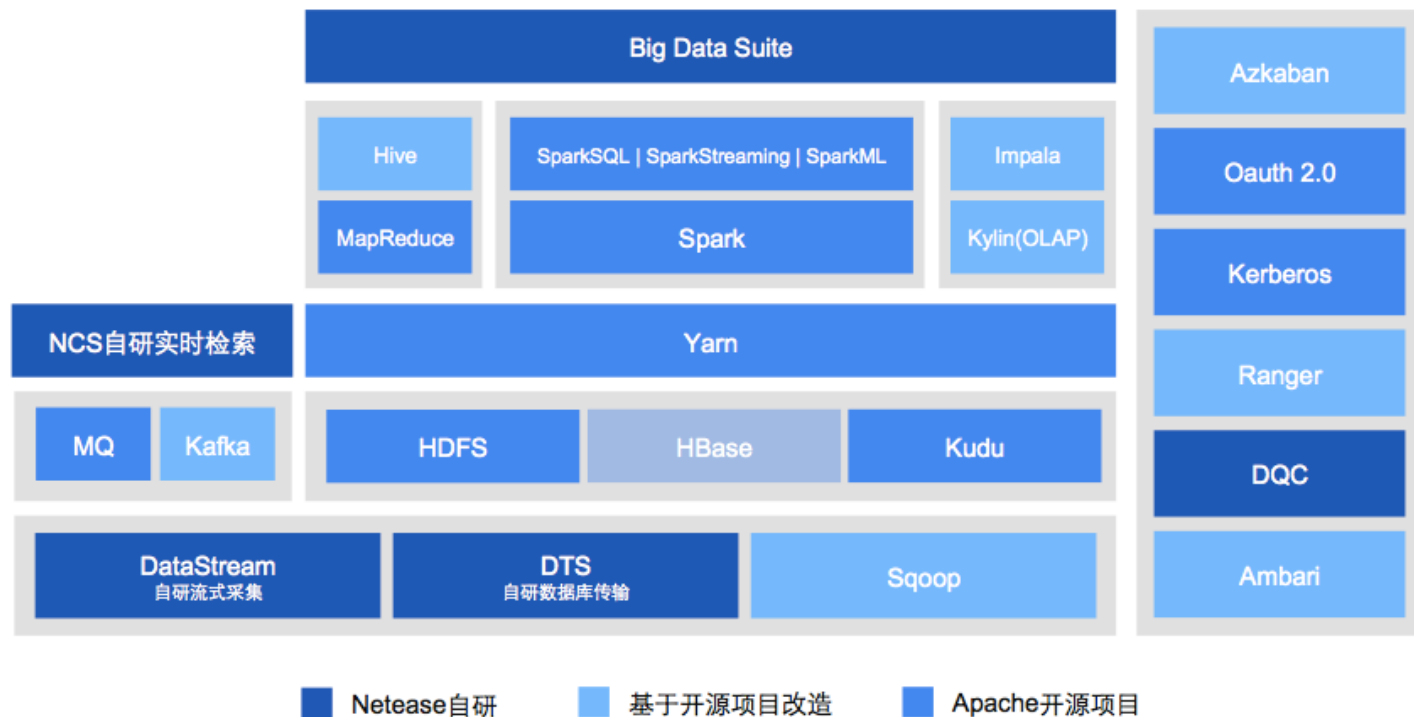
XINXIN FAN,HONGXIANG JIANG



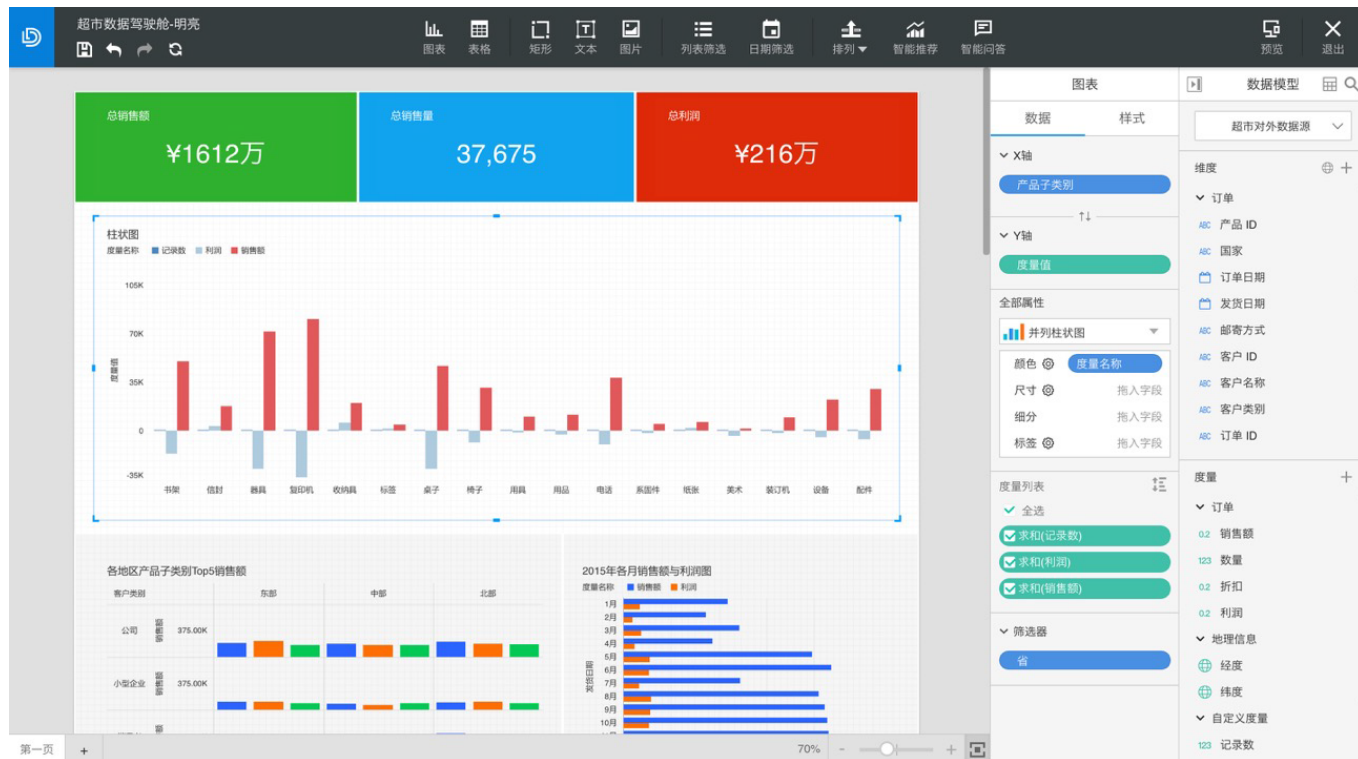
Agenda

- Overview HBase Service In Netease
- Key Practices Over HBase
- What We Have Done To HBase
- What We Are Doing Now

BigData System In Netease - mengma



BigData System In Netease - youdata

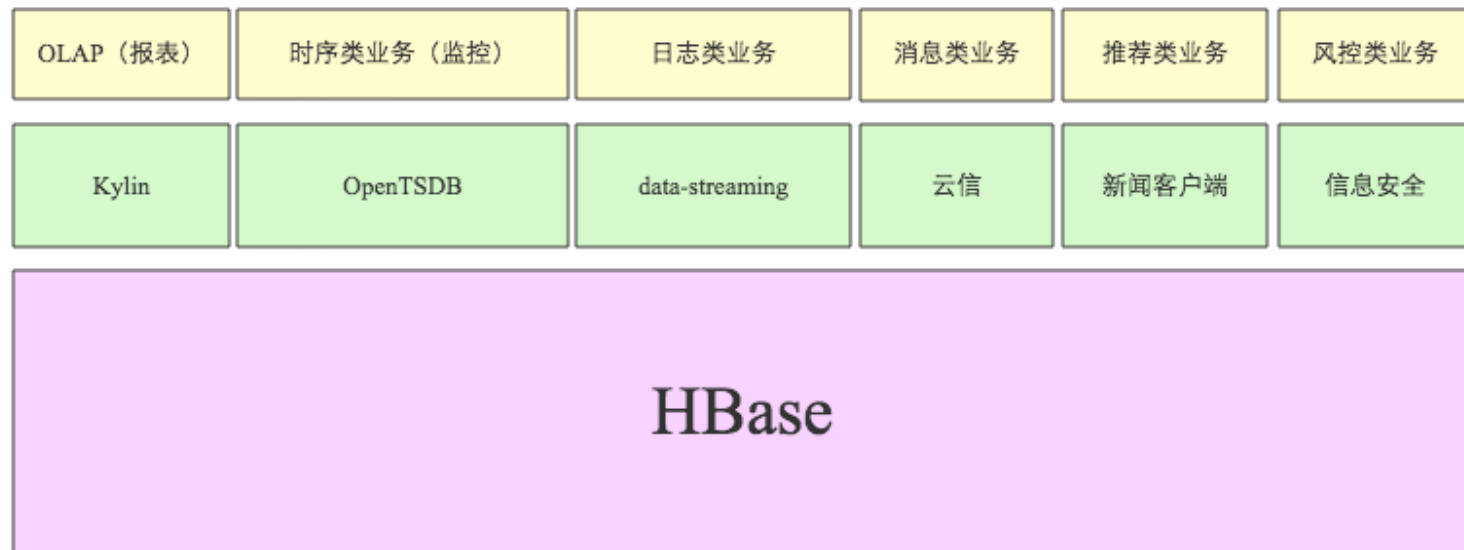


HBase In Netease

HBase Users come from 6 major departments, more than 40 different applications



HBase In Netease



HBase In Netease

7 HBase Clusters

200+ RegionServers

Hundreds of Terabytes Data

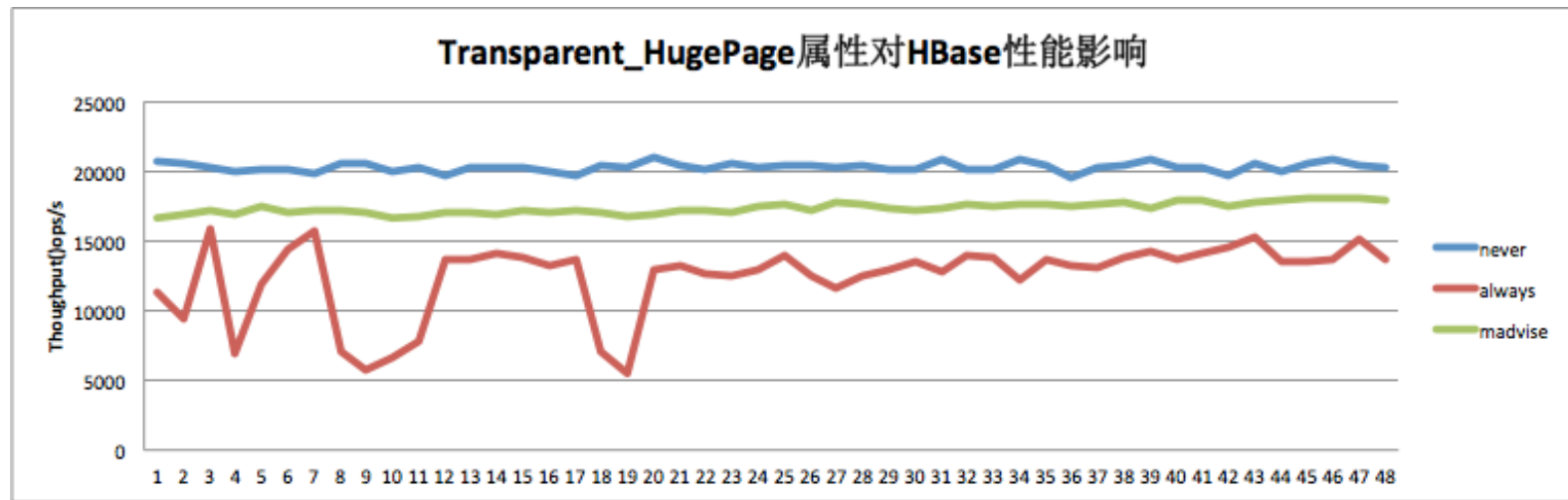
Agenda

- Overview HBase Service In Netease
- Key Practices Over HBase
- What We Have Done To HBase
- What We Are Doing Now

Key Practices - Linux System

- Tuning transparent huge pages (THP) off
- Set `vm.swappiness = 0`
- Set `vm.min_free_kbytes` to at least 1GB
- Disable NUMA zone reclaim with `vm.zone_reclaim_mode = 0`

Key Practices - Linux System



Key Practices - Schema

✧ Not Use PREFIX_TREE DATA_BLOCK_ENCODING !!!

- HBASE-12959 : compact never end
- HBASE-12817(fixed) : Data missing while scanning

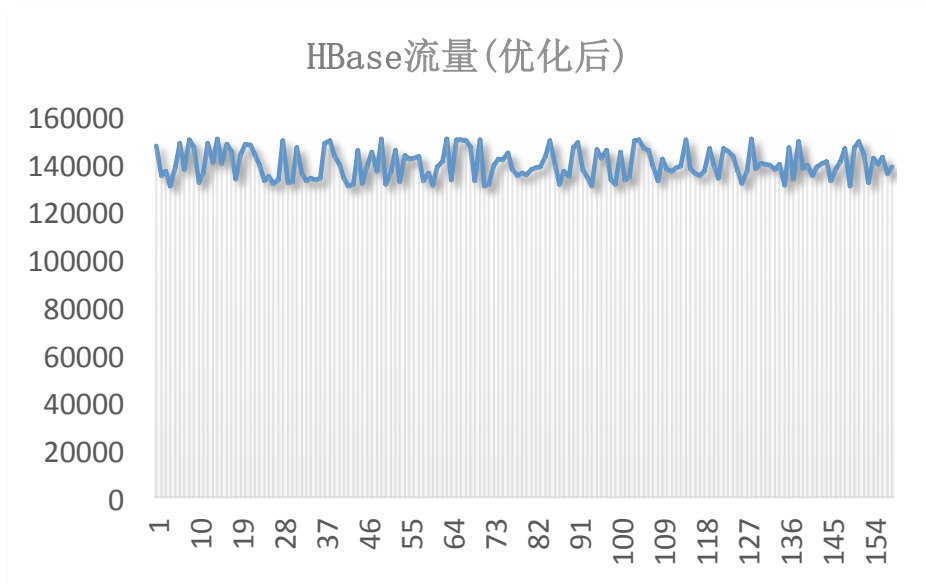
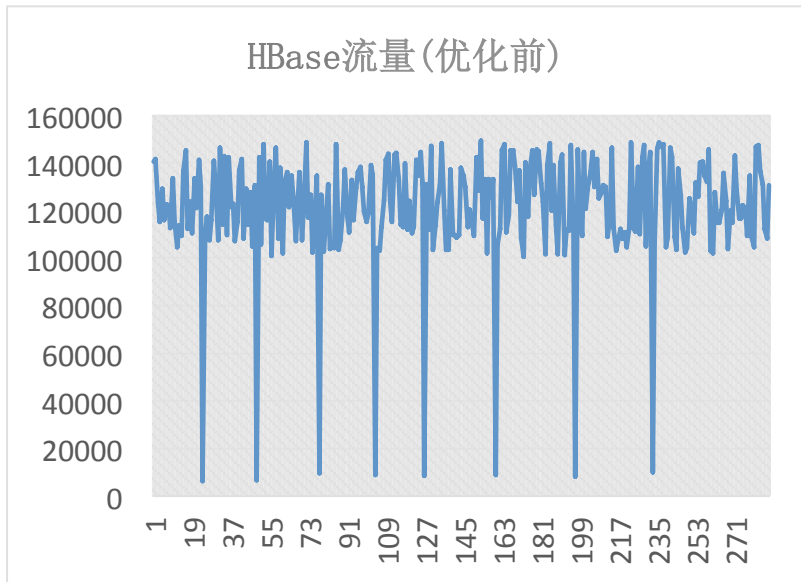
Key Practices - Schema

✓ Use More Useful Table-Level Configuration !!!

- MAX_FILESIZE
- MEMSTORE_FLUSH_SIZE
- DFS_REPLICATION

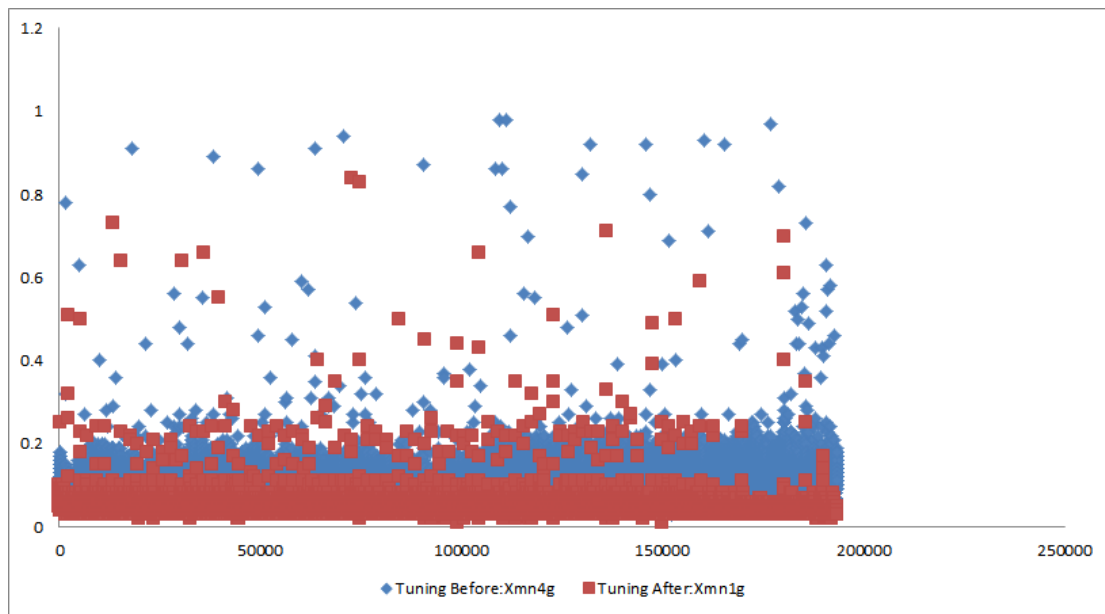
Key Practices - GC

✓ Use BucketCache(Offheap) Instead of LRUBlockCache !!!



Key Practices - GC

✓ CMS GC ---- Xmn : 1~3G, -XX:SurvivorRatio=2



Agenda

- Overview HBase Service In Netease
- Key Practices Over HBase
- What We Have Done To HBase
- What We Are Doing Now

Request Queue At Table-Level



Different workloads may influence each other frequently!

- The write requests with large fields may influence the small write requests
- The scan requests with high throughput may influence the other scan requests

active handlers preemption?

assign the independent request queue to the large requests

Request Queue At Table-Level

```
hbase(main):012:0* help 'create_resqueue'  
Create resourcequeue; pass resourcequeue name,  
and optionally a dictionary of resourcequeue configuration.  
Examples:
```

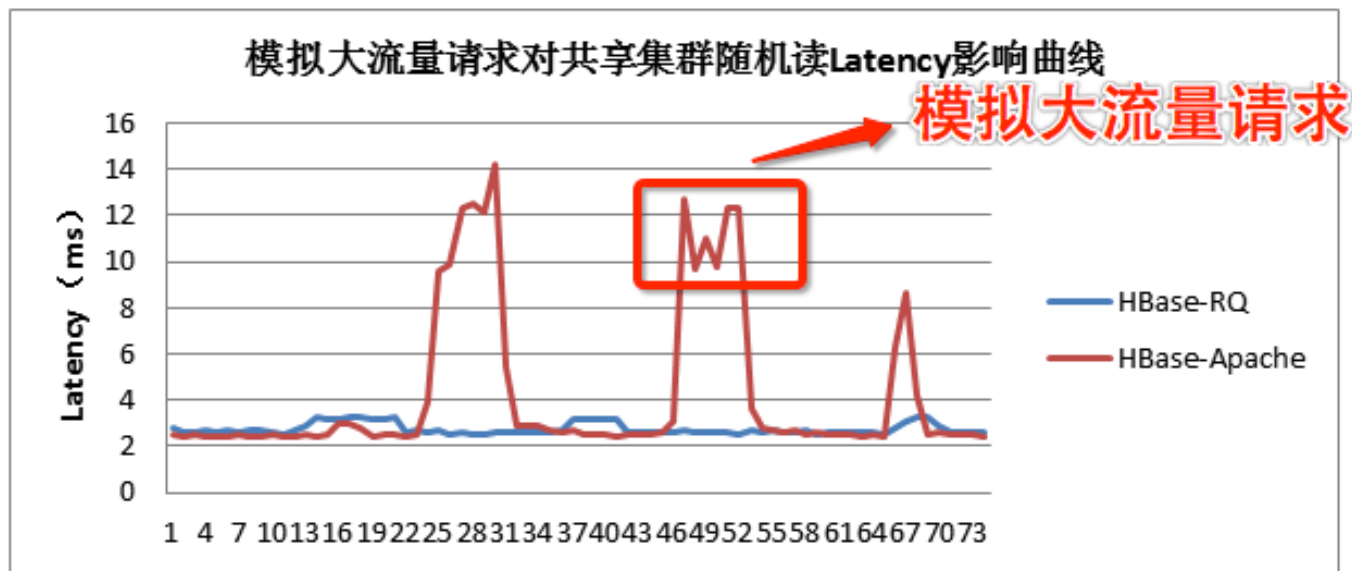
```
hbase> create_resqueue 'rq1'  
hbase> create_resqueue 'rq1', {'handler' => 5}
```

```
hbase(main):017:0* help 'set_resqueue'  
set a table request to a resourcequeue.  
Syntax : set_resqueue <table>,<request>,<resqueue>  
  
request is either zero or more letters from the set "GSW".  
GET('G'), WRITE('W'), SCAN('S')
```

For example:

```
hbase> set_resqueue 'table1', 'W', 'rq1'  
hbase> set_resqueue 'table1', 'GSW', 'rq1'
```

Request Queue At Table-Level



Improvement – Table Metrics View

- RegionServer Metrics? Region Metrics?
- Sometimes, Table Metrics is more Useful!

Improvement – Table Metrics View



Improvement – Table Metrics View

Tables

Request metricsStorefile MetricsMemstore MetricsCompaction MetricsCacheStats Metrics								
Table Name	Num. Stores	Num. Storefiles	Num. KV	Storefile Size Uncompressed	Storefile Size	Index Size	Bloom Size	
ar:corpora_meta_data	2	4	18997	1m	0m	1k	4k	
ar:corpora_uid	2	0	0	0m	0m	0k	0k	
cloud_music:user_portrait	42	313	4975568561	565150m	212316m	621728k	1982046k	
da:uda_group_members	7	1	3362	0m	0m	1k	4k	
da:uda_summary	14	0	0	0m	0m	0k	0k	
da:uda_user	6	16	233995552	21343m	4623m	25536k	18288k	
da:uda_user_events	210	1170	20530513174	9656560m	1224820m	13292659k	783238k	
da:uda_users	12	70	212931467	20765m	7826m	21895k	68424k	

Improvement – Table Metrics View

Tables

Request metrics

Storefile Metrics

Memstore Metrics

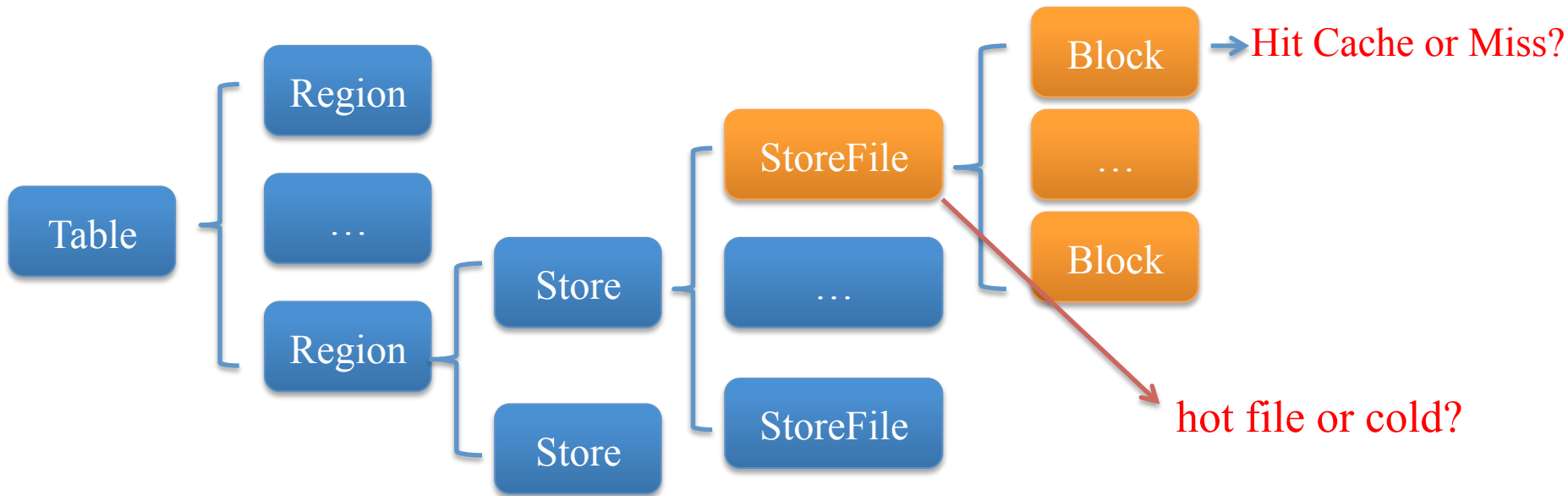
Compaction Metrics

CacheStats Metrics

if meta miss a lot, it's very dangerous

Table Name	Meta Hit Count	Meta Miss Count	Meta Hit Ratio	Data Hit Count	Data Miss Count	Seeks Per Second	Data Hit Ratio
ar:corpora_meta_data	24	4	85.71%	358	27	0	92.99%
ar:corpora_uid	0	0	0.00%	0	0	0	0.00%
cloud_music:user_portrait	8422300	23292	99.72%	0	10250534	0	0.00%

Improvement – Table Metrics View



Hot Files , You may do more

- Compaction Policy Based on Hot Files?
- Hierarchical Storage Policy Based on Hot Files?

Improvement – Table JMX Metrics

"Namespace_kaola_table_item_crm_metric_readRequestCount" : 29319660,
"Namespace_kaola_table_item_crm_metric_writeRequestCount" : 68654293,
"Namespace_kaola_table_item_crm_metric_totalRequestCount" : 97973953,
"Namespace_kaola_table_item_crm_metric_memstoreSize" : 5214610520,
"Namespace_kaola_table_item_crm_metric_storeFileSize" : 11222922684,
"Namespace_kaola_table_item_crm_metric_tableSize" : 16437533204,
"Namespace_kaola_table_item_crm_metric_storeFileCount" : 354,
"Namespace_kaola_table_item_crm_metric_storeFileSeekCount" : 3476326,
"Namespace_kaola_table_bi-baitiao-open_metric_readRequestCount" : 54253051,
"Namespace_kaola_table_bi-baitiao-open_metric_writeRequestCount" : 516266836,
"Namespace_kaola_table_bi-baitiao-open_metric_totalRequestCount" : 570519887,
"Namespace_kaola_table_bi-baitiao-open_metric_memstoreSize" : 10805889040,
"Namespace_kaola_table_bi-baitiao-open_metric_storeFileSize" : 115598190760,
"Namespace_kaola_table_bi-baitiao-open_metric_tableSize" : 126404079800,
"Namespace_kaola_table_bi-baitiao-open_metric_storeFileCount" : 415,
"Namespace_kaola_table_bi-baitiao-open_metric_storeFileSeekCount" : 38784299,



Others

- Check and Merge the empty region periodically
- Set the Request Priority per table
- More configuration set to **Table-Level**
 - ◆ **COMPACTION_THRESHOLD**
 - ◆ **MAJOR_COMPACTION_PERIOD**

What We Are Doing Now

- Inverted Index
- RegionServer Group
- Highly Available HBase

Improvement – InvertedIndex

	basic					
	username	age	school	city	...	
uid1						
uid2						
...						

✓ select * from table where uid = 'xxx'

select * from table where school = 'shenzhen' and age > 30

Full Table Scan

Improvement – InvertedIndex

InvertedIndex

{
shenzhen : <uid1, uid3, ... uidy>
wuhan : <uid2, uid5, ... uidx>
...
}

where city = "wuhan"

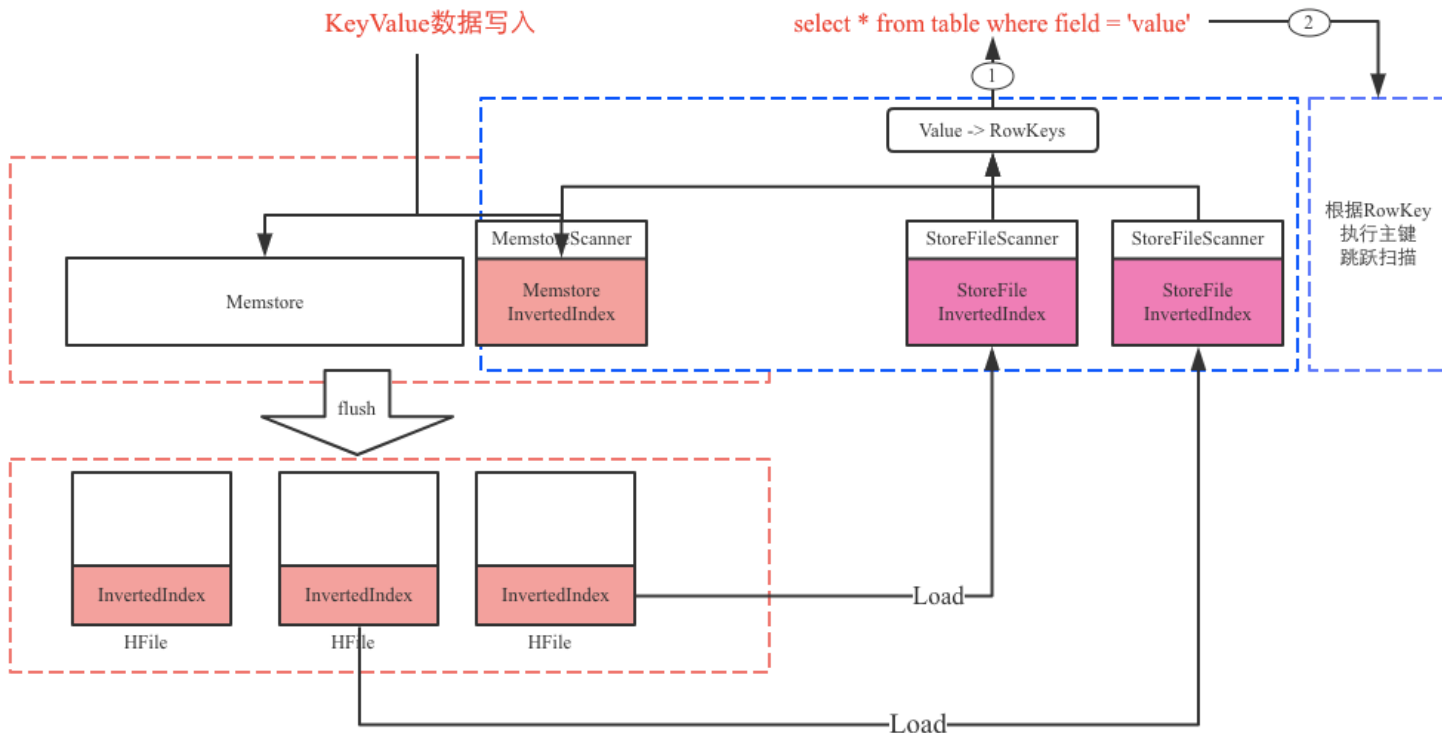
Rowkeys :

<uid2, uid5, ... uidx>

select * from user where uid in (uid2,uid5,...uidx)

<user2, user5,...userx>

Improvement – InvertedIndex



Improvement – InvertedIndex

- One billion KVs . 50 columns and 4 columns setted inverted index
- Write Throughput drops 5%~15%
- SingleColumnValueFilter Scan 
 - latency of without Index : 200s+
 - latency of with Index : 60ms



Thanks You!

<http://www.hbasefly.com>

<http://hbase-help.com>