

Jak postavit nástroj pro datovou analýzu

Lukáš Antalov, Vedoucí týmu vývoje



SEZNAM.CZ

Outline

- Čo to je webová analytika?
- Webová analytika @ Seznam.cz
 - Zber dát
 - Spracovávanie
 - Agregácia
- Hadoop
 - HDFS
 - MapReduce
 - Hive
 - HBase

**K čomu slúži
webová
analytika?**



Začiatky boli jednoduché

Návštevy	
Celkem	1133788
Týden	3635
Dnes	834
Online	427
Pořadí	
Celkem	3806
V kategorií	26
Počítáno v auditu	

Welcome to my web site.
You are visitor number

021529

Analýza logov webserveru

IP Address	Date	Request	Stat...	Size	Country	Referer
207.46.204.182	5/25/2010 7:52:45 PM	GET /A9F9FC5049156E18FDD64D47B06EA_00000.temp0085.htm HTT...	404	279	United States	-
123.125.68.91	5/25/2010 8:08:51 PM	GET /redir_help.php HTTP/1.1	302	20	China	-
207.46.13.97	5/25/2010 8:35:50 PM	GET /flock/beta/768B0A235FE1A432C564F7923D6D8C11_00000.temp0072...	404	289	United States	-
68.97.7.234	5/25/2010 8:41:37 PM	GET /robots.txt HTTP/1.1	200	66	United States	-
68.97.7.234	5/25/2010 8:41:39 PM	GET / HTTP/1.1	200	1428	United States	-
76.120.1.77	5/25/2010 8:56:39 PM	GET /robots.txt HTTP/1.1	200	66	United States	-
76.120.1.77	5/25/2010 8:56:40 PM	GET / HTTP/1.1	200	1428	United States	-
216.129.119.44	5/25/2010 8:59:07 PM	GET /faq.txt HTTP/1.1	200	943	United States	-
207.46.199.53	5/25/2010 8:59:49 PM	GET /A9F9FC5049156E18FDD64D47B06EA_00000.temp0034.htm HTT...	404	279	United States	-
207.46.204.186	5/25/2010 9:02:47 PM	GET /A9F9FC5049156E18FDD64D47B06EA_00000.temp000c.htm HTT...	404	279	United States	-
64.79.89.130	5/25/2010 9:47:18 PM	GET / HTTP/1.1	200	3759	United States	-
64.79.89.130	5/25/2010 9:47:19 PM	GET /favicon.ico HTTP/1.1	200	2238	United States	-
119.63.198.14	5/25/2010 9:57:51 PM	GET /robots.txt HTTP/1.1	200	64	Japan	-
207.46.13.47	5/25/2010 10:16:17 PM	GET /A9F9FC5049156E18FDD64D47B06EA_00000.temp0041.htm HTT...	404	279	United States	-
220.181.7.56	5/25/2010 10:33:20 PM	GET / HTTP/1.1	200	3759	China	-
123.125.66.84	5/25/2010 10:35:20 PM	GET / HTTP/1.1	200	3759	China	-
207.46.13.87	5/25/2010 10:37:46 PM	GET /A9F9FC5049156E18FDD64D47B06EA_00000.temp0010.htm HTT...	404	279	United States	-
207.46.13.44	5/25/2010 10:58:48 PM	GET /flock/beta/768B0A235FE1A432C564F7923D6D8C11_00000.temp000b...	404	288	United States	-
123.125.66.97	5/25/2010 11:38:46 PM	GET / HTTP/1.1	200	3759	China	-
220.181.7.67	5/25/2010 11:38:47 PM	GET / HTTP/1.1	200	3759	China	-
207.46.199.199	5/26/2010 12:09:34 AM	GET /robots.txt HTTP/1.1	200	66	United States	-
220.181.7.73	5/26/2010 12:35:57 AM	GET / HTTP/1.1	200	3759	China	-

Google Analytics

Google Analytics

[My Account](#) | [Help](#) | [Contact Us](#) | [Sign Out](#)

Dashboard

▶ Saved Reports

Visitors

Traffic Sources

Content

Goals

Settings

✉ Email

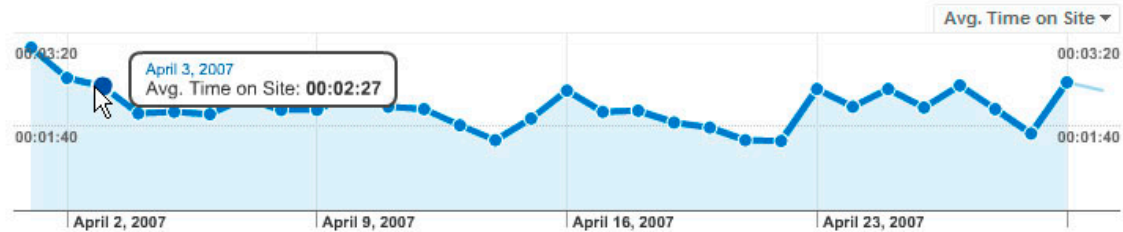
Help Resources

- 🔍 About this Report
- 🔍 Conversion University
- 🔍 Common Questions
- 🔍 Report Finder
- 🔍 Beta Feedback

Dashboard

Export ▾ ✉ Email

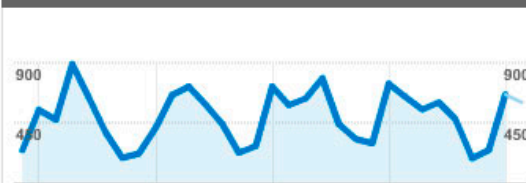
Apr 1, 2007 - Apr 30, 2007 ▾



Site Usage



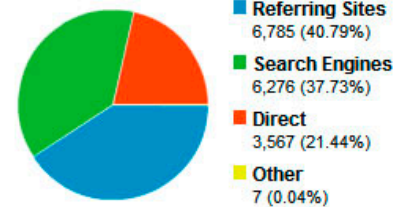
Visitors Overview



11,916 Visitors

[view report](#)

Traffic Sources Overview



[view report](#)

S akými údajmi pracujeme?

- Zdroj návštevy
 - Médium
- Charakteristika návštevníka
 - Prehliadač
 - Operačný systém
 - Hardware
 - Lokalita
- Priechod webom
 - odchody

Webová analytika

- Služi k pochopeniu chovania návštevníkov na stránkach
- Je základom stratégie pre internetový marketing

Seznam Analytika



Zber dát

- Klientskú časť zabezpečuje javascript
 - Velký vs malý zberač

Zber dát

- Klientuskú časť zabezpečuje javascript
 - Velký vs malý zberač
- Serverová časť
 - Logovanie dát v json formátu
 - Rotovanie logov po 5 minutách
 - bzip2 kompresia
 - Čas v UTC

Zber dát

- Klientuskú časť zabezpečuje javascript
 - Velký vs malý zberač
- Serverová časť
 - Logovanie dát v json formátu
 - Rotovanie logov po 5 minútách
 - bzip2 kompresia
 - Čas v UTC
- O zber sa stará 16 strojov
 - V špičke až 16000 req/sec
 - Denne 500 GB dát

Spracovávanie dát

- Regulárne výrazy
- Kategórie
- Rozsahy
- Stromová štruktúra

Spracovávajúce dát

- Regulárne výrazy
- Kategórie
- Rozsahy
- Stromová štruktúra

Konfigurace

Název	Regulární výraz
Opera/\$1	<code>^[^\\(]*opera/\\d+\\.\\d+ \\(.*\\) .* version/(\\d+)\\.\\d+.*\$</code>
Opera/\$1	<code>^.*opera[/](\\d+)\\.\\d+.*\$</code>
Chrome/\$1	<code>^[^\\(]*mozilla/\\d+\\.\\d+ \\(.*\\).*[\\d]chrome/(\\d+)\\.\\d+.*\$</code>
Safari	<code>^[^\\(]*mozilla/\\d+\\.\\d+ \\(.*\\).* version/\\d+\\.\\d+.* safari/\\d{1,4}\\d+.*\$</code>

Spracovávanie dát

- Regulárne výrazy
- Kategórie
- Rozsahy
- Stromová štruktúra

Konfigurace

Název	Min. hodnota	Max. hodnota
0-999	0	999
1000-1999	1000	1999
2000-2999	2000	2999
3000-3999	3000	3999

Spracovávanie dát – report

- Definícia dimenzií
- Výber z metrik
- Voľba granularit

Spracovávanie dát – report

- Definícia dimenzií
- Výber z metrik
- Voľba granularit
- Na základe konfigurácie sa vygeneruje HiveQL dotaz!

Príklad

```
SELECT
    COUNT(DISTINCT gsid), COUNT(1)
FROM hit
WHERE
    service = "novinky"
    AND year = 2013
    AND month = 2
    AND action = "impress"
```

Typické workflow

- Predpočítanie dát – návštevy (MapReduce)
- Samotný výpočet (Hive)
- Tvorba dimenzií z výsledku dotazu
- Agregácia (C++ alebo MapReduce)
- Zmazanie medzivýpočtov

Agregácia

- Data sú agregované až na 4 úrovniach
- Optimalizácia pre zobrazovanie v grafoch
- Počítajú sa kombinácie medzi dimenziami

Apache Hadoop



Apache Hadoop

- Open source software pre spoľahlivé, škálovateľné a distribuované výpočty
- Umožňuje spracovávať veľké objem dát pomocou clustru serverov
- Obsahuje:
 - Hadoop Common
 - Hadoop Distributed File System (HDFS™)
 - Hadoop YARN
 - Hadoop MapReduce

Apache Hadoop

- Škálovateľnosť
 - Škálovanie pridávaním ďalších strojov
 - Lokálne spracovávanie – nezahľcuje sieť dátami

Apache Hadoop

- Škálovateľnosť
 - Škálovanie pridávaním ďalších strojov
 - Lokálne spracovávanie – nezahľcuje sieť dátami
- Flexibilita
 - Akékoľvek typy dát (blobs, dokumenty, záznamy,)
 - V ľubovolnej forme (štrukturované, neštrukturované)

Apache Hadoop

- Škálovateľnosť
 - Škálovanie pridávaním ďalších strojov
 - Lokálne spracovávanie – nezahluje sieť dátami
- Flexibilita
 - Akékoľvek typy dát (blobs, dokumenty, záznamy,)
 - V ľubovolnej forme (štrukturované, neštrukturované)
- Účinnosť
 - Zjednotený storage, metadata, bezpečnosť
 - Spoľahlivý a výkonný distribuovaný súborový systém
 - MapReduce – programovací framework

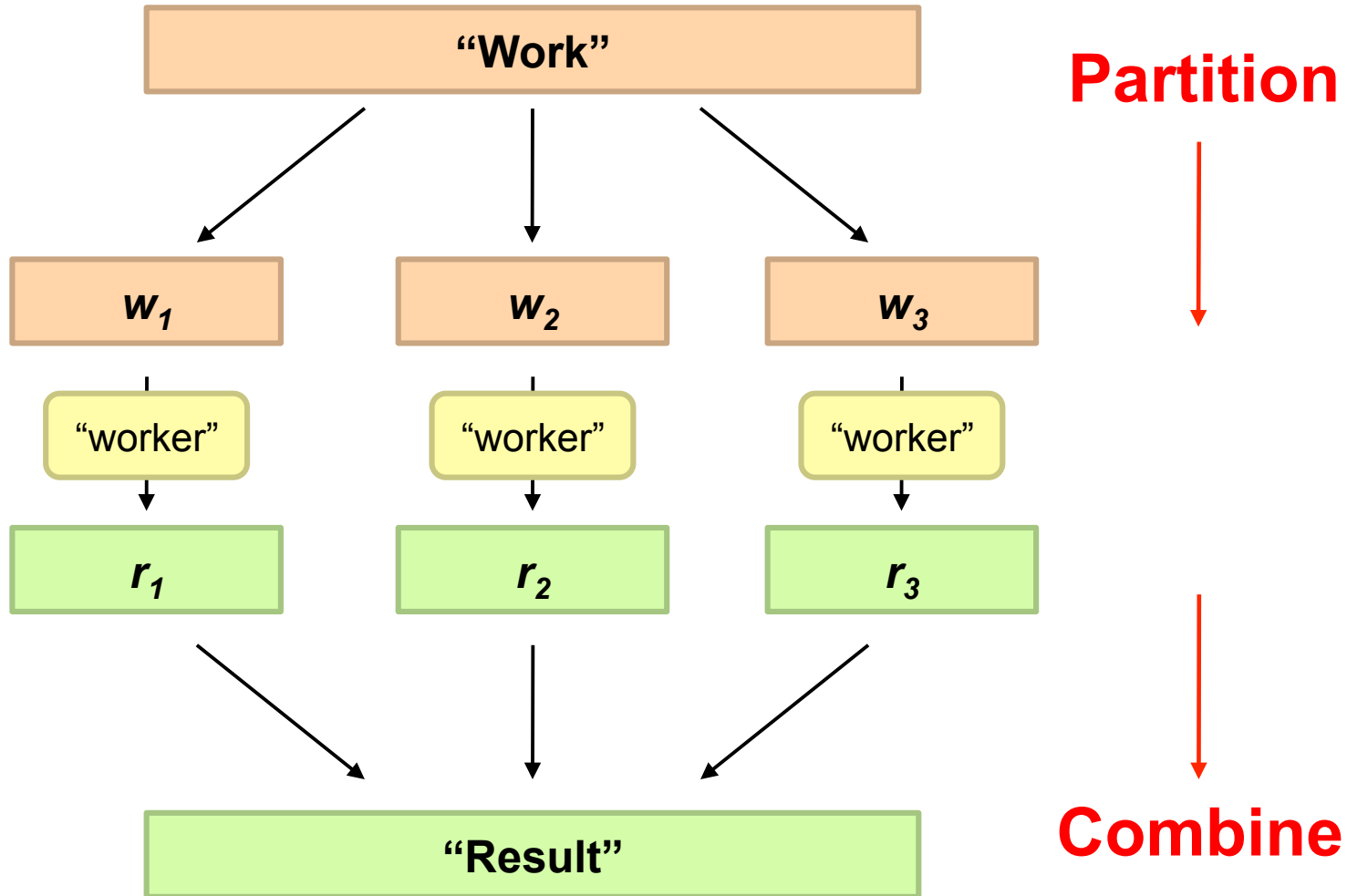
Spracovávanie dát

- Sériové spracovávanie je bottleneck

Spracovávanie dát

- Sériové spracovávanie je bottleneck
- Rýchla kalkulácia:
 - Typická rýchlosť čítania z disku 75 MB/sec
 - Na 100 GB súbor je potreba 100GB RAM a 22 minút

Rozdel' a panuj



Paralelné spracovávanie je komplikované

- Ako pridelíme prácu workerom?
- Čo ak je tej práce viac ako slotov?
- Ako sa zachovať ak práca skončí chybou?
- Ako riadiť distribuovanú synchronizáciu?

HDFS

- Distribuovaný súbakový systém
- Redundantný storage
- Navrhnutý tak, aby spoľahlivo ukladal data na bežnom stroji
- Očakáva hardwarové chyby
- Primárne určený pre veľké súbory

HDFS – súbory a bloky

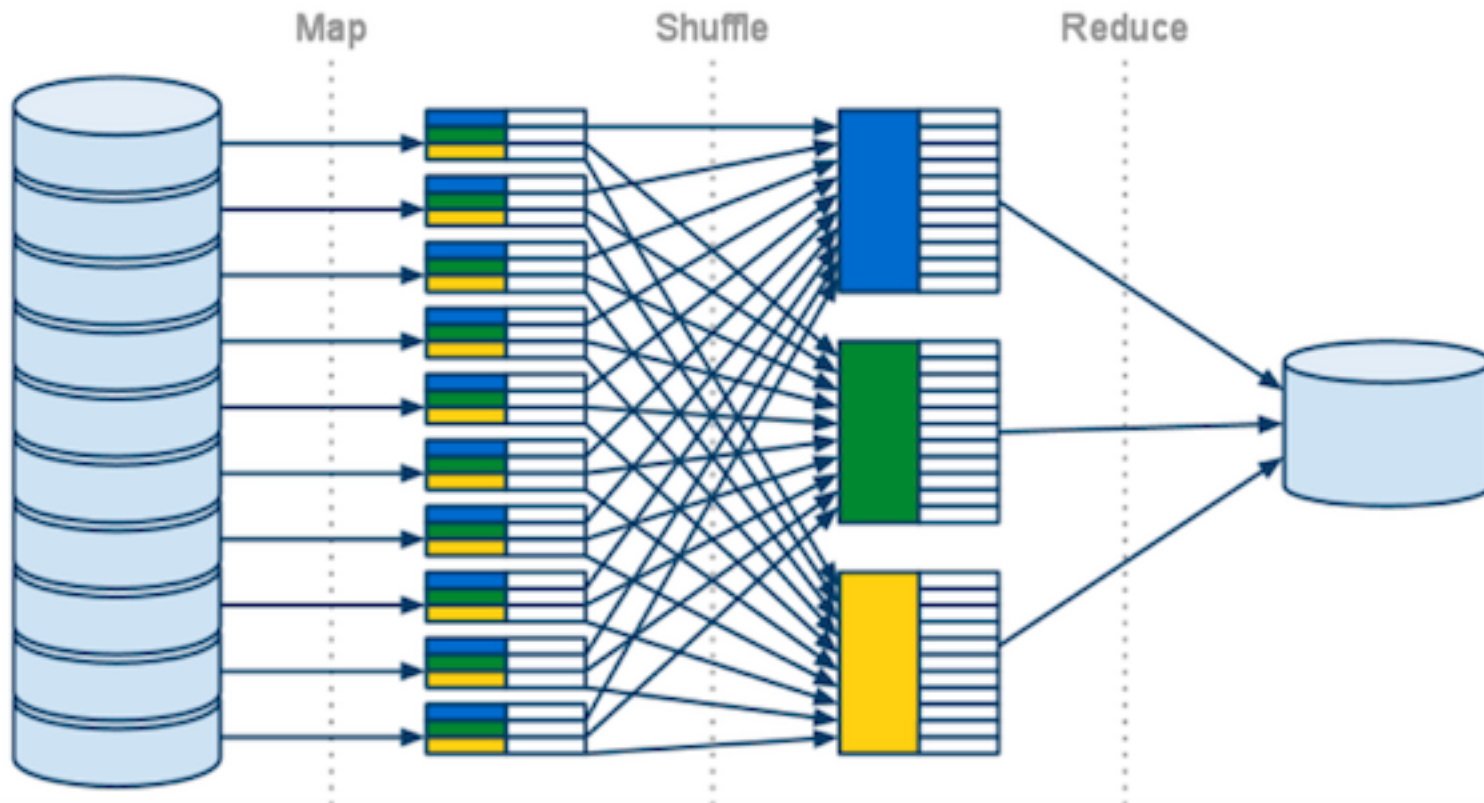
- Súbory su uložené ako sada blokov
- Bloky
 - 64 MB kusy súborov (konfigurovateľné)
 - Replikujú sa na 3 stroje (konfigurovateľné)

HDFS – súbory a bloky

- Súbory su uložené ako sada blokov
- Bloky
 - 64 MB kusy súborov (konfigurovateľné)
 - Replikujú sa na 3 stroje (konfigurovateľné)
- NameNode spravuje metadata o súboroch a blokoch
- SecondaryNameNode udržiava zálohu NameNode dát
- DataNode ukladá a vydáva bloky

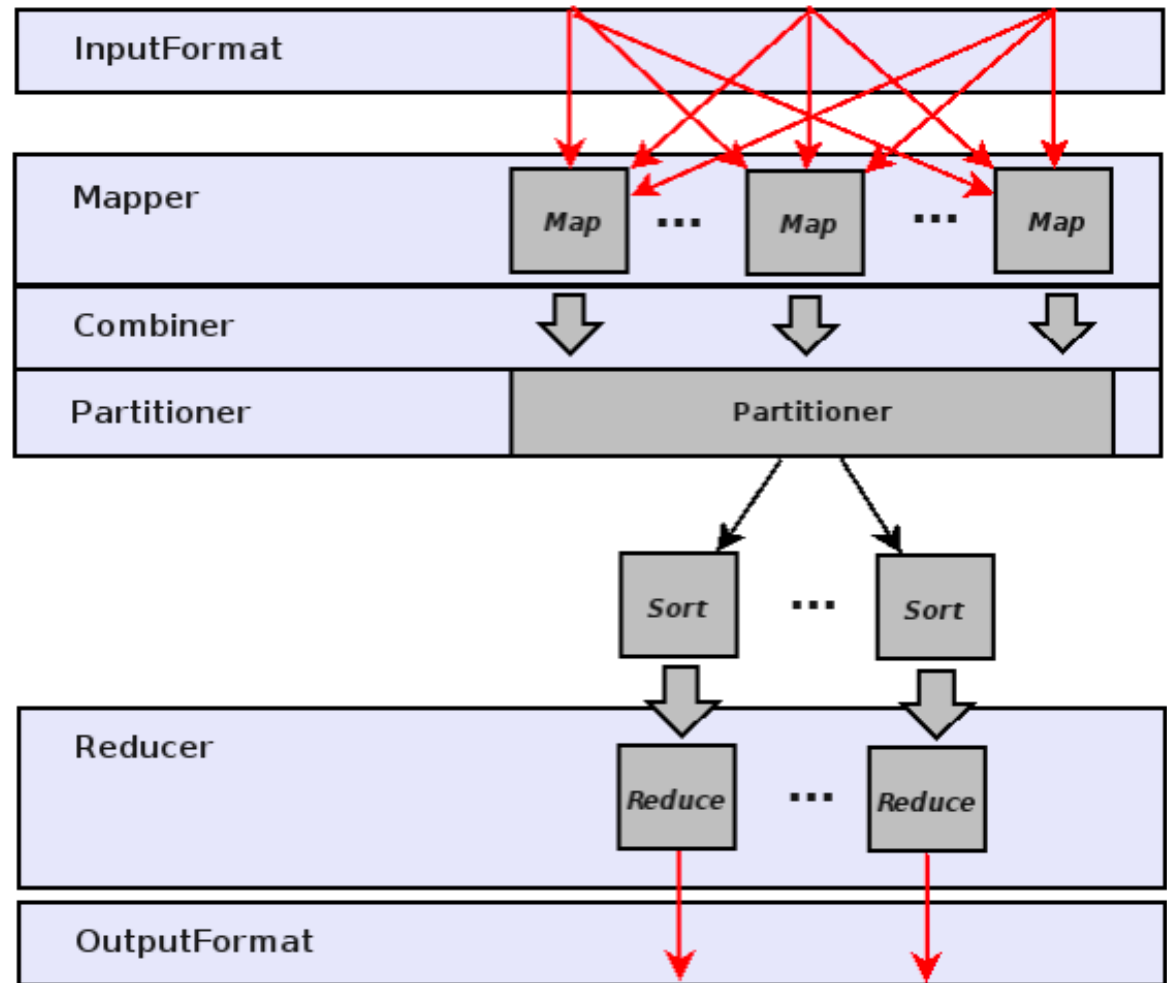
MapReduce

- Programovací model pre definovanie distribuovaných výpočtov
- Framework pre organizovanie a vykonávanie výpočtov



MapReduce – Classes

- Časti frameworku sú implementované ako abstraktné triedy alebo rozhrania



MapReduce – non java použitie

- Hadoop streaming
 - Libovolný jazyk
 - std:in
 - std:out
- Hadoop Pipes
 - C++ interface

MapReduce – word count příklad

```
function map(String name, String document):
```

```
    for each word w in document:
```

```
        emit(w, 1)
```

```
function reduce(String word, Iterator partialCounts):
```

```
    totalCount = 0
```

```
    for each count in partialCounts:
```

```
        totalCount += count
```

```
    emit(word, totalCount)
```

MapReduce – doplnkové info

- Výstup z map taskov sa zapisuje na disk
- Priebežné data sa môžu kopírovať skôr ako dobehne map task
- Samotny reduce začína až keď dobehnú všetky map tasky
- Každý reducer dostane kľúče zoradené

MapReduce – joby a tasky

- Job – uživateľom odoslaná implementácia map a reduce funkci nad určitými datami
- Task – jeden mapper alebo reducer
 - Neúspešné tasky sa automaticky pustia znova
 - Tasky v ideálnom prípade majú data lokálne k dispozícii
- JobTracker – spravuje prijaté joby a deleguje tasky medzi stroje
- TaskTracker – pýta si od JobTracker prácu a vykonáva tasky

Čo keď je task neúspešný

- Nie z dôvodu chyby v implementácii.
- Automatické spúšťanie až do N pokusov
- Po N pokusoch je job neúspešný

Čo keď' je task neúspešný

- Nie z dôvodu chyby v implementácii.
- Automatické spúšťanie až do N pokusov
- Po N pokusoch je job neúspešný
- Speculative execution
 - Spustenie rovnakej práce viackrát
 - Nie vždy žiadané
 - 1. dokončený výhraha, ostatné dostanú kill

MapReduce – záver

- Je vhodný pre
 - Paralelné algoritmy
 - Grupovanie, filtrovanie, joinovanie..
 - Offline dávkove joby nad množstvom dát
 - Analýza dát

MapReduce – záver

- Je vhodný pre
 - Paralelné algoritmy
 - Grupovanie, filtrovanie, joinovanie..
 - Offline dávkove joby nad množstvom dát
 - Analýza dát
- Nie je vhodný pre
 - Joby potrebujúce zdieľať stav
 - Low-latency joby
 - Joby nad malým množstvom dát
 - Hľadanie konkrétnych záznamov

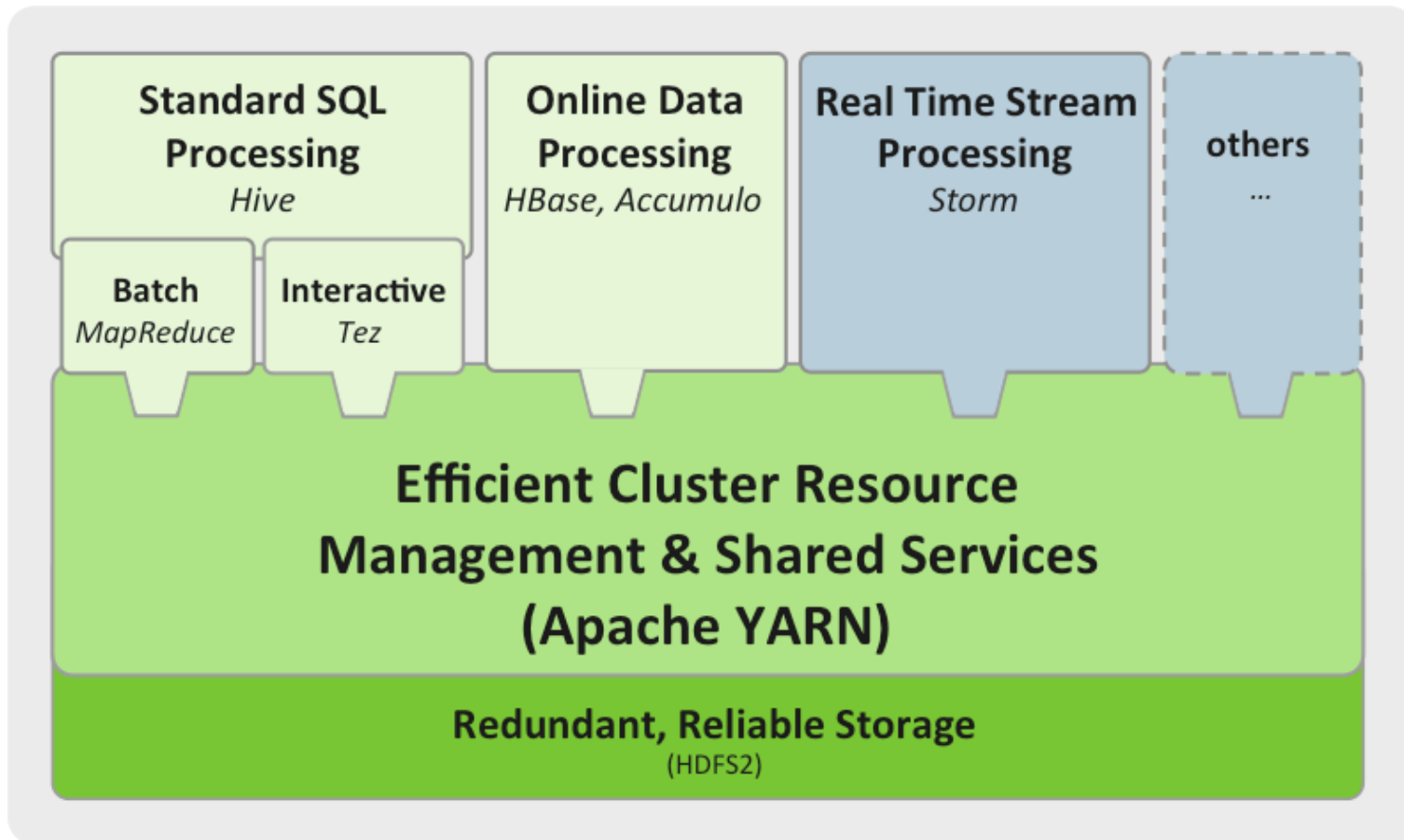
Hadoop Ecosystem

- Hive
 - DWH systém
 - SQL-like jazyk zvaný HiveQL
- HBase
 - Stípcovo orientovaná databáza
- ZooKeeper
 - Centralizovaná služba na udržiavanie informácií

Hadoop Ecosystem

- Hive
 - DWH systém
 - SQL-like jazyk zvaný HiveQL
- HBase
 - Stípcovo orientovaná databáza
- ZooKeeper
 - Centralizovaná služba na udržiavanie informácií
- Je toho viac:
 - Cascading, Scribe, Cassandra, Hypertable, Voldemort
 - Storm, Pig, Howl, Oozie, Chukwa, Mahout
 - Sqoop, Flume, Avro
 - ...

Hadoop Ecosystem



Hive

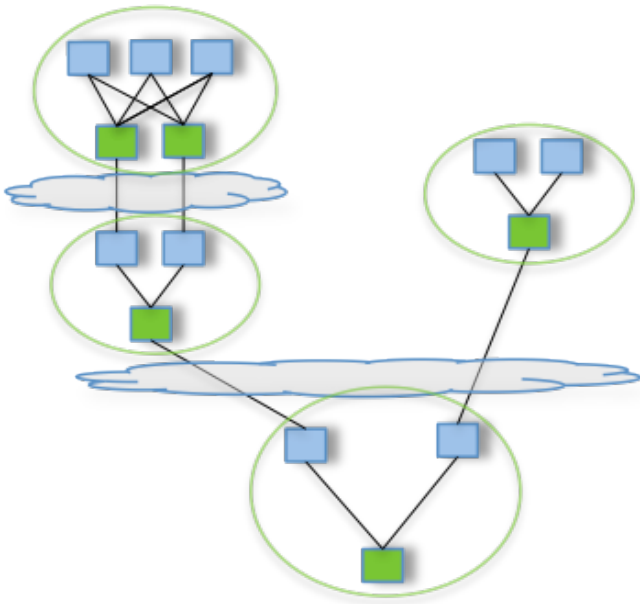


Hive

- Infrastruktúra pre dátový sklad
- Systém pre správu a dotazovanie nad štrukturovanými dátami
- Ukladá data na HDFS
- Používa MapReduce na výpočty
- Poskytuje SQL-like dotazovací jazyk zvaný HiveQL

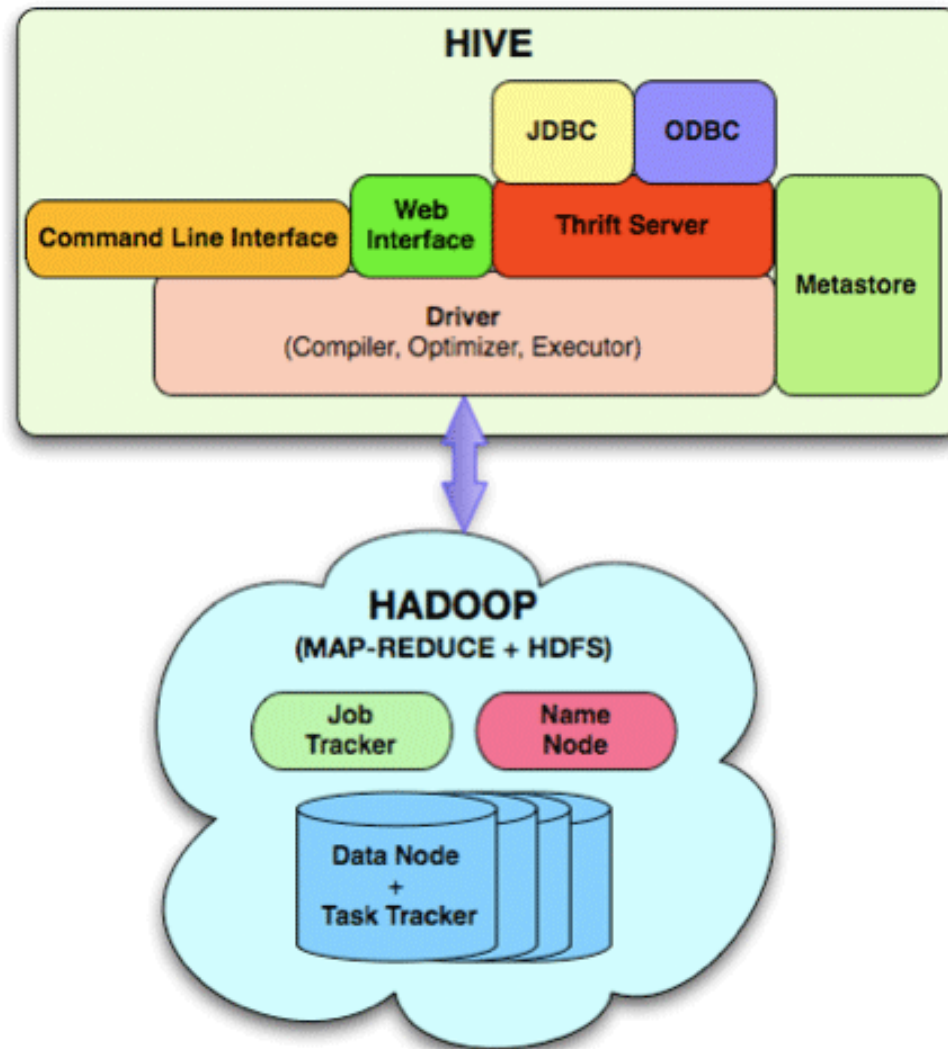
Hive – dotazy

- Hive engine dotazy konvertuje do MapReduce kódu
- Kompilátor vytvára zo zložitejších dotazov acyklický graf MapReducu jobov
- Tie joby sú potom postupne posielané na JobTracker



Pig/Hive - MR

Hive – architektúra



HiveQL

- Nespĺňa SQL-92 ťtandard

HiveQL

- Nespĺňa SQL-92 ťstandard
- DDL
 - DESCRIBE, CREATE, ALTER, DROP
 - DATABASE, TABLE, PARTITION, VIEW, FUNCTION, INDEX, COLUMN

HiveQL

- Nespĺňa SQL-92 ťstandard
- DDL
 - DESCRIBE, CREATE, ALTER, DROP
 - DATABASE, TABLE, PARTITION, VIEW, FUNCTION, INDEX, COLUMN
- DML
 - LOAD DATA, INSERT OVERWRITE, INSERT INTO

HiveQL

- Nespĺňa SQL-92 štandard
- DDL
 - DESCRIBE, CREATE, ALTER, DROP
 - DATABASE, TABLE, PARTITION, VIEW, FUNCTION, INDEX, COLUMN
- DML
 - LOAD DATA, INSERT OVERWRITE, INSERT INTO
- QL
 - SELECT, FROM, JOIN, WHERE, GROUP BY, HAVING, ORDER BY, LIMIT
 - Poddotazy sŭ povolené len v FROM klauzule
 - Equi-joins – outer joins a left semi joins
 - Transformačné scripty
 - Vlastné operátory a funkcie

Hive – dátové typy

- Primitívne typy
 - TINYINT, SMALLINT, INT, BIGINT, FLOAT, DOUBLE
 - BOOLEAN, STRING, TIMESTAMP
- Komplexné typy
 - ARRAY<data_type>
 - MAP<primitive_type, data_type>
 - STRUCT <col_name : data_type, ...>

Hive – schémata

```
CREATE TABLE user_info (  
    user_id BIGINT,  
    name STRING,  
    year INT  
);
```

```
LOAD DATA LOCAL INPATH 'sample/u.data'  
OVERWRITE INTO TABLE user_info;
```


Hive – schémata #2

```
CREATE TABLE user_info (  
    user_id BIGINT,  
    name STRING  
)  
PARTITIONED BY(year INT)  
CLUSTERED BY(user_id) INTO 256 BUCKETS;
```

Hive – schémata #3

```
CREATE EXTERNAL TABLE visitors_day (  
    gsid STRING,  
    time INT,  
    kind INT  
) PARTITIONED BY (year INT, month INT, day INT)  
LOCATION '/visitors/day/';
```

Hive – schémata #3

```
CREATE EXTERNAL TABLE visitors_day (  
    gsid STRING,  
    time INT,  
    kind INT  
) PARTITIONED BY (year INT, month INT, day INT)  
LOCATION '/visitors/day/';
```

- Expandovaná partícia:
 - /visitors/day/year=2013/month=3/day=11/

Hive – schémata #4

```
CREATE EXTERNAL TABLE hit (...)  
ROW FORMAT SERDE 'cz.seznam.analytics.serializer.Json'  
LOCATION '/www/wa/';
```

```
CREATE EXTERNAL TABLE visitors_hour (...)  
ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t'  
LOCATION '/visitors/hour/';
```

HBase



HBase

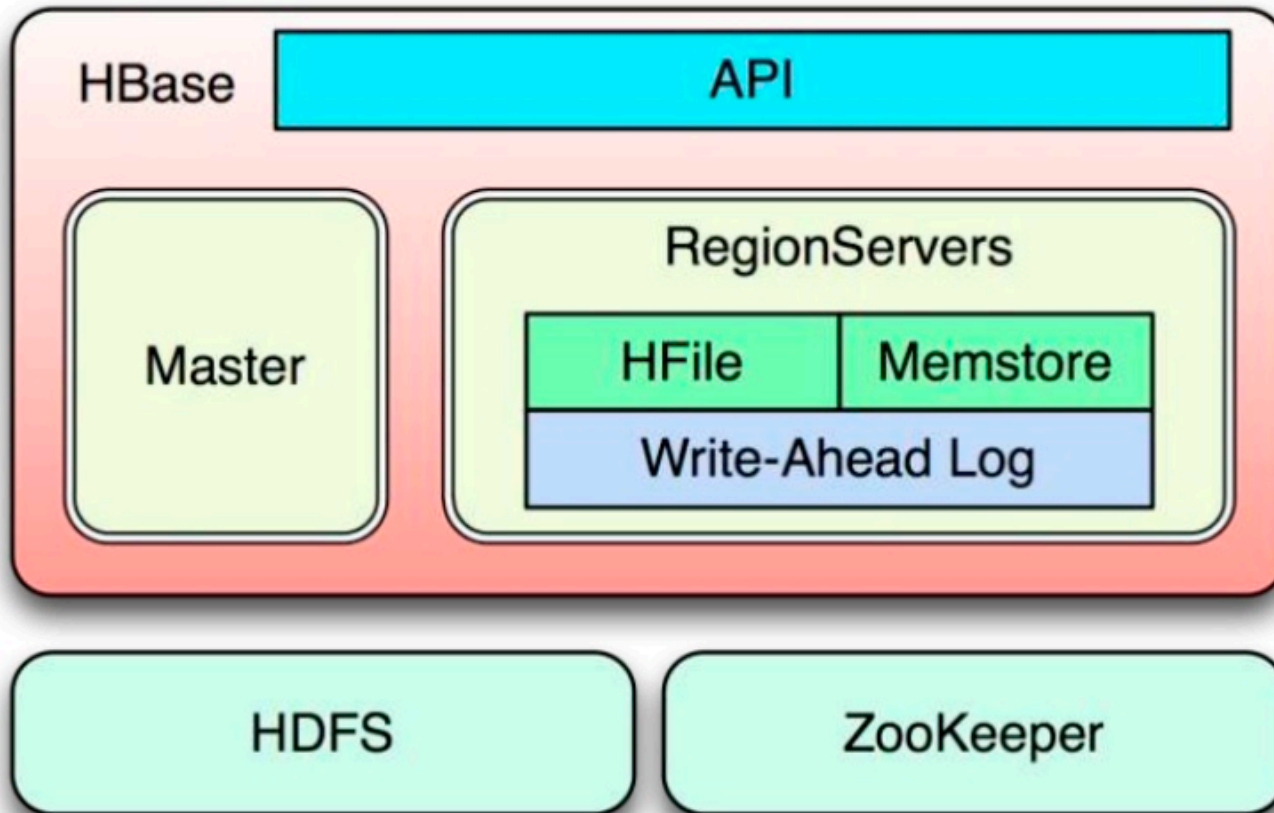
- Open-source nerelačná distribuovaná stĺpcovo orientovaná databáza
- Operuje nad Hadoop clustrom
- Ukladá data na HDFS
- Odolná voči výpadkom
- Horizontálne škálovateľná
- ACID garancia na úrovni riadkov
- Automatické shardovanie tabuliek

Velká zoradená mapa

	Row key	Column key	Timestamp	Cell	
Sorted by Row key and Column	Row1	info:aaa	1273516197868	valueA	2 Versions of this cell
	Row1	info:bbb	1273871824184	valueB	
	Row1	info:bbb	1273871823022	oldValueB	
	Row1	info:ccc	1273746289103	valueC	
	Row2	info:hello	1273878447049	i_am_a_value	
	Row3	info:	1273616297446	another_value	

Timestamp is a long value

HBase – architektúra



HBase – Master

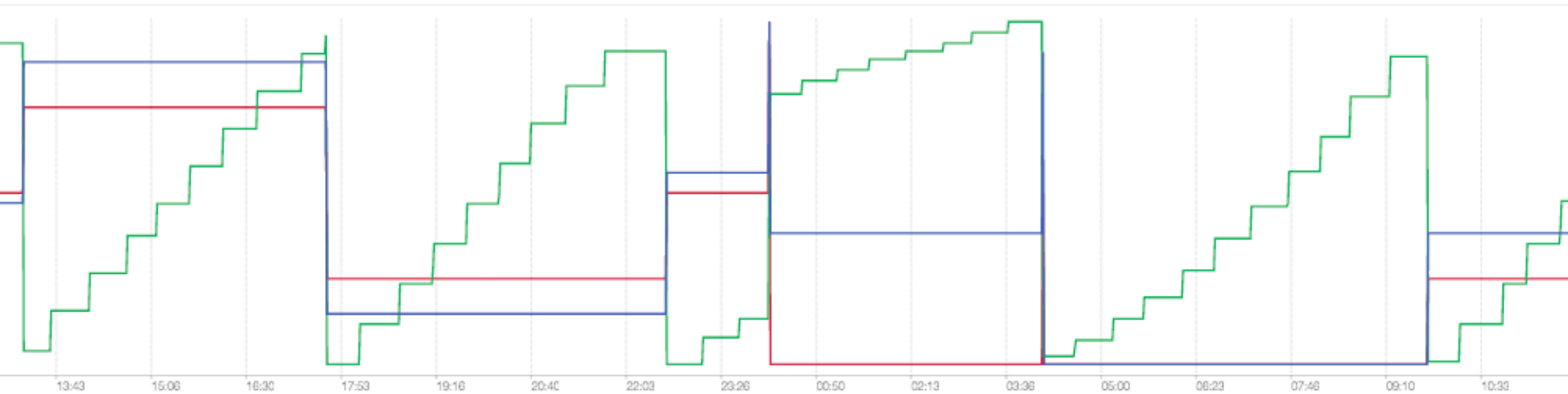
- Priraduje regiony RegionServerom
- Drží adresu RS obsluhujícího tabulku ROOT
- Stráží RS a případě výpadku:
 - provede split logu daného RS
 - přiřadí off-line regiony živým RS

HBase – RegionServer

- spracováva R/W požiadavky klienta
- zápisy sa logujú do Write Ahead Logu (WAL)

HBase – Compaction

- Zlúčenie množstvo HFiles do jedného
- Major compaction



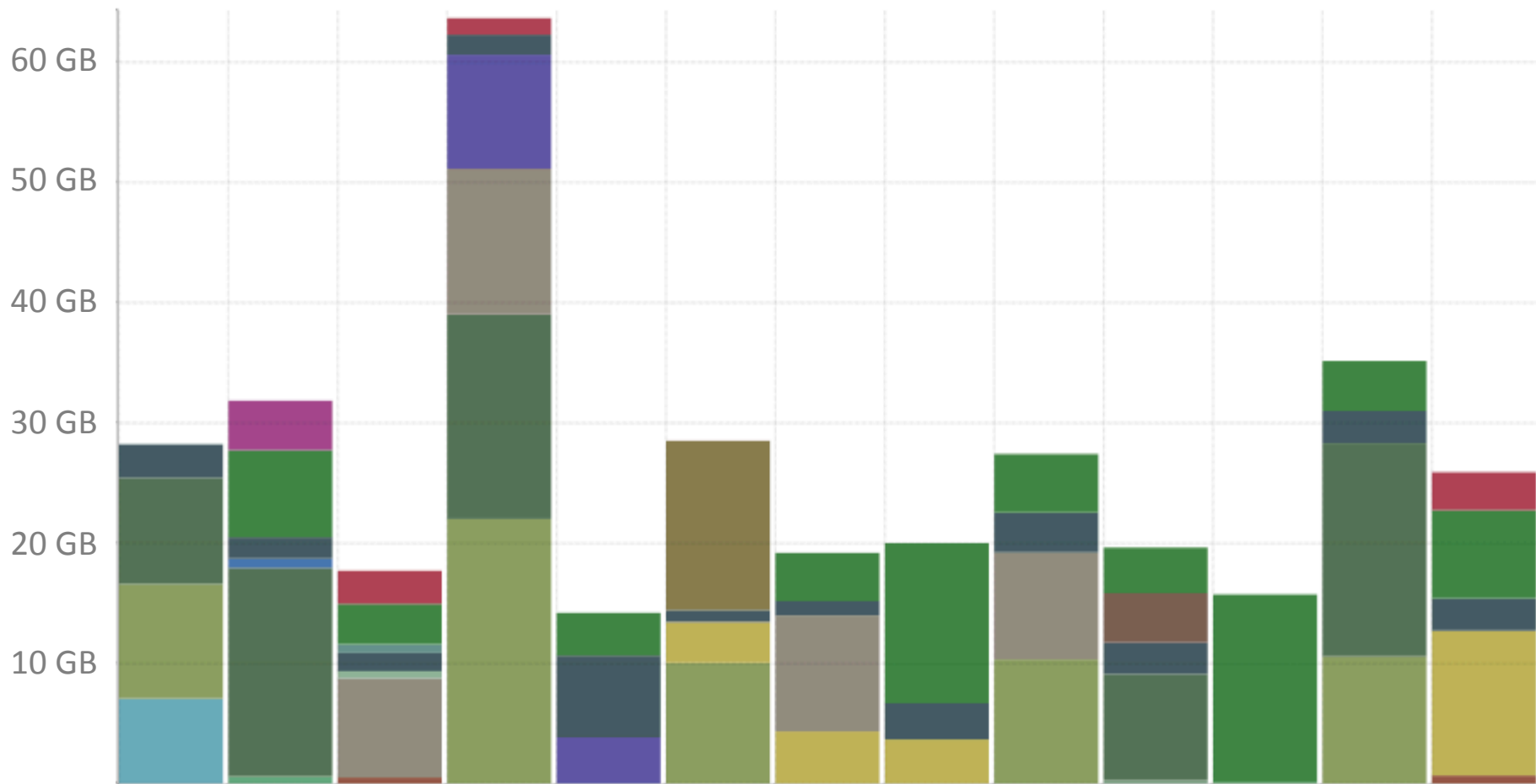
HBase API

- `get(row)`
- `put(row, map<columns, values>)`
- `scan(key range, filter)`
- `increment(row, columns)`
- `delete(...)`

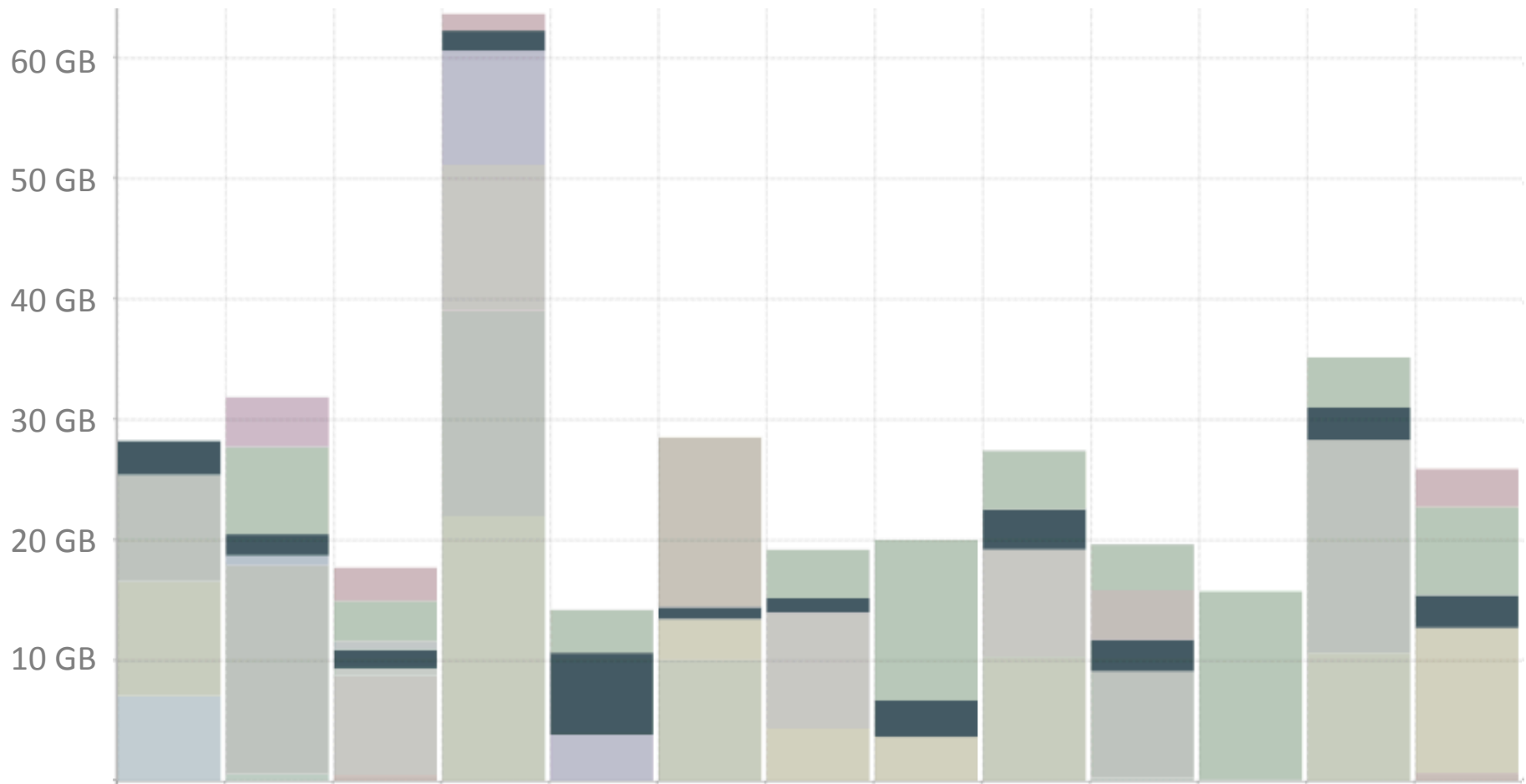
HBase – MapReduce

- Zdroj aj úložisko MR jobov
- Mapy majú data lokálne

Štatistiky Sklik.cz



Štatistiky Sklik.cz



HBase – Bloom filters

- Space-efficient probabilistic data structure
- Testuje prítomnosť prvku v množine
- Môže byť false positive, ale nie false negative

HBase – Bloom filters

- Space-efficient probabilistic data structure
- Testuje prítomnosť prvku v množine
- Môže byť false positive, ale nie false negative

- Index – zrýchľuje čítanie
- Access patterns – pomáha v prístupe k dátam ak preskakujeme veľa záznamov pri scanoch

HBase – Coprocessor & Filters

- Region-side funkcionálna pri práci nad dátami
- Custom filtrovanie záznamov
- Coprocessor - Observer
 - Hooky pri práci s dátami a regiónmi
 - pre/post metódy pre rôzne operácie
- Coprocessor – Endpoint
 - Nahradzujú stored procedúry v RDBMS



Děkuji za pozornost

Lukáš Antalov, Vedoucí týmu vývoje