Hadoop a spracovanie veľkého objemu dát – realita verzus potenciál

Trendy a stratégie

Hadoop a spracovanie veľkého objemu dát – realita verzus potenciál

Trendy a stratégie

Mgr. Martin Šeleng, PhD.,
Ústav Informatiky Slovenská Akadémia Vied

eFOCUS

Obsah

- Google Map/Reduce architektúra
- Hadoop
- Facebook
- Hive
- Yahoo!
- Pig
- Ďalšie zdroje
- Sumarizácia

Google infraštruktúra (2003)

- Dvojjadrové x86 procesory, beží na nich operačný systém Linux a majú 2-4GB pamäte
- Sieťové pripojenie s rýchlosťou 100Mb/s
- Kluster pozostáva zo stoviek až tisícok pracovných staníc
- Disky s IDE rozhraním

- **Bežné desktopové počítače!**

Google Map/Reduce architektúra

- Základný model – 2 funkcie
  - Map: Vstupom je súbor dát (textových), výstupom asociatívne pole (prvky nie sú indexované podľa postupnosti celých čísel, ale pomocou kľúčov) typu: kľúč/hodnota
  - Reduce: Vstupom je výstup z funkcie Map a výstupom je asociatívne pole typu: kľúč/hodnota, pričom platí, že polia s rovnakými kľúčmi sú spracovávané na jednom uzle
- Rozšírený model (navyše oproti základnému)
  - Combine: Po dobehnutí funkcie Map spustí lokálne funkciu Reduce
  - Partition: V prípade, že chceme robiť s inými ako textovými dátami, potrebujeme implementovať metódu, ktorá bude zodpovedná za rozkúskovanie vstupných dát na menšie bloky
- **Programátor nemusí poznať spôsob akým rámec distribuuje výpočty!**
- **Aplikácie je možné vyvíjať a ladiť na desktope**

Map/Reduce v Googli

- **80% úloh je spúšťaných na Map/Reduce architektúre**
  - Internetové vyhľadávanie a indexovanie
  - PageRank
  - Klastrovanie Google News
  - Spracovávanie satelitných snímkov pre Google Maps
  - Spracovanie jazykových modelov (štatistické prekladanie) Google Translate
  - Učiace sa algoritmy
  - A iné...
- **20% Pregel, a iné ...**

Ako vieme kedy sú dáta dosť veľké?

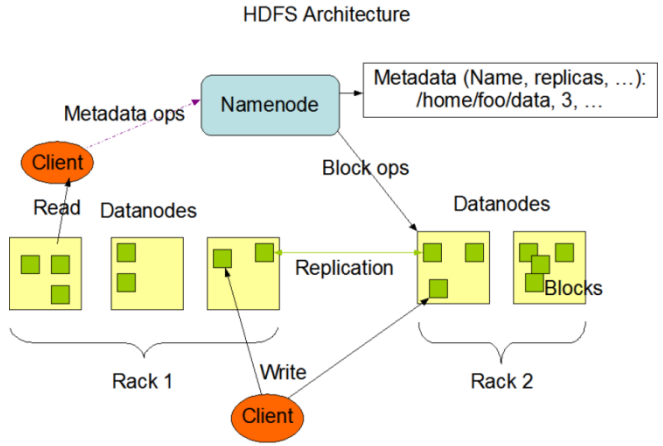
- Veľké dáta – neznamená, že sú naozaj veľké, môže ísť aj o veľa relatívne malých záznamov
  - Webové stránky majú v priemere 320KB (https://developers.google.com/speed/articles/web-metrics)
  - Transakčné dáta sú naozaj veľké len niekoľko KB, ale je ich veľa
- Satelitné snímky sú naozaj veľké dáta
- Štruktúrované
  - Logy webového servera
  - Transakčné dáta
  - Záznamy o telefónnych hovoroch
  - ...
- Neštrúkturované
  - Webové stránky
  - Literárne diela
  - Emaily
  - ...

Hadoop – história

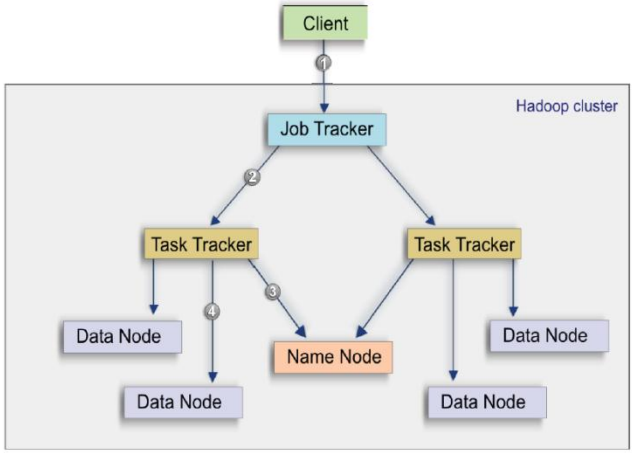
- Hadoop projekt začal vyvíjať v roku 2004 Doug Cutting (po publikovaní článku o Map/Reduce od spoločnosti Google)
- Hadoop bol prvý krát použitý ako distribuované prostredie pre internetový sťahovač Nutch (jún 2003), ktorý vyvinuli Doug Cutting a Mike Cafarella
- Indexer použitý v rámci projektu Nutch, vyvinul Doug Cutting a nazval ho Lucene (1999)
- Všetky projekty sú dnes top projektmi a sú spravované komunitou Apache Software Foundation

Hadoop – ekosystém

- **Hadoop základné komponenty**
  - Map/Reduce (spracovanie úloh a dát)
  - HDFS/Hadoop Distributed File System (distribuované dátové úložisko)
- **HDFS**
  - NameNode
  - SecondaryNameNode
  - DataNodes
- **Map/Reduce**
  - JobTracker
  - TaskTrackers
  - JobHistoryServer



Distribuované dátové úložisko Hadoop



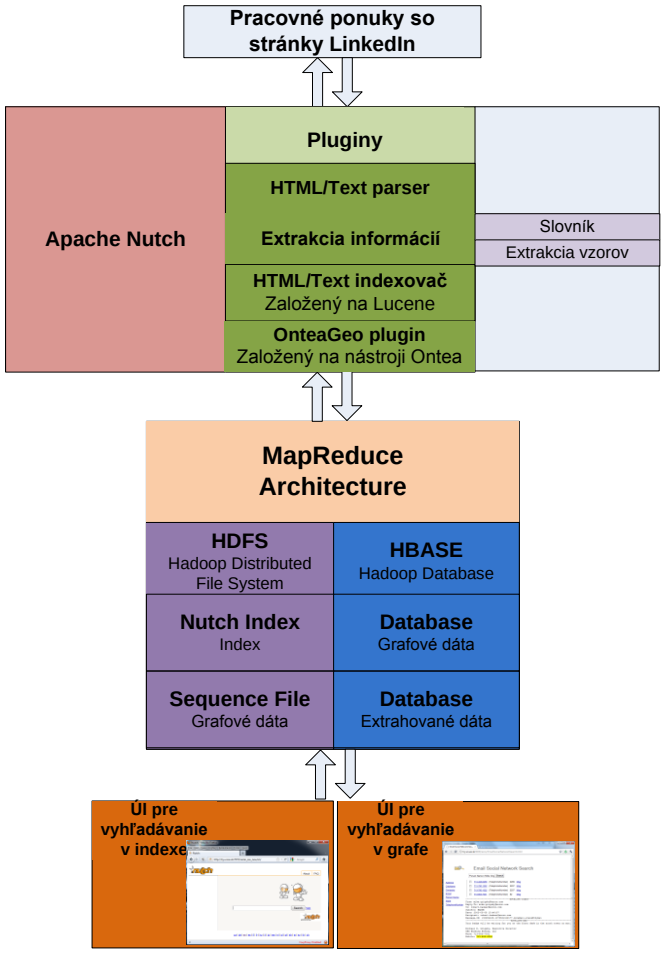
Spôsob spustenia úlohy v rámci Hadoop

Hadoop – ekosystém (pokračovanie)

- Pig – analýza veľkých dátových setov (Yahoo!)
- Hive – dátový sklad (FaceBook)
- HCatalog – meta úroveň pre prácu nad rámcami Hadoop/Hive/Pig
- Hbase – distribuovaná databáza
- Mahout - škálovateľné knižnica pre data mining
- Zookeeper – organizátor konfiguračných súborov
- Oozie – manažment úloh spustených nad Hadoop-om
- Sqoop – nástroj na prenášanie veľkých objemov dát medzi Hadoop-om a relačnými úložiskami
- Flume – zberač logov

Hadoop – ukážka

- **Softvér**
  - Apache Hadoop (distribúcia od spoločnosti Cloudera) 0.20.2+737 (Map/Reduce)
  - Apache Nutch 1.3 (internetový sťahovač)
  - Apache Solr 3.1.0 (indexovač /vyhľadávač)
  - Apache HBase (distribúcia od spoločnosti Cloudera) 0.89.20100924+28
- **Hardvér**
  - Testovacie prostredie – 8 uzlov + master
    - Procesor - Intel® Core™ 2 Quad CPU Q9550 2.83GHz
    - Systémová pamäť - 4 GB
    - HDD - WDC WD7500AACS-0 (750 GB)
    - OS - Linux 2.6.24-19-generic x86_64 GNU/Linux
  - Štatistiky celého clustra
    - Celková disková kapacita je cez 5TB
    - Základný replikačný faktor 3
    - 32 map úloh bežiacich paralelne
    - 16 reduce úloh bežiacich paralelne
  - Apache Solr na ďalšom PC
- **Aplikácia – vyhľadávanie nad pracovnými ponukami zo stránky LinkedIn**



Architektúra systému na vyhľadávanie pracovných ponúk

Hadoop – ukážka

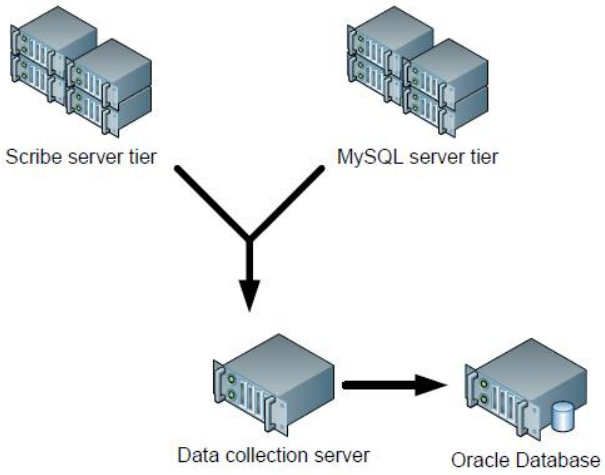
- **Fakty o stiahnutých pracovných ponukách**
  - 92 223 stránok z toho je 58 112 pracovných ponúk
  - Veľkosť indexu 591MB
    - 1 257 019 indexovaných termov
    - Viac informácií o indexe je možné nájsť tu:
      - http://try.ui.sav.sk:7070/2011-11-02/admin/schema.jsp
      - http://try.ui.sav.sk:7070/2011-11-02/admin/luke?wt=xslt&tr=luke.xsl
  - Veľkosť stiahnutých dát 1.8GB
    - Vrátane databázy URL liniek
    - Vrátane grafu prepojení medzi stránkami
    - Vrátane obsahu (1.2GB)
    - Parsovaných dát a metadát (309MB)
- **Používateľské rozhrania**
  - http://147.213.75.181:8080/2012-01-07/
  - http://147.213.75.181:8080/apache-solr-3.1.0/
  - http://try.ui.sav.sk:7070/2011-11-02/browse
  - http://try.ui.sav.sk:7070/apache-solr-3.1.0/browse

Kto používa/podporuje Hadoop

- **Používatelia**
  - Amazon (1 – 100 uzlov)
  - Adobe (80 uzlov)
  - EBay (532 uzlov)
  - Facebook (približne 3000 uzlov)
  - IBM (počet skladísk a uzlov nie je známy)
  - LinkedIn (približne 4000 uzlov)
  - Twitter (počet uzlov nie je známy)
  - Yahoo! (viac ako 40 000! uzlov)
- **Vývojári a podporovatelia**
  - Cloudera
  - Hortonworks
  - MapR
  - Yahoo!
  - Microsoft
  - IBM
  - Twitter
  - Facebook

Facebook

- Postup zbierania dát vyprodukovaný používateľmi v rámci sociálnej siete Facebook
  - Dáta boli zbierané pomocou úloh zadaných v plánovači a výsledky boli uložené do Oracle databázy (>24h)
  - ETL (Extract, transform a load) úlohy boli naimplementované v Pythone
- Množstvo spracovávaných dát (denne)
  - 2006 – 15GB
  - 2008 – 250GB
  - 2009 – 5TB komprimovaných dát
  - 2010 - 12TB komprimovaných dát
  - 2012 – 400TB nekomprimovaných dát

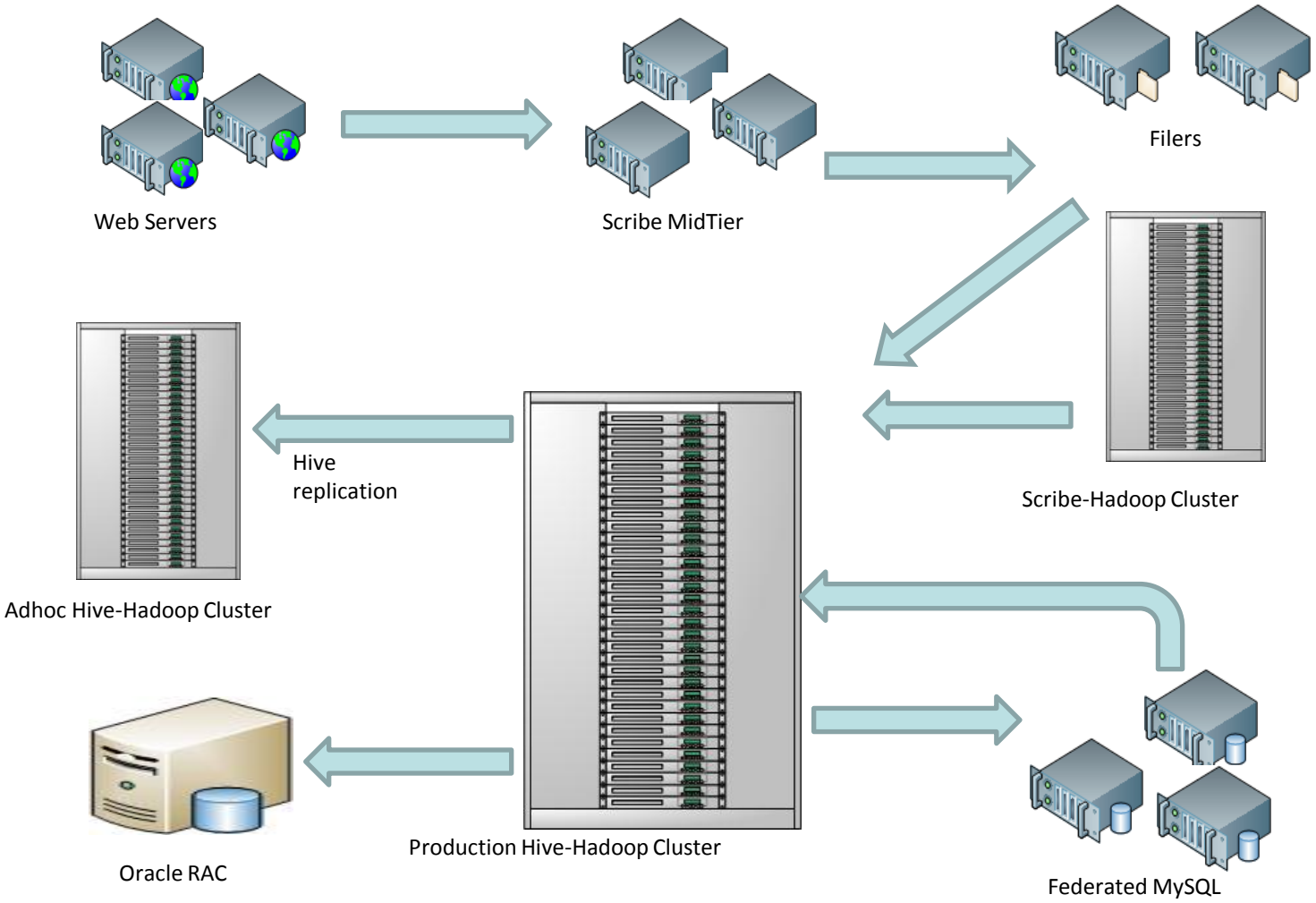


Spracovanie dát v spoločnosti Facebook

Facebook ekosystém

- Scribe – zberač logov
- Hive – dátové skladisko (dávkové pracovanie – HiPal UI)
- Cassandra – distribuovaná databáza (ukončená podpora)
- Hbase – distribuovaná databáza
- Data Freeway – škálovateľný dátový tok
- Puma – analýza v reálnom čase (spracovanie toku dát skoro v reálnom čase)
- Peregrine – jednoduché a rýchle dopyty (mimo Hadoop)

Facebook a Hive – architektúra



Spracovanie dát v spoločnosti Facebook pomocou Hadoop/Hive

Facebook a Hive

- Hadoop/Hive dátové skladisko (2010)
  - 22400 jadier (2000 uzlov), disková kapacita 21PB
  - 12TB/32GB RAM na uzol
  - Dve úrovne pripojenia topológie siete
    - 1 Gbit/s z uzla do switchu v danom racku
    - 4 Gbit/s do hlavnej úrovne zo switchu racku
- Štatistika (deň 2010):
  - 800TB prezeraných komprimovaných údajov
  - 25000 spustených Map/Reduce úloh
  - 80 000 hodín strojového času
  - ~200 ľudí (15% spoločnosti)/mesiac spúšťa úlohy nad Hadoop/Hive
  - Analytici (nie informatici) používajú Hadoop pomocou Hive
  - 95% úloh je Hive úloh
  - Nová úloha je spustená každú sekundu

Hive – výhody

- **Čo Hadoop neposkytuje:**
  - jazyk, ktorým by sa dalo jednoducho tieto úlohy písať (bez potreby písať Map/Reduce programy)
  - editor príkazového riadku, v ktorom by sa tieto úlohy mohli písať
  - schémy o jednotlivých tabuľkách v databázach
- **Čo Hive poskytuje:**
  - poskytuje vlastný editor príkazového riadku (tzv. hive>), ktorý je podobný MySQL editoru (mysql>)
  - jazyk, ktorým je možné písať dopyty (HQL – Hive query language, podobné SQL)
  - podporu pre JDBC klientov (pripojenie sa k ľubovoľnej SQL databáze)
  - uloženie metadát o databázach a tabuľkách
  - možnosť písať HQL dopyty, pričom Hive automaticky preloží tieto dopyty do Map a Reduce úloh
  - dáta sú štandardne csv súbory, ale je možné použiť ľubovoľné objekty

Hive – ukážka

- **Problém**: Najobľúbenejšie slovo Martina Kukučína.
- **Riešenie**: Prečítať všetky jeho diela a spočítať.
- **Inteligentnejšie riešenie**:
  - Máme rôzne internetové zdroje ako napr.: http://zlatyfond.sme.sk
  - Pre jednoduchosť, a preto aby nás SME nezakázalo sme stiahli 2 diela:
    - Dom v stráni
    - Rysavá jalovica

Hive – ukážka (pokračovanie)

- shell> hadoop jar /usr/lib/hadoop-mapreduce/hadoop-mapreduce-examples.jar grep dom_v_strani_diak.txt dom_v_strani_freq '\w+,
- hive> CREATE TABLE dom_v_strani (freq INT, word STRING) ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' STORED AS TEXTFILE;
- hive> LOAD DATA INPATH "dom_v_strani_freq" INTO TABLE dom_v_strani;
- hive> CREATE TABLE spojena (word STRING, dom_v_strani_f INT, rysava_jalovica_f INT);
- hive> INSERT OVERWRITE TABLE spojena SELECT d.word, d.freq, r.freq FROM dom_v_strani d JOIN rysava_jalovica r ON (d.word = r.word);
- hive> SELECT word, dom_v_strani_f, rysava_jalovica_f, (dom_v_strani_f + rysava_jalovica_f) AS s FROM spojena SORT BY s DESC LIMIT 10;

Všetky slová				Slová dlhšie ako 3 znaky			
Slovo	**Počet slov Dom v stráni**	**Počet slov Rysavá Jalovica**	**Počet slov spolu**	**Slovo**	**Počet slov Dom v stráni**	**Počet slov Rysavá Jalovica**	**Počet slov spolu**
sa	4360	465	4825	jeho	404	32	436
a	3917	473	4390	este	354	56	410
na	1948	276	2224	vsetko	297	21	318
v	1699	141	1840	pred	273	20	293
i	1373	131	1504	akoby	238	35	273
je	1285	135	1420	bude	252	13	265
ze	1144	153	1297	teraz	232	31	263
co	1138	117	1255	ktory	205	10	215
to	870	156	1026	nech	195	13	208
do	841	152	993	lebo	147	57	204

Yahoo! a Hadoop

- Yahoo! zamestnalo Douga Cuttinga v roku 2006 (odišiel v 2009) – začiatky Hadoop-u v Yahoo! (testovací klaster)
- 2007 – okolo 5000 uzlov (25PB)
- 2008 – nasadenie na vyhľadávanie v http://yahoo.com (okolo 20 000 uzlov – 60PB)
- 2009 – Yahoo! sa spojilo s Apache Foundation a stalo sa jedným s prispievateľov a podporovateľov projektu Apache Hadoop (okolo 30 000 uzlov – 110PB)
- 2010 – Yahoo! sa stalo spoluzakladateľom spinoffu Hortonworks (okolo 40 000 uzlov – 225PB)
- 2011/2012 – 42 000 uzlov (350PB)
- Yahoo! spolu s Hortonworks je najväčším prispievateľom do rámca Hadoop (>70% kódu od roku 2006)

Yahoo! a Hadoop (pokračovanie)

- Yahoo! mail
  - 450M mailových účtov
  - 40% spamu menej ako Hotmail a 55% spamu menej ako Gmail (analýza spamu beží nad Hadoop klastrom)
- Viac ako 20 rôznych klastrov
- Najväčší klaster 4 000 uzlov
- Viac ako 1000+ používateľov
- 1 000 000+ úloh spustených na klastroch/mesiac
- Projekt Pig
  - Potreba písať SQL dopyty databázovými špecialistami
  - Pig začalo vyvíjať Yahoo! (neskôr projekt prešiel pod Apache Foundation)
  - Dodnes sú hlavnými vývojármi zamestnanci Yahoo!

Pig – výhody

- **Čo Hadoop neposkytuje:**
  - jazyk, ktorým by sa dalo jednoducho tieto úlohy písať (bez potreby písať Map/Reduce programy)
  - editor príkazového riadku, v ktorom by sa tieto úlohy mohli písať
- **Výhody Pig:**
  - Poskytuje tzv. Pig Latin (kvázi SQL jazyk) na vykonávanie dopytov hlavne nad logmi z webových serverov
  - Umožňuje písanie Map/Reduce úloh (ne)odborníkom na Javu a technológiu Map/Reduce, resp. expertom na relačné DB
  - Pracuje nad csv súbormi z ľubovoľnou schémou
  - Podporuje implementáciu tzv. UDF (User-Defined Functions), čim umožňuje prácu nad ľubovoľnými objektmi a takisto podporuje ľubovoľné operácie nad nimi

Pig Latin - ukážka

- **Ukážka logov z webového servera Yahoo! (anonymizovaný používateľ, čas a dátum, vyhľadávaný reťazec)**

2A9EABFB35F5B954	970916105432	+md foods +proteins
BED75271605EBD0C	970916025458	yahoo caht
BED75271605EBD0C	970916090700	hawaii chat universe
BED75271605EBD0C	970916094445	yahoo chat
824F413FA37520BF	970916185605	exhibitionists
824F413FA37520BF	970916190220	exhibitionists
824F413FA37520BF	970916191233	exhibitionists
7A8D9CFC957C7FCA	970916064707	duron paint
7A8D9CFC957C7FCA	970916064731	duron paint
A25C8C765238184A	970916103534	brookings
A25C8C765238184A	970916104751	breton liberation front

Pig Latin - ukážka

- **Úloha: Zistiť počet dopytov používateľa na vyhľadávacom serveri.**
- grunt> log = LOAD ‘excite-small.log’ AS (user, time, query);
- grunt> grpd = GROUP log BY user;
- grunt> cntd = FOREACH grpd GENERATE group, COUNT(log) AS cnt;
- grunt> ordered = ORDER cntd BY cnt desc;
- grunt> STORE ordered INTO ‘output’;
  - 128315306CE647F6 78
  - 0B294E3062F036C3 61
  - 7D286B5592D83BBE 59
  - AAA6E4471629BC8F 47
  - EC6E91864359DD8D 47
  - 54E8C79987B6F2F3 46
  - 567854C718273984 46
  - 917FDFC55A0EA9ED 38
  - 52C93D97F5AD00AE 38
  - 3F59FEC0AD9851A5 38

Pig Latin – ukážka (pokračovanie)

- **Úloha: Zistiť používateľov, ktorí sa dopytujú viac ako 50x na vyhľadávacom serveri.**
- grunt> cntd = FOREACH grpd GENERATE group, COUNT(log) AS cnt;
- grunt> fltrd = FILTER cntd BY cnt > 50;
- grunt> STORE fltrd INTO ‘output50’;
  - 0B294E3062F036C3 61
  - 128315306CE647F6 78
  - 7D286B5592D83BBE 59

Pig Latin – ukážka (pokračovanie)

- **Úloha: Zistiť najvyhľadávanejšie slová v určitom čase.**
- 00 bluebird 4
- 00 cute 4
- 00 chested 4
- 00 diablo cheats 3
- 00 psygnosis 2
- 00 video camera 2
- ....
- **Úloha: Zistiť počet vyhľadávaní každej frázy doobeda a poobede.**
- weather 7 57
- web and 1 1
- webchat 4 11
- website 1 4
- webster 1 4
- wedding 3 19
- welcome 1 1
- western 2 14
- ....

Ďalšie zaujímavé zdroje

- **Dopytovacie jazyky**
  - Sawzall – http://static.googleusercontent.com/external_content/untrusted_dlcp/research.google.com/en//archive/sawzall-sciprog.pdf (Google)
  - JAQL – https://www.ibm.com/developerworks/mydeveloperworks/blogs/ibm-big-data/entry/why_jaql2?lang=en (IBM)
  - Scope – http://research.microsoft.com/en-us/um/people/jrzhou/pub/Scope.pdf (Microsoft)
  - YAML – http://www.greenplum.com (Greenplum)
- **Distribuované grafové rámce**
  - Pregel – http://kowshik.github.com/JPregel/pregel_paper.pdf (Google)
  - Apache Giraph – http://incubator.apache.org/giraph/
  - Apache Hama – http://hama.apache.org/
- **Práca nad relačnými dátami**
  - Stratosphere – https://www.stratosphere.eu (TU Berlin)
  - Cascading – http://www.cascading.org/ (Concurrent)

Sumarizácia

- Je zrejmé, že Hadoop je hojne využívanými veľkými spoločnosťami na spracovanie rôznych typov dát
- Jedna z ranných štúdií spoločnosti Google ukázala, že až 90% úloh spúšťaných vo svete sa dá prepísať na Map/Reduce úlohy
  - Toto nie je o optimalizácií, ale o možnosti spracovávať „veľké dáta „
- Potenciál Hadoop-u je hlavne v komunite, ktorá ho spravuje (Apache Foundation)
- Nevýhoda je, že dnes máme troch najväčších kontribútorov (aj propagátorov) do kódu a nie je jasné akým smerom sa bude Hadoop v budúcnosti uberať:
  - Hortonworks (Yahoo!)
  - Cloudera
  - MapR
- Open source komunita už nemá priamy dosah na uberanie sa vývoja Hadoop-u

Letná škola IR na ÚI SAV

Pondelok	Utorok	Streda	Štvrtok	Piatok
MapReduce, PACT	Information Retrieval	NoSQL, CAP theorem	Nutch 1	Complex Networks
Hadoop 1	Lucene	HBase	Nutch 2	Graph databases
Hadoop 2	Information Extraction	Cassandra	Machine Learning	Semantic Search
Hive	GATE	?	Weka, Mahout	?

Termín: ***15.-19.7.2013*** *alebo* ***22.-26.7.2013*** *Možnosť registrácie na konkrétny deň*

Ďakujem za pozornosť

eFOCUS