„current issues in data management, database and information systems research"

Platinum Sponsors

facebook Google Microsoft
ORACLE

Gold Sponsors

EMC² Greenplum
IBM Research SAP SYBASE An SAP Company

Silver Sponsors

hp NEC TPC Transaction Processing Performance Council
YAHOO! LABS

Workshop: QDB '11, Quality in Databases

- Ahmed K. Elmagarmid (Qatar Computing Research Institute): **Insight into Data Cleaning and Linkage**

- Anastasios Karagiannis, Panos Vassiliadis (University of Ioannina), Alkis Simitsis (HP Labs): **Macro-level Scheduling of ETL Workflows**

- Wenfei Fan (University of Edinburgh): **Data Quality: Theory and Practice**

- Melanie Herschel, Torsten Grust (Universitaet Tubingen): **Transformation Lifecycle Management with Nautilus**

Distributed Entity Resolution (ER): Paradigms and Tools

distributed Key-Value Stores:

- distributed B-tree index for all attributes
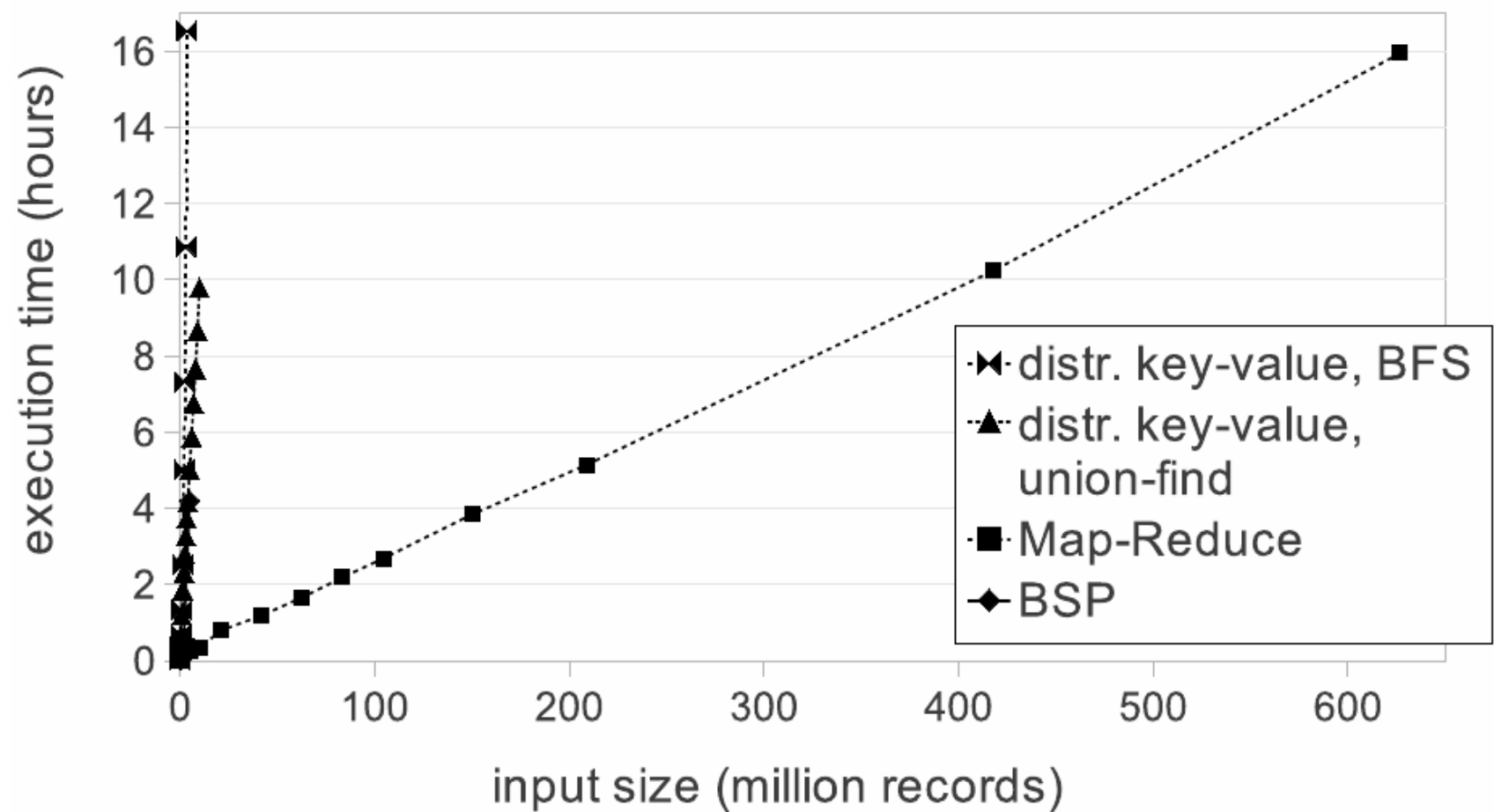- Project Voldemort

MapReduce:

- map → reduce operations
- Apache Hadoop

Bulk Synchronous Parallel:

- supersteps: computation → communication → barrier synchronization
- Apache Hama

Csaba István Sidló, András Garzó, András Molnár, and András A. Benczúr. **Infrastructures and bounds for distributed entity resolution**

Distributed ER Experiments

- 15 older blade servers, 4GB memory, 3GHz CPU each

- insurance client dataset (20 million records, ~ 2 records per entity)

- match logic: multiple attributes, combined in 5 features

Csaba István Sidló, András Garzó, András Molnár, and András A. Benczúr.
Infrastructures and bounds for distributed entity resolution

A Large-scale ER Application

- Intelius: „intelligently integrating information online for personal security and to inform the decision-making process"

- ehhez: publikus adatbázisok integrálása → large-scale ER

  - saját pipeline-szerű keretrendszerük:

    - attribútumok gazdagítása (külső tudás: pl. nevek gyakorisága), rekord-rekord párokra:

      - birthday_difference, regional population, street_address_match etc.

    - blocking

    - klasszifikátorok

  - előadásban: cost-sensitive Alternating Decision Tree

    - a tévedések irányának különbözőségére

    - érthető kimenet

    - bemenet: „several billion personal records"

Sheng Chen (Stevens Institute of Technology), Andrew Borthwick, and Vitor R. Carvalho(Intelius Data Research): **The Case for CostSensitive and Easy-To-Interpret Models in Industrial Record Linkage**

VLDB 2011: Research Sessions

- Database Design

- Query Processing

- Distributed Systems x 2

- Transaction Processing

- Integrity Maintenance

- Uncertain Data

- Causality, Quality, and Dependencies

- Privacy and Protection

- Statistical Methods x 2

- Data Integration

- Streams and Events

- Skyline and String Matching

- Entity Matching

- Web + Web Data

- Ranking + Searching and Ranking

- Graph Data x 3

- GeoSpatial

- Causality, Quality, and Dependencies

- New Hardware Architecture

- Cloud Computing and High-Availability

- MapReduce and Hadoop

- GPU-based Architectures and Column-store Indexing

- Human-Computer Interaction

- Social Networks

- Recommender Systems

VLDB 2011: Research Sessions

- Database Design
- Query Processing
- Distributed Systems x 2
- Transaction Processing
- Integrity Maintenance
- Uncertain Data
- Causality, Quality, and Dependencies
- Privacy and Protection
- Statistical Methods x 2
- **Data Integration**
- Streams and Events
- Skyline and String Matching

- **Entity Matching**
- **Web + Web Data**
- **Ranking + Searching and Ranking**
- **Graph Data x 3**
- GeoSpatial
- Causality, Quality, and Dependencies
- New Hardware Architecture
- **Cloud Computing and High-Availability**
- **MapReduce and Hadoop**
- **GPU-based Architectures and Column-store Indexing**
- Human-Computer Interaction
- Social Networks
- **Recommender Systems**

7

Comparison: ADBIS 2011 Research Sessions

Advances in Databases and Information Systems

- Central and East European countries

- theory and/or applications of database technology and information systems

- Query Processing x 2

- Business Process Management

- Data Warehousing x 2

- DB Systems

- Spatial Data

- Information Systems x 2

- Physical DB Design

- Evolution, Integrity, Security

- Data Semantics

- System Issues

„Global Brain"
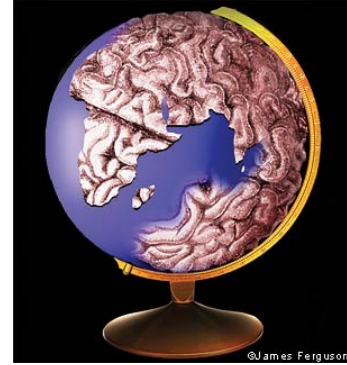
- „worldwide intelligent network formed by people together with the information and communication technologies that connect them into an "organic" whole" (Wikipedia)

  - computers + human problem solving

- Danny Hillis: "Global consciousness is that thing responsible for deciding that pots containing decaffeinated coffee should be orange."

- speed of knowledge propagation increases

- now developing with web applications: AI + collective intelligence of users → massive, interlinked cloud databases
  „we are the ants leaving feromone trails, that google is following" (*forrás: google keresés*)

  - eg.: human memory + google (megj., Benczúr A.: kérdések hossza, prefixek, emberi korlátok); sensing with mobile sensors (photo upload); ...

  - new kind of global brain, global consciousness

  - morality?, repsonsibility?, … (eg.: financial market automatization)

Keynote Talk: Tim O'Reilly (O'Reilly Media). **Towards a Global Brain.**

„Big Data"

- what „Big Data" means? are existing tools relevant?

  - modeling: bottleneck (vs. size) (more data we can model)

  - instead: automatic model generation („data frame theory"),

- big data patterns (example: personal GPS tracker):

  - digital shoebox: raw data + sourceID + instanceID

  - information production

  - model development

  - monitor, mine, manage (fraud detection, ...)

Keynote Talk: David Campbell (Microsoft). **Is it still „Big Data" if it fits in my pocket?**

Entity Matching: collective EM

- framework to scale any generic EM algorithm

  - collective EM: based on relations (eg. cites)

    - poor scalability

- running multiple instances of the EM algorithm on small neighborhoods of the data and passing messages across neighborhoods to construct a global solution

  - multiple instances + message passing

  - EM algorithms: black-boxes + properties (monotonicity)

  - accurate & sound algorithm

- experiments:

  - 2003 KDD Cup, HEPTH: papers, 58,515 author references, 13,092 authors, 29,555 papers
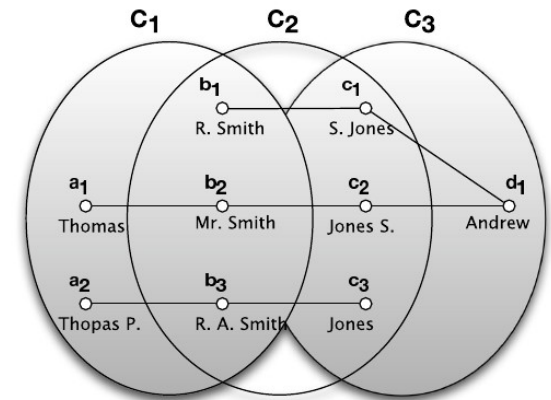
  - Hadoop impl.

Figure 2: A Cover of Entities.

Vibhor Rastogi (Yahoo! Research), Nilesh Dalvi (Yahoo! Research), Minos Garofalakis (Technical University of Crete): **Large-Scale Collective Entity Matching**

Main Memory Hybrid Storage: HYRISE

- predict the performance of different partitionings → select the best

- workload: OLTP vs. OLAP

- goal: real-time analytics → no separate OLAP system

- in memory: „we belive that many future databases … will fit into the memory of a small number of machines"

- similar: Data Morphing

- other vendors, hybrid storage layouts: Vertica FlexStore, VectorWise, Oracle, GreenPlum

$$r = (a_1 \ldots a_8)$$

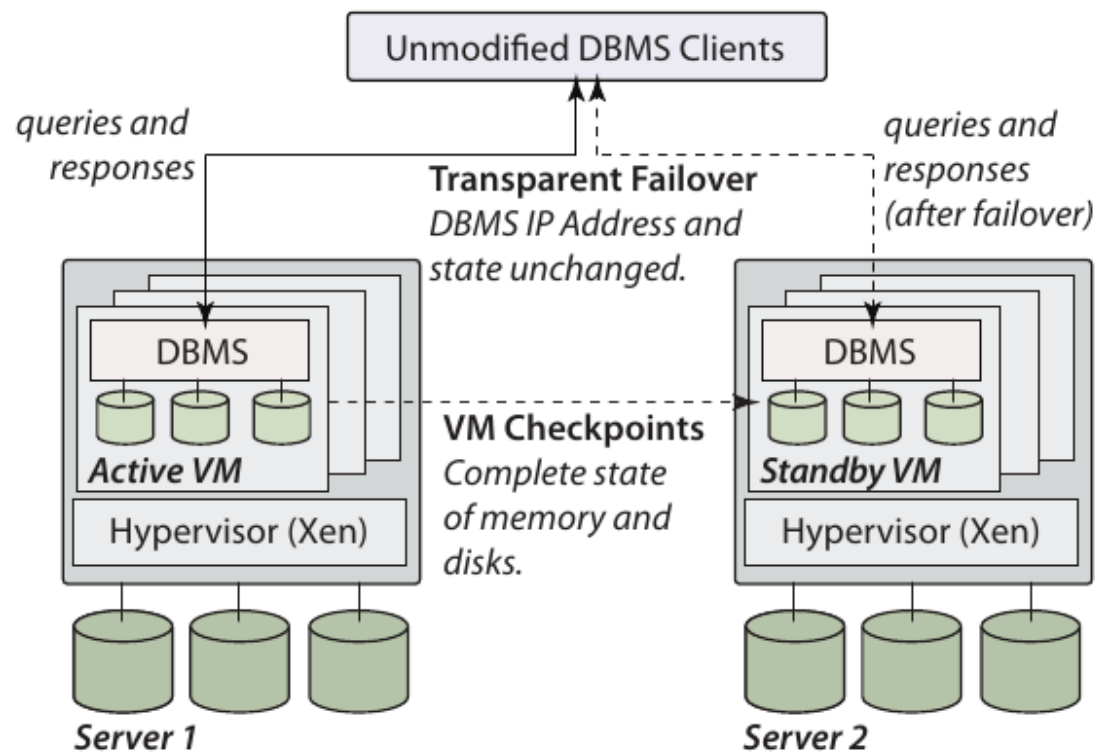| $C_1 (a_1)$	$C_2 (a_2 .. a_6)$	$C_2 (a_7 .. a_8)$
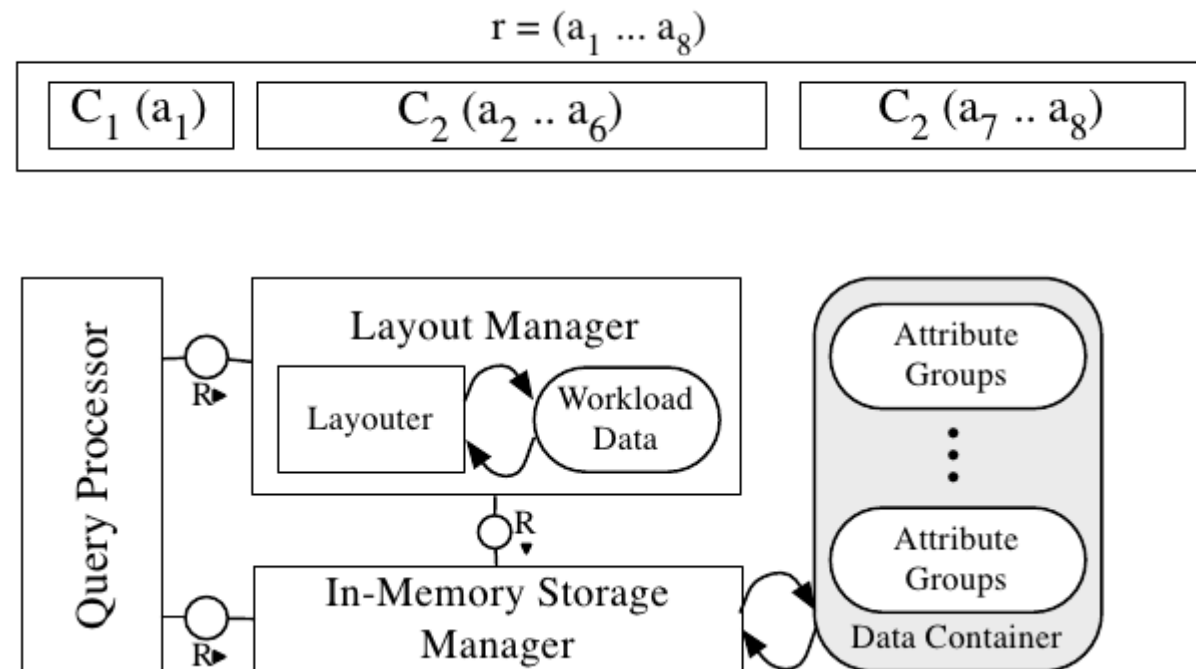
Martin Grund (Hasso-Plattner-Institute), Jens Krüger (Hasso-Plattner-Institute), Hasso Plattner (Hasso-Plattner Institute), Alexander Zeier (Hasso-Plattner Institute), Philippe Cudre-Mauroux (MIT), Samuel Madden (MIT). **HYRISE - A Main Memory Hybrid Storage Engine**

Column-Oriented MapReduce Storage

- parallel DBMS vs. MapReduce

- problems:

  - complex data types

  - writing map and reduce functions, no declarative query language

  - default Java

- solution:

  - lazy record construction

  - compression (LZO, simple dictionaries)

  - Pig, Hive, Jaql: declarative → further optimization

  - Column-oriented storage formats for HDFS blocks

- experiments, HDFS replication:

  - file → binary format: 3x speedup

  - file → column-based containers: 1-2 magnitudes of speedup

Avrilia Floratou, Jignesh Patel, Eugene Shekita, Sandeep Tata. **Column-Oriented Storage Techniques for MapReduce** (IBM, ...)
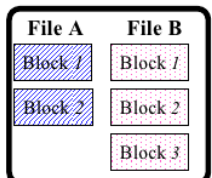
High-Availability RemusDB (best paper)

- HA: implemented in DB ↔ service below DB?

- traditional HA: high performance overhead

- active-standby HA with DBs in virtual machines

  - failover with virtualization layer → transparent to DB

  - Remus: VM replication + PostgreSQL, MySQL

- DBMS-aware VM checkpointing system: RemusDB

  - reduce states transfered during checkpoints

  - reduce latency: minor DB code modifications

- experiments: TCP benchmarks

  - low overhead, fast failover

Unmodified DBMS Clients

queries and responses

Transparent Failover DBMS IP Address and state unchanged.

queries and responses (after failover)

DBMS

Active VM

Hypervisor (Xen)

VM Checkpoints Complete state of memory and disks.

DBMS

Standby VM

Hypervisor (Xen)

Server 1

Server 2

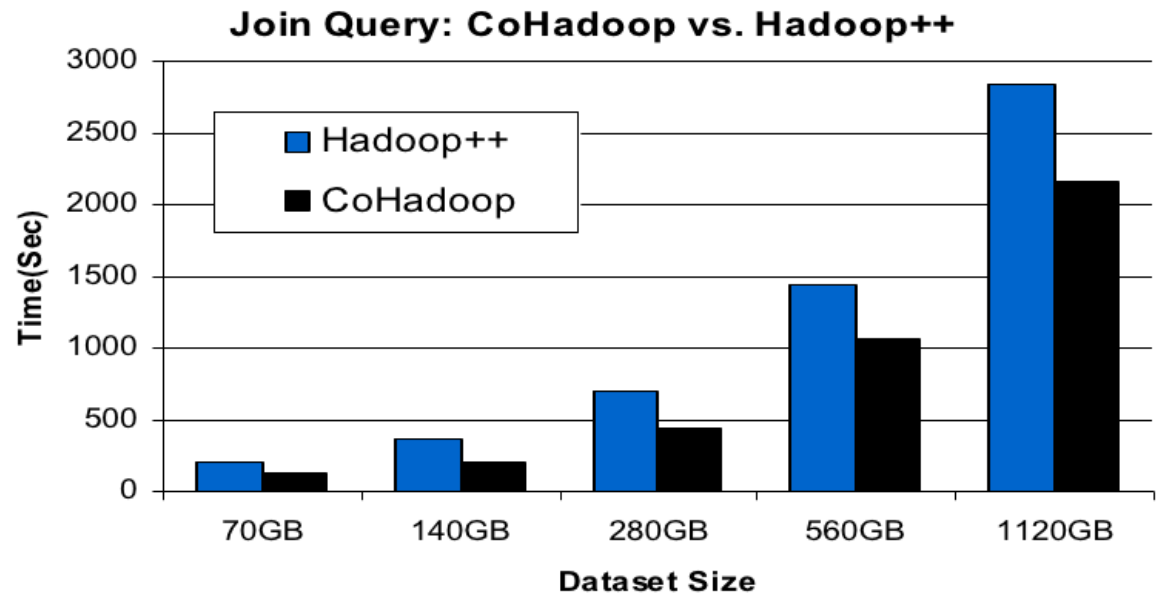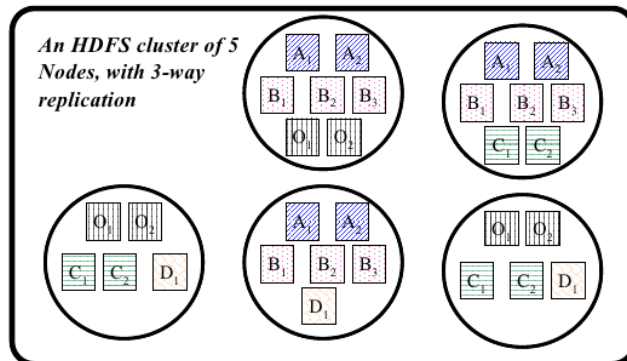# Data Placement in Hadoop

- relocate related data on the same set of nodes

- CoHadoop: lightweiht extension (to HDFS)

  - no data format requirement

  - applicaton hints: related data

- HDFS: random block placement

- motivation: event log processing (sessionization)

Mohamed Eltabakh, Yuanyuan Tian, Fatma Ozcan, Rainer Gemulla, Aljoscha Krettek, John McPherson. **CoHadoop: Flexible Data Placement and Its Exploitation in Hadoop** (IBM, Max-Planck)

Main Memory Column-stores: Indexing

- preliminary index structures: low cost, specific key ranges

- adaptive indexing: based on usage of key ranges

  - database cracking: costly init, fast convergence,

    - adaptive AVL tree - „incremental quicsort"

  - adaptive merging: low init cost, slow conv.,

    - „incremental mergesort"

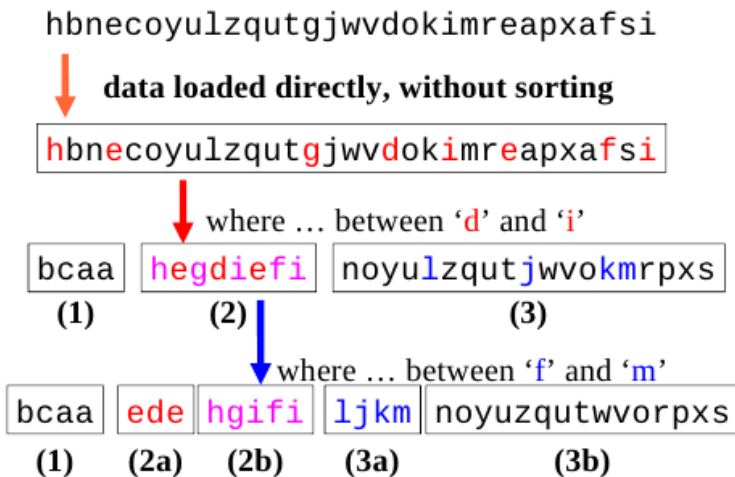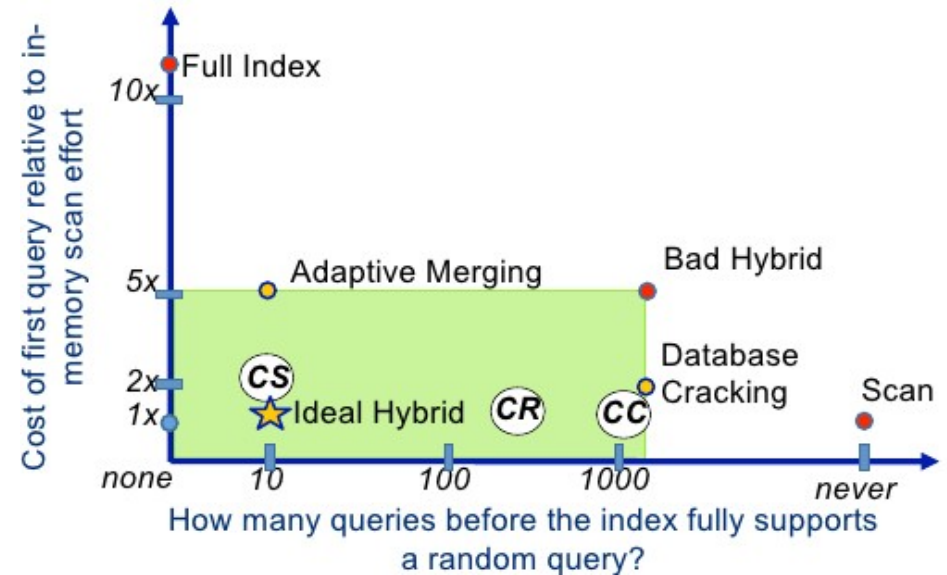- hybrid method: low initialization cost, rapid convergence
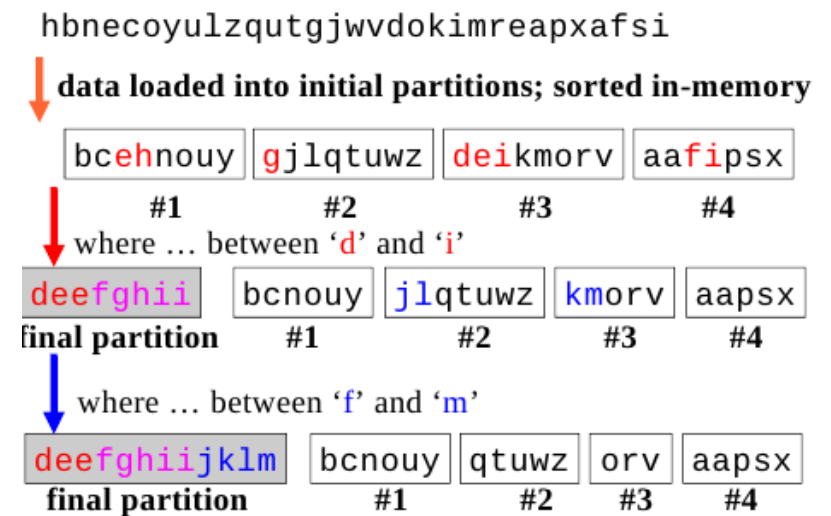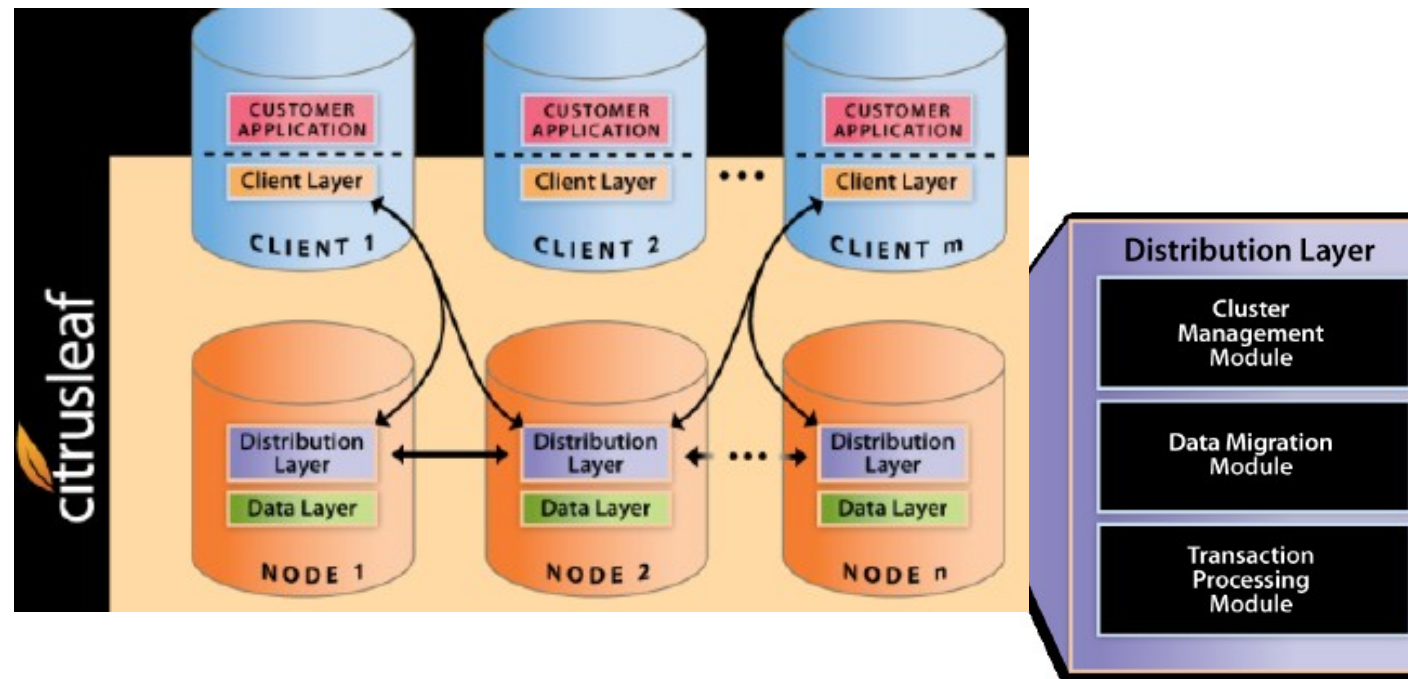

Figure 2: Database cracking.

Figure 3: Adaptive merging.

Stratos Idreos, Stefan Manegold, Harumi Kuno, Goetz Graefe. **Merging What's Cracked, Cracking What's Merged: Adaptive Indexing in Main-Memory Column-Stores** (HP Labs, CWI)

citrusleaf™ - ACID NoSQL (industrial)
the real-time noSQL database

- application: realt-time bidding for advertising, campaign management, social networking (gaming)

- goal: linear scalability + ACID, consistency, backup/restore, high availability, …

- key-value store, no SQL support

- állítás: jobb, mint VoltDB, Clustrix, Cassandra, MongoDB, Redis, …

V. Srinivasan (Citrusleaf), Brian Bulkowski (Citrusleaf). **Citrusleaf: A Real-Time NoSQL DB which Preserves ACID**

Tenzig: SQL on MapReduce (industrial)

- mostly complete SQL, high performance, scalability, reliability, low latency, columnar data support, …

- 1,000+ users, 10,000+ queries / day, 1.5 PB data

- SQL → MapReduce (Sawzall, Flume-Java, PIG, HIVE, HadoopDB); MapReduce → DBs (AsterData, GreenPlum, Paraccel, Vertica)

- motivation: Google Ads data warehouse

  - costly scalability, rapidly increasing loading times, limited analyst creativity

- implement DB optimizations with slight MapReduce modifications

- impl.: distributed worker pool (→ low latency), query server, metadata server

- optimization:

  - for projection, filtering, joins (boradcast, sort-merge, hash)

  - be aware of the heterogeneous sources (BigTable)

Biswapesh Chattopadhyay, Liang Lin, Weiran Liu, Sagar Mittal, Prathyusha Aragonda, Vera Lychagina, Younghee Kwon, Michael Wong. **Tenzing - A SQL Implementation on the MapReduce Framework** (Google)

Distributed Systems: Paxos

- Paxos: „family of protocols for solving consensus in a network of unreliable processors"

- „Spinnaker" experimental datastore

  - Paxos: replication protocol (vs. two-phase commit)

    - fast (-er!) for reads, 5-10% slower for writes as alternatives
    - works if majority of nodes are alive

- may be used for (?): Amazon Dynamo, Google Bigtable, Yahoo PNUTS, MS SQL Azure, FAWN
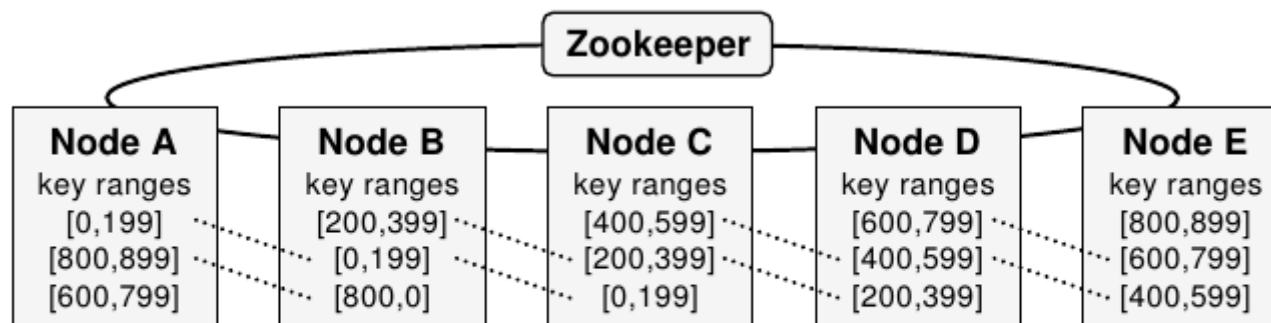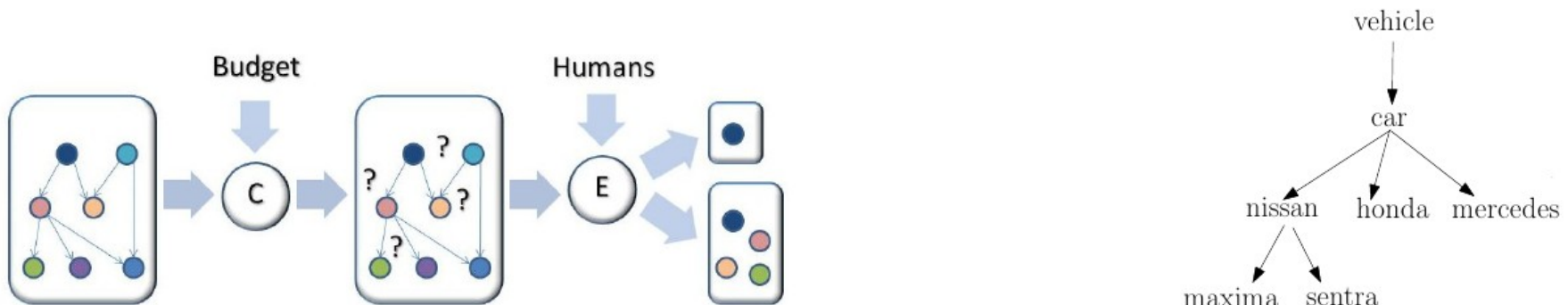
Figure 2: Example of a Spinnaker cluster.

Jun Rao, Eugene Shekita, Sandeep Tata. **Using Paxos to Build a Scalable, Consistent, and Highly Available Datastore** (LinkedIn, IBM)

Human-Assisted Graph Search

- DAG + "Is there a target node that is reachable from the current node?"

- applications: interactive search, image segmentation, debugging workflows + other „crowd sourcing": text summarization, labeling, ranking … (CrowDB!)

  - human computation optimizer!

  - single / multi, bounded / unlimited, DAG / downward-forest / upward forest

- experiments: DMOZ concept hierarchy + internet director

Aditya Parameswaran, Anish Das Sarma, Hector Garcia-Molina, Neoklis Polyzotis, Jennifer Widom. **Human-Assisted Graph Search: It's Okay to Ask Questions**
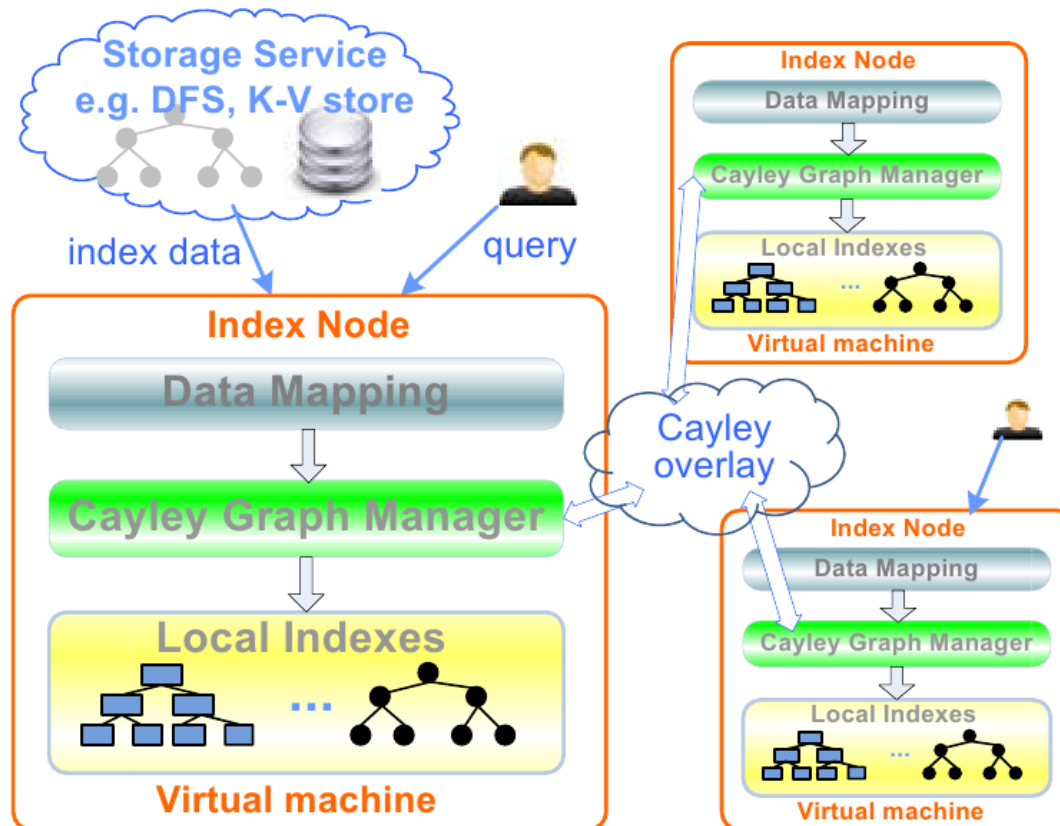
Social Networks

- Structural Trend Analysis For Online Social Networks (UCSB)

  - use of friend information (relations)

  - detection methods for coordinated and uncoordinated (viral) topics

- On Social-Temporal Group Query with Acquaintance Constraint

  - find activity time and attendees with minimum total social distance

  - NP-hard problem → efficient pruning

- Social Content Matching in MapReduce (Yahoo, Max-Planck)

  - match content of suppliers and consumers (Flickr, Yahoo! Answers, …) ; bipartite graph (user - content)

  - maximize overall relevance

Cloud and Indexing

- „database as a service"
  - requires: distributed indexes for clouds
- framework for users: define their own idexes; P2P model + Cayley graph
- implementations: distr. B+, multi-dim., distr. hash
- experiments: Amazon EC2

Gang Chen, Hoang Tam Vo, Sai Wu, Beng Chin Ooi, M. Tamer Özsu. **A Framework for Supporting DBMS-like Indexes in the Cloud**

BIRTE '11: Real Time Business Intelligence

- Guy Lohman (IBM Almaden Research Center, USA): **Blink: Not Your Father's Database**

- Qiming Chen, Meichun Hsu, Ren Wu (HP Labs, USA): **A cost-aware strategy for merging differential stores in column-oriented in-memory DBMS**

- José Blakeley (Microsoft Corporation): **Microsoft SQL Server Parallel Data Warehouse – Architecture Overview**

- Sang Kyun Cha (Seoul National University & SAP, Korea): **SAP HANA: Breaking Vertical and Horizontal Tiers in Enterprise with High-Performance Distributed In-Memory Database**

- Shilpa Lawande, Andrew Lamb, Lakshmikant Shrinivas (Vertica/Hewlett Packard): **Scalable Social-Graphing Analytics with the Vertica Analytic Platform**

Challenges and Vision

- Data Markets in the Cloud: An Opportunity for the Database Community

- Data is Dead... Without What-if Models (IBM)

- Antropocentric Data Systems

- ...

További érdekes cikkek

- Fast Sparse Matrix-Vector Multiplication on GPUs: Implications for Graph Mining

- Graph Indexing of Road Networks for Shortest Path Queries with Label Restrictions (UCR)

- Entity Matching: How Similar is Similar

- Compression Aware Physical Database Design (Microsoft)

- Jaql: A Scripting Language for Large Scale Semistructured Data Analysis (IBM)

- Serializable Snapshot Isolation for Replicated Databases in High-Update Scenarios

- Automatic Optimization for MapReduce Programs

- Generating Efficient Execution Plans for Vertically Partitioned XML Databases