NoSQL

New approaches for data management

Outline

- What's NoSQL?
- Introduction to column-stores
- MonetDB

What's NoSQL

- Not only SQL
- Term used first in 1998, then in 2009 in a NoSQL meeting at San Francisco
- Need to solve a problem that relational databases are bad fit for
- No to «one size fits all» thinking of traditional DBMS

New data processing markets

- Data warehouses
- stream processing
- text processing
- scientific-oriented databases
- semi-structured data

Relational database systems

- Developed in ‘70s
- Frequently unneeded complexity (i.e., ACID properties)
- Hardware has evolve, DBMS don’t
- No suitable for new large-scale applications

Goal of NoSQL

- Scale out data
- Performance on single servers
- Rigid schema design
- Design considerations
  - Main memory
  - Multi-threading
  - Grid computing
  - High availability
  - No knobs but self-everything

Stonebraker, Michael: The "NoSQL" Discussion has Nothing to Do With SQL. November 2009. – Blog post of 2009-11-04.
http://cacm.acm.org/blogs/blog-cacm/50678-the-nosql-discussion-has-nothing-to-do-with-sql/fulltext

Types of NoSQL data stores (1/2)

- Document stores. The central concept of a document store is the notion of a "document" that encapsulate and encode data (or information) in some standard formats or encodings. Encodings in use include XML, YAML, and JSON as well as binary forms like BSON.

- Key-value stores. In this model, data is represented as a collection of key-value pairs, such that each possible key appears at most once in the collection. The key-value model is one of the simplest non-trivial data models, and richer data models are often implemented on top of it. Key-value stores can use consistency models ranging from eventual consistency to serializability. Some support ordering of keys. Some maintain data in memory (RAM), while others employ solid-state drives or rotating disks

Types of NoSQL databases (2/2)

- Graph database. This kind of database is designed for data whose relations are well represented as a graph (elements interconnected with an undetermined number of relations between them). The kind of data could be social relations, public transport links, road maps or network topologies.

- Column-oriented database. It is a database management system (DBMS) that stores data tables as sections of columns of data rather than as rows of data. In comparison, most relational DBMSs store data in rows.

Categorization of NoSQL data stores

	Performance	**Scalability**	**Flexibility**	**Complexity**	**Functionality**
Key-Value Stores	high	high	high	none	variable (none)
Column stores	high	high	moderate	low	minimal
Document stores	high	variable (high)	high	low	variable (low)
Graph databases	variable	variable	high	high	graph theory
Relational databases	variable	variable	low	moderate	relational algebra

NoSQL implementations

- key-value stores
  - http://en.wikipedia.org/wiki/NoSQL#Key-value_stores
- Document stores
  - http://en.wikipedia.org/wiki/Document-oriented_database
- Graph databases
  - http://en.wikipedia.org/wiki/Graph_database
- Column-oriented databases
  - http://en.wikipedia.org/wiki/List_of_column-oriented_DBMSes

Column-store

Introduction and main concepts

- Extract of the VLDB Tutorial of 2009
- MonetDB at a glance

Bibliographie

- Strauch, C., Sites, U. L. S., & Kriha, W. (2011). NoSQL databases. Lecture Notes, Stuttgart Media University. http://home.aubg.bg/students/ENL100/Cloud%20Computing/Research%20Paper/nosqldbs.pdf

- Abadi D, Boncz P, Harizopoulos S, Idreos S, Madden S. The Design and Implementation of Modern Column-Oriented Database Systems. Foundations and Trends in Databases. 2013;5(3):197-280. http://stratos.seas.harvard.edu/files/stratos/files/columnstoresfntdbs.pdf