

NoSQL (Not Only SQL)

NoSQL ou Not Only SQL

- ▶ **Contexto:** SGDBs tradicionais estão com **baixa performance** em aplicações com **uso intensivo** de dados, indexação de **quantidades gigantescas de documentos**, rapidez para fornecer dados para sites de **altíssimo tráfego** e arquivos **multimídia**
- ▶ **Exemplos:** 3 TB do Digg, 50 TB caixa de entrada do Facebook, 2 PB do eBay etc
- ▶ NoSQL entendido como “**non-relational**”
Conferência em 2009 sobre armazenamento de dados de maneira não relacional (open source, distributed non relational databases) retomou nomenclatura
"select fun, profit from real_world where relational=false;"

NoSQL, trocando o *modelo*

- ▶ Implementações NoSQL normalmente não fornecem todas as garantias de consistência típicas de SGDBs relacionais tradicionais
- ▶ A grande maioria não satisfaz completamente as propriedades **ACID** (atomicity, consistency, isolation, durability) para a garantia das transações

Adotam o modelo **BASE** com três princípios:

- ▶ basic availability: abordagem distribuída com alta replicação => foco na disponibilidade
- ▶ soft state: consistência de dados é problema do programador e não do banco
- ▶ eventual consistency: não há garantia de consistência no tratamento das transações (ex: início e fim)

Arquitetura

- ▶ A arquiteturas são **distribuídas** com dados **redundantes** em diversos servidores (usando tabelas Hash distribuídas)
- ▶ As interfaces são baseadas em **arrays associativos** ou **pares chave–valor**
- ▶ Implementações XML adotam Xquery como linguagem de consulta

Categorias de projetos NoSQL

- ▶ **Document-oriented database**
 - Exemplos: [MongoDB](#), [Apache CouchDB](#)
- ▶ **Graph database**
 - Exemplos: [Neo4J](#), [AllegroGraph](#), [Core Data](#), [OrientDB](#) (Mac/iPhone)
- ▶ **Pares Chave–Valor (key–value pairs)**
 - Várias implementações: disco, Ram, ordenadas etc
 - Exemplos principais:
 - [BigTable](#): proprietário do Google
 - [Dynamo](#): proprietário da Amazon
 - [Cassandra](#): projeto open source (Apache)

Apache



Cassandra

- ▶ Banco de dados distribuído usado pelo Facebook que iniciou seu desenvolvimento (feito em **Java**) baseado do modelo **BigTable** e no Dynamo da Amazon
- ▶ Projeto open source da Apache Foundation
- ▶ Twitter migrou de MySQL para Cassandra
- ▶ **Similares**: MapReduce, Hadoop, Hbase, Hypertable e outros baseados no BigTable
- ▶ **Usuários**: Facebook, Netflix, Digg, Twitter, eBay, Reddit entre outros
- ▶ Maior projeto em produção com Cassandra:
100TB de dados distribuídos em **150 servidores**
- ▶ **<http://cassandra.apache.org/>**

MongoDB



- ▶ Banco de dados open source orientado a documentos desenvolvido em C++ (10gen)
- ▶ Nome vem de **humongous** (extremamente grande)
- ▶ MongoDB gerencia documentos JSON-like (JavaScript Object Notation)
- ▶ Forte integração com JavaScript
- ▶ **Similares**: Jackrabbit , Apache CouchDB, RavenDB etc
- ▶ **Usuários**: foursquare, bit.ly, SourceForge etc
- ▶ <http://www.mongodb.org/>

CouchDB



- ▶ DB de código-aberto que foca cuja proposta é "um banco de dados que abrange a Web"
- ▶ Orientado a documentos, mantém certa consistência sendo fácil de usar e replicar
- ▶ Utiliza JSON para armazenamento, JavaScript para consulta usando MapReduce
- ▶ RESTful web interface, Padrão HTTP como API: utiliza os métodos POST, GET, PUT, DELETE para as operações CRUD (Create, Read, Update, Delete) em todos os recursos
- ▶ **Usuários**: apps do Facebook, BBC, CERN, ShareGrove, etc
- ▶ <http://couchdb.apache.org/>

Considerações finais

- ▶ Comparação: <http://kkovacs.eu/cassandra-vs-mongodb-vs-couchdb-vs-redis>
- ▶ Considerar a possibilidade de experimentar/usar um database **NoSQL**
 - Pelo perfil da **categoria** de implementação
 - Pelo modelo de **arquitetura** distribuída
 - Por questões de **performance**
- ▶ **Projetos em evidência**
 - Cassandra (grandes qtdes dados, logs etc)
 - MongoDB (ruby) x CouchDB (js, jquery)
 - Neo4J (java)

Referências (1)

- ▶ [The Definitive Guide to NoSQL](#)
- ▶ [http://nosql-databases.org/](#)
- ▶ [The NoSQL Alternative](#) (artigo na Dr. Dobb's)
- ▶ [Bigtable: A Distributed Storage System for Structured Data](#) (artigo publicado no OSDI 2006 – Symposium on Operating System Design and Implementation)
- ▶ [Dynamo: Amazon's Highly Available Key-value Store](#) (artigo publicado no ACM Symposium on Operating Systems Principles – SOSP 2007)



Referências (2)

- ▶ Up and Running with Cassandra
- ▶ **MongoDB: The Definitive Guide**
Chodorow/Dirolf (O'Reilly)
- ▶ **MongoDB for Web Development**
Pirtle (Developer's Library)
- ▶ **CouchDB: The Definitive Guide**
Anderson/Lenhardt/Slater (O'Reilly)

