NoSQL

Tamas Piros

NoSQL databases

- Different types of databases exist under the label of NoSQL
  - Document based - CouchDB, MongoDB
  - Graph based - Neo4J
  - Key/value store - Aerospike
  - Multi-model databases - MarkLogic, ArrangoDB

Key challenges with RDBMS

- Strict / rigid schema
- Ever changing data requirements
  - Internal changes
  - Acquisitions
- Data marts, data lakes
  - How to update them?

Firstname	Lastname	Email
Joe	Smith	joe@corp.org
Susan	Black	susan@corp.org

First	Given	Email
Adam	Smith	adam@corp.org
Kate	Brown	kate@corp.org

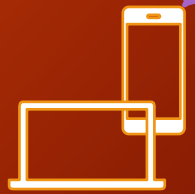
Firstname	Lastname	Email
Joe	Smith	joe@corp.org
Susan	Black	susan@corp.org

First	Given	Email
Adam	Smith	adam@corp.org
Kate	Brown	kate@corp.org

Data Lake

First	Last	Email
Joe	Smith	joe@corp.org
Susan	Black	susan@corp.org
Adam	Smith	adam@corp.org
Kate	Brown	kate@corp.org

New App?
Update to data?
Multiple apps updating data?



(De)Normalisation

- Remove duplicate data - place them in separate tables
- NoSQL is schema agnostic
- In NoSQL - keep everything related about an entity in a single document
  - Better read performance
  - If updates are crucial, consider document joins
  - Better represents the final data structure for the app (no ORM)
  - Minimise One-To-Many relationships

NoSQL vs RDBMS

RDBMS	NoSQL (Document based)
Database	Database
Table	Collection
Row	Document
Column	JSON property
Index	Index
Query	Query

Representation of a resource

```
{
  "employee": {
    "name": "Steven",
    "salary": "£25000",
    "email": "steven@corp.org",
    "department": {
      "name": "Engineering",
      "location": "London"
    }
  }
}
```

Primary Key in NoSQL

- Unique way to identify a row of data
- In NoSQL:
  - URI
  - ObjectID
  - Document ID

Table vs Collection

Employees

First	Last	Email
Joe	Smith	joe@corp.org
Susan	Black	susan@corp.org

Collection: employees

```
{
  "employee": {
    "first": “Joe”,
    "last": “Smith”,
    "email": "joe@corp.org",
  }
}
```

```
{
  "employee": {
    "first": “Susan”,
    "last": “Black”,
    "email": “susan@corp.org",
  }
}
```

Types of NoSQL databases

- Key-value store (ArangoDB, Redis, DynamoDB)
- Document store (Couchbase, DocumentDB, MongoDB, MarkLogic)
- Graph database (Neo4j, MarkLogic, Virtuoso)
- Object database (ObjectDB, ZODB)
- Tabular (BigTable, Hbase)
- RDF database (MarkLogic, Virtuoso)
- Multi-value database (Infinity DB)
- Multi-model database (MarkLogic, Couchbase)

NoSQL is Designed to Handle change

- NoSQL systems have flexibility out of the box
- No requirements for a schema; schema agnostics approach
- NoSQL is great for heterogeneous data
  - No schema means, any document with any structure can be inserted
- NoSQL databases prefer a data-first approach vs a schema first approach

NoSQL is designed for scale

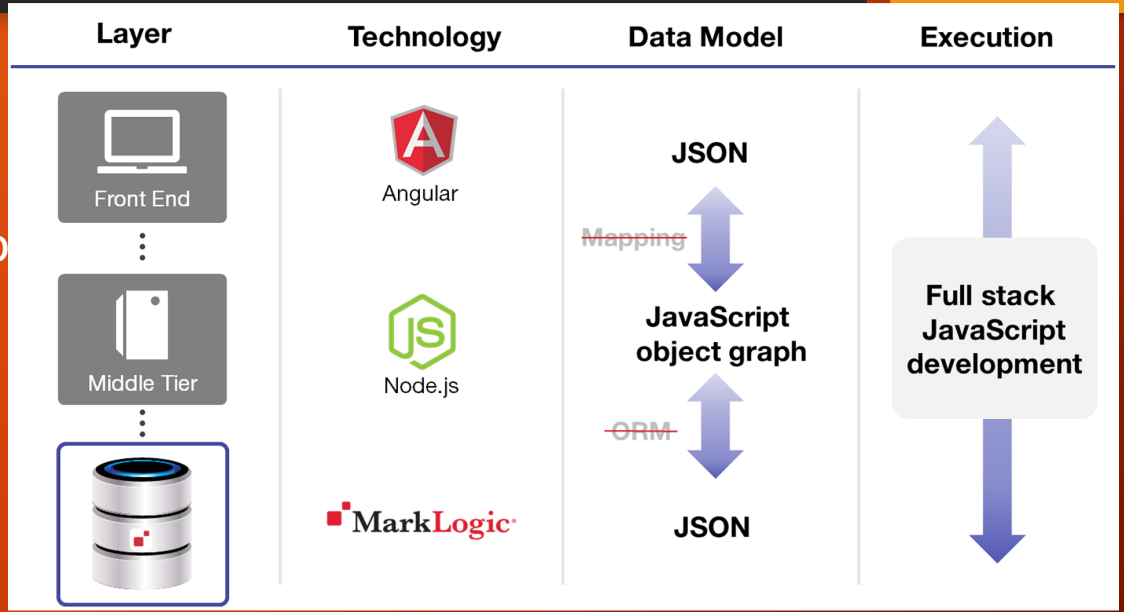
- NoSQL databases are designed for massive scale on distributed systems
- Horizontal scalability
  - Database is run on multiple servers that work together, each sharing part of the load
- Elasticity
  - Add or remove nodes in the cluster
  - Data rebalancing is automatic

NoSQL is for mixed workloads

- NoSQL databases can handle both operational and analytical workloads
- This setup requires fewer architectural components, less maintenance, less complexity and faster responses
- Using a great variety of indexes help to handle such workloads
- Allows to have a single source of truth

NoSQL is for modern app development

- No ORM required
- Simple, API access
- Allows for a more agile appro



Differentiator between NoSQL databases

- NoSQL databases have different properties
- They can be rated by performance, scalability, flexibility complexity and functionality
- ”NoSQL cannot support ACID” – incorrect statement
  - Some NoSQL databases, including MarkLogic, do support ACID
- NoSQL database are great for real-time alerting, real-time search, data enrichment, integrating data silos

Services that you use everyday

- Facebook (Messaging) - HBase
- LinkedIn - Espresso
- Google Mail / Inbox - BigTable
- Yell.com - MarkLogic
- HealthCare.gov (“Obamacare” – Marketplace & Exchanges) - MarkLogic
- Top Investment Banks – MarkLogic
- NBC (Saturday Night Live) - MarkLogic