Multi-Model databases

Aurelijus Banelis

Aurelijus Banelis

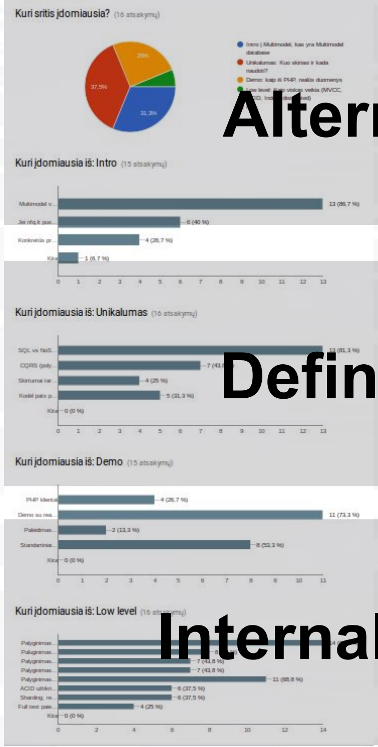
aurelijus.banelis.lt

aurelijus@banelis.lt

.NFQ

Auginte

Multi-Model databases



WHY

Alternatives, problems

WHAT

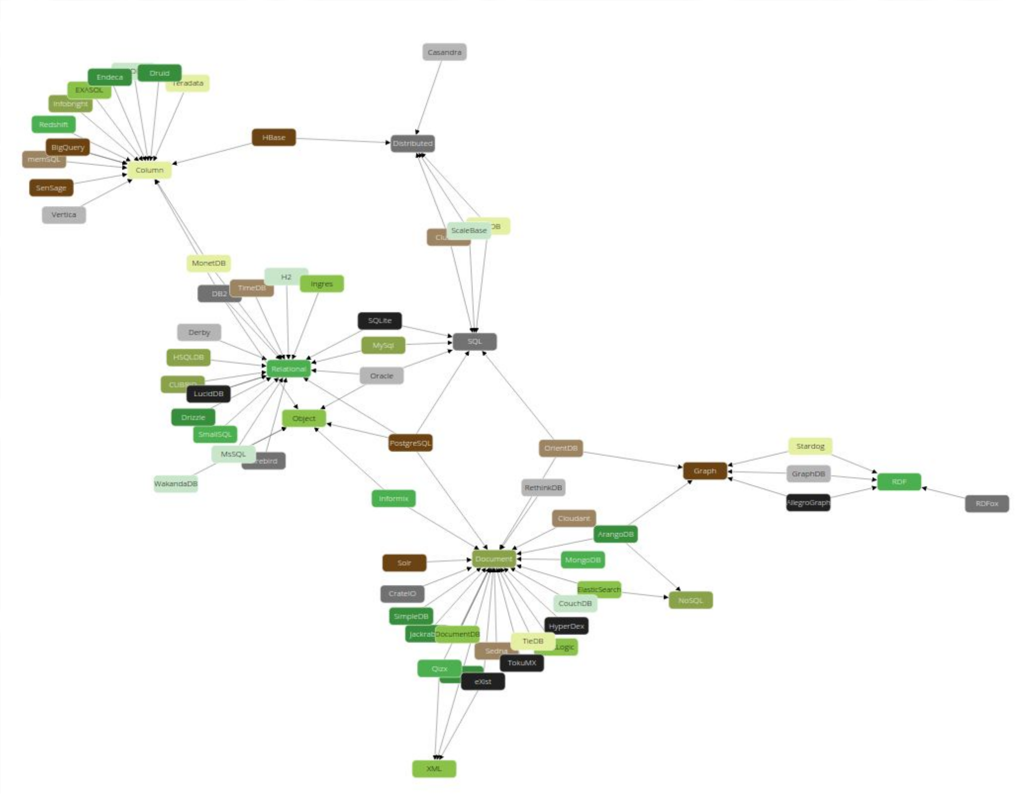
Definition, basic usage

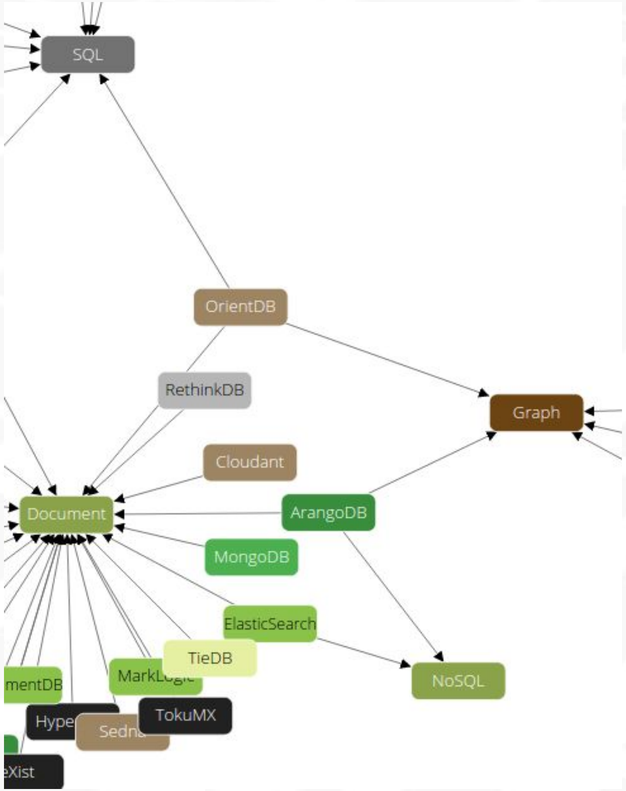
HOW

Internals, pros & cons, tips

Do we need another DB?

There are already a lot of relational, graph, column, RDF, key-value, distributed, SQL, noSQL, newSQL databases (and variation of those)





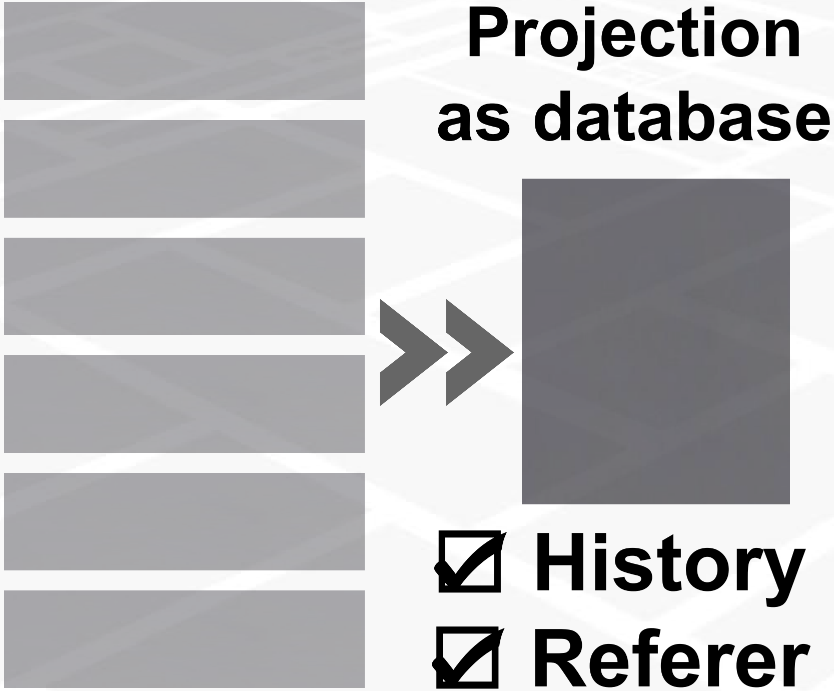
Trends in DBMS

Own	SQL	NoSQL	CQRS	Multi-Model
Red-Black, binary tree R-tree	Tables Joins Indexes	Collections Nesting Indexes	Event, Command, Projection	Relations Nesting Indexes
Custom structures Speed	ACID Transactions	Scaling Async	Speed Reports	Choice for structures and ACID
Reinventing the wheel	Sharding Replication	Relations	Technology zoo, tools	Not silver bullet

Events vs Multimodel

CQRS versus multimodel database

Events



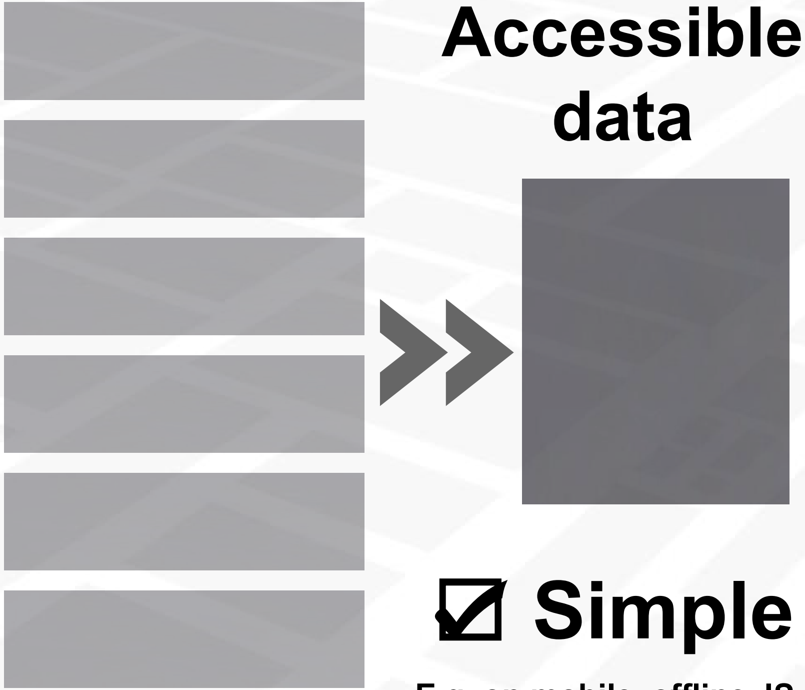
Projection as database

☑ History

☑ Referer

Solving conflicts later

WAL



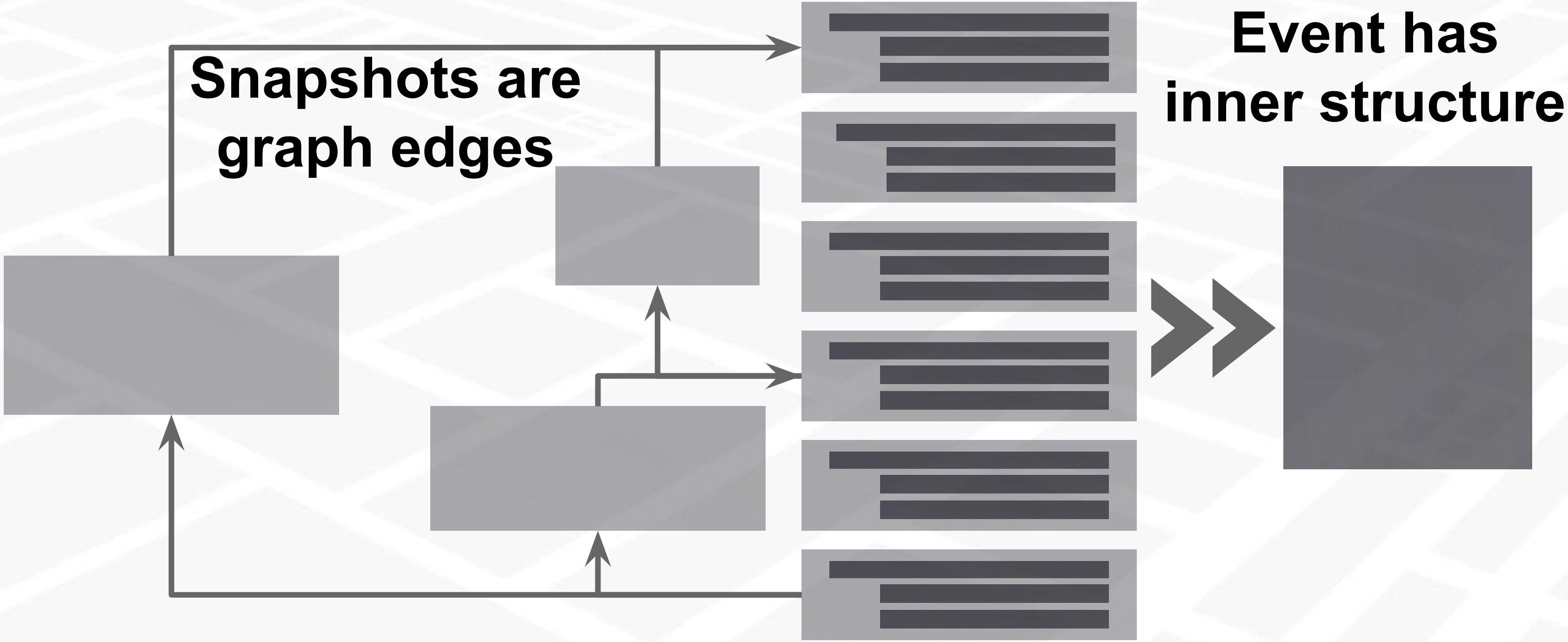
Accessible data

☑ Simple

E.g. on mobile, offline JS

CQRS with Multimodel

Implementation of snapshots in Event database. Made my choice



Snapshots are graph edges

Event has inner structure

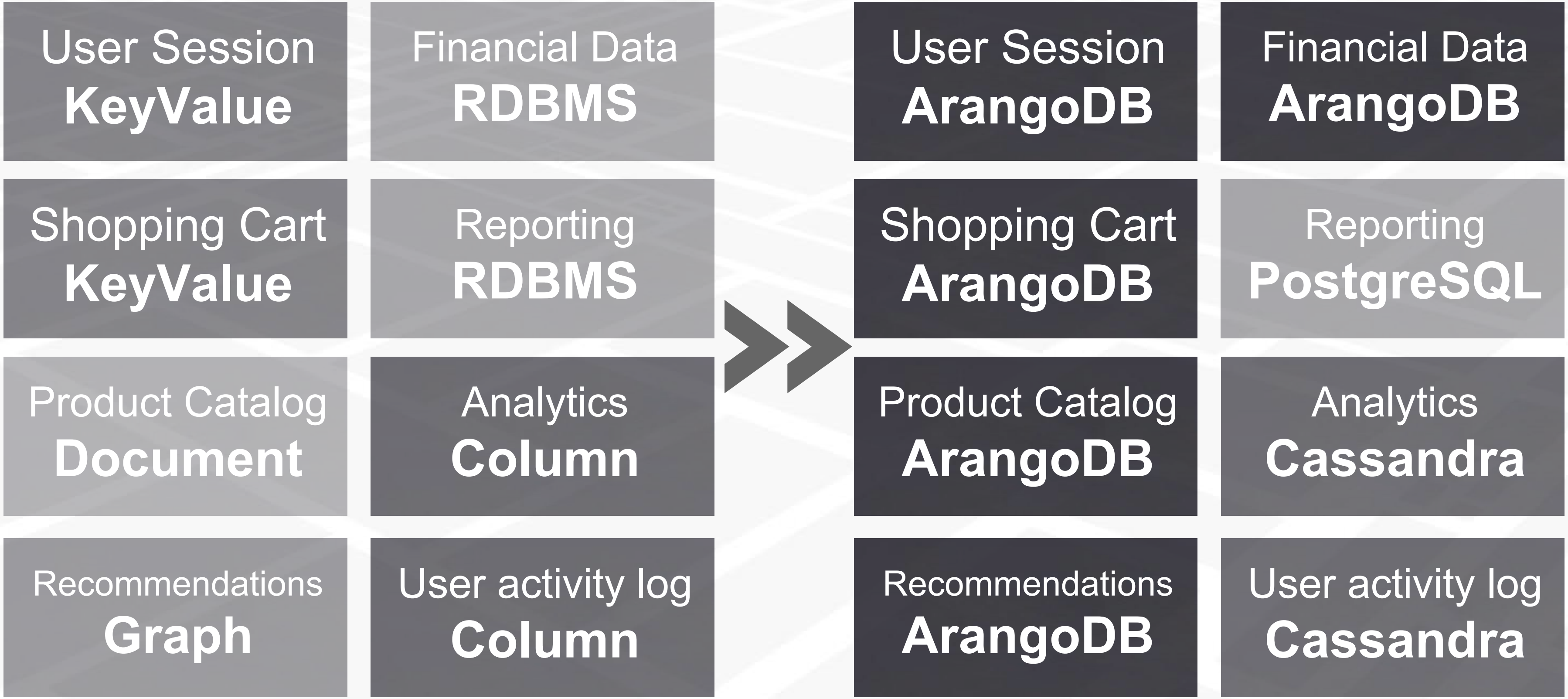
WHY

WHAT

HOW

Polyglot Persistence

CQRS is very powerful only if data can be split correctly



User Session
KeyValue

Financial Data
RDBMS

Shopping Cart
KeyValue

Reporting
RDBMS

Product Catalog
Document

Analytics
Column

Recommendations
Graph

User activity log
Column

User Session
ArangoDB

Financial Data
ArangoDB

Shopping Cart
ArangoDB

Reporting
PostgreSQL

Product Catalog
ArangoDB

Analytics
Cassandra

Recommendations
ArangoDB

User activity log
Cassandra

I chose MMD because

1. **Multiple data structures** and operations used in application (especially graph based)
2. Needed common data pattern for desktop, web, offline frontend and mobile: **common implementation of basic functionality, easy to synchronize**
3. Needed a way to track changes/**history** (CQRS)
4. Stable and **reliable**
5. Possibilities for **scaling**/distributing
6. Split data by user: **autonomous database**

MMD - MultiModel database

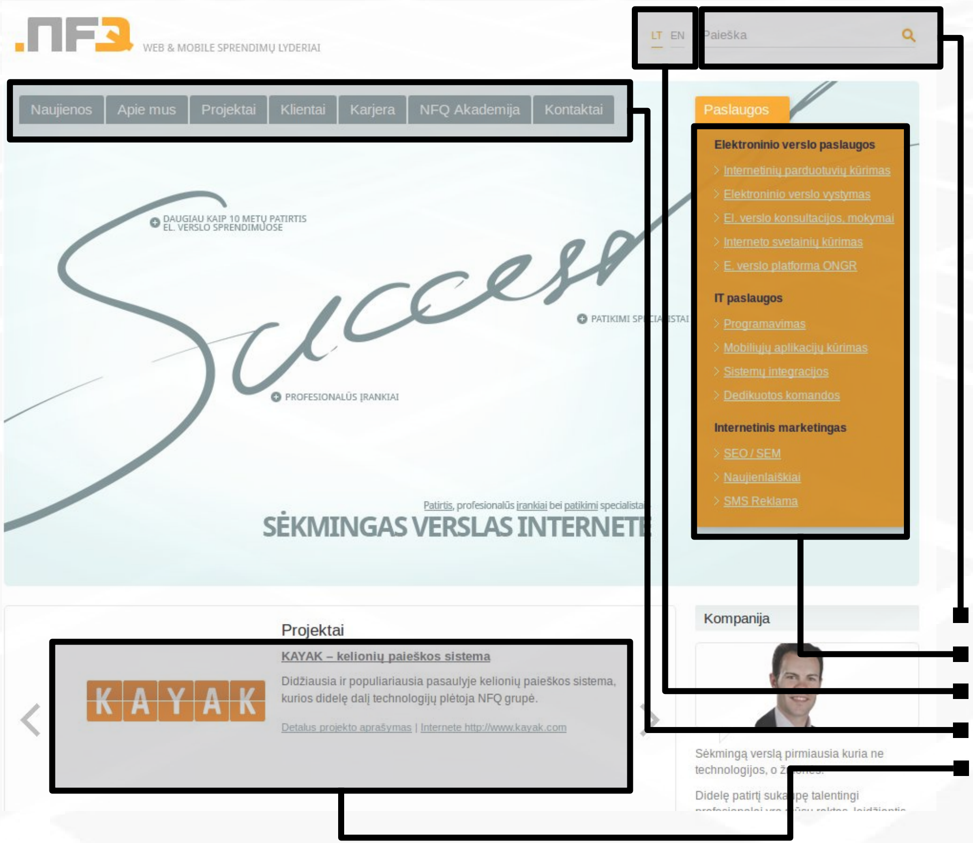
Definition

... designed to support

multiple data models against a single, integrated backend

WIKIPEDIA
The Free Encyclopedia

Single backend



.NFQ WEB & MOBILE SPRENDIMŲ LYDERIAI

LT EN

Paieška

Naujienos | Apie mus | Projektai | Klientai | Karjera | NFQ Akademija | Kontaktai

DAUGIAU KAIP 10 METŲ PATIRTIS EL. VERSLO SPRENDIMUOSE

PATIKIMI SPECIALISTAI

PROFESIONALŪS ĮRANKIAI

Patirtis, profesionalūs įrankiai bei patikimi specialistai

SĖKMINGAS VERSLAS INTERNETE

Paslaugos

Elektroninio verslo paslaugos
- Internetinių parduotuvių kūrimas
- Elektroninio verslo vystymas
- El. verslo konsultacijos, mokymai
- Interneto svetainių kūrimas
- E. verslo platforma ONGR

IT paslaugos
- Programavimas
- Mobiliųjų aplikacijų kūrimas
- Sistemų integracijos
- Dedikuotos komandos

Internetinis marketingas
- SEO / SEM
- Naujienlaiškiai
- SMS Reklama

Projektai

KAYAK – kelionių paieškos sistema

Didžiausia ir populiariausia pasaulyje kelionių paieškos sistema, kurios didelę dalį technologijų plėtoja NFQ grupė.

Detalus projekto aprašymas | Internete http://www.kayak.com

Kompanija

Sėkmingą verslą pirmiausia kuria ne technologijos, o ž

Didelę patirtį sukaupę talentingi

> “We want to prevent a deadlock where the team is forced to switch the technology in the middle of the project because it doesn’t meet the requirements any longer”

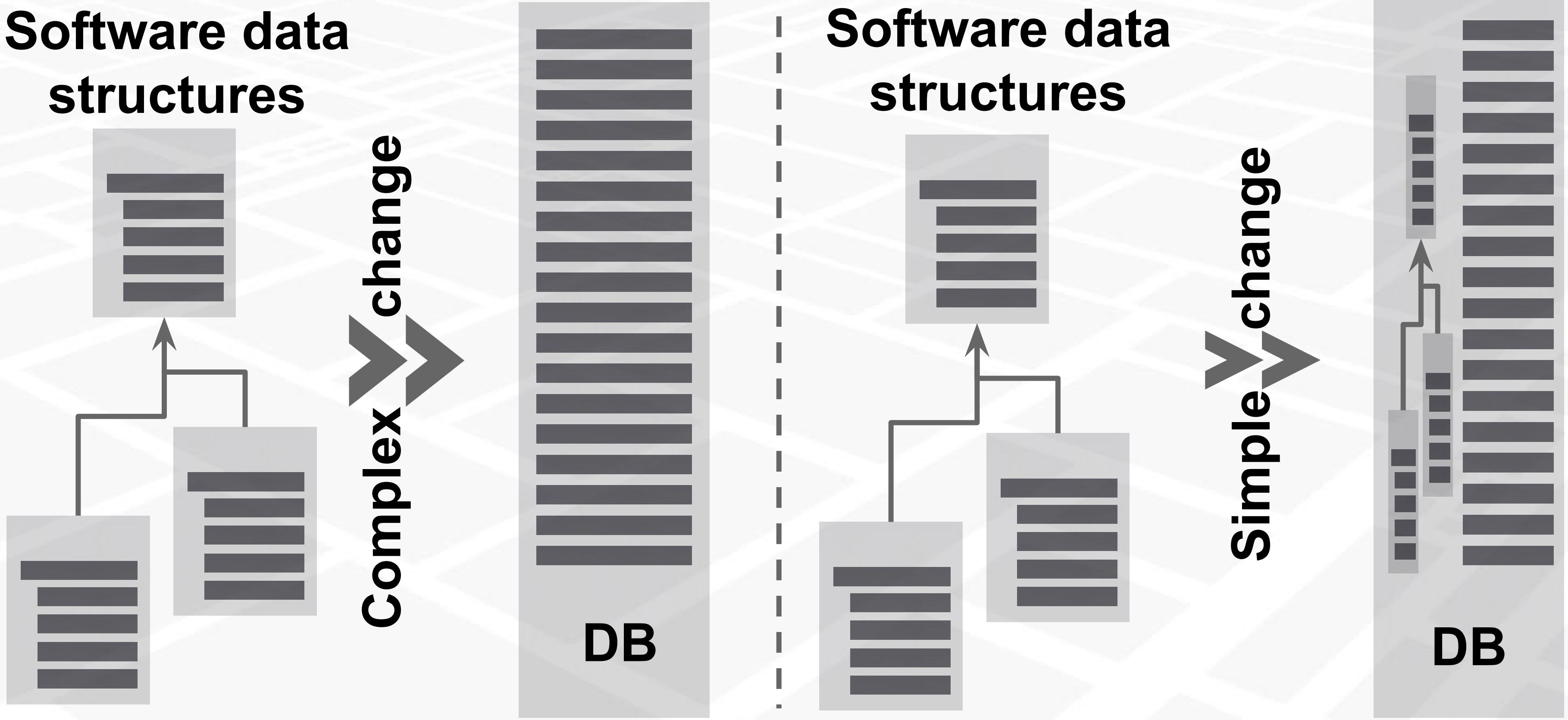
- Martin Schönert and Frank Celler (ArangoDB)

- **Fulltext** - Search field could be optimised
- **Graph** - Services and Projects have bidirectional relation
- **Key-value** - SEO urls need high performance
- **Table** - Category has predefined structure, multiple fields
- **Nested** - Structure of projects differs (Facebook link, services, client, multiple paragraphs)

WHY
WHAT
HOW

ORM vs MultiModel

Software structure for easy use, database structure for fast read. ORM work on DB side



Software data structures

Complex change

DB

Software data structures

Simple change

DB

WHY
WHAT
HOW

Multiple models

Graph	Tables	Events	Object	Key-value
Relations in HDD	Query language for JOINS	Append-only, GC later	Schema for each document	Id/hash calculation per cluster
Fast jumping between records	Server-side optimised scan	Faster write on SSD	Faster development or migration	Async writes
Gremlin wrappers	SQL, JSONiq like syntax	MVCC	Shapes Schema-hybrid	Mapreduce per clusters

Limitation for Multi-Model

ACID	Distributed	Full text
Transaction must fit in memory	Low stability Depending on other tools	Depending on other tokenizers
MVCC Multiversion Concurrency Control	**Hazelcast** **Etcd**	**LibCU** **Lucene**

DEMO

1. PHP clients
2. Multimodel: OrientDB, ArangoDB
3. Async MVCC: @id, _id, @version, _rev

https://github.com/aurelijusb/demo-multimodel-databases

Real world query examples:
https://github.com/Auginte/zooming-based-organizer/blob/master/auginte-distribution/src/main/scala/com/auginte/distribution/orientdb/ReferWrapper.scala#L60

Conclusion

Single-model	Solving common problem Data structures are stable
Multi-Model	Exploring new markets Relations intensive data
Many (CQRS)	High load or big data Dedicated SysOps / Cloud

Questions?

WHY	Alternatives, problems
WHAT	Definition, basic usage
HOW	Internals, pros & cons, tips

Feedback is always welcome:

https://docs.google.com/forms/d/1qLHPlA4GlZSI5MuBEyFhBMQiTMn4_RtlJ89oMbyDrBg/viewform

References and useful links

- https://en.wikipedia.org/wiki/Multi-model_database
- http://orientdb.com/orientdb/
- https://www.arangodb.com/
-
- http://www.odbms.org/blog/2013/10/on-multi-model-databases-interview-with-martin-schonert-and-frank-celler/
- https://www.arangodb.com/key-features/
- https://lostechies.com/jimmybogard/2013/06/06/acid-2-0-in-action/
- http://www.slideshare.net/arangodb/multi-modeldatabases-41917934
- http://www.slideshare.net/LuigiDellAquila/orientdb-time-representation
- https://youtu.be/JHGkaShoyNs?t=57m7s
- https://en.wikipedia.org/wiki/Entity%E2%80%93attribute%E2%80%93value_model
- http://www.infoworld.com/article/2861579/database/the-rise-of-the-multimodel-database.html
- http://www.jamesserra.com/archive/2015/07/what-is-polyglot-persistence/
- http://de.slideshare.net/MichaelHackstein/multi-modeldatabases
- http://aws.amazon.com/about-aws/whats-new/2015/08/amazon-dynamodb-titan-graph-database-integration/
- https://mesosphere.com/blog/2015/11/30/arangodb-benchmark-dcos/