# GRAPH DATABASE NEO4J

Maria Grigorieva for the DKB Project
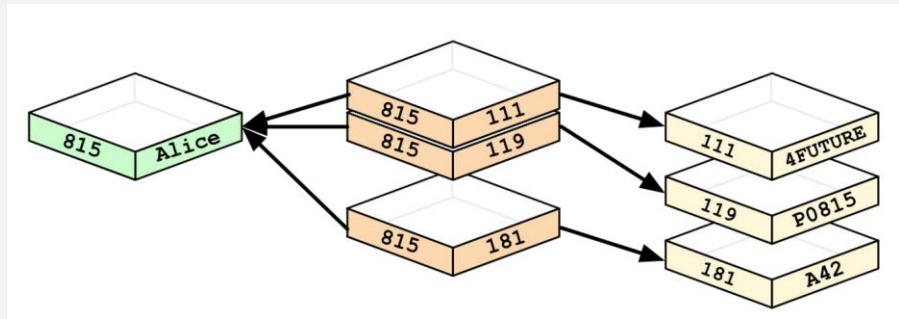
07 Sept 2017

# FOCUS ON DIFFERENT DATA MODELS



Scale To Size

Key/value Stores

Column Stores

Document Database

Graph Database

Typical Relational Database

RDBMS Performance Line

Scale To Complexity

# RELATIONAL DATABASE

## SQL Query with JOINs



Persons — Dept_Members — Department
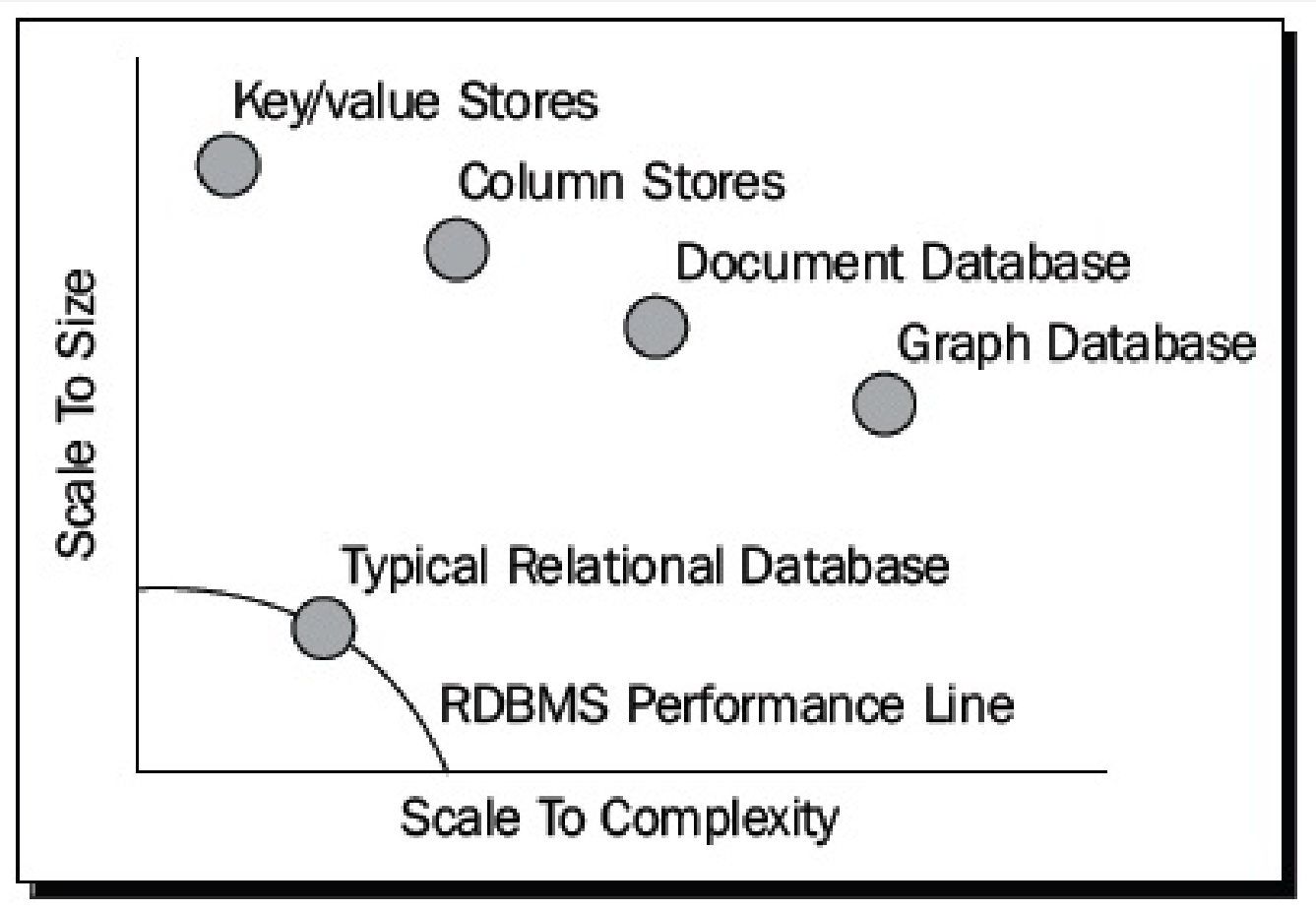
# GRAPH DATABASES



**Relational Versus Graph Models**  neo4j

**Relational Model**          **Graph Model**
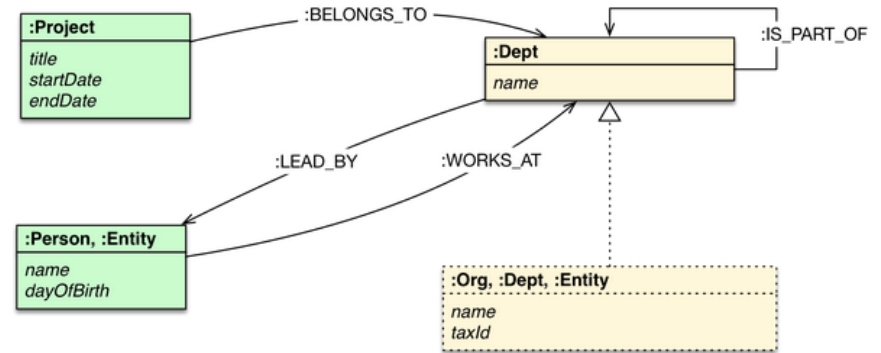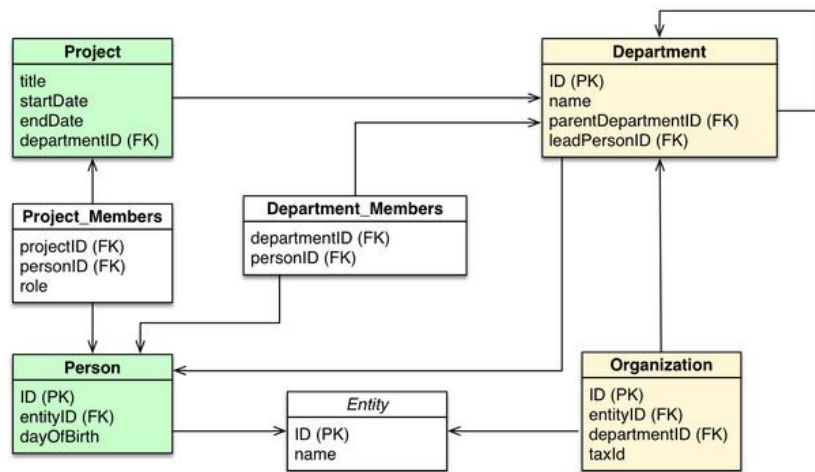
# RELATIONAL MODEL VS GRAPH

# NEO4J SPEED

- EBAY

- "We found that Neo4j to be literally 1000s of times faster than our prior MySQL solution, with queries that require 10-100 times less code. Today, Neo4j provides eBay with functionality that was previously impossible"

http://info.neo4j.com/rs/neotechnology/images/neo4j-casestudy-ebay.pdf?_ga=2.98504223.122420050.1504721209-2003509311.1496440766
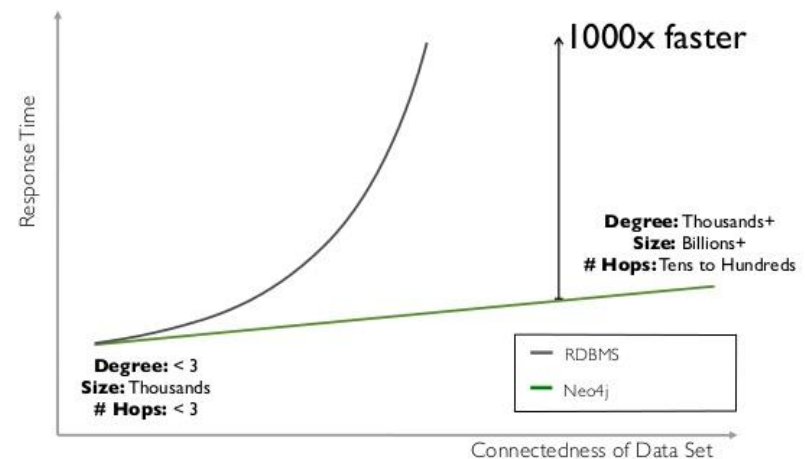
# GRAPH DATABASES CHARACTERISTICS

- Intuitiveness
  - Possibility to quickly and easily understand the data

- Speed
  - Development process
  - Speed of querying data

- Agility = A Naturally Adaptive Model + Query Language Designed for Connectness



Connected Data & Query Performance — neo technology

RDBMS vs. Native Graph Database

1000x faster

Degree: Thousands+
Size: Billions+
# Hops: Tens to Hundreds

Degree: < 3
Size: Thousands
# Hops: < 3

RDBMS
Neo4j

Response Time

Connectedness of Data Set

Neo Technology, Inc Confidential

# NEO4J USE-CASES

- Real Time Recommendations
- Master Data Management
- Fraud Detection
- Graph Based Search
- Network & IT Operations
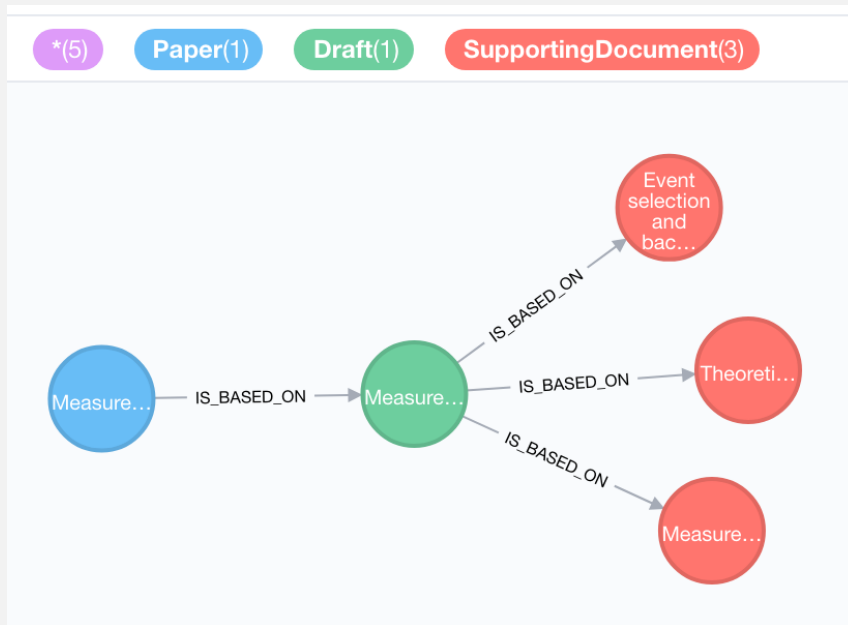- Identity & Access Management
- Social Networks
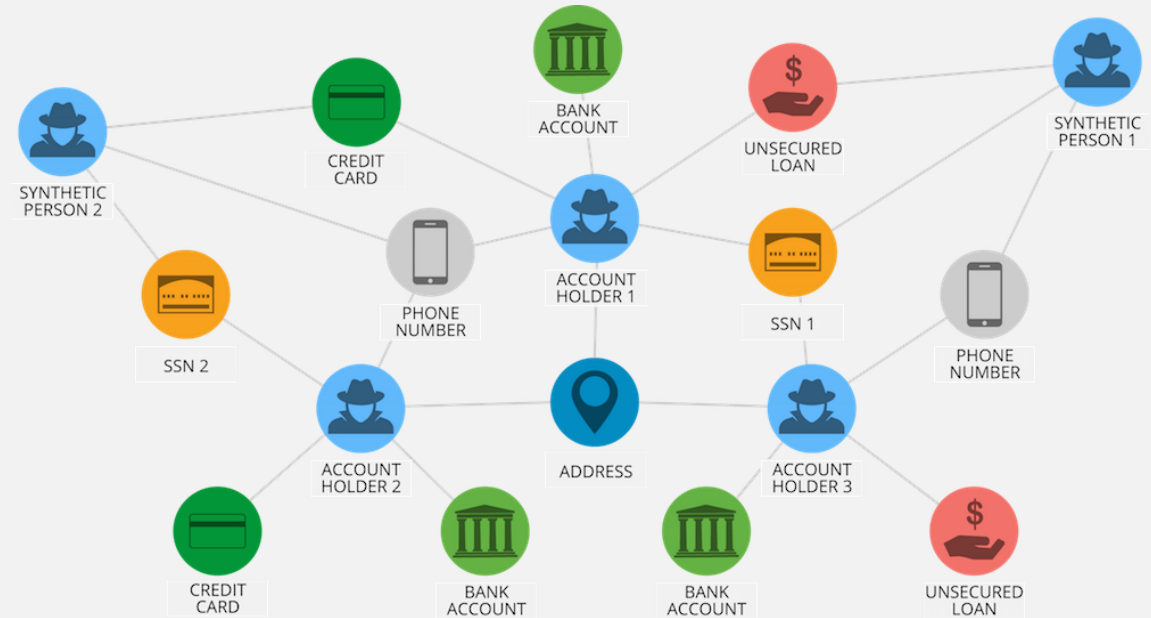
# CERN DOCUMENT SERVER PUBLICATIONS HISTORY

- ## Simple model:

  - (paper) - [:IS_BASED_ON] -> (draft) - [:IS_BASED_ON] -> (supporting note)



## Cyper model:

# CDS INVENIO

- CDS Invenio is available under GPL http://cdsware.cern.ch/
  - Free download and usage
  - Instances across the globe: sciences & humanities
    - Administrative documents, librettos, an art collection
- Standards
  - MARC21 metadata format http://www.loc.gov/
  - Multi-lingual; UNICODE
  - Compliancy with all browsers; web standards
- Flexibility
  - Format support
    - SiteMap for GoogleWeb, similar for GoogleScholar
    - Export as: BibTeX, MARC, MARCXML, DC, EndNote, NLM
    - Subscribe to: RSS (& email alerts) –
  - Available in 20 languages (external contributions)

## MARC

```
002210393 001__ 2210393
002210393 005__ 20170412182032.0
002210393 037__ $$aATL-COM-PHYS-2016-1194
002210393 041__ $$aeng
002210393 088__ $$9CERN-ATL-COM-PHYS-2016-1194
002210393 100__ $$0AUTHOR|(SzGeCERN)683170$$0AUTHOR|(INSPIRE)INSPIRE-00219079$$aLi, Haifeng$$gStatistics,
002210393 245__ $$aSupporting note: Search for the dimuon decay of the Higgs boson in $pp$ collisions at $
002210393 260__ $$c2016
002210393 269__ $$aGeneva$$bCERN$$c25 Aug 2016
002210393 300__ $$amult. p
002210393 520__ $$aA search for the dimuon decay of the Higgs boson has been performed using data correspo
002210393 595__ $$aCERN CDS-Invenio WebSubmit
002210393 6531_ $$9CERN$$aHIGGS
002210393 65017 $$2SzGeCERN$$aParticle Physics - Experiment
002210393 65027 $$2SzGeCERN$$aPhysics
002210393 690C_ $$aCERN
002210393 690C_ $$aINTNOTE
002210393 690C_ $$aPRIVATLAS
002210393 693__ $$aCERN LHC$$eATLAS
002210393 700__ $$0AUTHOR|(SzGeCERN)751740$$0AUTHOR|(INSPIRE)INSPIRE-00436733$$aLiu, Yanlin$$gOptimization
002210393 700__ $$0AUTHOR|(SzGeCERN)446640$$0AUTHOR|(INSPIRE)INSPIRE-00132531$$aTsybychev, Dmitri$$gSuperv
002210393 700__ $$0AUTHOR|(SzGeCERN)764063$$aFray, Antony$$gCutflow, spurious signal systematics$$ma.fray@
002210393 700__ $$0AUTHOR|(SzGeCERN)708109$$0AUTHOR|(INSPIRE)INSPIRE-00376466$$aHoenig, Friedrich$$gMuon f
002210393 700__ $$0AUTHOR|(SzGeCERN)414244$$0AUTHOR|(INSPIRE)INSPIRE-00300183$$aMehta, Andrew$$gVH,h->mumu
002210393 700__ $$0AUTHOR|(SzGeCERN)670199$$0AUTHOR|(INSPIRE)INSPIRE-00120467$$aRizvi, Eram$$gSupervisor$$
002210393 700__ $$0AUTHOR|(SzGeCERN)666716$$0AUTHOR|(INSPIRE)INSPIRE-00185894$$aThompson, Paul$$gSignal th
002210393 700__ $$0AUTHOR|(SzGeCERN)515237$$0AUTHOR|(INSPIRE)INSPIRE-00226382$$aZhao, Zhengguo$$gSupervis
002210393 700__ $$0AUTHOR|(SzGeCERN)401247$$0AUTHOR|(INSPIRE)INSPIRE-00138410$$aZhou, Bing$$gSupervisor$$m
002210393 700__ $$0AUTHOR|(SzGeCERN)384990$$0AUTHOR|(INSPIRE)INSPIRE-00212987$$aDai, Tiesheng$$gPrivate MC
002210393 700__ $$0AUTHOR|(INSPIRE)INSPIRE-00237882$$0AUTHOR|(SzGeCERN)698137$$aWu, Yusheng$$gAnalysis, sp
```

| METADATA CONCEPT | PROPOSED MARC 21 REPRESENTATION |
|---|---|
| Abstract | 520 $a |
| Author, first | 100 $a |
| Author(s), additional | 700 $a |
| Collection identifier | 980 $a |
| Email | 8560 $f |
| Imprint | 260 $a,b,c; 300 $a |
| Keywords | 6531 $a |
| Language | 041 $a |
| OAI identifier | 909CO $o |
| Publication info | 909C4 $* [many subfields] |
| References | 999C5 $* [many subfields] |
| Primary report number | 037 $a [unique throughout the system!] |
| Additional report number(s) | 088 $a |
| Series | 490 $a,v |
| Subject | 65017 $a |
| Title | 245 $a |
| URL (e.g. to fulltext) | 8564 $u, $z |

# CDS TO NEO4J DATAFLOW



CDS Papers [MARC]

01.01.2016 – now

CDS Papers [CYPER]

(papers)

CDS Drafts [JSON]

CDS Drafts [CYPER]

(papers)-[:IS_BASED_ON]->(drafts)-[:IS_BASED_ON]->(supporting documents)

CDS Supporting Documents [CYPER]

CDS Supporting Note [JSON]

Extract Support Docs ID from 1st page content

(papers)-[:IS_BASED_ON]->(drafts)