



Ariadne: a Tracking System for Relationships in LHCb Metadata

Illya Shapoval^{1,2,3,4}, Marco Clemencic¹, Marco Cattaneo¹

¹ CERN, ² KIPT, ³ UNIFE, ⁴ INFN-FE

20th International Conference on Computing in High Energy and Nuclear Physics

14-18 October 2013

Amsterdam

Content

- Objectives
- Approach
- Choice of DBMS
- Ariadne: design
- Interaction with the system

Introduction

- LHCb data processing implies handling of heterogeneous metadata entities
 - Versions of applications (in 2 dimensions)
 - Conditions Database states (in 2x3 dimensions)
 - Real data reconstruction types
 - MC data simulation types
 - Many others: trigger configurations, arch. specifiers, etc.

Objectives: how did it start

What entities are compatible with a **CondDB state**?

Is a **CondDB state** consistent?

What entities are compatible with an **application**?

What entities are compatible with a data **processing type**?

What entities are **compatible**?

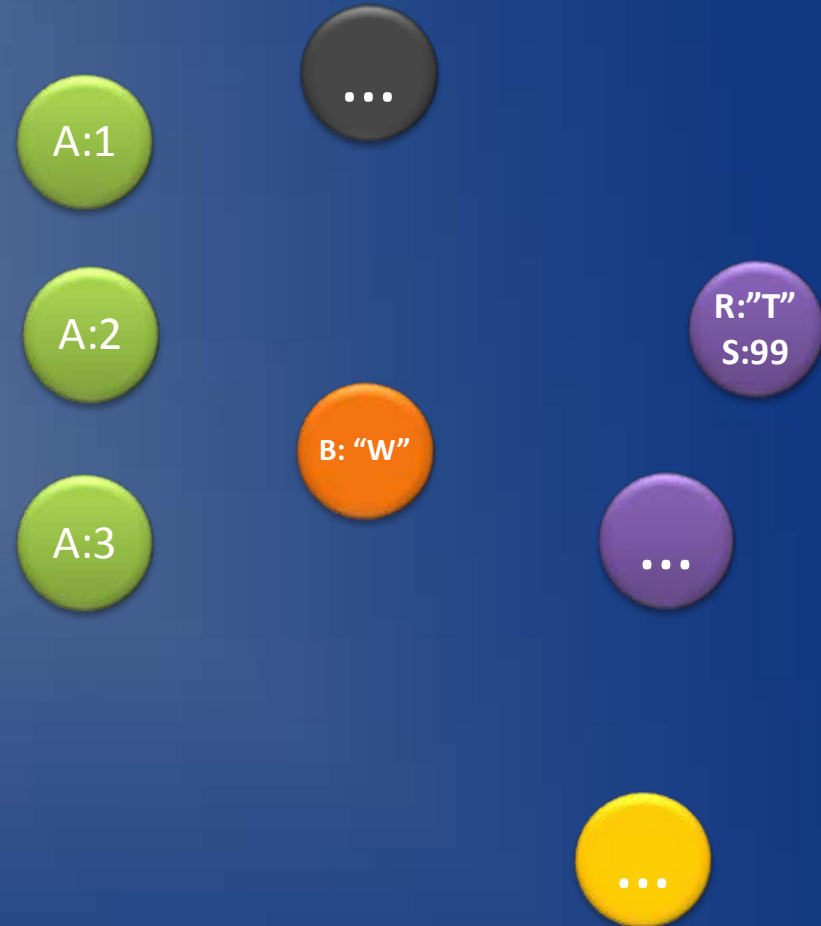
How are entities **related**?

Requirements

- An operational space with generic way of
 - Expressing relationship constraints
 - Tracking relationships
 - Extracting solutions
- Ease of data management
- Flexibility (to extend the area of application)

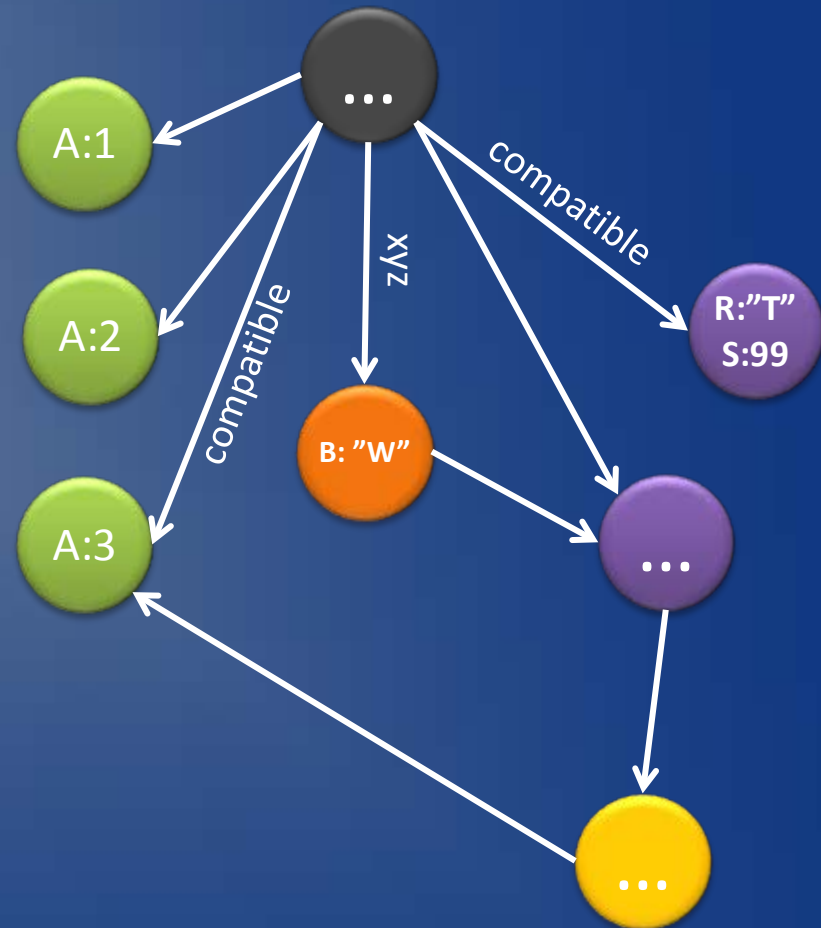
Approach: property graphs

- Modeling structured metadata
 - as nodes with attributes (key+value)



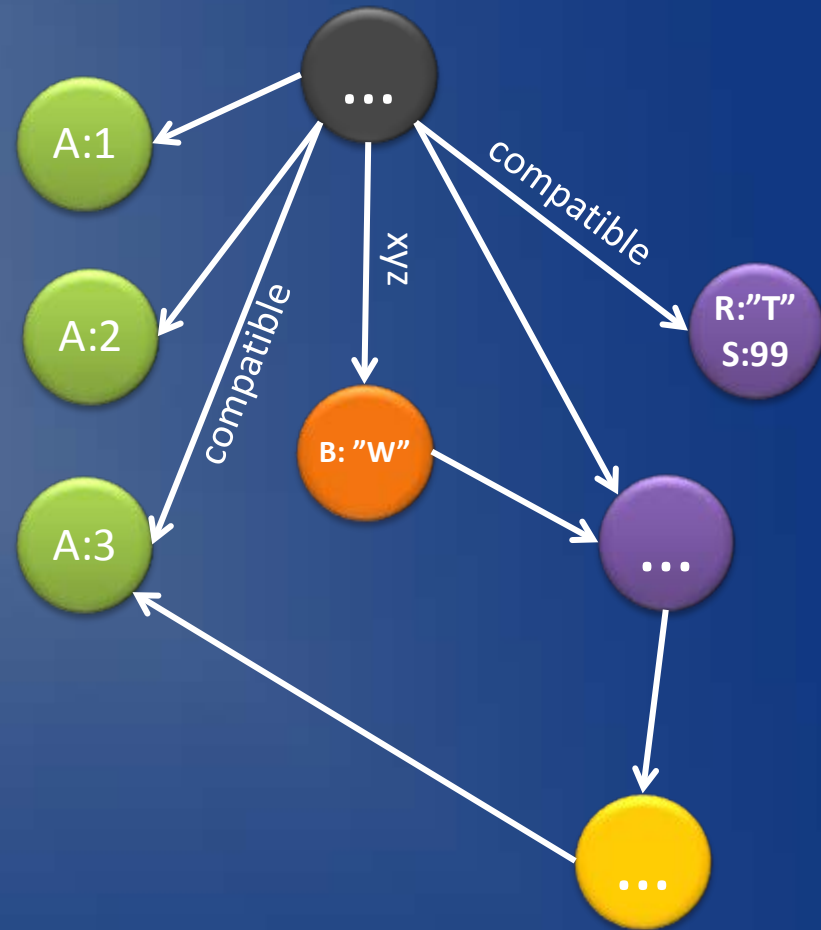
Approach: property graphs

- Modeling structured metadata
 - as nodes with attributes (key+value)
 - with typed connections



Approach: property graphs

- Modeling structured metadata
 - as nodes with attributes (key+value)
 - with typed connections
- Tracking?
- Extracting?



Choice of DBMS model

Relevant characteristics	Relational solution	Graph solution
Object-relational impedance mismatch problem ^[1]	^[2] Suffers of	^[2] Free of
Flexibility of schema	No	Yes
Performance of structural queries	Poor ^[3]	Better ^[3]
Scaling to data complexity	Poor ^[4]	Better ^[4]
Ease of data management	Doable	Better

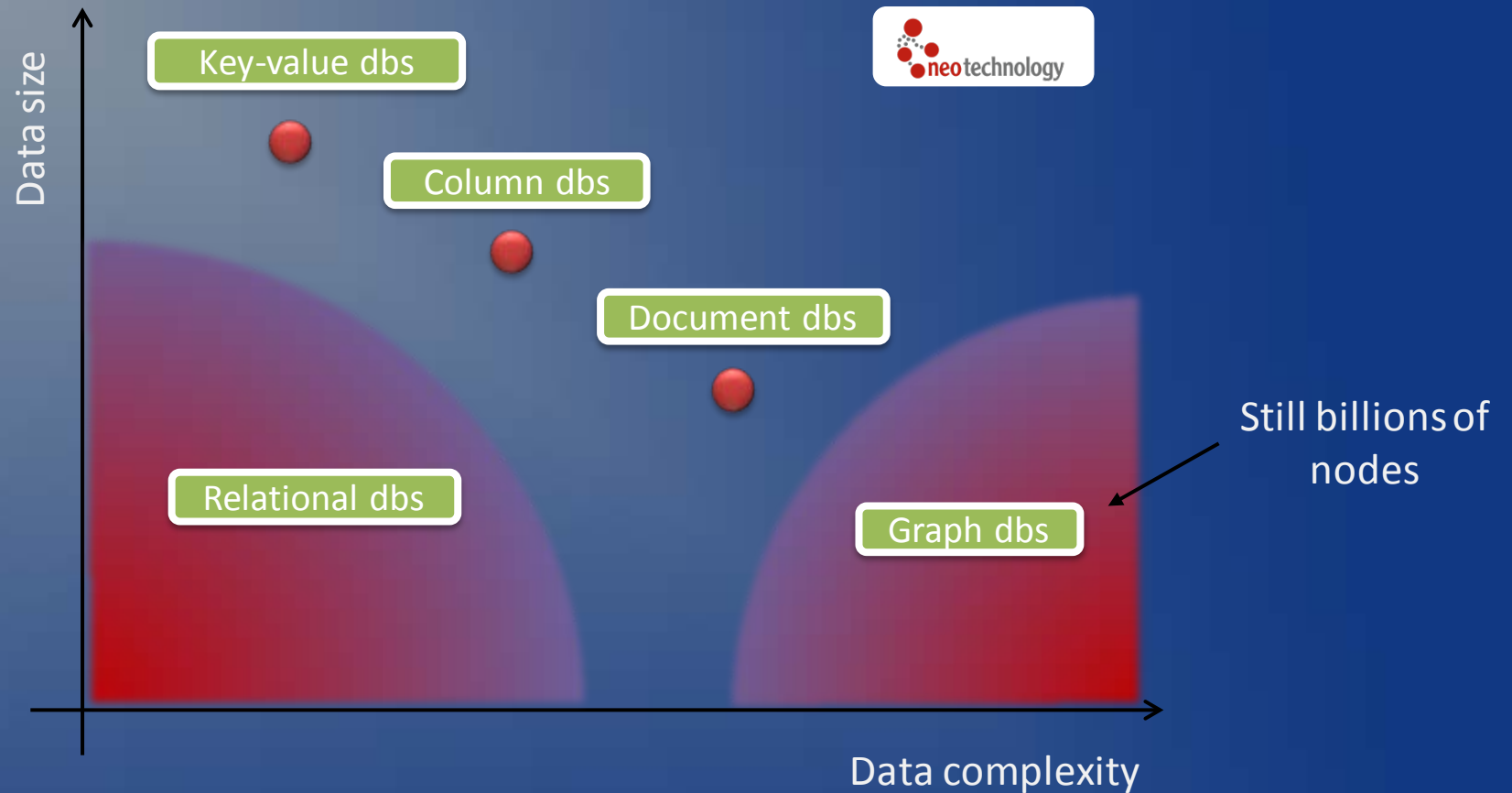
[1] C. Ireland et al, A Classification of Object-Relational Impedance Mismatch, DBKDA '09.

[2] Correlated with the "Performance of structural queries" row of the table

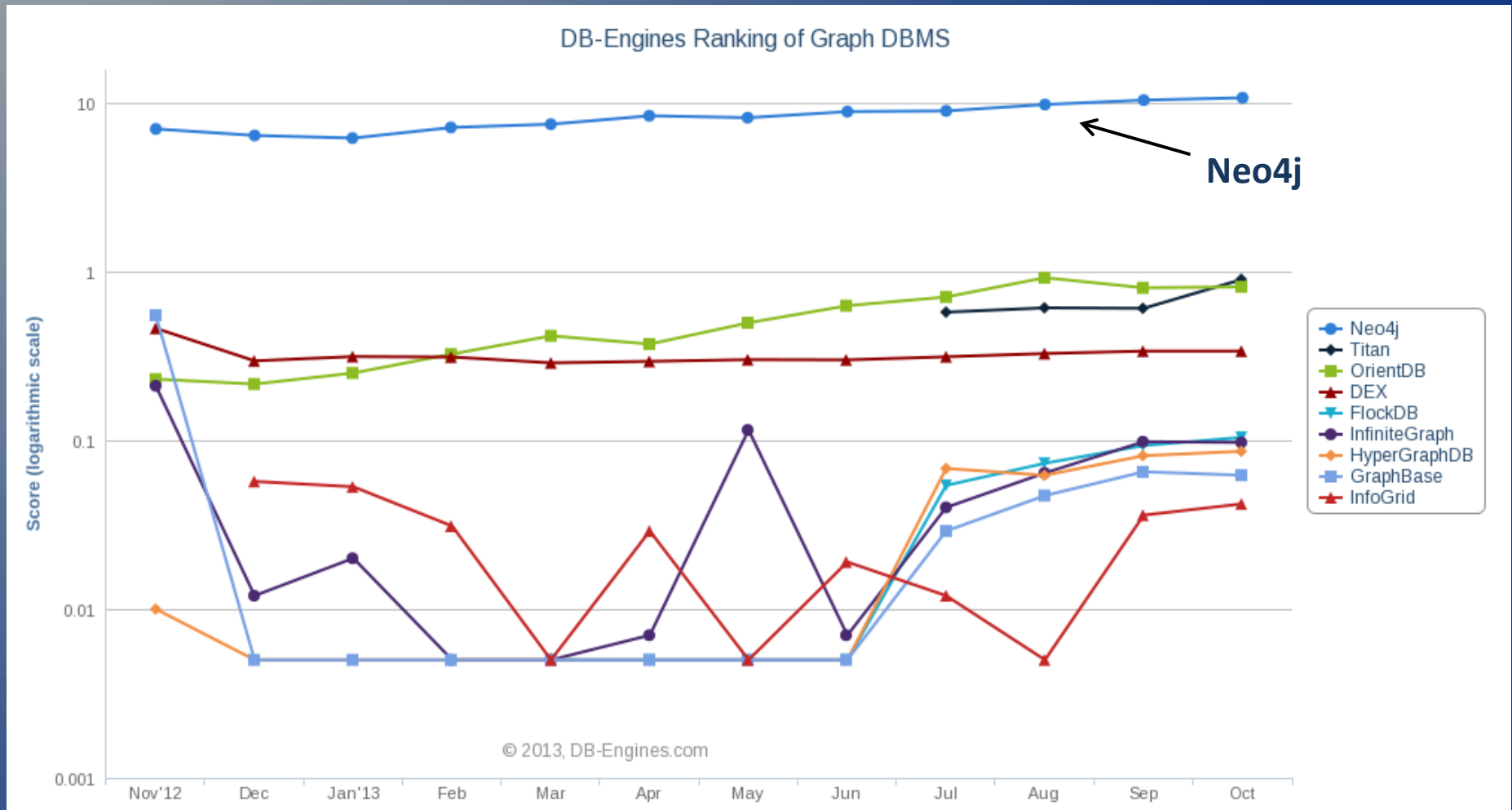
[3] C. Vicknair et al, A Comparison of a Graph Database and a Relational Database, ACM SE '10, NY.

[4] See the next slide

Other NoSQL DBMS models



Choice of graph DBMS



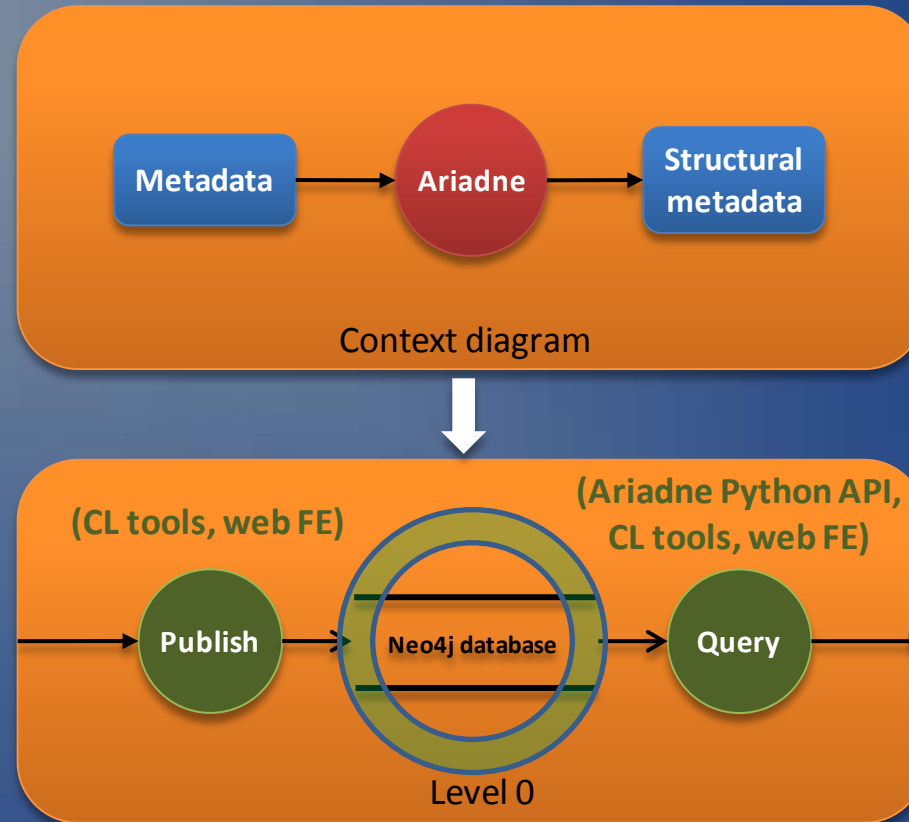
Taken from the "Knowledge Base of Relational and NoSQL DBMS" at:
<http://db-engines.com/en/ranking/graph+dbms>

Neo4j customers

 Adobe	 aikux.com	 Global 500 Telcommunication		 Global 500 Manufacturing	 CareerArcGroup
 careerbuilder.com	 CHIP ONLINE	 CISCO	 classmates-com	 compete	 Curaspan HEALTH GROUP
 Global 500 Logist cs		 splink	 DRAKER MONITOR. MANAGE. CONTROL.	 DRW TRADING GROUP	 Dshini
	 ePals	 EQUILAR	 Era7 bioinformatics		 FOVEA
 <fuseworks/>	 gamesys	 gen	 glassdoor	 Glowbl	 HealthUnlocked
 Hedge	 hp	 HUAWEI	 Humanwest.co	 ICE	 entropy
 ID MISSION	 indiatimes	 InfoJobs	 intuit	 isar software	 janssen
 Juni sphere	 Justdial	 KeyLines NETWORK INSIGHT	 kitedesk	 LAUREATE EDUCATION, INC.	 LIFECHURCH.TV
 LifeWay	 Live@Play	 Global 500 Media	 LOCKHEED MARTIN	 Global 500 Transportation	 maaii Let's connect
 megree	 migRaven	 minettabrook	 NetApp	 NexLP	 OPEN Tree of Life
 Perigee	 Pitney Bowes	 Global 500 Manufacturing	 Pro Integra S.L.	 research now	 Global 500 Retail
 SERENA	 SevenBridges	 SFR	 Share Practice	 SRM SOCIETY FOR HUMAN RESOURCE MANAGEMENT	 shuti
 SODIFRANCE	 springcm	 SQUIDOO	 SUPERSERIOUS	 teachscape	 telenor

Ariadne: system design











A tracking system for relationships in LHCb metadata



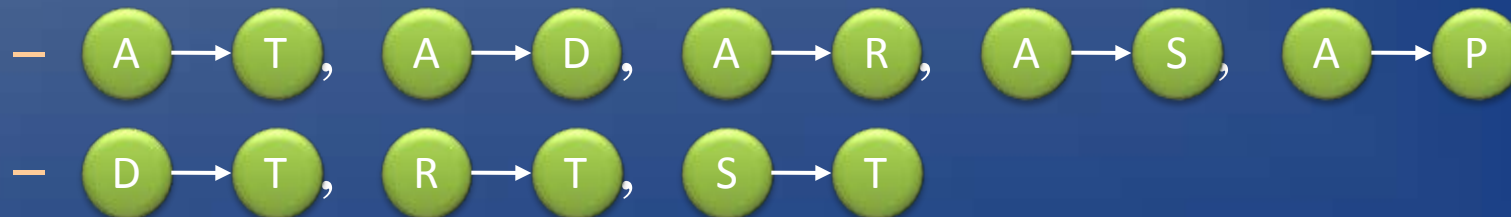
(Leveled data flow diagram, shown in the Yourdon-DeMarco DFD notation)

Ariadne: current knowledge graph

- Metadata entities that current graph contains (>500)

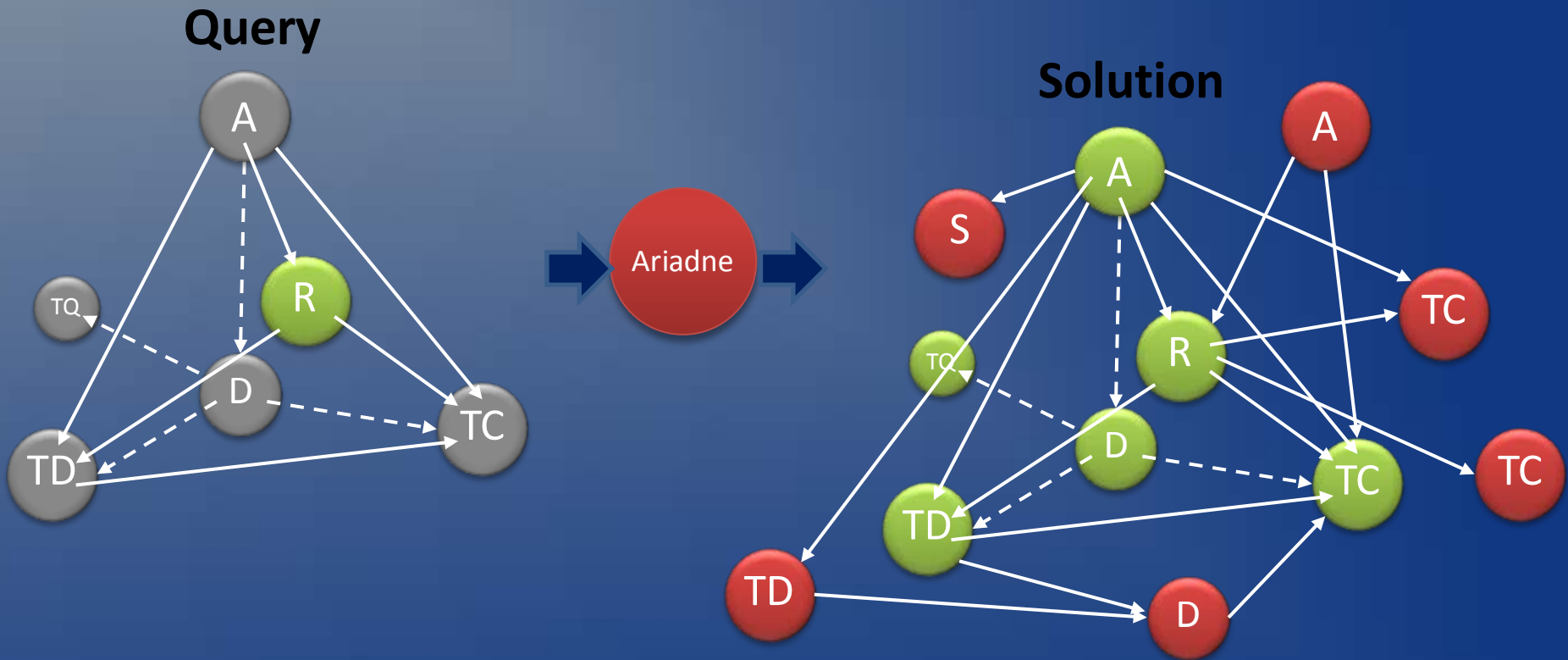
- Applications 
- CondDB tags  (or ,  and  to specify partitions)
- DetectorTypes(\leftrightarrow DataTypes) , RecoTypes , SimTypes 
- Platforms , GRID sites 

- Relationships between those entities (~50k)



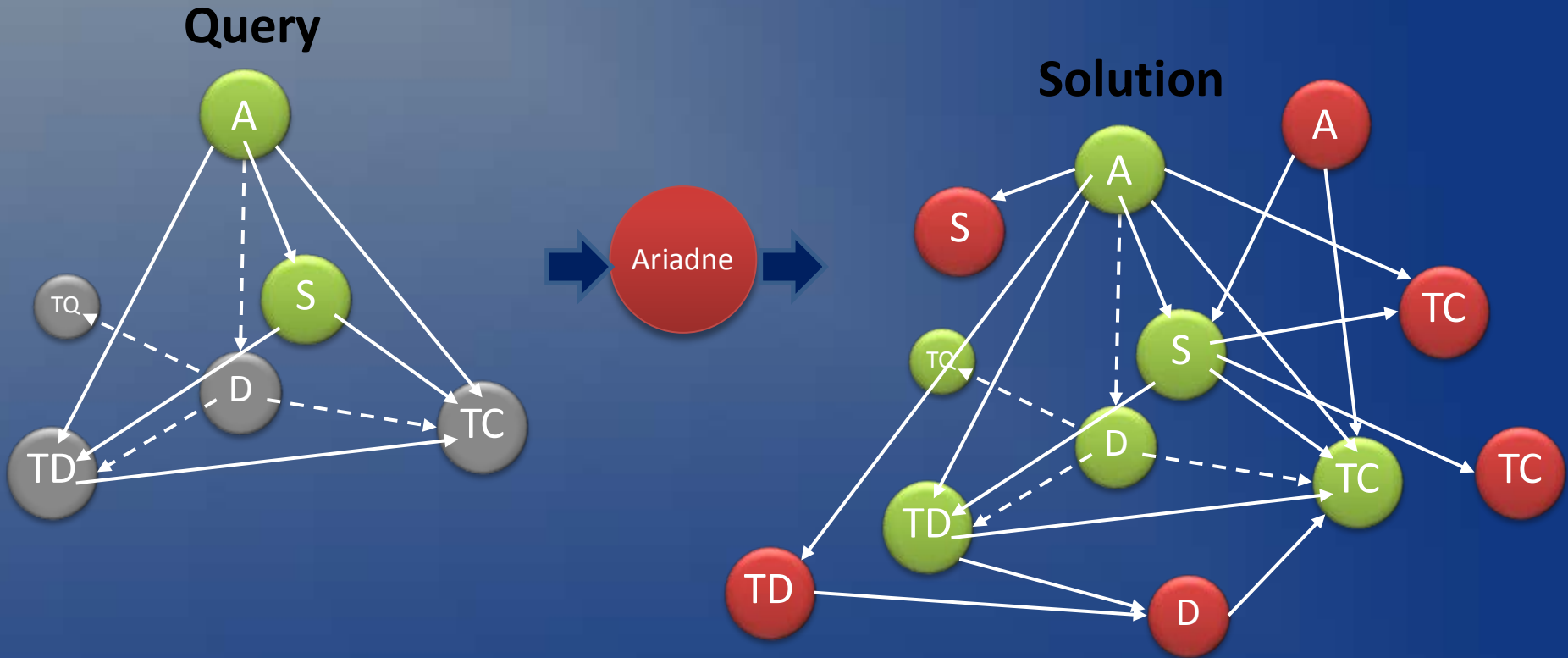
Tracking Relationships: Matching Patterns (1)

- What is the full compatible set of entities for concrete **real data processing type**?



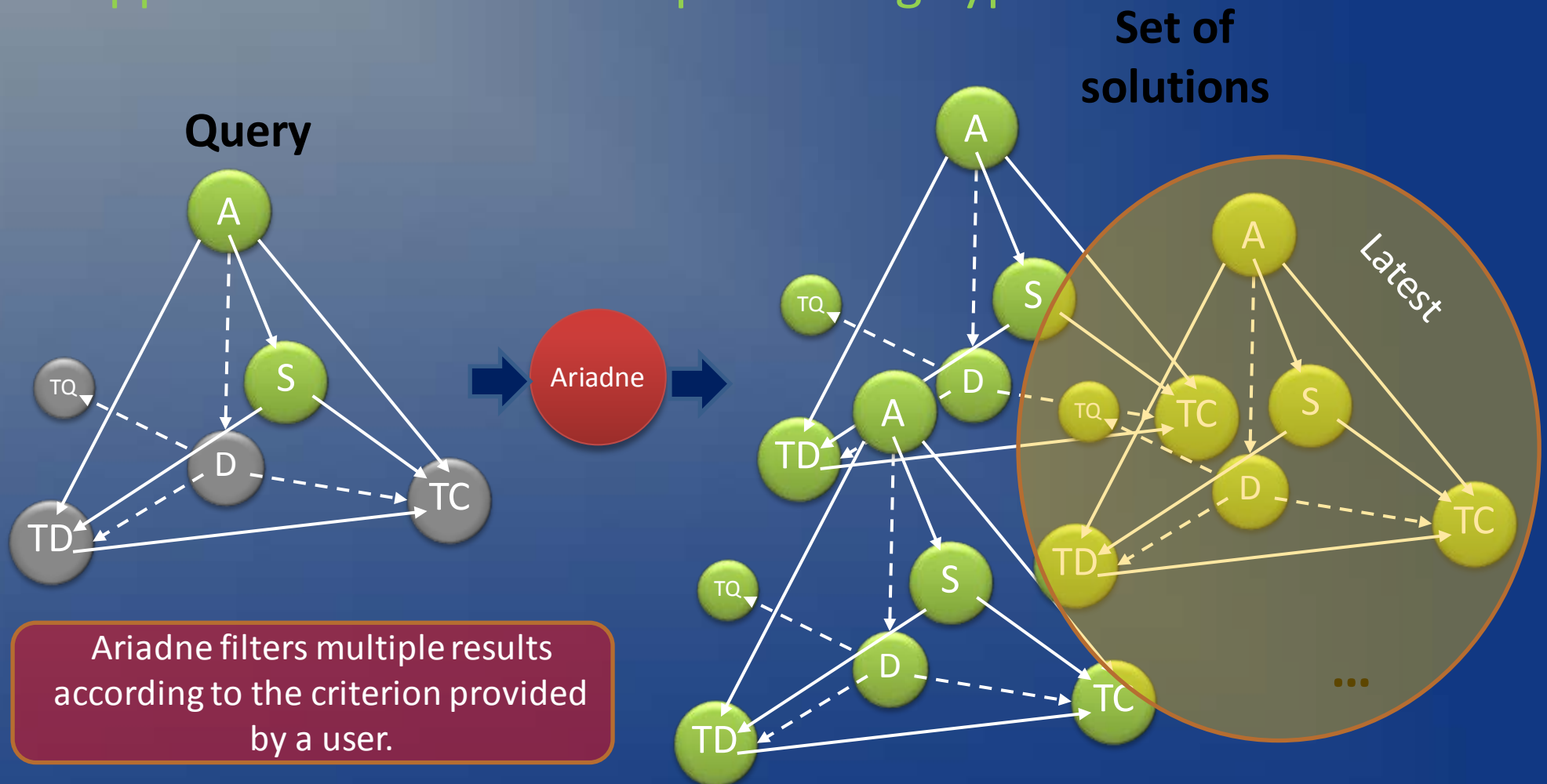
Tracking Relationships: Matching Patterns (2)

- What is the full compatible set of entities for concrete application and MC data processing type?



Tracking Relationships: filtering multiple solutions (3)

- What is the full compatible set of entities for concrete application and MC data processing type?



Other application domains

What entities are compatible with a **CondDB state**?

Is a **CondDB state** consistent?

What entities are compatible with an **application**?

What entities are compatible with a data **processing type**?

What does an application **depend on**?

Where an entity is **deployed to**?

What platform an application was **built for**?

...?

What entities are compatible?

How are entities **related**?

Summary

- Ariadne – a **generic** tracking system for **relationships** in LHCb metadata
 - Provides **generic** UI layer for **heterogeneous** metadata;
 - Based on the **novel** Neo4j **graph** database;
 - Provides powerful **expressiveness** when dealing with **complex data** (lots of relationships);
 - **Scalable** and **high performant** solution for complex data.

BACKUP



Ariadne: system requirements

(narrows down to the Neo4j's requirements)

	Minimum	Recommended	Actual
CPU	Intel Core i3	Intel Core i7	Intel(R) Xeon(R) CPU L5640 @ 2.27GHz
Memory	2GB	16—32GB or more	48GB
Disk	SATA	SSD w/ SATA III	SATA II
Filesystem	ext4 (or similar)	ext4, ZFS	ext4
Software	Oracle Java 7, OpenJDK 7	Oracle Java 7	Oracle Java 7

Compatibility of Entities: Definition

- Two entities are declared compatible if a job, that uses them simultaneously:
 - A. **does not crash** because of the combination
 - B. **is configured** by the combination **to work exactly in the way a developer anticipated** it.

Implication:

- A set of entities is declared compatible if and only if each pair of the entities (the compatibility is tracked between) out of the set is compatible.

Ariadne: interaction

- Ariadne management
 - Admin command line tools
 - Visual interaction with the Ariadne graph
- 'DDDDBConf' already got an Ariadne-based feature (>=LHCB_v35r6)

```
from Configurables import DDDDBConf  
DDDDBConf( DataType = "2012", AutoTags = True )
```

- Advanced interaction for users through the XMLRPC endpoint:

```
from xmlrpclib import ServerProxy  
proxy = ServerProxy("http://ariadne-lhcb.cern.ch/xmlrpc/", allow_none = True)  
response = proxy.queryAriadne(question = {...}, criterion = '...')
```

Ariadne: interaction (2)

```
critterion = 'latest_LHCBCOND_DDDB'
```

```
question = { 'application': 'Brunel',  
             'application_version': 'v43r2p2',  
             'detector_type': '2012',  
             'reco_type': 'Reco14',  
             'DDDBTag': None,  
             'LHCBCONDTAG': None,  
             'DQFLAGSTag': None }
```



```
response = { 'pathExists': True,  
             'application': 'Brunel',  
             'application_version': 'v43r2p2',  
             'detector_type': '2012',  
             'reco_type': 'Reco14',  
             'DDDBTag': 'dddb-20120831',  
             'LHCBCONDTAG': 'cond-20130114',  
             'DQFLAGSTag': 'dq-20121016' }
```

```
question = { 'application': 'Brunel', 'application_version': 'v43r2p2', 'detector_type': '2012', 'reco_type': 'Reco14',  
             'DDDBTag': 'dddb-20120831', 'LHCBCONDTAG': 'cond-20130114', 'DQFLAGSTag': 'dq-20121016' }
```

```
question = { 'detector_type': '2012', 'reco_type': 'Reco14', 'DDDBTag': None, 'LHCBCONDTAG': None, 'DQFLAGSTag': None }
```

```
question = { 'detector_type': '2012', 'DDDBTag': None, 'LHCBCONDTAG': None, 'DQFLAGSTag': None }
```

```
question = { 'reco_type': 'Reco14', 'DDDBTag': None, 'LHCBCONDTAG': None }
```