GRAPH DATABASE NEO4J

Maria Grigorieva for the DKB Project

07 Sept 2017

FOCUS ON DIFFERENT DATA MODELS



RELATIONAL DATABASE



SQL Query with JOINs

LAS T

UNION

FROM

(SELECT TdiredReportees AS diredReportees, sum(T.count) AS count FROM (

PROBE SELECT managerpid AS directReportees, 0 AS count FROM person_reportee manager WHERE managerpid = (SELECT dI FROM person WHERE name = "IName IName") UNION

SELECT manager.pidAS directReportees, count/manager.directly_manages)AS count FROM person reportee manager WHERE manager.pid = (SELECT id FROM person WHERE name = "fName IName") GROUP BY directReportees

UNION SELECT manager.pid.AS directReportees, count (reportee directly manages) AS count

FROM person reportee manager JOIN person reportee reporte e ON manager directly_manages = reportee.pid

WHERE manager pid = (SELECT id FROM person WHERE name = "fName IName") GROUP BY diredR eponees UNION

SELECT manager.pid AS directReportees, count (L2Reportees directly_manages) AS count

FROM person_reportee manager JOIN person_reportee L1 Reportee s ON manager directly_manages = L1Reportees.pld JOIN person reportee L2 Reportees

ON LTRe potees.directly_manage s = L2Reportees.pid WHERE manager.pid = (SELECT id FROM person WHERE name = "fName IName")

GROUP BY directReportees

) AS T GROUP BY diredReportees)

UNION (SELECT T.dire.dReportees AS dire.dReportees, sum(T.count) AS count FROM (

SELECT manager.directly_manages AS directReportees, 0 AS count

FROM person reportee manager WHERE manager.pid = (SELECT id FROM person WHERE name = "(Name IName") UNION SELECT reporte e.pid AS directReportees, count(reportee directly manages) AS count

FROM person_reportee manager

JOIN person reportee reporte e

ON manager dredty manages = reportee.pid WHERE manager.pid = (SELECT id FROM person WHERE name = "fName IName")

GROUP BY directReportees UNION

(continued from previous page...) SELECT depth1 Reporte es pid AS diredReportees. count/depth2 Reporte es.directly_m anages) AS count FROM person_reportee manager JOIN person reportee L1Reportees ON manager.directly_manages = L1Reportees.pid JOIN person reportee L2R eportees ON L1Reportees.directly_manages = L2R eportees.pld WHERE manager.pld = (SELECT Id FROM person WHERE name = "Name IName") GROUP BY directReportees GROUP BY directReportees) (SELECT T.directReportees AS directReportees, sum (T.count) AS count SELECT reportee. directly_manages AS directReportees, 0 AS count FROM person reportee manager JOIN person_reporter reporter ON manager directly_manages = WHERE manager, kie (SELECT id FROM person WHERE name = "IName IName")

GROUP BY directReportees UNION SELECT L2Reportees.pidAS directReportees, count)L2Reportees.directly_mana.ges)AS

toun! FROM person_reportee manager JOIN person reportee L1R eportees ON manager directly_manages = L1Reportees pid JOIN person reportee L2R eportees ON LTReportees.directly_manages = L2Reportees.pid WHERE manager.pid = (SELECT Id FROM person WHERE name = "Name IName")

GROUP BY directReportees (AS T GROUP BY directReportees)

UNION (SELECT L2Reportees.directly_manages AS directReportees, 0AS count FROM person reportee manager JOIN person_reportee L1Reportees ON manager directly manages = L1Reportees pid ON person reporties L2Reportees ON L1Reporties. directly manages = L2Reportees pid WHERE manager pid = (SELECTI of ROM person WHERE name = "Name IName")

3

GRAPH DATABASES







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



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

Cyper model:

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

*(5)	Paper(1)	Draft(1)	SupportingDocument(3)
Measu	re IS_BASED	_ON Measu	re IS_BASED_ON Theoreti
Maria Gri	gorieva		

CREATE (ATLAS_HIGG_2016_25_003:Paper { type: 'paper', title: 'Measurement of inclusive and differential cross sections in the \$H \\rightarrow ZZ** creation_date: '2017-08-09T12:20:05', primary_report_number: 'arXiv:1708.02810', keywords: 'Higgs physics, particle and resonance production, experimental results', number_of_authors: 2876, email_message: 'fmalek@lpsc.in2p3.fr', recid: 2277731}) CREATE (ATLAS_HIGG_2016_25_002:Draft { type: 'draft', title: 'Measurement of inclusive and differential cross sections in the \$H \\rightarrow ZZ** creation_date: '2017-06-14T23:58:08', primary_report_number: 'ATLAS-HIGG-2016-25-002', keywords: 'Higgs physics, particle and resonance production, experimental results', email_message: 'stephane.willocq@cern.ch', recid: 2269060}) CREATE (CERN_ATL_COM_PHYS_2016_1599:SupportingDocument { type: 'internal note', title: 'Event selection and background estimation in the \$H\\rightarrow ZZ^{(*)}\\rightarrow creation_date: '2016-11-09T15:32:09', primary_report_number: 'ATL-COM-PHYS-2016-1599', keywords: 'HIGGS', recid: 2231576}) CREATE (CERN_ATL_COM_PHYS_2016_1605 SupportingDocument { type: 'internal note', title: 'Measurement of the fiducial, differential and total Higgs production cross sections creation_date: '2016-11-09T18:04:02', primary_report_number: 'ATL-COM-PHYS-2016-1605', keywords: 'higgs', email_message: 'eleni.mountricha@cern.ch, R.D.Schaffer@cern.ch, andrea.gabrielli@cern.ch, bij recid: 2231597 CREATE (CERN_ATL_COM_PHYS_2016_1632:SupportingDocument { type: 'internal note', title: 'Theoretical Uncertainties Occurring In \$H\\rightarrow ZZ\$ analyses', creation_date: '2016-11-11T18:16:49', primary_report_number: 'ATL-COM-PHYS-2016-1632', keywords: 'higgs', email_message: 'eleni.mountricha@cern.ch, R.D.Schaffer@cern.ch, eleni.mountricha@cern.ch', recid: 2232040 CREATE (ATLAS_HIGG_2016_25_003)-[:IS_BASED_ON]->(ATLAS_HIGG_2016_25_002), (ATLAS_HIGG_2016_25_002)-[:IS_BASED_ON]->(CERN_ATL_COM_PHYS_2016_1599), (ATLAS_HIGG_2016_25_002)-[:IS_BASED_ON]->(CERN_ATL_COM_PHYS_2016_1605), 9 (ATLAS_HIGG_2016_25_002)-[:IS_BASED_ON]->(CERN_ATL_COM_PHYS_2016_1632)

CDS INVENIO

https://cds.cern.ch/help/hacking/search-engine-api

- 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)
 Maria Grigorieva

MARC

02210393 001 2210393		
J02210393 005 20170412182032.0		
02210393 037 \$\$aATL-COM-PHYS-2016-1194		
)02210393 041 \$\$aeng	MERADARA CONCEPT	DEODOCED MADC 21 DEDDECENTATION
02210393 088 \$\$9CERN-ATL-COM-PHYS-2016-1194	METADATA CONCEPT	FROFOSED MARC 21 REFRESENTATION
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 \$	Abstract	520 \$a
102210393 260\$\$c2016	Author, first	100 Şa
102210393 269 \$\$aGeneva\$\$bCERN\$\$c25 Aug 2016	Author(s), additional	700 \$a
102210393 300 \$\$amult. p	Collection identifier	980 \$a
102210393 520 \$\$aA search for the dimuon decay of the Higgs boson has been performed using data correspo	Paril	
J02210393 595 \$\$aCERN CDS-Invenio WebSubmit	Bindit	8300 \$1
102210393 6531_ \$\$9CERN\$\$aHIGGS	Imprint	260 \$a,b,c; 300 \$a
002210393 65017 \$\$28zGeCERN\$\$aParticle Physics - Experiment	Keywords	6531 \$a
02210393 65027 \$\$2SzGeCERN\$\$aPhysics	Language	041 \$a
02210393 690C_ \$\$aCERN	OAI identifier	909CO \$o
02210393 690C_ \$\$aintnote	Publication info	909C4 St [many subfields]
102210393 690C_ \$\$aPRIVATLAS		Source of [many sublicitus]
J02210393 693 \$\$aCERN LHC\$\$eATLAS	References	999C5 \$* [many subfleids]
/02210393 700_ \$\$0AUTHOR (SZGECERN)751740\$\$0AUTHOR (INSPIRE)INSPIRE-00436733\$\$aLiu, Yanlin\$\$gOptimization	Primary report number	037 \$a [unique throughout the s
102210393 700\$\$0AUTHOR (SZGECERN)446640\$\$0AUTHOR (INSPIRE)INSPIRE-00132531\$\$aTsybychev, Dmitri\$\$gSuperv	Additional report number(s)	088 \$a
102210393 700\$\$0AUTHOR (SzGeCERN)764063\$\$aFray, Antony\$\$gCutflow, spurious signal systematics\$\$ma.fray@	Series	490 \$a,v
102210393 700 \$\$0AUTHOR (SZGECERN)708109\$\$0AUTHOR (INSPIRE)INSPIRE-00376466\$\$aHoenig, Friedrich\$\$gMuon f	Subject	65017 \$a
102210393 700 \$\$0AUTHOR (SZGECERN)414244\$\$0AUTHOR (INSPIRE)INSPIRE-00300183\$\$aMehta, Andrew\$\$gVH,h->mumu	mitle	245 \$2
102210393 700 \$\$0AUTHOR (SZGECERN)670199\$\$0AUTHOR (INSPIRE)INSPIRE-00120467\$\$aRizvi, Eram\$\$gSupervisor\$\$	TILLE	245 58
J02210393 700\$\$0AUTHOR (SZGECERN)666716\$\$0AUTHOR (INSPIRE)INSPIRE-00185894\$\$aThompson, Paul\$\$gSignal th	URL (e.g. to fulltext)	8564 Şu, Şz
J02210393 700\$\$0AUTHOR (SZGECERN)515237\$\$0AUTHOR (INSPIRE)INSPIRE-00226382\$\$aZhao, Zhengguo\$\$gSuperviso	https://cds.cern.ch/help/admin/howto-marc	
J02210393 700 \$\$0AUTHOR (SZGECERN)401247\$\$0AUTHOR (INSPIRE)INSPIRE-00138410\$\$aZhou, Bing\$\$gSupervisor\$\$m		
JUZZIUJJJ 700_ ŞŞUAUTHOR (SZGECERN)J8499UŞŞUAUTHOR (INSPIRE)INSPIRE-00212987ŞŞADAİ, TieshengŞŞGPrivate MC		
JUZZIUJJJ /UU \$\$UAUTHUK (INSPIRE)INSPIRE-00237882\$\$0AUTHOR (SZGECERN)698137\$\$aWu, Yusheng\$\$gAnalysis, sp		

https://media.readthedocs.org/pdf/invenio-client/vo.1.o/invenio-client.pdf

the system!

CDS TO NEO4J DATAFLOW

