



AllegroGraph for Flexibility in the Enterprise and on the Web

Jans Aasman

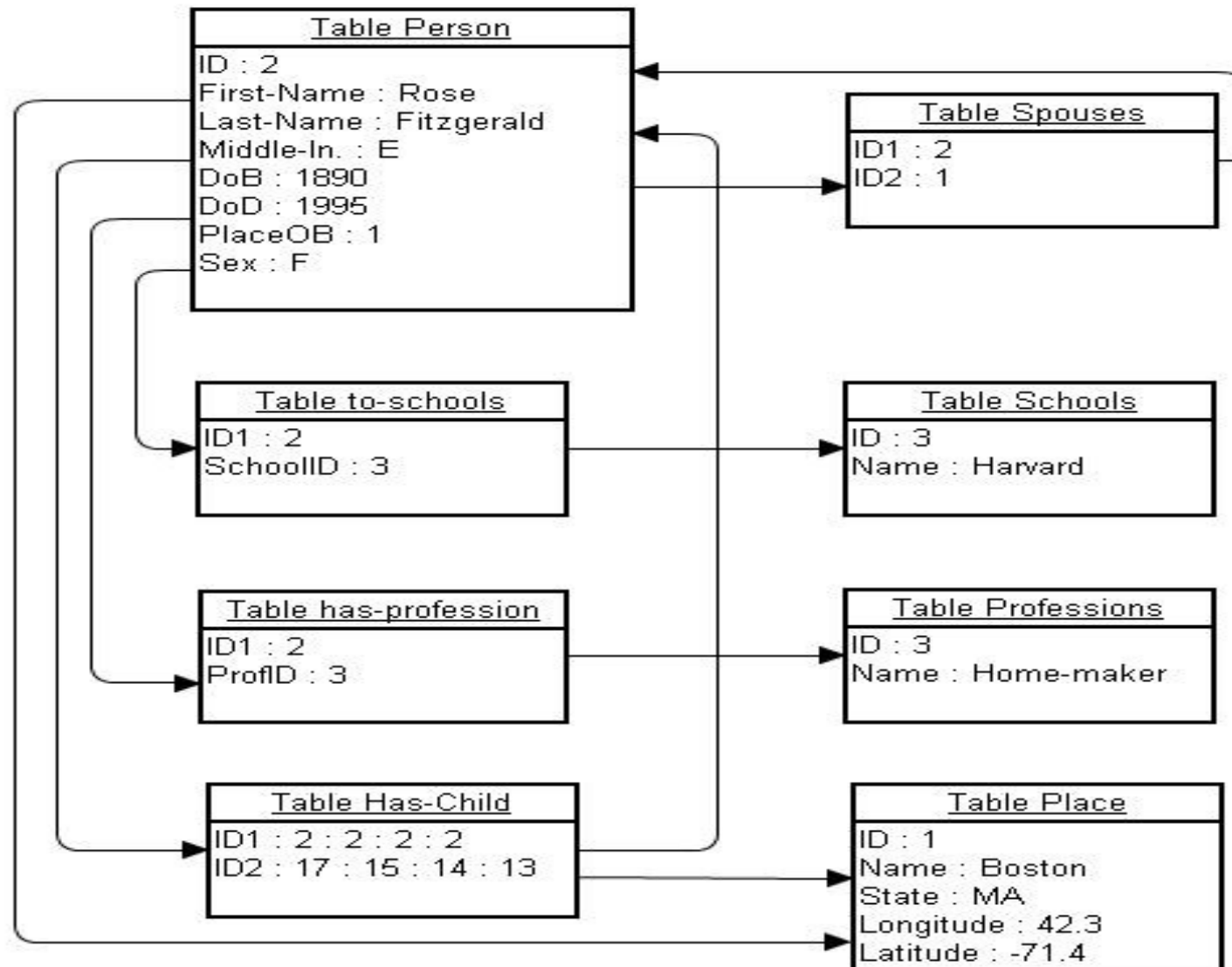
Franz Inc

ja@franz.com



What is a triple store

Person Database Diagram



subject	predicate	object
person2	type	person
person2	first-name	Rose
person2	middle-initial	E
person2	last-name	Fitzgerald
person2	suffix	none
person2	alma-mater	Harvard
person2	birth-year	1890
person2	death-year	1995
person2	sex	female
person2	spouse	person1
person2	has-child	person17
person2	has-child	person15
person2	has-child	person13
person2	has-child	person11
person2	has-child	person9
person2	has-child	person7
person2	has-child	person6
person2	has-child	person4
person2	has-child	person3
person2	profession	home-maker

(1 (2 3) (4 5) (6 7) (8 9) (10 11) (12 13) (14 15)(16 17) (18 19 20 21 22 23 24 27 28) (29 30))



How is it different and why is it more flexible?

- No **Schema**.
 - Say whatever you want to say but
 - ontologies may constrain what you put in triple store
- No **Link Tables**
 - because you can do one-to-many relationships directly
- No **Indexing Choices**
 - add new data attributes (predicates) on-the-fly that will be real-time available for querying
- Takes **anything** you give it: it is trivial to consume
 - Rows and columns from RDB, XML, RDF(S), OWL, Text and Extracted Entities



And how is it the same

- AllegroGraph has all the enterprise features of a relational database
 - 24/7, robust, real time processing
 - ACID, transactional
 - Replication, On-line Backup, Point in time recovery
 - High Availability
 - Many kinds of full text indexing
 - REST Protocol

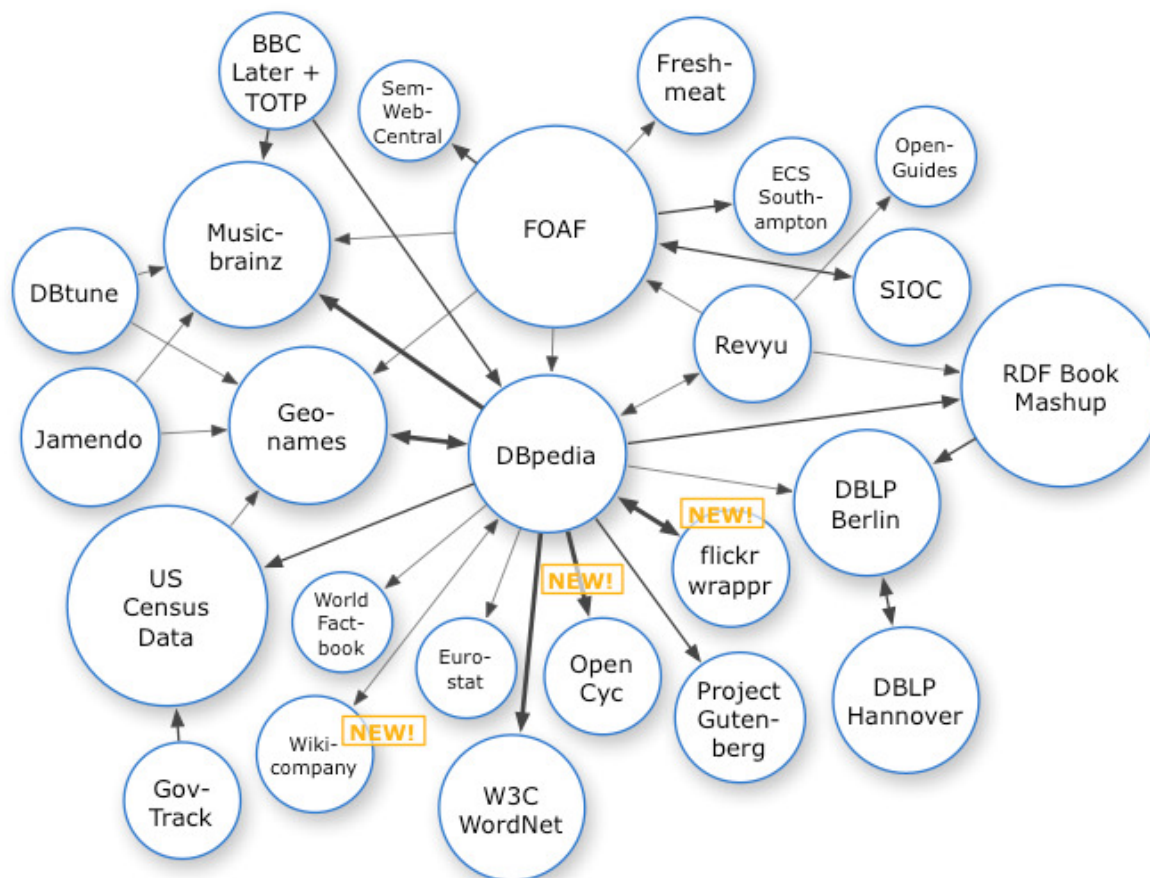


Why Triple Stores are here to stay.

- There is now a standard for meta data or graphs
 - data representation: RDF/RDFS
 - logic and inferencing: RDFS/OWL
 - querying: SPARQL
 - rules: RuleML, Swirl
- With vertical compression we can get this smaller than RDB
- Comparable to RDB and NoSQL wrt speed and scalability
- And an enormous impetus to share data via RDF



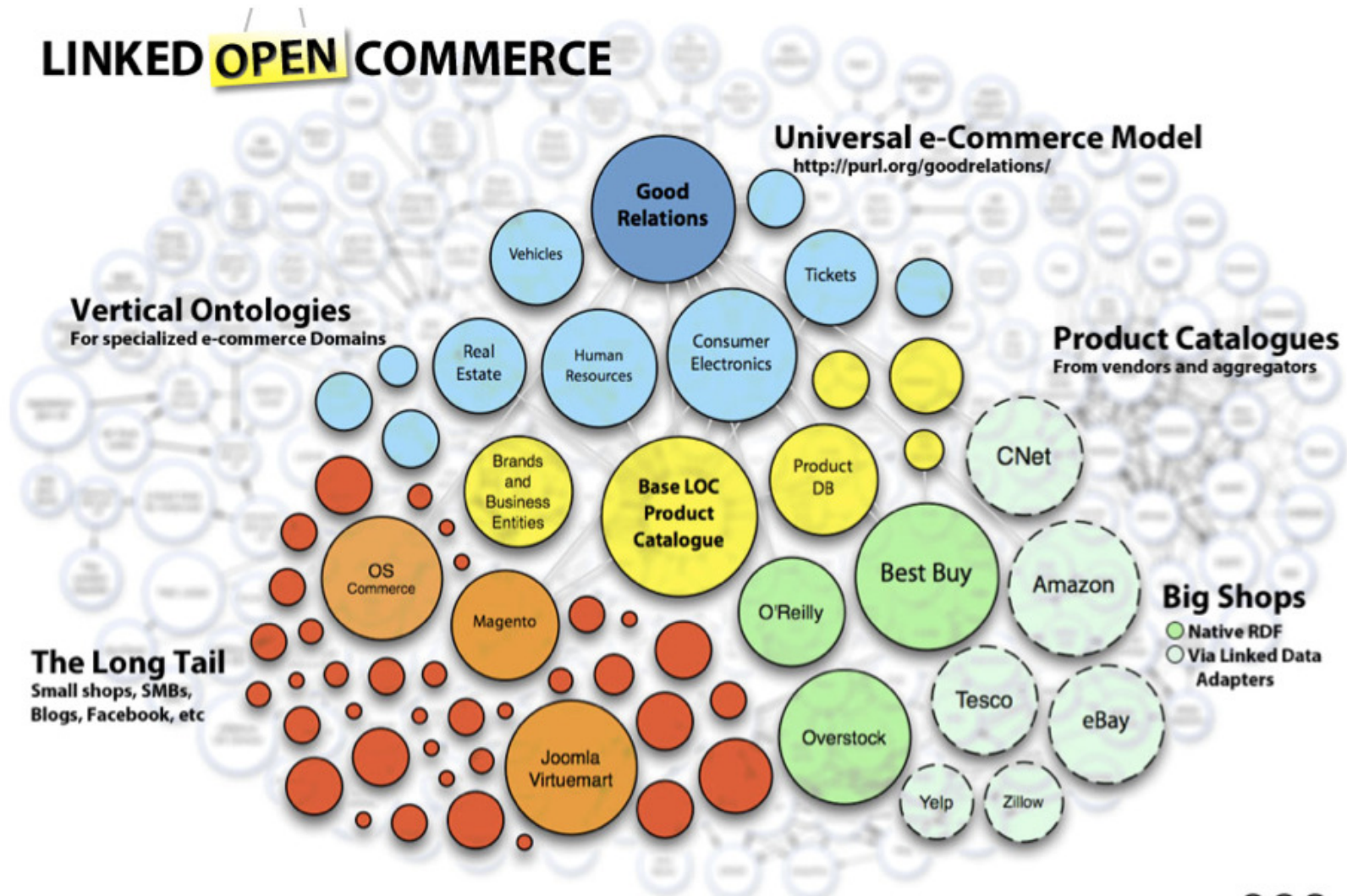
Oct 2007





LOD meets LOC

LINKED **OPEN** COMMERCE





Just look at this

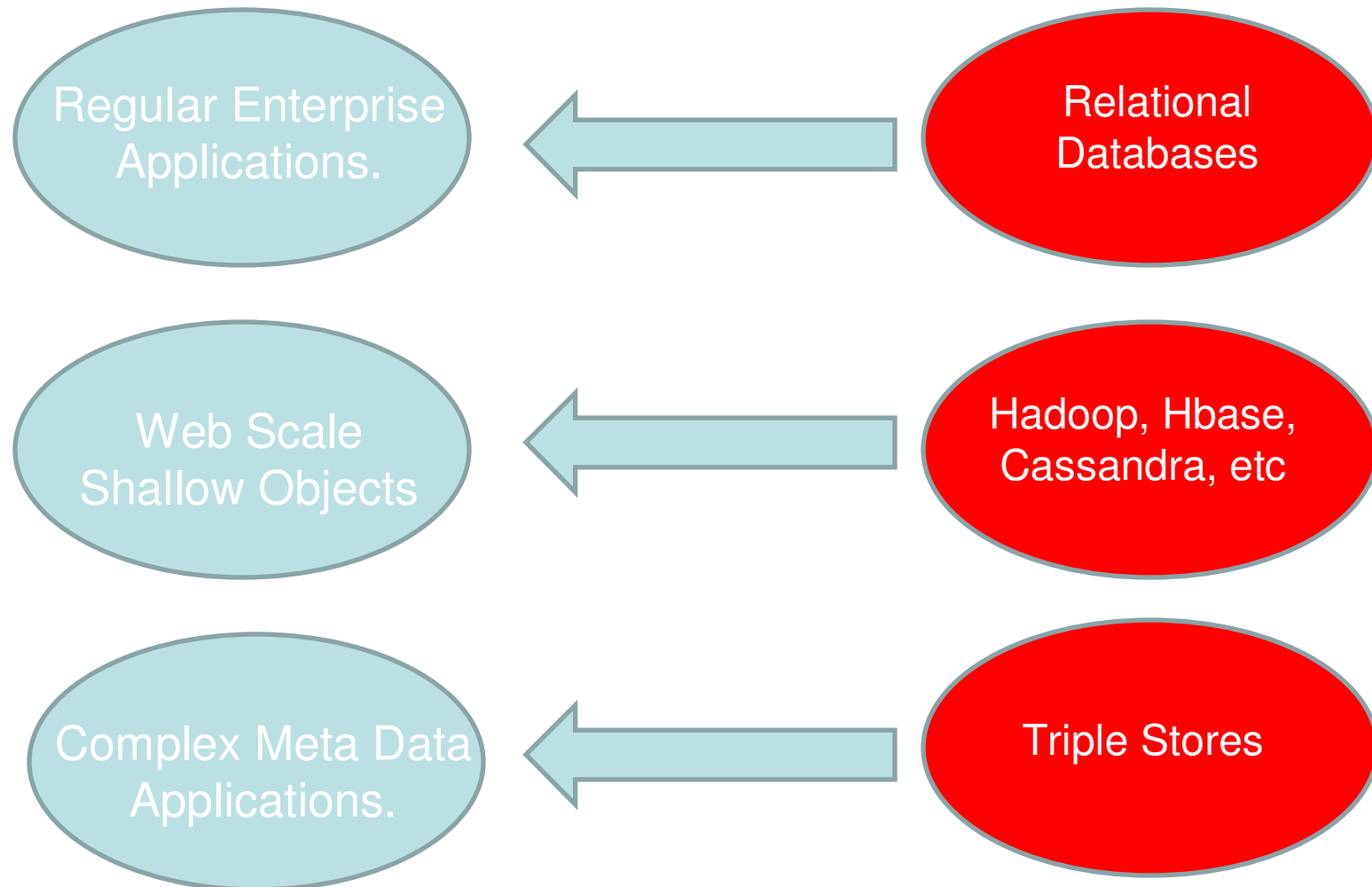


But what about NoSQL?

- And by that we mean
 - Cassandra
 - Hbase
 - CoucheDB
 - Etc..

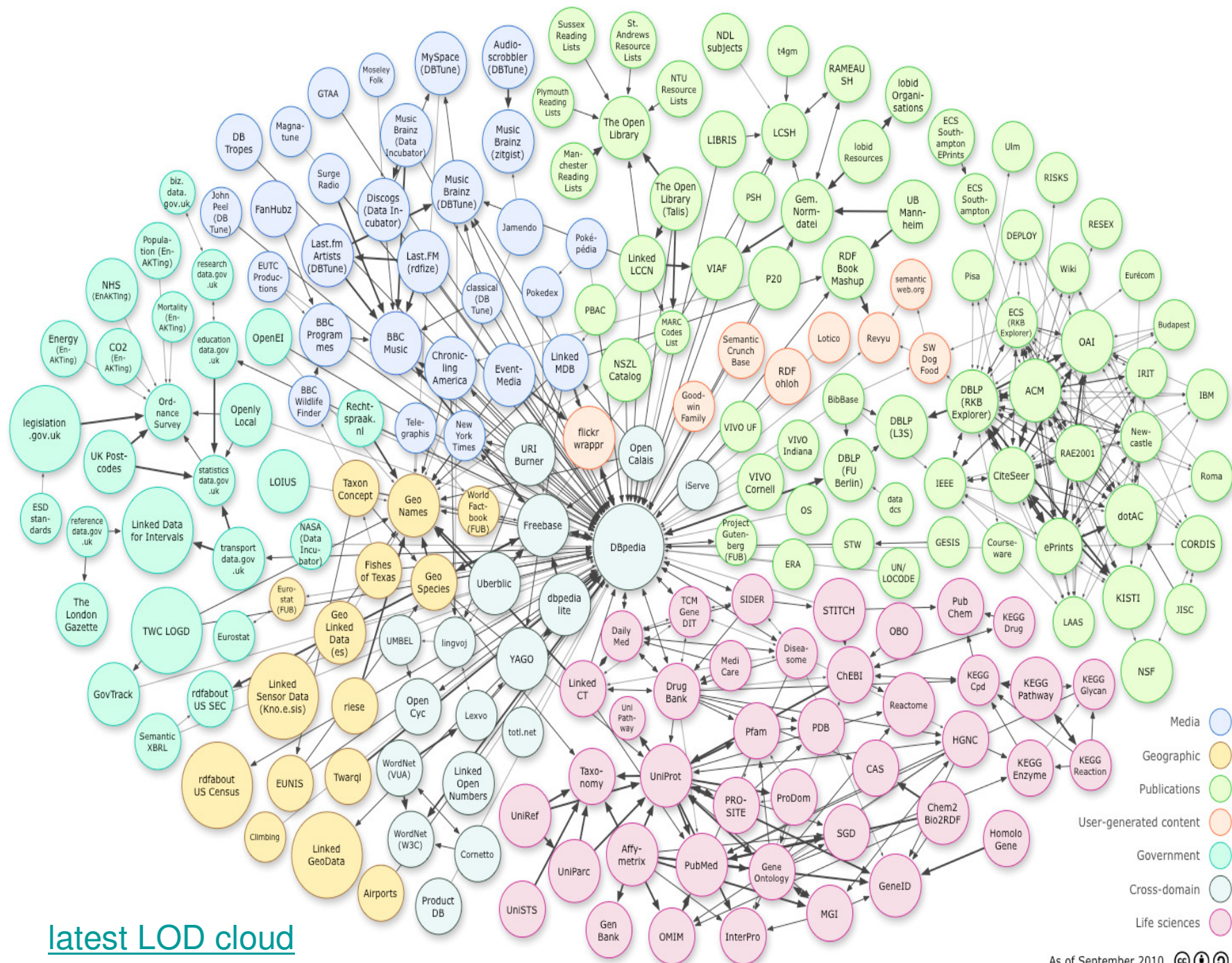


Classes of applications and their natural database.





Examples of things that work better in triple stores



latest LOD cloud

A SPARQL query spanning 4 sources

```
prefix go: <http://purl.org/obo/owl/GO#>
prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
prefix owl: <http://www.w3.org/2002/07/owl#>
prefix mesh: <http://purl.org/commons/record/mesh/>
prefix sc: <http://purl.org/science/owl/sciencecommons/>
prefix ro: <http://www.obofoundry.org/ro/ro.owl#>

select ?genename ?processname
where
{
  graph <http://purl.org/commons/hcls/pubmesh>
  {
    ?paper ?p mesh:D017966 .
    ?article sc:identified_by_pmid ?paper.
    ?gene sc:describes_gene_or_gene_product_mentioned_by ?article.
  }
  graph <http://purl.org/commons/hcls/goa>
  {
    ?protein rdfs:subClassOf ?res.
    ?res owl:onProperty ro:has_function.
    ?res owl:someValuesFrom ?res2.
    ?res2 owl:onProperty ro:realized_as.
    ?res2 owl:someValuesFrom ?process.
  }
  graph <http://purl.org/commons/hcls/20070416/classrelations>
  {
    { ?process <http://purl.org/obo/owl/obo#part_of> go:GO_0007166 }
    union
    { ?process rdfs:subClassOf go:GO_0007166 }
    ?protein rdfs:subClassOf ?parent.
    ?parent owl:equivalentClass ?res3.
    ?res3 owl:hasValue ?gene.
  }
  graph <http://purl.org/commons/hcls/gene>
  {
    ?gene rdfs:label ?genename
  }
  graph <http://purl.org/commons/hcls/20070416>
  {
    ?process rdfs:label ?processname
  }
}
```

Mesh: Pyramidal Neurons



Pubmed: Journal Articles



Entrez Gene: Genes



GO: Signal Transduction

Inference required



A Financial Application

- Find a cyclic path between two companies that are engaged in revenue boosting



Q1: A reasonable hard query for horizontally scaling stores and rdb, a straight forward query for vertical/parallel stores

Select ?a ?b ?c ?d ?e

where {

Franz send-money ?a

?a send-money ?b

?b send-money ?c

?c send-money Cray

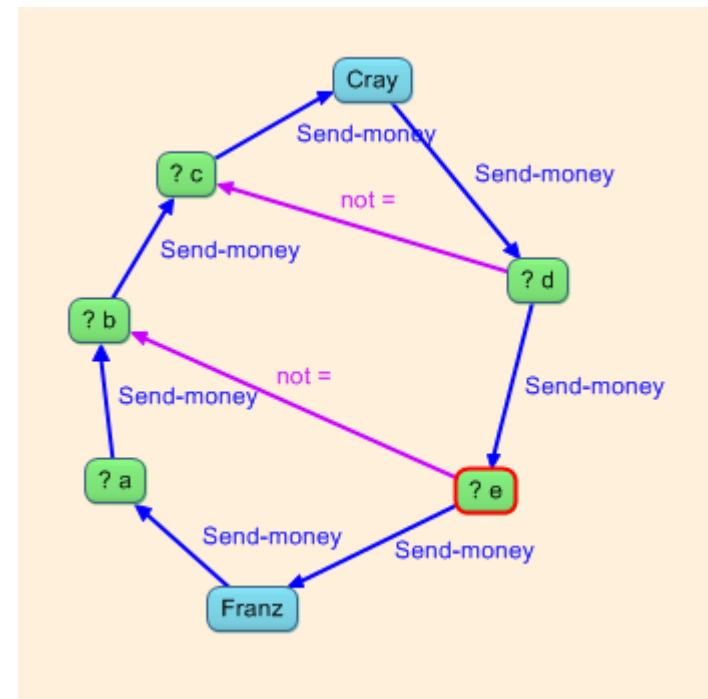
Cray send-money ?d

Not (?d = ?c)

?d send-money ?e

Not (?e ?b)

?e send-money Franz}





Q2: A very hard query for horizontally scaling stores and rdb, a straight forward query for vertical/parallel store

Find a money trail from Franz to Cray that is two or more steps, find another money trail from Franz Cray that is also two or more steps where the two trails are completely different

```
(Select (?path1 ?path2)
(path Franz Cray <send-money> >= 2 ?path1)
(path Cray Franz <send-money> >= 2 ?path2)
(empty (intersection ?path1 ?path2)))
```



Why is this hard in SQL

- Relational databases very good at straight joins but less optimal for self-joins of unpredictable length
- Try writing this as a sql query 😊



Why is this hard in distributed key/value stores.

- Databases like Cassandra are extremely good at retrieving nested objects in a sea of billion of objects but are less optimal for joins.
- Relatively hard to write these as map reduce expressions. Every query has to be expressed as program, ad hoc is therefore discouraged'



Why is this easier in vertical scaling triple stores

- These queries can be easily parallelized
- Graph databases are made to do graph search
- Ad hoc queries easy to write in Sparql or Prolog

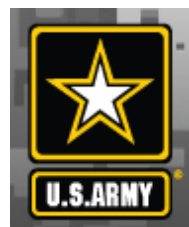


Summarizing

	RDB	Cassandra, Hbase	AllegroGraph 4.0
Schemaless	-	+/-	+
Transactions	+	+/-	+
ACID	+	+/-	+
Concurrent & Dynamic	+	+	+
Random Access	+	+	+
High Availability	+	+	+
Complex graph search	-	-	+
Structured + Unstructured	-	+/-	+
Scalability	+	++	+
General Flexibility	-	+/-	+



Is it also used in the enterprise?



The Redlands Institute



amdocs

CUSTOMER EXPERIENCE SYSTEMS INNOVATION

AMDOCS STATEMENTS OF DIRECTION ANNOUNCING AMDOCS CES 8

OUR NEWEST PORTFOLIO RELEASE.

Discover the market imperatives, design principles, and how they will be implemented. Hear Dov Baharav, Amdocs CEO, and Emily Green, President and CEO of Yankee Group, along with other Amdocs executives discuss Amdocs CES 8.

[View this online event now >](#)

THE UNIVERSITY OF TEXAS

MD Anderson ~~Cancer~~ Center

Making Cancer History[®]



HOOVERS[™]
A D&B COMPANY

Try Hoover's Today. Get Business Done. Fast.



UNITED STATES

DEPARTMENT OF DEFENSE

www.defense.gov



Why do they say they use triple stores?

- **Flexibility**
 - In changing your data schemas on the fly and do real time processing with (meta) data
 - Add new predicates and use them in queries the next microsecond
- **Flexibility**
 - In writing ad-hoc queries
 - In finding complex patterns and working with rules
- **Flexibility**
 - Integrating various data sources without complicated masterdata management.

and there is a standard



An event example



A Simple Event Ontology

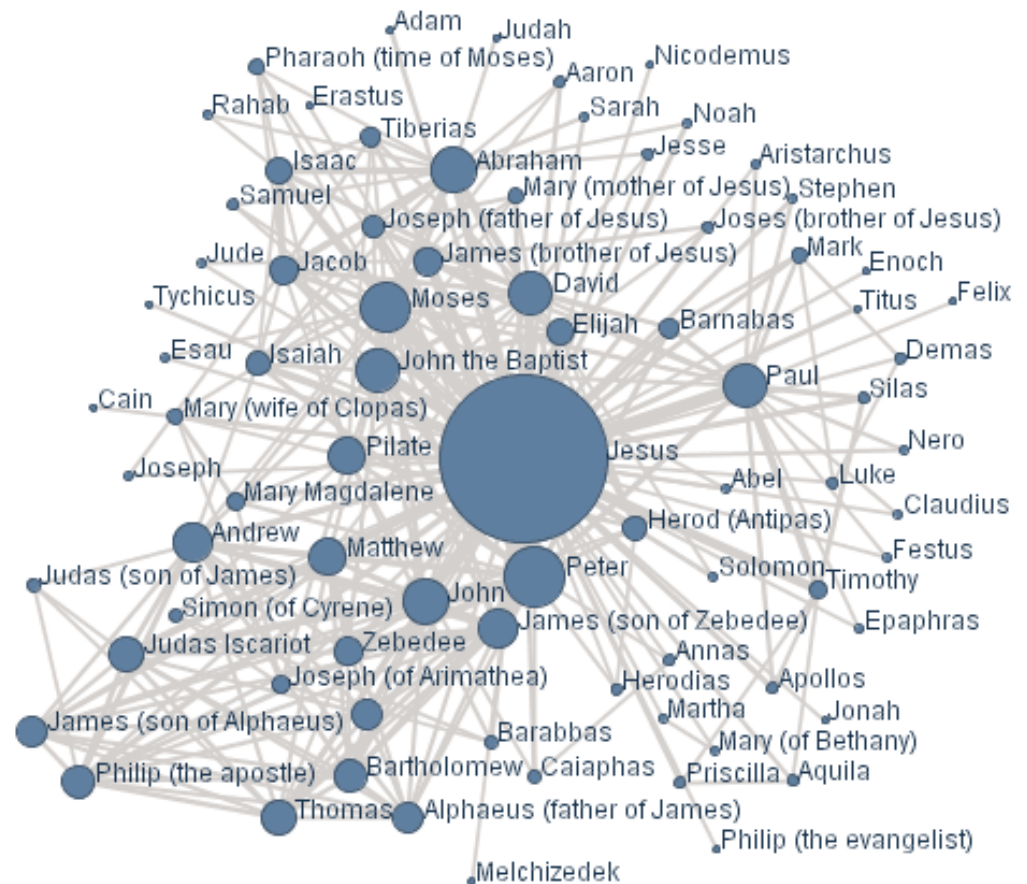
- A type
 - Meetings, communications event, financial transactions, visit, attack/truce, an insurance claim, a purchase order
 - RDFS++ reasoning
- A list of actors
 - Social Network Analysis
- A place
 - GeoSpatial Reasoning
- A Start-time and possible an end-time
 - Temporal Reasoning
- Anything else that describes the event
 - Goods that changed hands



Social Network Analysis

Answers 4 questions

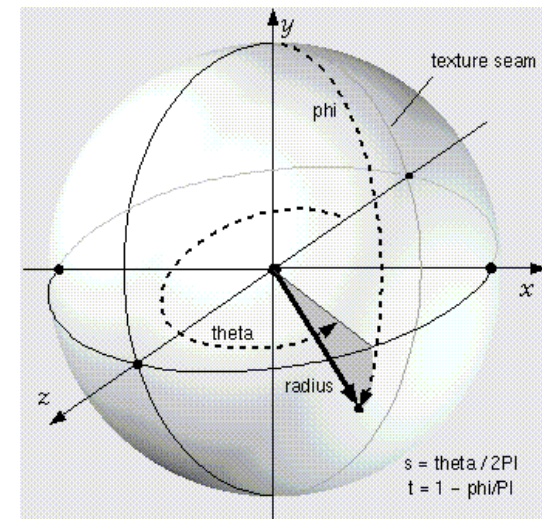
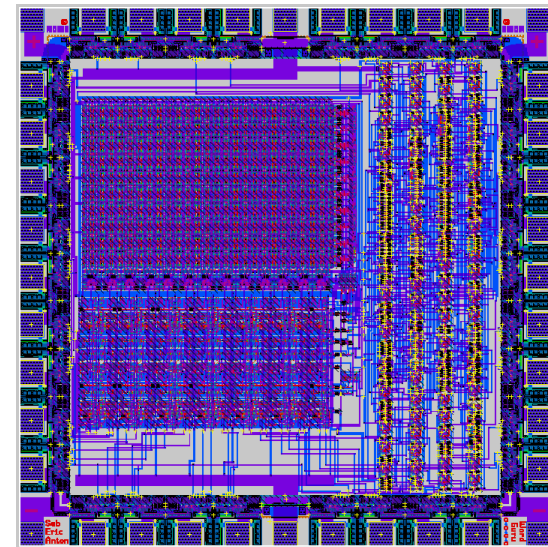
- How far is P1 from P2 (and how strong is the relation?)
- To what groups does this person belong (ego groups, cliques?)
- How important is this person in the group?
- Does this group have a leader, how cohesive are they?





GeoSpatial


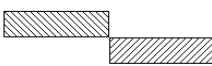
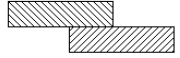

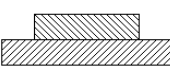


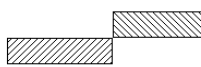
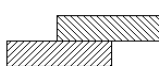

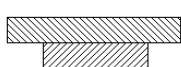
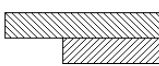
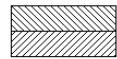
- Make the following super efficient
 - Where did something happen?
 - How far was event1 from event2?
 - Find all the events that occurred in a bounding box or radius of M miles?
 - Do these two shapes overlap?
 - Find all the objects in the intersection of two shapes
- On a very large scale
 - when things don't fit in memory
 - millions of events and polygons





Temporal Reasoning

- Adhere to our convention to encode StartTimes and EndTimes and enjoy efficient temporal primitives
- Implementation of Allen's interval logic primitives

	(interval-before ?e1 ?e2)
	(interval-meets ?e1 ?e2)
	(interval-overlaps ?e1 ?e2)
	(interval-starts ?e1 ?e2)
	(interval-during ?e1 ?e2)
	(interval-finishes ?e1 ?e2)
	(interval-after ?e1 ?e2)
	(interval-met-by ?e1 ?e2)
	(interval-overlapped-by ?e1 ?e2)
	(interval-started-by ?e1 ?e2)
	(interval-contains ?e1 ?e2)
	(interval-finished-by ?e1 ?e2)
	(interval-cotemporal ?e1 ?e2)



Activity Recognition

- Our customers use AllegroGraph as an event database with social network analysis and geospatial and temporal reasoning

Find all meetings that happened in November within 5 miles of Berkeley that was attended by the most important person in Jans' friends and friends of friends.

```
(select (?x)
  (ego-group person:jans knows ?group 2)
  (actor-centrality-members ?group knows ?x ?num)
  (q ?event fr:actor ?x)
  (qs ?event rdf:type fr:Meeting)
  (interval-during ?event "2008-11-01" "2008-11-06")
  (geo-box-around geoname:Berkeley ?event 5 miles)
!)
```

SNA

SNA

DB Lookup

RDFS

Temporal

Spatial



Thank you
And questions?