YarcData
A DIVISION OF CRAY INC.

launches

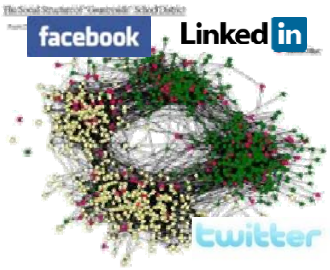
uRiKA

Big Data Graph Appliance for Relationship Analytics

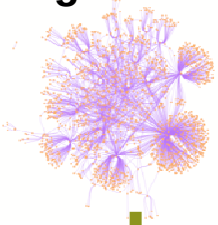
Feb 29, 2012

Many Big Data problems are based on Graphs

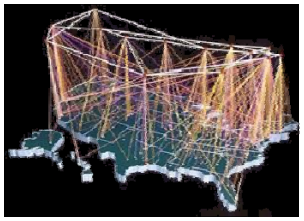
Social Networking

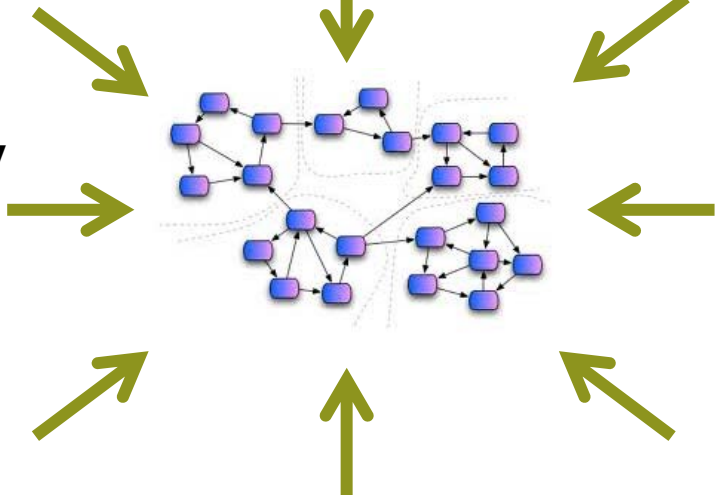


Intelligence/Security

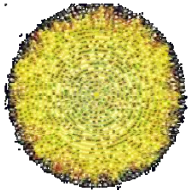


Telecom/Mobile

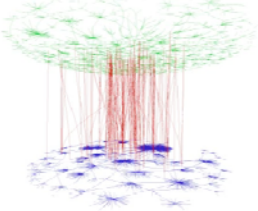




Life Sciences/Biology



Healthcare/Medicine

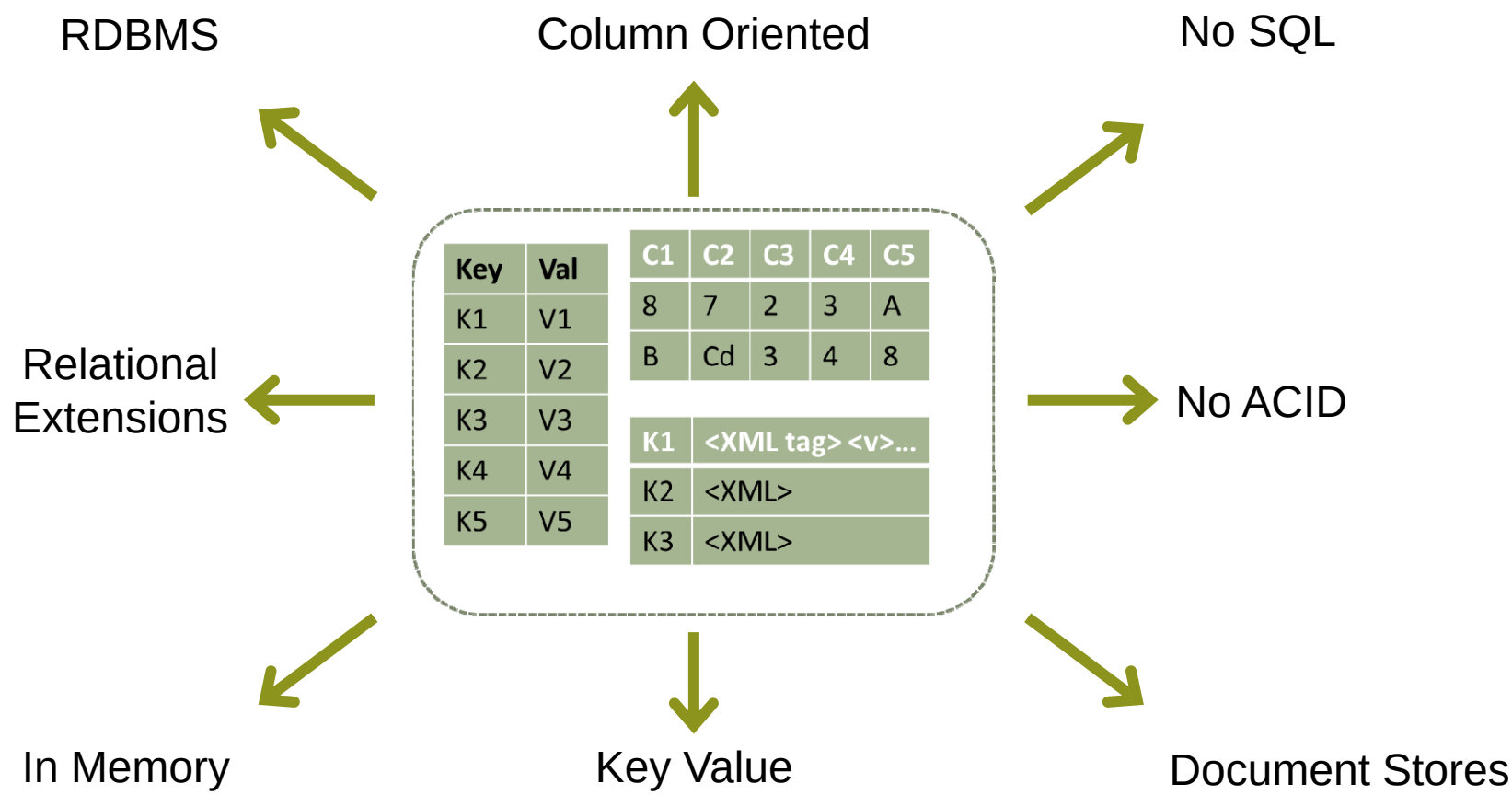


Finance

Internet/WWW

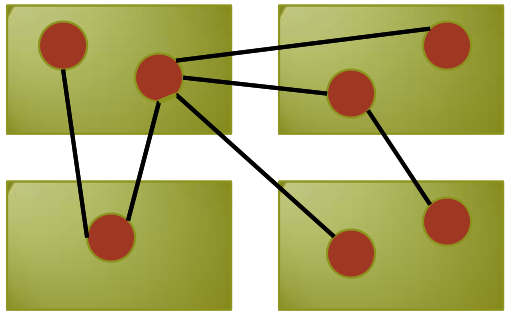
Supply Chain

But most Big Data solutions are based on distributed, index-based data structures that scale out on clusters



Current Big Data approaches (including graph databases) result in low performance on Graphs…

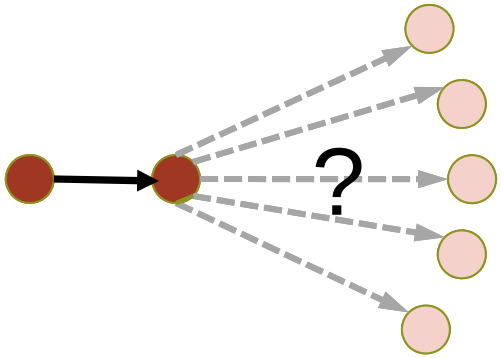
Graphs are hard to Partition



High cost to follow relationships that span Cluster Nodes

Network is 100 times SLOWER than Memory*

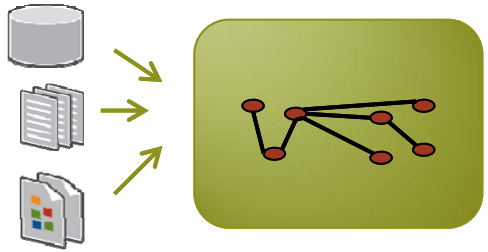
Graphs are Non-Deterministic



High cost to follow multiple competing alternatives which cannot be pre-fetched/cached

Memory is 100 times SLOWER than Processor*

Graphs are highly Dynamic



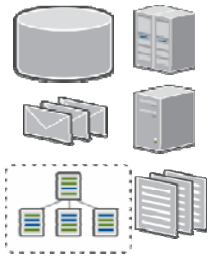
High cost to load multiple, constantly changing datasets into in-memory graph models

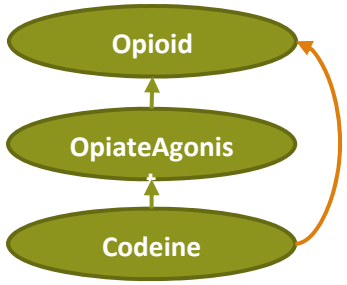
Storage I/O is 1000 times SLOWER than Memory I/O*

*Source: Hennessy, J. and Patterson, D., "Computer Architecture: A Quantitative Approach", 2012 edition

The uRiKA Moment: Discovery of Unknown/Hidden Relationships in Big Data

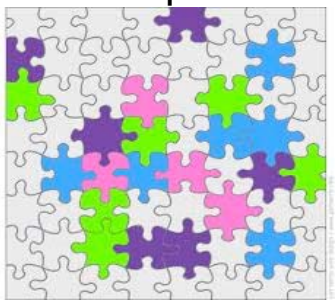
1. Aggregate data and relationships from multiple sources



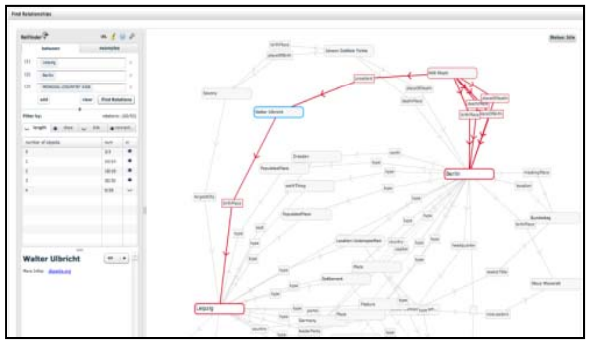


2. Augment Relationships through automated inference and deduction

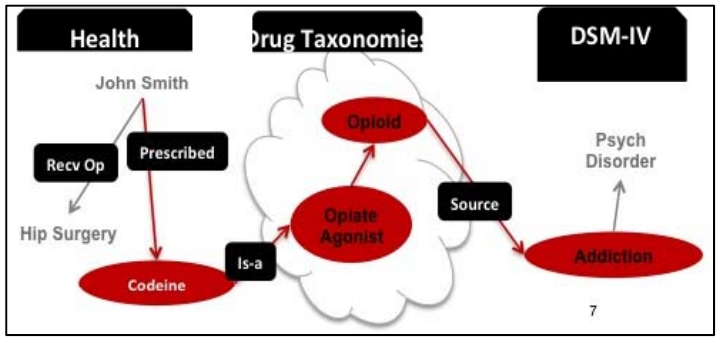
3. Build a Dynamic Relationship Warehouse



Visualize relationships for real time, interactive Discovery



Search for relationships based on partially specified Patterns/Templates



YarcData
A DIVISION OF CRAY INC.

uRiKA Customer Use Case: Government Organization

"Connecting the dots" to identify Persons of Interest

The Challenge

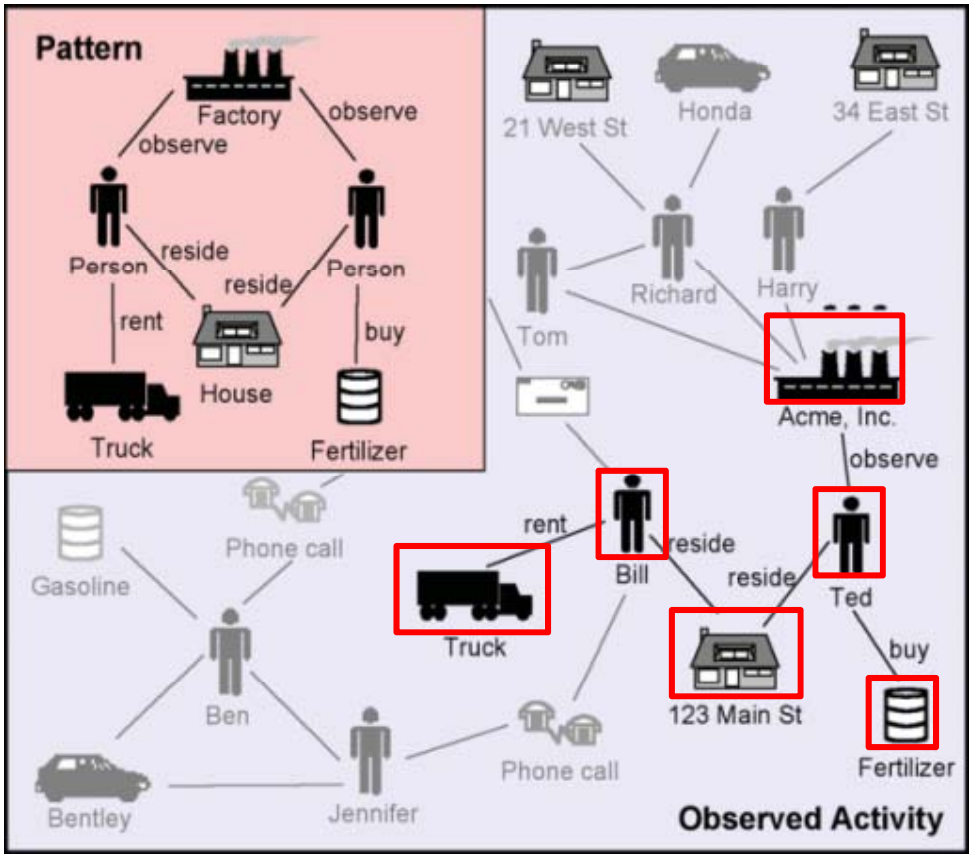
- Massive data stores of multiple data types from multiple sources
- Inaccurate, Incomplete and Falsified data
- Continuous stream of incoming data

uRiKA Solution

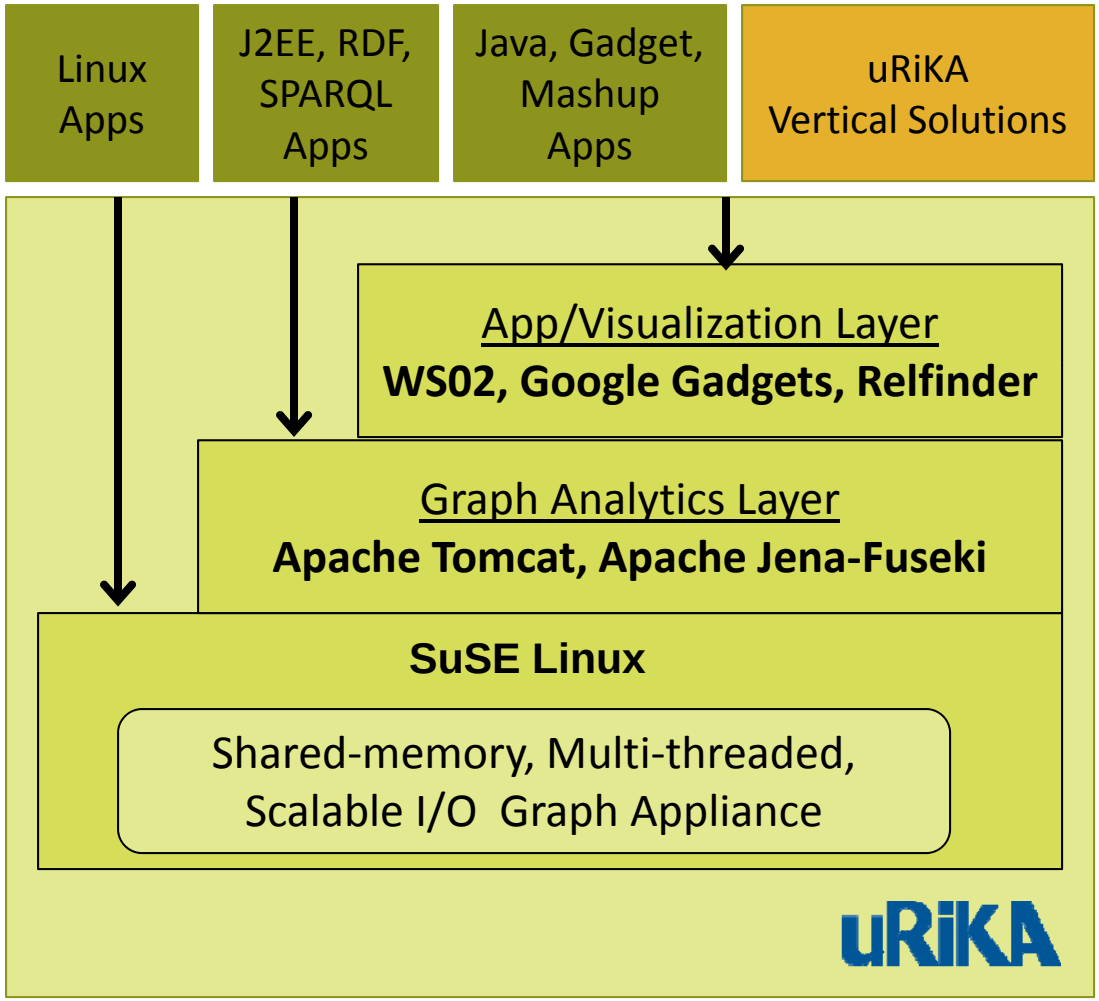
- uRiKA holds entire relationship graph in memory – updated constantly
- Search for Patterns of suspicious behavior and activities
- Graphical interactive exploration of relationships between people, places, things, organizations, communications, etc.

Business Value

- Proactive identification of terrorists, criminals and plots



uRiKA: …with enterprise-ready Application Platform



- **Industry-standard, Open-source Software Stack**
  - Linux, Java, Apache, WS02, Gadgets, Mashups…
- **Reusable Existing Skillsets**
  - OSGI, App Server, SOA, ESB, Web toolkit…
- **No Lock-in**
  - All applications and artifacts built on uRiKA can be run on other platforms
- **Vertical Solutions coming soon…**
  - Life Sciences, Financial Services, Healthcare…

uRiKA: Big Data Graph Appliance for Relationship Analytics

Discover Unknown and Hidden Relationships in Big Data

- *Relationship Warehouse supporting Inferencing/Deduction, Pattern-based queries and Intuitive Visualization*

Perform Real-time Analytics on Big Data Graphs

- *High-performance, Graph Appliance with large shared-memory, massive multi-threading and scalable I/O*

Realize Rapid Time to Value on Big Data Solutions

- *Ease of Enterprise adoption with industry-standards, open-source software stack enabling reuse of existing skillsets and no lock-in*

YarcData
A DIVISION OF CRAY INC.

Differences between KDT/CombBLAS and RDF/SPARQL Worldviews

- **(SPR's) KDT/CombBLAS worldview**
  - Few types of vertices, few types of edges
  - Minimal infrastructure to get data into KDT
  - Typical calculation explores global structure
- **RDF/SPARQL**
  - Many types of vertices, many types of edges
  - Tremendous community effort to get data into RDF form and into "SPARQL endpoints" (== semantic database instance)
  - Core calculation is subgraph isomorphism
  - No real capability to do global-structure calculations