Center for Applied High Performance Computing

Will Mitchell, Software Architect
William.Mitchell@noblis.org

Ben Pecheux, Software Architect
benjamin.pecheux@noblis.org

Gil Miller, Noblis CTO
hgmiller@noblis.org

HPC User Forum
April 2012

noblis
For the best of reasons

Google CEO Eric Schmidt

I don't think of this as my phone, it is my personal super computer that can answer any question that I care about in life.

Noblis

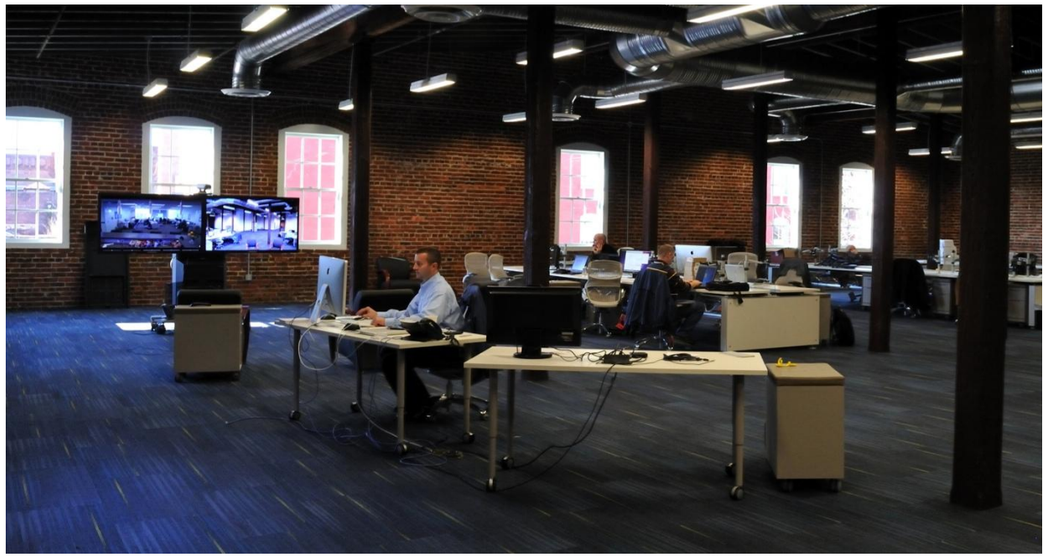
Non-profit science, technology, and strategy organization

Noblis is a science, technology, and strategy organization that helps clients solve complex scientific systems, process, and infrastructure problems in ways that benefit the public:

- Noblis delivers the best of scientific thought, management, and engineering know-how
- In an environment of independence and objectivity to find solutions that are practical, forward thinking, and sustainable

Center for Applied High Performance Computing

A transformational center



Vision

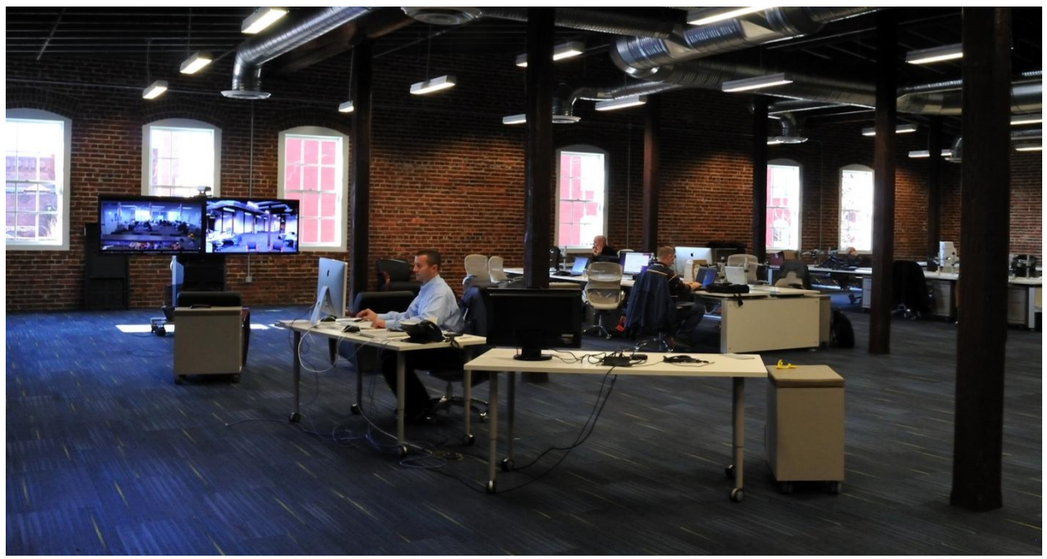
- A world class center for high performance computing that accelerates the development of applications requiring graph analytics for the benefit of the U.S.

Mission

- Promote development of HPC applications to solve problems of national importance
- Reduce the cost barriers to entry to HPC
- Enable U.S. competitiveness
- Create sustainable high-tech jobs
- Enable small businesses to innovate using HPC

Center for Applied High Performance Computing

A transformational center



We continue to seek and welcome additional partners and collaborators

Center for Applied High Performance Computing

Dan River—Where Innovations Flow

Noblis establishes transformational center

Vision

- A world class center for high performance computing that accelerates the development of applications requiring graph analytics for the benefit of the U.S

Mission

- Promote development of HPC applications to solve problems of national importance
- Reduce the cost barriers to entry to HPC
- Enable U.S. competitiveness
- Create sustainable high-tech jobs
- Enable small businesses to innovate using HPC

Center for Applied High Performance Computing

Selected areas of research and development

- New platform architectures, including Extreme Data Framework development
- Large transactional data sets – analytics and fraud detection
- Mobile applications connected to HPC
- Integration and semantic formulations of legacy data bases
- Infectious disease interaction and discovery
- Probabilistic networks
- Network resilience for critical networks

Center for Applied High Performance Computing

HPC resources

danriver — Cray XMT 2

(aka “uRiKa Relationship Analytics Appliance”)

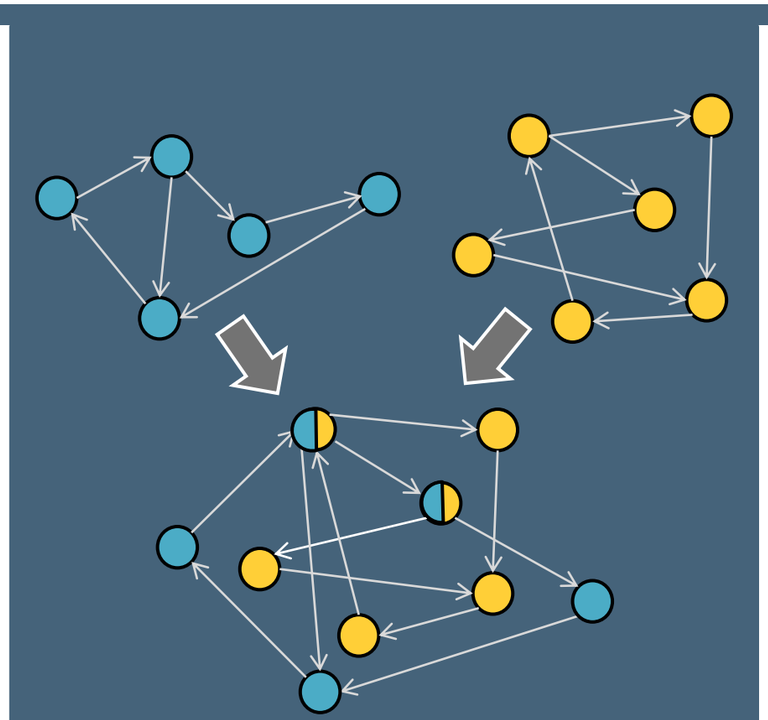
- 128 node configuration with high bandwidth 3D Torus interconnect
- 4 terabytes of shared memory
- 100 terabytes of Lustre storage
- 10 gigabit networking

NICC — HPC cluster

- 144-core Torque and MPI-based traditional Linux cluster
  - Can be augmented to double that capacity
- Over 20 terabytes of usable storage is also available to the cluster
  - Additional storage provisioned as needed
- 10 gigabit networking

Center for Applied High Performance Computing

Cray XMT 2 and the uRiKa appliance



eXtreme MultiThreaded architecture optimized for:

- Graph analytics: Problems that can be represented as trees, networks, or graphs
- Sifting through large amounts of data and finding commonalities between seemingly unrelated data

Cray XMT 2 enables:

- Problems represented as trees, networks, or graphs
- Semantic data and web formulations
- Applications that need access to large amounts of memory in an unpredictable manner
- Finding real-time answers to complex queries within large data sets using graph analytics

Yarc Data uRiKa software provides:

- Sparql endpoint
- Inference engine
- Gadget-based web UI

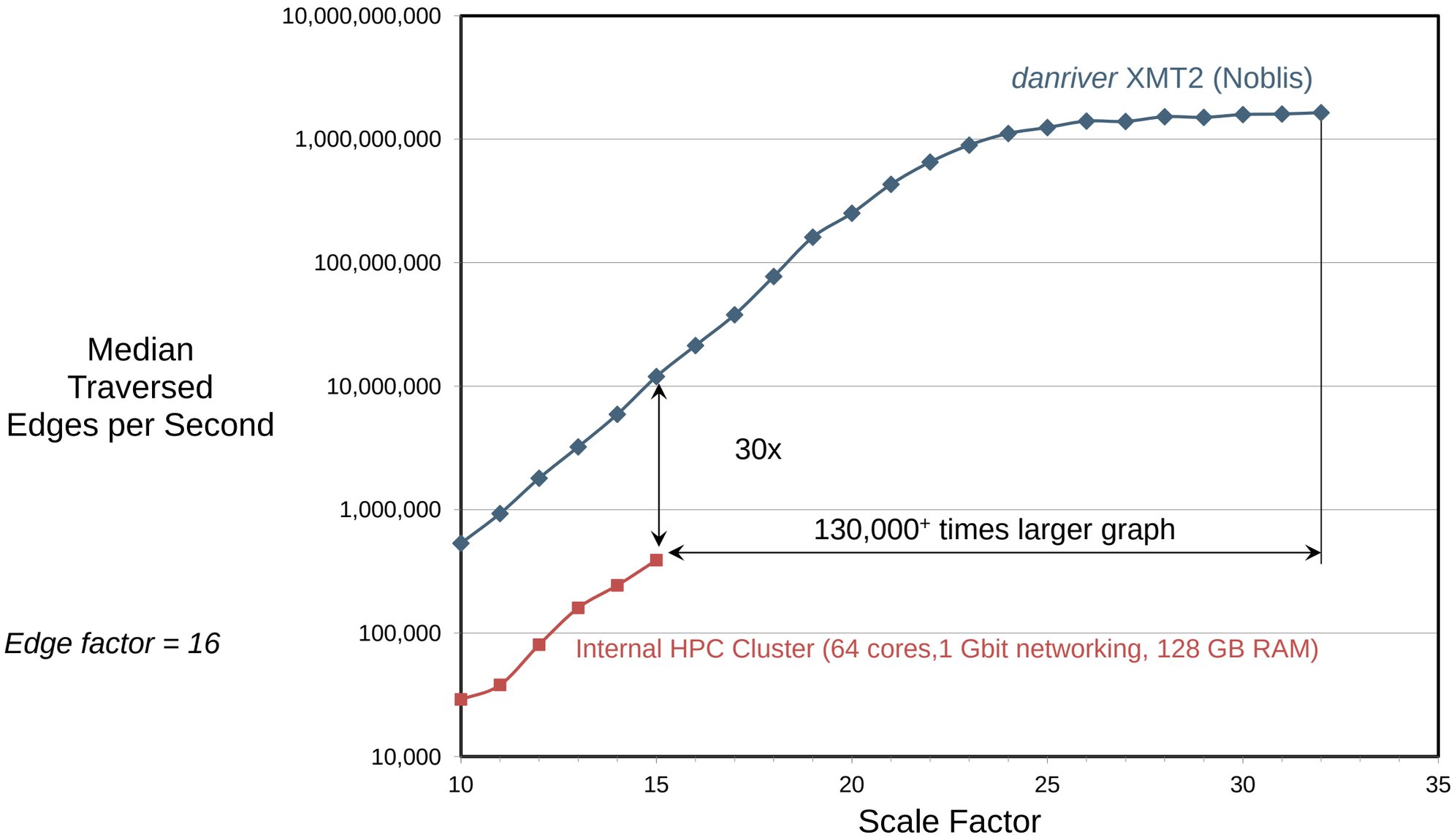
Graph 500 Benchmark

November 2011 Official Results

Rank	Machine and owner	Graph Size	TEPS	Category
1	NNSA/SC Blue Gene/Q Prototype II (4096 nodes / 65,536 cores) (IBM Research, T.J. Watson)	32	254,349,000,000	Custom
2	Hopper (1800 nodes / 43,200 cores) (LBL)	37	113,368,000,000	Custom
2	Lomonosov (4096 nodes / 32,768 cores) (Moscow State University)	37	103,251,000,000	Custom
3	TSUBAME (2732 processors / 1366 nodes / 16,392 CPU cores) (GSIC, Tokyo Institute of Tech)	36	100,366,000,000	Custom
4	Jugene (65,536 nodes) (Forschungszentrum Jülich)	37	92,876,900,000	Custom
5	Intrepid (32,768 nodes / 131,072 cores) (ANL)	35	78,869,900,000	Custom
6	Endeavor (Westmere) (320 nodes / 640 sockets / 3840 cores) (Parallel Computing Lab / Intel Labs)	32	57,567,900,000	Custom
7	IBM BlueGene/Q (512 nodes) (IBM Research, T.J. Watson)	30	56,523,400,000	Custom
8	Franklin (4000 nodes / 16,000 cores) (LBL)	36	19,955,100,000	Custom
9	SGI Altix ICE 8400EX (256 nodes / 1024 cores) (SGI)	31	14,085,400,000	Reference
10	SGI Altix UV 1000 (2048 cores) (SGI)	32	10,161,300,000	Reference
11	Red Sky (512 nodes / 4096 cores) (SNL)	33	9,470,000,000	Custom
12	Endeavor (Sandy Bridge) (256 nodes / 4096 cores) (Intel)	34	9,250,000,000	Custom
13	Lonestar (1024 nodes / 12 cores) (TACC)	35	9,240,000,000	Custom
14	DAS-4/VU (128 processors) (VU University)	30	7,135,130,000	Reference
15	BlueGene/P (2048 nodes / 8192 cores) (Moscow State University)	32	6,930,560,000	Custom
16	Jaguar PF (512 nodes / 1024 cores) (ORNL)	33	6,260,000,000	Custom
17	mirasol (Quad-Socket Intel E7-8870 - 2.4 GHz - 10 cores / 20 threads per socket) (Georgia Tech)	28	5,125,652,789	Custom
18	Blacklight (512 processors) (PSC)	32	4,452,270,000	Custom
19	Todi (176 AMD Interlagos, 176 NVIDIA Tesla X2090) (CSCS)	28	3,059,970,000	Custom
20	Dingus (Convey HC-1ex - 1 node / 4 cores, 4 FPGAs) (SNL)	28	1,758,682,718	Custom
21	Wingus (Convey HC-1ex - 1 node / 4 cores, 4 FPGAs) (SNL)	27	1,754,693,945	Custom
22	Vortex (Convey HC-1ex - 1 node / 4 cores, 4 FPGAs) (Convey Computer Corporation)	28	1,675,401,731	Custom
23	Convey01 (Convey HC-1ex - 1 node / 4 cores, 4 FPGAs) (Bielefeld University, CeBiTec)	28	1,614,891,176	Custom
24	Hc1-d (Convey HC-1 - 1 node / 4 cores, 4 FPGAs) (Convey Computer Corporation)	28	1,601,346,927	Custom
25	Convey2 (Convey HC-1 - 1 node / 4 cores, 4 FPGAs) (LBL/NERSC)	28	1,597,872,397	Custom

Graph 500 Benchmark

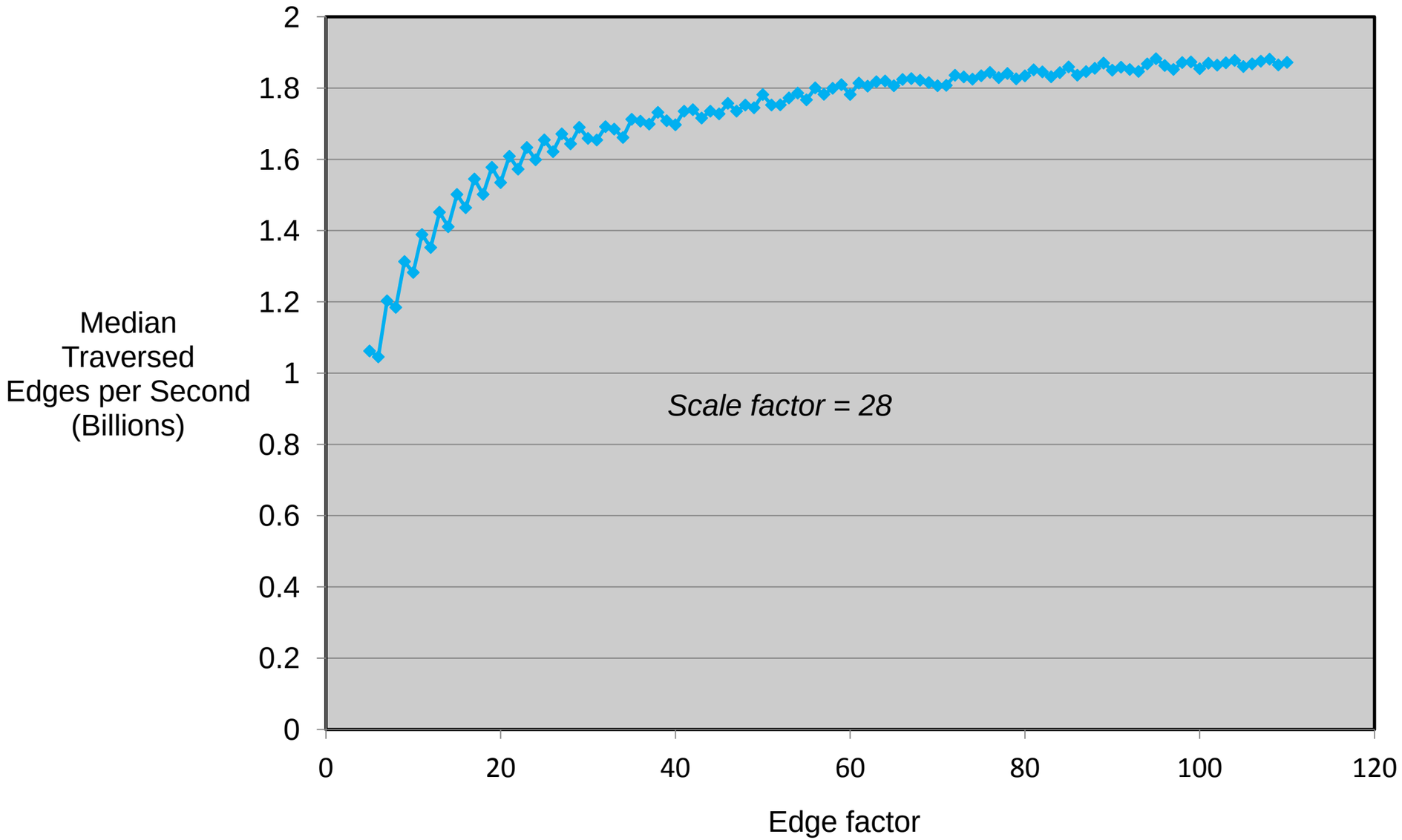
Preliminary Results for the Noblis Danriver XMT2



Edge factor = 16

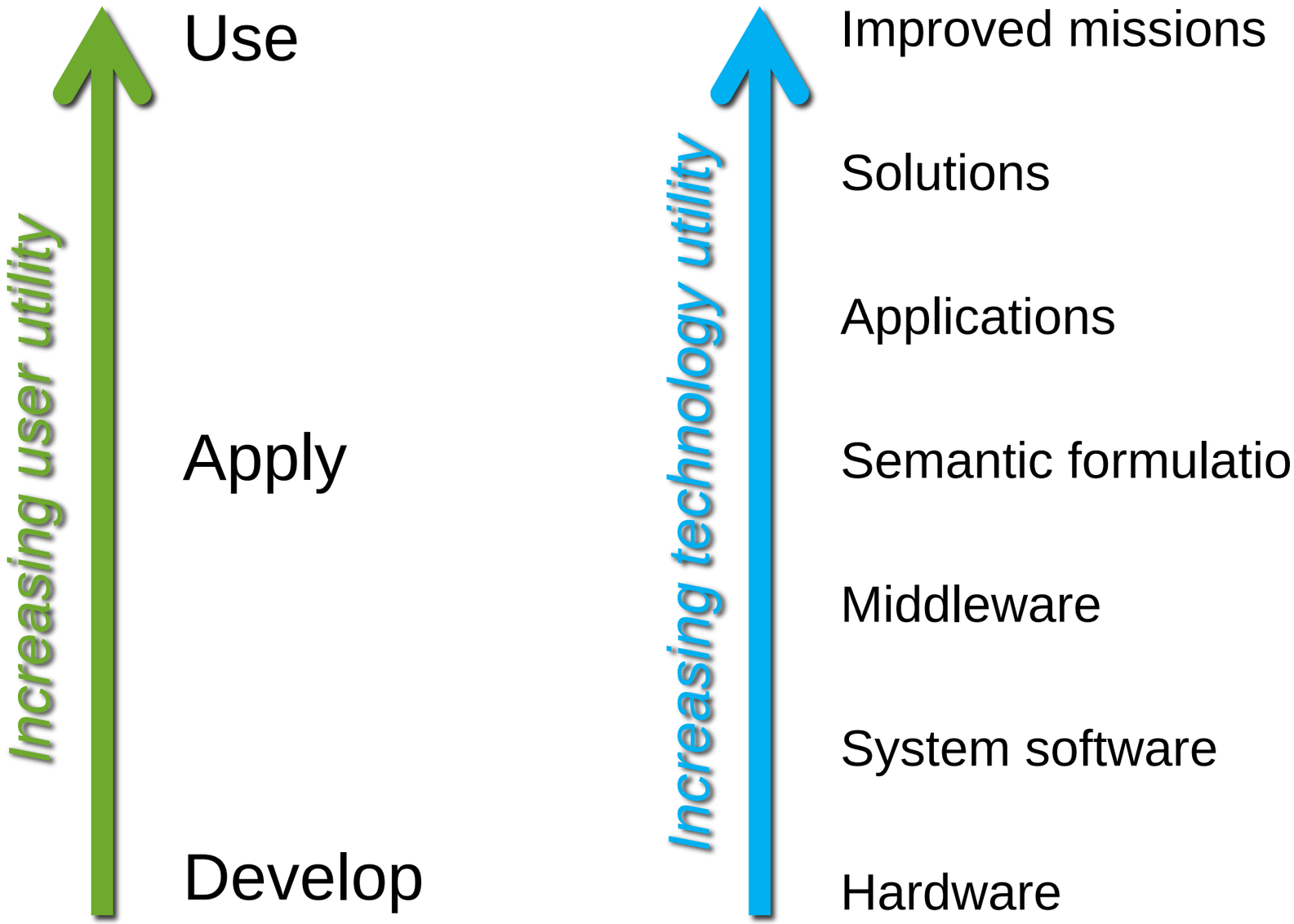
Graph 500 Benchmark

XMT2 TEPS Score increasing with Edge Factor (density)



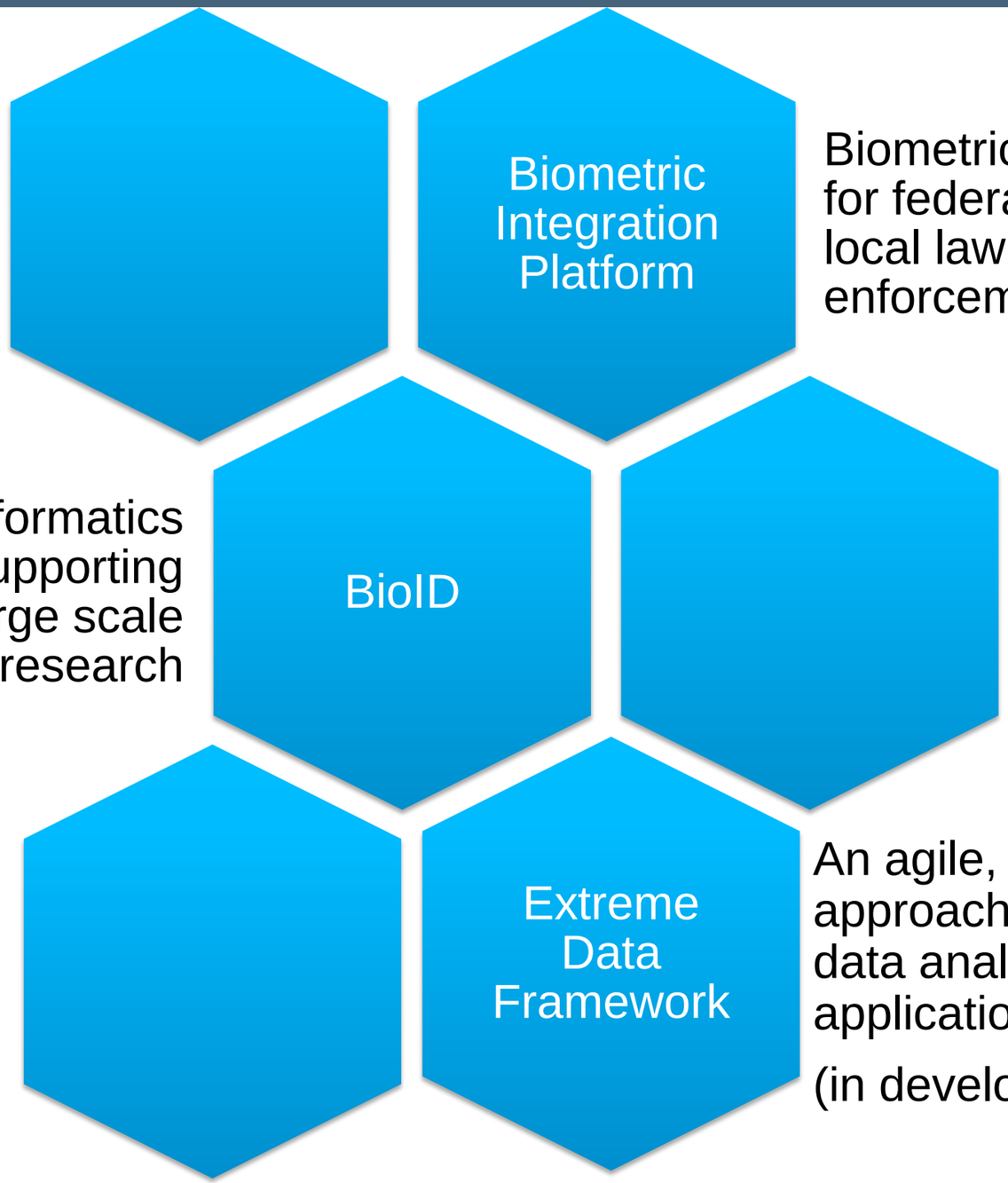
Center for Applied High Performance Computing

HPC value chain



Center for Applied High Performance Computing

Selected Noblis solutions and platforms in the field



Biometric forensics for federal, state, and local law enforcement

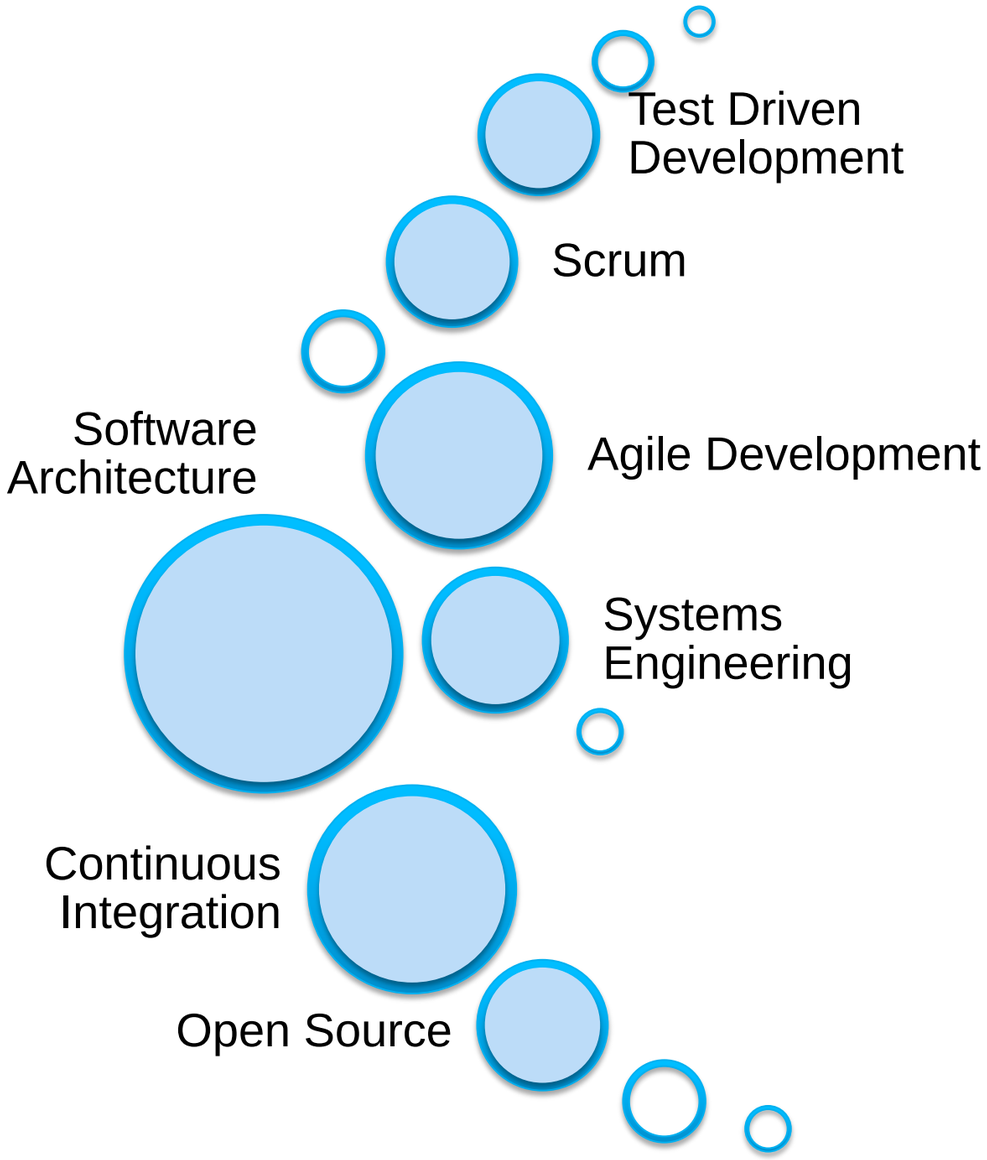
Bioinformatics platform supporting large scale pathogenic research

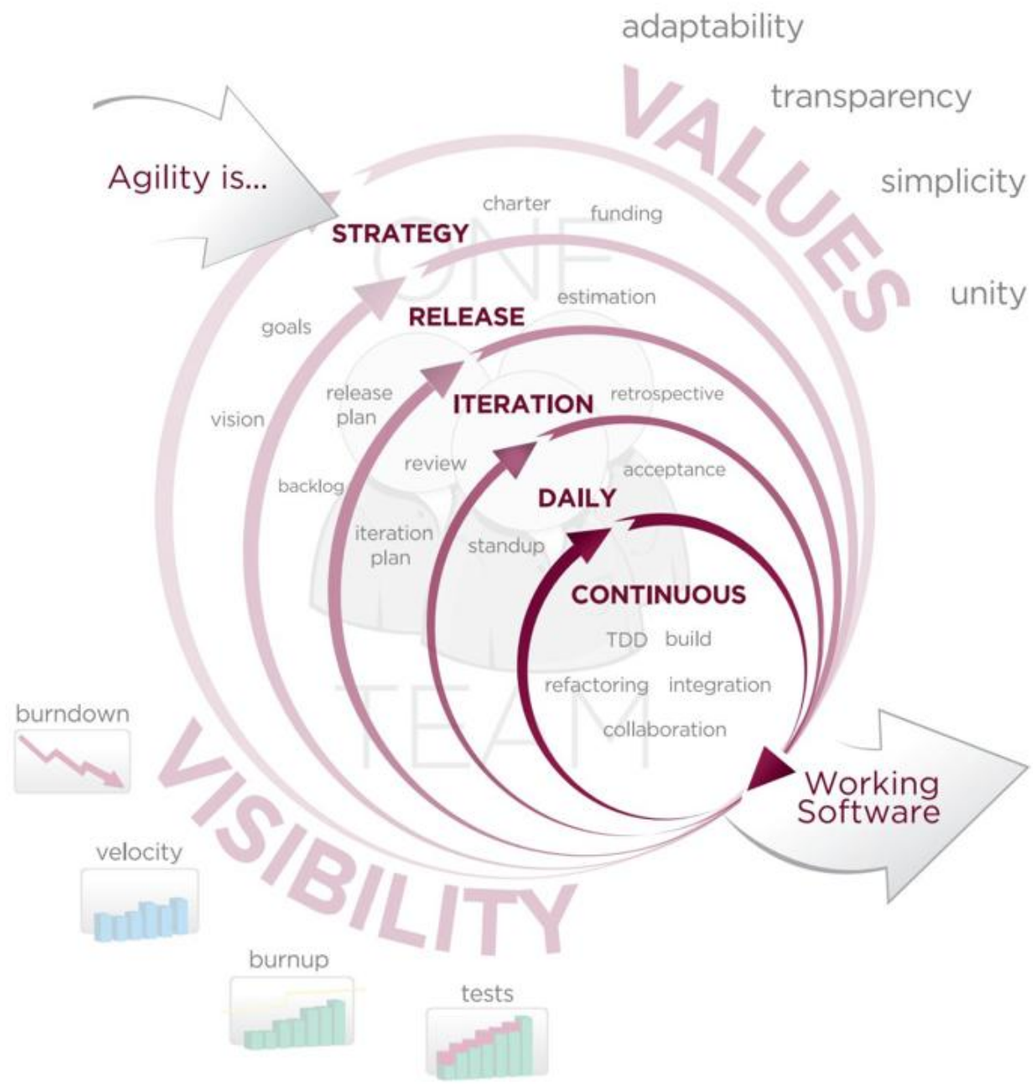
An agile, open source approach to developing big data analytics/semantics applications

(in development)

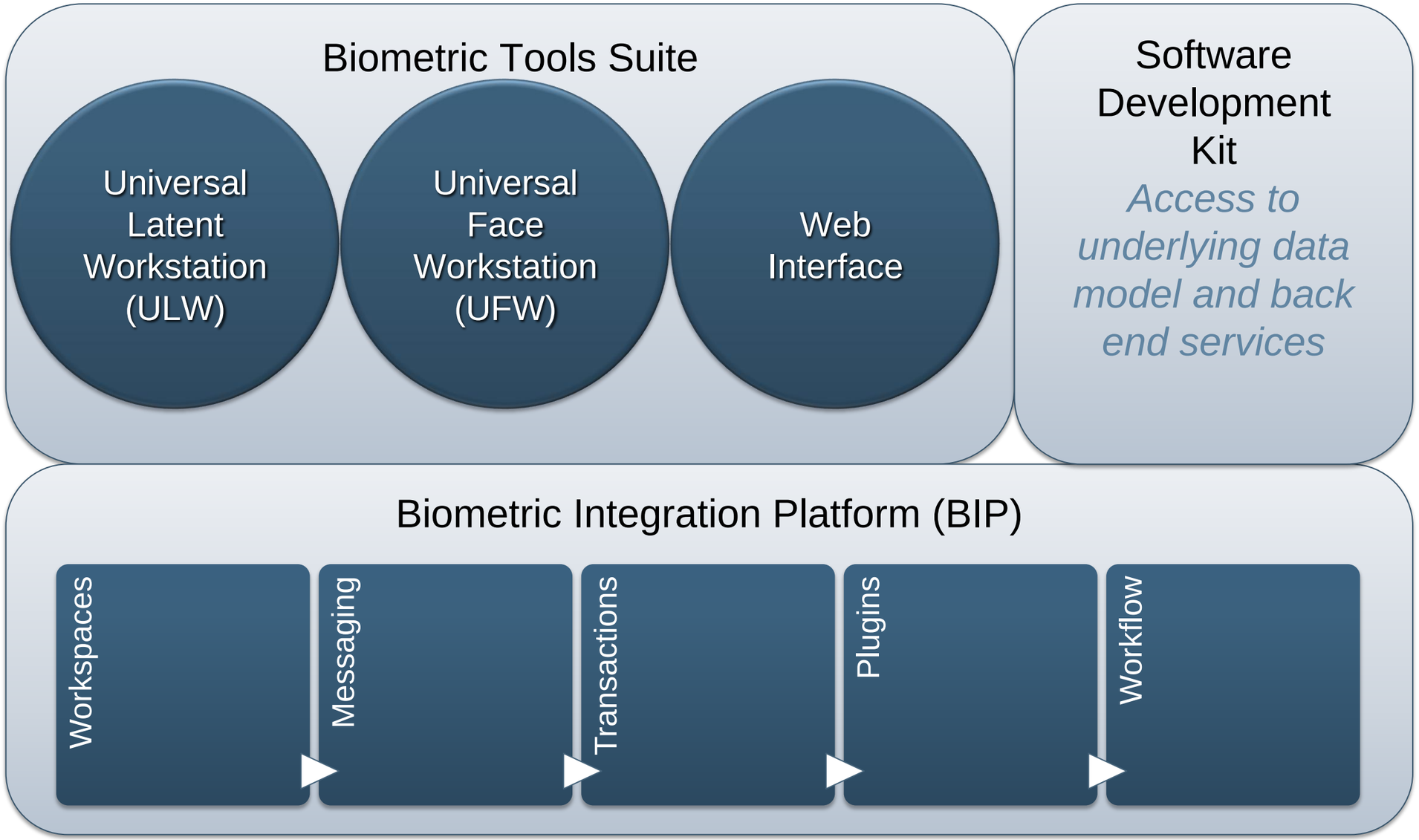
Center for Applied High Performance Computing

Relevant Noblis technical processes and values





A Government Open Source Platform for Biometric Forensics

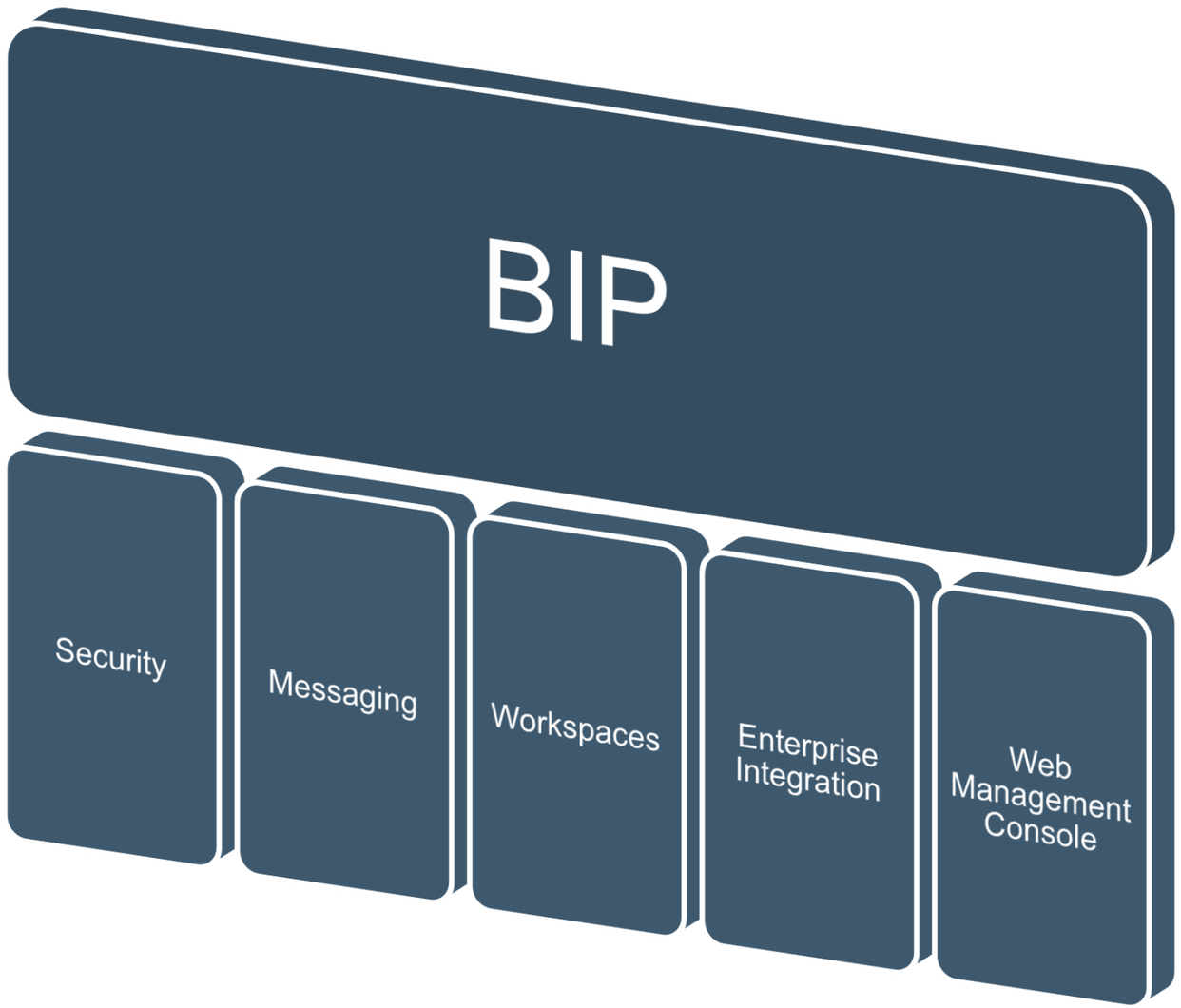


BIP Enables:

- *Large scale deployments*
- *Layered, pluggable security model*
- *Cross-agency collaboration*
- *Integration with back-end systems*
- *An architectural reference point for future procurements*

Biometric Integration Platform

- A Service Oriented Architecture for Biometric Forensics
- Enables large scale deployments of applications with enterprise grade security, connectivity, and management
- Designed to allow costly components to be pluggable
- Built on best of breed open source (LGPL) Java EE components
- Web Services encapsulate critical business functionality within a protected application server
- Web Management Console provides full control over security and platform configuration



Center for Applied High Performance Computing

BIP open source software map

Object-Oriented Programming Language with dynamic language features

Groovy

Server-side AJAX web application development framework + client-side JavaScript engine

vaadin

GRAILS

Groovy-based web application framework

BIP

Distributed, fault-tolerant and schema-free document-oriented database

Apache CouchDB

git

Distributed Revision Control System

Gradle
a better way to build

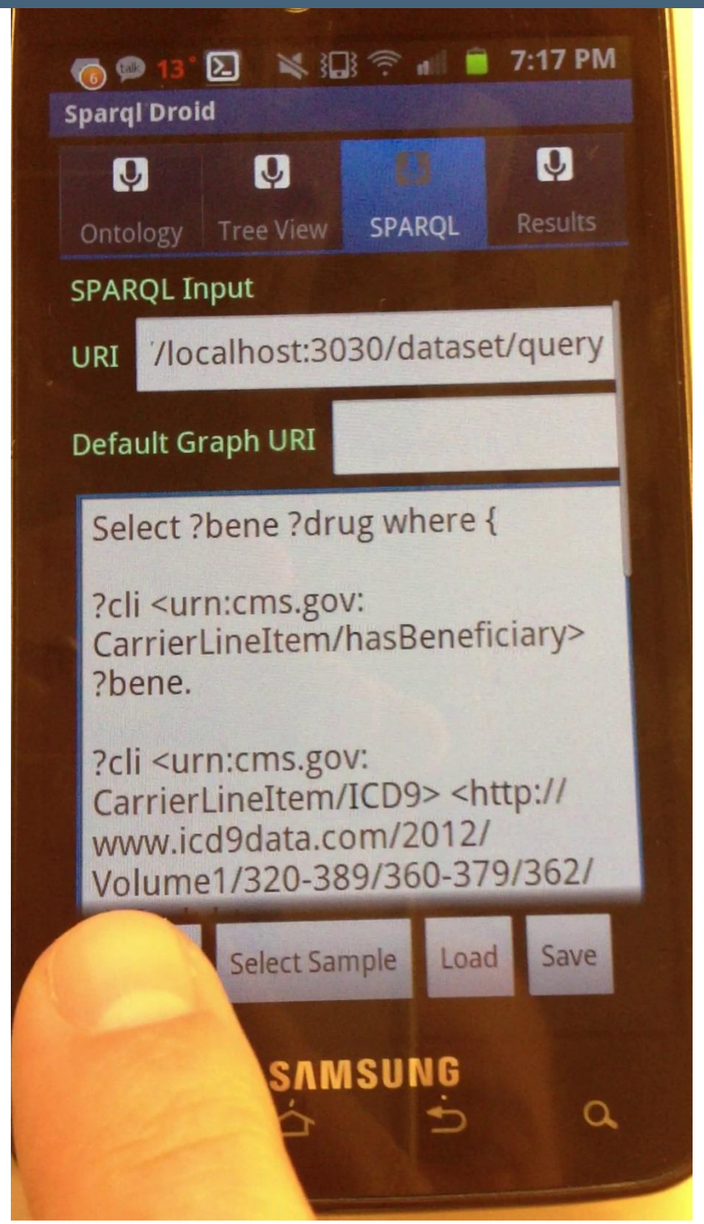
Groovy-based Build System

Component of the deployed product

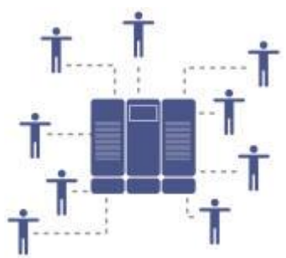
Used during development

Center for Applied High Performance Computing

Mobile Demo



- Center for Medicare & Medicaid Services legacy data set
- Size: 690 millions triples
- Representing about 20 years worth of patient data
- Doctor treating patient for retinal disorder
- What drugs are people diagnosed with “other retinal disorders” taking?
- Capable of querying millions of past relevant prescription in a few seconds



Centralized

Personal

Pervasive

Center for Applied High Performance Computing

An open invitation

- **We continue to seek and welcome additional partners and collaborators**
  - Come visit us in Danville or in Falls Church
  - Noblis CAHPC is interested in collaborating with start-ups, small businesses, researchers, and government agencies to build new solutions
- We are optimistic about the value of semantic integration and graph analytics
- We expect a wave of platform-level software to be created to support this field, and we expect to see many vertical market solutions

Center for Applied High Performance Computing

Will Mitchell, Software Architect
William.Mitchell@noblis.org

Ben Pecheux, Software Architect
benjamin.pecheux@noblis.org

Gil Miller, Noblis CTO
hgmiller@noblis.org

HPC User Forum
April 2012

noblis
For the best of reasons

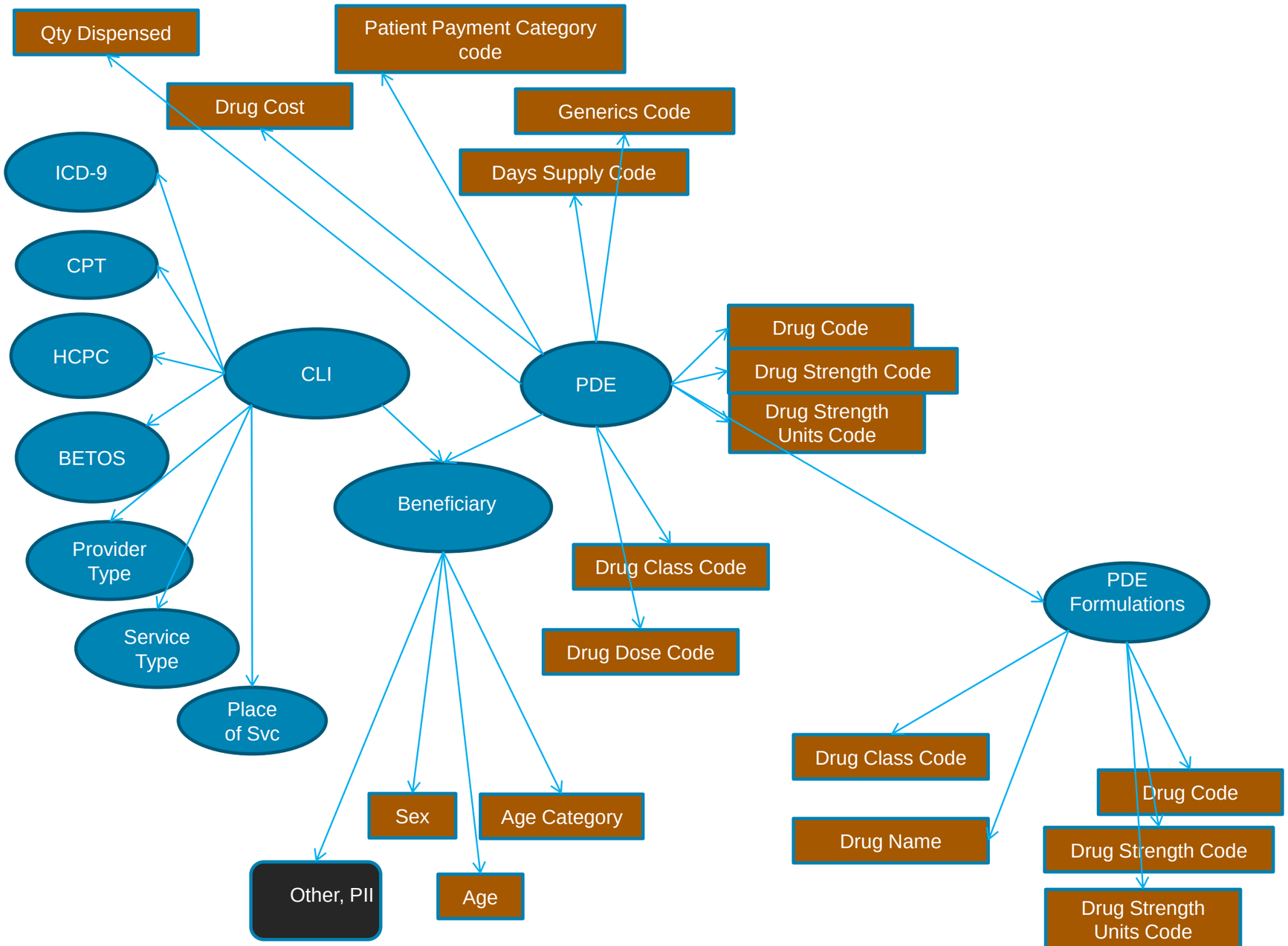
BACKUP SLIDES

CMS Relationship Analytics

Model of Center for Medicare & Medicaid Services (CMS) data

Relationships expressed in dozens of ontologies

Xxx rows
Preprocessing
Load data into uRiKa…



CMS Relationship Analytics: Properties Considered

<urn:cms.gov:PrescriptionDrugEvent/daysSupplyCode>
<urn:cms.gov:CarrierLineItem/ICD9>
<urn:cms.gov:PrescriptionDrugEvent/genericsFlag>
<urn:cms.gov:PrescriptionDrugEvent/PDElineID>
<urn:cms.gov:PrescriptionDrugEvent/formulation>
<urn:cms.gov:CarrierLineItem/dateOfSvc>
<urn:cms.gov:CarrierLineItem/hasProviderType>
<urn:cms.gov:CarrierLineItem/hcpcsPaymtAmt>
rdf:type [http]
<urn:cms.gov:PrescriptionDrugEvent/doseCode>
<urn:cms.gov:Beneficiary/ageCategory>
<urn:cms.gov:CarrierLineItem/hasCMSTypeSvc>
<urn:cms.gov:CarrierLineItem/carLineID>
<urn:cms.gov:PrescriptionDrugEvent/drugCost>
<urn:cms.gov:PrescriptionDrugEvent/drugQtyDispensed>
<urn:cms.gov:Beneficiary/age>
<urn:cms.gov:PrescriptionDrugEvent/patientPayCode>
<urn:cms.gov:PrescriptionDrugEvent/hasBeneficiary>
<urn:cms.gov:CarrierLineItem/hasBeneficiary>
<urn:cms.gov:Beneficiary/gender>
<urn:cms.gov:CarrierLineItem/carLineSvcCnt>
<urn:cms.gov:CarrierLineItem/hasPlaceOfSvc>
<urn:cms.gov:CarrierLineItem/CPT>
<urn:cms.gov:Beneficiary/beneID>
<urn:cms.gov:CarrierLineItem/BETOS>
<urn:cms.gov:CarrierLineItem/HCPC>
<urn:cms.gov:hcpc2010/actionEffectiveDate>
rdfs:domain [http]
<urn:cms.gov:PDEFormulary/drugStrengthCode>
<urn:cms.gov:hcpc2010/shortDescription>
<urn:cms.gov:hcpc2010/codeTerminationDate>
<urn:cms.gov:PDEFormulary/drugStrengthUnitCode>
rdfs:range [http]
dc:description [http]
<urn:ama-assn.org:cpt2008/postTime>
<urn:ama-assn.org:cpt2008/workRVU>
rdfs:subClassOf [http]
rdfs:label [http]
<urn:ama-assn.org:cpt2008/preTime>
<urn:cms.gov:hcpc2010/betos>
<urn:cms.gov:PDEFormulary/drugClassCode>
<urn:ama-assn.org:cpt2008/CPTcode>
<urn:ama-assn.org:cpt2008/totalTime>
<urn:cms.gov:PDEFormulary/drugCodeName>
<urn:ama-assn.org:cpt2008/intraTime>
owl:versionInfo [http]
owl:imports [http]
dc:creator [http]
rdfs:comment [http]
<urn:cms.gov:hcpc2010/code>
<urn:cms.gov:PDEFormulary/drugCode>
<urn:cms.gov:hcpc2010/codeAddDate>

CMS Relationship Analytics: Queries

What drugs are people diagnosed with “other retinal disorders” taking?

```
select  ?bene ?drug where {
?cli <urn:cms.gov:CarrierLineItem/hasBeneficiary> ?bene.
?cli <urn:cms.gov:CarrierLineItem/ICD9> <http://www.icd9data.com/2012/Volume1/320-389/360-379/362/default.htm>.
?pde <urn:cms.gov:PrescriptionDrugEvent/hasBeneficiary> ?bene .
?pde <urn:cms.gov:PrescriptionDrugEvent/formulation> ?form.
?form <urn:cms.gov:PDEFormulary/drugCodeName> ?drug .
} LIMIT 100
```

What diagnoses did people taking Amlodipine (36) have?

```
select ?bene ?diag where {
?cli <urn:cms.gov:CarrierLineItem/hasBeneficiary> ?bene.
?cli <urn:cms.gov:CarrierLineItem/ICD9> ?diag.
?pde <urn:cms.gov:PrescriptionDrugEvent/hasBeneficiary> ?bene .
?pde <urn:cms.gov:PrescriptionDrugEvent/formulation> ?form.
?form <urn:cms.gov:PDEFormulary/drugCode> "36"@en .
} LIMIT 100
```

Cray XMT2 and the uRiKa Appliance

- YarcData is a new subsidiary of Cray
- Cray XMT2 rebranded as the "YarcData uRiKa Big Data Graph Appliance"
- uRiKa software provides:
  - Sparql endpoint
  - Inference engine
  - Gadget-based web UI
  - Enterprise service bus