

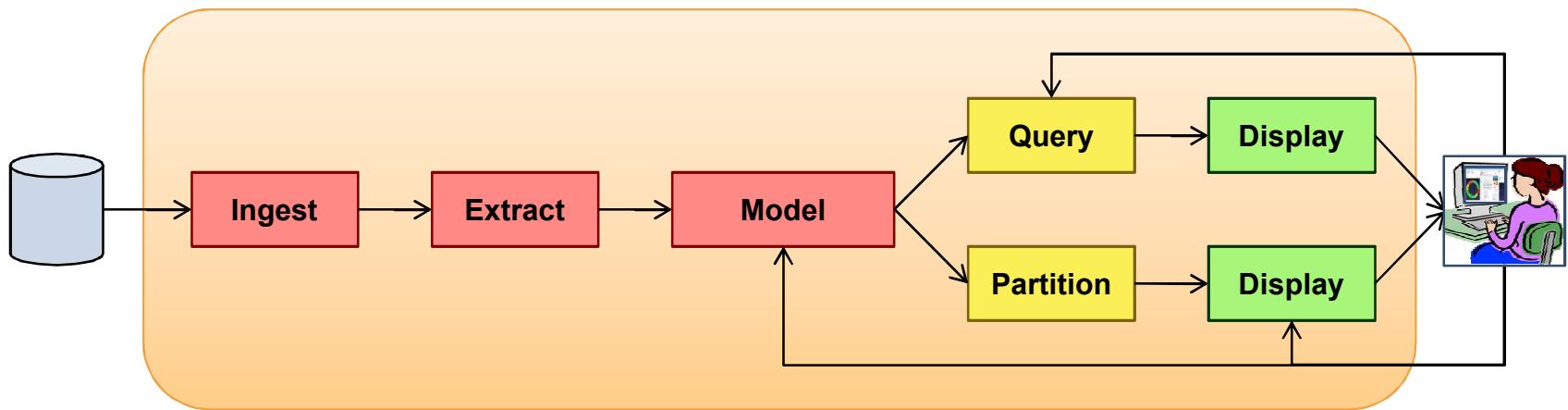
Titan / NGC Impact

August 12, 2009

**Timothy M. Shead
NGC Integration Lead**

What is Titan?

A flexible parallel pipeline architecture, written in C++



Provides a direct mapping between whiteboard and code.

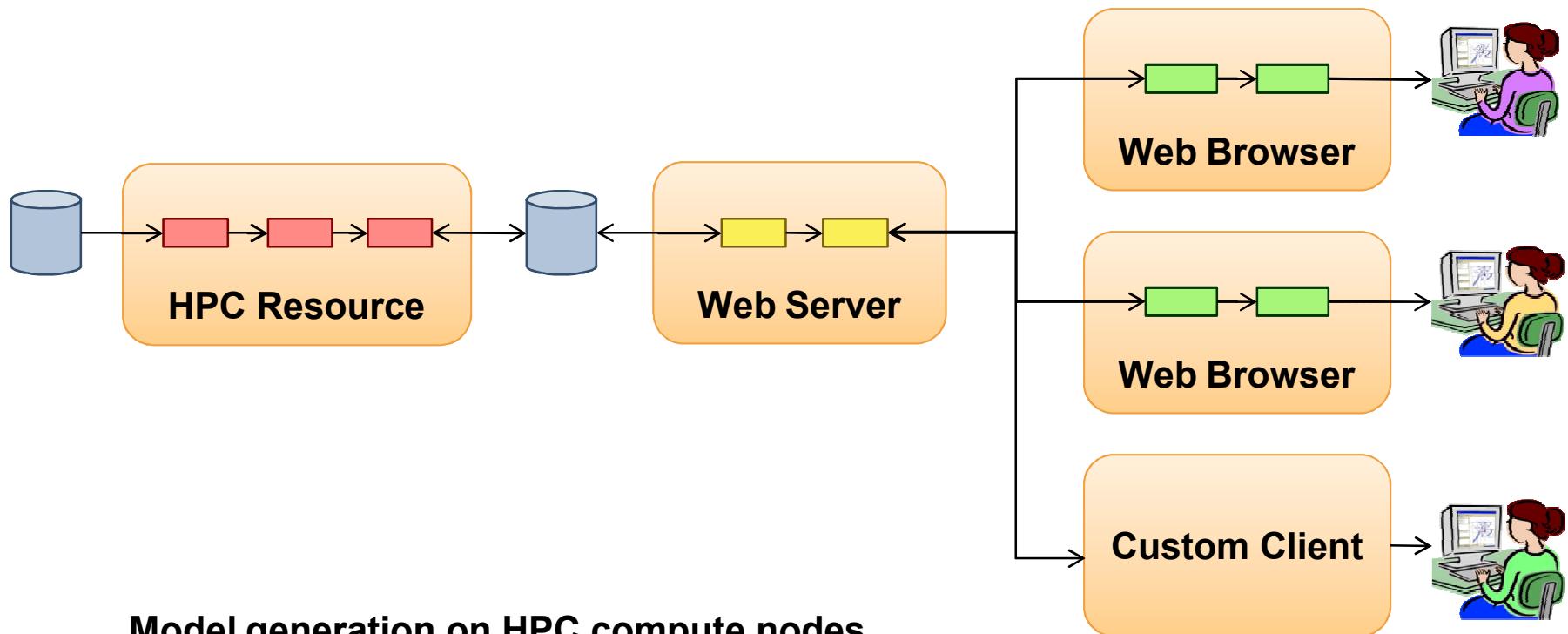
Manages execution:

- Manages the “flow” of data through components.
- In what order are components executed?
- What needs to be recomputed when parameters change?

Serial or parallel

What is Titan?

How the NGC delivers “HPC informatics capabilities that are both *usable* and *useful* to analysts.”



**Model generation on HPC compute nodes,
analysis on a web server, visualization in client.**



What is Titan?

Data Structures

- Table
- Tree
- DAG
- Directed Graph
- Undirected Graph
- Sparse N-way Array
- Dense N-way Array
- Unicode Text

Database Drivers

- MySQL
- Postgres
- Oracle
- SQLite
- ODBC
- Netezza
- **Berkeley DB**
- Hypertable

MTGL Algorithms

- Community Finder
- ST Search
- CSG Search
- Temporal Search
- **wCNM**
- Subgraph Isomorphism

Readers

- Dimacs
- DOT
- GXL
- Chaco
- XML
- Tulip
- DelimitedText
- Unicode Delimited Text
- FixedWidth
- ISI, RIS
- Palantir XML
- **Matrix Market**
- Protobuf
- **HDF5 (network packets)**
- OWL (ontologies)
- S-Expressions
- XML-RPC

Linear Algebra / Tensors

- TPP / PARAFAC
- Trilinos Integration

BGL Graph Algorithms

- Breadth First Search
- Connected Components
- Biconnected Components
- Brandes Centrality

Parallel Statistics Algorithms

- Descriptive
- Order
- Correlative
- Contingency

Clustering

- Partitioning
- Hierarchical
- Soft Clusters

PBGL Integration

MATLAB Integration

R Integration

Linear-Time Graph Layouts

- GSpace
- Hierarchical
- Clustered
- Three tree-based variants

Text Analysis

- Latent Semantic Analysis
- Latent Dirichlet Allocation

Multiple View Types

- Render (3D)
- Graph
- Hierarchical Graph
- Tree
- Treemap
- Georeferenced
- Parallel Coordinates
- Radial Tree Ring
- **Protovis (dozens of new visualizations)**

Multiple Platforms / Languages

- Windows, Linux, OSX, HPC
- Write components in C++
- Use with C++, TCL, Python, Java, .NET, COM
- **Apache Modules**
- **HTML + JavaScript**



What is Titan?

A set of libraries you can use today ...

Home Page:

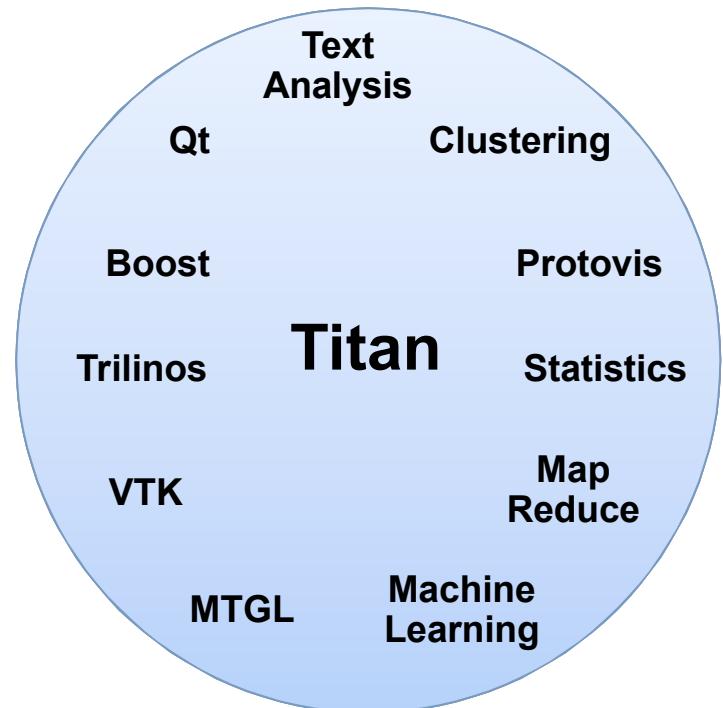
<http://titan.sandia.gov>

Public Git Repository:

<git://public.kitware.com/titan.git>

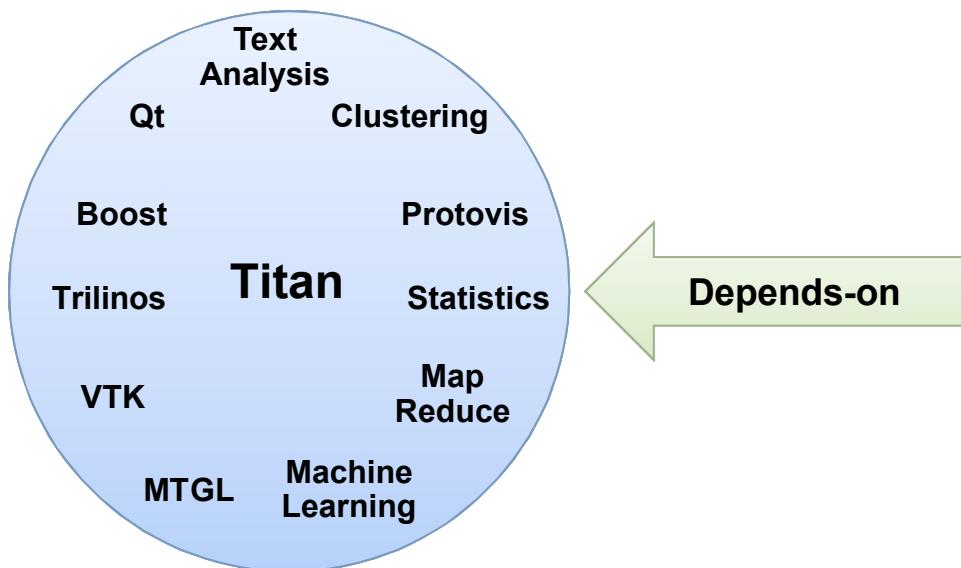
Reference Documentation:

<http://public.kitware.com/titan-docs>





Sidebar: Using Titan within Sensitive Domains





Projects Using Titan

- **Network Grand Challenge** (Hendrickson / Kegelmeyer)
- **La Luz - SOCOM IATA** (Benz / Johnson)
- **HPC PubMed** (Wilson / Shepherd / Otahal / 5900)
- **MegaTux Network Emulitics** (Minnich)
- **NA-22 Satellite Image Analysis for Non-Proliferation** (Watson)
- **Streaming Data for Cybersecurity** (Plimpton)
- **ParaSpace Simulation Ensemble Characterization** (Crossno / Martin / Hunt / Shead)
- NIH SBIR: "Flexible information visualization and analysis platform for biomedical research" (**Kitware**)
- Army SBIR: "An open framework for analyzing mobile ad hoc networks" (**Kitware**)
- DARPA SBIR: "An open, adaptive geospatial visual information system" (**Kitware**)
- Navy STTR: "Using stylistic topic models to detect deception through unusual linguistic activity" (**Kitware**)
- DOE SBIR: "A scalable visual system for proliferation analysis" (**Kitware**)
- ParaText Scalable Text Ingestion (**Dunlavy / Crossno / Shead / Stanton**)
- CARGIO: Hard drive Analysis (**Kegelmeyer / Desapio**)
- Nuclear Materials Attribution (**Crossno**)
- OVIS Cluster Monitoring (**Brandt**)
- ThreatView (**Summers / Linebarger / Kerr / Wolfenbarger**)
- Graph-Based Informatics for Counter-Terrorism LDRD (**Wylie**)

NGC Prototype 2 Recap

Titan.P2.0.9.8 (alpha) - C:/data/IIR_20_no_empty

Document Clusters (View 1)

Clusters **Search Results** **Selected Documents**

Clusters

- 1:0.588:/tritone/names/internet
- Header_216_2014_93_Body_222_0055_97.txt
- 7:0.893:/submarine/ammonia/pharmacycards
- 3:0.669:/concerns/stated/pharmacycards
- 4:0.654:/concerns/stated/basic/victims
- 5:0.854:/concerns/stated/political
- IIR 2 222 0006 96.txt
- IIR 2 222 0034 97.txt
- 8:0.917:/billion/basic/offshore
- 2:0.647:/political/seekers/asylum
- 0:0.902:/victims/concerns/stated
- IIR 6 842 0030 92.txt
- IIR 6 842 0259 92.txt

Document Text (View 5)

Filename: Header_216_2014_93_Body_222_0055_97.txt Find: Go + 100%

SERIAL: IIR 2 216 2014 93.

UNCLAS

SUBJECT: IIR 2 216 2014 93/CZ MILITARY INTELLIGENCE - FUTURE PLANS.

SOURCE: AN OPEN SOURCE ELECTRONIC INFORMATION SERVICE WHICH PROVIDES DIRECT ACCESS TO FOREIGN COMPUTER NETWORKS, DATABASES AND ON-LINE INFORMATION SERVICES WORLDWIDE. SOURCE RELIABILITY CANNOT BE DETERMINED. -

DIG

- R 101341Z NOV 92
- R 021732Z NOV 92
- R 301300Z OCT 92

COUNTRIES

- CZECHOSLOVAKIA (CZ) (4)
- CHINA (3)

SERIAL

- SERIAL: IIR 2 216 2014 93.

WARNINGS

- THIS IS AN INFORMATION REPORT, NOT FINALLY EVALUATED INTELLIGENCE. REPORT CLASSIFIED U N C L A S S I F I E D. -
- DEPARTMENT OF DEFENSE
- REPORT CLASSIFIED U N C L A S S I F I E D.

Message Text

TEXT: 1. THE CHINESE GOVERNMENT THIS WEEK INTENDED TO PUT AN END TO THE SPECULATIVE DOMAIN NAMES AND PROFITEERING FROM THE RES OFFICIALS SAID. IN ADDITION, THE GOVERNMENT INTERNET INFORMATION CENTER, WHICH WILL BE MANAGEMENT, OPERATION AND SERVICE OF THE INTERNET, WILL BE THE ONLY ORGANIZATION IN THE COUNTRY TO OVERSEE THE REGISTRATION OF DOMAIN NAMES, ACCORDING TO THE STATE COUNCIL INFORMATION ADVANCEMENT, WHICH OVERSEES THE NETWORKS WITH THE INTERNET.

2. THE CENTER, WHICH FALLS UNDER THE CHINESE GOVERNMENT, WILL BE THE ONLY ORGANIZATION IN THE COUNTRY TO OVERSEE THE REGISTRATION OF DOMAIN NAMES, ACCORDING TO THE STATE COUNCIL INFORMATION ADVANCEMENT, WHICH OVERSEES THE NETWORKS WITH THE INTERNET.

3. THE NEW REGULATIONS SPECIFY THE STRUCTURE OF THE DOMAIN NAMES AND THE PROCEDURES BY WHICH THE NAMES ARE APPLIED FOR AND APPROVED. APPLICANTS WILL

REGISTERED ORGANIZATIONS WITH THE ABILITY TO REGISTER DOMAINS WITH THE CHINESE GOVERNMENT.

4. CHINA ESTABLISHED ITS OWN DOMAIN NAME MONTH AFTER IT WAS HOOKED UP TO THE INTERNET. MORE THAN 50,000 COMPUTERS IN CHINA ARE CONNECTED TO THE INTERNET TO DATE, AND THE COUNTRY IS PROMOTING THE DEVELOPMENT OF THE COUNTRY'S BUSINESS AND CULTURAL ESTABLISHMENTS. THE NAMES AND THE SCALPING OF THOSE NAMES HAVE BECOME A MAJOR SOURCE OF INCOME FOR THE CHINESE GOVERNMENT.

COMMENT: (SOOTH MI BDE HQ COMMENT) - THIS INTELLIGENCE INFORMATION REPORT PRODUCED BY THE CHINESE GOVERNMENT IS MEANT TO MEET THE CHINESE GOVERNMENT'S THEATER INTELLIGENCE REQUIREMENTS.

Entities (View 2)

Filter:

- CITY
- COUNTRY
- ORGANIZATION
- PERSON
- NATIONAL MIDDLE SCHOOL
- NATIONAL PEOPLE (1)
- NAVAL HOSPITAL (1)
- NAVY (1)
- NAVY REPRESENTATIVE C
- NEAREST ADMINISTRATIVE C
- NEW CHINA D
- NEWS AGENCY (1)

Entity Hotlist (View 3)

CUSTOM

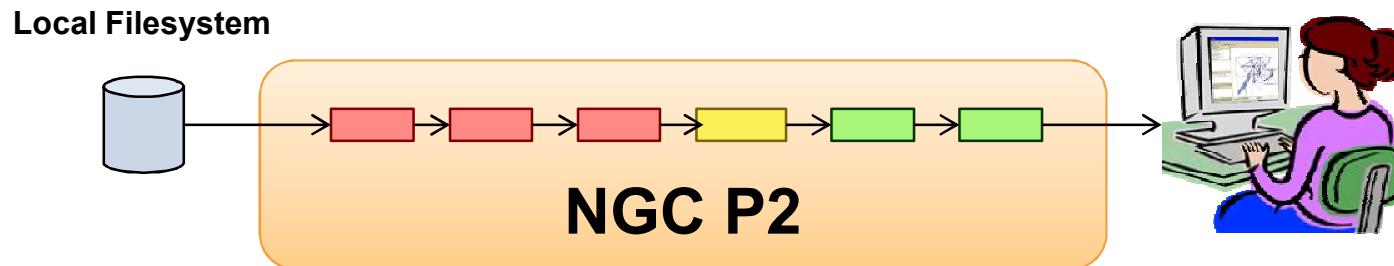
- HEALTH MINISTRY
- THAILAND DEVELOPMENT RESEARCH INSTITUTE (TRDI) STATES
- YALA HOSPITAL
- NAVAL HOSPITAL

Entity Map (View 4)

Graph View Table View



P2 Pipeline



Serial All The Way: put the entire pipeline in one process.
Zero administration, easy to use, online computation, small data.

LaLuz – SOCOM IATA

- Using an “evolved P2” to address undisclosed project domain.
- An example of building sensitive applications on Titan.

Here we pulled in a social network from Palantir. We have a "Red" network and a "Green" network.

The screenshot displays the LaLuz application interface, which includes several panels:

- Document Clusters (View 1):** Shows a tree view of document clusters, including "Security Bali bombers", "U.S. claims SE Asia terrorism progress", "What is Jemaah Islamiyah", "girls/schools/mortenson", "Fighting Terrorism Through Education", "Greg Mortenson", "It Takes a School", "Military Finds Unlikely Advisor", "School of Dreams", "Schools for Girls", "Teacher Can We Leave Now", "asyr/bakar/qaeda", "J1 military chief in Mindanao", "Profile- Abu Bakar Ba'syir", "Profile- Jemaah Islamiyah", "Students Build Schools", "Terrorist Groups in Southeast Asia", "Thailand Islamist Insurgency", "The al-Qaeda of Southeast Asia", "taiba/lashkar/india", and "mumba/lashkar/taiba".
- Document Text (View 5):** Displays a detailed text document about the September 11 terror attacks planned in part by al-Qaeda operatives in Malaysia.
- Entities (View 2):** Lists entities categorized as ORGANIZATION (BAGHLAN, BAGMARA, BAHARAK, BAHAWALPUR, BAHIRAHAT, BAHRAIN, BAJAUR, BALAKRISHNAN, BALAKANDI, BALKH, BALOCHISTAN, BALTISTAN, BALUCHI, BALUCHISTAN), PERSON (CHRISTOPHER KOLEND, GREG MORTENSON, HAFIZ MUHAMMAD SAEED, JEAN HOERNI, LASHKAR-E TOIBA), MISC (None), and LOCATION (None).
- Entity Hotlist (View 7):** Lists entities including CENTRAL ASIA INSTITUTE, CHRISTOPHER KOLEND, GREG MORTENSON, HAFIZ MUHAMMAD SAEED, JEAN HOERNI, and LASHKAR-E TOIBA.
- Hotlist Map (View 4):** A network graph showing connections between entities. A red arrow points to "LASHKAR-E TOIBA", and a green arrow points to "HAFIZ MUHAMMAD SAEED".
- Corpus Maps (View 3):** A circular map showing the distribution of terms across various documents.
- Bottom Panel:** Contains buttons for "Cluster View" and "Graph View".

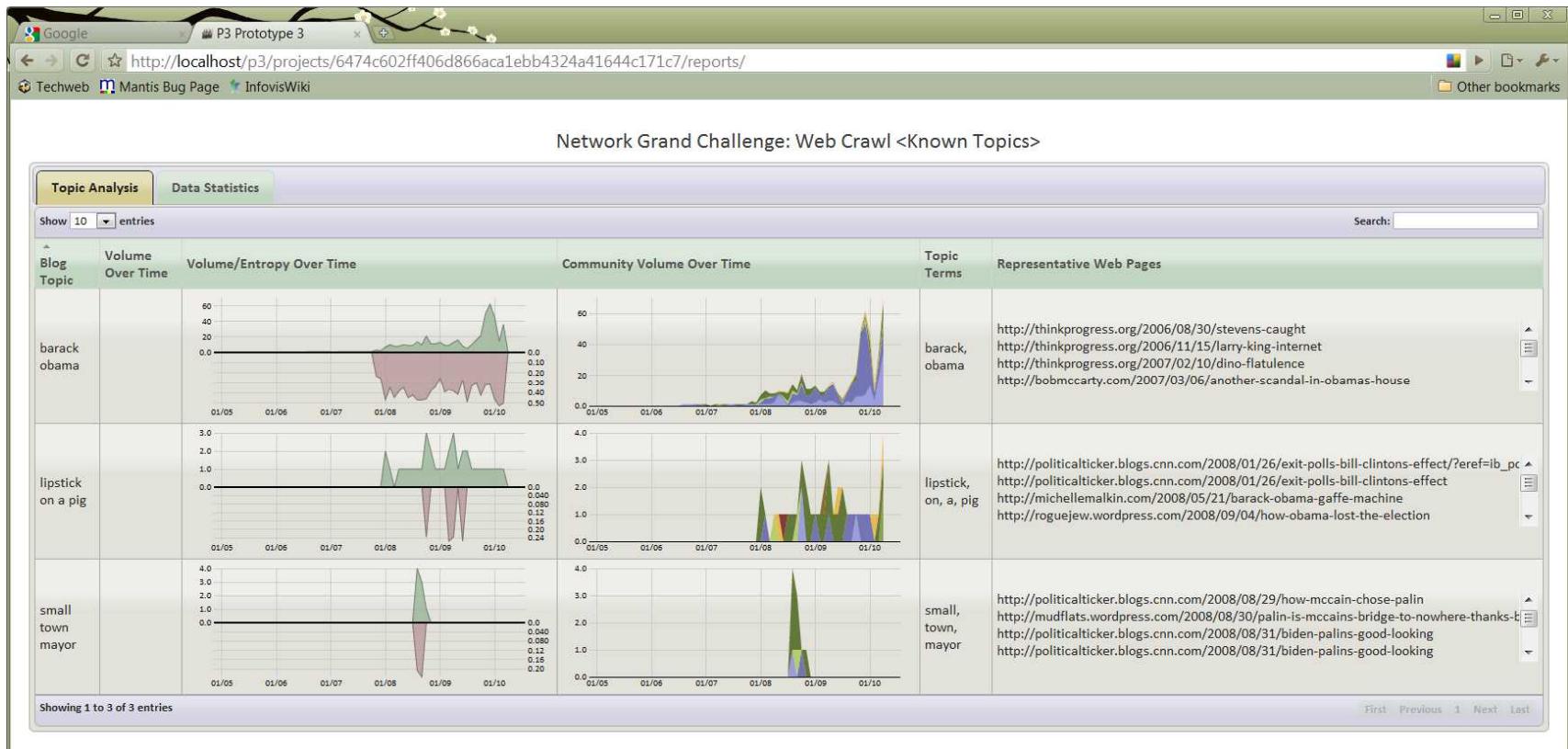


A modest P2 endorsement ...

“P2 has been used as a test case on real data which led to NEW leads and analysis of Social Networking activity ... further answers were derived when new development capabilities were coordinated with 5900 for P2”

- Bryan Ingram, SNL

Network Grand Challenge - P3

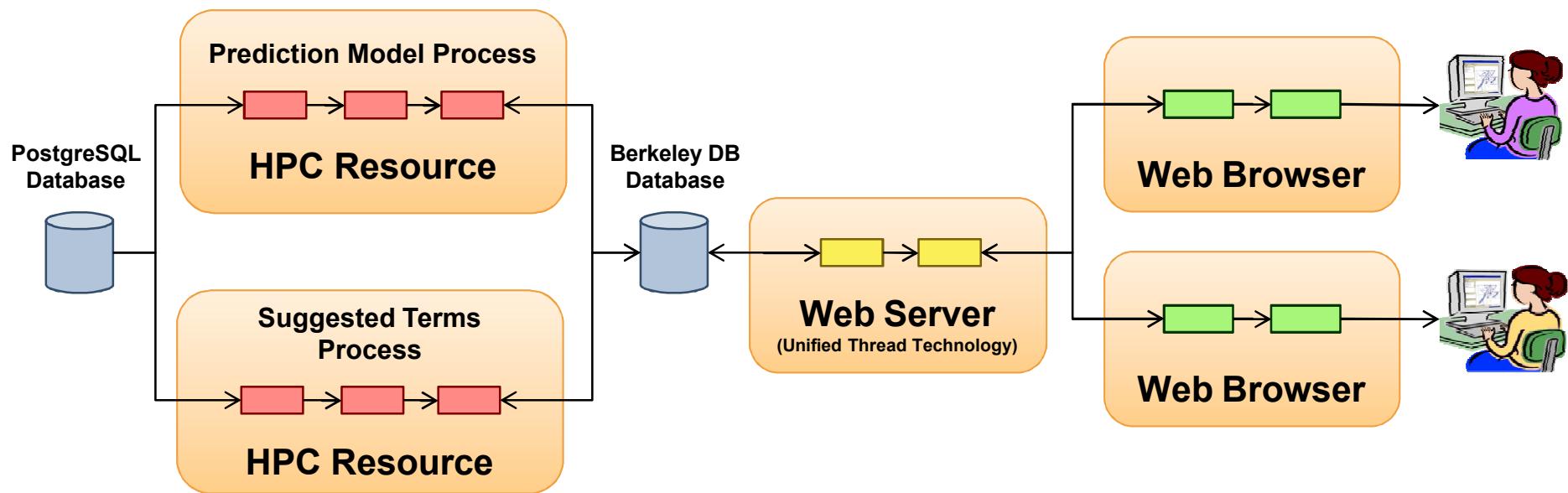




Network Grand Challenge - P3

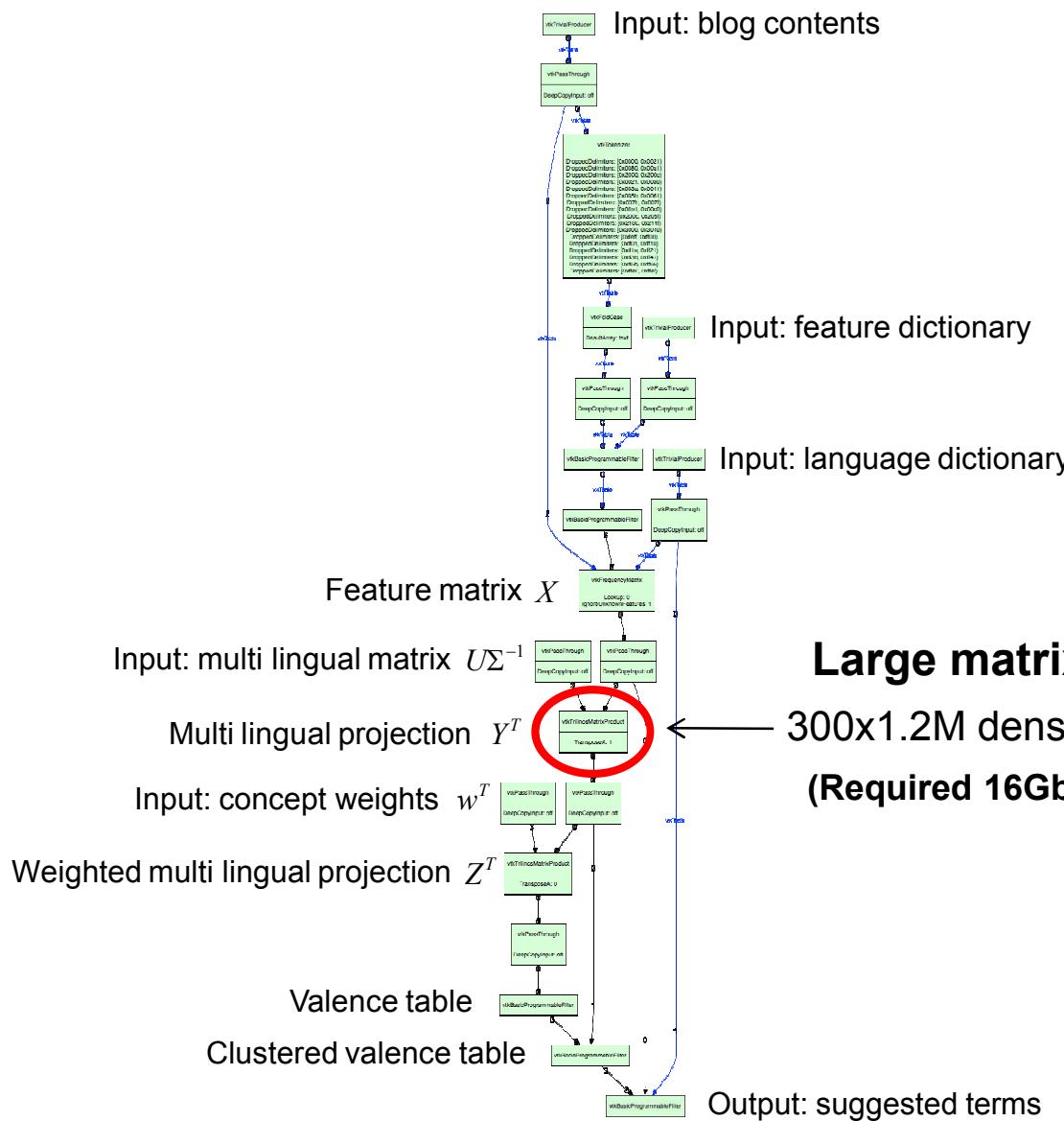
Lovely Video Here

P3 Pipeline



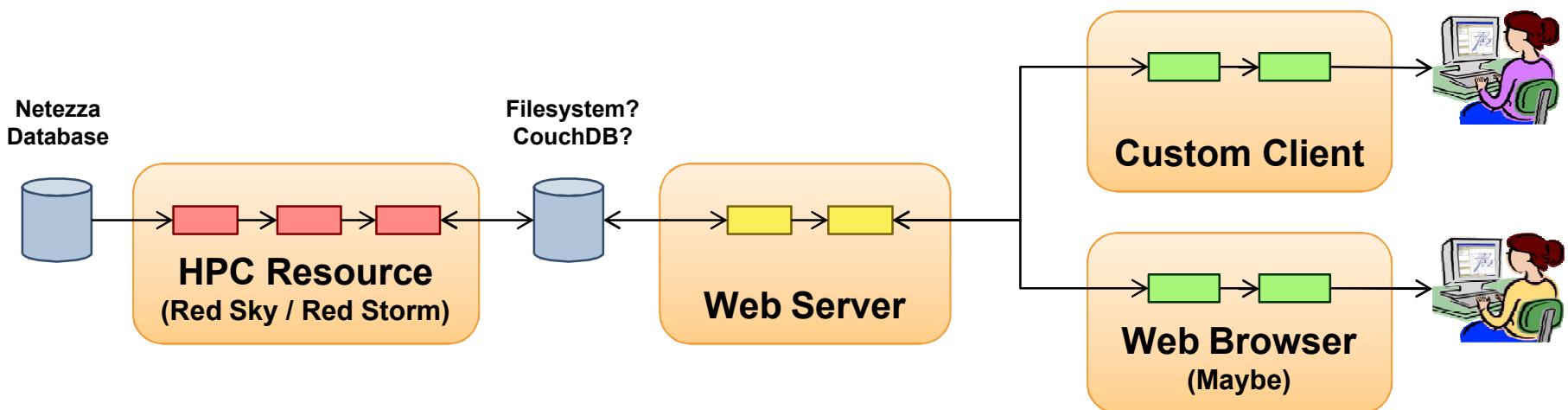
Model generation / analysis on HPC compute nodes, visualization in client.
Modest administration, modest data, mostly offline computation.

P3 Pipeline Detail: Suggested Terms



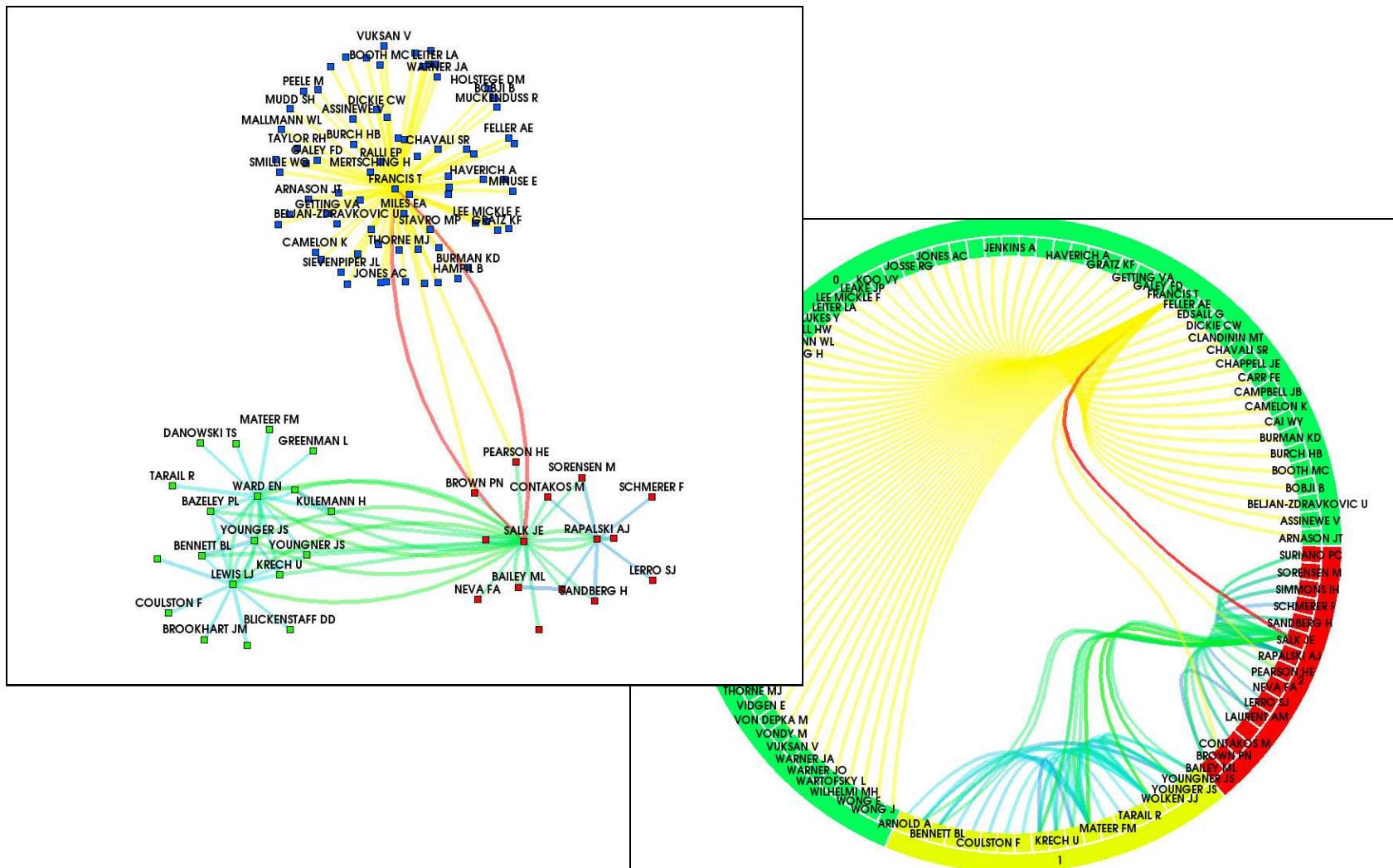
HPC PubMed Pipeline

- Using Netezza, LDA, and other text analysis techniques.
- Benefiting from NGC multi-tier experience.



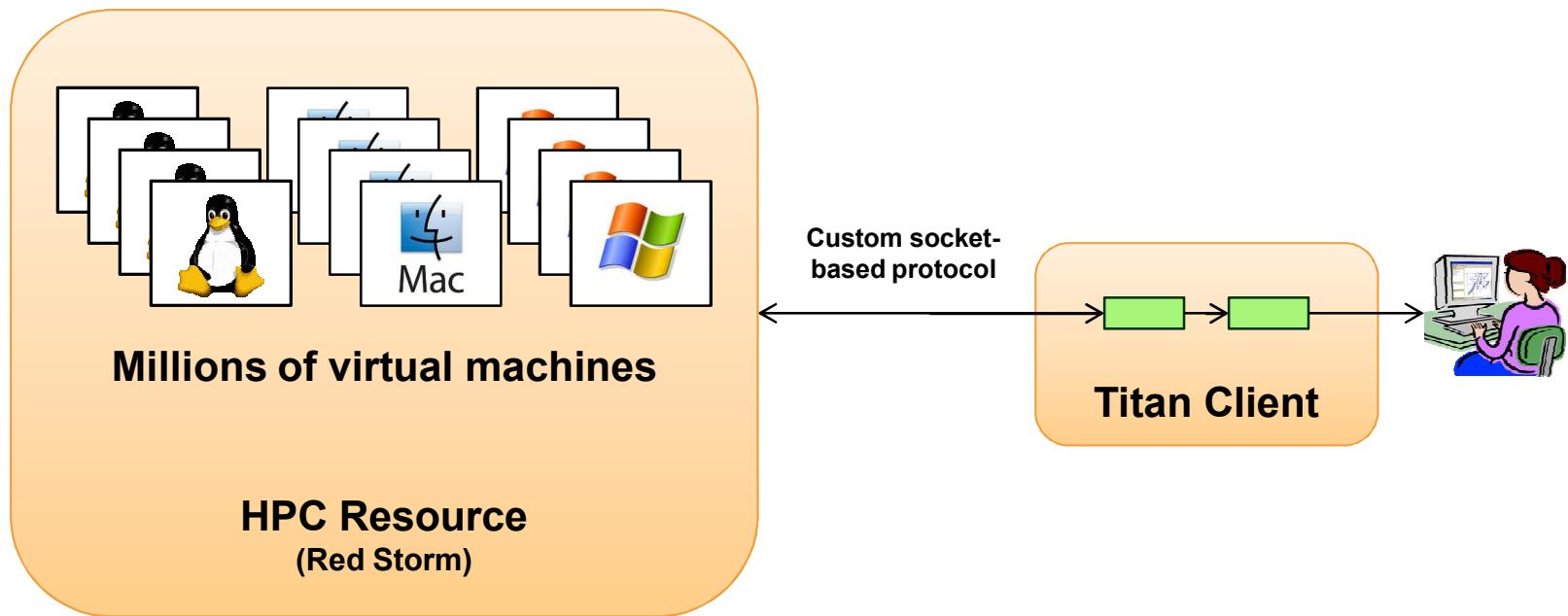
Model generation on HPC compute nodes, analysis on the web server, visualization in client.

HPC PubMed Visualizations



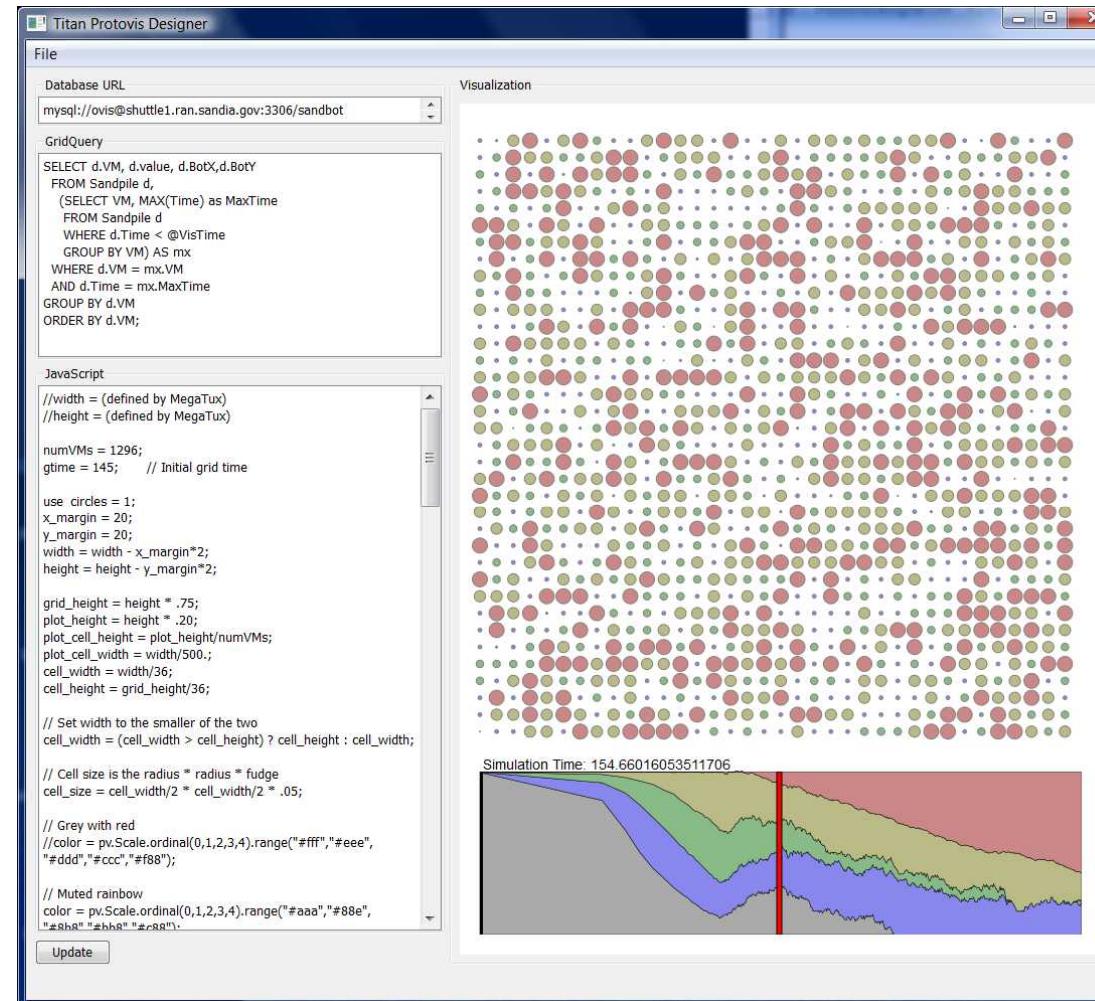
Megatux Network Emulitics

Utilizing the latest in lightweight virtualization techniques, investigate methods to configure and collect data from emulations employing millions of VMs on HPC clusters.



Live, streaming data produced on HPC compute nodes, analysis and visualization in client.

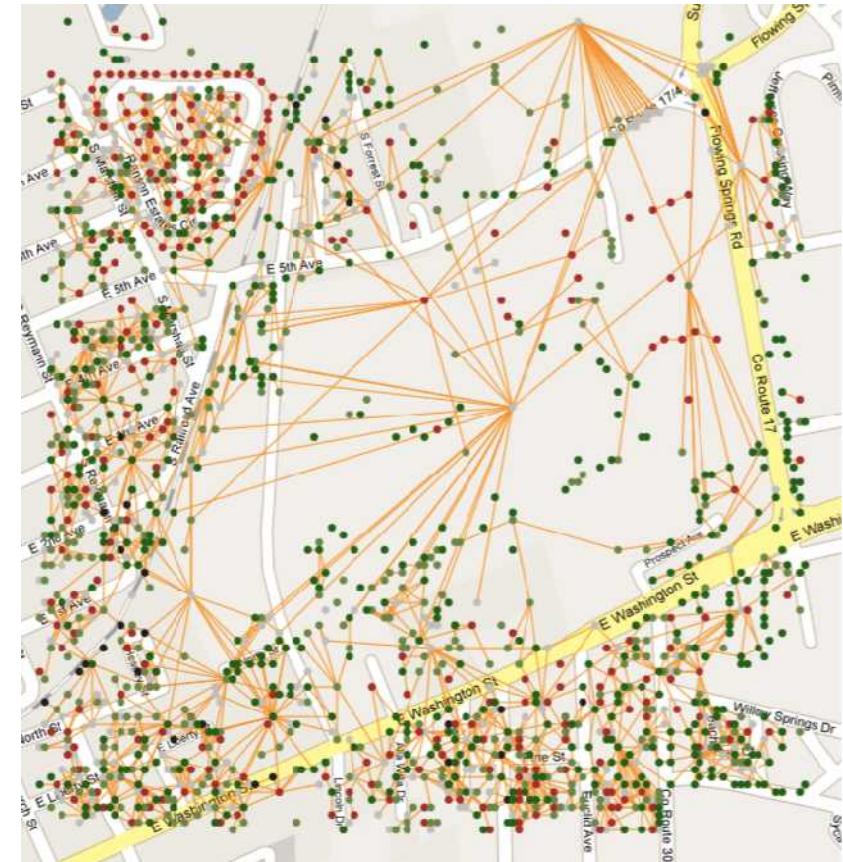
Megatux Network Emulitics



Prototype Megatux client using Protovis

NA-22 Satellite Image Analysis

A graph-based approach to image feature analysis.



NA-22 Satellite Image Analysis

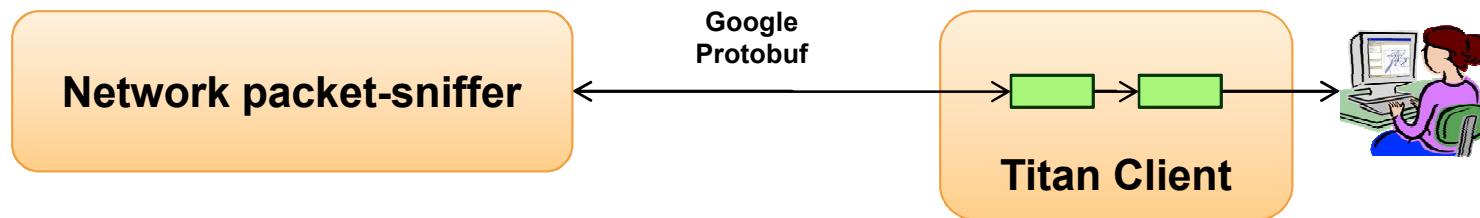
- Graph communities identified using wCNM, converted to KML polygons
- Size proportional to feature size, color map based on community ID



Streaming Data for Cybersecurity

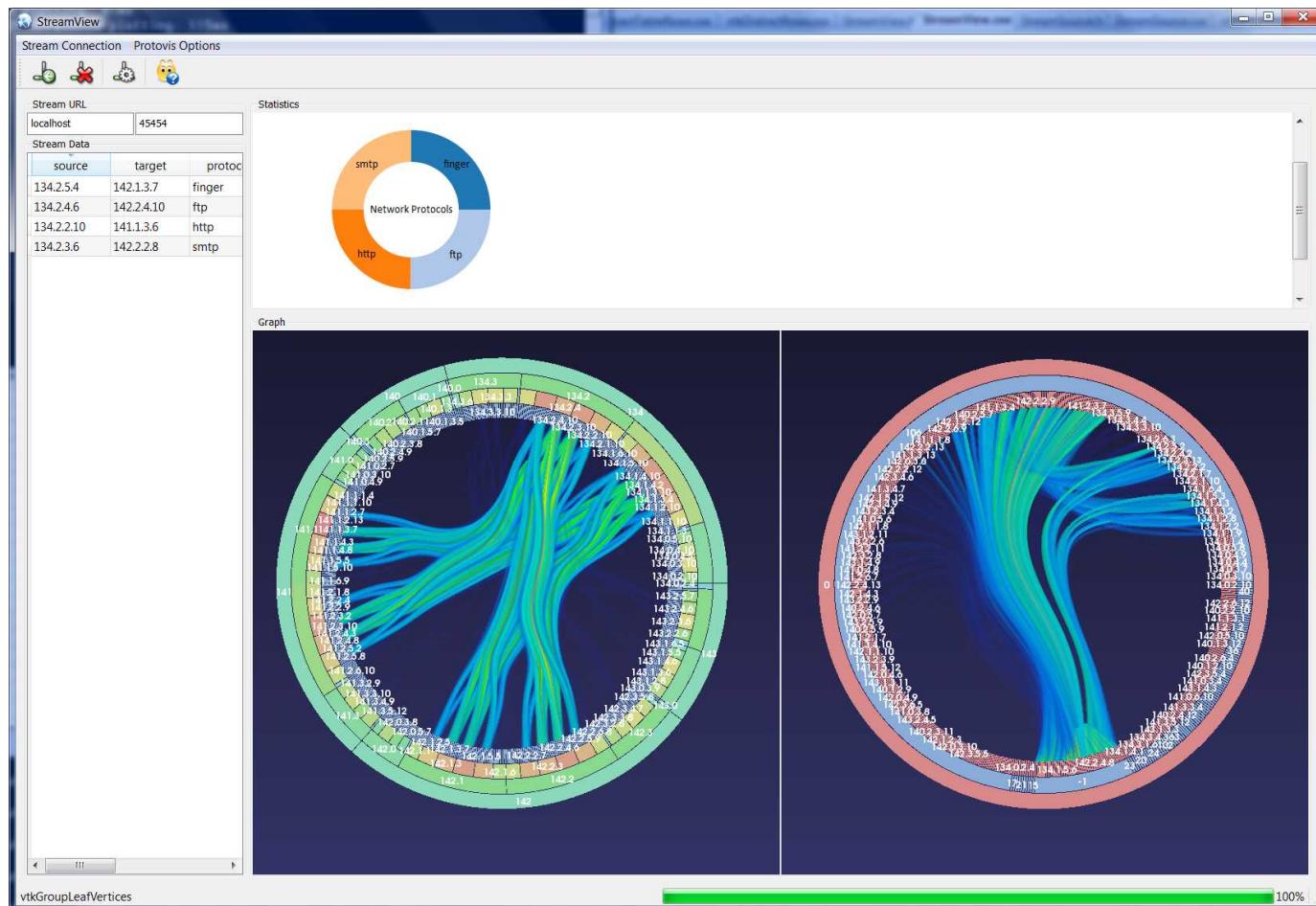
Developing tools and techniques for realtime packet analysis.

- Titan streaming pipeline executive.
- Task-based multithreading.
- 15k packets / sec real-time network data.



Live, streaming data collected by packet-sniffer hardware,
analysis and visualization in client.

Streaming Data for Cybersecurity

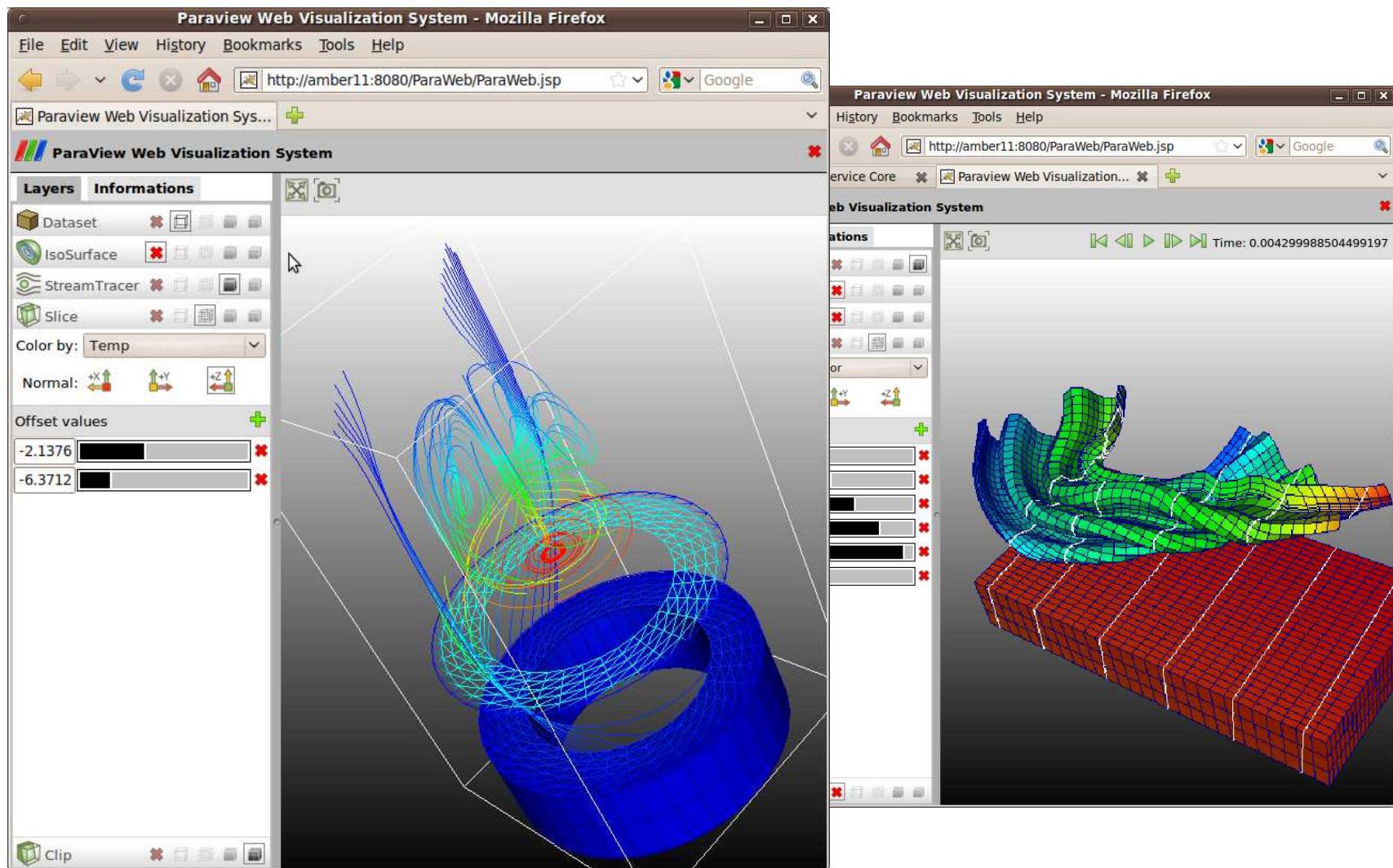


Prototype streaming Titan client
Rendered and Provis views.

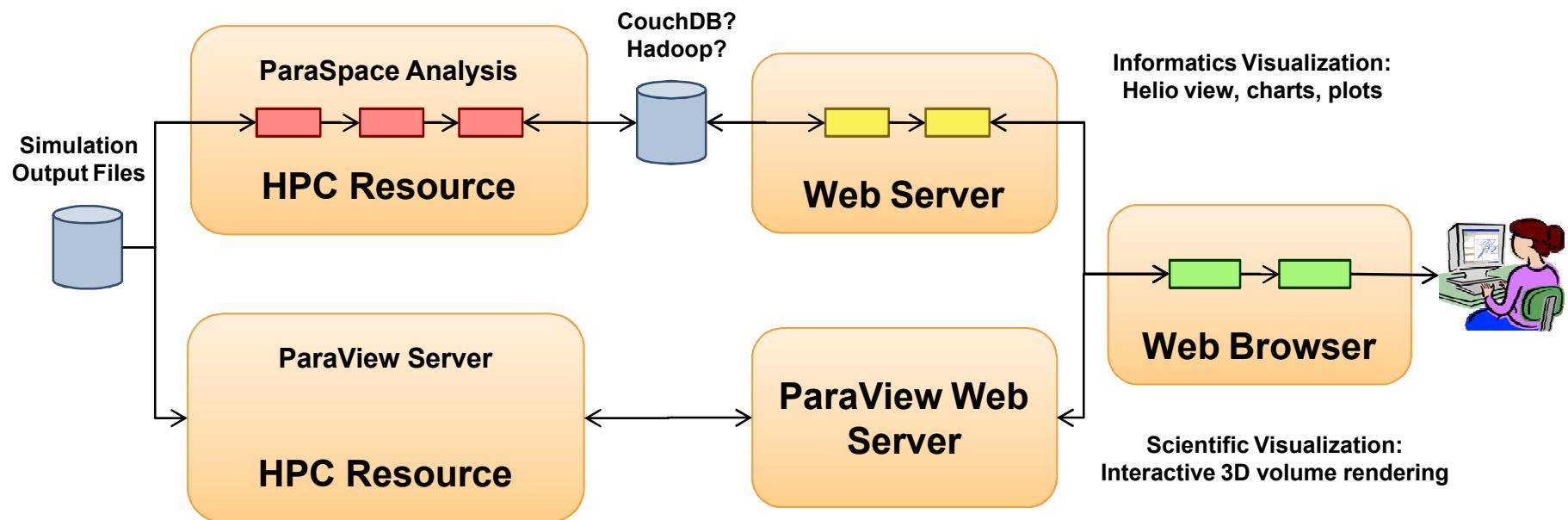
ParaSpace



ParaSpace



ParaSpace Pipeline



A combination of rendered and informatics views, delivered to the browser.
Modest administration, huge data, mostly offline computation.



Titan / NGC Impact Summary

- **Titan repository makes a substantial set of NGC techniques immediately available for researchers and application developers.**
- **Multiple Sandia IC customers have committed resources to develop P2.**
- **The depth and breadth of Titan capabilities are being used by multiple internal and external research projects.**