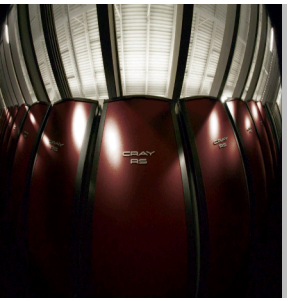
SAND2013-3322C



Hey, You got your DWA in my HPC!

Experiences Integrating Analytics and HPC

Approved for Public Release: SAND2013-XXXX

Ron A. Oldfield

Scalable System Software
Sandia National Laboratories
Albuquerque, NM, USA

Salishan Conference on High-Speed Computing
April 2013

Sandia National Laboratories

Exceptional service in the national interest

U.S. DEPARTMENT OF ENERGY

NNSA

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Why HPC and Analytics?

Sandia National Laboratories

1. Increasing desire to use HPC for analytics
   - Process ***massive amounts of data***
     - Current approaches cull data before processing
     - HPC could identify global relationships in data
     - Time-series analysis to identify patterns (requires large time windows)
   - Some problems have strong ***compute requirements***
     - Eigensolves, LSA, LMSA (lots of matrix multiplies)
     - Graph algorithms
   - Some problems have short ***time-to-solution*** requirements
     - Short response-time is critical
   - National security interest

2. Increasing desire to use DWA in HPC-app workflow
   - Post-processing sim data (e.g., economic modeling)
   - I/O system metadata (fast indexing, searching)
   - Feature selection/detection for “data triage” in DWA

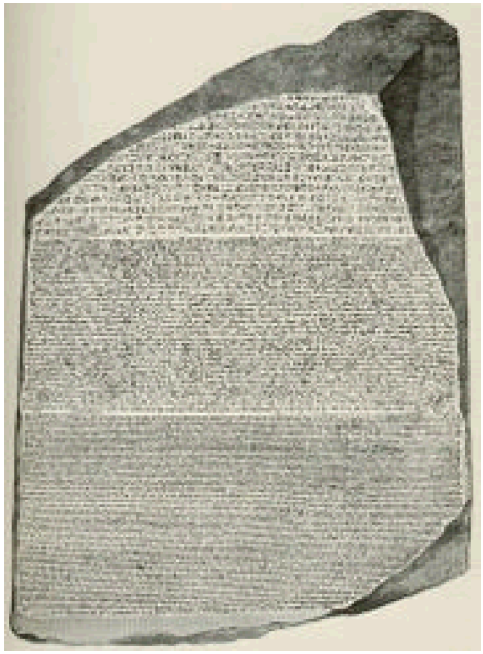
Multilingual Document Clustering

Linking threats across different languages

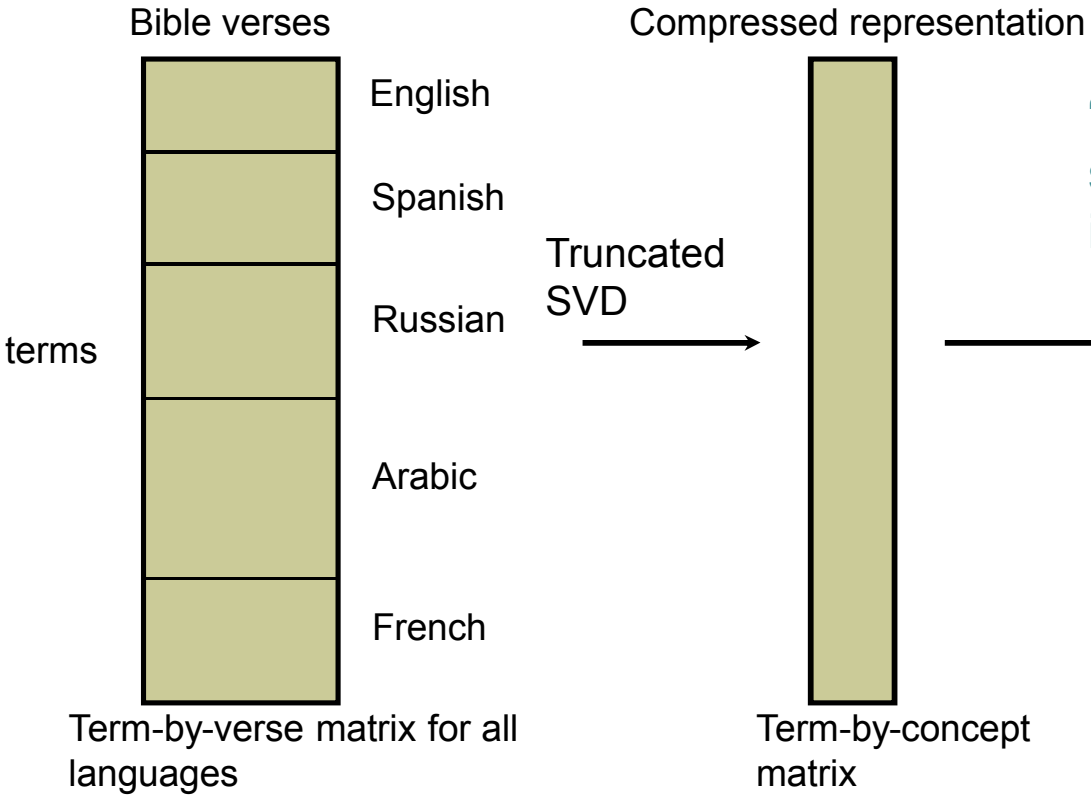
- Threats buried in many languages
  - Web documents
  - Internet traffic (email, packet payloads)
  - Journal publications
- Sandia has technology for handling massive data sets in 54 different languages

Afrikaans	Estonian	Norwegian
Albanian	Finnish	Persian (Farsi)
Amharic	French	Polish
Arabic	German	Portuguese
Aramaic	Greek (New Testament)	Romani
Armenian Eastern	Greek (Modern)	Romanian
Armenian Western	Hebrew (Old Testament)	Russian
Basque	Hebrew (Modern)	Scots Gaelic
Breton	Hungarian	Spanish
Chamorro	Indonesian	Swahili
Chinese (Simplified)	Italian	Swedish
Chinese (Traditional)	Japanese	Tagalog
Croatian	Korean	Thai
Czech	Latin	Turkish
Danish	Latvian	Ukrainian
Dutch	Lithuanian	Vietnamese
English	Manx Gaelic	Wolof
Esperanto	Maori	Xhosa

Rosetta stone

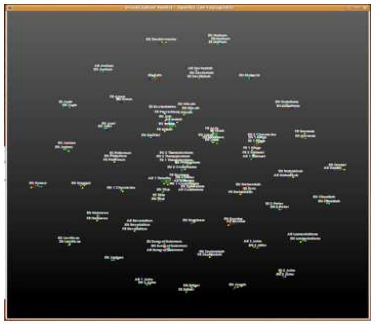


Latent Morpho-Semantic Analysis



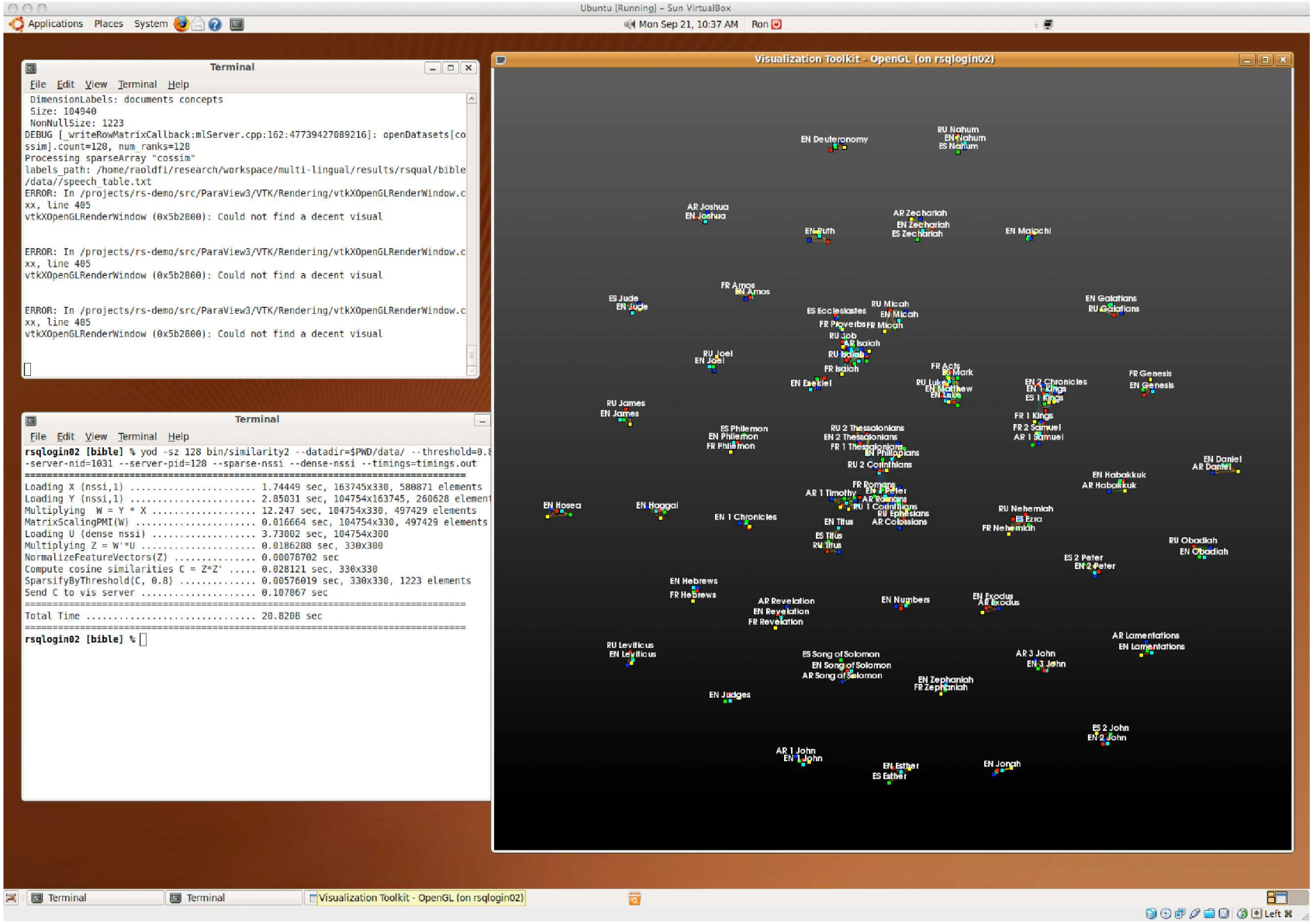
"Translate" new documents into a small number of language-independent features

Books of the Bible clustered by topic regardless of language



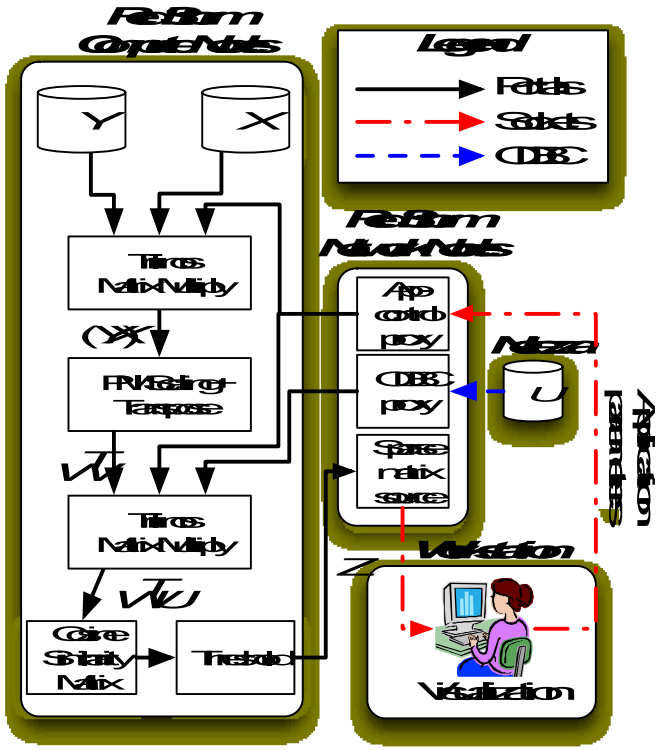
Peter A. Chew, Brett W. Bader, and Ahmed Abdelali. Latent Morpho-Semantic Analysis: Multilingual information retrieval with character n-grams and mutual information. In *Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008)*, pages 129–136, Manchester, UK, August 2008.

Latent Morpho-Semantic Analysis



Multilingual Document Clustering is a Great Candidate for HPC

- Satisfies HPC requirements
  - Scale of Data
    - Millions of elements (Bible, Quran, Wikipedia, Europarl)
  - Computationally expensive
    - Matrix multiplies for large matrices
  - Time to Solution
    - Interactive control/vis is a motivating factor
    - Focus on “strong scaling” capabilities of HPC platform

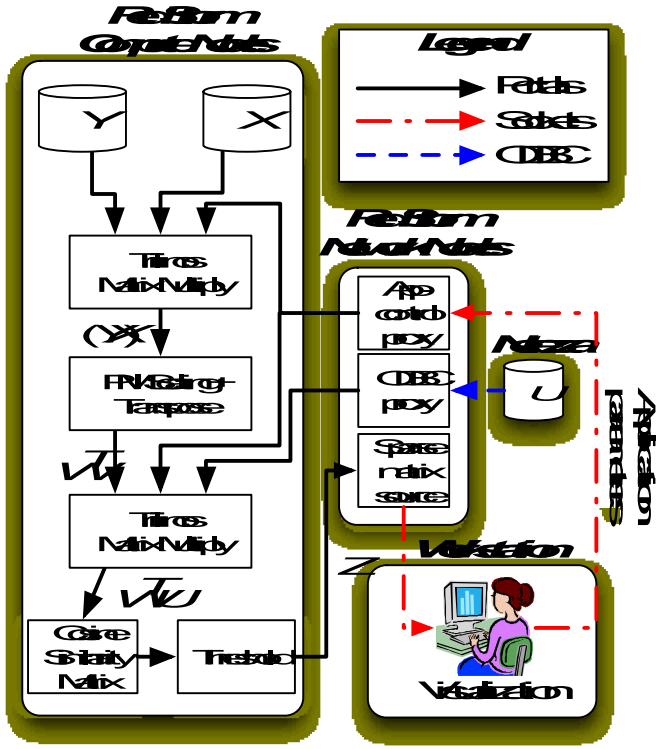


- Leveraging Existing Sandia Libraries
  - LMSA for dataset generation
  - Trilinos for computation
  - VTK/Titan for visualization
  - Nessie for data services (provides “glue” to integrate systems)

Architectural Challenges

Exploiting specialized architectures

- Cray for numerics
- Clusters/Workstations for vis and interactive control
- Data Warehouse Appliances for database functionality



Integrating these systems for interactive jobs has never been done

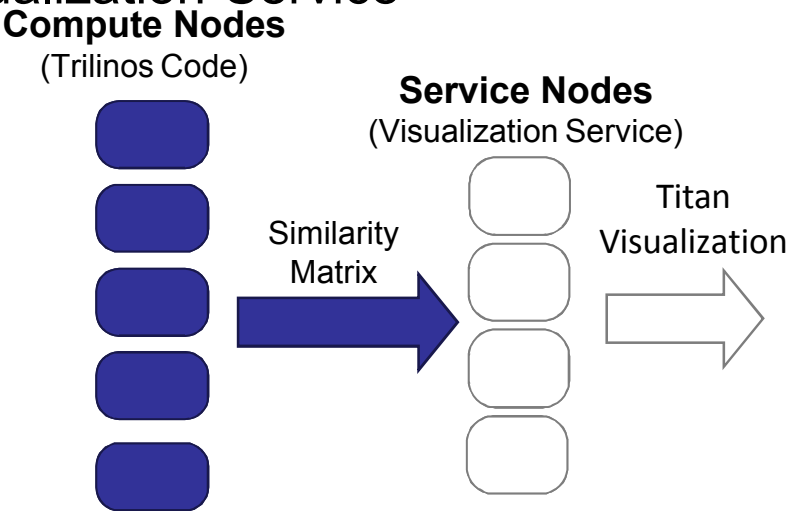
Bridging Architectures with Data Services

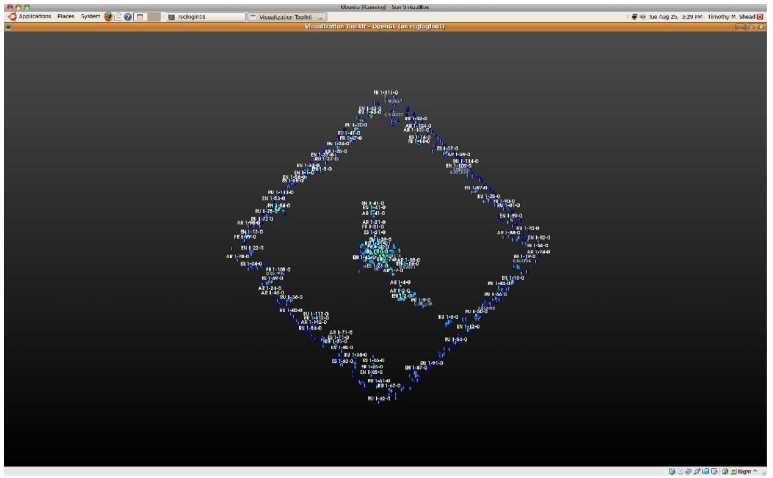
Sandia National Laboratories

Network Scalable Service Interface (Nessie)

- Developed for the Lightweight File Systems Project
- Framework for HPC client/server services
- Designed for scalability and bulk data movement
- Portals, InfiniBand, Gemini (Cray), DCMF (IBM) Implementations

Visualization Service





Scaling Challenges for Document Clustering

Sandia National Laboratories

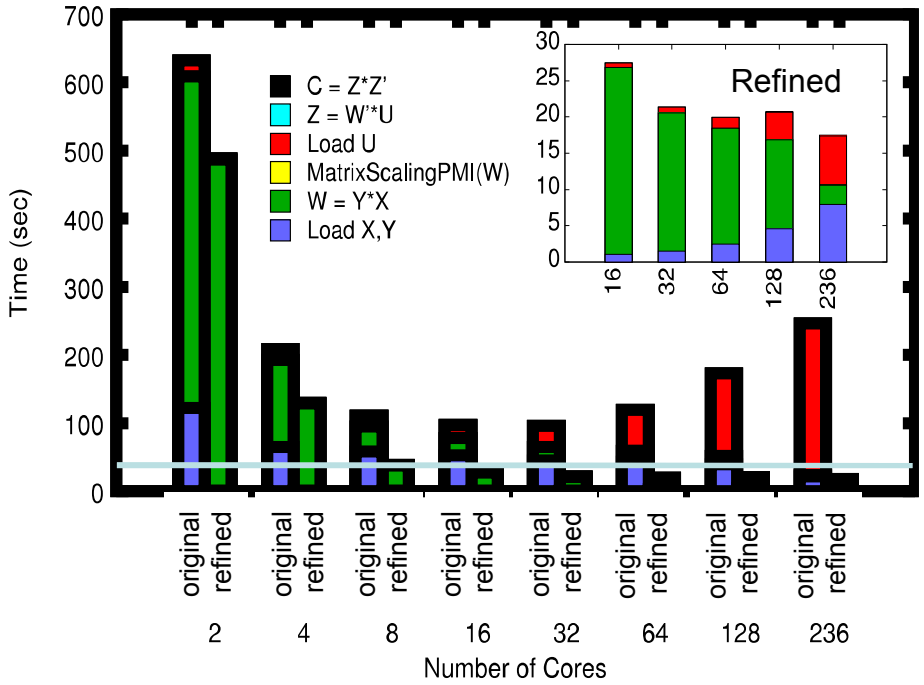
Strong scaling exposes weaknesses

- Original methods for loading were not designed for production use.

Improvements to enable large scale

- Sparse Reads
  - Keep track of mapping
  - Parallel I/O
- Dense Reads
  - Convert to binary format
  - Parallel I/O
  - Data ordering
- Memory efficient algorithms
  - Multi-pass dense-matrix multiply for cosine-similarity allows calculation of full dataset
  - Previous version could not cluster 400K docs (one matrix had to be resident in memory on each process)

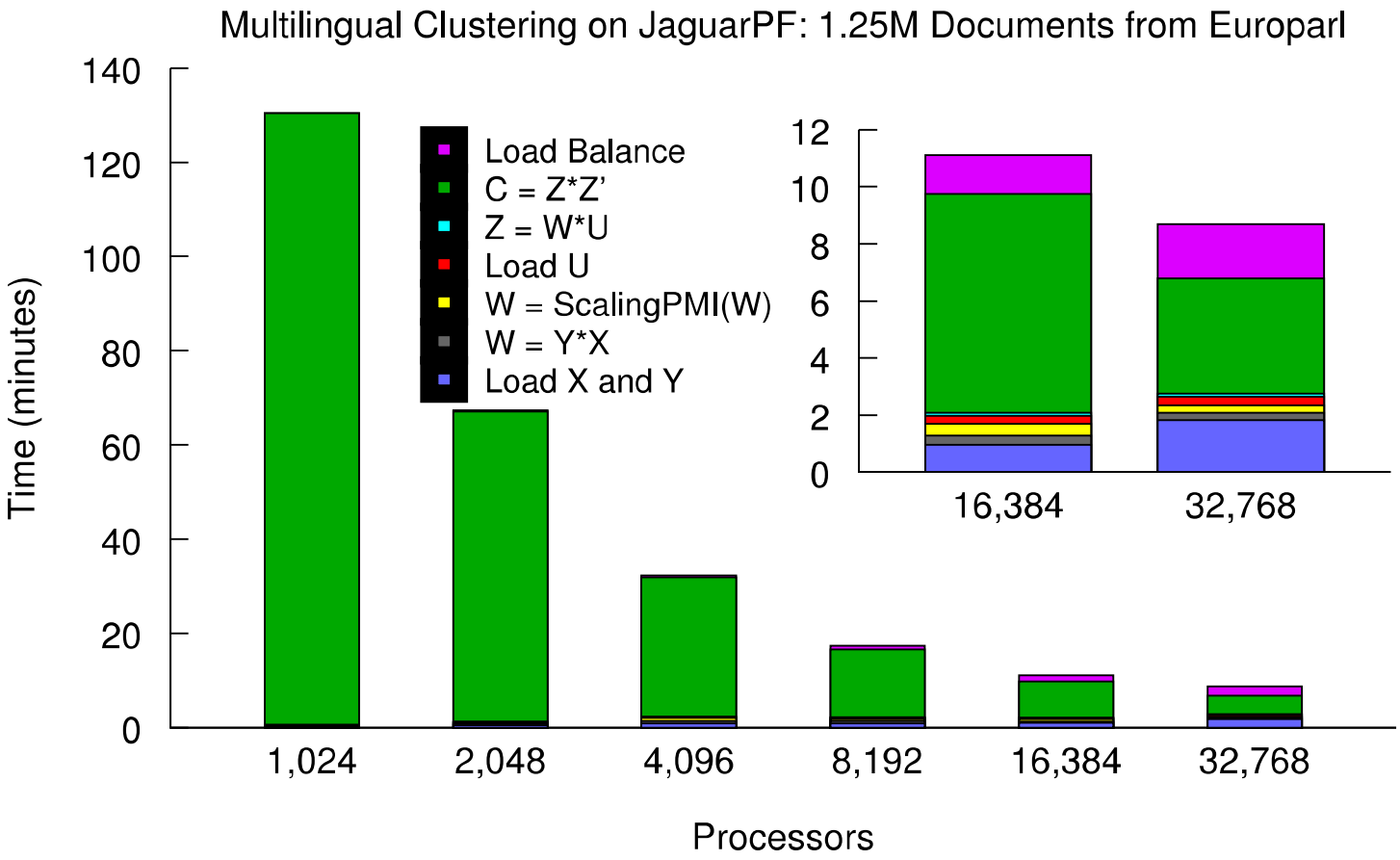
Performance Results: Bible Dataset



Ron A. Oldfield, Brett W. Bader, and Peter Chew. Supporting multilingual document clustering on the Cray XT3. In *SIAM Conference on Parallel Processing and Scientific Computing*, February 2010.

Multilingual Document Clustering

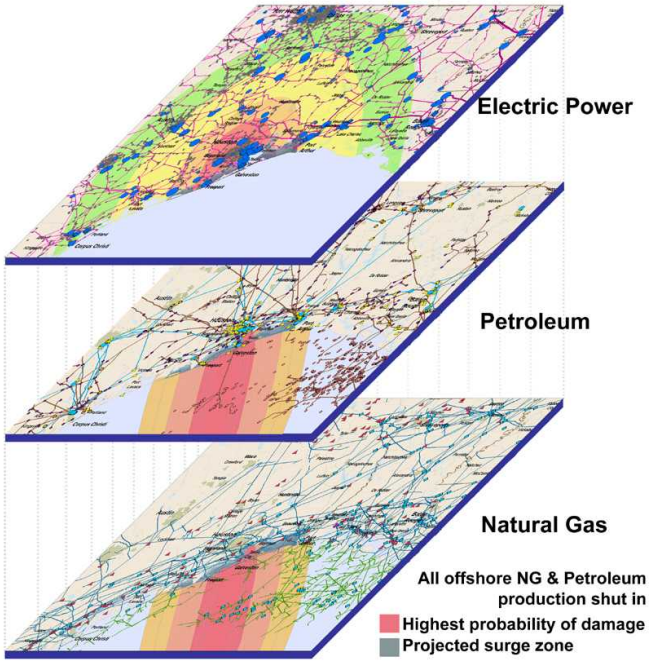
Performance on JaguarPF



NISAC/N-ABLE

Modeling Economic Security

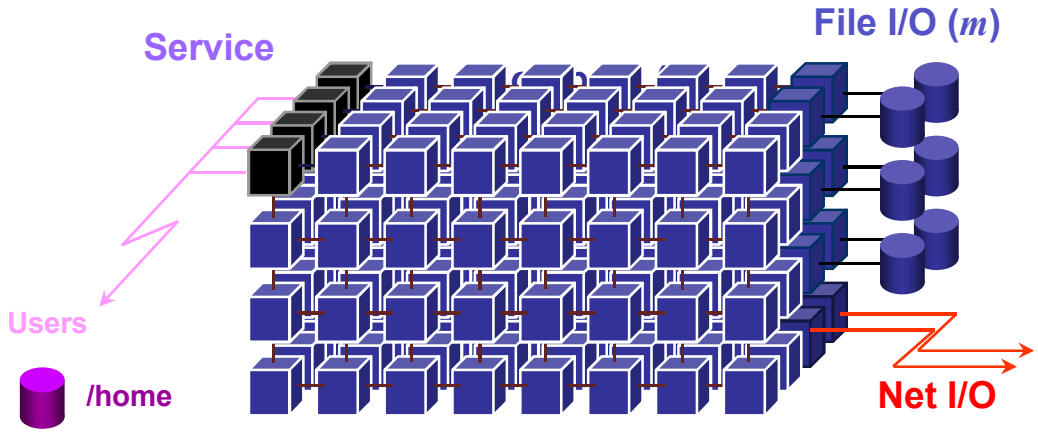
- Model economic impact of disruptions in infrastructure
  - Changes in U.S. Border Security technologies
  - Terrorist acts on commodity futures markets
  - Transportation disruptions on regional agriculture and food supply chains
  - Optimized military supply chains
  - Electric power and rail transportation disruptions on chemical supply chains
- Compute and data challenges
  - Models economy to the level of the individual firm
  - Model transactions from 10s of millions of companies
  - Simulation data ingested into DB for analysis
  - ***DB ingest is bottleneck*** (10x time to simulate data)
  - Time to solution is critical… want answers in hours



NISAC identifies potential consequences of disruptions to infrastructures and analyzes cascading impacts due to interdependencies

Integration of Cray XT and Netezza

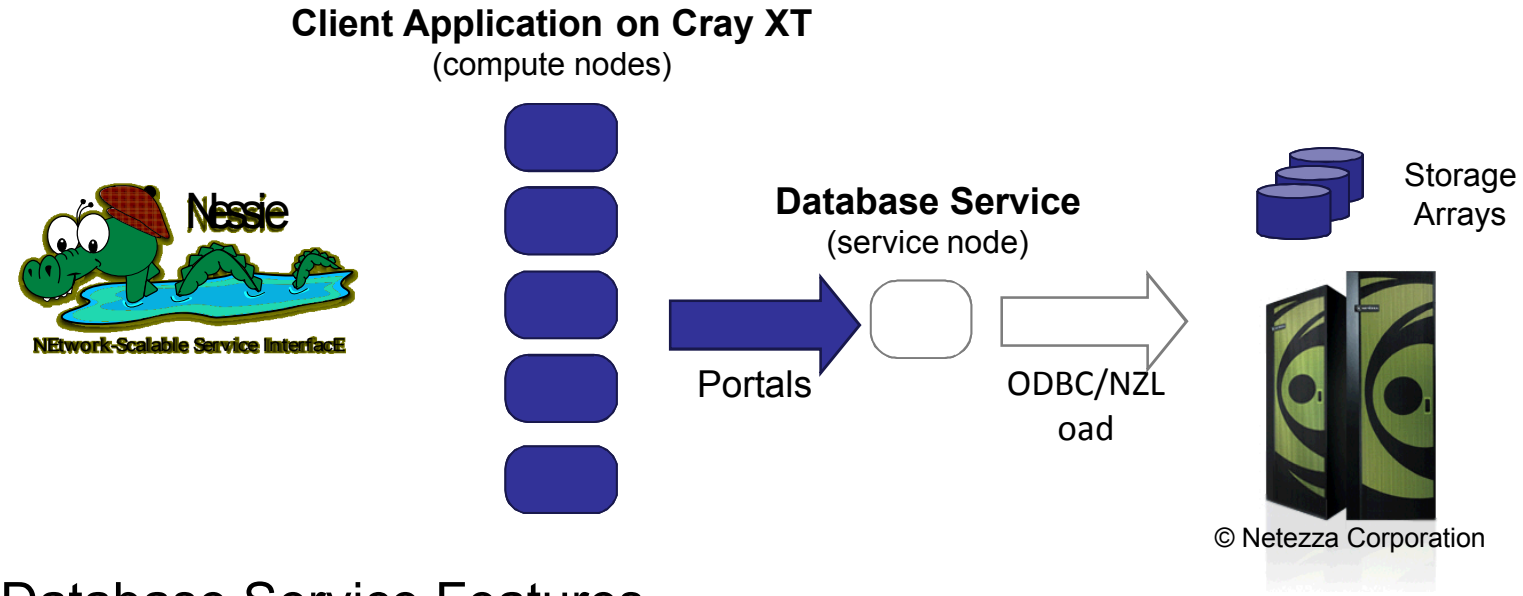
Cray XT for Simulation, Netezza for Analysis



- Potential Benefits
  - Cray XT provides memory and compute resources for large-scale simulations
  - Netezza provides fast queries for post-analysis of data
- Software and Hardware Challenges for HPC
  - Specialized internal network APIs (Portals) for Cray
    - No support for standard DB interfaces (e.g., ODBC)
    - Standard database APIs not designed for “bursty” input patterns
  - Fast network internally (2 GB/s/link), slow externally (1 Gb/s)
    - Networked integration of systems could lead to a severe I/O bottleneck

Database Service

A Network Proxy Between Cray and Netezza



Database Service Features

- Provides “bridge” between parallel apps and external DWA
- Runs on Cray XT/XE/XMT network nodes
- Applications communicate with DB service using Nessie (over Portals/uGNI)
- Service-level access to Netezza through standard interface (e.g., ODBC)

Ron A. Oldfield, Andrew Wilson, George Davidson, and Craig Ulmer. Access to external resources using service-node proxies. In *Proceedings of the Cray User Group Meeting*, Atlanta, GA, May 2009.

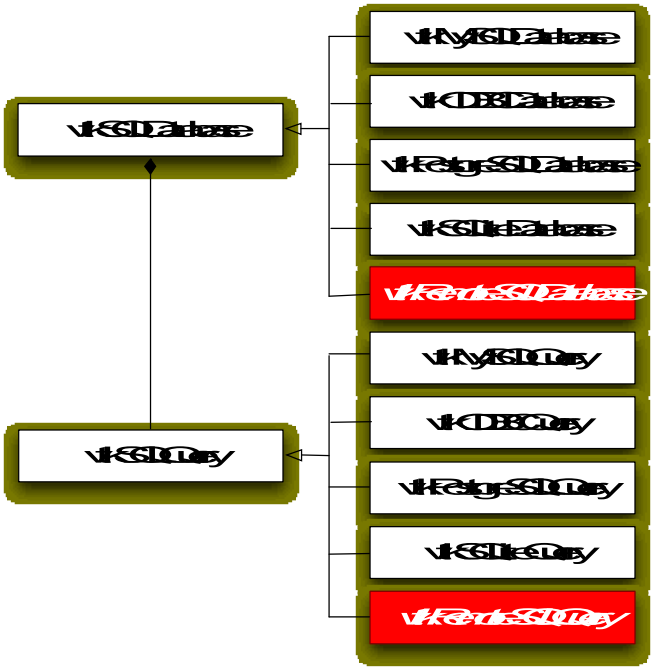
Database Service Implementation

SQL Service

- Client
  - Extensions of vtkSQL{Database,Query) classes
  - Construct direct SQL command for remote DB
  - Marshal args for remote func, send to server using NSSI (on top of Portals)
- Server
  - De-serialize request
  - Execute SQL command on behalf of application using ODBC.

Bulk Ingest Service

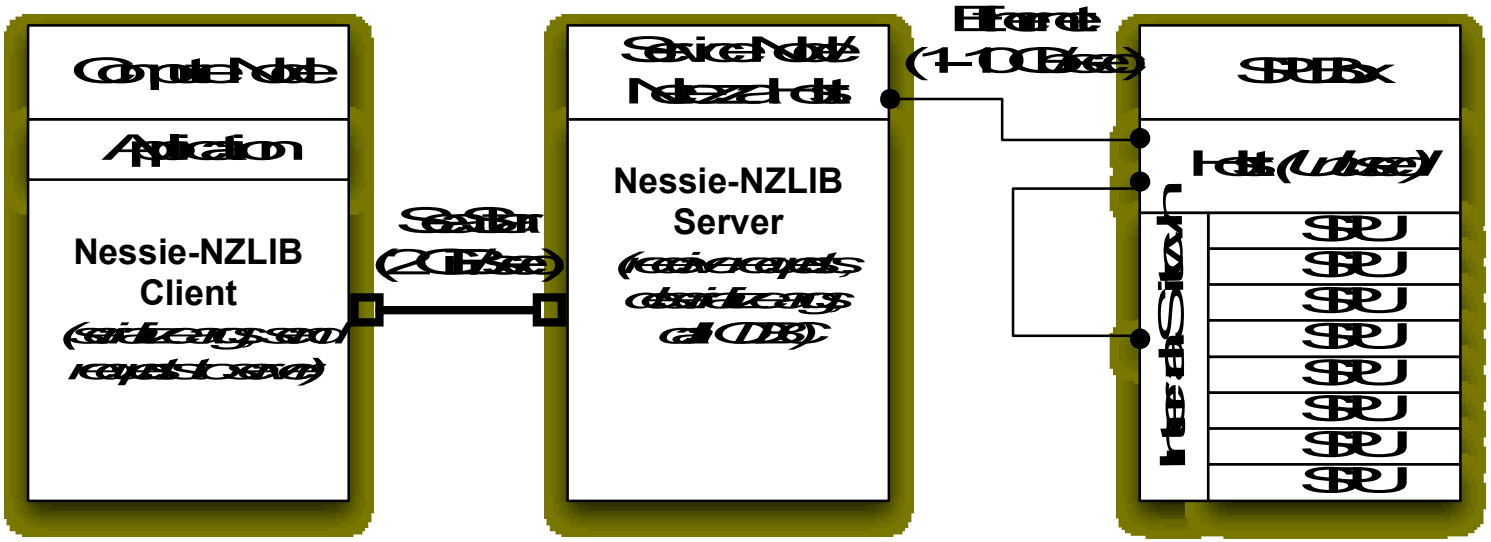
- Client
  - Simple interface to dump large tables
  - Options to transport data with request or fetch data using RDMA
- Server
  - De-serialize request, fetch data (if necessary)
  - Options for File, SQLite, ODBC, NZLoad



Tight Coupling of Cray and Netezza

A hardware solution to improving ingest rates

Original Networked Implementation

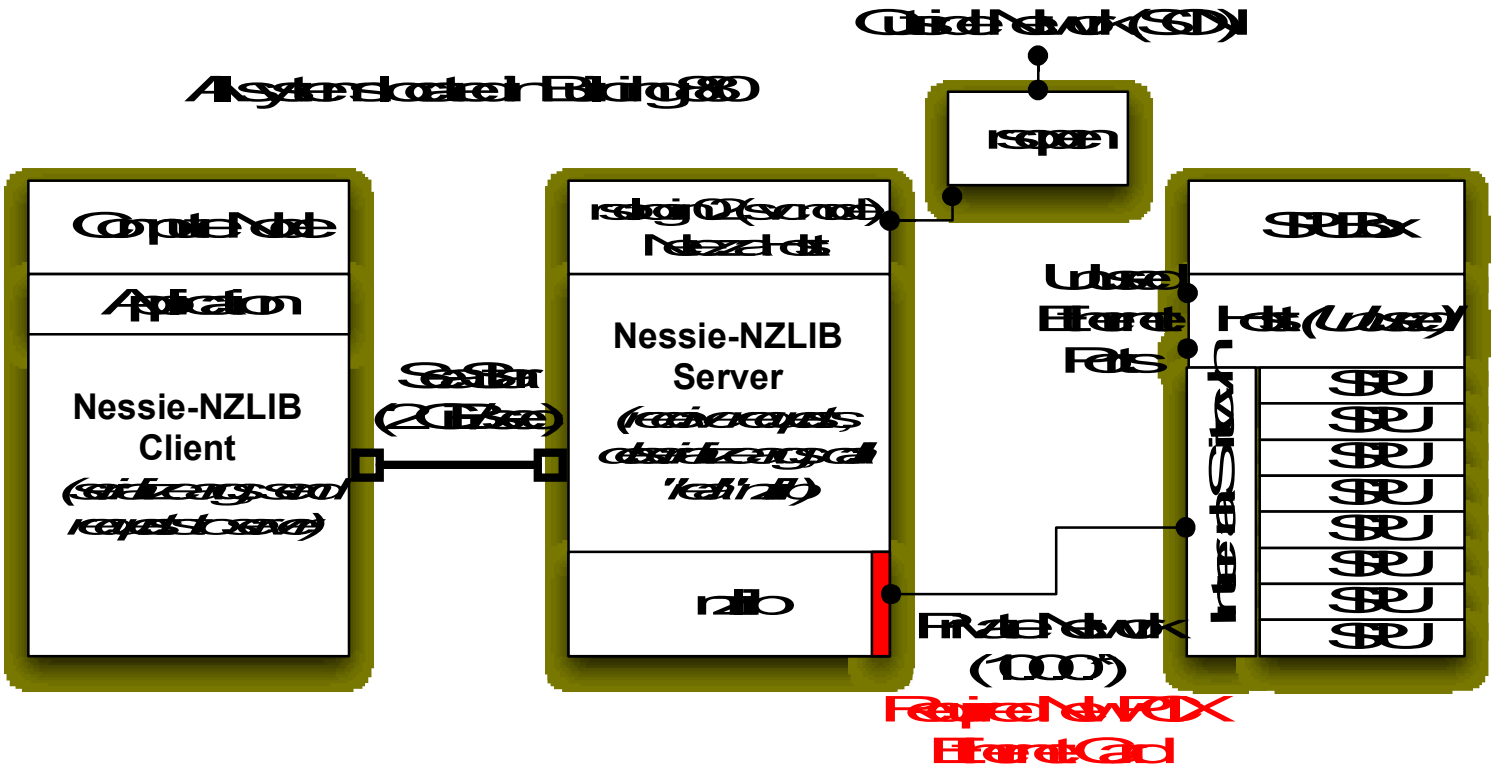


Would like to use Netezza library in the DB service instead of piping everything through the nzsql command-line tool.

Tight Coupling of Cray and Netezza

A hardware solution to improving ingest rates

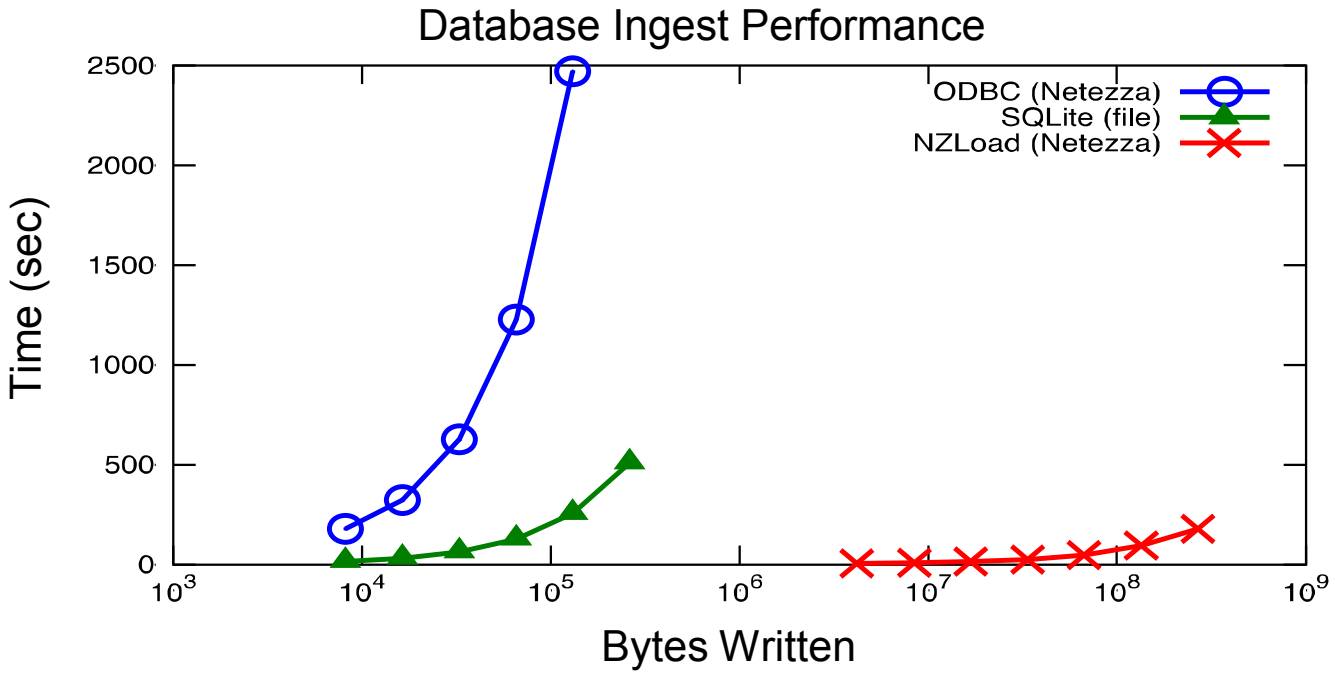
Move Host to Service Node



Use Netezza library in the DB service instead of piping everything through the nzsql command-line tool.

Results (Bulk Ingest)

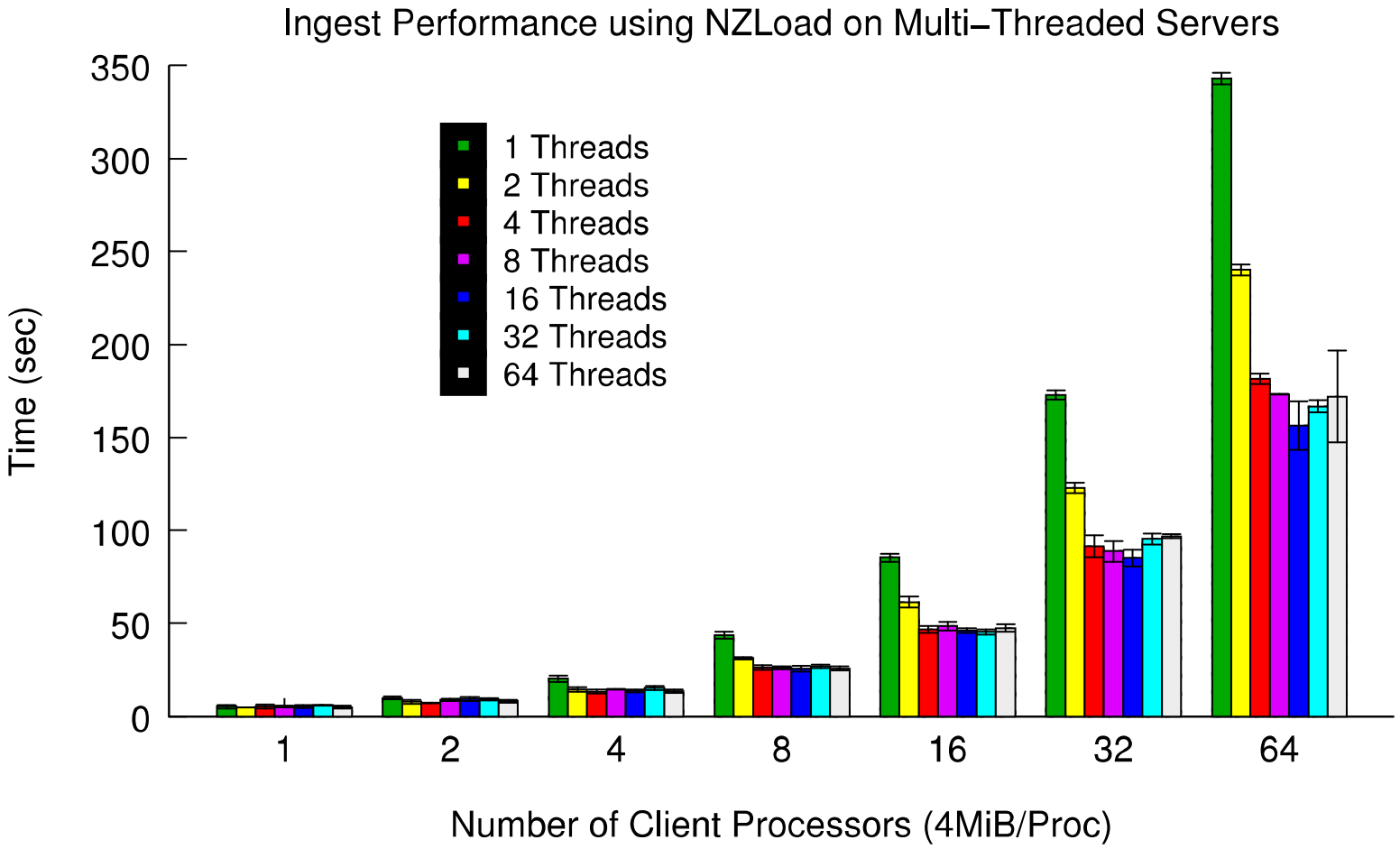
- Benchmark: Generate 4GiB of records/process, scale data by increasing number of clients.



- ODBC is the only available interface for remote access. It's the interface, not the network that's the bottleneck
- NZLoad only runs on the Netezza host, but it is many order of magnitude faster

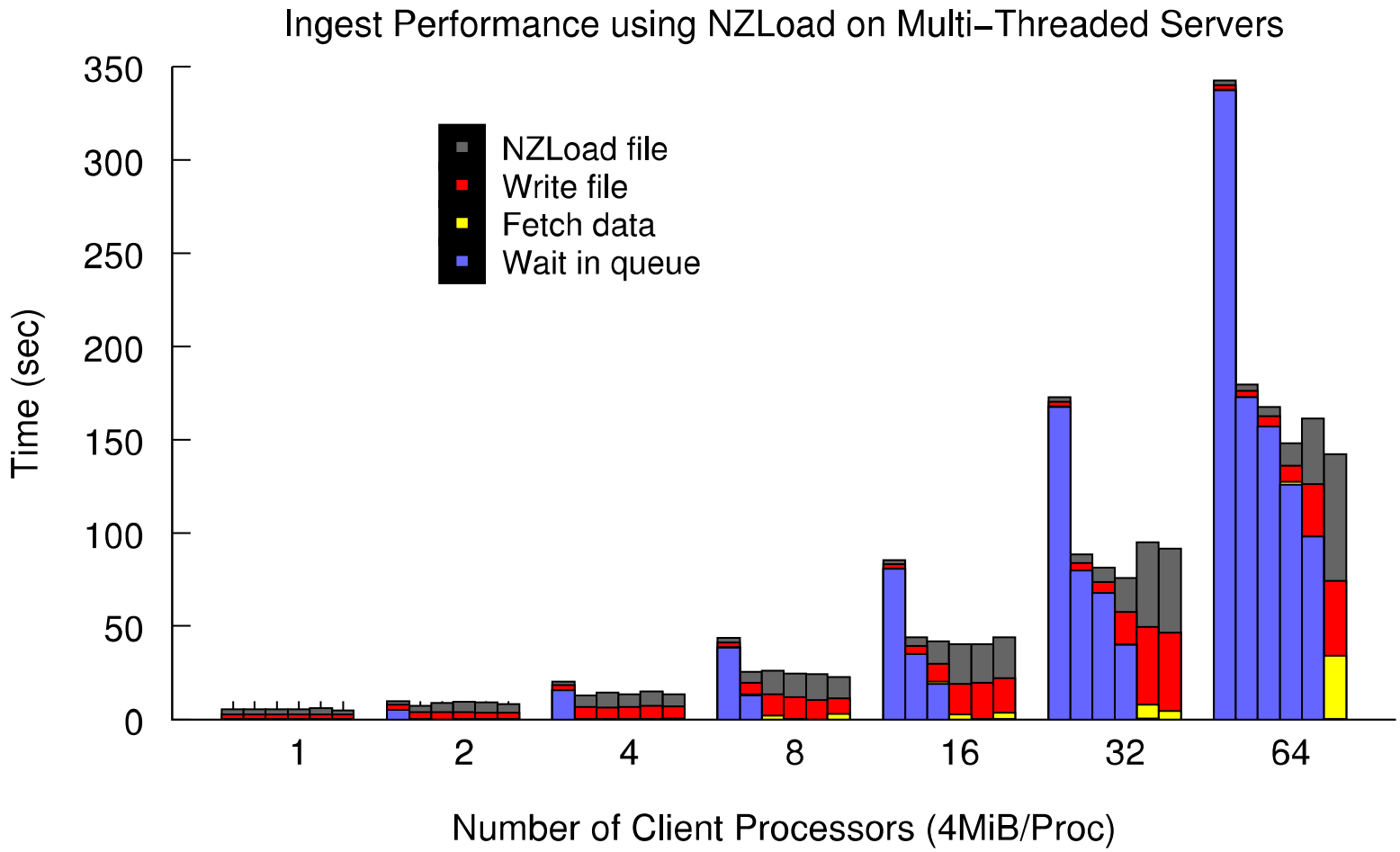
Evaluation of Hybrid System

Using NZLoad from the DB Service

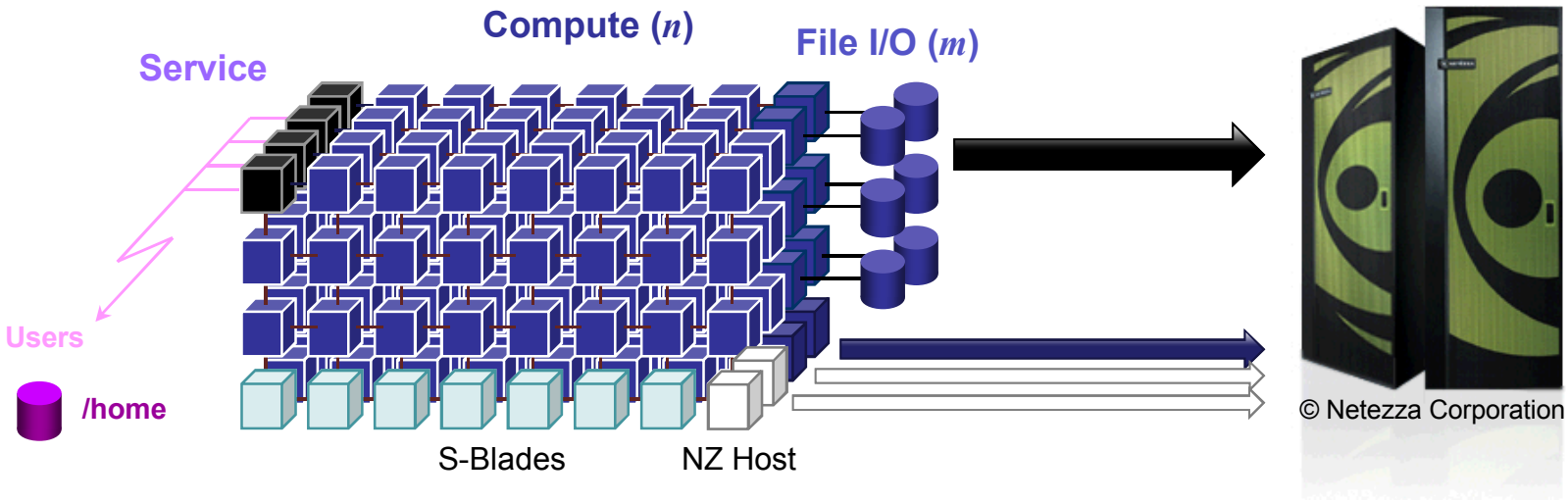


Evaluation of Hybrid System

Using NZLoad from the DB Service



Hybrid Architecture Evolution



Research/Engineering Questions

- What ingest rates will keep up with scientific workloads?
- Where are bottlenecks? Between host and S-BLADE?
- What software/networking infrastructure will resolve the bottlenecks?

An evolving architecture to support rapid ingest for HPC workloads

1) Stage data to FS during sim, ftp to host, bulk load to DB. (post-processing)
2) DB Service forwards ODBC requests to remote Netezza (slow network to host)
3) Service node becomes NZ host (fast access to host, slow to S-BLADEs)
4) Multiple service-node hosts (parallel access to back-end S-BLADEs)
5) Really wacky! Hosts and S-BLADEs on fast network (fully integrated)

Summary and Lessons

Sandia National Laboratories

- Data size, compute reqs, and time-to-solution motivate HPC for analytics
- Multilingual Document Clustering (Analytics Code on HPC)
  - Large Data Sets
  - Strong scaling for short time-to-solution
  - Requires interactive use by analyst
- Economic Modeling (Integration of HPC and Analytics Hardware)
  - Integrates HPC code with Data Warehouse Appliance
  - Time-to-solution includes data generation, DB ingest, post-analysis
  - Response time is critical to maintain relevance
- Data Services Play and Important Role
  - Provides “glue” for system integration
  - Enable data-warehouse integration
  - Enable interactive visualization
- Scaling Challenges Demand Attention to Detail
  - Moving cluster-based analytics codes to HPC system is not a trivial task
  - Strong scaling exposes weaknesses not seen in cluster

Acknowledgements

- Multilingual Document Clustering
  - Peter Chew, Brett Bader
- N-ABLE
  - Eric Eidsen, John Masciatoni
- Netezza Integration
  - George Davidson, Craig Ulmer, Andrew Wilson, Kevin Pedretti
- Funding Support
  - LDRD: Networks Grand Challenge
  - ASC/CSRF: HPC System Support for Informatics