



High-Performance Data Analytics

Recognizing the growing challenges in complex data analysis, Pacific Northwest National Laboratory (PNNL) is studying new ways to advance high-performance computing technologies that accelerate data analysis and predictive analytics to address complex challenges in energy, national security, and fundamental science.

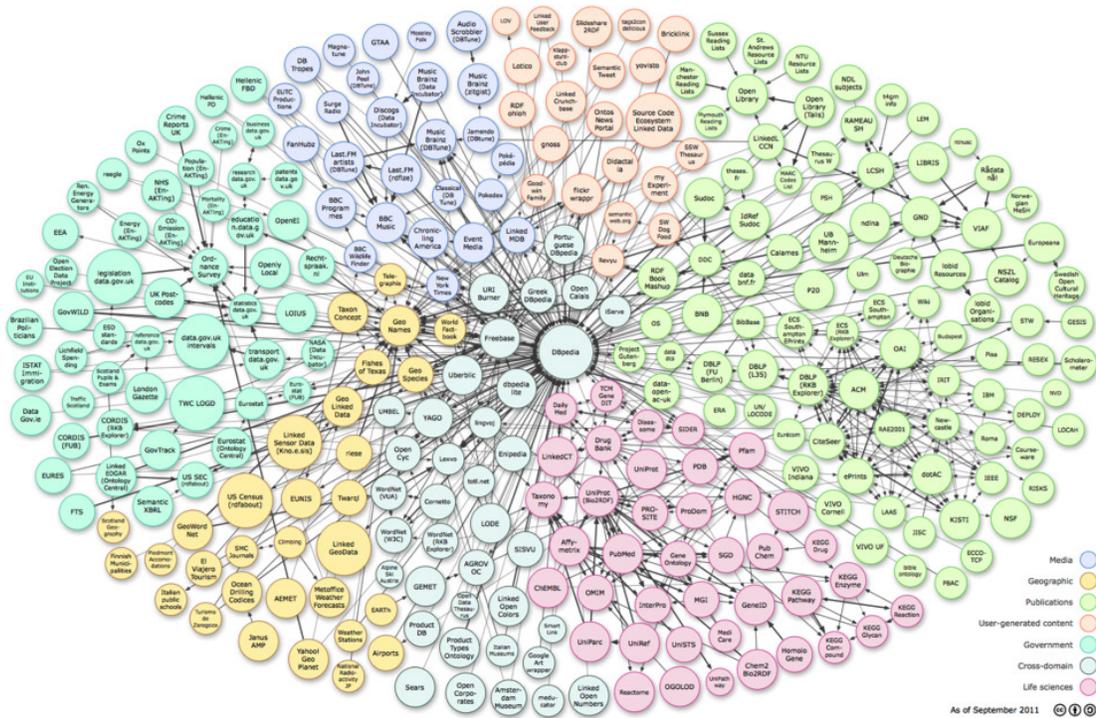
CHALLENGE

To address complex challenges of national importance, the rapid and effective analysis of large-scale data is crucial. However, the volume, variety, and unstructured nature of big data available to analysts have outpaced the capabilities of conventional analytic methods. Development of new data intensive applications is necessary. Unfortunately, these applications are memory bound on conventional commodity computing systems due to the increasing gap between memory and processor speeds. That is, the speed of the memory determines performance. Slow memory speed prevents

the processor from working while it waits for data to arrive. Consequently, new programming models, runtime systems, and even computer architectures are needed to enable the emerging class of data intensive applications.

SOLUTION

PNNL research staff are developing multithreaded graph algorithms and the programming models and runtime systems necessary to support them on conventional commodity computing systems. Their work provides a scalable, cost-effective solution for big data problems that accommodates unstructured, in-memory datasets and allows for high performance of irregular applications. In particular, PNNL has designed a database that is closely aligned with future data trends and discovers complex relationships in unstructured data that other databases fail to find. The Graph Engine Multithreaded System (GEMS) database runs on commodity platforms from desktop computers to the cloud – requiring no special system requirements.



The semantic web creates challenges to effective data analysis. PNNL is focused on the development of new methods to enable data intensive applications to accommodate such unstructured data. *Linking Open Data cloud diagram, by Richard Cyganiak and Anja Jentzsch. <http://lod-cloud.net/>*

IMPACT

PNNL has been at the vanguard of high-performance computing for over five years with investments across the entire computing stack, ranging from hardware evaluation and emulation to systems-level research, application development, and the creation of new algorithms. It has developed close ties with several leading vendors and engaged in extensive outreach to academia and other national laboratories.

More recently, a new focus has been placed on high-performance computing for the rapid and effective analysis of large-scale data analytics problems. Beginning in 2014, PNNL will be leading a new high-performance data analytics program where they will continue to invest in collaborative partnerships for a continuous impact in meeting the nation's new and evolving national security needs. The new program will focus on four technical areas: streaming, graphs, compute intensive, and exploratory data analysis. In addition, it will build out a geographically-distributed testbed for evaluating new methods for the processing, storage, and management of big data.

For more information, contact:

Rich Quadrel

Project Manager
rich.quadrel@pnnl.gov
(509) 375-5933

Dave Thurman

Manager, Analytics Programs
National Security Directorate
dave.thurman@pnnl.gov
(206) 528-3221



Proudly Operated by **Battelle** Since 1965

