XMT BOF SC09
XMT Status And Roadmap

Shoaib Mufti
Director Knowledge Management

CRAY
THE SUPERCOMPUTER COMPANY

CRAY THE SUPERCOMPUTER COMPANY

Outline

- XMT Application and Overview
- XMT Status and Roadmap
- Knowledge Management Practice

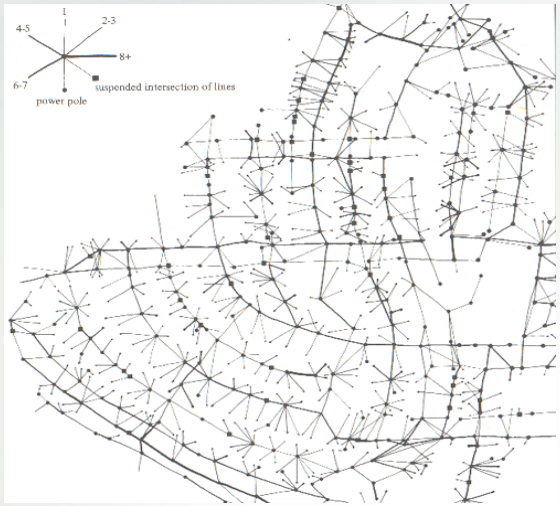
Knowledge Management and Discovery

- **Growing size of on-line content, new frontiers in science, and national security needs are creating applications that require processing of a massive amount of unstructured data**

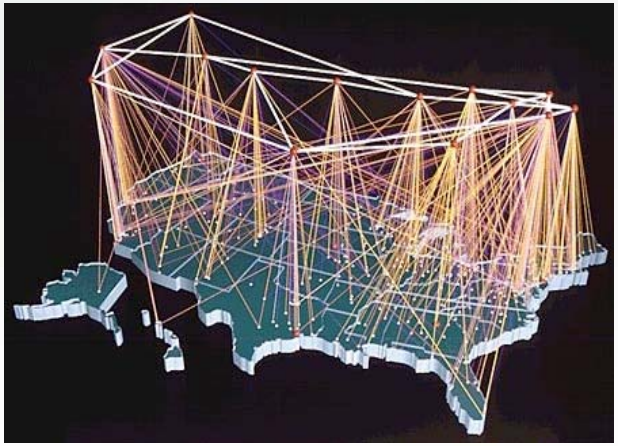
- **These problems require finding useful information and gathering knowledge from massive amount of seemingly unrelated data**
  - **"Finding needle in a haystack" problems**
  - **"Connecting the dots" problems**

- ***Examples:***
  - **Intelligence** → knowledge from massive homeland security data, cyber security by real time intrusion detection, tracking suspicious activities in billions of financial transactions
  - **National Defense** → battle field analysis
  - **Energy** → electric power grid failure analysis, energy conservation by rerouting electric power in an electric grid
  - **Health care** → disease spread, detection and prevention of epidemics/pandemics (e.g. Avian flu) by doing social networking analysis
  - **Systems Biology** → understanding complex life systems, drug design, microbial research, unravel the mysteries of the HIV virus

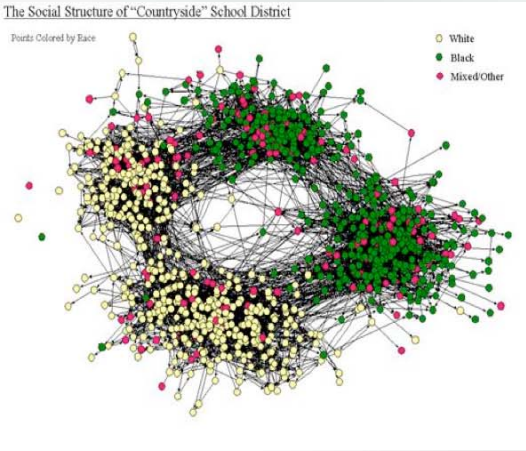
Power Distribution Networks



Internet backbone

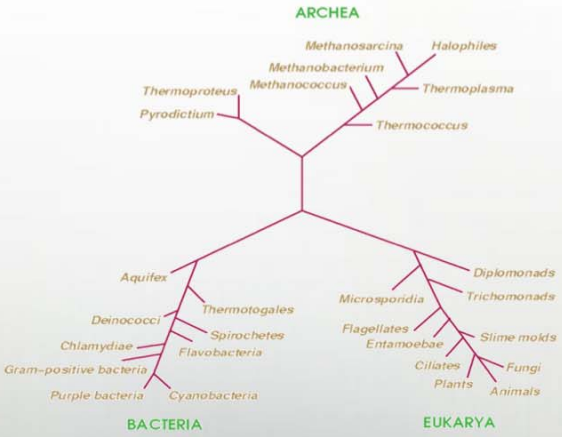


Social Networks

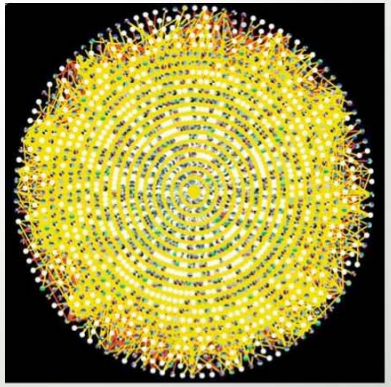


Graphs are everywhere!

Ground Transportation



Tree of Life



Protein-interaction networks

How do we process these Graphs?

- Challenges:
  - Runtime is dominated by latency
    - Random accesses to global address space
    - Perhaps many at once
  - Access pattern is data dependent
    - Prefetching unlikely to help
    - Usually only want small part of cache line
  - Potentially abysmal locality at all levels of memory hierarchy

- Desired Features:
  - Low latency / high bandwidth
    - For small messages!
  - Light-weight synchronization mechanism
  - No dependence on cache for performance
  - Global address space
    - No graph partitioning required
    - No local/global numbering conversions

One machine with these properties is the Cray XMT

Cray XMT Overview

Cray XMT

❖ Background

- With government support, Cray developed the eXtreme MultiThreading (XMT) system and technology to solve unstructured data analysis problems

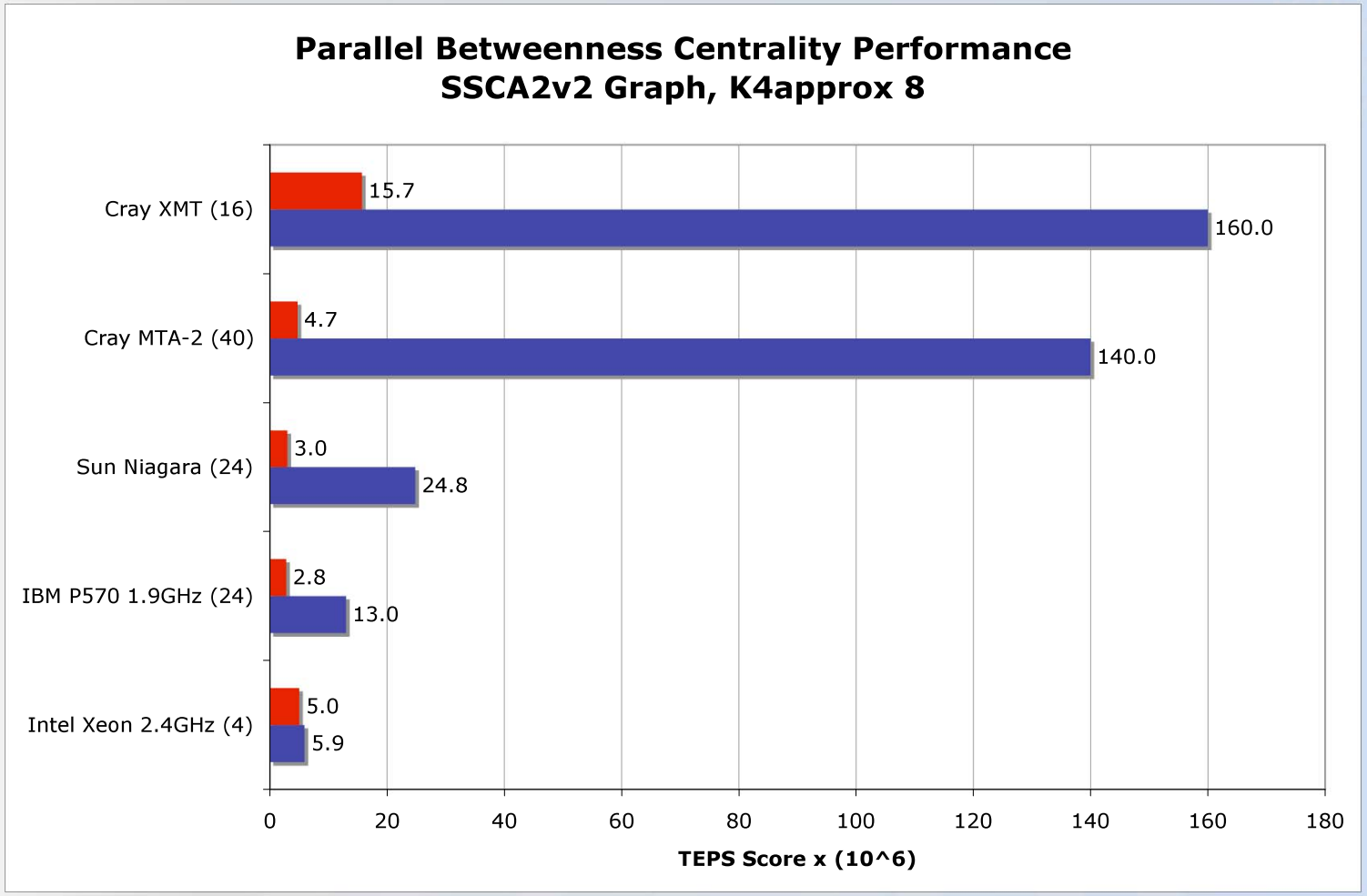
❖ Characteristics

- Extreme multithreading
  - Architecture supports 8000 processors
  - 128 hardware threads per processor
  - Practically unlimited virtual threads
- Very large shared memory
  - Architecture supports 128TB of memory
- Very low power
  - Less than 30 watt processors
- Ease of use
  - Compiler and Runtime makes parallel programming easy
- Superior price/performance for Data Intensive Computing
  - E.g. Graph Analytics, "Connecting the *Dots*"

Cray XMT Characteristics

- **Remote memory requests do not stall processor**
  - Each processor has hardware support for 128 streams of unlimited threads
  - No cache or local memory
  - Context switch on every clock cycle
  - Multiple outstanding loads
  - Other streams work while your request gets fulfilled
- **Light-weight, word-level synchronization**
  - Minimizes access conflicts
- **Hashed Global Shared Memory**
  - Minimizes hotspots
- **Hardware bit manipulation functions**
  - Bit matrix multiply
  - Shift left/right

SSCA2 TEPS Performance Comparison

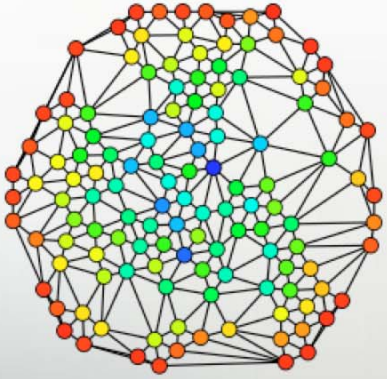


Single Processor **All Processors**

courtesy of David Bader, GA Tech

Betweenness Centrality

- **Application Significance:**
  - ✓ Betweenness is a centrality measure of a vertex within a graph. Vertices that occur on many shortest paths between other vertices have higher betweenness than those that do not.
- **XMT vs. Opteron Cluster:**
  - ✓ 64 processor Cray XMT vs. 64 processor Opteron Cluster
- **XMT performed 350 times better than an Opteron Cluster**



Cray XMT Status and Roadmap

- **First 16P XMT1 was shipped in 2Q 2008**
  - Available today with 512P and 4TB memory!
- **Multiple XMT1 Systems at Customer Sites**
  - Five customer sites with eight installations
- **XMT beginner and advanced training courses available**
- **SC09 Activities**
  - XMT BOF, DEMO PNNL Booth, Talk LexisNexis Booth (Sandia, Lexis, Cray)
- **Next generation XMT development underway**
  - Increased Memory Capacity by more than four times
  - Improved Reliability, Availability, Serviceability (RAS)
  - Reduced Footprint per TB memory – Power and Space
  - Improved Price/Performance
  - System and User Software Improvements

Community Momentum

Pacific Northwest National Laboratory (CASS-MT)

- Center for Adaptive Supercomputing Software – Multithreading Architectures
  http://cass-mt.pnl.gov/default.aspx
- Research Areas
  - Algorithms, System Software and Applications Kernels
    - Social Network Analysis
    - Statistical Textual Document Analysis
    - Dynamic Network Analysis
    - Sparse Graph Network-of-Network Algorithms
    - Contingency Analysis
  - Applications/Solutions
    - E.g. Electric Power Grid

Sandia National Laboratory

- Research Areas
  - Algorithms and Applications Kernels
  - Applications/Solutions
    - E.g. Informatics

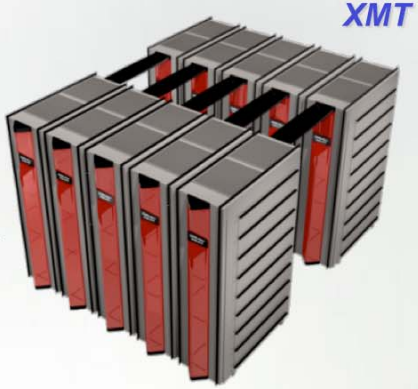
National Science Foundation

- Univ. of Notre Dame, Univ. of Delaware, UC Santa Barbara, CalTech, Georgia Tech, UC Berkeley, SNL

Knowledge Management Practice

- **Cray created the Knowledge Management Practice as a part of Cray Custom Engineering initiative in 2009 to build solutions to meet the growing demand of large scale data analysis and mining**
- **Builds an ecosystem around the Cray XMT and other products**
- **Builds business for Cray's informatics technology**
- **Go beyond offering "just hardware"**
  - Develop applications and solutions with partnerships
  - Expertise, Training, Consulting, Application development
- **Leverages Cray's vast experience**
  - Supercomputing
  - Custom engineering

KM Solution Architecture (Example)

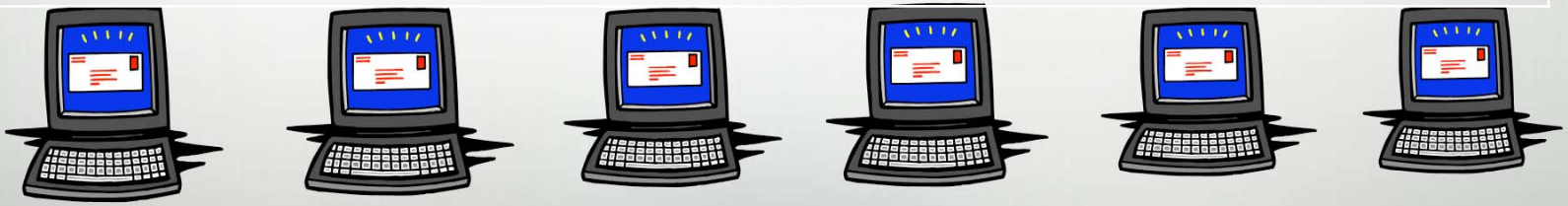


Graphical Queries

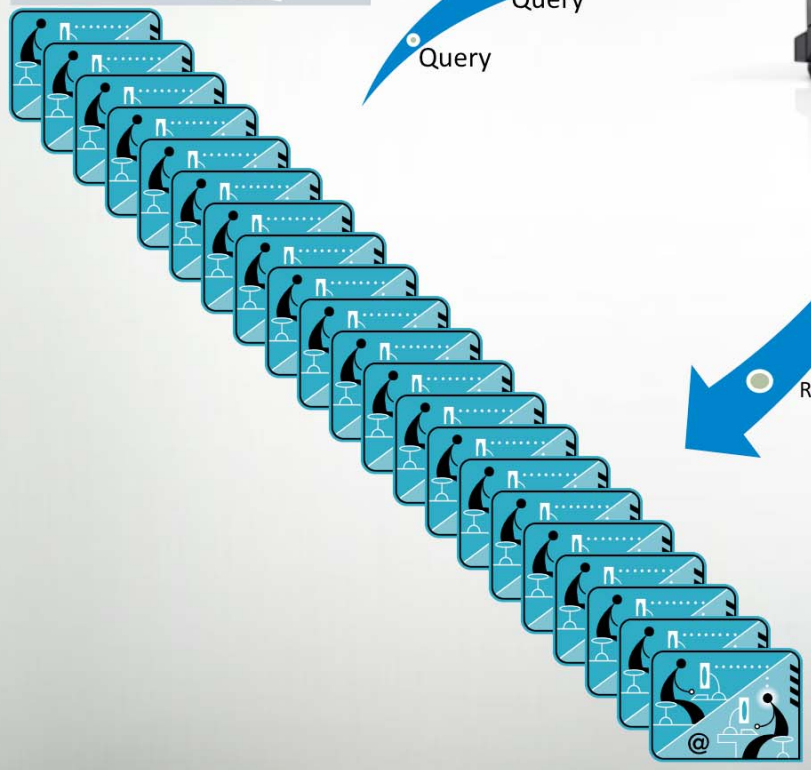
Graph Resides in XMT Memory

Analysts' Queries

Transaction Processing System



Interactive Analytics



Query

Query

Query

Response

Response

Response

Semantic Data

World Wide Web

Acknowledgements

- David Bader -- Georgia Tech
- Kamesh Madduri -- Lawrence Berkeley National Laboratory
- John Feo, Daniel Chavarria – PNNL
- Jon Berry, Bruce Hendrickson – Sandia National Labs


CRAY
THE SUPERCOMPUTER COMPANY