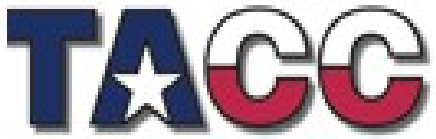
Sun Grid Engine at TACC

- **Roland Dittel**
  - Software Engineer
- Sun Microsystems GmbH

About TACC



- Research center at the University of Texas at Austin
- Provide advanced computing resources and services
- Deploys several HPC systems from different vendors
  - SUN, DELL, IBM
- Partner of NSF's TeraGrid

Copyright Sun Microsystems 2008

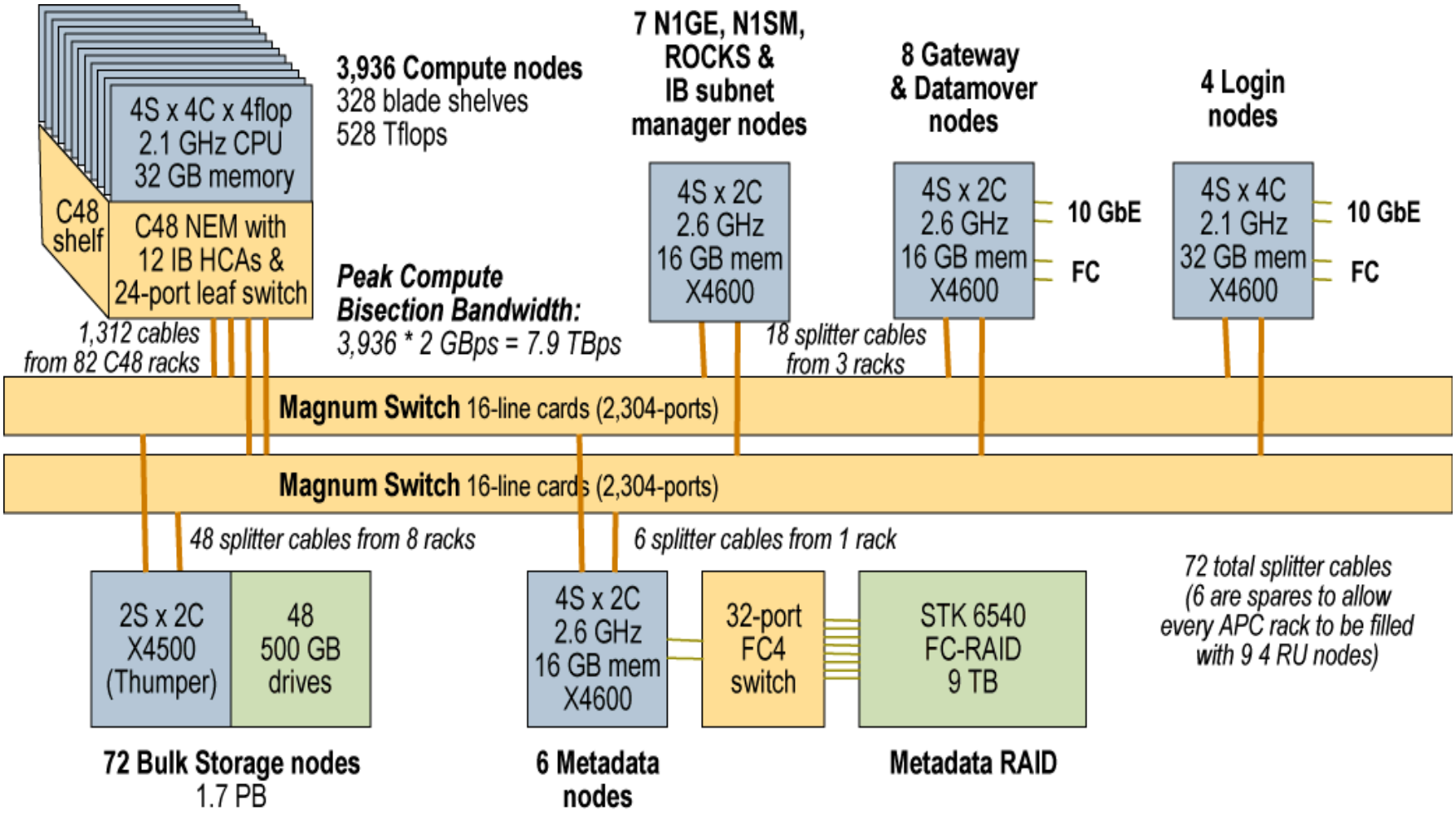
TACC's Ranger Cluster

- One of the largest HPC systems for open science research
- First of the new NSF Track2 HPC acquisitions
- 504 TFlops Peak Performance
- In production since February 4, 2008

uses Sun Grid Engine

Copyright Sun Microsystems 2008

HPC Architecture



Copyright Sun Microsystems 2008

Ranger Software Stack

- Cent OS 4.4
- Linux Kernel 2.6.9
- Lustre 1.6.4
- Mvapich2
- Rocks
- Globus
- SGE 6.1AR_snapshot3

Copyright Sun Microsystems 2008

SGE 6.1 AR_snapshot3

- Codebase somewhere between 6.1 and 6.2
  - 6.1 including Advance Reservation
  - Additional bug fixes
  - Additional scalability enhancements
    - Reduced load reports
    - Single-stage Parallel Scheduling Algorithm
    - Communication Layer enhancements
    - General qmaster 'diet'
    - ...

Copyright Sun Microsystems 2008

Sun Grid Engine Setup

- Classic spooling on Lustre
  - OK because only few spooling operations (just massive parallel jobs)
  - shadowd fail-over
- 6 Queues with different restrictions
  - max runtime, max. procs
- Project based functional Ticket Policy
- SSH for tight integrated PE tasks

Copyright Sun Microsystems 2008

Advance Reservation

- Hard requirement to use SGE
- Reason for own code branch
- Use cases
  - System Reservations
    - Reserve for maintenance jobs
  - User Reservations
    - Predictable access to exclusive resources

Copyright Sun Microsystems 2008

Execution Daemon CPU usage

- Complaint: “we are loosing 5-8% of one core to sge_execd”
- Analysis: time is spend in Portable Data Collector (PDC)
  - Collects job usage
  - PDC executed every second
- Introduced new tuning parameter PDC_INTERVAL
  - Set to load report interval
  - Usage now <1%

Copyright Sun Microsystems 2008

Qmaster CPU usage

- Complaint: “qmaster becomes unresponsive as more nodes are connected”
- Analysis: qmaster is busy with handling load reports
  - Load report interval was 15s, resulting in ~300 LR/s
  - Short interval necessary to react on file system outages
  - Reduced load report interval was broken in this build
- Set load report interval to 3 minutes
  - On file system outage host put himself into a outage hostgroup
  - Outage hosts are limited to zero slots by a ResourceQuota

Copyright Sun Microsystems 2008

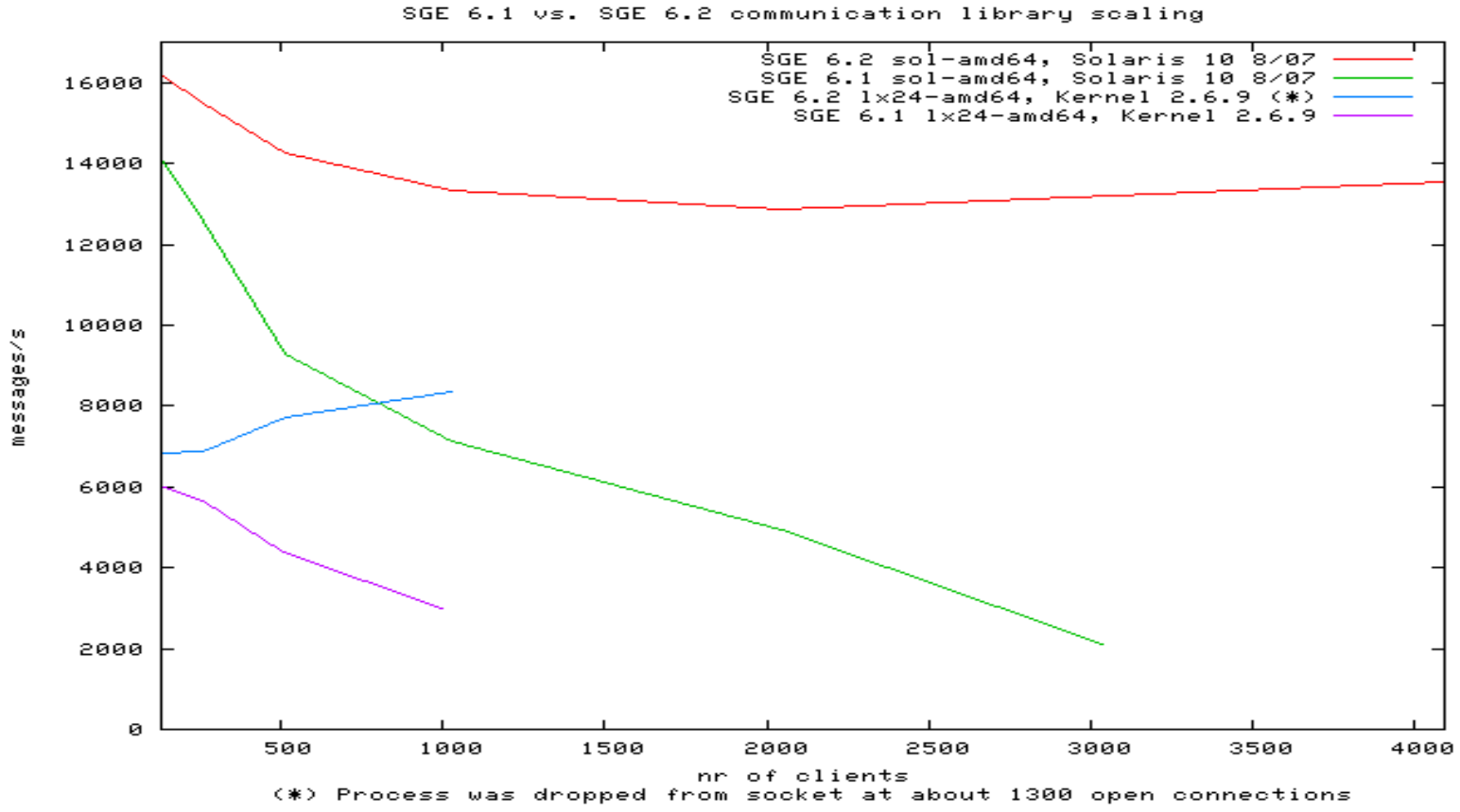
Massive parallel jobs (1)

- Complaint: “qmaster becomes unresponsive when massive parallel job gets started”
- Analysis: qmaster sends too many data to every execd
  - Every slave node needs to know about the MPI task
  - But every slave node got whole job queue instance list
    - Data growth non linear
- Fix: send only required data to every execd
  - Growth is now linear

Copyright Sun Microsystems 2008

Massive parallel jobs (2)

- Added communication layer improvements



Copyright Sun Microsystems 2008

Qmaster startup

- Complaint: “qmaster startup took ~40 s after crash”
  - Crash because of file system outage
- Analysis: startup time grows with amount of jobs and amount of queue instances
- Removed n*m relation and replaced linear by hashed searches
  - Scenario: 3000 nodes, 5 queues, 1000 running jobs, 60 pending jobs
    - Old: 72s
    - New: 6s

Copyright Sun Microsystems 2008

Ongoing

- Further reduction of CPU/memory usage
- Further work on massive MPI jobs (>32k tasks)
  - New Interactive Job Support
- Implement desired features
  - e.g Parallel job task limitation
- React on new occuring issues
- Move TACC to 6.2 code base

Copyright Sun Microsystems 2008

Sun Grid Engine at TACC

- **Roland Dittel**
  - Software Engineer
- Sun Microsystems GmbH