IBM Blue Gene

System Management Update

Pat McCarthy

System Management Update Agenda

- **Current System Management**
  - MMCS console
  - Web Interface
  - Hardware monitor
- **New Systems Management features (Release 2 and Release3)**
  - Blue Gene Navigator
  - New Performance Monitor
  - Coreprocessor debug tool
  - CSM support for BG/L
  - Mixed memory size partitions
  - New toolchain

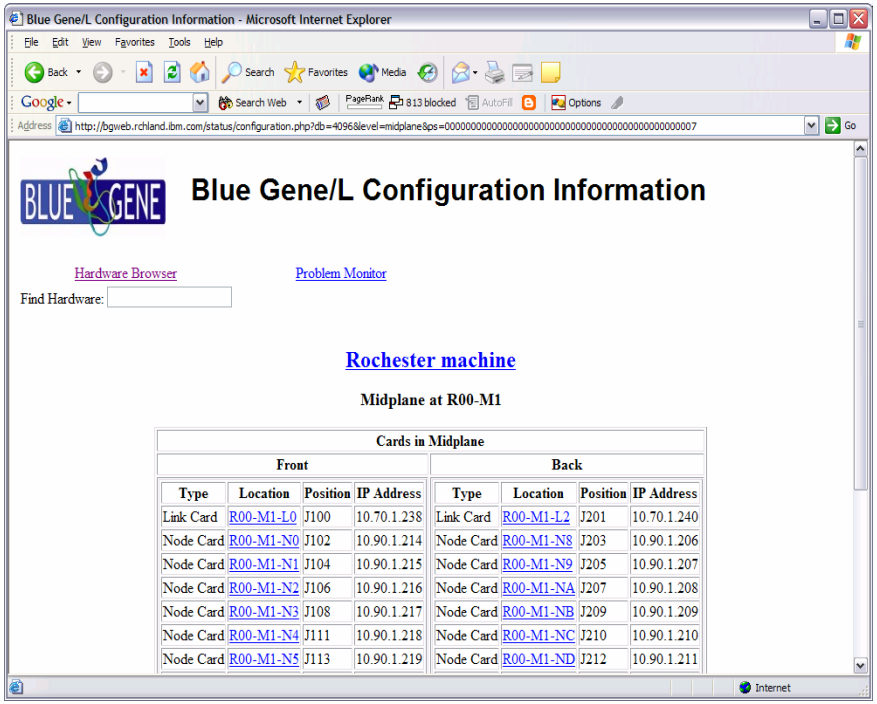
Current System Management

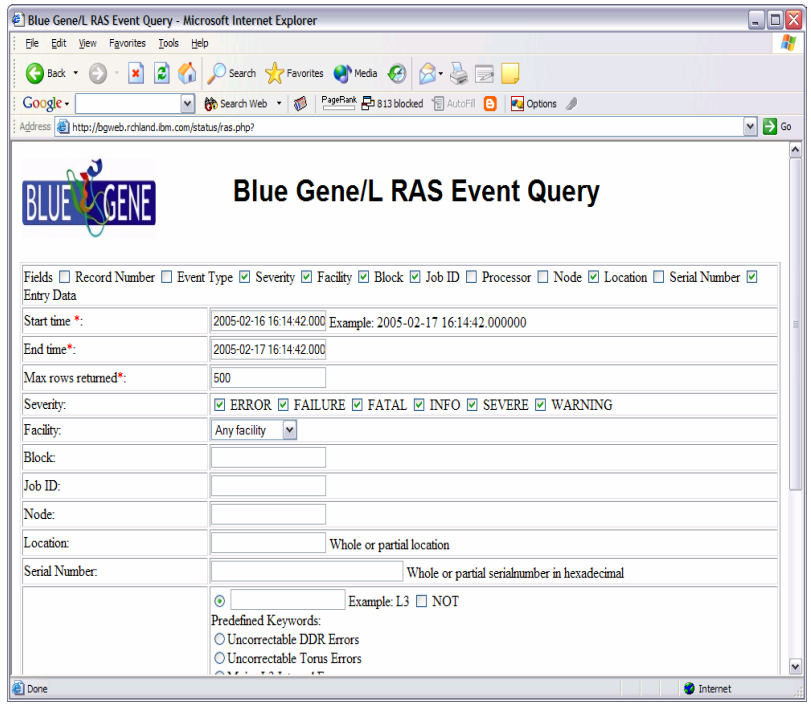
- **Main user interfaces for system management**
  - MMCS console
  - Web Interface
  - Hardware monitor
- **The DB2 database is the repository for both configuration and operational data used for system management**
  - RAS database
  - Configuration database
  - Operational database
  - Environmental database
- **Detailed presentation by Mark Megerian at February 2005 System Software workshop**

Web Interface to DB

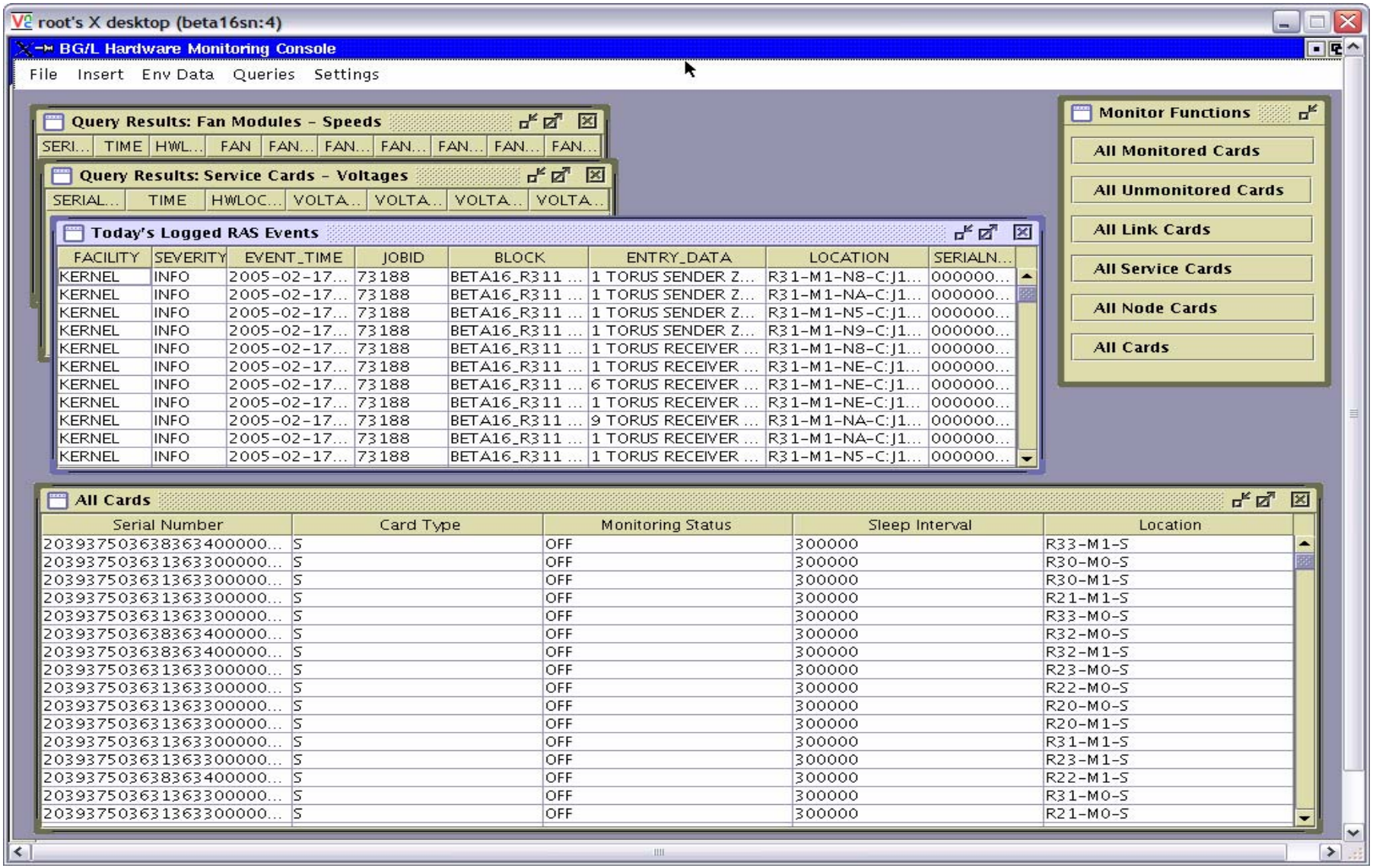
- **A front-end that runs via browser to view DB2 data.**
- **Supports the viewing of RAS data, configuration data, diagnostics data, and operational data.**
- **Can be used to see how the hardware fits together**
- **Can be used to find trouble areas, hardware anomalies**
- **Eliminates the need to have SQL expertise to view basic BlueGene information from the database.**

Web Page Screen Shots





Hardware Monitor Screen Shot



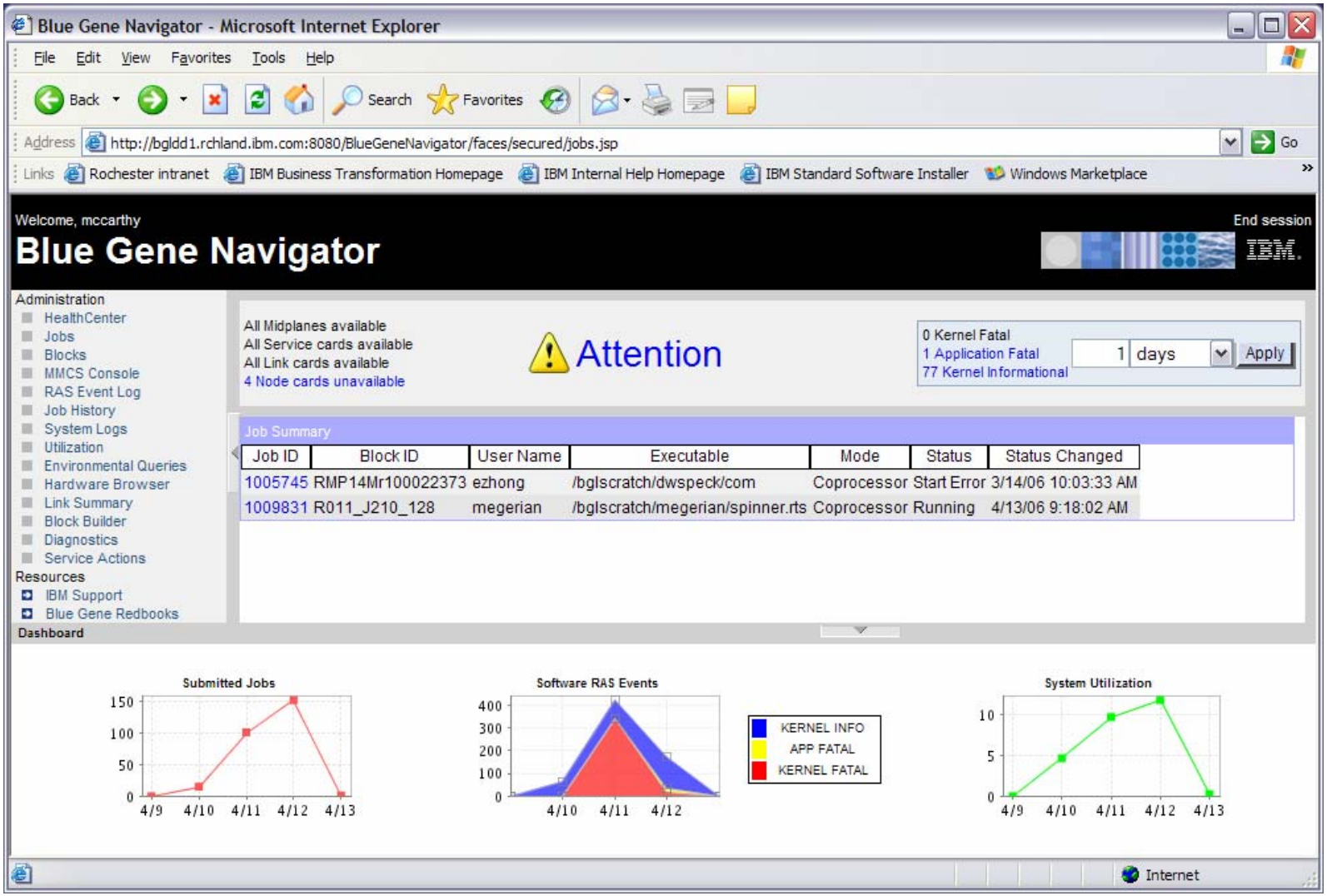
New Systems Management features (Release 2 and Release 3)

- Blue Gene Navigator
- Improved External Performance Monitor
- Coreprocessor debug tool
- CSM support for BG/L
- Mixed memory size partitions
- New toolchain

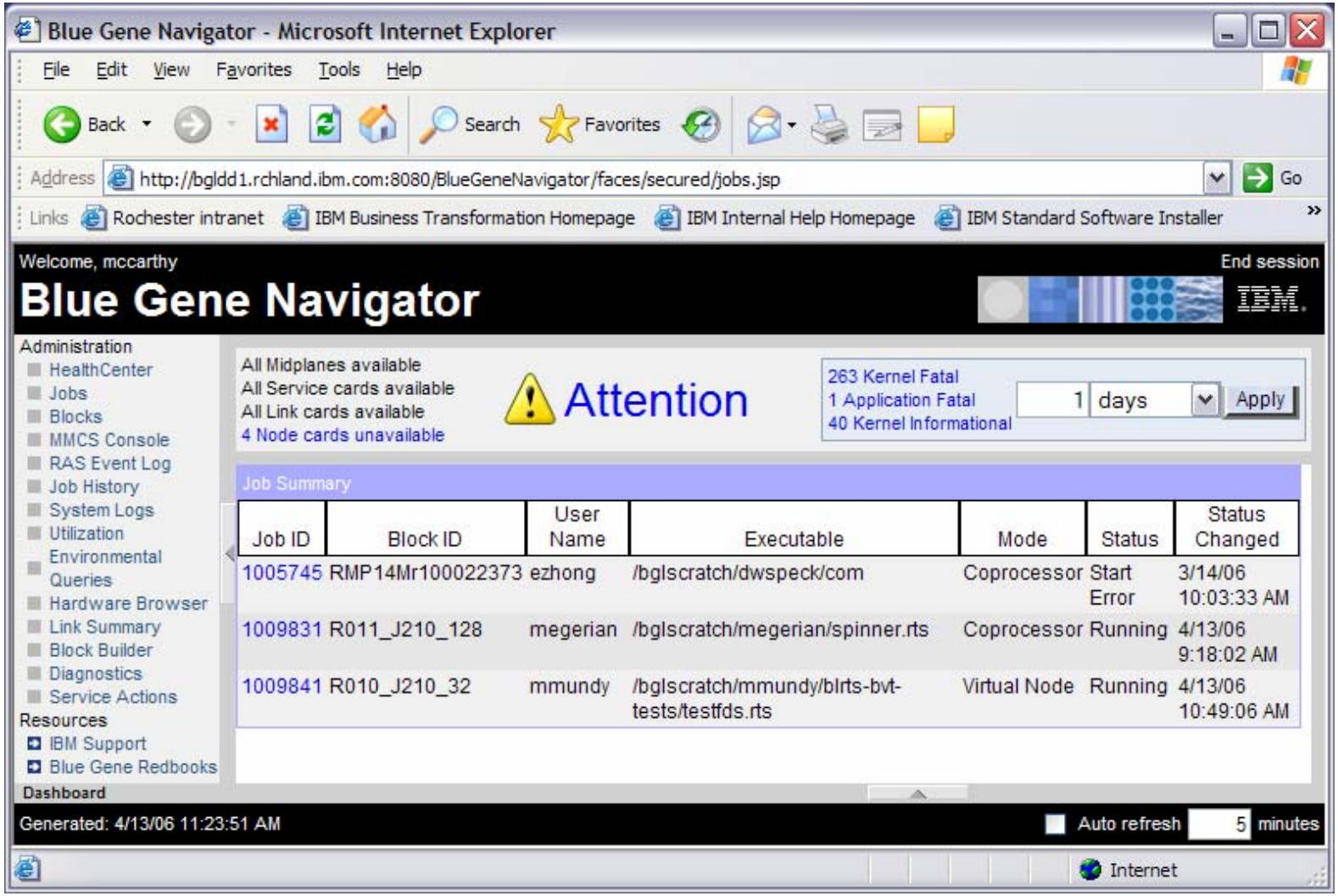
Blue Gene Navigator Overview

- Based on NIWS console work done by IBM and funded by NIWS
  - IBM enhanced NIWS console for BG/L
- Browser based admin console
  - Intended for Blue Gene administrators (not end users)
  - Mozilla Firefox 1.5 and IE 6 are the supported browsers
  - Uses Apache Tomcat on Service Node
- Monitor Blue Gene system health and utilization
- Browse hardware
- View blocks, running jobs and completed jobs
- Drill down to see details of interest, such as job details and RAS events
- Access the Midplane Management Control System (mmcs) console
- Find, browse, and search system logs
- Run system diagnostics, and examine results
- …. and more

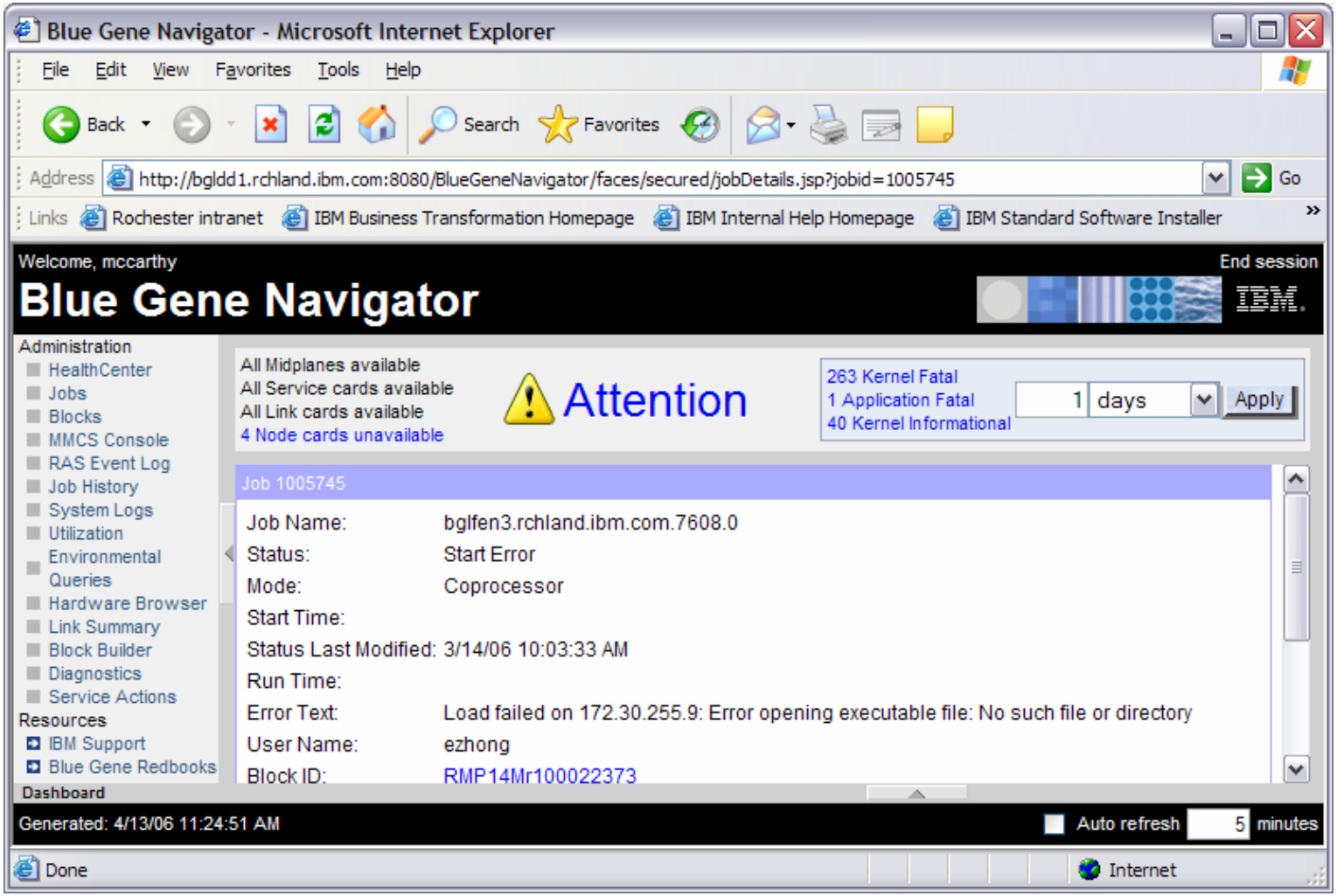
Blue Gene Navigator – Main screen



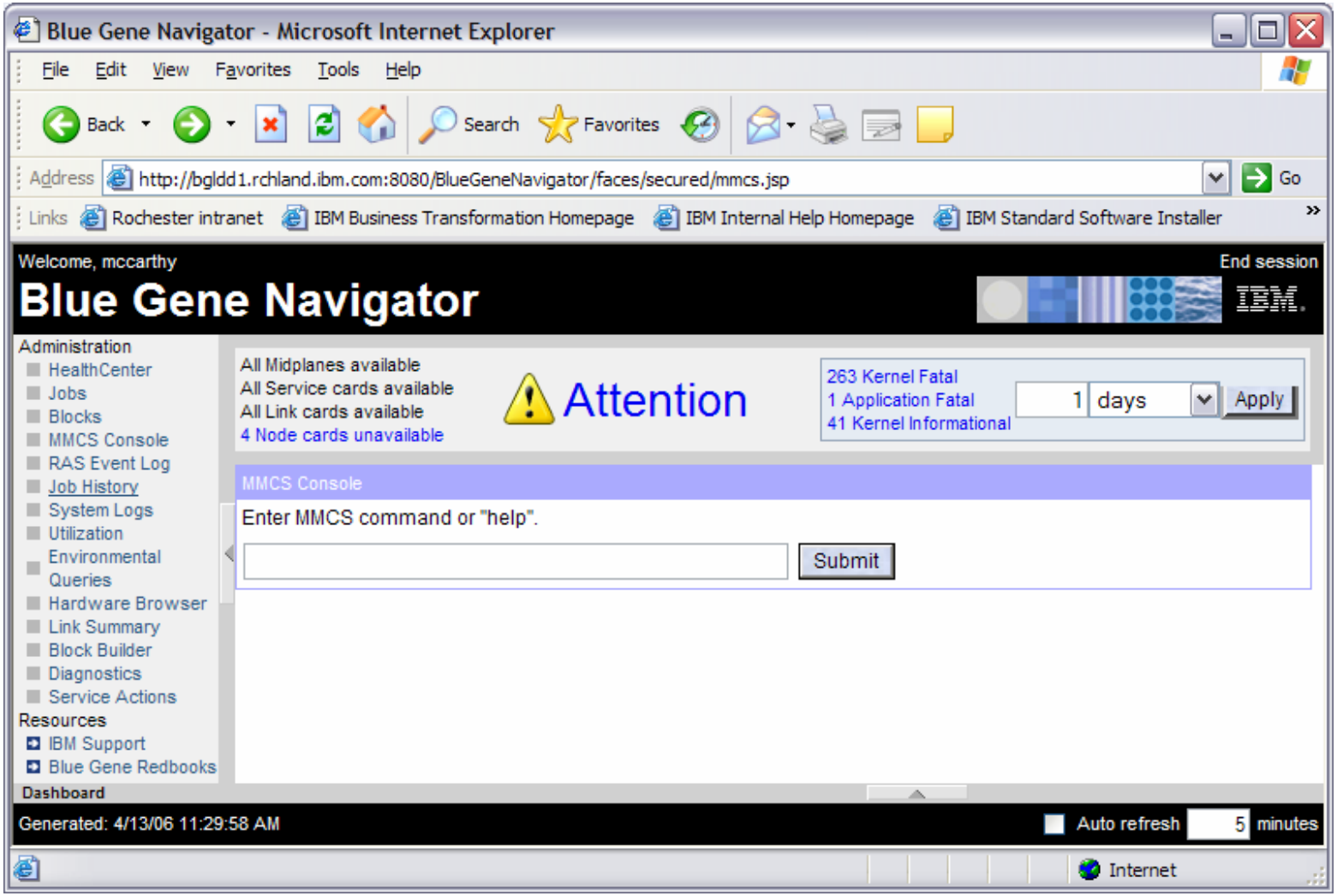
Blue Gene Navigator – Jobs



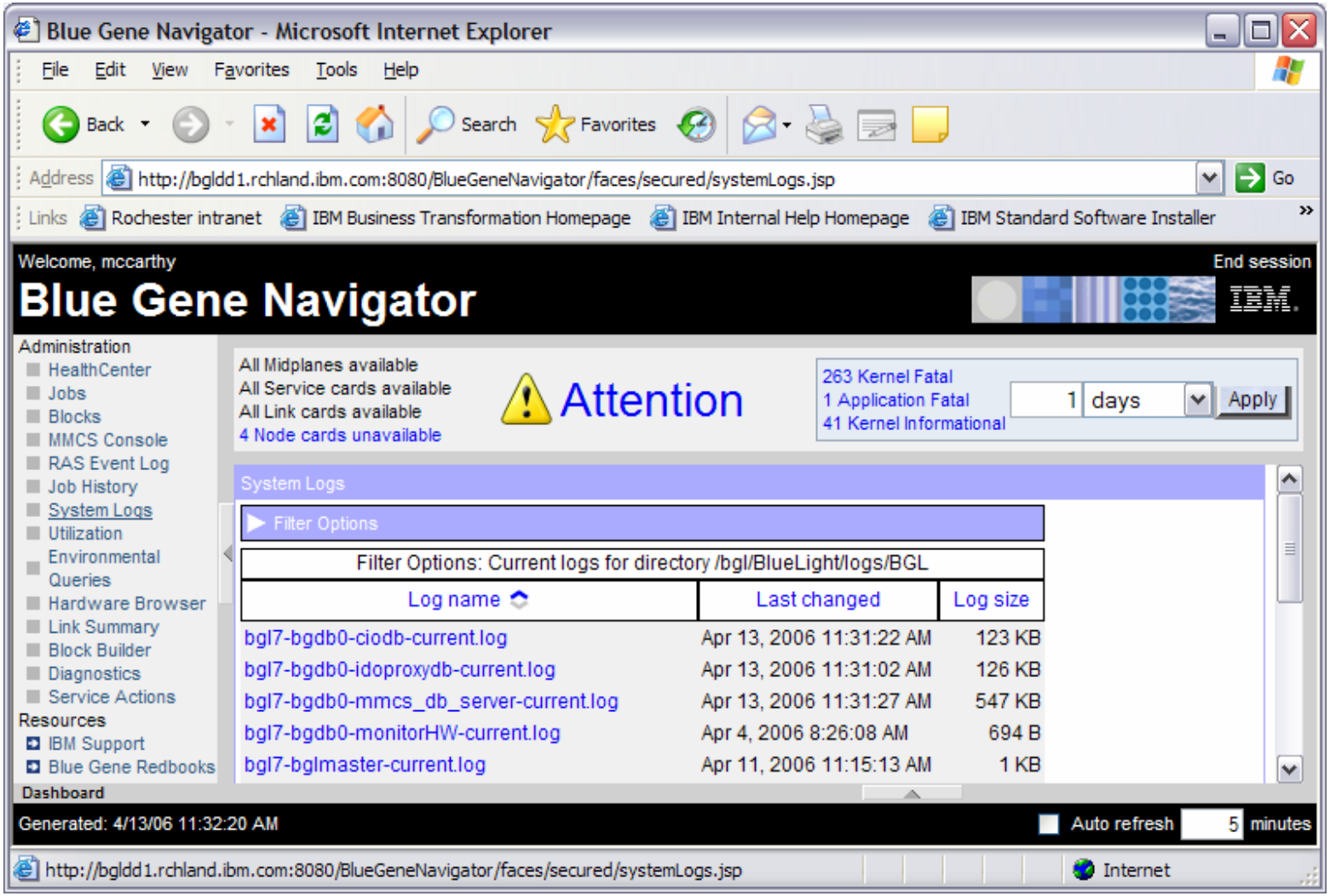
Blue Gene Navigator – Jobs drilldown



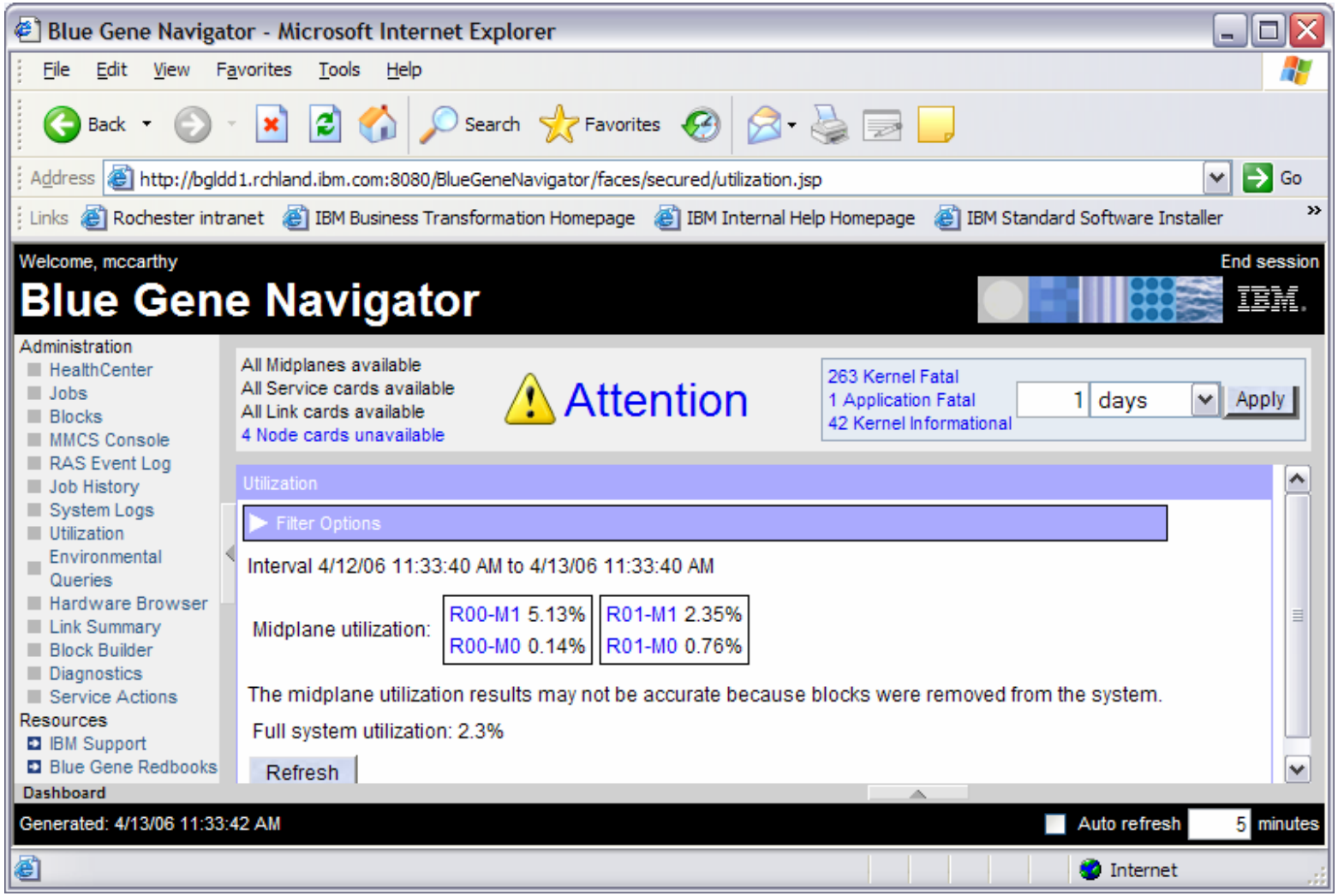
Blue Gene Navigator – MMCS Console



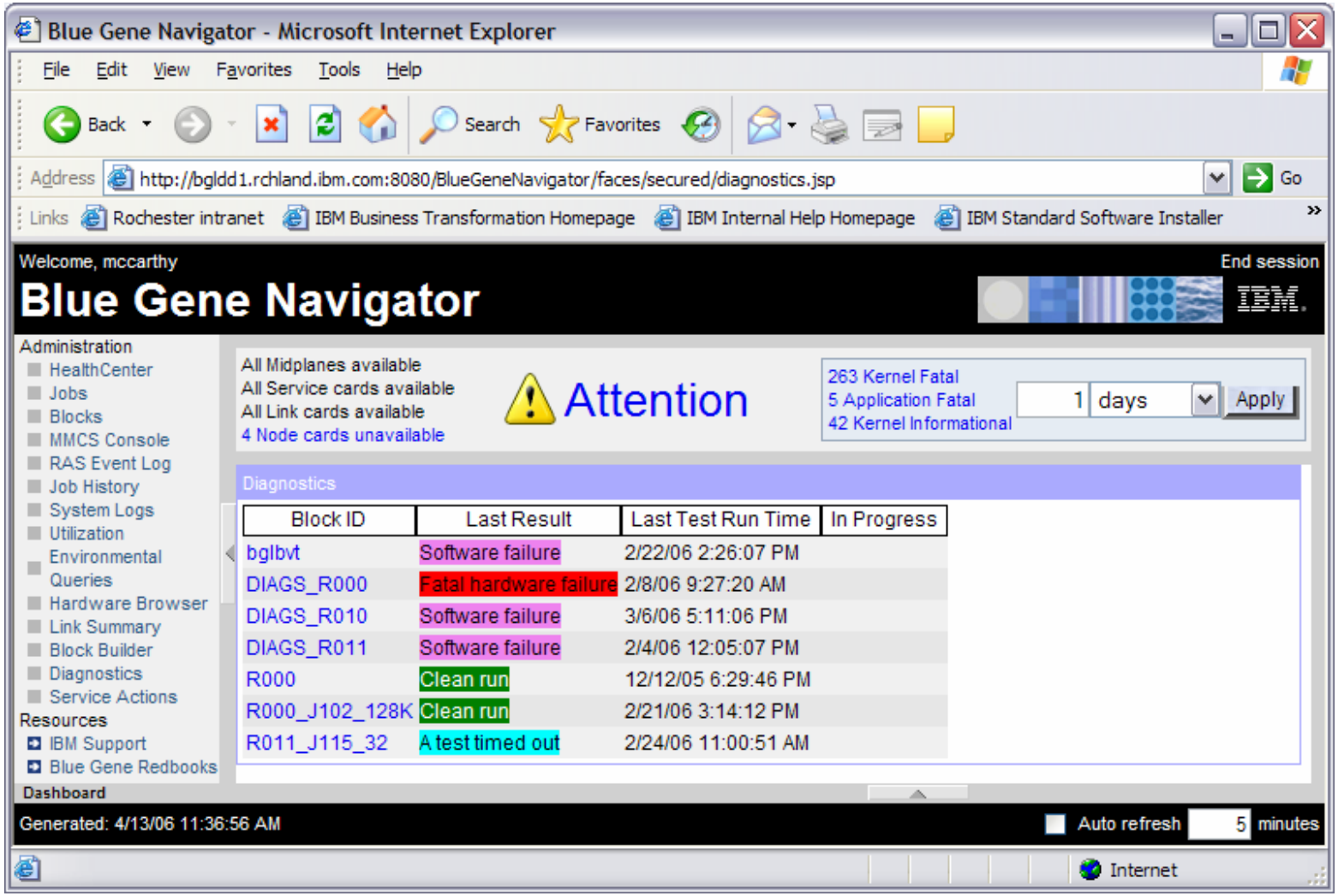
Blue Gene Navigator – System Logs



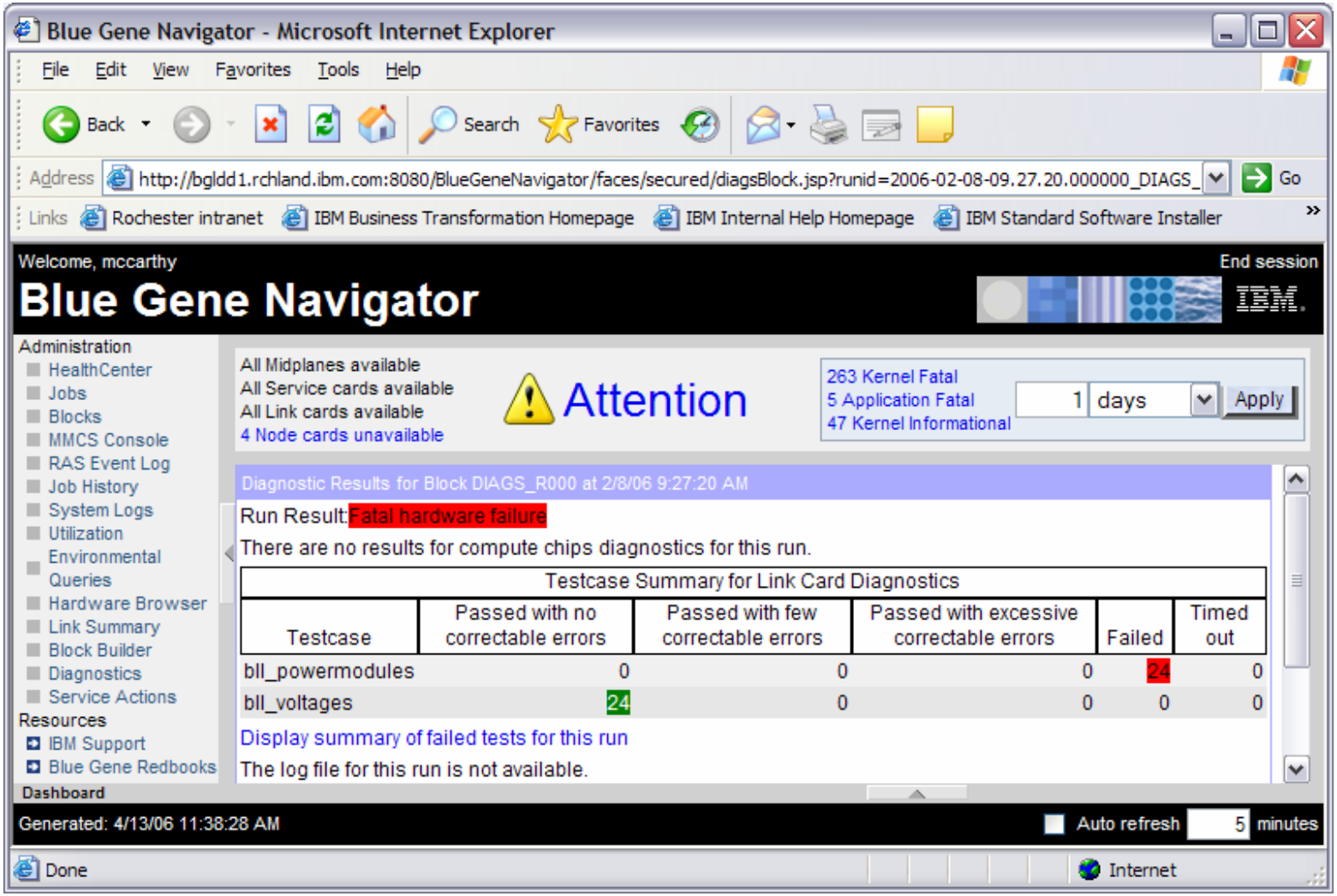
Blue Gene Navigator - Utilization



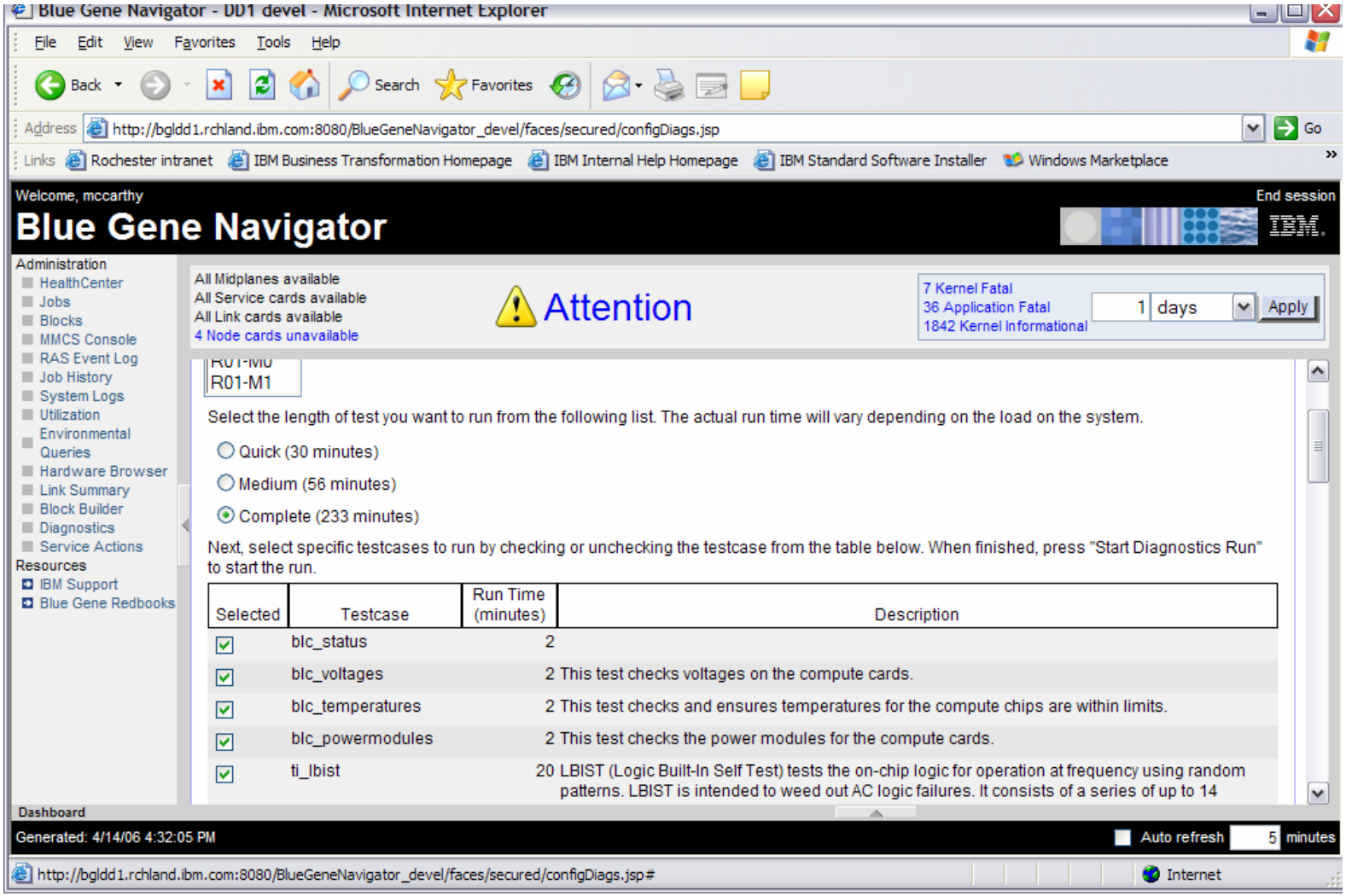
Blue Gene Navigator - Diagnostics



Blue Gene Navigator – Diagnostics drilldown



Blue Gene Navigator – Run Diagnostics



Improved External Performance Monitor

- **All release 2 functions still supported**
- **Performance monitor now multithreaded**
  - More efficient (faster)
- **Twenty two different groups of performance counters that can be monitored**
  - Match the counter definitions of the HPC toolkit
- **Derived floating point operation (flop) counter**
  - All inclusive counter for floating point activity on all nodes in the application
  - Helps to understand floating point percentage of peak
- **Collection of difference performance counters for a given application without recompiling the app**

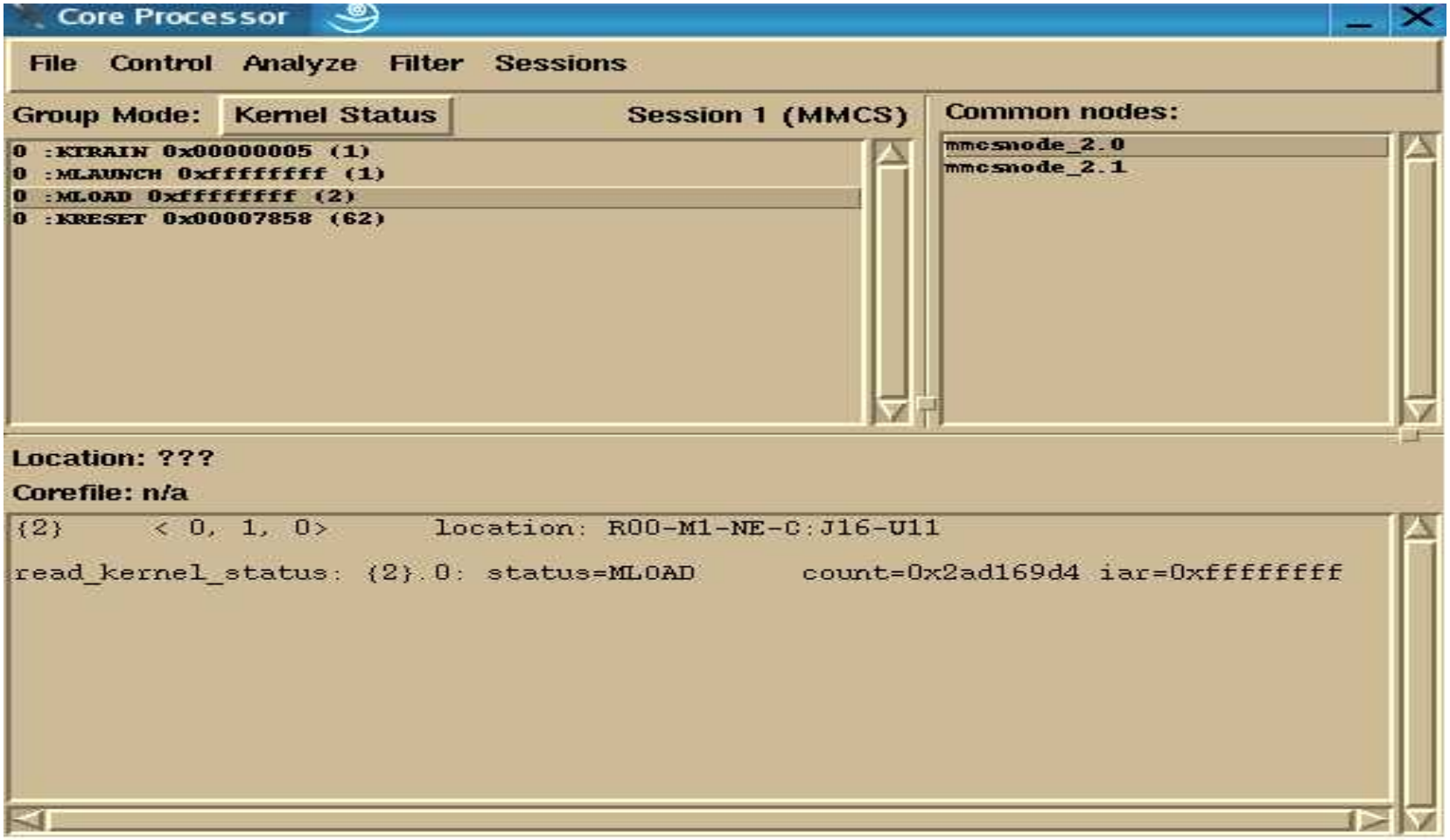
External Performance Monitor

- **Summary sample type**
  - Previously all sample intervals were saved
  - Greatly reduce the required storage
- **Selective of BG/L jobs based on block id and/or user name**
- **More then one instance of the monitor can run at the same time**
- **Data can be directed to two different data stores**
  - Flat file (can be imported into database later)
  - SQL database
- **GUI interface**
  - Real time analysis
  - Statistical analysis

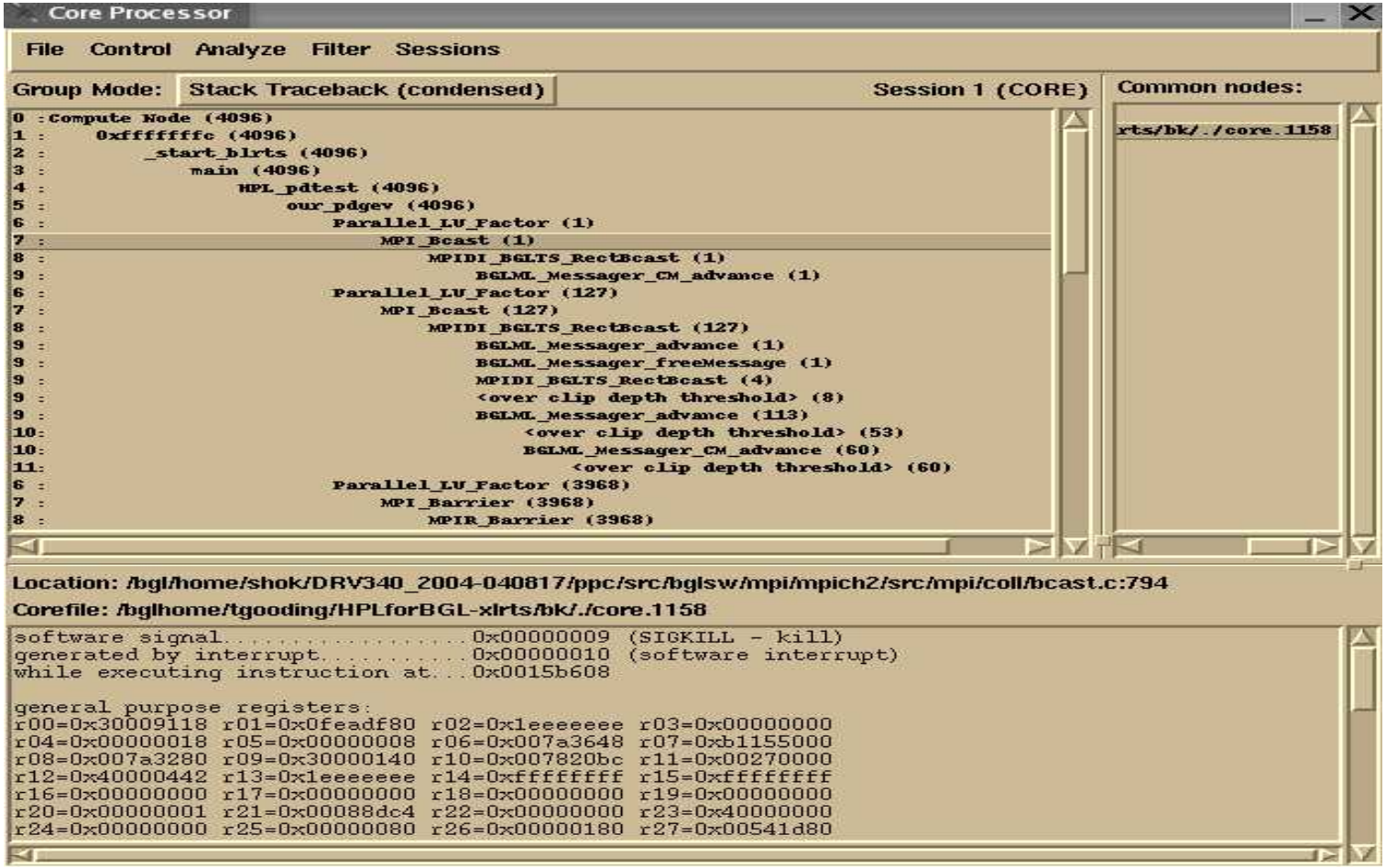
Coreprocessor debug tool

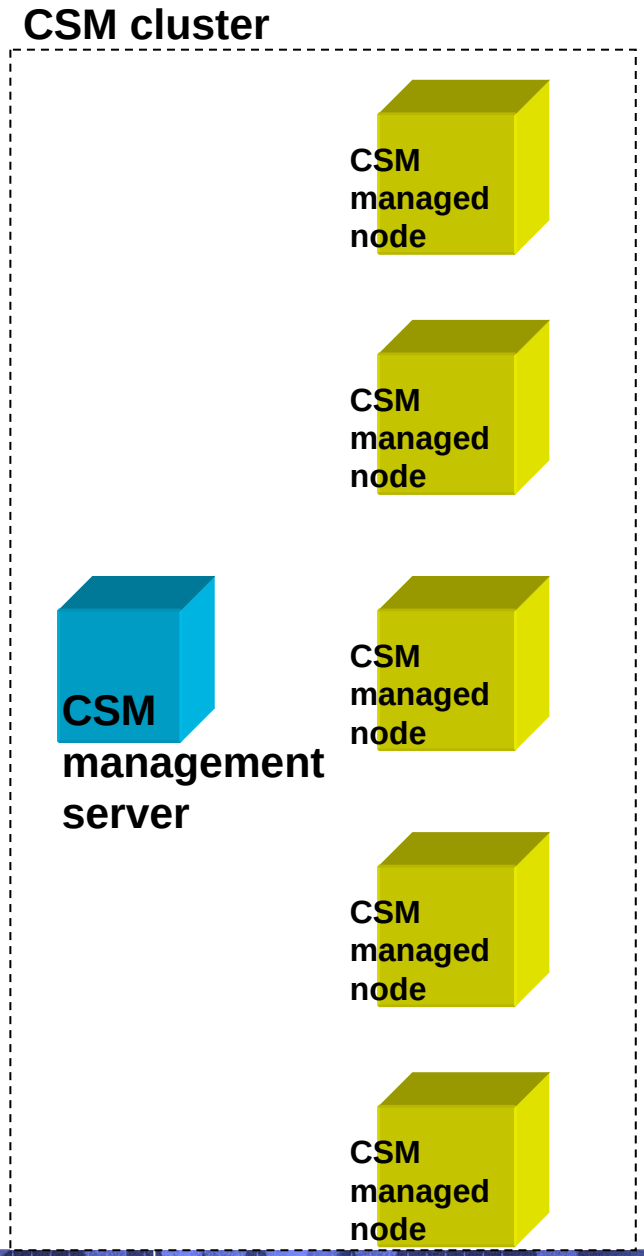
- **Enables parallel debug of problems at all levels**
  - Hardware
  - Kernel
  - Application
- **A debug tool based on the low level hardware JTAG interface to read and organize hardware information**
  - IARs, GPRs, SPRs, DCRs
  - Originally developed to process BG/L compute node core files but has evolved into a parallel debugger
- **No dependency on code running on the node**
- **Can sort nodes based on stack traceback and kernel status**
  - This feature allows isolation of a failing/problem node very quickly
- **Stack dumping on a per node basis**
- **Scales to 64K nodes (coprocessor mode) or 128 K nodes (virtual node mode)**
  - Used extensively during bringup of 64K nodes at LLNL

Coreprocessor – sorted by kernel status – 64 nodes



Coreprocessor – sorted by IAR – 4096 nodes



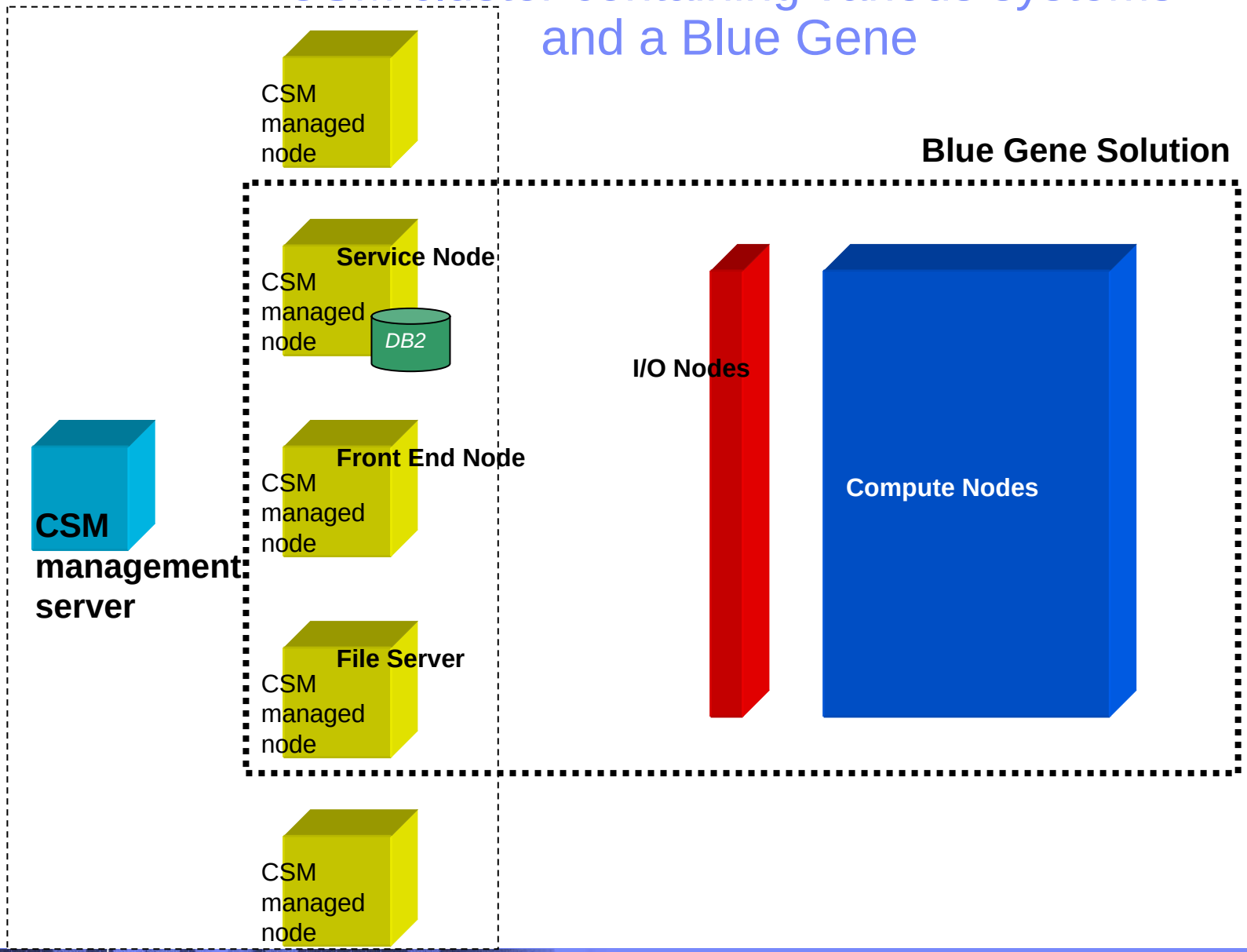


Cluster Systems Management (CSM) for AIX and Linux

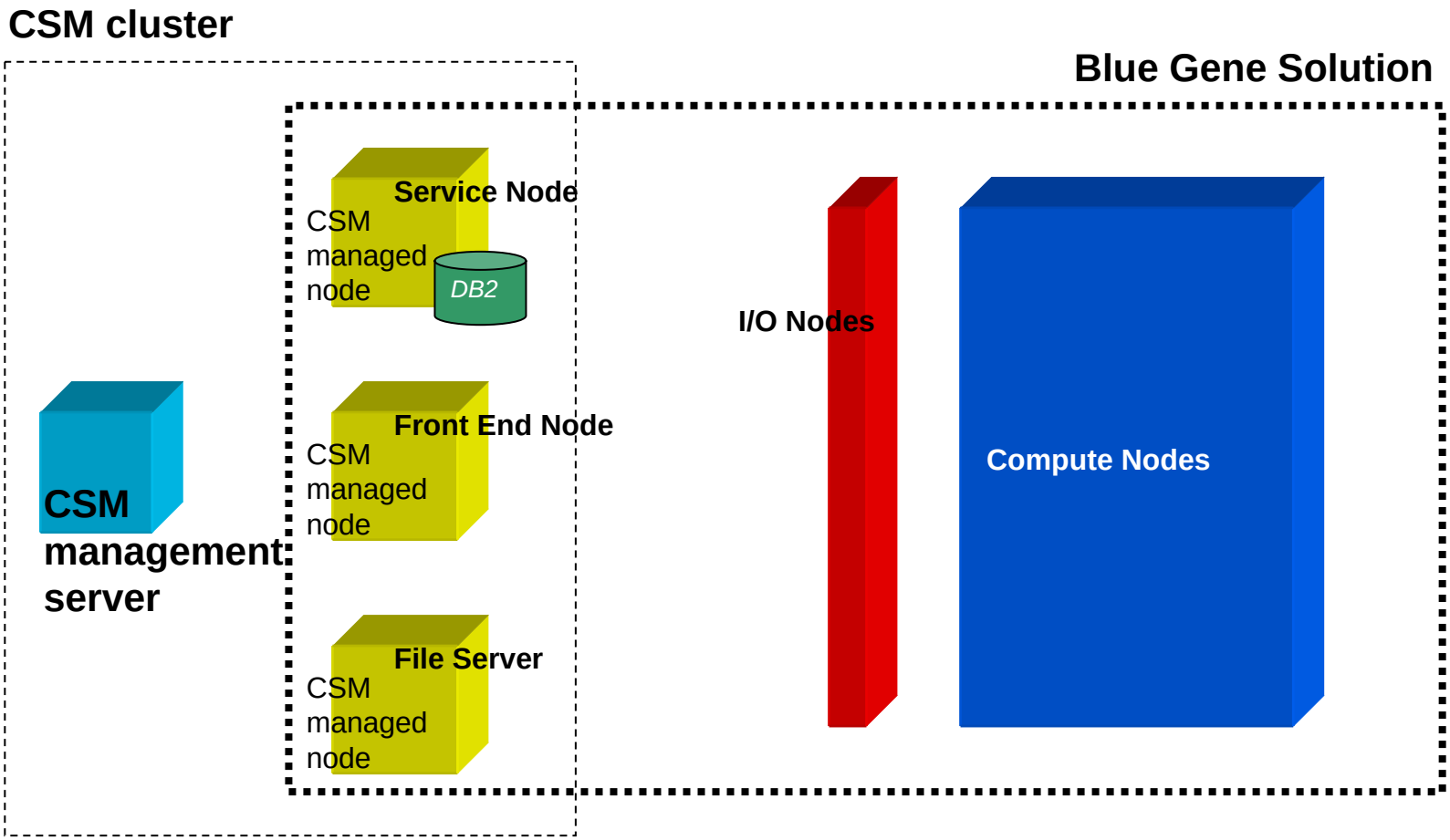
- **Single point of management for a broad range of pSeries & xSeries systems running AIX and/or Linux.**
- **Main capabilities include:**

- OS Install
- Software maintenance
- Distributed shell
- File Synchronization
- Remote Hardware Control
- Node Groups
- HA management server option
- Node monitoring
- **Blue Gene DB monitoring**

CSM cluster containing various systems and a Blue Gene



CSM cluster containing just a Blue Gene



(Note: The CSM management server need not be a separate system; it could be installed on the Service Node, a Front End Node, or a File Server.)

Why use CSM with your Blue Gene?

- Perform routine systems management of the Service Node, Front End Nodes, and File Servers from a single point of control.
- Monitor the Service Node, Front End Nodes, and File Servers for events of interest (e.g. daemon failure, /var full, etc.) and have responses of your choosing occur automatically when such events occur.
- Monitor the Blue Gene database in the same way without having to write complex DB2 Triggers and Stored Procedures. For example, CSM provides simple commands for things like this:
  - Watch the database for critical RAS events.
  - Whenever such events occur, send an e-mail containing detailed event information to the Blue Gene administrator.

Mixed Memory size partitions

- **Compute nodes can have either 512M or 1G DDR memory**
- **All nodes in the same midplane are same memory size**
- **Partitions can contain midplanes of nodes of different memory sizes**
- **Each node will be able to utilize all the memory on its own node**
- **Applications can determine memory size from the personality and adjust to utilize all memory**

New toolchain

- **New toolchain support**
  - Upgrades of several components:
    - gcc 3.2 to 3.4.3
    - binutils 2.13 to 2.16.1
    - glibc 2.2.6 to 2.3.6
  - Applications can not have libraries built with old and new tool chain
    - If the customer recompiles any application, all other associated programs, libraries, etc. must also be recompiled and relinked
  - Compilers upgraded with a PTF to BG/L V1R3:
    - IBM XL Fortran Advanced Edition V10.1.0.1 for Blue Gene
    - IBM XL C/C++ Advanced Edition V8.0.0.1 for Blue Gene

Summary

- **IBM continues to enhance the system management functions for Blue Gene**
- **The use of DB2 as the repository for all systems management information has enabled the development of new system management tools**
  - Blue Gene Navigator
  - CSM integration of Blue Gene
  - Multiple schedulers