High Application Availibility for HPEC

On Board for Mission Success

SKY COMPUTERS
A SUBSIDIARY OF ANALOGIC CORPORATION

High Application Availability

- **Percentage of time primary application is available**

$$\frac{MTBF}{MTBF+MTTR}$$

- **Common requirement is "5-9's" - 99.999%**
  - About 5 minutes down time per year
- **Failures caused by hardware, software or user**
- **Large HPEC system may have an MTBF of a few weeks**

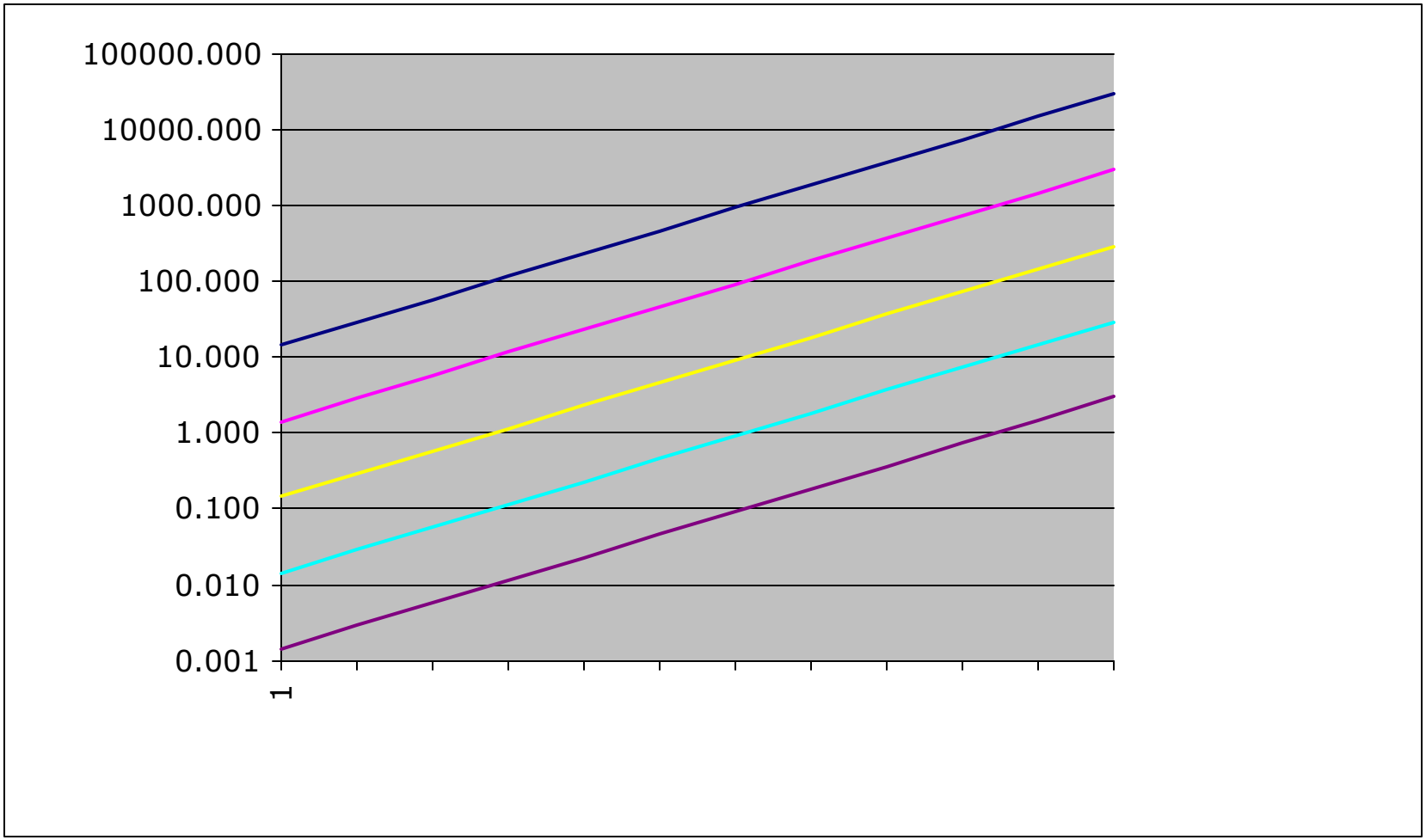
SKY COMPUTERS
A SUBSIDIARY OF ANALOGIC CORPORATION

System Design

- **Total system - hardware, system software and application - must be designed for HAA**
- **Typically n+m design in HPEC**
  - For each resource, n required for application
  - m additional provided for redundancy
  - Resources must be carefully identified - processors, memory, fans, power supplies, fabric connections, …
- **Recovery MUST be "automatic"**
- **Don't have time for human involvement**

SKY COMPUTERS
A SUBSIDIARY OF ANALOGIC CORPORATION

Availability



SKY COMPUTERS
A SUBSIDIARY OF ANALOGIC CORPORATION

Maximize MTBF

- **Prevent failures**
  - Careful electrical design
  - ECC/CRC error detection/correction
  - Good mechanical design including cooling
  - Good software design
  - Exhaustive test/debug
- **Preempt failures**
  - Online testing
  - Health monitoring
    - Temperatures, fan speeds, voltages
  - Opportunity to repair system before actual failure

SKY COMPUTERS
A SUBSIDIARY OF ANALOGIC CORPORATION

Fault Management

- Detection - determine that fault exists
- Diagnosis - identify failing component
- Isolation - protect rest of system from failures
- Recovery - get application running again
- Repair - replace or restart failing component

SKY COMPUTERS
A SUBSIDIARY OF ANALOGIC CORPORATION

Fault Management

- **Detection**
  - Hardware detected - ECC/parity errors, link status change, …
  - Software detected - timeouts, inconsistent answers, …
  - Must be detectable by reliable resource
- **Diagnosis**
  - Identify failed resource(s)
  - Repair not needed if n resources still available

SKY COMPUTERS
A SUBSIDIARY OF ANALOGIC CORPORATION

Fault Management

❒ Isolation

- InfiniBand “automatic path migration” to use alternate path through fabric
- Software re-configuration of routing tables in InfiniBand switches
- Remove processors from CORBA scheduler
- Other application specific choices

❒ Recovery

- Restart/resume the application with reduced configuration
- Detection to Recovery can be accomplished in a fraction of a second, perhaps milliseconds depending on failure

SKY COMPUTERS
A SUBSIDIARY OF ANALOGIC CORPORATION

Fault Management

❒ Repair

- Since most likely root cause is software fault, reset/restart may be all that is required
- Run detailed diagnostic
  - Verify failure and locate Field Replaceable Unit (FRU)
- Return still functional resources to use
- Technician replaces FRU
  - InfiniBand supports “Live Insertion”
- Return repaired component to use

SKY COMPUTERS
A SUBSIDIARY OF ANALOGIC CORPORATION

Some Resources

- **Service Availability Forum**
  http://developer.intel.com/platforms/applied/eiacomm/saforum.htm
- **Linux High Availability Project**
  http://linux-ha.org
- **Real-time CORBA, Dynamic Scheduling**
  http://www.omg.org
- **Telco oriented High Availability**
  http://www.goahead.com/products/products.htm
  http://www.ccpu.com/telco_products/middleware.html
  http://www.mvista.com/cge/index.html

SKY COMPUTERS
A SUBSIDIARY OF ANALOGIC CORPORATION

Conclusion

- HAA requires careful SYSTEM level analysis/design - hardware, system software and application must ALL cooperate
- Emerging fabrics like InfiniBand enable HAA capabilities not available previously for HPEC applications
- 5 step fault management process useful for design of HAA applications