

open



USE



IMPROVE



EVANGELIZE

Making Grid Engine HA with **Open** High Availability **Cluster** and **OpenSolaris**

Availability Engineering
Sun Microsystems

開
放
的
열린
مفتوح
libre
मुक्त
ಮುಕ್ತ
livre
libero
ముక్త
开放的
açık
open
nyílt
πικρό
オープン
livre
ανοικτό
offen
otevřený
öppen
открытый
வெளிப்படை



Outline

- What is “Open High Availability Cluster”?
- Why is High Availability Important?
- How Solaris Cluster Provides High Availability
- Cluster Agents
- Grid Engine without Open HA Cluster
- Grid Engine with Open HA Cluster
- Benefits of combining Grid Engine with Open HA Cluster



What is Open High Availability Cluster?



Open High Availability Cluster

- The open-source code base for Solaris™ Cluster
 - Based on Solaris Cluster 3.2
- Hosted by the HA Clusters Community Group
 - <http://www.opensolaris.org/os/community/ha-clusters/ohac/>
- Source Code available under the CDDL
- Binary Distribution available
 - Solaris Cluster Express 02/08 runs on top of Solaris Express Developer Edition 01/08 (build 79b)
 - <http://opensolaris.org/os/community/ha-clusters/ohac/Documentation/SCXdocs/SCX/>



Why is Sun Open-Sourcing Solaris Cluster?

- Share
 - Public should have access to the source-base
 - We hope developers will contribute back their agents and other modifications
- Open Development
 - Increase confidence in the cluster code-base
 - Enhance the code-base with third-party contributions
- Shorten Feedback Loop
 - Easier to incorporate feedback
 - Find and fix bugs quicker
 - Early adoption of new features



Why Is High Availability Important?



Why is HA Important?

- Downtime is costly
- Failures are inevitable
 - Hardware, Software, Human Error, Disaster
 - Automated recovery the goal
- With a single physical system, single points of failure are catastrophic
 - Network card dies, CPU misbehaves, Disk drive crashes ...

HA Clusters automate the recovery process from inevitable failures, minimizing downtime and cost



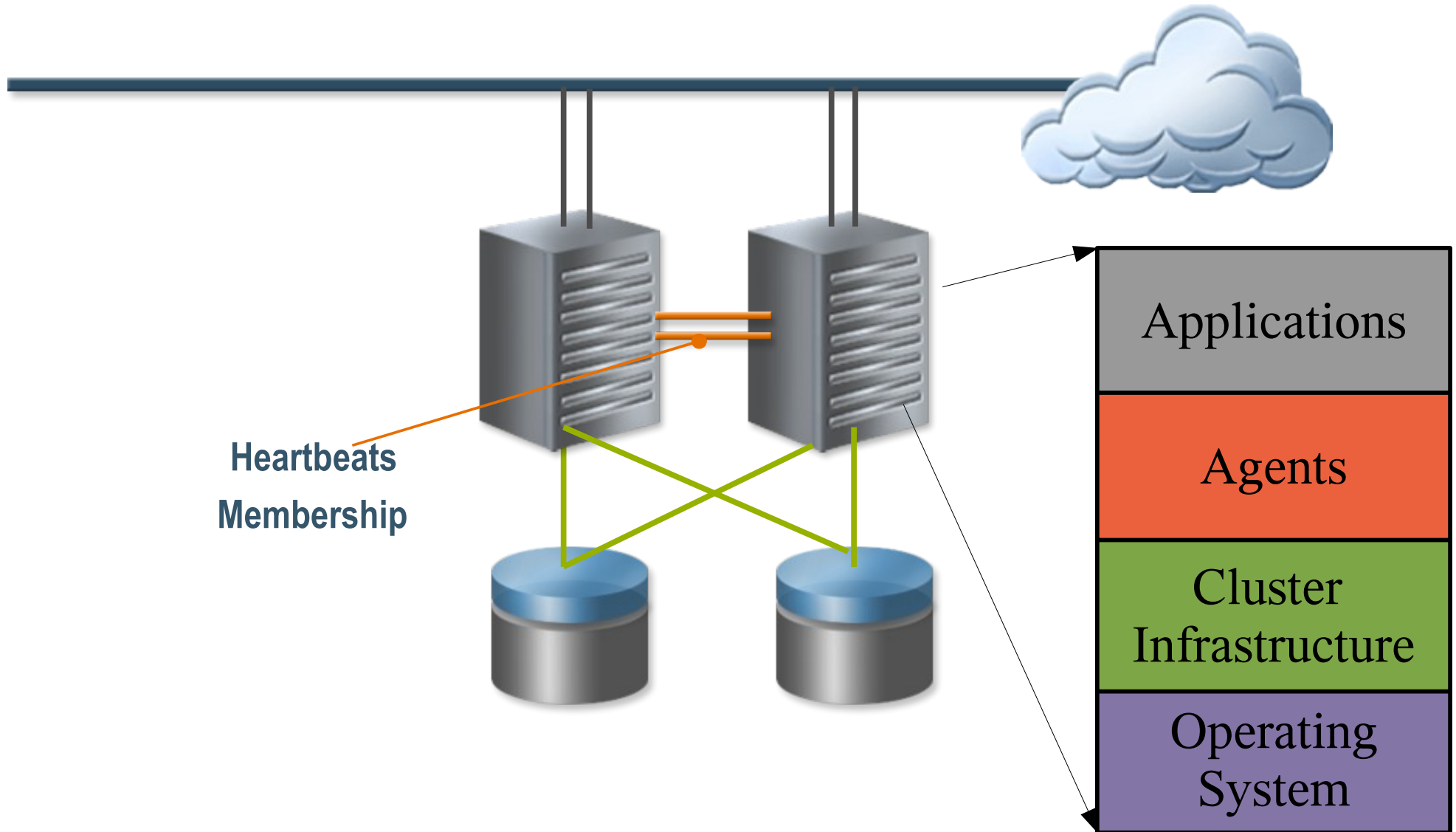
How Solaris Cluster Provides High Availability



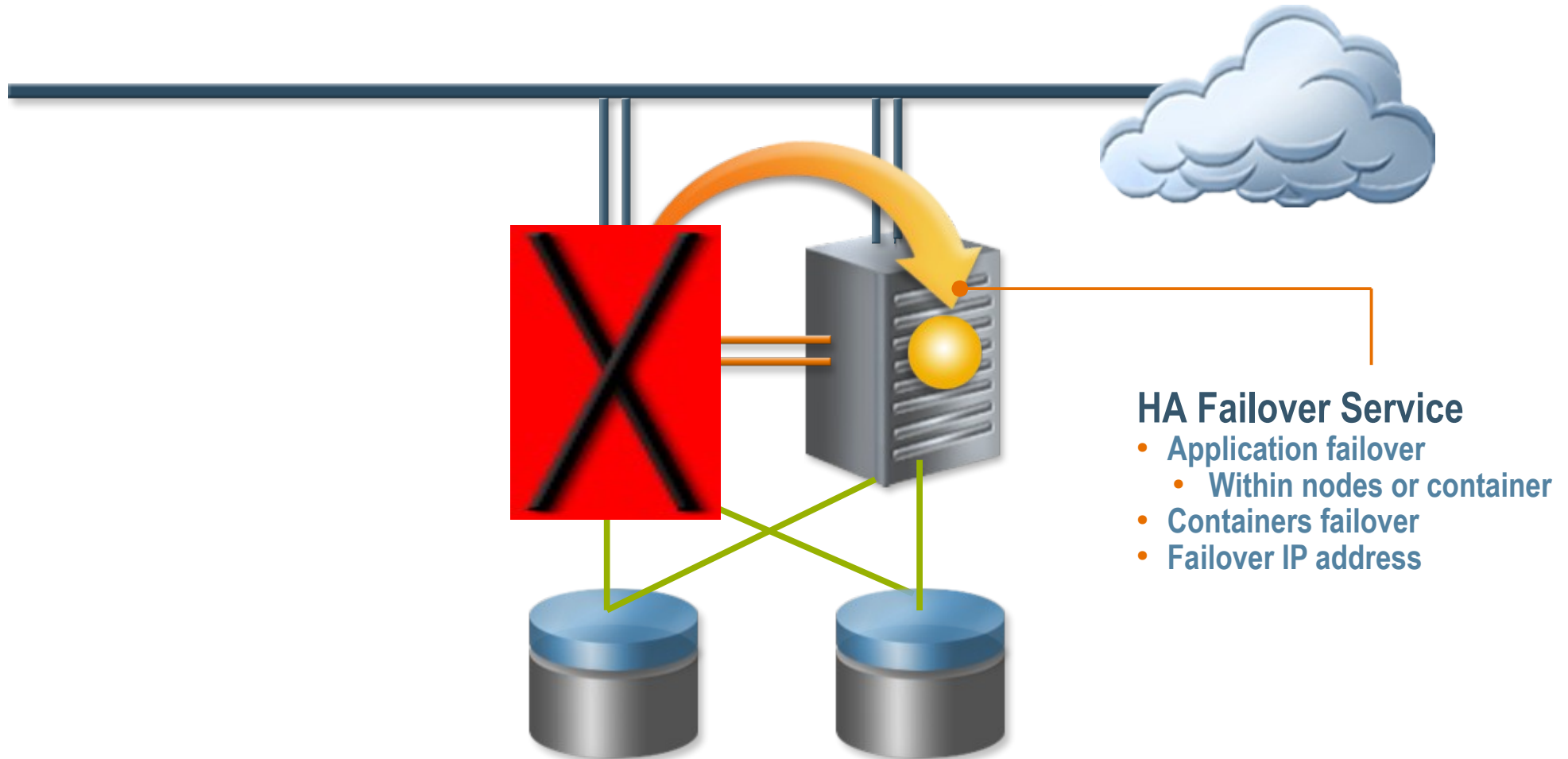
Platform for High Availability

- Tolerates Single Points of Failure (and some double failures)
- Hardware redundancy with off-the-shelf hardware
- Robust software HA infrastructure
 - Monitors the health of cluster infrastructure (hosts, shared storage, and network)
 - Orchestrates recovery of applications and cluster infrastructure (shared storage and IP networking)
 - Integrated tightly with Solaris Operating System
 - Robust membership including quorum to prevent partitions
 - Disk Fencing ensures Data Integrity in spite of failures

Solaris Cluster Stack

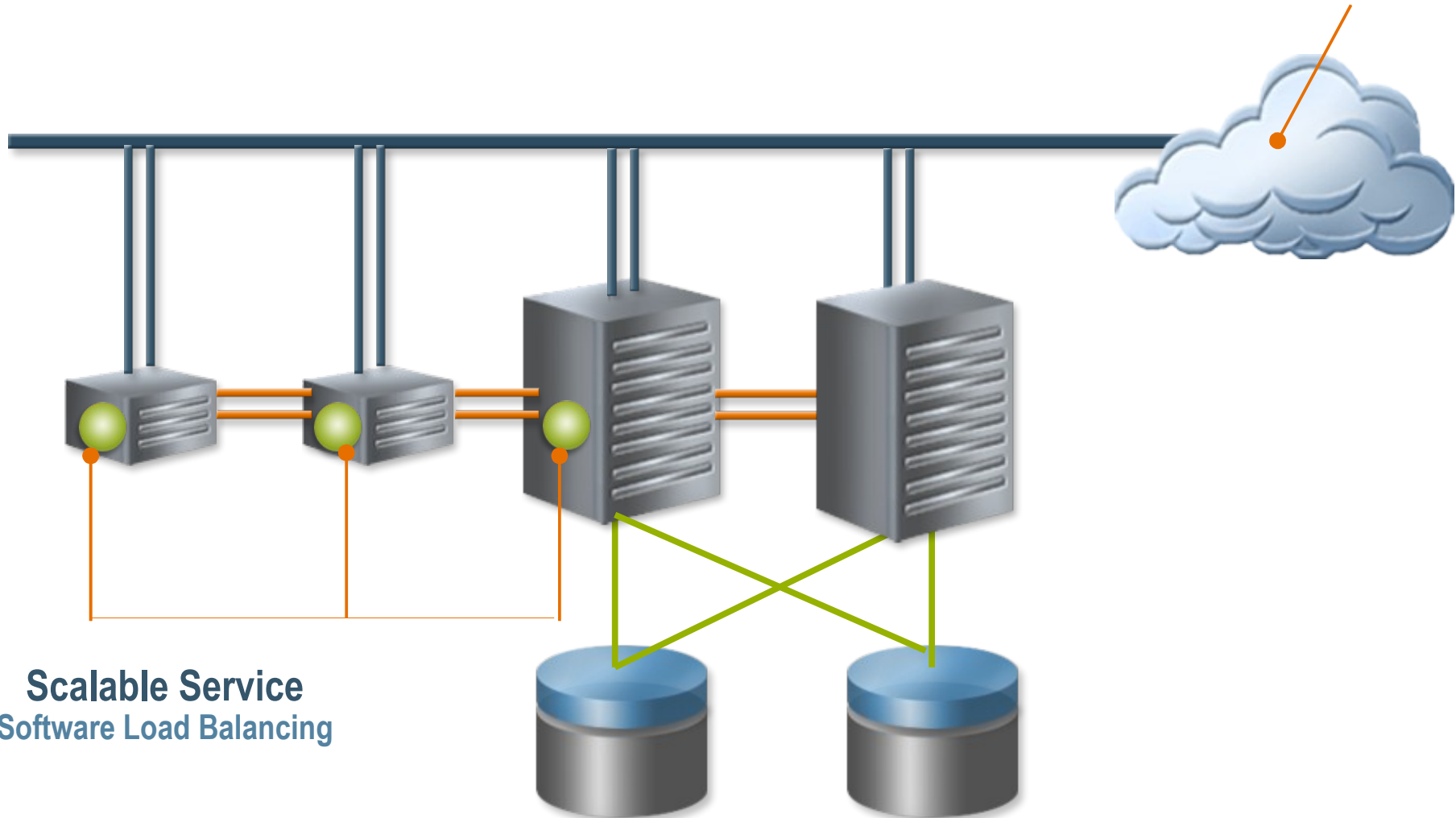


Failover Service



Scalable Service

Global Network Service
Provides Global IP address
with failure protection



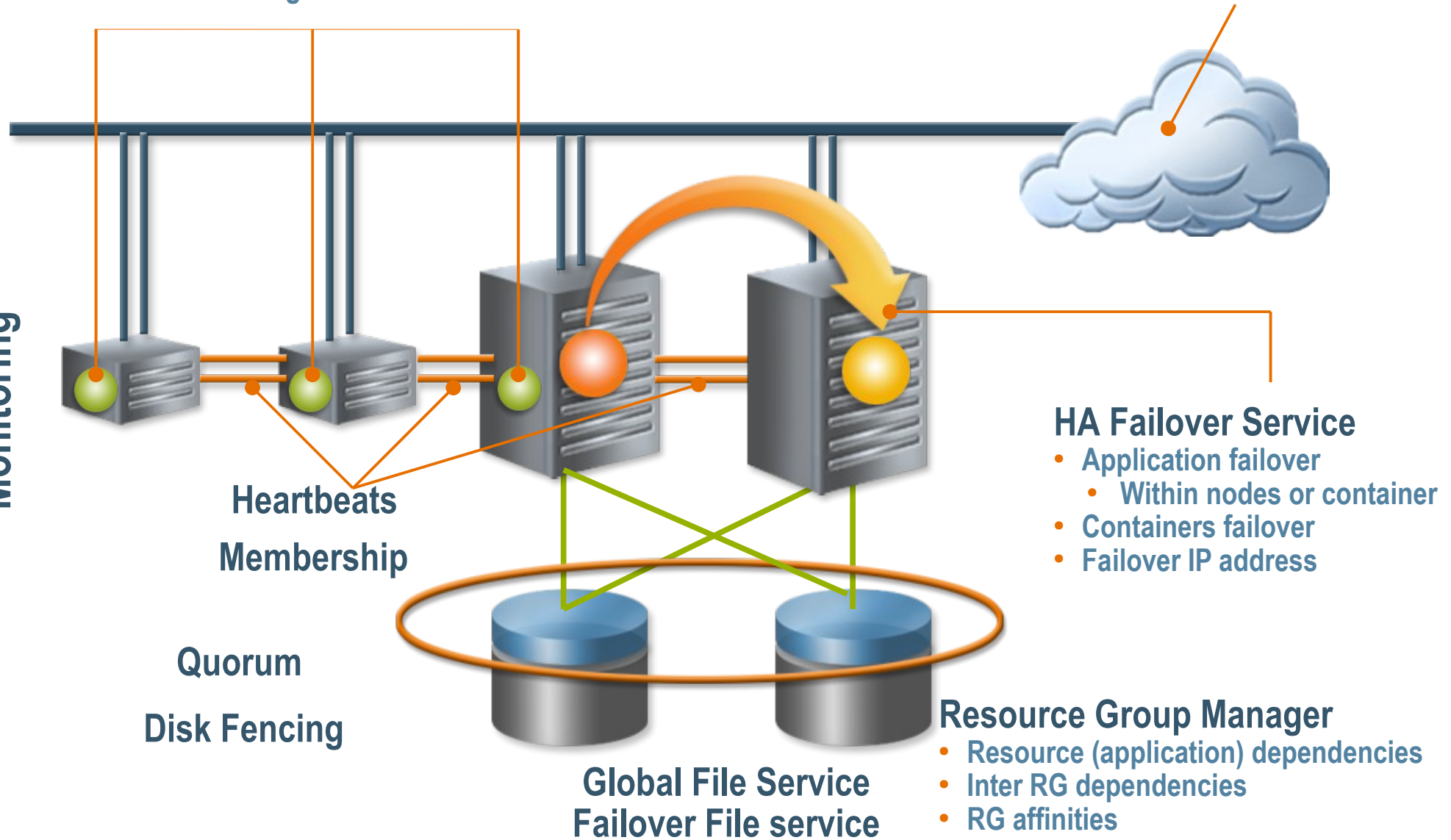
Scalable Service
Software Load Balancing

Solaris Cluster Architecture

Scalable Service
Software Load Balancing

Global Network Service
Provides Global IP address
with failure protection

Monitoring



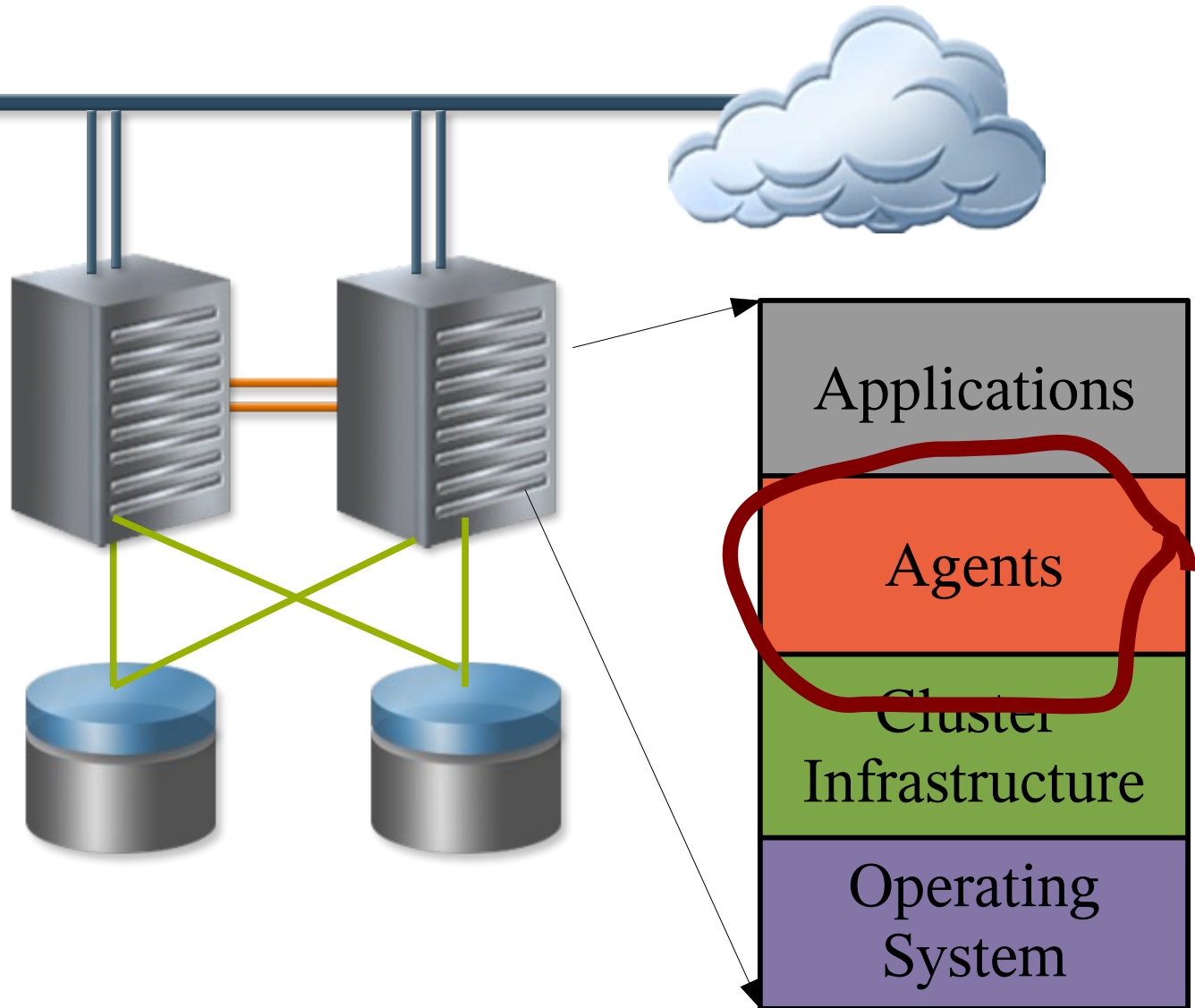


Cluster Agents

Solaris Cluster Stack

HA Data Services used:

- HA Grid Engine (SUNW.gds)
- HA NFS (SUNW.nfs)
- SUNW.HAStoragePlus
- SUNW.LogicalHostname





Cluster Agents (Data Services)

- Applications run on cluster unmodified (off-the-shelf)
- Cluster Agents are the “glue” layer between applications and cluster infrastructure
 - Custom agent for each application
 - ♦ Interacts with cluster core through APIs
 - ♦ Provides start, stop, and other commands specific to the application to be called by the cluster framework
 - ♦ Provides monitor daemon specific to the application
- Application cannot break into the high availability business on Solaris Cluster without an agent!



Grid Engine without Open HA Cluster

Grid Engine HA without Cluster

Master Host

- sge_qmaster
- sge_schedd
- IP Master Host

Shadow Host

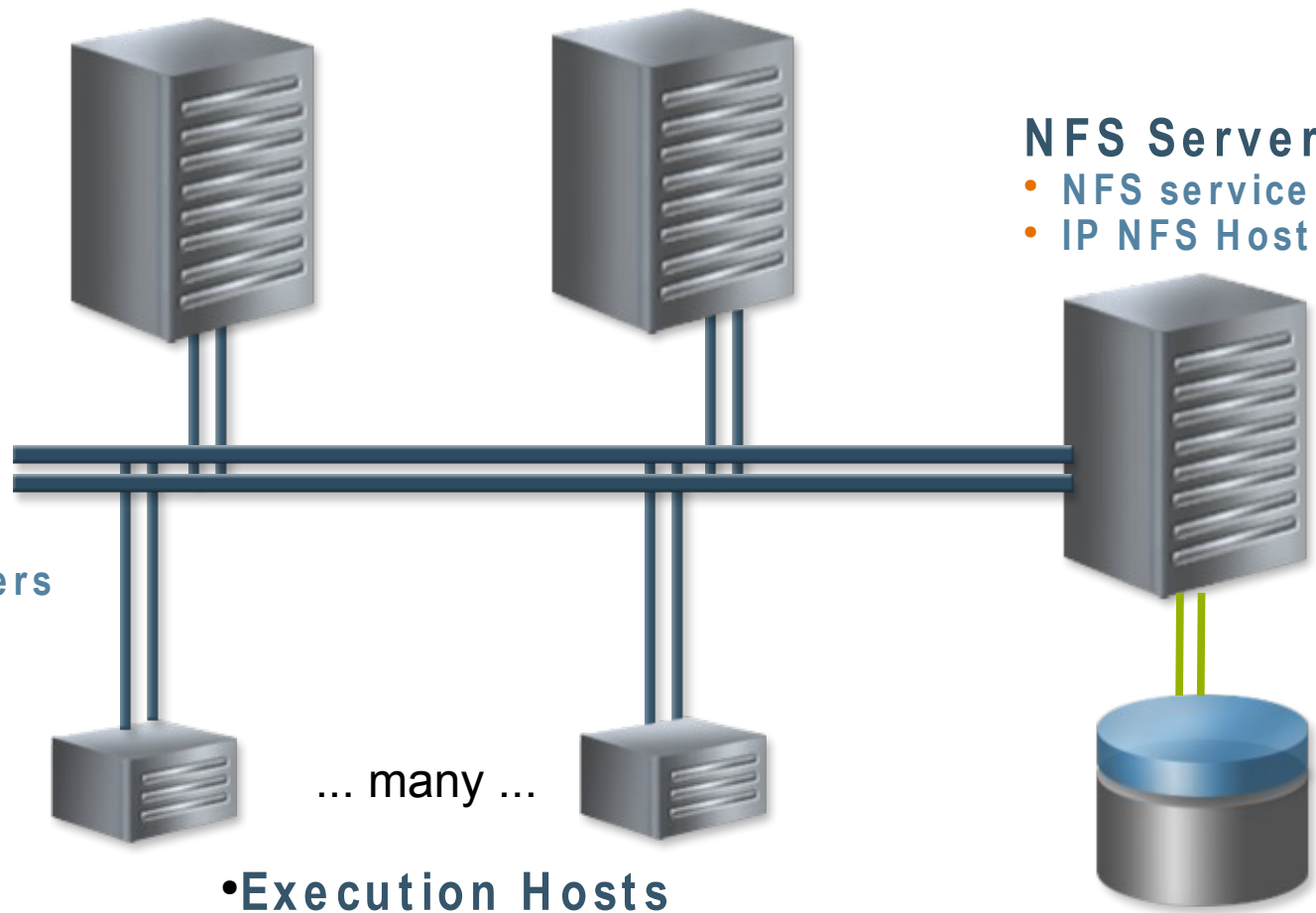
- sge_shadowd
- IP Shadow Host

NFS Server

- NFS service
- IP NFS Host

NFS Share

- \$SGE_ROOT
- contains files
 - act_qmaster
 - shadow_masters



Execution Hosts

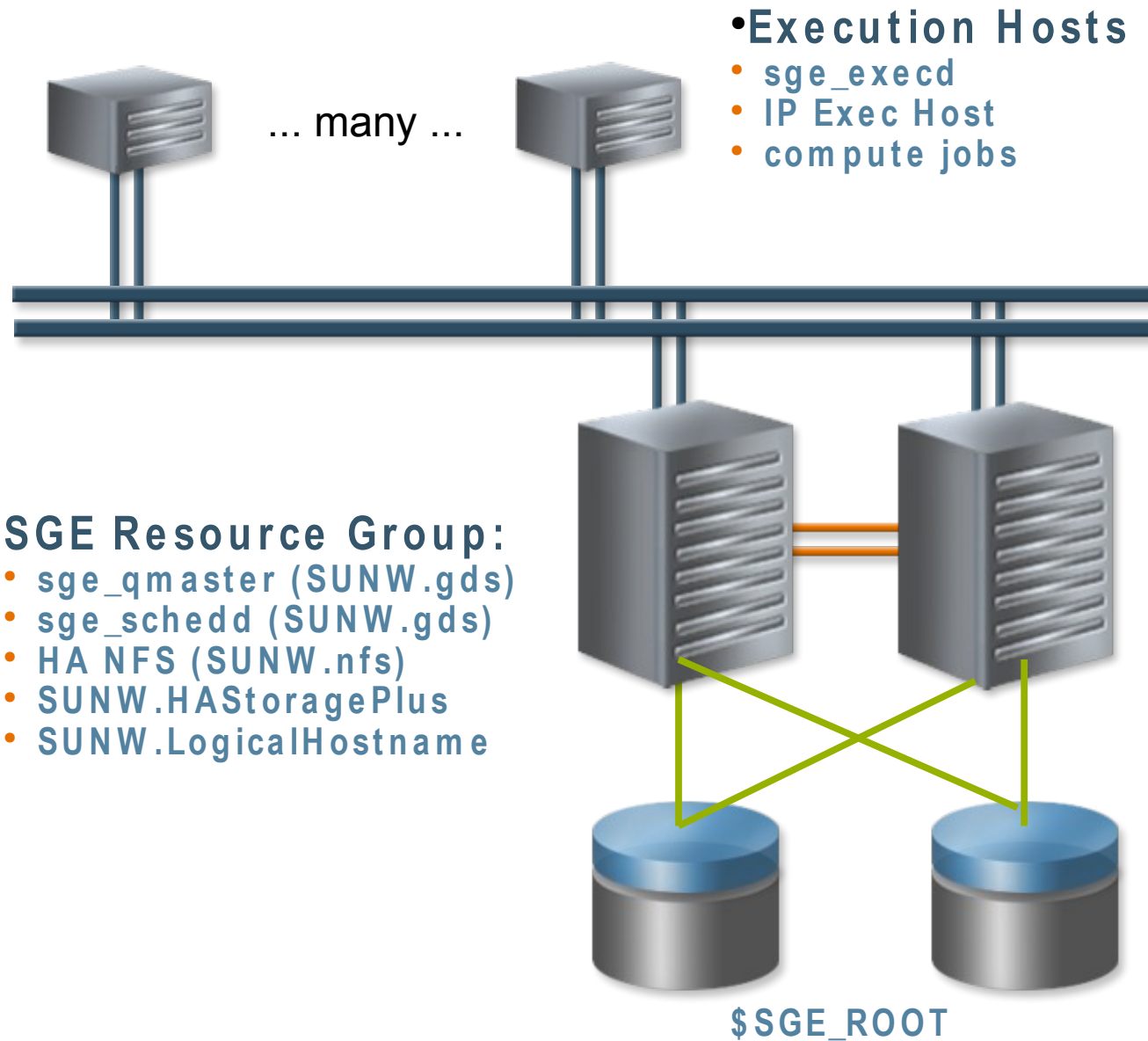
- sge_execd
- IP Exec Host
- compute jobs

\$SGE_ROOT



Grid Engine with Open HA Cluster

Grid Engine with Open HA Cluster





Benefits of combining Grid Engine with Open HA Cluster



Benefits

- No additional HA-NFS filer needed
- Grid Engine binaries for execution hosts can be shared via NFS from same Cluster
- Reliable and robust failover mechanisms
- Unique and reliable service IP address for execution hosts
- Mechanism to reflect necessary service dependencies
- Faster and finer grained reaction upon individual service failures



References



References

- Sun Cluster Data Service for Sun Grid Engine Guide
 - <http://docs.sun.com/app/docs/doc/820-3702?a=expand>
- Sun Cluster Data Service for NFS Guide
 - <http://docs.sun.com/app/docs/doc/820-2565?a=expand>
- Blog about Grid Engine on Solaris Cluster
 - http://blogs.sun.com/SC/entry/support_of_sun_grid_engine
- Shortcut to Date Services Source
 - <http://opensolaris.org/os/community/ha-clusters/ohac/Documentation/Agents/open-agents/>
- Sun N1 Grid Engine 6.1 Administration Guide
 - <http://docs.sun.com/app/docs/doc/820-0698/eqqgz?a=view>

open



USE



IMPROVE



EVANGELIZE

Questions?

Thank you!

Ashutosh Tripathi

ashutosh.tripathi@sun.com

<http://blogs.sun.com/sc/>

“open” artwork and icons by chandan:

<http://blogs.sun.com/chandan>

開
放
的
열린
مفتوح
libre
मुक्त
ಮುಕ್ತ
livre
libero
ముక్త
开放的
açık
open
nyílt
•••••
πικρ
オープン
livre
ανοικτό
offen
otevřený
öppen
ОТКРЫТЫЙ
வெளிப்படை