

# Emerging Standards: The Open Cluster Framework

## Defining Standard APIs for Clustering

**Alan Robertson**  
**IBM Linux Technology Center**  
***alanr@unix.sh***

# Agenda

- ◆ **Why Standard APIs for Clustering?**
- ◆ **What is Open Clustering Framework?**
- ◆ **API areas being addressed**
- ◆ **OCF Cluster Conceptual Model**
- ◆ **API details**
- ◆ **Future Work**

# Why Standard Clustering APIs?

- ◆ **At least 4 OSS HA products**
- ◆ **10-20 Proprietary HA Products**
- ◆ **A large number of HPC "offerings"**
- ◆ **Each has it's own API and expectations**
- ◆ **No "900 pound gorillas" in the Linux market to create/dictate "standards"**
- ◆ **Confusion, impairing the progress of clustering on Linux**

# What is the OCF?

- ◆ **An open group of industry and academic providers and users of clustering services who are defining some standard APIs for clustering**
- ◆ **Most OCF APIs are generally intended to be usable by both high-performance and high-availability clustering platforms**
- ◆ **A working group of the Free Standards Group**

# Open Clustering Framework

- ◆ **Two-pronged approach**
  - **Define standard cluster APIs**
  - **Create component-based reference implementation**
- ◆ **Both proceed together**
- ◆ **The standards will be a 900-pound penguin...**



# Project Structure

## API Definition

**Select Areas of Interest**  
**Create Subteams**  
**Define APIs**  
**Reach agreement**  
**Publish APIs for review**  
**Refine APIs**

## Reference Implementation

**Create Plumbing/Infrastructure**  
**Coordinate with API definition**  
**Define Framework components**  
**Implement components**  
**Test result**  
**Provide as Open Source**

# Properties of the APIs

- ◆ **Implementation Neutral (agnostic)**
- ◆ **Royalty-Free**
- ◆ **For OSS or proprietary software**
- ◆ **Creates opportunities for interoperability**
- ◆ **Focused on Linux, but not limited to Linux**

# API Areas of Interest

- ◆ **Resource Services**
- ◆ **Event services**
- ◆ **Node services**
- ◆ **Recovery**
- ◆ **Group Services**
- ◆ **Low Level Communication Services**
- ◆ **Fencing**
- ◆ **DLM**
- ◆ **External Interfaces (GUI, SNMP, logging, etc.)**

# OCF Cluster Conceptual Model

- ◆ **A cluster is a collection of nodes (computers)**
- ◆ **Failures in the cluster occur asynchronously and are observed stochastically and independently**
- ◆ **Each cluster is divided into zero or more partitions (by communication failures, etc.)**
- ◆ **Each active node belongs to exactly one partition at a time**
- ◆ **One of these partitions may be named the “primary” partition. This partition is said to “have quorum”**

# Partitioning Example

## *Partition 1 (primary)*

**Node  
A**

**Node  
B**

**Node  
C**

## *Partition 2*

**Node  
D**

**Node  
E**

# OCF Conceptual Model (continued)

- ◆ **The method of determining membership is defined by the implementation – not by OCF standards**
- ◆ **The method of electing a primary partition is defined by the implementation – not by OCF standards**
- ◆ **The OCF generally defines the properties an implementation must have, not how they are achieved**

# OCF “API Objects”

- ◆ **Event stream**
- ◆ **Node ID (128 bits)**
- ◆ **Membership id (160 bits)**
- ◆ **Resource type**
- ◆ **Resource instance**

# Resource Agents

- ◆ **Integrates non-cluster-aware services into an HA cluster**
- ◆ **Basic operations are start, stop, status, monitor, metadata**
- ◆ **Analogous to a cluster-wide init process**
- ◆ **This is where most current HA applications are run**
- ◆ **One resource agent per resource type**

# Resource Agents (continued)

- ◆ **Patterned after SystemV (LSB) init scripts**
- ◆ **Extensions to LSB init scripts for:**
  - **Metadata**
  - **Multiple instances  
(via environment parameters)**
  - **Service monitoring**
  - **Configuration validation**
- ◆ **Resource agent can also serve as init script**

# Event Services

- ◆ **Provide a common method of receiving events for OCF APIs**
- ◆ **Notification through file descriptors**
- ◆ **Key to membership, group services, recovery, and any API providing asynchronous event notification**

# Membership Services

- ◆ **Largely event-oriented**
- ◆ **Provides membership change notification**
- ◆ **Informs applications of membership of current cluster partition**
- ◆ **Tells of added, deleted, constant members**
- ◆ **Can be received in incremental updates, or full information each time**
- ◆ **Not intimately tied to quorum**

# Current Status

- ◆ **> 130 members on mailing list representing ~100 organizations**
- ◆ **Active participation by IBM, SuSE, OSDL, Sun, HP, Intel, Steeleye, Oracle, BigStorage, Linux-HA, University of Delft**
- ◆ **Effort also endorsed by Free Standards Group, Conectiva, MSC Software, OSCAR, Red Hat, SGI, Bald Guy Software, UnitedLinux**

# Current Status

- ◆ **Now a working group of the FSG**
- ◆ **Productive working group meeting held Jan, 2003 in NYC, and June 2003 in Ottawa**
- ◆ **Preliminary Draft APIs available for**
  - **Event Services**
  - **Membership**
  - **Resource agents**
- ◆ **Resource Agent API near release**

# Future Plans

- ◆ **Release of Resource Agent APIs: 4Q2003**
- ◆ **Next areas: membership, events, recovery, group services, fencing**
- ◆ **Refine and add APIs => official spec**
- ◆ **Formal review period**
- ◆ **Complete and release reference implementation**

# How To Participate?

- ◆ **Commit staff to help define standards**
- ◆ **Join a subteam**
- ◆ **Commit staff to implement, test ref. model**
- ◆ **Announce your support for the OCF**
- ◆ **Encourage your suppliers, customers to support the OCF**
- ◆ **Join the mailing list**
- ◆ **Evolve your product to conform to the APIs**

# Conclusions

- ◆ **Much confusion currently exists in Linux clustering**
- ◆ **The Open Cluster Framework will provide some structure and order**
- ◆ **This has the potential of making Linux the strongest clustering environment available anywhere**

# References

- ◆ <http://opencf.org/>
- ◆ <http://linux-ha.org/>

Alan Robertson *[alanr@unix.sh](mailto:alanr@unix.sh)*

*<http://linux-ha.org/>*