Open Clustering Framework

Lars Marowsky-Brée <lmb@suse.de>
Engineering Lead HA/Clustering
Research and Development, SuSE Linux AG
Agenda

- Introduction
- HA on Linux
- One solution
- Status
- The future
Introduction to SuSE Linux AG

The company:
- Founded in 1992
- In year 2002:
  - 62% market share in Europe
  - No 2 in the US retail market:
    - Increased Sales by 71%
    - Market share at 27%
- Subsidiaries in Nuremberg, Prague, Oakland
- 350 employees

Partner companies:
- Siemens, IBM, Intel, HP/Compaq, AMD
- SAP World-wide technology partner, Oracle
Introduction to SuSE Linux AG, cont’d

Areas of expertise:

- Consumer Linux
- Business Linux
  - SuSE Linux Enterprise Server
    - IA32, IA64, AMD 64bit, z-Series (31/64bit), p/i-Series (PowerPC)
  - SuSE Firewall on CD
  - SuSE Linux E-Mail Server
  - SuSE Linux Groupware Server
- Core Linux Development
  - SuSE Labs
- United Linux
- Linux consulting
- Training
Abundance of riches

High Availability Clustering on Linux?

A solved problem!

• About ten or more Open Source projects

• Many commercial products available
  ▪ Ports from native to Linux
  ▪ Well tested in the field

• Every niche occupied somehow
Madness

Unfortunately, they do not interoperate:
- Different concepts
- Different APIs
- Different architecture
- Different language

Thus:
- Impossible to assemble a coherent solution
  - Cooperation only by chance
- ISVs can’t rely on a common infrastructure
- Heterogeneous market
- Qualified resources spread thinly across too many projects
But how different is "different"?

Actually:

- Solving mostly the same problem
- Largely the same ideas
- Largely similar architecture

- No reason why APIs should be different for telco or data center
  - Implementation details should be hidden.
One solution

- Adopt a relevant POSIX/IEEE/... standard
- Adopt a de-facto industry standard
- Adopt a successful API from another platform
- Start from scratch.
One solution

Start from scratch:

- Create a model to abstract the existing solutions into a coherent framework.
  - Identify common components and functionality
  - Modular design
- Define / Adopt APIs for the identified components one by one
  - Best practices should be reused
- Abstract enough to encompass all current implementations.

- Challenge: Political as well as technical
Goals

- Must be IP and royalty free
- Accessible both from kernel and user-space
- Implementation independence
- Aspires to be Operating System-agnostic

- At least one Open Source reference implementation
  - Implementation and standardization are different projects!
Architecture overview

Cluster Resource Management
Failover policy

Resource Management
Fencing Monitoring
Instantiation

Cluster-aware applications
Databases OLTP HPC
Clustered LVM / Filesystems

Group Services
Membership Messaging
Transactions Barriers

Node Services
Liveness Membership
Communications Quorum

Monitoring
SNMP
CIM
Logging
Node and Group Services

• Node Services:
  ▪ Node Liveness
  ▪ Node Communication
  ▪ Node Quorum

• Group Services:
  ▪ Membership
  ▪ Messaging
  ▪ Synchronization
    › Barriers
    › Transactions

• MPI
Node and Cluster Resource Services

- Resource Services
  - Instantiation
  - Fencing
  - Monitoring
  - Agents

- Cluster-wide Resource Services
  - Policy
    - Fail-/Switch-Over

- Logging

- Plumbing
External APIs and cluster-aware applications

- Monitoring
  - SNMP
  - CIM

- Application layer:
  - Clustered Volume Management
  - Cluster-aware filesystems
  - Databases
  - OLTP
Technical teams

- Event System
  - Joe DiMartino (OSDL)

- Node Services
  - Alan Robertson (IBM)

- Group Services
  - Ram Pai (IBM)

- Resource Services
  - Lars Marowsky-Brée (SuSE)

- Organizational Council
  - David Ham (University of Delft)

- Steering Committee
  - Formed by technical team leads
  - Head: Alan Robertson
Related efforts

- Free Standards Group
  - Technology Forum

- Open Source Development Labs
  - Data Center Linux
  - Carrier Grade Linux

- Service-Availability Forum

- IEEE Taskforce on Cluster Computing

- United Linux
Status

• Due this year:
  ▪ Node Services API
  ▪ Resource Agent API
  ▪ Event API

Next:
• Take on more APIs (logging, Group Services, DLM)
• heartbeat transformed to a reference implementation
• API consumers: EVMS, OpenGFS etc

• Faster progress ;-)
The blessed Orb of Detection (0:3)

- Get commitment instead of approval
  - Everyone agrees it is needed, but everyone else first please

- Support and input is crucial

- Linux will be the first

- Linux will be the role model for other platforms
Questions and answers

http://www.opencnf.org/

? 

Lars Marowsky-Brée <lmb@suse.de>