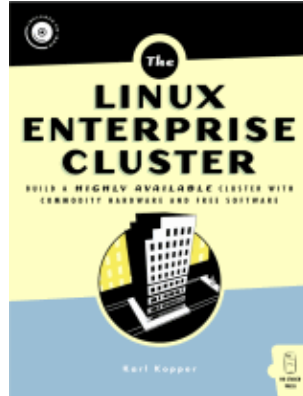


Book review: The Linux Enterprise Cluster



by Tom Uijldert
<tom.ujldert(at)linuxfocus.org>

About the author:

Tom is a member of the Dutch linuxfocus team and he has been wrestling with clusters ever since Digital came up with the idea.

Abstract:

"The Linux Enterprise Cluster" is published by No Starch (<http://nostarch.com/>). ISBN: 1-59327-036-4. Author: Karl Kopper.

Clustering

The art of having a couple of commodity computer systems behave as one.

This, in essence, is what clustering is all about.
So why would you want to do that?

Well, 2 possible reasons here:

1. You don't have enough processing power (nor enough money to get it), so you combine some cheap boxes to get the power just the same.
2. There's some critical programs running on your system and you *really* can't afford to have system downtime (time is money). Hence, you devise a configuration whereby a commodity computer (or part thereof) may fail but the system as a whole stays up (high availability).

The first issue is in the realm of universities and R&D departments. They never have enough funds but they *do* need the processing power. Projects like beowulf or ParallelKnoppix take care of these.

The second issue is more in the realm of companies (or *enterprises*). Funding is usually not a problem but system downtime *is*.

Imagine having your development department of 60-odd people waiting on the system to become available

again. Or an office of dozens of administrative staff, waiting for the database to come up again. System downtime can thus become very expensive and that's where this book comes in.

The book focuses on building *enterprise class* clusters –which is synonym for highly available– that will keep on going.

In many ways, this is a very unforgiving book. Don't expect any airy–fairy talk on the theories behind clustering or high availability. This book gets right down to business: build the damned thing and do your maintenance on it.

So if you're looking for some abstract, high–level introduction into highly available systems, stop right here and go look somewhere else. You will not find it here.

If, however, you have the boss breathing down your neck right now screaming: “The system is down again and it is costing me a fortune! What the hell am I paying you for!?!” then this is the book you'll need.

So read on, for some more details.

History

Legend has it that the computer company HP (formerly known as Compaq, formerly known as Digital Equipment Corporation, “what's in a name?”) could not, in those days of mini– and mainframe computers, come up with a processor rivaling the power of the IBM mainframe processors. Hence they came up with the idea of clustering their minicomputers so that they could offer customers something that rivaled the mainframe–bids.

I doubt whether they ever persuaded an IBM customer with this scheme but they *did* find something else. That if a minicomputer crashed or went down, the others would still function. The worst that could happen was that a user would have to login again because he was attached to the crashing computer. A highly available system was born.

Now I'm not sure how much of this is true and/or urban legend, but it's a nice story so I'll stick with it until a better story comes along.

Clustering is a specialised and complex matter. This is proven by the fact that no commercial Unix–vendor has –until now– come up with a clustering–solution that could rival the (Open)VMS solution (yes, not even the Unix systems of Digital itself were up to par with the VMS clusters). And now the open source community has a stab at it.

The book

Has a bit of an odd build–up. You would expect it to start with a general description of what the goal is, some background and theory and then gradually moving down to the bits–and–bytes stuff.

Not this book. This one states: “We're going to build us a cluster, and this is the recipe:...”. And so part one starts with some Linux basics like compiling kernels, installing packages and basic network configuration, that you'll need to master before a next building step is taken.

That next step essentially contains more basics, but now focused on packages and configurations that deal with high availability. Included here are subjects like system cloning, the heartbeat package and stonith-devices.

Part 3 then combines all these basics to implement highly available clusters using load-balancing. And here, finally, everything comes together, with some added cluster-theory as well (we're by now almost 200 pages into the book already).

The final part deals with how to keep a cluster running. How to administer maintenance and monitor it's performance.

So, like I said, a bit of an odd build-up but then again, who said you need to read a book sequentially, front to back?

The verdict

If anything, this is a *practical* book.

I can see a battered copy of it, always lying around in the server room. Battered, because it is so frequently referenced by the sysadmins in maintaining their clusters. Like a cookbook indeed.

This practicality can for instance be seen in the substantial part (4) that is devoted to cluster maintenance and monitoring. Where many a book would stop once a cluster-configuration has been built, this book does not forget that all-important phase that follows after implementation.

Another indication is the tremendous amount of notes, footnotes, tips and tricks that this book is littered with. What about a gem like (page 322): *“You can write a script that calls another script, but just be sure to pass the exit status of the child script back as the exit status of the main script or SNMPD will not see it”*;

You can only come up with these kind of notes once you have experienced them yourself and bumped your head on them before.

And this book *oozes* that kind of blood, sweat and tears. I can well imagine the author, for months hacking away at installing and configuring new systems, finding out what problems there are, trying to solve them or come up with other alternatives. All the while meticulously making notes of every glitch or problem he encounters.

And finally, definitive proof was given just a few days ago, after reading the book, when I could help a colleague with a problem by pointing to some excerpts from the book.

In conclusion

The practicality of this book is also one of its weaknesses. It will not age well. A lot will need to be rewritten in 2 –3 years due to ongoing developments in the open source community, configuration changes etc. Then again, you don't buy a book only to use it years from now.

In addition, the book contains a handful of appendices and a complementary CD-ROM with more on downloading, troubleshooting, configuring, packages and scripts, enough for you to experiment with clustering to your hearts content.

What I would *really* like though (hint hint) is for the author, after this landmark work, to take the book and knowledge gained a step further and come up with a complete distribution. Something like *EnterpriseKnoppix*, containing all the goodies needed to make a highly available cluster.

<p><u>Webpages maintained by the LinuxFocus Editor team</u> © Tom Uijldert "some rights reserved" see linuxfocus.org/license/ http://www.LinuxFocus.org</p>	<p>Translation information: en --> -- : Tom Uijldert <tom.ujldert(at)linuxfocus.org></p>
--	---

2005-07-25, generated by lfparsr_pdf version 2.51