

ParallelKnoppix

Linux Gazette n°110 — Janvier 2005

Majid Hameed

Copyright © 2005 Majid Hameed

Copyright © 2005 Gauthier Savart

Copyright © 2005 Joëlle Cornavin

Article paru dans le n°110 de la Gazette Linux de janvier 2005.

Traduction française par Gauthier Savart <gauthier.POINT.savart@chez.laposte.POINT.net>.

Relecture de la traduction française par Joëlle Cornavin <jcornavi@chez.club.TIRET.internet.POINT.fr>.

Article publié sous Open Publication License (<http://linuxgazette.net/copying.html>). La Linux Gazette n'est ni produite, ni sponsorisée, ni avalisée par notre hébergeur principal, SSC, Inc.

Table des matières

1. Résumé.....	??
2. Arrière-plan.....	??
3. Description.....	??
4. Qu'est-ce qu'un démarrage PXE ?	??
5. Que dois-je faire si ma carte d'interface réseau n'est pas compatible avec PXE ?	??
6. Téléchargement de ParallelKnoppix.....	??
7. Comment cela fonctionne-t-il ?.....	??
8. Que dois-je faire si de multiples serveurs DHCP fonctionnent ?	??
9. Récapitulatif	??
10. Mon expérience personnelle.....	??
11. Conclusion	??
12. Références.....	??

1. Résumé

ParallelKnoppix est un Live CD fondé sur Knoppix, lequel est également un Live CD inspiré lui-même de la distribution Linux Debian. ParallelKnoppix vous permet de créer une grappe (*cluster*) Linux dotée d'outils/de bibliothèques de programmation parallèle comme MPI (*Message Passing Interface*) simplement en quelques minutes. Il économise beaucoup de temps, que nous emploierons à la configuration de l'environnement de calcul. L'environnement existant n'est pas perturbé par l'utilisation

de ParallelKnoppix, puisqu'une distribution Live CD s'exécute sans installation (un répertoire est créé sur le nœud maître et peut être supprimé après redémarrage si nécessaire).

[Extrait de <http://pareto.uab.es/mcreel/ParallelKnoppix/> :]

ParallelKnoppix est un rematriçage de Knoppix qui permet de mettre en place une grappe de machines pour du traitement parallèle utilisant les implémentations LAM-MPI (*Local Area Multicomputer*) et/ou MPICH (*Message Passing Interface/Chameleon*) de MPI. Monter la grappe et la rendre opérationnelle prend moins de 15 minutes, si les machines disposent de cartes réseau PXE (*Preboot eXecution Environment*).

2. Arrière-plan

La mise en grappe est une des techniques les moins chères pour faire du traitement en parallèle. C'est d'ailleurs une des forces de Linux. Un certain nombre d'universités et d'autres organisations simulent les super-calculateurs en reliant des PC via des cartes Ethernet sous Linux. Ce système d'exploitation a été largement adopté par la communauté scientifique pour effectuer leurs travaux de recherche, du fait qu'il offre bon nombre d'outils scientifiques tels que LAM, MPI, PVM (*Parallel Virtual Machine*), et bien plus encore. Linux est convient donc parfaitement pour le calcul parallèle. Cependant, le problème est que les scientifiques et les programmeurs doivent réaliser beaucoup de travail de préconfiguration de l'environnement Linux. Ceci ralentit et complique leur tâche.

Maintenant, les gourous Linux ont résolu ce problème en développant des Live CD. Le chercheur peut dorénavant choisir un Live CD pour faire de la programmation parallèle sans la nécessité d'une longue configuration : la grappe est opérationnelle en quelques minutes.

Parmi les Live CD dédiés à la programmation parallèle, se trouve ParallelKnoppix (<http://pareto.uab.es/mcreel/ParallelKnoppix/>). D'autres Live CD, comme BCCD (<http://bccd.cs.uni.edu/>) et ClusterKnoppix (<http://bofh.be/clusterknoppix/>), sont également disponibles.

3. Description

Tout comme son prédécesseur Knoppix, ParallelKnoppix détectera tout le matériel et les périphériques automatiquement. Je l'ai testé avec une carte mère Intel® P-IV D865GBF et une autre carte mère Intel® 810C (P-III) et, dans les deux cas, ParallelKnoppix a configuré automatiquement l'ensemble du matériel sans qu'il soit nécessaire d'intervenir. Les ordinateurs qui sont configurés à l'aide de ParallelKnoppix partagent un répertoire commun créé sur le nœud principal via NFS (*Network File System*). Le nœud maître est amorcé par le cédérom et les esclaves le sont à l'aide du réseau (le nœud maître lance un serveur DHCP). Les esclaves ont un BIOS avec activation de la fonction PXE avec des cartes d'interface réseau (NIC, *Network Interface Card*) compatibles avec PXE.

Tout service nécessaire pour LAM/MPI est configuré automatiquement. DHCP, NFS, SSH (avec *logins* sans mot de passe) sont tous configurés et opérationnels : vous êtes prêt à expérimenter les programmes MPI et autres applications parallèles.

Une fois configuré, ParallelKnoppix n'est pas très sécurisé : le mot de passe du Live CD pour à la fois l'utilisateur normal et le super-utilisateur (*root*) sont publiquement connus. Quiconque ayant un minimum de connaissances sur ParallelKnoppix peut facilement accéder à la grappe ParallelKnoppix. Dans ce cas, la facilité de configuration est obtenue moyennant un compromis au détriment de la sécurité.

D'une manière générale, si la sécurité est pour vous un impératif, votre réseau ParallelKnoppix devra être isolé de l'Internet, voire de votre intranet.

— Ben

4. Qu'est-ce qu'un démarrage PXE ?

PXE est l'acronyme de *Preboot eXecution Environment* (environnement d'exécution de prédémarrage), une technologie utilisée pour amorcer un PC à distance via un réseau. PXE doit être pris en charge par le BIOS du système et la carte d'interface réseau doit être compatible avec PXE.

5. Que dois-je faire si ma carte d'interface réseau n'est pas compatible avec PXE ?

Vous devrez soit installer une puce ROM (*Read-Only Memory*, mémoire morte) avec une image Etherboot (<http://www.etherboot.org/>) sur votre carte d'interface réseau, soit graver un cédérom avec l'image. ROM-o-matic.net (<http://rom-o-matic.net/>) génère dynamiquement des images de ROM Etherboot.

6. Téléchargement de ParallelKnoppix

Le fichier ISO est disponible sur les sites suivants :

- Via FTP : <ftp://volcano.uab.es/pub/parallelknoppix.iso>
- Via HTTP : <http://pareto.uab.es/mcreel/ParallelKnoppix/parallelknoppix.iso>
- MD5SUM pour l'image :
<http://pareto.uab.es/mcreel/ParallelKnoppix/parallelknoppix-2004-12-16.iso.md5>
- Consultez la page d'accueil (<http://pareto.uab.es/mcreel/ParallelKnoppix/>) si les liens ci-dessus ont expiré.

Après avoir téléchargé les images ISO, vérifiez les sommes de contrôle MD5 avec les images ISO pour vous assurer que votre téléchargement s'est déroulé sans erreur. Il suffit de lancer le programme **md5sum** à l'invite d'un terminal et de comparer les valeurs retournées :

Dans la commande ci-dessus, remplacez *nom-de-fichier.iso* par le nom exact du fichier.

Si pour une raison quelconque, vous n'utilisez pas Linux, il existe également md5summer (<http://www.md5summer.com/>) pour Windows®. Un calculateur MD5 pour DOS est également disponible.

Si la somme MD5 est correcte, gravez les images ISO sur des CD-R ou des CD-RW.

Graver des images ISO sur un cédérom nécessite un programme tel que cdrecord.

7. Comment cela fonctionne-t-il ?

Vous trouverez un bon tutoriel sur Parallel Knoppix rempli d'instructions pas à pas, de copies d'écran du processus d'installation, etc. ici (<http://pareto.uab.es/wp/2004/62604.pdf>) au format PDF. Si vous avez exporté votre cédérom vers les nœuds, Parallelknoppix prendra en charge 50 nœuds sans problème, nombre limite au-delà duquel aucun test n'a été réalisé. En ce qui me concerne, je n'ai testé moi-même que 5 nœuds.

8. Que dois-je faire si de multiples serveurs DHCP fonctionnent ?

Si vous utilisez cette configuration dans une université (ce qui est mon cas), vous allez probablement entrer en collision avec le serveur DHCP « officiel » et, éventuellement, un autre serveur PXE. Quand vous essayez de démarrer les nœuds à l'aide d'un terminal serveur, les nœuds démarrent souvent à partir du serveur PXE préexistant et obtiennent leurs adresses IP depuis le serveur officiel, non pas depuis le serveur DHCP opérationnel sur l'ordinateur qui a été démarré avec le cédérom de ParallelKnoppix. La solution que j'ai pour le moment consiste à déconnecter physiquement les ordinateurs à utiliser en tant que nœuds à partir des serveurs PXE et/ou DHCP préexistants. Sinon, il faut obtenir de l'aide des administrateurs pour désactiver temporairement ces serveurs. Si quelqu'un a une solution plus élégante à proposer, j'aimerais en être informé. J'imagine qu'il faut faire intervenir `miniroot.gz` et utiliser Rom-o-Matic pour créer la ROM de démarrage PXE.

9. Récapitulatif

Le Live CD ParallelKnoppix est utilisé pour démarrer un nœud maître. Dès que le nœud maître est opérationnel, un *script* est exécuté, qui configure un serveur DHCP, partage un répertoire de travail commun avec tous les nœuds utilisant NFS, génère les clés publiques pour que SSH fonctionne correctement (*logins* sans mot de passe), etc. Une fois que le serveur DHCP est opérationnel sur le nœud maître, les clients (esclaves) sont lancés via le démarrage PXE. Après avoir réussi ce lancement, le répertoire test est copié dans le répertoire NFS commun partagé, et les programmes commencent à s'exécuter en parallèle sur des PC multiples.

10. Mon expérience personnelle

Étant étudiant de premier cycle en informatique, on m'a donné un projet consistant à résoudre un problème mathématique à l'aide de MPI dans le laboratoire de calcul parallèle. J'ai choisi ParallelKnoppix comme alternative pour faire une démonstration de mon programme MPI dans l'environnement Linux. Quand le nœud principal est démarré avec le cédérom ParallelKnoppix, à un stade du démarrage, vous devez choisir la résolution : saisissez simplement '6', parce que c'est le mode de résolution maximal pris en charge. Une fois mon nœud maître démarré, j'ai lancé le *script* (**K -> ParallelKnopix -> Setup ParallelKnoppix**, comme l'indique le tutoriel ci-dessus). Quand le *script* a créé le serveur DHCP, j'ai démarré mes nœuds esclaves et les ai fait démarrer à l'aide de PXE. Tous les nœuds ont démarré sans problème.

J'ai ensuite copié mon programme dans `parallel_knoppix_working` et exécuté mon programme MPI en parallèle. C'est aussi simple que cela.

Pour la compilation, j'utilise :

Pour l'exécution :

11. Conclusion

[Extrait du tutoriel ParallelKnoppix :]

« Le cédérom ParallelKnoppix offre un moyen très simple et rapide de configurer une grappe de PC hétérogènes à architecture IA-32. Il n'est pas destiné à fournir une grappe stable pour des utilisateurs multiples ; c'est plutôt un outil pour la création rapide d'une grappe en vue d'une utilisation individuelle. Comme il est possible de personnaliser le cédérom, on peut réactualiser les fichiers de configuration et de travail, il peut constituer une solution à long pour un utilisateur individuel ».

— Michael Creel

12. Références

- Page d'accueil de ParallelKnoppix : <http://pareto.uab.es/mcreel/ParallelKnoppix/>
- *Discussion Paper on Parallel Knoppix By Michael Creel (14 octobre 2004)* : <http://pareto.uab.es/wp/2004/62504.pdf>
- *High Performance Linux Clusters with OSCAR, Rocks, OpenMosix, and MPI By Joseph D. Sloan, Publication: O'Reilly Associates - Date de publication: Novembre 2004* :
 - <http://safari.oreilly.com/0596005709>
 - <http://safari.oreilly.com/?XmlId=0596005709>
- Page d'accueil de Knoppix :
 - <http://www.knopper.net/knoppix/index-en.html>
 - <http://www.knoppix.org/>
 - <http://www.knoppix.com> (<http://www.knoppix.com/>)
 - <http://www.Knoppix.net> (<http://www.knoppix.net/>)

- *ROM-o-matic.net* : <http://rom-o-matic.net/>
- *LAM/MPI Parallel Computing* : <http://www.lam-mpi.org/>
- *LAM/MPI User's Guide* : <http://www.lam-mpi.org/download/files/7.1.1-user.pdf>

Majid Hameed est étudiant de premier cycle du *Department of Computer Science* de l'université de Karachi, Sindh, au Pakistan. Ses centres d'intérêts principaux sont l'intelligence artificielle, les systèmes d'exploitation, le réseau, la programmation et l'infographie.

Passionné de Linux, j'utilise Linux en tant que système d'exploitation depuis plus de 3 ans. J'ai essayé et utilisé les distributions suivantes : Red Hat Linux 9, 8,7.3, 7.2, Slackware Linux 10,, 9.1, Slax, Mandrake Move 2, Knoppix 3.4, Vector Linux 4.3, et quelques autres.