

# Cray marie supercalculateur et Big Data analytique avec Intel Hadoop

Cray, le constructeur historique de supercalculateurs, a annoncé la disponibilité de la nouvelle distribution Hadoop d'Intel sur sa technologie CS300 de supercomputing. La combinaison des deux donne naissance à une puissante solution de Big Data analytique.

## Cray CS300

Le Cray CS300 est un cluster sous Linux, piloté par l'application Cray ACE (*Advanced Cluster Engine*), secondée par une version adaptée de SLURM (*Simple Linux Utility for Resource Management*), et qui supporte désormais Infiniband.

Répondant aux attentes du marché en matière de standardisation, la machine embarque une plateforme 2 sockets, lesquels accueillent des processeurs Intel Xeon. Une partie de son innovation provient de son refroidissement liquide et de son architecture qui optimise l'efficacité énergétique.

## Intel Distribution Hadoop

Le constructeur se fait ici remarquer avec l'adoption de la plateforme Big Data Intel Distribution for Apache Hadoop. Intel a conçu cette distribution afin d'optimiser le code de Hadoop pour sa plateforme Xeon (voir la vidéo de [Stanislas Odinet, consultant technique pour Intel : « Big data : nous avons notre propre distribution Hadoop »](#)). Une démarche qui vient confirmer l'intérêt du fondateur pour le marché du Big Data et sa volonté d'aller au-delà des serveurs.

Intel travaille avec la communauté Hadoop depuis 2009. Mais le fondateur a franchi un nouveau pas en février avec l'annonce de sa propre distribution Hadoop. Parmi les apports d'Intel figure l'optimisation de l'usage de ses processeurs par le support des technologies qui y sont intégrées, à l'image de AES-NI (*Advanced Encryption Standard New Instructions*) qui permet d'accélérer le chiffrement sur le système de fichier HDFS (*Hadoop Distributed File System*).

## Une architecture de performance

La combinaison du HPC et de l'analytique devrait permettre de franchir un nouveau pas dans la puissance du Big Data. Elle pourrait surtout être une étape incontournable pour atteindre le temps réel. Et elle confirme que le Big Data analytique a trouvé dans les clusters sa plateforme de prédilection pour l'exécution d'algorithmes complexes.

Avec Intel Distribution for Apache Hadoop, Cray renforce la sécurité de sa plateforme, améliore le temps réel dans le traitement de la donnée, ainsi que les performances de son architecture de stockage. La solution devrait également intégrer les applications Hadoop qu'Intel n'a pas souhaité rendre open source, comme Intel Manager for Apache Hadoop pour la configuration et le déploiement, ou Active Tuner for Apache Hadoop qui améliore la performance du compute sur le

cluster où s'exécute sa distribution.

Les architectures de référence sur lesquelles tourne Hadoop montrent des faiblesses tant du côté du stockage de données que du traitement de ces données. Le HPC, dans sa configuration désormais classique en clusters, pourrait donc se révéler comme étant une architecture de référence pour l'analytique. D'autant plus que le HPC offre des capacités d'évolution adaptées à cette approche.

---

### **Voir aussi**

[Silicon.fr étend son site dédié à l'emploi IT](#)

[Silicon.fr en direct sur les smartphones et tablettes](#)