

Avantages et Inconvénients des clusters Hadoop

Avantages

Le principal atout des clusters Hadoop est leur adéquation à l'analyse de gros volumes de données. Le Big Data est le plus souvent non structuré et largement distribué.

Si Hadoop convient bien à l'analyse de ce type de données, c'est parce qu'il possède un mode de fonctionnement qui vise à fractionner les données en éléments et attribue chacun d'entre eux à un nœud donné du cluster. Les données n'ont pas à être uniformes, car chaque élément de donnée est traité par un processus distinct d'un nœud distinct du cluster.

Second avantage des clusters Hadoop réside dans leur évolutivité. Un des problèmes de l'analyse du BigData est sa croissance continue. De plus, le BigData est surtout utile s'il est analysé en temps réel ou dans un délai aussi proche que possible du temps réel. Les capacités de traitement en parallèle de Hadoop participent à la rapidité de l'analyse mais à mesure que le volume de données à analyser augmente, la puissance de traitement du cluster peut devenir insuffisante. Heureusement l'ajout de nœud peut faire évoluer le framework.

Le troisième avantage de ce type d'architecture réside dans son coût. En effet, l'analyse du BigData fait partie de l'informatique d'entreprise qui est traditionnellement coûteuse, pourtant, les clusters Hadoop peuvent s'avérer être une solution économique. Ils sont généralement peu onéreux pour 2 raisons, d'abord le logiciel est Open Source. La distribution Apache Hadoop est même proposée en téléchargement gratuit. De plus, l'utilisation de matériel standard contribue à limiter mes coûts d'un projet Hadoop. On peut donc construire un cluster puissant sans dépenser beaucoup en serveurs.

Un autre avantage de ces clusters est leur résistance aux pannes. Quand un élément de donnée est envoyé à un nœud pour analyse, les données sont également copiées vers d'autres nœuds : en cas de défaillance sur un nœud, d'autres copies existent ailleurs dans le cluster et il est toujours possible de les analyser.

Inconvénients

Cependant les clusters Hadoop ne conviennent pas aux besoins d'analyse de données de toutes les organisations. Notamment, pour des petits volumes de données, les bénéfices seront presque inexistantes.

L'un des inconvénients des clusters Hadoop réside dans le principe même de la solution de clustering, qui suppose que les données puissent être fractionnées et analysées par des processus parallèles exécutés sur des nœuds cluster distincts. Si l'analyse ne peut se dérouler dans un environnement de traitement parallèle, un cluster Hadoop n'est tout simplement pas le bon outil.

L'un des plus grands désavantages reste probablement les difficultés d'apprentissage associées à la construction, au fonctionnement et à la prise en charge de la solution.