

Big Data Hadoop : MapR M7 est sur AWS

Les applications Apache Hadoop et NoSQL sont les bienvenues sur la plateforme analytique EMR (*Elastic MapReduce*) du cloud Amazon, depuis que AWS (*Amazon Web Services*) a annoncé la disponibilité de MapR M7, la dernière version de l'une des trois distributions historiques de Hadoop (avec Hortonworks et Cloudera).

Des clusters Hadoop sur AWS

Concrètement, pour stocker, traiter et analyser de gros volumes de données, les utilisateurs d'Amazon Elastic MapReduce vont pouvoir déployer et dimensionner des clusters M7 sur des instances de stockage hs1.8xlarge.

Celles-ci sont destinées aux applications nécessitant un accès séquentiel à des jeux de données particulièrement volumineux. Elles offrent 48 To de capacité de stockage sur 24 disques durs, 16 cœurs virtuels qui peuvent fournir 35 unités de calcul EC2 (ECU) en termes de performances CPU, 117 Go de RAM et 10 Gbits/s en mise en réseau.

La plateforme Hadoop de MapR rend plus accessibles les technologies Hadoop et le déploiement de HBase, avec toutes les fonctionnalités attendues. Amazon décrit HBase comme un magasin de données distribué et basé sur des colonnes, qui fournit des écritures et lectures cohérentes, le partage automatique des tables et un stockage efficace de volumes importants de données dispersées.

Il est conçu pour fonctionner en transparence avec Hadoop, via le partage de son système de fichiers et en assurant les entrées et sorties pour l'exécution des opérations MapReduce dans Hadoop.

Amazon a choisi sa distribution Hadoop

L'architecture MapR M7 sur EMR offre aux utilisateurs de HBase jusqu'à 100 Kops/s par nœud sur les instances HS1, l'absence de compactations, la division par région, la restauration instantanée après échec, la faible latence consolidée, le full HA, la cohérence des snapshots et la récupération point-in-time, et la reprise après sinistre (*mirroring*).

Cette annonce vient confirmer le choix d'Amazon de déployer sa stratégie Big Data Hadoop en partenariat avec MapR. Ce choix repose en particulier sur la capacité de la distribution Hadoop d'offrir aux applications Linux un accès direct aux données dans le cluster via l'interface NFS développée par MapR.

Voir aussi

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

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