

Big data : VMware intègre Hadoop à sa plateforme Spring

SpringHadoop est un nouveau module de l'univers Java de VMware, qui réalise l'intégration de la plateforme **Apache Hadoop** et du *framework* de développement Spring de VMware.

Il vise à aider les développeurs à « *simplifier la conception de solutions de traitement réparti avec Apache Hadoop* ». Ils doivent ainsi plus facilement « *construire, déployer, gérer et ajuster la nouvelle vague d'applications axées sur les données.* » Le choix en matière de volumes et d'accès aux données s'est très considérablement élargi, constate VMware. Les développeurs Java peuvent ainsi intégrer de nouveaux modèles d'accès aux données dans leurs applications.

SpringHadoop permet de combiner facilement Apache Hadoop, le Distributed File System de Hadoop et divers *add-ons* comme le langage d'analyse de données Pig ou encore le *datawarehouse* Hive et SQL.

Pour rappel, la suite Spring Data permet au *framework* de développement d'accéder à diverses bases de données, des *data grids*, des *document stores* et donc, désormais, des outils MapReduce comme Hadoop. (cf. notre article '[Les solutions d'intégration de Talend partent à l'assaut du big data](#)').

Les principales caractéristiques de SpringHadoop :

- prise en charge de la configuration, de la création et du traitement via MapReduce, Streaming, Hive, Pig, et Cascading grâce au conteneur Spring ;
- prise en charge complète de l'accès aux données HDFS grâce aux langages de script Java (Groovy, JRuby, Jython, Rhino, etc.) ;
- prise en charge de la configuration déclarative de HBase ;
- prise en charge dédiée de Spring Batch pour le développement de puissantes chaînes de traitement intégrant les opérations HDFS et tous types de tâches Hadoop ;
- prise en charge de Spring Intégration permettant l'accès simplifié à une large gamme de systèmes existants grâce à une architecture extensible de connecteurs orientés événements et de filtres ;
- de puissantes options de configuration et des mécanismes de création de modèles pour les connexions client à Hadoop ;
- prise en charge déclarative et programmatique de Hadoop Tools (FsShell et DistCP inclus).

SpringHadoop est disponible dès aujourd'hui en téléchargement sous la licence *open source* Apache 2.0.