

Extension du framework SparkSQL pour une prise en charge efficiente des données spatiales massives dans les télécommunications.

El Hassane Nassif *, Hicham Hajji *,
Reda Yaagoubi *, Hassan Badir **

* Institut Agronomique et Vétérinaire Hassan II , Rabat
nassif.hassane@gmail.com, h.hajji@iav.ac.ma, r.yaagoubi@iav.ac.ma
<http://www.iav.ac.ma>

** Ecole de Sciences Appliqués de Tanger
hbadir@uae.ac.ma
<http://www.ensat.ac.ma>

Résumé. Nous assistons ces dernières années à une évolution explosive des données spatiales issues des activités : appels, sms et navigation internet des abonnés Télécoms. Ce type de données arrivent sous divers formats et à une fréquence de plus en plus élevée, à tel point que les exploiter relève du défi. En pratique, les opérateurs font face à un manque de solutions spécialisées dans l'analytique géospatiale sur des données massives. Ce qui rend difficile l'exploitation de la dimension spatiale dans les télécoms. Dans cet article, nous présentons une architecture Big Data intégrée pour la gestion des données spatiales massives dans les télécoms. Et nous expliquons les extensions effectuées sur le framework SparkSql afin de supporter les types et les requêtes spatiales.

1 Introduction

La concurrence et les nouvelles tendances du marché poussent les opérateurs de télécommunication à chercher des moyens innovants pour réaliser plus de marge. Un de ces moyens consiste à mettre en place une approche intégrée pour la géolocalisation de leurs clients. Ceci est possible à travers l'exploitation des traces géolocalisées des abonnés dans la réalisation des cas d'utilisations analytiques géospatiales. Pour ce faire et au vu des caractéristiques Big Data liées à l'évolution explosive du volume, de la vitesse et de la variété des données générées par les différentes activités des clients, il est nécessaire de s'orienter vers les technologies Big Data (Phaneendra et Reddy (2013)). Dans ce sens, nous avons proposé dans un précédent travail (E.Nassif et al. (2018)), une architecture Big Data intégrée pour la gestion des données spatiales massives dans le domaine des télécommunications. Vu que les outils de traitement Big Data qui existent sur le marché ne supportent pas le type spatial. Nous proposons dans ce papier d'étendre la nouvelle API SparkSQL (Armbrust et al. (2015)) du moteur de calcul parallèle Apache Spark ¹ pour une prise en charge des données spatiales massives provenant des

1. <https://spark.apache.org>