

Hayat BELLAFKIH

## **Analyse de données à grande échelle : application au dépôt RIPE Atlas**

RIPE Atlas est un projet créé et géré par l'organisme RIPE NCC. Ce projet consiste à déployer de simples dispositifs, appelés sondes, que l'on connecte à un routeur. Ces sondes consomment une quantité faible d'électricité et de bande passante. Elles ont la capacité d'effectuer des mesures liées aux réseaux informatiques. La répartition abondante des sondes Atlas engendre quotidiennement une quantité importante de données qui dépasse la capacité des outils traditionnels de stockage et de traitements. On parle de données massives.

Certaines problématiques dans le domaine des réseaux informatiques nécessitent une exploration en profondeur des données réseaux afin d'aboutir à des résultats significatifs. Dans ce mémoire, des technologies adaptées aux données massives ont été évaluées pour étudier un des problèmes liés aux performances des réseaux informatiques. Cette évaluation adopte un outil existant qui utilise un nombre très important de traceroutes.

La manipulation des données massives concerne la collecte de ces données, leur stockage et leurs traitements. De nombreuses technologies dédiées aux données à grande échelle sont disponibles pour répondre à ces besoins. Le choix d'une technologie s'appuie sur différents facteurs. Comme les frais appliqués suite à l'utilisation d'une technologie, sa performance en terme de temps d'analyse, l'évolutivité de la solution adoptée, etc. Ces éléments et autres ont été évalués sur un sous-ensemble de technologies. Et ce, en vue d'analyser des traceroutes du projet RIPE Atlas dans le cadre de l'outil adopté. Ces technologies sont MongoDB, Amazon S3, Amazon Athena et Apache Spark.

**Mots clés :** RIPE Atlas, Traceroute, Big Data, données massives, MongoDB, Amazon Web Service, Apache Spark.