

Nouveautés de la technologie SIG

Bulletin d'information de l'OSIG 01/2016

Vitesse (Fast Data) et intelligence (Smart Data)

Le dernier article émanant de notre groupe de travail traitait des mégadonnées (Big Data) dans le contexte des SIG (bulletin d'information 3/2015). Il indiquait, dans sa conclusion, que les Big Data appartenaient d'ores et déjà au passé et qu'il était de plus en plus question de données rapides (Fast Data) et de données intelligentes (Smart Data).

Les Fast Data et les Smart Data sont utilisées de manière générale comme des sous-catégories des Big Data, appartenant aux Big Data dites «V» comme vitesse (Fast Data) ou valeur (Smart Data).

S'il s'agit, pour les Big Data dans leur ensemble, de technologies dédiées au recueil et à l'analyse de très gros volumes de données en croissance rapide, les Smart Data visent à maîtriser l'art de filtrer le «flot» des Big Data à l'aide d'algorithmes intelligents pour ne retenir que les informations essentielles, celles qui ont vraiment un sens (Information that actually makes sense), autrement dit, pour déduire des données recueillies les informations qui permettent d'agir ensuite.

Les Fast Data désignent la capacité à tirer des enseignements des données au moment même où elles sont créées. Il est donc question ici d'informations en temps réel. L'adage selon lequel la vitesse prime sur le volume (Data is Fast before it's Big) s'applique bien souvent. Dans quel contexte les Fast Data sont-elles générées? A titre d'exemples, on citera les applications mobiles, les réseaux de capteurs, les médias sociaux ou l'Internet des objets.

La combinaison des Fast Data et des Smart Data constitue l'un des défis majeurs des Big Data, consistant à filtrer les données, si possible en temps réel, pour ne retenir que celles qui sont pertinentes et lancer sans délai les actions qu'elles impliquent. Ces données doivent-elles être enregistrées ou non? Une réflexion est à mener à ce sujet, sachant que bon nombre de données perdent rapidement leur valeur et ne sont plus utilisables (pour des prévisions ou la prise de décisions) au bout d'un certain temps – c'est par exemple le cas des données météorologiques ou de celles relatives au trafic.

Si les données doivent effectivement être enregistrées, il est fréquent que les banques de données (relationnelles) usuelles se montrent insuffisantes. Elles sont par exemple suppléées par des systèmes dits NewSQL. Il s'agit de systèmes de gestion de banque de données qui cherchent à compléter les forces des SGBDR classiques (modèle de données relationnel, support SQL, sécurité des transactions ACID) par les propriétés des systèmes NoSQL (scalabilité au service du traitement en ligne des transactions – OLTP). De tels systèmes sont par exemple obtenus à l'aide d'architectures en mémoire (In-Memory, gestion des données déplacée du disque dur vers la mémoire centrale) ou de services sur le Cloud.

C'est dans ce contexte que se posent alors les questions suivantes:

- «Existe-t-il des Fast Data dans les SIG?»
- «Les SIG font-ils partie des Smart Data?»

Nous estimons que les géodonnées sont clairement des Smart Data lorsqu'elles sont traitées dans le cadre d'analyses intelligentes et que des décisions sont prises ou que de nouveaux processus sont lancés sur cette base. On peut donc aussi parler de Fast Data dans les SIG, lorsque des flux de

données transmis par des réseaux de capteurs géoréférencés sont par exemple traités dans des applications d'informatique décisionnelle (business intelligence).

Nous voulons stimuler le débat sur ces questions et nous réjouissons de votre point de vue. Discutez avec nous!

Luzius Ammann

Groupe de travail Technologie SIG

technologie@sogi.ch