



Effet collatéral : stocker les données générées par l'Internet des objets ?

Publié le jeudi 13 avril 2017

Voir en ligne : <https://www.france-science.org/Effet-collateral-stocker-les.html>

L'Internet des objets (IoT) connaît ces derniers temps un regain de popularité après avoir été éclipsé par les récentes montées en puissance des domaines de l'intelligence artificielle. En tant que générateur de données visant à être exploitées par la sphère de l'IA, l'Internet des objets poursuit son développement en introduisant de nouvelles problématiques liées à ces grands volumes de données produites par des périphériques toujours plus nombreux. De la même manière que d'autres types de données, la sphère IoT a un besoin grandissant de technologies de stockage à même de gérer des débits de données très élevés tout en restant compatibles et intégrables dans les écosystèmes visant à exploiter ces données.

L'Internet des objets

L'Internet des objets se présente comme une forme de prolongement de l'Internet dans le monde réel et tangible : les événements physiques sont ainsi mesurés et transcrits en signaux compréhensibles pour l'écosystème de l'Internet (des ordinateurs, des serveurs etc.). Ce réseau particulier, considéré comme le Web 3.0 et comprenant également le Web sémantique, se présente comme l'itération suivante du Web 2.0 qui est considéré comme l'ère du Web social.

Constitué de périphériques présents dans le monde réel, l'Internet des objets a des applications très diverses couvrant les domaines de la quantification de l'être humain à la e-santé en passant par des domaines comme la domotique ou encore la conduite de véhicules autonomes. Cette profusion de dispositifs connectés engendre des volumes massifs de nouvelles données avec un impact important sur l'augmentation du trafic Internet.

Cette quantité de données produites a des effets non seulement sur le trafic généré et sur l'infrastructure même de l'Internet, mais également sur la nécessité de créer des technologies permettant de stocker ces données pour pouvoir les exploiter. Dans la baie de San Francisco, des entreprises comme Crate proposent des solutions technologiques adaptées au stockage de ces données de l'Internet des objets.

Crate, combiner la puissance de l'algèbre relationnelle à la rapidité des SGBD non structurés

Créée en 2013 et installée dans la baie de San Francisco, Crate.io a été distinguée lors du Tech Crunch Disrupt de Berlin en 2014 et vient tout juste de recevoir le *Big Data Tech Trailblazers Award* pour sa technologie innovante de système de gestion de bases de données pour le Big Data. La startup a développé

une technologie hybride offrant la rapidité et les possibilités de passage à l'échelle de l'écosystème de données non structurées NoSQL en lui alliant la simplicité et la puissance des requêtes traditionnelles des systèmes de bases de données structurées via SQL. La solution que l'entreprise propose est Open Source et donc disponible gratuitement si le déploiement se fait sur un serveur externe. Les technologies développées par les entreprises comme Crate.io visent à prendre en charge tous les types de données possibles, qu'ils soient structurés ou non.

Spécialement conçue pour les données produites par toutes sortes d'objets connectés (capteurs, wearables, équipements industriels, véhicules connectés, supervision d'infrastructures ou encore machine learning), CrateDB fait la synthèse entre bases de données distribuées de type NoSQL sur lesquelles elle se fonde et l'exploitation de SQL, langage standard de requêtes des systèmes de bases de données structurées comme MariaDB et MySQL. CrateDB assoit ses capacités de passage à l'échelle en exploitant des micro services et des conteneurs comme Docker, Kubernetes ou Mesos et peut ainsi évoluer en fonction du volume des données et du nombre de nœuds de calculs. La prise en charge des conteneurs traditionnels du Big Data permet à CrateDB de s'intégrer facilement dans une infrastructure logicielle conçue pour le machine learning tout en distribuant les données en fonction de l'évolution de la configuration.

Remplacer les SGBD traditionnels

Le défi auquel se confrontent les entreprises œuvrant dans le stockage de données dédié à l'Internet des objets, consiste à remplacer des technologies ayant déjà fait leurs preuves dans la gestion de bases de données. Ces entreprises de l'Internet des objets viennent concurrencer directement les systèmes de gestion de bases de données structurées MySQL, Postgres et Oracle sur des questions de performances et de coûts et défier les systèmes de gestion de bases non structurées comme MongoDB, Cassandra, Elastic Splunk ou encore Hadoop (plus généralement les systèmes exploitant des moteurs de type Map Reduce) sur leurs facilités d'utilisation et d'intégration.

Un véhicule connecté génère environ 2000 enregistrements par seconde, soit plus de cent milliards par jour, générant un volume de données considérable qu'il faut pouvoir stocker et traiter. Des entreprises comme Waterline Data, Trifacta, GridGain, InfluxData ou encore Crate.io (pour ne citer que les entreprises de la baie de San Francisco) créent des technologies pour ce faire. En se positionnant sur ce segment, ces acteurs et les technologies qu'ils développent mettent en lumière la difficulté pour les systèmes existants de gérer simplement ce flot de données. Cela prouve que l'écosystème de l'Internet des objets nécessite des solutions nouvelles et dédiées en réponse au besoin croissant de gérer des débits de données toujours plus élevés.

Rédacteur :

- Marc-Emmanuel Perrin, Attaché adjoint pour la Science et la Technologie, San Francisco, deputy-sf@ambascience-usa.org

Sources :

- <http://www.computerweekly.com/feature/Silicon-Valley-startups-aim-to-make-big-data-capture-and-prep-slicker>
- <http://www.businesswire.com/news/home/20170228005748/en/>
- <https://crate.io/>
- <https://www.marsouin.org/IMG/pdf/Usages-Proulx2-2005.pdf>
- <http://www.informatiquenews.fr/crate-io-sgbd-oriente-internet-objets-51329>