

# Mesa : Google dévoile son datawarehouse géant

Quand on s'appelle Facebook, Twitter ou **Google**, les questions d'infrastructures IT prennent en général une tournure un peu particulière. La gestion des données fait partie de ces spécificités et chacune des sociétés s'en occupe différemment. Chez Google, le système de **datawarehouse** se nomme **Mesa** et il vient de faire l'objet [d'une publication](#) de la part d'ingénieurs de la firme de Mountain View. Ils présenteront leurs travaux le mois prochain à la **conférence Very Large Database en Chine**.

Le document indique que Mesa est un système de datawarehouse avec des fonctions analytiques à grande échelle qui stocke les données critiques provenant de la division publicité en ligne de Google. Mesa est donc conçu pour répondre à des contraintes utilisateurs et systèmes spécifiques comme l'intégration et l'indexation en temps réel des données, la haute disponibilité, la tolérance aux pannes et l'évolutivité en fonction du volume d'informations et de requêtes. Concrètement, le service gère **des Po de données**, réalise **des millions de mises à jour par seconde** et sert des milliards de requêtes générant elles-même des milliers de milliards de lignes par jour. Il est **géo-répliqué à travers plusieurs datacenters** et assure une réponse rapide aux requêtes même quand un des centres de calcul est indisponible.

## Une offre proche de Vertica de HP

Pour cette solution, les ingénieurs expliquent qu'ils ont créé une base de données compatible ACID (atomicité, cohérence, isolation et durabilité). Elle a été façonnée pour les besoins métiers spécifiques de Google, mais elle pourrait très bien être proposée comme un service de datawarehouse traditionnel. Les scientifiques expliquent dans leur document qu'ils ont testé plusieurs bases de données existantes comme **BigTable, MegaStore, Spanner, F1**, mais sans être convaincus sur certains sujets comme la géo-réplication ou les montées en charge des requêtes. Ils apparentent la solution Mesa « à l'offre Vertica de HP pour la mise à jour dynamique et la gestion en temps réel des données transactionnelles ».

Dans les détails donnés par le document de travail, nos confrères de [GigaOM](#) soulignent que Google s'appuie sur des équipements de l'infrastructure Cloud maison, avec probablement du hardware conçu par lui. En effet, le système ayant précédé Mesa qui fonctionnait sur du matériel de classe entreprise était très onéreux à faire évoluer.

A terme, le service Mesa pourrait **servir à d'autres projets comme Hadoop** et des discussions sont en cours avec Cloudera. Par ailleurs, on peut imaginer qu'un tel système trouve sa place dans l'offre Cloud de Google au côté de BigQuery et plus récemment de DataFlow. Mais il faudra attendre encore un peu.

**Crédit Photo @Google**

**A lire aussi :**

[Google I/O : Dataflow alimente en temps réel Big Query en Big Data](#)

[Big Data : Google investit 80 millions de dollars dans MapR](#)