

Cloud : avec Kinesis, Amazon veut démocratiser le Big Data

Dans le cadre de sa **conférence re:Invent** organisée cette semaine à Las Vegas, Amazon a présenté un service visant à faciliter l'usage des applications Big Data. En complément des instances Hadoop ou NoSQL que proposent déjà tous les acteurs du Cloud, AWS a conçu un service clef en main, visant à **faciliter le traitement en temps réel des flux de données**.

Le service, baptisé Kinesis, permet « de stocker et de traiter des centaines de téraoctets de données par heure à partir de centaines de milliers de sources », assure le cyberlibraire dans un communiqué. Sa vocation consiste à **agréger et préparer les données pour des applications analytiques**.

Des bibliothèques de développement

Les applications envisagées vont de l'analyse du parcours de navigation des internautes, au suivi du sentiment des internautes sur les réseaux sociaux, en passant par le développement de jeux vidéos en tenant compte du comportement des « gamers » ou à l'analyse de flux financiers ou de données issues de capteurs.

Le point commun des applications imaginées par AWS ? **L'analyse temps réel**. Les frontaux analytiques peuvent être bâtis grâce aux bibliothèques Java fournies par Amazon. Grâce à ces dernières, les développeurs pourront **concevoir de nouvelles applications** (qui peuvent elles mêmes être hébergées sur AWS...) ou venir **enrichir la BI de l'entreprise**, pour générer de nouveaux tableaux de bord, des alertes, voire déclencher des réactions automatiques (ajustement de prix par exemple).

Dans son descriptif du service, AWS précise que Kinesis **s'adapte à la volée** (à la hausse comme à la baisse) – sans interruption de services – au volume de données « streamées » à un instant t. Le service est également intégré au stockage S3, au datawarehouse Redshift et à la base de données NoSQL DynamoDB, tous trois déjà au catalogue d'AWS.

Données éphémères

Concrètement, l'entreprise cliente réserve via la console de gestion d'AWS, une API ou un SDK une taille de flux de données (par blocs de 1 Mo/s), sans se soucier des infrastructures sous-jacentes. Par le même procédé, elle peut ajuster à la volée la taille du service.

Les données sont alors disponibles pour analyse et **stockées pour 24 heures**. Pour assurer la persistance des données, il faut alors les déplacer sur un stockage de long terme (comme S3 ou Redshift). Selon AWS, Kinesis embarque des fonctions de tolérance de panne et de réplication des données.

La [tarification](#) couple un **accès horaire au service** (0,015 dollar pour une heure avec 1 Mo/s en

entrée et 2 Mo en sortie) et un **coût au volume de données** (0,028 dollar par million d'entrées de données).

En complément : les autres nouveautés de la conférence AWS re:Invent

[- Amazon s'attaque au poste de travail virtualisé avec WorkSpaces](#)

[- AWS s'apprête à supporter PostgreSQL](#)