

Ingestion de l'univers visible dans une base de données distribuée Cloud-Native

jeudi 16 mars 2023 14:30 (30 minutes)

Le télescope Rubin Observatory va bouleverser l'astronomie.

Il permettra de constituer le plus large et le plus profond relevé de l'univers. Les objets célestes et leurs propriétés physiques seront identifiés et catalogués dans une base de données qui sera composée, à terme, de milliers de milliards d'entrées. D'un volume de l'ordre de plusieurs dizaines de pétaoctets, ce catalogue jouera un rôle majeur dans l'exploitation scientifique des données produites par le télescope.

Pour satisfaire ce besoin, un logiciel spécifique de base de données, nommé Qserv, est activement développé par une équipe d'une dizaine d'experts basée à l'Université de Stanford aux Etats Unis et par Fabrice.

Nous vous présenterons l'opérateur Qserv, basé sur le framework operator-sdk, qui permet de déployer et d'orchestrer cette base de données distribuée en quelques secondes sur des milliers de nœuds, aussi bien en architecture conventionnelle, au centre de Calcul de l'IN2P3, que Cloud-Computing.

De plus, nous détaillerons la manière dont nous pilotons et nous supervisons, grâce à Kubernetes et au framework ArgoWorkflow, les algorithmes qui permettent de charger dans Qserv, de manière automatique et reproductible, des millions de fichiers et des téraoctets de données cosmologiques en quelques minutes.

Nous conclurons en présentant notre feuille de route qui a pour objectif l'ingestion de centaines de millions de fichiers et de pétaoctets de données.

Orateurs: JAMMES, Fabrice (CNRS); MAINETTI, Gabriele (CC-IN2P3)

Classification de Session: Kubernetes dans l'ESR