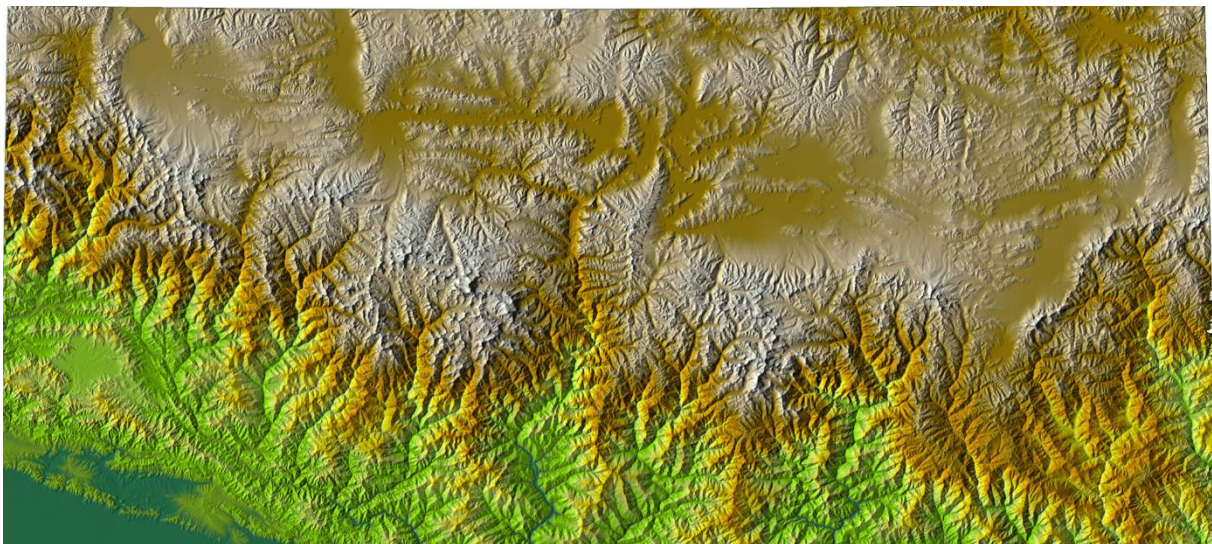


Projet Big Data

L'objectif du projet est de travailler à partir de données géolocalisées de grandes tailles pour fournir des visualisations interactives via une interface Web ou une application desktop connectée à l'infrastructure Big Data. Les données sont stockées sur HDFS et nous disposons de Spark ou Map Reduce, HBASE et express.js (+ npm hbase) java. Les jeux de données que nous pouvons utiliser sont les suivants :

- Des villes du monde géolocalisées avec leurs populations
- Des relevés satellitaires de la NASA indiquant le niveau du sol (hauteur) sur l'ensemble du globe. Ces données ont été récupérées sur le site <http://viewfinderpanoramas.org/>.
- Des positions GPS diverses obtenues par la communauté des utilisateurs d'Open-StreetMap. <http://wiki.openstreetmap.org/wiki/Planet.gpx>

L'objectif de ce projet est de proposer des visualisations de ces données sous la forme d'une carte chaleurs interactive. Une carte de chaleur est obtenue en agrégeant toutes les données qui tombent dans une cellule (binning) de la visualisation et en affectant la couleur de la cellule en fonction de la valeur agrégée. L'image ci-dessous a été obtenue en agrégeant tous les points des hauteurs sur la région de l'Himalayas. Les couleurs sont obtenues en utilisant un dégradé allant du vert d'eau (mer 0 mètre) au blanc (neige 8000 mètres) en passant par le jaune (4000 mètres). Nous pouvons générer le même genre de carte pour la population des villes mais aussi pour l'activité des utilisateurs d'OpenStreetMap.



La génération de ces cartes nécessite de lire toutes les données (plusieurs Giga) à chaque fois afin d'agréger toutes les valeurs qui tombent dans une cellule (un pixel pour une image). Nous voulons créer un système interactif dans lequel nous pourrions nous promener dans ces cartes comme nous pouvons le faire dans google maps (zoom and pan). Pour cela, nous voulons utiliser la puissance de calcul et de stockage de notre infrastructure Big Data. L'idée est précalculer toutes les agrégations possibles et les stocker de manière indexée afin de pouvoir y accéder très rapidement sans aucun recalcul.

Les jeux de données sont disponibles sur le serveur HDFS. L'objectif est d'effectuer l'étape de pré-calcul le plus rapidement possible en utilisant tous les moyens offerts par l'écosystème Hadoop. Votre système doit pouvoir procurer au final la précision la meilleure possible et être accessible via un serveur express.js s'exécutant dans la salle de TP ou bien une application Java s'exécutant elle aussi dans la salle de TP.

Consignes pour l'évaluation du Projet

Packaging

Vous devez pour votre projet : - Vous mettre par groupe de 2 obligatoirement - Ecrire un rapport de deux pages minimums, 8 pages maximums. Ce rapport doit expliquer ce que vous avez fait et comment vous y êtes parvenus. - Produire un jar exécutable directement sur le cluster utilisé en TP. - Donner le code source complet de votre projet.

Dépôts du projet

Tous les fichiers doivent être mis dans un répertoire portant vos deux noms préfixé de "ProjetPLE2017". Exemple ProjetPLE2017_Auber_David. Archivez ce répertoire en **zip** uniquement. Vous déposerez ensuite votre unique archive ayant pour nom ProjetPLE2016_Nom1_Nom2.zip sur le moodle ou bien par envoi d'email à auber@labri.fr si vous ne pouvez pas accéder à moodle.

Deadline !!!

Attention vous avez jusqu'au 28 janvier pour remettre votre projet.

Notation

La note se décompose en 5 parties :

- 5 points sur la qualité du rapport
- 5 points sur la qualité du code et de la solution
- 4 points sur la compréhension et la mise en œuvre du projet
- 4 sur la présentation en soutenance.
- 2 et plus sur la performance de votre solution.

Soutenance

Pour votre soutenance vous aurez 6 minutes pour présenter 6 slides, soit 3 minutes et 3 slides chacun, détaillant votre solution, ses performances, ses défauts et les choix que vous avez effectués.

Vous déposerez vos 6+1 slides (il faut 1 slide de titre avec vos noms et prénom) au format pdf sur moodle ou par email si vous n'avez pas d'accès à moodle. Votre fichier devra s'appeler SlidesPLE2017_Nom1_Nom2.zip, vous devez remettre vos slides 24heurs avant la soutenance.