
Big GeoData - Implantación de un entorno Big Data para la integración y análisis de datos geoespaciales en el centro de supercomputación extremeño CénitS-COMPUTAEX

En el proyecto se ha implementado un proceso completo de Big Data utilizando datos meteorológicos con el objetivo de conseguir información útil de determinadas zonas de interés. Los datos fueron obtenidos de las bases de datos de la NASA, siendo estructurados para aprovechar la información de sus satélites sobre la actividad del viento en la superficie de mares y océanos. Adicionalmente, estos datos fueron transformados y almacenados en una base de datos NoSQL de código abierto, MongoDB. La información fue seleccionada siguiendo estudios de viabilidad de trabajos previos, con tratamiento indexado de los datos georeferenciados mediante el uso de un índice 2dsphere. Asimismo, los datos directamente descargados de los repositorios NASA se fueron transformados para realizar procesamientos selectivos por regiones de interés con técnicas MapReduce, debido al gran volumen de datos a tratar, utilizando para ello Apache Hadoop.

El algoritmo MapReduce desarrollado por Dávila tiene dos tareas principales, por un lado, Map se encarga de tomar un conjunto de coordenadas y le asigna una clave dependiendo de la región a la que pertenezca y, en segundo lugar, Reduce toma el conjunto de elementos proporcionado por Map y combina los elementos en un conjunto más pequeño, realizando un filtrado de datos. Por último, para ejecutar tareas analíticas sobre los datos seleccionados mediante MapReduce o por consulta directa sobre la base de datos MongoDB, se permite la invocación a los algoritmos alojados en el entorno WEKA (Waikato Environment for Knowledge Analysis), facilitando así el empleo de algoritmos de análisis de cada una de las regiones obtenidas. Todo el proyecto se lleva a cabo en una máquina virtual de SpatialHadoop, analizando su escalabilidad con diferentes configuraciones de cara a su posterior implantación en un entorno HPC (High-Performance Computing).

Fuente de la publicación:

- Fátima Dávila Benítez, Félix Rodríguez Rodríguez y José-Luis González-Sánchez. *Big GeoData: Implantación de un entorno Big Data para la integración y análisis de datos geoespaciales en el centro de supercomputación extremeño CénitS-COMPUTAEX*. Trabajo Final de Grado (TFG) en Ingeniería Informática en Ingeniería del Software de la Universidad de Extremadura.

Documentos relacionados:

Noticias relacionadas:

- Becaria de investigación de CénitS presenta su Trabajo de Fin de Grado [[Cénits](#) [1]].

URL del

envío: <https://web.computaex.es/enlaces/publicaciones/big-geodata-implantacion-entorno-big-data-integracion-analisis-datos>

Enlaces

[1] <http://www.cenits.es/noticias/26112020-becaria-investigacion-cenits-presenta-su-trabajo-fin-grado>