

Towards Large Scale Environmental Data Processing with Apache Spark

Título Towards Large Scale Environmental Data Processing with Apache Spark

Autores Ferrón D., Villarroya S., Viqueira J.R.R., Pena T.F.


Tipo Comunicación para congreso

Fonte 20th Pacific Asia Conference on Information Systems, Chiayi, Taiwan, 2016.

Abstract Currently available environmental datasets are either manually constructed by professionals or automatically generated from the observations provided by sensing devices. Usually, the former are modelled and recorded with traditional general-purpose relational technologies, whereas the latter require more specific scientific array formats and tools. Declarative data processing technologies are available both for relational and array data, however, the efficient declarative integrated processing of array and relational environmental data is a problem for which a satisfactory solution has still not been provided. Due to the above, an integrated data processing language called MAPAL has been proposed. This paper provides a brief description of the design decisions and challenges, related to data storage and data processing that arise during the ongoing implementation of MAPAL on top of the Apache Spark large scale data processing framework.

Palabras clave Environmental Data, Data Processing, Big Data, Apache Spark

DESCARGAS

 Referencia BibTex