International Journal of Agricultural and Environmental Information Systems
Volume 11 • Issue 1 • January-March 2020

Hadoop Paradigm for Satellite Environmental Big Data Processing

Badr-Eddine Boudriki Semlali, LIST Lab, Faculty of Sciences and Techniques of Tangier, Abdelmalek Essaâdi University, Morocco
https://orcid.org/0000-0003-0671-4808

Chaker El Amrani, LIST Lab, Faculty of Sciences and Techniques of Tangier, Abdelmalek Essaâdi University, Morocco

Guadalupe Ortiz, UCASE Research Group, Escuela Superior de Ingeniería, University of Cadiz, Spain

ABSTRACT

The important growth of industrial, transport, and agriculture activities, has not led only to the air quality and climate changes issues, but also to the increase of the potential natural disasters. The emission of harmful gases, particularly: the Vertical Column Density (VCD) of CO, SO_2 and NOx, is one of the major factors causing the aforementioned environmental problems. Our research aims to contribute finding solution to this hazardous phenomenon, by using remote sensing (RS) techniques to monitor air quality which may help decision makers. However, RS data is not easy to manage, because of their huge amount, high complexity, variety, and velocity, Thus, our manuscript explains the different aspects of the used satellite data. Furthermore, this article has proven that RS data could be regarded as big data. Accordingly, we have adopted the Hadoop big data architecture and explained how to process efficiently RS environmental data.

KEYWORDS

Big Data, Hadoop Architecture, Remote Sensing, Satellites Sensors

1. INTRODUCTION

In this last decade, the world has suffered from various environmental problems, and several natural disasters, including air pollution, abnormal climate change, earthquakes, and so on (Smith et al., 2014). In this sense, it is important to supervise the climatic and pollution data such as temperature, humidity, wind speed and concentration of trace gazes in atmospheric layers, particularly in the troposphere. For this purpose, satellite data and Remote Sensing (RS) can be of great utility. In this investigation, we aim to apply some of this satellite is applications in Morocco areas (Badr-eddine Boudriki Semlali, Amrani, & Denys, 2019). We are going to track pollutants plumes emitted from the agricultural zones and the wildfire some using appropriate satellite sensors, see Table 6. Furthermore, we are looking forward to use satellites data in order to supply Moroccan forecasting agencies by providing the daily processed datasets and imageries, for instance we could combine satellites data with ground monitor datasets to produce a daily report forwarded to the forecasters. In addition, RS data will help us to monitor anthropogenic

DOI: 10.4018/IJAEIS.2020010102

pollutant emissions and climate changes of Morocco in Near-Real Time (NRT), to prevent damages and help in the decision makers (Badreddine Boudriki Semlali, & Chaker, 2019). These data can be also used as an input of some climate or Air Quality (AQ) models such as CALMET/CALLPUFF and AEROMOD (Holnicki, Kałuszko, & Trapp, 2016). Recently, satellites data support many potential applications such as in, pollutant plumes tracking, AQ monitoring and weather forecasting (Duncan et al., 2014). Generally, RS technique refers to the use of the technologies measuring the specifications of earth surface, ocean and atmosphere components without making a physical contact with it through the electromagnetic energy (EME) (Chijioke, 2012). This technique employs plenty of sensors. These satellites produce daily a large number of datasets, coming from various sources and diverse sensors within a different spatial, temporal and spectral resolutions (Yakubailik, Romas'ko, & Pavlichenko, 2019), these data have also different file formats and are continuously increasing storage spaces (Ma et al., 2015). Accordingly, RS data are regarded as Remote Sensing Big Data (RSBD) (Oussous, Benjelloun, Ait Lahcen, & Belfkih, 2017). Thus, the processing of RSBD includes several challenges in term of data collection, storage and handling (Ma et al., 2015). As a result, it is necessary to develop a Big Data (BD) platform enabling data collection, sort, categorizing, analyze and storage. In this study we will cover the basics of the RS techniques, the use of satellites and sensors. For this purpose, we will collect satellites data from the Mediterranean Dialogue Earth Observatory (MDEO) terrestrial station installed in Abdelmalek Essaâdi University of Tangier (El Amrani, Rochon, El-Ghazawi, Altay, & Rachidi, 2012), the National Aeronautics and Space Administration (NASA), the National Oceanic and Atmospheric Administration (NOAA), the European Space Agency (ESA), some Meteorological Ground Station (MGS) and from a Raspberry PI ground sensor in NRT. These satellites produce daily a large number of datasets, coming from various sources and diverse sensors within a different spatial, temporal and spectral resolutions, these data have also different file formats and are continuously increasing storage spaces (Ma et al., 2015). Accordingly, and according to the attribute definition of Big Data (BD) based in the 4Vs (volume, variety, velocity and veracity and so on), RS data are regarded as RSBD. Thus, the processing of RSBD includes several challenges in term of data collection, storage and handling (Sun, Liu, Ma, Liu, & Sun, 2016). As a result, it is necessary to develop a BD platform enabling data collection, sort, categorizing, analyze and storages. Moreover, we will prove that the received data are BD. For these purposes, we will adopt the Hadoop architecture to process RSBD.

2. ISSUES

There are many environmental problems and RS data management challenges which are:

- The apparition of the natural disasters and the environmental issues including: forest fire, climate changes and air pollution;
- The RS data are complex, have huge volume, and high velocity and veracity;
- Current and architectures of RS data processing are limited and face many challenges.

3. MAIN FOCUS OF THE ARTICLE

This investigation has several goals which are:

- Presenting a brief survey about the used satellites and sensors for air pollution monitoring;
- Performing a RS data analysis of four satellite data sources including the MDEO, the NASA, the NOAA and the Copernicus platform built by the ESA;
- Proofing that the used RS data are BD based in the 4Vs of BD;
- Adopting the Hadoop paradigm for a RS data processing;
- Exploiting the RS data in air pollution in Morocco.