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.

4. RELATED WORKS

Table 1 includes some related works focusing in RS data processing and environmental issues.

5. REMOTE SENSING: DEFINITION, APPLICATIONS, SATELLITES AND SENSORS

In this section, we will explain RS technique, the primary use of satellite and the sensor specifications.

5.1. What is the Remote Sensing Technique?

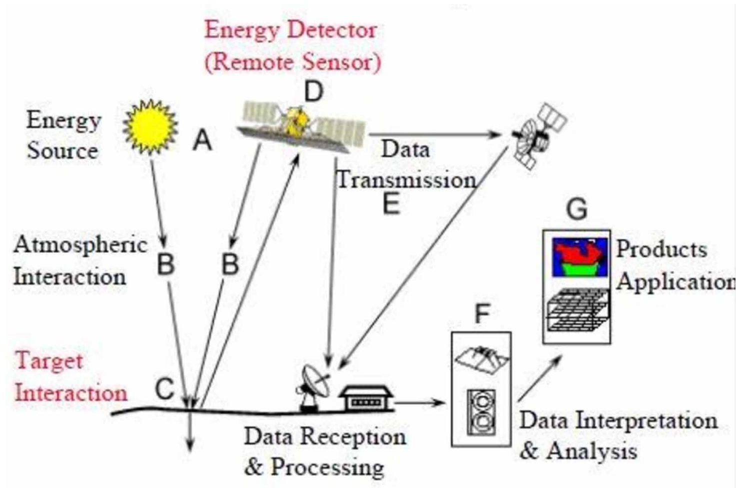
RS generally refers to the use of the technology for measuring the specification of surface, ocean and atmospheric components without making a physical contact with them, by using the Electromagnetic Energy (EME) (Chijioke, 2012). This technique monitors earth continuously using active sensors (Sandeep Gupta, n.d.), while the signal is emitted and detected by the same sensor. Consequently, images are formed by the scattered or reflected microwaves. Moreover, each target has a unique spectral signature (see Figure 1), as explained below:

A: Sun or satellites illuminate the target.
B: Effect of passing through the atmospheric layer.
C: Ground target reflects the energy.

Table 1. Some related works and their focusing

Investigation	Computer and IT Topics	Environmental Application
1. Monitoring of atmospheric composition using the thermal infrared IASI/MetOp sounder (Clerbaux et al., 2009)	Satellite sensors calibration and satellite data sets validation and optimization.	The use of the Infrared Atmospheric Sounding Interferometer (IASI) satellite sensors to monitor atmospheric pollution and changing composition.
2. Thermal infrared geostationary satellite sensor data application for prediction and monitoring earthquake in Algeria (Hassini & Belbachir, 2016)	The processing of the Meteosat Second Generation (MSG) data and the application of the algorithms of datamining for prediction.	This approach depicts some anomalous increases in surface temperature that occur before an earthquake and the case study was the earthquake of Algeria.
3. Real-Time Big Data Analytical Architecture for Remote Sensing Application (Rathore et al., 2015)	The proposition and the development of a new parallel architecture for RSBD processing based to MapReduce in NRT.	In this paper author acquires and process ENVISAT data because ENVISAT satellite mission has been continuously providing global measurements for the earth including sea, land, ice, and forest since 2002.
4. Development of a Java-based application for environmental remote sensing data processing (Badr-eddine Boudriki Semlali et al., 2019)	The development of a Java-based application software to collect, process and visualize several environmental and pollution data, acquired from the MDEO platform.	This research has focused in air pollution monitoring in Morocco using the MDEO data sources and authors found out a correlation between the emitting sources and the densities of traces gases.
5. Air4People: A Smart Air Quality Monitoring and Context-Aware Notification System (Garcia-de-Prado, Ortiz, Boubeta-Puig, & Corral-Plaza, 2018)	Development of a mobile application checking current air quality values at any available station, receiving notifications concerning current air quality at one particular station of interest.	This research presents a context-aware notification system, which submits personalized alerts to citizens based on several types of context, whenever air quality-related health risks are detected for their particular context.

Figure 1. RS process (Chijioke, 2012)



D: Reflected energy recorded by satellite is sensor.
E: Transmission of data to ground stations.
F and G: Conversion of data into digital format including maps, images or raster files.

The second type is the passive sensor, when the detection is based only to the reflection of the sunlight.

RS technique helps also to collect data from dangerous or unreachable areas such as oceans, forests, deserts or atmospheric layers. The gathered RS data are processed in order to remove imperfections, ensure geometric corrections and apply data calibrations (Li, Ding, & Long, 2015). In 1986 NASA defined four levels of RS processing beginning from level0 to level3, see Table 2 (Parkinson, Ward, & King, 2005).

Currently RS data are becoming more widely used in the decision making and the environmental management activities. This can be achieved thanks to the gathering of a large volume of datasets with high velocity, and this is why RS data are currently regarded as RSBD (Ma et al., 2015). In this research, we have collected RS data from a large number of satellites equipped with different sensors from the MDEO, NASA NOAA, ESA and some MGS.

5.2. Primary Use of Satellites

Nowadays, we notice several environmental issues, such as outdoor pollution due to the emission of anthropogenic gases by industrial and transport activities (Badr-eddine Boudriki Semlali et al.,

Table 2. RS data processing levels

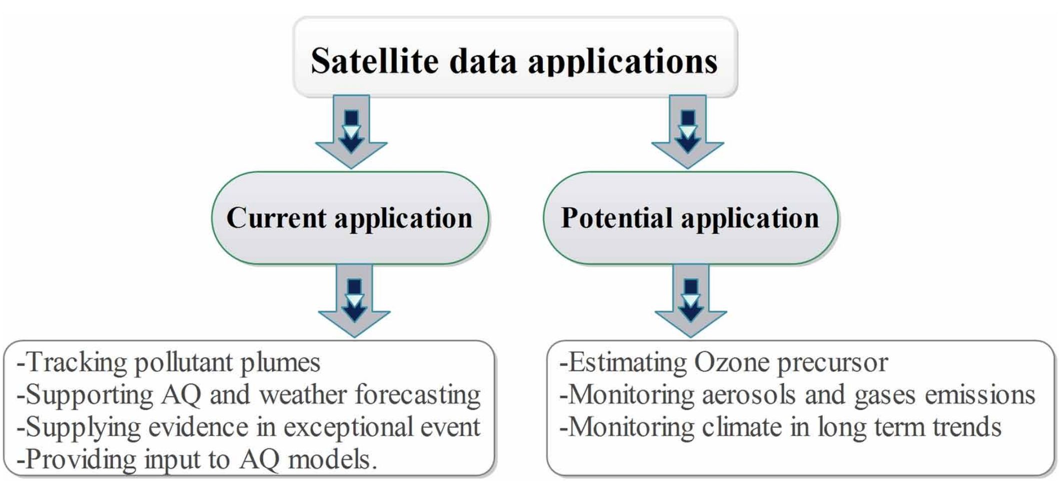
Level	Description
1a	Reconstructed, unprocessed instrument data at full resolution, time-referenced, and annotated with ancillary information, including radiometric
1b	Level 1a data that have been processed to sensor units.
2	Derived geophysical variables such as Aerosol Optical Depth (AOD) or the Vertical Column Density (VCD) of some trace gases at the same resolution and location as Level 1 source data.
3	Variables mapped on uniform space time grid scales, usually with some completeness and consistency.

2019). Furthermore, we have noticed an increase of natural disasters caused by the climate changes, particularly storm, flood and heatwave (EL AMRANI Chaker, 2015). So, there is a wealth of atmospheric composition of satellite data for AQ applications and climate monitoring in NRT (see Figure 2). Besides, RS data are further used to measure the density of the traces gases in the different atmospheric layers (Duncan et al., 2014). This role is recognized by the Environmental Protection Agency (EPA) and several AQ agencies.

5.3. Satellite and Sensor Specifications

Satellite is regarded as an artificial machine which has been placed into a specified orbit. The story of satellite was beginning in October 4, 1957 by the lunch of the first Russian satellite Sputnik-1. Currently, there are more than three thousand satellites in orbits. Approximately, five hundreds are polar satellites placed in a Low Orbit (LEO), fifty satellites are in the Height Orbit (HEO) and the rest are geostationary satellites (How many satellites are orbiting the Earth in 2018?, 2018). There are many purposes of the use of these satellites: military, earth observation, weather monitoring, forecasting and so on. Some of these satellites help in the scientific researches providing meteorological information, land survey, AQ monitoring. All of these satellites are placed into a geocentric orbit with a specific altitude, synchronous specifications and classification (Wilkin, 2010). In this study, we collect RSBD from the MDEO platform, The Earth Observing System Data and Information System (EOSDIS) of NASA, the National Environmental Satellite Data and Information Service (NESDIS) of NOAA and the Copernicus operated by ESA. The MDEO platform gives us access to the European Organization for the Exploitation of Meteorological Satellites (EUMETSAT) data in NRT using the ground station of acquisition placed at Abdelmalek Essaâdi University of Tangier (El Amrani et al., 2012). Wherein, all of these satellites are for environmental monitoring, meteorological information and for scientific researches purposes. We find out that all of these satellites are polar (Schmetz et al., 2002), passing by a sun synchronous orbit, with approximately eight hundred Kilometers (Km) of altitude into LEO, expecting the Meteosat Second Generation (MSG) is a geostationary satellite. Satellites have usually a specified life time ranging between four to twenty years, for instance the MetOp A was launched in 2006 and will be in orbit until 2020 (Zaytar & Amrani, 2017). We have used also newer satellites, particularly the Sentinel series, thus we have collected data from Sentinel-3A and Sentinel-5P lunched in 2016 (Showstack, 2014). We notice that all of these polar satellites make around fourteen to sixteen orbits daily, within one hundred min per orbit. However, they cross Morocco areas twice a day into a specified range time. The unique geostationary satellite used in this

Figure 2. Satellite data applications



research is the MSG: it passes over the equator with an altitude of thirty-six thousand Km (Schmetz et al., 2002). Moreover, all these satellites are equipped with several sensors that enable measurements. Table 3 shows all the used satellites in our research with their full specifications.

Satellites sensors are regarded as instruments measuring many earth, atmospheric and ocean variables. Generally, there are two kinds of satellites sensors, the first are active instruments such as the Light Detection and Ranging (LIDAR) and the Radio Detection and Ranging (Radar), they illuminate the scanned object with their energy, emit radiation over target and then, detect the reflected or backscattered radiation (see Table 4). The second type are passive instruments, they detect natural radiation emitted from target, particularly sunlight and photons (see Table 5).

Satellite sensors have some other characteristics which are listed below:

- **Spatial Resolution (SPR):** SPR means the earth surface covered by the instrument, there are a high and low resolution as shown in Figure 3;
- **Temporal Resolution (TPR):** TPR refers to the revisiting frequency of satellite sensors to the same location, normally we distinct three TPR as illustrated in Figure 3;
- **Spectral Resolution (STR):** STR means the number of spectrum bands in which sensor can collect the reflected radiance. Figure 4 shows the electromagnetic spectrum. Accordingly, satellite sensors offer a relation between the spectral and the spatial resolution. Where the STR increases, the SPR decreases.

Table 3. Used satellites specifications

Satellite Name	Organization	Lunching date	Weight (Kg)	Altitude classification	Altitude (Km)	Synchronous Classification	Orbits/ day	Instruments
MSG	EUMETSAT (MDEO project)	2002	2000	High Earth orbit (HEO)	36 000	Geostationary orbit (GEO)	1	SEVIRI
MeTop A		2006	4100	Low Earth orbit (LEO)	830	Sun-synchronous orbit (SSO)	14	IASI-MHS-GOME2-ASCAT-AVHRR-AMSU-HIRS
MeTop B		2012	4101		831		14	
TERRA	NASA (EOSDIS project)	1999	5190		705		16	ASTER-MISR-MODIS-MOPIT
AQUA		2002	2934		705		16	AIRS-AMSU-MODIS-HSB
AURA		2004	2970		708		16	MLS-OMI-TES
POES	NOAA (IDEA project)	2009	1479		850		16	AVHRR-HIRS-AMSU-MHS
NPP		2011	2100		832		16	VIIRS-ATMs-OMPS
Sentinel 5P	ESA (Copernicus project)	2017	820		824		14	TROPOMI
Sentinel 3		2016	1250		814		14	OLCI-SLSTR-SRAL
GOSAT-2		2009	2000		680		15	FTS- Imager

Table 4. Active satellites sensors feature

Sensor Name	Feature
Hyper Spectral radiometer	Advanced multispectral sensor that detect a very narrow spectral bands, visible, near infrared and mid infrared
Radiometer	Measure the intensity of electromagnetic radiation within the spectrum, radiometer is identified by a portion of spectrum like such as (visible, infrared and microwave)
Sounder	Measure vertical distribution of atmospheric parameters like temperature, pressure, VCD traces gases
spectrometer	Detect, measure and analyze spectral content f incident electromagnetic radiation
Spectra radiometer	Measure the intensities in multiple wavelength bands

Table 5. Passive satellites sensors feature

Sensor Name	Feature
LIDAR	Light detection using laser to measure distance between a target
Radar	Emit electromagnetic energy like microwave to detect distance or range of target
Scatter meter	Height frequency microwave radar, measure backscattered radiation over ocean Surface, detect the wind speed and wind direction.
Sounder	Instrument measuring vertical distribution of precipitation and atmospheric composition (Temperature, humidity, cloud composition, VCD of gases)
Laser Altimeter	Measure the height of platform.

Figure 3. Satellite sensors resolutions

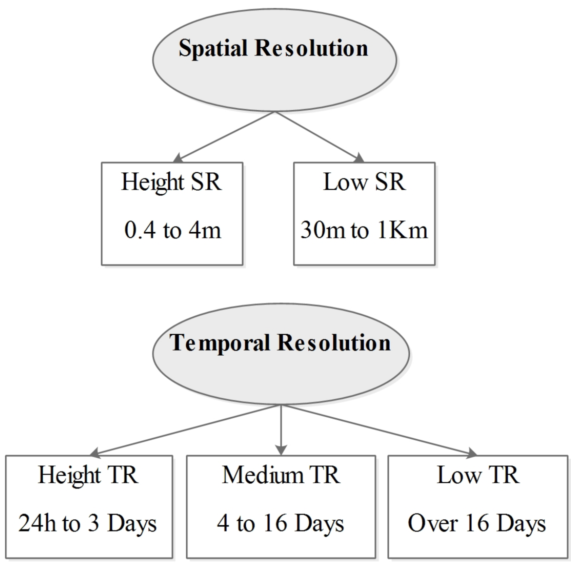
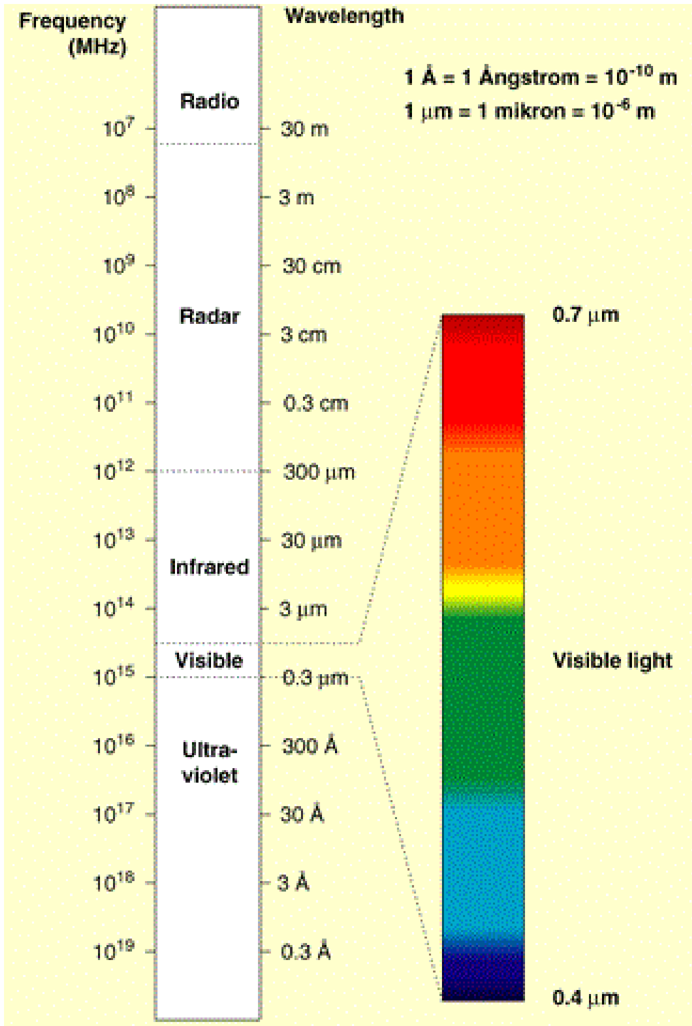


Figure 1 explains the mechanism of satellite sensors, in resume, instruments emit the EME like microwave passing through atmospheric layers, afterward, and reflected energy by the measured target is recorded. Then, values are transmitted to the ground stations to be converted into raster images, map or scientific data files. Table 6 summarizes the most used sensors in this research, the majority of these sensors are active sounders like: the AMSU, IASI (*BUFR Descriptors for IASI Level 2 Data*, 2017), MHS, MLS and AIRS. There are also some active spectrometers such as the TROPOMI, TANSO-FTS, MODIS, MIRS and the GOME-2 (Rosemary, 2016). The SPR of the aforementioned sensors ranges between high to low. In addition, all of these satellite sensors have a high TPR feeding from one to two days.

6. METHOD

This section shows the main aspects and specifications of the used satellite data presenting the diverse data sources and format of the acquired data.

Figure 4. The electromagnetic spectrum



6.1. Satellites BD Aspects and Application to the Case Study

BD is a term used in 1950 by John Mashey, at present, BD is become more popular, and it attracts diverse attentions from both technological experts and the public in general. Commonly, BD refers to a large, diverse and complex data that current applications and architectures are not able to manage them efficiently (Oussous et al., 2017). However, the definition of BD is rather diverse, and reaching a consensus is difficult (Hu, Wen, Chua, & Li, 2014). Consequently, there aren't a unique and a perfect definition of BD concept. BD does not mean only the massive data but also other features that differentiate it from the concept of "massive data" (Russom, 2011). As a result, there are three different definitions: the attribute definition that delineates the four salient features of BD, particularly: the volume, the variety, the velocity and the value. Accordingly, the 4Vs definition is widely used to characterize BD. The second one is the comparative definition that refers to datasets that go beyond the ability of storage, managing, and processing. The last denotation is the architectural definition that suggest that BD limits the ability to perform effective within the traditional relational approaches, or requires the use of significant horizontal scaling for efficient processing (Hu et al., 2014). These

Table 6. Used satellites sensors specifications

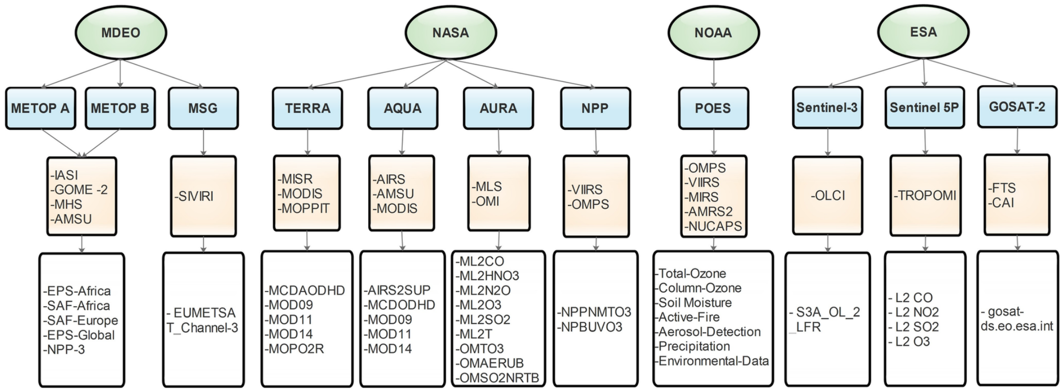
Instrument	Abbreviation	Type (Spectral Resolution)	Spatial Resolution
TROPOMI	Tropospheric Monitoring Instrument	Spectrometer (NIRUV)	High
OLCI	Ocean and Land Color Instrument	Optical	High
TANSO-FTS	Fourier Transform Spectrometer	Spectrometer (NIR)	High
MISR	Multi-angle Imaging Spectro Radiometer	Spectroradiometer (NIR)	High
MODIS	Moderate Resolution Imaging Spectrometer	Spectrometer (IR)	High - Low
AIRS	Atmospheric Infrared Sounder	Sounder (IR)	Low
AMSU	Advanced Microwave Sounding Unit	Sounder (Microwave)	Low
MLS	Microwave Limb Sounder	Sounder (Microwave)	Low
OMI	Ozone Monitoring Instrument	Hyperspectral Imaging (Visible)	Low
MHS	Microwave Humidity Sounder	Sounder (Microwave)	Low
VIIRS	Visible Infrared Imaging Radiometer Suite	Radiometer (NIR)	Low
IASI	Infrared Atmospheric Sounding Interferometer	Sounder (IR)	Low
ASCAT	Advanced Scatter Meter	Scatter meter	Low
GOME-2	Global Ozone Monitoring Experiment–2	Spectrometer (UV)	Low
SEVIRI	Spinning Enhanced Visible and Infrared Imager	Radiometer (NVIR)	Low

definitions concern also RSBD. On the other hand, RS data are becoming more widely used from distinctive fields (e.g., environmental monitoring, climate forecasting, and scientific researches), by the advanced of satellites, sensors, the RS is undergoing an explosive growth (Ma et al., 2015). This advances in sensors technologies have given rise also, to the increase of the complexity of RS data, particularly the diversity and the dimensionality. Furthermore, the large scale environmental are exploiting regional to global covered temporal and multi sensors RS data for processing. Moreover, RS technique deals with a large collection of datasets with a huge volume. So, the first definition is more suitable to describe RSBD. In the next subsection, we will discuss several aspects and features of the used RSBD, the diversity of data sources, and their huge volume, complexity and velocity.

6.2. The Diverse RS Data Sources

Figure 5 shows several sources of RS data. This has become a common barrier to end users, because of the increasingly large number of data access form and datasets types. In this study, we have used the MDEO data as the main RS data sources (Badr-eddine & El Amrani, 2019). The MDEO platform provides NRT data from multiple EUMETSAT satellites. These data support scientific researches such as pollution monitoring, early warning against disasters, particularly, storm and flood, and it supervises the climate changes. EUMETSAT satellites send data, to the ADC in Antarctica and Green Land after being measured. Then, data is transmitted into the Central Facility (CF) in Darmstadt to produce data of level b2. Afterward, data is distributed to the regional and global EUMETCast ground station like the MDEO ground station installed at Abdelmalek Essaâdi University (El Amrani, Rochon, El-Ghazawi, Altay, & Rachidi, 2013). Finally, RSBD data are ready to be downloaded and processed. We have not only used the MDEO RS data but also some NASA satellites, these data are provided and managed by the Earth Observation System Data and Information System (EOSDIS) and distributed into twelves datacenters in the world (“EOSDIS,” 2017). To make the research and the filter of the satellite is products easier, we have used the Reverb search engine to discard unnecessary RS data. We collected also pollution and aerosols products from the Godard Earth sciences Data

Figure 5. Used RS data sources



and Information Services Center (GESDISC). We have collected RS data also from other federal agencies providing free and NRT RS data such as the Infusing Satellite Data into Environmental Applications (NESDIS) of NOAA ("NOAA National Environmental Satellite, Data, and Information Service (NESDIS) |," 2018). Moreover, we have acquired RS data from the Copernicus platform built by ESA. We have gathered also data from some Meteorological Ground Station (MGS) through these websites: www.weatherunderground.com and www.worldweatheronline.com in order to apply a data validation. In addition, we have developed a Raspberry PI robot, equipped with MQ sensors in order to support RS data validation. Figure 5 shows the different data sources with their satellites, instruments and their products details. Table 7 contains the Unified Resources Locator (URLs) of the used satellites data sources.

6.3. Ground Sensor Data of Raspberry Robot

Internet of Things (IOT) is the one of latest revolutions of internet. The story was beginning in the early of 1980s. The concept of IOT is to interconnect a big number of things or sensors through internet and will be able to communicate with each other (Zanella, Bui, Castellani, Vangelista, & Zorzi, 2014). IOT become more exploited in many applications such as: detecting wildfire in forest, waste management, air pollution and noise monitoring, traffic congestion and city energy consumption (Perera, Zaslavsky, Christen, & Georgakopoulos, 2012). However, our research focusses only to the integration of the IOT in the environmental issues, particularly, climate changes and air pollution monitoring. The European Union (EU) officially aims to reduce greenhouse gas emissions by 2020, thus, an urban IOT might help monitoring the AQ in crowded area and could optimize the health of people by finding the healthiest area (Zanella et al., 2014).

Table 7. RS data sources

Data Source	Download Link
MDEO	https://www.eumetsat.int/website/home/Data/DataDelivery/EUMETCast/index.html
EOSDIS	https://earthdata.nasa.gov/earth-observation-data/near-real-time/download-nrt-data https://www.ospo.noaa.gov/
NESDIS	ftp://ftp-npp.bou.class.noaa.gov/
Copernicus	https://scihub.copernicus.eu/

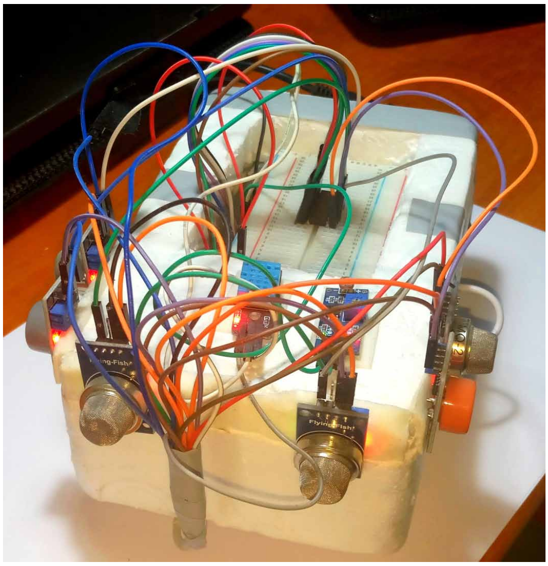
In our application, we have built and mounted a Raspberry PI 3 robot, see Figure 7, well equipped with many MQ sensors measuring traces gases and climatic variables such as temperature, humidity and sunlight. The battery of this robot can delay sixteen hours, could communicate with internet via Wireless-Fidelity (Wi-Fi) (Arya, Bhadoria, & Chaudhari, 2018) or mobile networks, and could be located with a high accuracy using the longitude and latitude calculated by the Global Position System (GPS) dongle. The MQ gas sensors series measure and detect many gases particularly (CO, $CO_{2,}$ $CH_{4,}$ $NH_{4,}$ and smoke), see Table 8. However, the Raspberry robot cannot read the digital signal from sensor; to fix this problem, we have used the MCP3008 microchip showed in Figure 6. The MCP3008 is a low-cost channel 10-bit-analog to digital converter (Bhadoria & Chaudhari, 2019). This chip could read the digital output from sensors with a good precision to that in an Arduino card. Thus, our purpose is to measure the

Table 8. Used MQ sensors specifications

Sensor is Name	Detected Gas
MQ-2	LPG
	CO
	SMOKE
MQ-3	CO
	CH_4
	LPG
MQ-4	LPG
	CH_4
	ALCHOOL
MQ-6	CH_4
	C_3H8
	C2H5OH
MQ-7	CO
	CH4
	H2
MQ-135	CO2
	CO
	NH4

Figure 6. The MCP3008 microchip and Raspberry PI 3

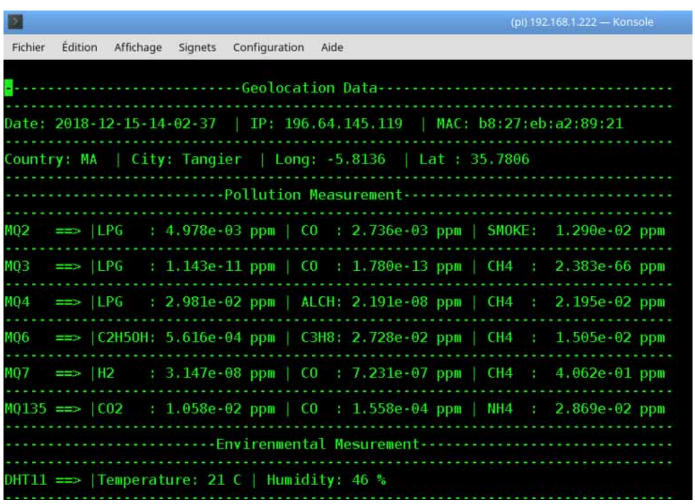
Figure 7. Raspberry PI robot



density of pollutant gases in the ground area. The obtained values from the robot are unit less. Consequently, we have used some mathematical equations to convert them into the Part per Million (PPM) unit. Finally, all the measured data are stored into a remote relational database to be exploited after in NRT.

Figure 8 shows a screenshot of the measured values by the Raspberry Pi robot using the aforementioned MQ sensors. The connection was done by the Secure Shell (SSH) protocol. We notice that there are many pollutant datasets such as the concentration of the CO_2 CH_4, CO, NH_4, Liquefied Petroleum Gas (LPG), and the smoke with the PPM unit. Moreover, this robot measures also the temperature and the humidity with high accuracy. All these datasets are recalculated after a delay of ten second and stored into a remote database.

Figure 8. A screenshot of the Raspberry PI measured values



7. STATISTICAL RESULTS

This section includes the statistical results concerning the variety, volume, and velocity of the received RS data.

7.1. The Variety of RS Data

Variety refers to the data form: structured, semi-structured or unstructured. RSBD are becoming more diverse due to the wide range of their use in earth sciences disciplines such as, environmental monitoring, climate supervision and oceanography. RSBD are commonly stored in semi-structured files including the Binary Universal Form for the Representation of meteorological data (BUFR) assumed by the World Meteorological Organization (WMO) (Zaytar & Amrani, 2017), the Hierarchical Data Format (HDF5) developed by the National Center for Supercomputing Applications (NCSA), the Network Common Data Form (NetCDF) created by the University Corporation for Atmospheric Research (UCAR) university, and so on. These formats have various physical structures. Accordingly, decrypting these kinds of files demand some specified libraries and interfaces, that make the read of RSBD data more difficult for non-expert end users. Figure 9 shows the most used file is formats in this research. We have to process forty two percent of the binary files, thirty-one of NetCDF files, twenty percent of HDF5 files and only five percent of the BUFR files. To decrypt this format, we have made in use some Python libraries such as the BUFRextract (BUFREXC) (Francis Breame, 2018) and the pybufr_ecmwf (ECMWF) (Siemen, Lamy-Thepaut, Li, & Russell, 2007). Furthermore, the MGS provide datasets in NRT with JSON and XML format, including useful values of temperature, humidity and so one. In addition, the Raspberry Pi 3 robot supplies us with the NRT data.

7.2. Measuring RS Data Volume

With the recent advances in sensors and earth observation, there are a deluge of RS data. Currently, more than two hundred satellites are in orbit capturing multi-spatial and multi spectral data from sensors. Figure 10 shows the used data volume acquired from the MDEO, the NASA, the NOAA, the ESA satellites, some MGS and the Raspberry Pi sensors, respectively. We notice that NetCDF and the HDF5 products are heavier than the BUFR and the binary data. However, data acquired from the MGS and the Raspberry Pi sensors have very low size approximately one Gigabytes (GB) because they are flat and JSON files. The total downloaded volume from the aforementioned sources is more than one hundred GB stored inside more than eleven hundred files per day and sumps up thirty-seven Terabytes (TB) per year. This amount of generated data is considered as low big and manageable comparatively with other pioneer datacenter such as the EOSDIS and the Chinese Academy of Sciences (CAS) reaching six TB per day.

Figure 9. Daily percent of RS data is format (%)

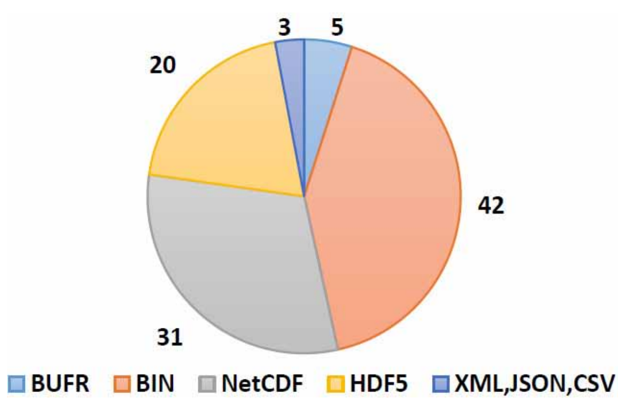
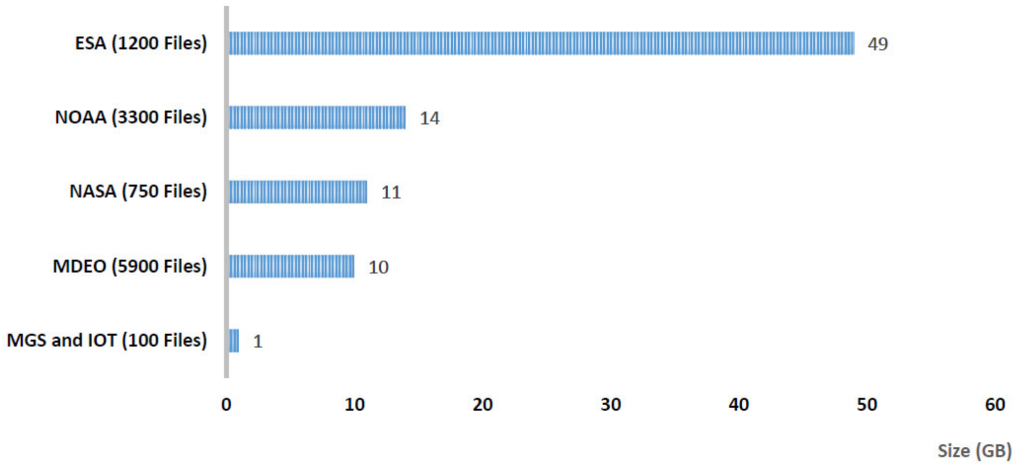


Figure 10. Daily RS data is volume (GB)



7.3. The Velocity of RS Data

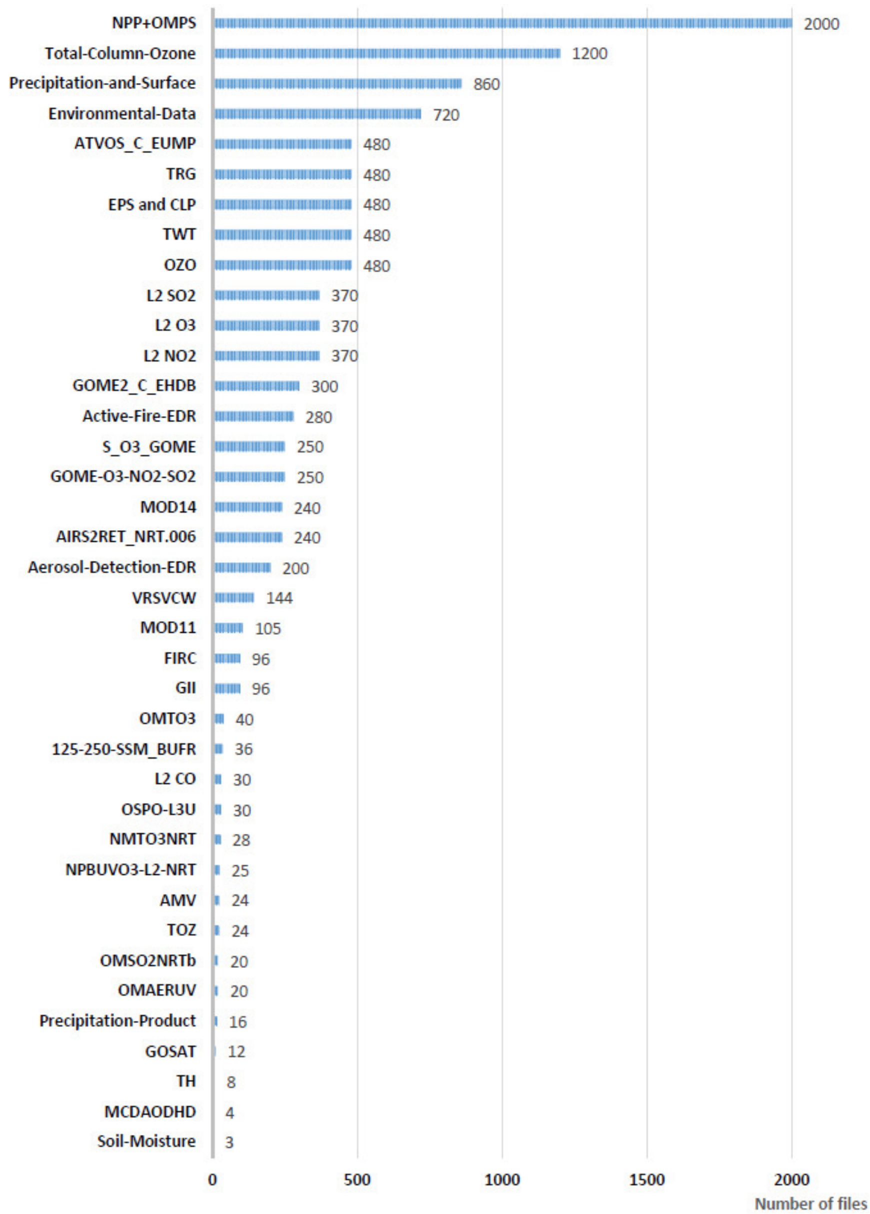
Velocity assigns to the rate of which data come, and how fast must be processed, by the age of Internet and the wide range of using satellites sensors and IOT, a rapid processing must be achieved to deduce useful deep insights in NRT, heling in decision making. Velocity varies in function of data category. There are two kind of data processing: the NRT or streaming data and the batch data. Streaming data are treated as soon as possible in order to keep potential value and data freshness. Moreover, data arrive continuously occurring enormous volumes (Sun et al., 2016). However, batch data are stored in chunks within a specified, finite size and their processing takes a long time, and it needs multiple Central Processing Unit (CPU) and memories (Russom, 2011). Figure 11 shows the MDEO, NASA, NOAA and ESA RSBD velocity. We notice an important daily number of files coming with a big size. For instance, the Suomi National Polar-orbiting Partnership (NPP) satellite provides data of the OMPS sensors with a high rate reaching two thousand file per day, EPS-Africa channel, provides about four thousand files daily, the Sentinel-5P affords also a huge daily data encompassing forty-eight GB. Similarly, the NASA and NOAA data have also an important velocity, see Figure 10 and 11. Thus, RSBD come in NRT, MDEO data are available with a frequency and latency of thirty minutes (Min). However, the NASA data have a frequency of five Min and two hours of latency. As a result, RSBD are regarded as batch datasets, because, they need a big capacity of storage, CPU and memories to be processed. Figure 12 displays the daily size of RS data.

From the aforementioned statistical results and from the attribute definition based in the four salient (Venue, Volume, Velocity and Variety) of big data we confirm that RS data are big data as illustrated in Figure 13.

8. THE PROPOSED BIG DATA ARCHITECTURE

Designing a BD architecture is the best way to split the problems of BD processing. We have to design and make in place well all the essential components of BD, where each layer has a specific function (Erraissi, 2017b). These architecture helps to trace the data pipeline, both batch and streaming processing. Generally, we find six layers in the BD architecture. The data sources layer, the ingestion layer, the Hadoop storage, management, infrastructure, security and the monitoring layer, see Figure 14. In the following subsections we explain the architecture in detail: Section 8.1 explains data sources, Section 8.2 describes ingestion layer, Section 8.3 describes storage layer, Section 8.4 details

Figure 11. Daily number of RS file

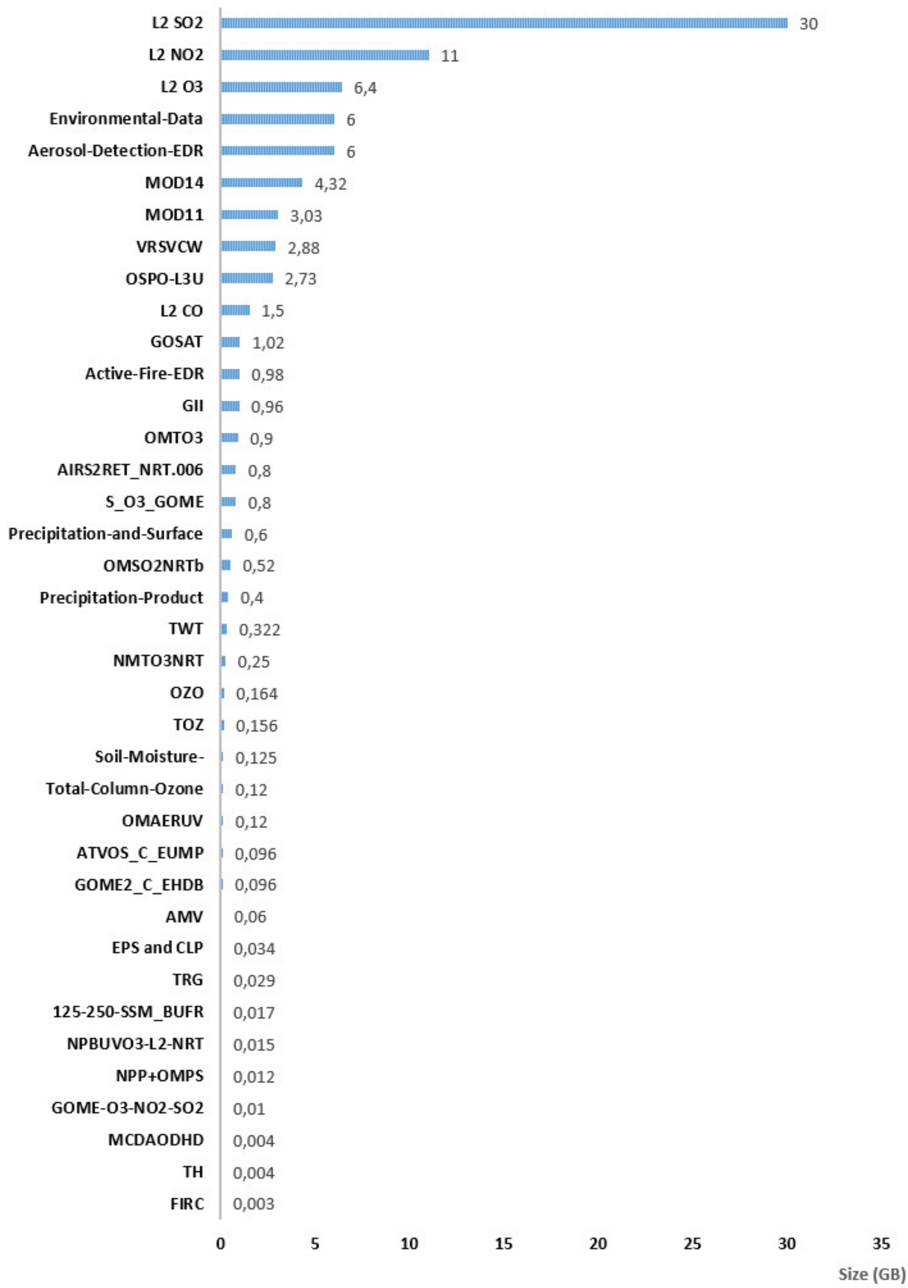


processing and query layer, Section 8.5 depict visualization layer and finally we can see monitoring layer in Section 8.6.

8.1. Data Sources

This layer defines the various entries of RS data acquired from the different venues with a huge volume, high velocity and variety, that need an efficient treatment in the ingestion layer. In our case, there are five main data sources providing in NRT RSBD. Which are the MDEO ground station of EUMETSAT data, the EOSDIS data access of NASA, the NESDIS data access of

Figure 12. Daily size of RS data (GB)



NOAA, the Copernicus platform built and operated by ESA, the MGS and the Raspberry PI robot sensors data. The role of this layer is to automatically connect with many File Transfer Protocol (FTP) or Hypertext Transfer Protocol (HTTPS) protocols using the WGET and CURL, doing the authentication, selecting the interesting channels and products to be downloaded and finally making data available in NRT to be preprocessed.

Figure 13. The four salient of remote sensing big data

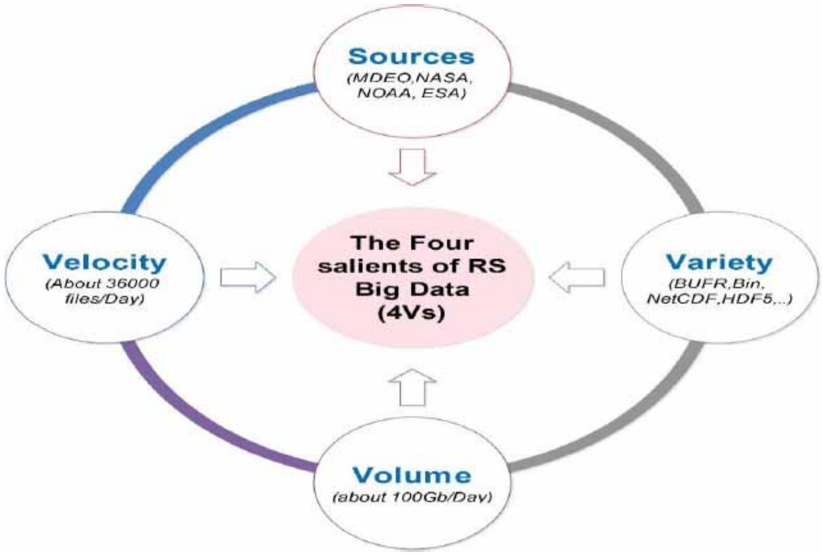
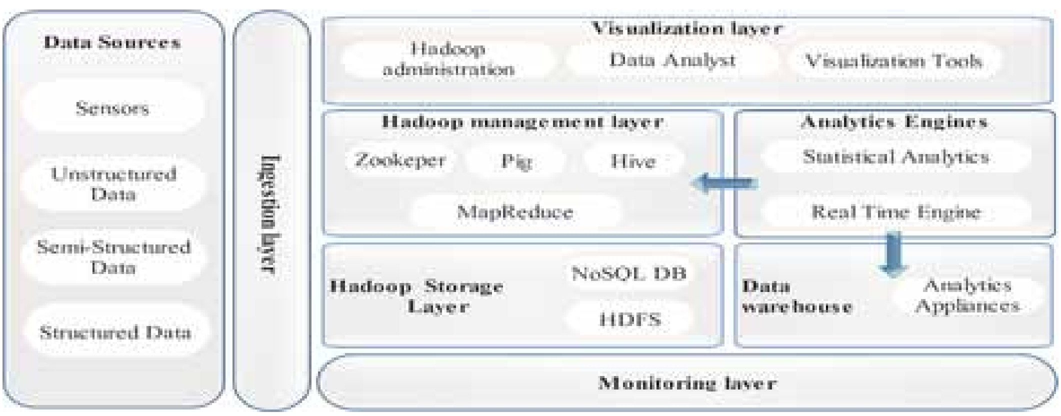


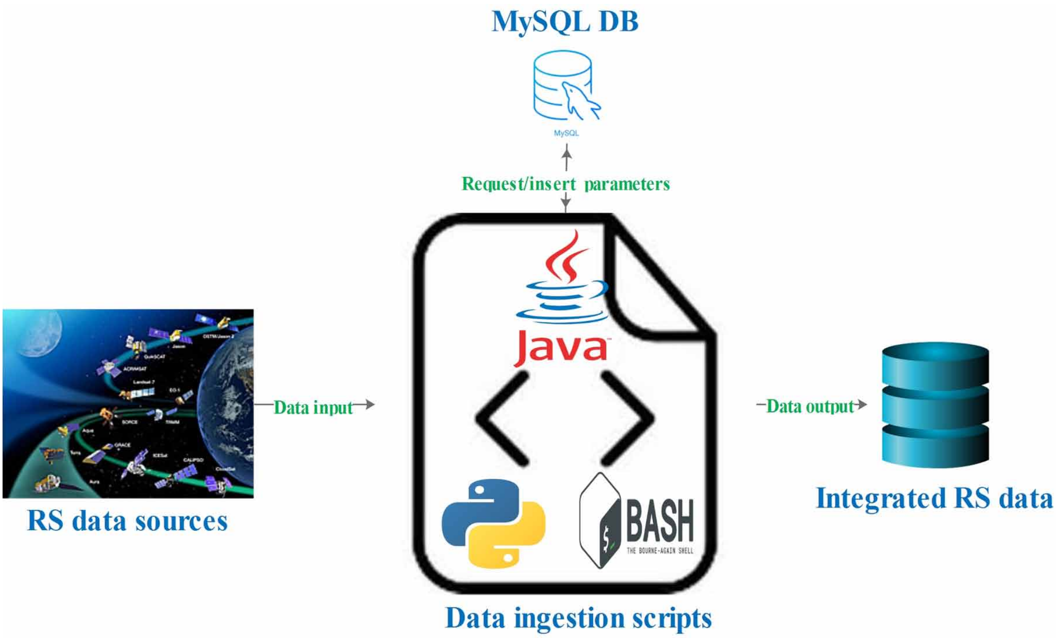
Figure 14. The proposed BD architecture



8.2. Ingestion Layer

This part of BD architectures is very important since it involves connection to plenty data sources, filters, subsets, extracts and removes noised and inaccurate datasets (Erraissi, Belangour, & Tragha, 2018). From Figure 15, we notice that the main input of the ingestion layer are the unprocessed satellite data. The output contains the refined and accurate datasets to be integrated and stored and processed. In addition, the ingestion layer communicates with a MySQL database including all required information and database of the acquired data. Our ingestion layer decompresses and filters RSBD of the selected countries using their satellite orbits cross times, and bounding rectangle of longitude and latitude. Then, many environmental and pollutants variables such as temperature, humidity and the concentration of several trace gazes particularly CO, $CO_{2,}$ $NO_{x,}$ CH_4 are extracted from plenty converted files. The final step in this layer is to subset and remove inaccurate values before to be stored inside distributed files system especially the Hadoop HDFS, Amazon S3 or GlusterFS. This ingestion layer has been designed for RSBD preprocessing, it is able to handle enormous input of

Figure 15. The input and output of the RS ingestion layer



data, extract the meaningful information from satellite data. This ingestion layer has been developed by many separated and cooperative scripts. Java, Python and BASH are the principal programming languages used in this investigation. The Bash was exploited to connect automatically, to download RSBD from the various sources and to manage the big number of files. Python scripts are mainly used to extract, explore and write the final output datasets.

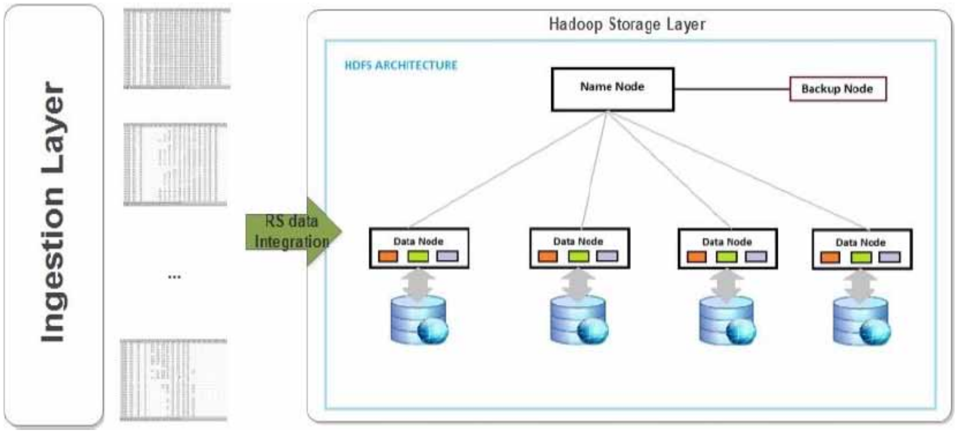
8.3. Storage Layer

This layer is dedicated to store BD inside distributed file system among a lot of clusters. In our study, we are looking forward to storing the integrated data occurred from the output of the ingestion layer to a Hadoop Distributed File System (HDFS) because it is one of the best tools at now that enables holding huge volume of data, provides easier access and allows data redundancy to prevent data loses in case of failure or damage. Furthermore, the HDFS facilitates parallel data processing main master/slave is topology (Wang et al., 2015). Figure 16 shows the general architecture of the HDFS that will be used to store the preprocessed satellite data. The cluster contains a Name node as a Master that keeps the directory tree of all files in the file system, and tracks where across the cluster the file data is kept. And a Data node as a slave storing data. And a Backup node as the name states its main role is to act as the dynamic backup for the Filesystem Namespace (Metadata) in the Primary Name node of the Hadoop Ecosystem.

8.4. Processing and Query Layer

This layer includes all necessary tools for streaming and batch processing. In our case, for instance, we can site: MapReduce, Storm, and Spark are the key elements in the Hadoop management layer. These tools are rapid, scalable, reliable and easy to operate. Thus, Storm and Spark are commonly used in stream processing (Erraissi, 2017a). We will use too the MapReduce because it is very suitable for batch processing such as RSBD. The processing layer contains also query languages such as: Pig, Hive, Sqoop and so on, in order to access unstructured, semi-structured and structured data. In

Figure 16. The general architecture of the HDFS



our research, this layer will contain some novel algorithms helping in data optimization, validation and interpolation. Furthermore, it will contain a prediction software based in Artificial Intelligent (AI) algorithms helping in decision-making concerning the AQ and climate changes issues. Figure 17 illustrates the general architecture of the MapReduce formwork that will be exploited to process the stored data. A MapReduce program is composed of a map procedure which performs filtering and sorting (such as sorting the temperature by latitude, longitude and time into queues), and a reduce function, which performs a summary operation (such as calculating the average temperature in each queue). In our research MapReduce will be used mainly to enhance the quality of satellite data, interpolating, fusing and validating remote sensing data and finally applying some AI model to predict climate changes and pollution data.

8.5. Visualization Layer

The final step of RSBD processing is the visualization, it helps to show the final results into charts, maps that help analysts and scientists understanding faster and clearly and deducing potential insights. There are several tools dedicated to the data visualization, particularly, the custom dashboard, Kibana.

Figures 18 and 19 show respectively the Vertical Mixing Ration (VMR) with the Part per Million or Billion unit (PPMv-PPBv) of the Ozone (O3), Carbon Monoxide (CO) of Morocco in 11/06/2018. The altitude of this measurement is between 0-200 meters. These maps are an example of the visualization tools showing the low tropospheric air pollution of Morocco helping in the decision

Figure 17. The general architecture of the MapReduce

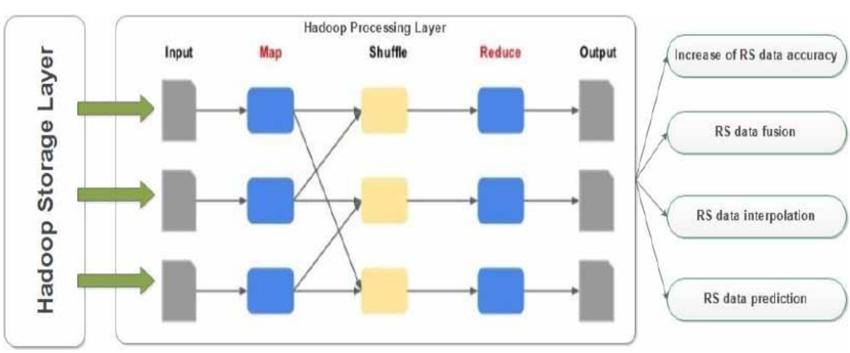
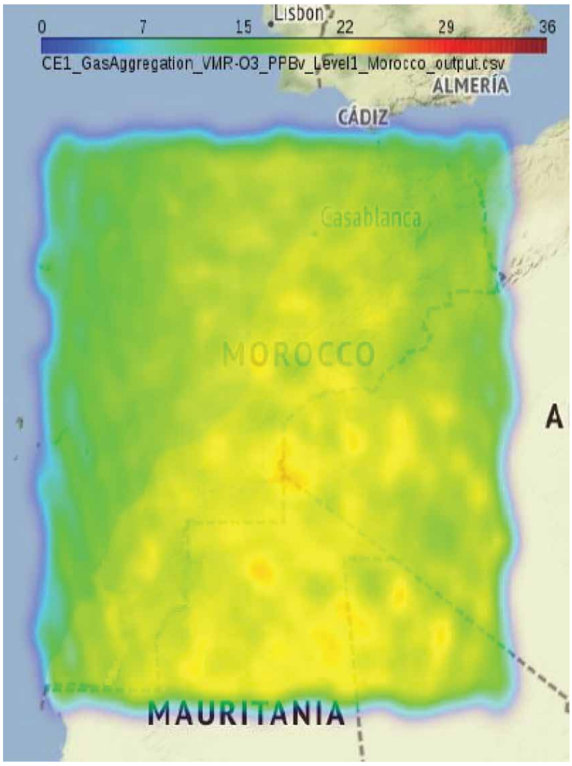


Figure 18. VMR of O_3 (PPBv)



makers. We notice that the density of the trace gases is significant near industrial zones located in Casablanca, Safi and Tangier. More over the concentration of the CO and O3 are high in the coastal areas near ports due to the high activities of the maritime transport activities.

8.6. Monitoring Layer

BD platform is composed with plenty frameworks, tools and physical materials that must be managed and well configured, this is why, and it exists several monitoring systems for each BD distribution like: Ambari for Hortonworks data platform, Cloudera manager for Cloudera distribution and Web Console for IBM Infosphere Big Insights. This monitoring system increase surely the performance of BD platform.

9. COMPARISON

This section includes a brief comparison in term of the used RS data, processing and applications between our study and some others. From Table 9 we notice that all the aforementioned studies cited in the related work is section (see Table 1) process RS data for an environmental application, particularly: air pollution monitoring, forest fire detection, climate changes supervision and earthquakes prediction. Thus, our investigation aims to apply RS techniques in all these topics.

We remark also that all of these studies acquire data from one satellite or more. However, in this paper we acquire RS data from eleven satellites and sixteen sensors. As a result, our input of satellite data is very consistent with less gaps. We notice that the studies 1-2-4 use a single machine for processing making an important execution time. However, in this study, we have

Figure 19. VMR of CO (PPMv)

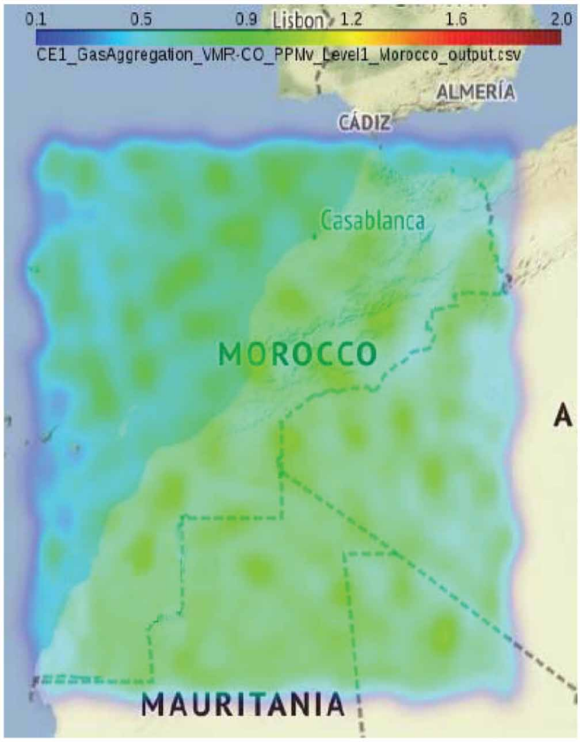


Table 9. Comparison among related works

Study	RS Data					RS Data Processing		RS Application
	Sources	Satellites	Sensors	Size/Day	Velocity/Day	Architecture	Tools	
This study	4	11	16	> 37 Gb	> 36000 Files	Single & distributed (Hadoop)	Java Python Bash	Air pollution Forest fire Climate changes
1	1	1	1	~ 566 Mb	~ 2000 Files	Single	Python	Air Pollution
2	1	1	1	~ 10 Mb	~ 8 Files	Single	AEPA	Earthquake prediction
3	1	1	1	~ 500 Mb	~ 5 Files	Distributed (Hadoop)	Pig	Climate changes
4	2	2	4	> 1 Gb	> 2200 Files	Single	Java	Air pollution

developed a distributed architecture based in the Hadoop paradigm. Accordingly, the processing is scalable and optimized.

10. CONCLUSION

During this last decade, air pollution and climate changes have been two phenomenon that affect environment is safety and human is health. This is due to the emission of pollutant gases particularly, CO, $CO_{2,}$ NO_x and so on, from industrials, transport and agriculture activities. Thus, the continuous monitoring of atmospheric composition in NRT become highly significant. The key solution is to employ RS techniques that provide a global scale satellites data in NRT. Our research aims to monitor abnormal climate changes and supervise AQ especially in Morocco. We collected RSBD in NRT from six sources which are the MDEO ground station of EUMETSAT data, the EOSDIS data of NASA, the NESDIS data of NOAA, the Copernicus platform, some MGS data and the Raspberry PI sensors data. The handicap is that these datasets have not only a huge volume and velocity but also come with different file formats since they cover a global scale, and they come from different satellites sensors with a wide spectral resolution. Accordingly, we have proved that RS data are BD according to the four salient: volume, variety, velocity and veracity. In our case, RSBD are heavy in term of size reaching one hundred GB per day, stored in different file is extensions including (BIN, BUFR, NetCDF, HDF5, etc.) and have a high velocity averaged within eleven thousand daily files. There is no doubt that the existing systems and architectures are so limited to handle the NRT RSBD. As a result, we have adopted the Hadoop architecture to RSBD that processes efficiently this kind of data. This architecture is composed of six main layers as follow: the data sources, data ingestion, data storage, data processing, data visualization and the monitoring layer. The aforementioned architecture automatically collects, filters, extracts and stores data into HDFS. Moreover, the processing layer will include some AI algorithms increasing data accuracy. Finally, the visualization layer will show data in NRT into interactives maps and charts helping in decision makers. This architecture should be supplied with another cloud computing architecture that may optimize the time execution. This task would be an interesting work to be conducted in the future.

ACKNOWLEDGMENT

The authors of this work are thankful to the CNRST (Centre National pour la Recherche Scientifique et Technique) of Morocco for the financial support provided within the PPR1 project.

REFERENCES

Arya, K. V., Bhadoria, R. S., & Chaudhari, N. S. (Eds.). (2018). Emerging Wireless Communication and Network Technologies: Principle, Paradigm and Performance. Springer. doi:10.1007/978-981-13-0396-8

Badr-eddine, B. S., & El Amrani, C. (2019). Towards Remote Sensing Datasets Collection and Processing. In A. Hameurlain, R. Wagner, & T. K. Dang (Eds.), *Transactions on Large-Scale Data- and Knowledge-Centered Systems XLI* (pp. 286–294). doi:10.1007/978-3-030-11196-0_26

Bhadoria, R. S., & Chaudhari, N. S. (2019). Pragmatic Sensory Data Semantics With Service-Oriented Computing. *Journal of Organizational and End User Computing*, *31*(2), 22–36. doi:10.4018/JOEUC.2019040102

Breame, F. (2018). BUFRextract User Guide v30.

BUFR Descriptors for IASI Level 2 Data. (2017).

Chijioke, G. E. (2012). Satellite remote sensing technology in spatial modeling process: Technique and procedures. *International Journal of Science and Technology*, *2*(5), 309–315.

Clerbaux, C., Boynard, A., Clarisse, L., George, M., Hadji-Lazaro, J., Herbin, H., & Wespes, C. et al. (2009). Monitoring of atmospheric composition using the thermal infrared IASI/MetOp sounder. *Atmospheric Chemistry and Physics, 9(16), 6041-6054.*

Duncan, B. N., Prados, A. I., Lamsal, L. N., Liu, Y., Streets, D. G., Gupta, P., & Ziemba, L. D. et al. (2014). Satellite data of atmospheric pollution for U.S. air quality applications: Examples of applications, summary of data end-user resources, answers to FAQs, and common mistakes to avoid. *Atmospheric Environment*, *94*, 647–662. doi:10.1016/j.atmosenv.2014.05.061

El Amrani, C. (2015). Remote Sensing for Real-time Early Warning of Environmental Disasters and WRF Modelling. *e-AGE 2015*, 11.

El Amrani, C., Rochon, G. L., El-Ghazawi, T., Altay, G., & Rachidi, T. (2012, July). Development of a real-time urban remote sensing initiative in the Mediterranean region for early warning and mitigation of disasters. *Proceedings of the 2012 IEEE International Geoscience and Remote Sensing Symposium* (pp. 2782-2785). IEEE. doi:10.1109/IGARSS.2012.6350855

El Amrani, C., Rochon, G. L., El-Ghazawi, T., Altay, G., & Rachidi, T. (2013, July). System architecture of the Mediterranean Dialogue Earth Observatory. *Proceedings of the 2013 IEEE International Geoscience and Remote Sensing Symposium-IGARSS* (pp. 600-603). IEEE. doi:10.1109/IGARSS.2013.6721228

Wikipedia. (n.d.). EOSDIS. Retrieved from https://en.wikipedia.org/w/index.php?title=EOSDIS&oldid=808578344

Erraissi, A. (2017b). Digging into Hadoop-based Big Data Architectures. *International Journal of Computer Science Issues*, *14*(6), 52–59. doi:10.20943/01201706.5259

Erraissi, A., Belangour, A., & Tragha, A. (2017a). A Comparative Study of Hadoop-based Big Data Architectures. *IJWA*, *9*(4), 129–137.

Erraissi, A., Belangour, A., & Tragha, A. (2018). Meta-Modeling of Data Sources and Ingestion Big Data Layers. SSRN Electronic Journal. doi:10.2139/ssrn.3185342

Garcia-de-Prado, A., Ortiz, G., Boubeta-Puig, J., & Corral-Plaza, D. (2018). Air4People: A Smart Air Quality Monitoring and Context-Aware Notification System. *Journal of Universal Computer Science*, *24*(7), 846–863.

Gupta, S. (n.d.). Active and Passive Remote Sensing.

Hassini, A., & Belbachir, A. H. (2016). Thermal infrared geostationary satellite sensor data application for prediction and monitoring earthquake in Algeria. *International Journal of Information Technology and Management*, *15*(4), 293–312.

Holnicki, P., Kałuszko, A., & Trapp, W. (2016). An urban scale application and validation of the CALPUFF model. *Atmospheric Pollution Research*, *7*(3), 393–402. doi:10.1016/j.apr.2015.10.016

Hu, H., Wen, Y., Chua, T.-S., & Li, X. (2014). Toward Scalable Systems for Big Data Analytics: A Technology Tutorial. *IEEE Access*, *2*, 652–687. doi:10.1109/ACCESS.2014.2332453

Li, Y., Ding, Z., & Long, T. (2015). Improved SPECAN algorithm based on Doppler rate equalization. *Proceedings of the IET International Radar Conference 2015*. Academic press. doi:10.1049/cp.2015.1443

Ma, Y., Wu, H., Wang, L., Huang, B., Ranjan, R., Zomaya, A., & Jie, W. (2015). Remote sensing big data computing: Challenges and opportunities. *Future Generation Computer Systems*, *51*, 47–60. doi:10.1016/j.future.2014.10.029

NOAA National Environmental Satellite, Data, and Information Service (NESDIS). Retrieved from https://www.nesdis.noaa.gov/

Oussous, A., Benjelloun, F.-Z., Ait Lahcen, A., & Belfkih, S. (2017). Big Data technologies: A survey. *Journal of King Saud University - Computer and Information Sciences*. doi:<ALIGNMENT.qj></ALIGNMENT>10.1016/j.jksuci.2017.06.001

Parkinson, C. L., Ward, A., & King, M. D. (2006). *Earth science reference handbook: a guide to NASA's earth science program and earth observing satellite missions*. National Aeronautics and Space Administration.

Perera, C., Zaslavsky, A., Christen, P., & Georgakopoulos, D. (2012). CA4IOT: Context Awareness for Internet of Things. *Proceedings of the 2012 IEEE International Conference on Green Computing and Communications* (pp. 775–782). IEEE Press. doi:10.1109/GreenCom.2012.128

Rathore, M. M. U., Paul, A., Ahmad, A., Chen, B.-W., Huang, B., & Ji, W. (2015). Real-Time Big Data Analytical Architecture for Remote Sensing Application. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *8*(10), 4610–4621. doi:10.1109/JSTARS.2015.2424683

Rosemary, M. (2016). GOME-2 ON METOP.

Russom, P. (2011). Big data analytics. *TDWI best practices report, 19*(4), 1-34.

Schmetz, J., Pili, P., Tjemkes, S., Just, D., Kerkmann, J., Rota, S., & Ratier, A. (2002). An Introduction to Meteosat Second Generation (MSG). *Bulletin of the American Meteorological Society*, *83*(7), 977–992. doi:10.1175/BAMS-83-7-Schmetz-2

Semlali, B. B., & Chaker, E. A. (2019). Towards Remote Sensing Datasets Collection and Processing. *International Journal of Embedded and Real-Time Communication Systems*, *10*(3), 49–67. doi:10.4018/IJERTCS.2019070104

Semlali, B. R. B., El Amrani, C., & Denys, S. (2019). Development of a Java-based application for environmental remote sensing data processing. *Iranian Journal of Electrical and Computer Engineering*, *9*(3), 1978.

Showstack, R. (2014). Sentinel Satellites Initiate New Era in Earth Observation. *Eos (Washington, D.C.)*, *95*(26), 239–240. doi:10.1002/2014EO260003

Siemen, S., Lamy-Thepaut, S., Li, F., & Russell, I. (2007). *The next generation of ECMWF's meteorological graphics library – Magics*. doi:10.21957/f0s1fwajmj

Smith, K. R., Bruce, N., Balakrishnan, K., Adair-Rohani, H., Balmes, J., Chafe, Z., & Rehfuess, E. et al. (2014). Millions Dead: How Do We Know and What Does It Mean? Methods Used in the Comparative Risk Assessment of Household Air Pollution. *Annual Review of Public Health*, *35*(1), 185–206. doi:10.1146/annurev-publhealth-032013-182356 PMID:24641558

Sun, X., Liu, P., Ma, Y., Liu, D., & Sun, Y. (2016). Streaming Remote Sensing Data Processing for the Future Smart Cities: State of the Art and Future Challenges. *International Journal of Distributed Systems and Technologies*, *7*(1), 1–14. doi:10.4018/IJDST.2016010101

Wang, C., Hu, F., Hu, X., Zhao, S., Wen, W., & Yang, C. (2015). A hadoop-based distributed framework for efficient managing and processing big remote sensing images. ISPRS Annals of Photogrammetry. *Remote Sensing and Spatial Information Sciences*, *II-4*(W2), 63–66. doi:10.5194/isprsannals-II-4-W2-63-2015

Wilkin, J. (2010). Lectures 1 and 2: Satellite orbits and Measurement Geometry. Rutgers.

Yakubailik, O., Romas'ko, V., & Pavlichenko, E. (2019). Complex for reception and real time processing of remote sensing data. *E3S Web of Conferences*. Academic Press. doi:10.1051/e3sconf/20197503003

Zanella, A., Bui, N., Castellani, A., Vangelista, L., & Zorzi, M. (2014). Internet of Things for Smart Cities. *IEEE Internet of Things Journal*, *1*(1), 22–32. doi:10.1109/JIOT.2014.2306328

Zaytar, M. A., & Amrani, C. E. (2018). MetOp Satellites Data Processing for Air Pollution Monitoring in Morocco. *International Journal of Electrical & Computer Engineering, 8*.

Badr-Eddine Boudriki Semlali was born in Tangier, Morocco. He obtained his BSc degrees (2015) from Faculty Sciences and Techniques of Tangier, department of computer engineering at Abdelmalek Essaadi University. Then he got a Master (2017) degree in computer system and network engineering from the same faculty. Currently, he is a researcher in big data analytics of a remotely sensing earth observatory. Specialized in remote sensing big data collecting, storing and processing, using big data technologies, and cloud computing techniques.

Chaker El Amrani is a Doctor in Mathematical Modelling and Numerical Simulation from the University of Liège, Belgium (2001). He worked at Belgocontrol in Brussels as Air Traffic Control Engineer. Then he joined Abdelmalek Essaadi University, where he is currently Full Professor. He is the NATO Partner Country Project Director of a real-time remote sensing initiative for early warning and mitigation of disasters and epidemics in Morocco. He served as Chair of the Computer Engineering Department from 2009 to 2014. He lectures distributed systems and is promoting High Performance Computing education at the University. He is an IEEE Senior member, and a Fulbright Visiting Scholar at the George Washington University. He has been a Visiting Professor at the University of Graz, the University of Bristol and the TU Dortmund University. His research interests include Cloud Computing, and environmental information systems.

Guadalupe Ortiz received her PhD degree in computer science from the University of Extremadura, Spain, in 2007. From 2001 to 2009, she was an Assistant Professor and a Researcher with the Computer Science Department, University of Extremadura. In 2009, she joined the Department of Computer Science and Engineering, University of Cadiz, as an Associate Professor. She has authored or co-authored numerous peer-reviewed papers in international journals, workshops, and conferences. Her research interests embrace service context awareness and their adaptation to mobile devices, and complex event processing integration in service-oriented and micro-service architectures and the Internet of Things.