

Chapter 53

Big Data for Satellite Image Processing: Analytics, Tools, Modeling, and Challenges

Remya S.

VIT University, India

Ramasubbareddy Somula

VIT University, India

Sravani Nalluri

VIT University, India

Vaishali R.

VIT University, India

Sasikala R.

VIT University, India

ABSTRACT

This chapter presents an introduction to the basics in big data including architecture, modeling, and the tools used. Big data is a term that is used for serving the high volume of data that can be used as an alternative to RDBMS and the other analytical technologies such as OLAP. For every application there exist databases that contain the essential information. But the sizes of the databases vary in different applications and we need to store, extract, and modify these databases. In order to make it useful, we have to deal with it efficiently. This is the place that big data plays an important role. Big data exceeds the processing and the overall capacity of other traditional databases. In this chapter, the basic architecture, tools, modeling, and challenges are presented in each section.

DOI: 10.4018/978-1-6684-3662-2.ch053

1. INTRODUCTION

Day by day, we see the data is rapidly increasing in many forms. We have some traditional data processing software to process small quantity of data. But as trillions of bytes of information is being processed per second, the traditional software techniques fail in processing this data. We need to re-think of a solution which can process this data. Now Big Data gives us a solution. Big Data is a term used for creating, capturing, communicating, aggregating, storing and analyzing large amounts of data. Many attempts encountered to quantify the growth rate in the volume of data is called as Information Explosion.

Major milestones took place in the history of sizing data volumes plus the evolution of the term Big Data. The following are some of them:

- In 1971, Arthur Miller stated in “The Assault on Privacy” that:

Too many information handlers seem to measure a man by the number of bits of storage capacity his dossier will occupy.

- In April 1980, I.A.Tjomsland gave a talk titled “Where Do We Go From Here?” at “Fourth IEEE Symposium on Mass Storage Systems” in which he says:

Data expands to fill the space available, I believe that large amounts of data are being retained because users have no way of identifying obsolete data, the penalties for storing obsolete data are less apparent than are the penalties for discarding potentially useful data.

- In 1997, Michael Lesk publishes “How much information is there in this world?” in which he concludes that:

There may be a few thousand petabytes of information all told, and the production of tape and disk will reach that level by the year 2000. So in only a few years, (a) we will be able to save everything- no information will have to be thrown out, and (b) the typical piece of information will never be looked at by a human being. (<https://www.forbes.com/sites/gilpress/2013/05/09/a-very-short-history-of-big-data/2/#1c3097c24343>).

The term Big Data was coined in 1998 by Mr. John Mashey, Chief Scientist at SGI. Even though Michael Cox and David Ellsworth seem to have used the term ‘Big Data’ in print, Mr. Mashey supposedly used the term in his various speeches and that’s why he is credited from coming up with Big Data. But some various sources say that the first use of the term Big Data was done in an academic paper- Visually Exploring Gigabyte Datasets in Realtime(ACM) (OECD, 2015; Mark A. Beyer & Douglas Laney, 2012).

The following are the differentiators of Big Data over Traditional Business Intelligence solutions:

- Data is retained in a distributed file system instead of on a central server.
- The processing functions are taken to the data rather than data being taken to the functions.
- Data is of different formats, both structured as well as unstructured.
- Data is both real-time as well as offline data.
- Technology relies on massively parallel processing(MPP) concepts.

13 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/big-data-for-satellite-image-processing/291028

Related Content

Improvement in Task Scheduling Capabilities for SaaS Cloud Deployments Using Intelligent Schedulers

Supriya Sawwashere (2021). *International Journal of Big Data and Analytics in Healthcare* (pp. 1-12).
www.irma-international.org/article/improvement-in-task-scheduling-capabilities-for-saas-cloud-deployments-using-intelligent-schedulers/287104

Developing a Method to Valuate the Collection of Big Data

Colleen Carraher Wolverton, Brandi N. Guidry Hollier, Michael W. Totaro and Lise Anne D. Slatten (2022).
Research Anthology on Big Data Analytics, Architectures, and Applications (pp. 188-196).
www.irma-international.org/chapter/developing-a-method-to-valuate-the-collection-of-big-data/290983

Bug Handling in Service Sector Software

Anjali Goyal and Neetu Sardana (2017). *Applying Predictive Analytics Within the Service Sector* (pp. 54-73).
www.irma-international.org/chapter/bug-handling-in-service-sector-software/177316

Data Envelopment Analysis for Improving the Microgrid Operations

Uetutiza C. C. Kuzatjike and K. S. Sastry Musti (2024). *Data Envelopment Analysis (DEA) Methods for Maximizing Efficiency* (pp. 184-206).
www.irma-international.org/chapter/data-envelopment-analysis-for-improving-the-microgrid-operations/336945

Data Mining Problems Classification and Techniques

Nayem Rahman (2018). *International Journal of Big Data and Analytics in Healthcare* (pp. 38-57).
www.irma-international.org/article/data-mining-problems-classification-and-techniques/209740