**Big Data**

Wikipedia defines[[1]](#endnote-1) Big Data as “an all-encompassing term for any collection of data sets so large and complex that it becomes difficult to process using traditional data processing applications.”

Big Data has often been looked at as a problem but it should be looked at as an opportunity. It enables data systems to process, archive, manage, and manipulate large amounts of disparate data.

The main characterization of "Big Data" is described by "Volume(How much data)", " Velocity(Speed of data processing)" and " Variety (Various types of data)". But the definition of “Big Data” varies depending on the data system’s capabilities and the capabilities of the services running at the data system manage the data. What is “Big” to some, may be small to others and what is considered "Big" today will not be in the future. And nowadays, due to the increase of its volume, Earth Observation data is also considered as "Big Data".

Big Data is NOT a single technology. It refers to multiple technologies and initiatives that involve large complex data sets and infrastructures. The real issue is not that agencies are now archiving and processing large amounts of data. It's what you do with the data that counts. Big Data, when harnessed properly, allows real-time analysis and data mining to produce better science.

Cloud Computing enables Big Data processing for data systems by relieving a number of complex scenarios that Big Data can introduce.

**Cloud Computing**

Cloud Computing is an emerging information technology and computing architecture that seeks economies of scale for storage and processing based on the incremental use of computing resources.

Cloud Computing Types[[2]](#endnote-2) and Services[[3]](#endnote-3) are categorized as shown in Table 1 and Table2.

Table 1 : Cloud Computing Types

|  |  |
| --- | --- |
| Types | Description |
| Private | Typically deployed within an organization's own internal ecosystem, often leveraging the organization's own private datacenter. |
| Public | Hosted by a third party datacenter located off premise at multiple locations outside of an organization's building. Public clouds are often hosted on virtualized multi-tenancy datacenters where different organizations have access to shared pooled hardware and power resources, yet can run their applications and data in secure, isolated environments. |
| Hybrid | A combination of using some services delivered via a private cloud internally and other services delivered via a public cloud externally. |

Table 2: Cloud Computing Services

|  |  |
| --- | --- |
| Service Category | Description |
| Infrastructure  as a Service  (IaaS) | The most basic cloud-service model, which provides the user with virtual infrastructure, for example servers and data storage space. Virtualization plays a major role in this mode, by allowing IaaS-cloud providers to supply resources on-demand extracting them from their large pools installed in data centers. |
| Platform  as a Service  (PaaS) | Cloud providers deliver to the user development environment services where the user can develop and run in-house built applications. The services might include an operating system, a programming language execution environment, databases and web servers. |
| Software  as a Service  (SaaS) | The cloud provides the user with access to already developer applications that are running in the cloud. The access is achieved by cloud clients and the cloud users do not manage the infrastructure where the application resides, eliminating with this the way the need to install and run the application on the cloud user’s own computers. |
| Network  as a Service  (Naas) | The least common model, where the user is provided with network connectivity services, such as VPN and bandwidth on demand. |

Cloud Computing has several Pros and Cons and some examples[[4]](#endnote-4) are shown in Table3.

|  |  |
| --- | --- |
| Pros | Cons |
| * Cost Efficiency * Convenience and continuous availability * Backup and Recovery * Environmentally friendly * Resiliency and Redundancy * Scalability and Performance * Quick deployment and ease of integration * Increased Storage Capacity * Device Diversity and Location Independence * Smaller learning curve | * Cost Efficiency * Security and privacy * Dependency and vendor lock-in * Technical Difficulties and Downtime * Limited control and flexibility * Increased Vulnerability |

**In to the Future**

Currently, the massive growth in both the variety and volume of Earth Observation data presents a number of challenges. For data providers or data management section, these may include hardware (e.g. storage systems, processing systems) scalability, software capability and/or timely data service. For data users, these may include data discoverability and accessibility.

These new technologies, Big Data and Cloud Computing, can support providing solutions for these challenges and contribute to create various applications which will solve social issues more steadily.

At CEOS/WGISS (Working Group on Information Systems and Services), experiences and best practices of these technologies are actively discussed and shared by members. For more details, please visit the WGISS Technology Exploration Interest Group page [http://ceos.org/ourwork/workinggroups/wgiss/interest-groups/technology-exploration/]

1. http://en.wikipedia.org/wiki/Big\_data [↑](#endnote-ref-1)
2. http://www.synergygs.com/Solutions/CloudServices/ [↑](#endnote-ref-2)
3. http://www.synergygs.com/Solutions/CloudServices/http://www.synergygs.com/Solutions/CloudServices/http://www.javacodegeeks.com/2013/04/advantages-and-disadvantages-of-cloud-computing-cloud-computing-pros-and-cons.html [↑](#endnote-ref-3)
4. http://www.javacodegeeks.com/2013/04/advantages-and-disadvantages-of-cloud-computing-cloud-computing-pros-and-cons.html [↑](#endnote-ref-4)