Geophysical Research Abstracts
Vol. 13, EGU2011-11506-1, 2011
EGU General Assembly 2011

Distributing Data to Scientific Communities

Daniel Higuero, Juan M. Tirado, and Jesus Carretero
University Carlos III, Spain

Different institutions around the world face the problem of distributing data to their users. In particular, scientific institutions such as ESA, NASA or CERN need to distribute vast amounts of data produced by different missions, satellites or projects to the interested scientists. This is a complex problem, as it includes different geographically distributed archives, heterogeneous architectures, variable volume of requests, different user profiles and even SLA between different partners.

We propose a data distribution architecture (HIDDRA) that alleviates this problem aiming to reduce administration overhead and user intervention in the data acquisition process. HIDDRA is a modular system based on three main components. First, a highly efficient parallel multiprotocol download engine to be installed at user sites, provides access using different common used protocols (HTTP, HTTPS or GridFTP) to the data archives and performs the required transfers. Using different protocols and parallel connections to different source servers reduces workload and hotspots in data servers and improves bandwidth utilization in user sites, thus reducing the time to receive the desired data. A second component manages the user subscriptions, starting transparent downloads of their subscribed products. This link between the users and the data provider organization eliminates the need for the users to check for new products that may be of their interest. Finally, the third component tries to dynamically optimize the size of distribution infrastructure in order to reduce power consumption and cost of ownership by introducing cloud-based solutions.