Towards exascale distributed data management

Giovanni Aloisio and Sandro Fiore
Euro-Mediterranean Centre for Climate Change - CMCC

Abstract

"Exascale eScience infrastructures" will face important and critical challenges both from computational and data perspectives. Increasingly complex and parallel scientific codes will lead to the production of huge amount of data. The large volume of data and the time needed to locate, access, analyze and visualize data will greatly impact on the scientific productivity of scientists and researchers in several domains. Significant improvements in the data management field will increase research productivity in solving complex scientific problems. Next-generation eScience infrastructures will start from the assumption that exascale HPC applications (running on million of cores) will generate data at a very high rate (terabytes/s). Hundreds of exabytes of data (distributed across several centres) are expected by 2020, to be available through heterogeneous storage resources for access, analysis, post-processing and other scientific activities.

Main data challenges at exascale

The design of Exascale data infrastructures for eScience must take into account several challenges needing special attention and new solutions at such large scale. In the following, some of the most relevant ones will be presented.

Distribution: the exascale scenario will involve a lot of distributed data available at an international level across several countries (data coming from simulations, observations, experiments, etc.). Collections of data will be stored at different sites and made available to the users for further analysis and studies. Aggregation capabilities as well as parallel access to data will help "client-side" applications in several eScience domains (i.e. climate change) in getting the needed data from distributed storages.

Replication: replication of data represents a key point to increase data locality, availability and fault tolerance. In the exascale scenario data replication must be able to deal with several kind of transient failures (i.e. network and storage failures) recovering from them. Replication of data will need further investigation on new algorithms, protocols, replication schemas, placement strategies is strongly needed. It will be necessary to manage consistency issues between replica catalogs and storage devices as well as replica lifetime issues. Advanced monitoring system will allow checking the status of each replica in terms of availability, network latency/bandwidth, usage, etc. New replicas should be automatically created based on user's needs and historical access information.

Access to the data environment: the involved components into an exascale environment (storages, metadata services, registries, analysis tools, ontologies, etc.) should interoperate and cooperate seamlessly. Unified access interfaces should hide the heterogeneity of the underlying systems and simplify the access to the available resources. Besides, the access to the underlying infrastructure should be secure, pervasive and ubiquitous. From a "user interface" point of view, scientific gateway solutions will help users in easily finding, accessing and integrating (cross-domain) data, managing metadata, exploiting visualization and analysis tools, etc. Integrated web-based environments will play a key role in changing and improving the daily activities of scientific users. Moreover, advanced workflow management capabilities will allow researchers composing complex scientific activities working on distributed storages, computational resources, databases, etc. Next generation data portals will be characterised by a high level of collaborative functionalities in the daily working activity of researchers and scientists. Social networking capabilities will increase the

level of discussions, feedback, exchange of scientific results and dissemination among groups, scientific teams, etc.
Finally, scientific gateways should be wider in “target” and easier in “access” because they should be designed and conceived not only for scientists (the targeted primary users are domain experts), but also for non-expert users (i.e. decision makers).

Storage and caching: storage technology is a vital part to manage large volume of data. Both hardware and software components will play a critical role in managing big scientific datasets. Parallel I/O will be crucial for several reasons: data generation rate (critical for storages when parallel applications are running on millions of cores as at exascale) and data extraction (carried out by analysis tools) are just two examples. New efficient and scalable caching algoritms will be needed to increase performance in managing data.

Scientific Discovery of data: in a exascale environment search & discovery of data will continue to play a critical role. Several metadata sources will be in the scene contributing to describe the available data collections. Metadata hierarchies and indexes will help in speeding up search and discovery phases at such large scale. It will be fundamental to query, combine, retrieve and filter metadata information producing query results in a fast, scalable and efficient way. Efficient harvesting protocols will gather metadata information from different sites, indexing the available distributed collections. Metadata standards and the associated standardization processes will be fundamental to describe data through widely accepted, known and adopted set of information. Automatic extraction of metadata will play a key role at exascale in easing and speeding up the ingestion metadata process. Provenance information will increasingly become more important to identify, tracing and recording the history of a data, the related processing and analysis steps and so forth.

Data analysis: at exascale, data analysis will need new mathematical approaches, algorithms and related parallel implementations able to scale with the high number of available processors as well as to efficiently access and manage data at the file system and storage level. To avoid huge data transfer across the network, several computations should be carried out close to the storage devices, that is inside the Supercomputing/HPC Centres. Avoiding huge data movement moving the analysis to the storage resources will be a relevant point of active storage processing studies.

Client side tools: on the client side, new tools should take advantage of today available desktop capabilities. This implies the need to have parallel implementations of tools to access, visualize and analyze data. Parallel applications could provide a high level of responsiveness highly needed in the future scenarios. Since it is impractical to download large volume of data on the client side, it will be fundamental to increase client-side data reduction capabilities to extract the needed and relevant (from the user point of view) subset of data.

Cloud-based data services: storages, metadata and scientific domain ontologies could be managed through cloud-based services able to offer a high level of reliability, scalability (via dynamic provision of resources) and security. Hosting of metadata and data on cloud services will increase the availability of both data and access services. For reliable serving of data, the cloud should be cheaper than traditional storage resources hosted at the HPC Centres and could both provides additional storage resources and represents an alternative way to host/manage data.

Interoperability: due to the large scale environment, the heterogeneity of the platforms and the complexity of the exascale system, interoperability plays an important role to make the interaction among all of the involved components, services and actors feasible and productive. Interoperability can be achieved through strong standards adoption. On the other side standard processes must be strongly encouraged to address real needs, leading to lighweight, effective and widely agreed documents. Interoperability makes really “open” an exascale environment. Standards should address among the others metadata issues, data protocols and storage interfaces. A notable example in the area of storage management is the Storage Resource Manager (SRM). The interoperability among multiple implementations of the Storage Resource Manager interface will improve the data access, movement, replication among several (heterogeneous) storages working at different and geographically spread sites.

Education and training: at exascale, strong expertise is needed to improve the capabilities of the overall system at several levels. Advanced courses (including training sessions) at the Universities and collaboration/master opportunities at HPC Centres would speed up the formation process of “exascale scientists and researchers”.