Earth System Grid Federation
On-Demand IDX Data Server

Cameron Christensen - Giorgio Scorzelli - Valerio Pascucci - Peer-Timo Bremer

The on-demand IDX data server is designed to provide streaming hierarchical versions of equivalent NetCDF climate data volumes in a user-directed manner such that specific timestep fields are converted just-in-time. This permits the bulk of the data to remain on the server and facilitates interactive analysis and visualization by immediately sending results for specific data requests. Initial conversions are cached for future use, amortizing the cost across future requests.

The IDX data format utilizes an hierarchical z-order to facilitate fast loading of coarse resolution data as well as better spatial locality for more efficient sub-region reads. Quickly streaming coarse resolution data provides a preview of the final data and facilitates interactivity.

Figure 1 shows the design of the overall system. Beginning at the ESGF search page (top-left of the diagram), the user can download a ViSUS configuration file describing how to load the selected climate dataset. When the user selects a dataset its corresponding IDX metadata is created and registered with an associated ViSUS server. The data can be loaded in a ViSUS client or compatible component.

When a request for data is received by the ViSUS Server, the server first checks to see if the data is available in the local cache. If so, it sends the cached data immediately. Otherwise, the server issues a request to convert the smallest portion of data that contains the region of the desired field at the given timestep. This data is read from the ESGF archive and written to the local cache. The process takes less than a second for most requests in the datasets currently tested. After conversion the data is available in the cache and a repeated attempt to access the data will succeed. During conversion, a copy of the original climate data is reordered but not modified in any way. Loading the full resolution data provides identical results to loading the original NetCDF volume.

Our on-demand conversion system facilitates the delivery of specific climate data in seconds that would have previously required long transfers. Combined with our user-specified dynamic analysis system and remote streaming, large multi-ensemble data comparisons can be achieved which were previously extremely cumbersome or impossible.

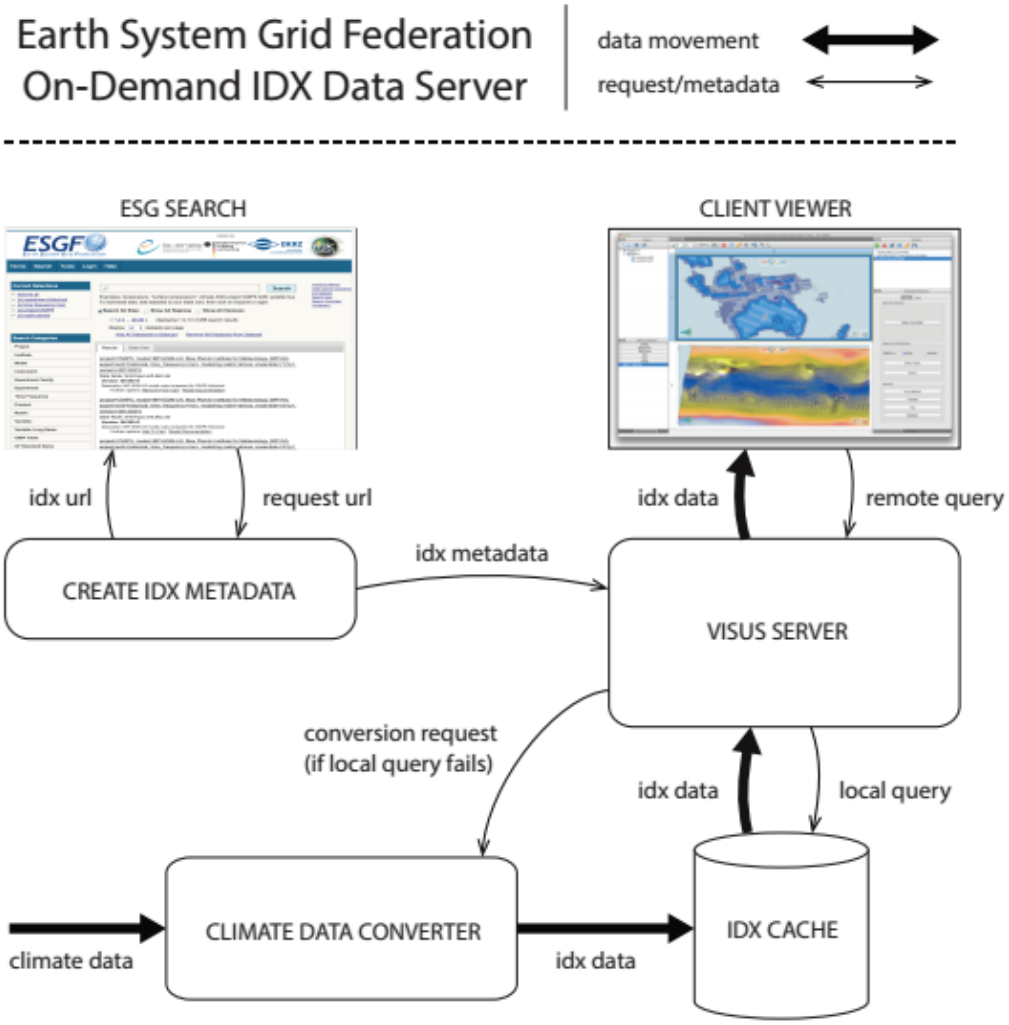

Figure 1. Current design of the interface between ESGF data provider node and ViSUS streaming data distribution server. A web request (top-left) allows to generate a configuration file that is added to the ViSUS server and downloaded to the client for reference in the bookmarks. A ViSUS client (top-right) can refer to the downloaded configuration to initiate specific requests to fields and time steps to visualize. The ViSUS server back-end (bottom) checks for each request if the data is available and, if not, it starts a process converting data at the smallest granularity possible.