Vortex2 Metadata Management on PRAGMA Cloud: A GeoPortal Experience

Umashanthi Pavalanathan, Yuan Luo, Beth Plale
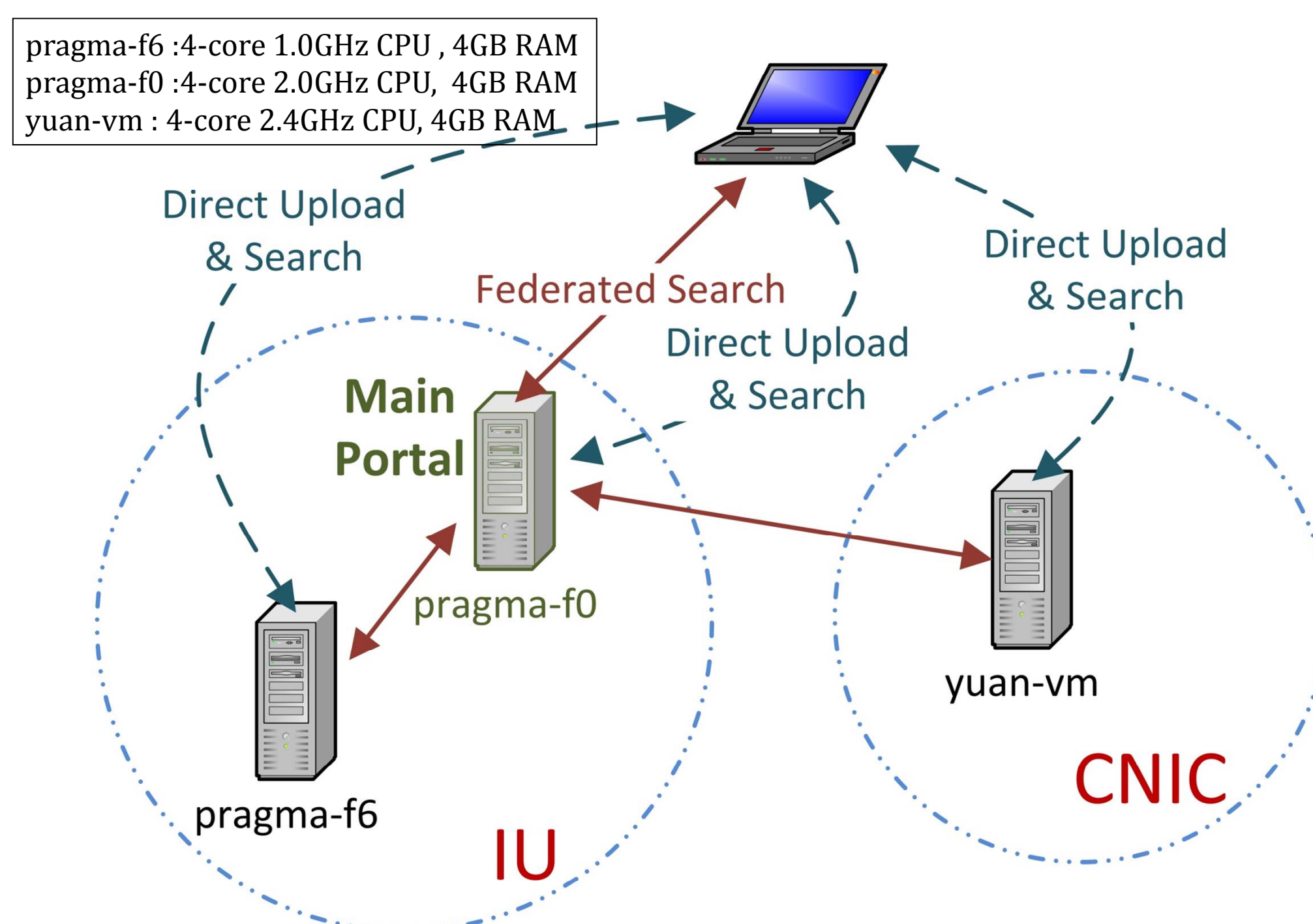*Indiana University*
umapaval@indiana.edu

DATA TO INSIGHT CENTER
INDIANA UNIVERSITY
Pervasive Technology Institute

- Metadata is crucial to preserving scientific investments in data by enabling metadata collection, long-term preservation, and reuse of scientific data.

- Vortex2, a large NSF funded 6-week field effort in Spring 2010, had well-instrumented vehicles driving nightly from Texas to Wisconsin to observe and measure potential tornadic activity. Our center generated 5 short-term forecasts per day for the 6-week duration. The forecasts were then turned into images and made available to field researchers on their cell phones. The collection, now curated, contains over 9000 products from 175 weather forecasts.

- ESRI GeoPortal is an open source metadata server that enables discovery and use of geospatial resources.

- We experimented with metadata catalogs on PRAGMA by federating GeoPortal instances and using the IU Vortex2 collection. A GeoPortal server instance can be registered as a resource under another instance. Each GeoPortal instance collects metadata provided by its host and exposes itself to the main portal via the CSW protocol. Metadata files are directly uploaded to the GeoPortal instances

- We set up 3 instances of GeoPortal server. We stored the metadata for the 175 weather forecasts to each instance.

Experiment Architecture

pragma-f6 : 4-core 1.0GHz CPU , 4GB RAM
pragma-f0 : 4-core 2.0GHz CPU, 4GB RAM
yuan-vm : 4-core 2.4GHz CPU, 4GB RAM

We set up our demonstration using a hierarchical organization (above) where a primary GeoPortal server on prgama-f0 is aware of two secondary GeoPortal servers on pragma-f6 and yuan-vm. During federated search, primary portal searches itself and secondary portals, combines results, and returns results to the client. Search results are formatted as a metadata abstract which contain links to downloadable metadata files as shown in image below.


Experiments

We implemented a RESTful client for metadata upload and search. We examined two search methods:

- Direct search - client queries multiple GeoPortal instances separately

- Federated search - a client queries multiple metadata repositories through a primary instance
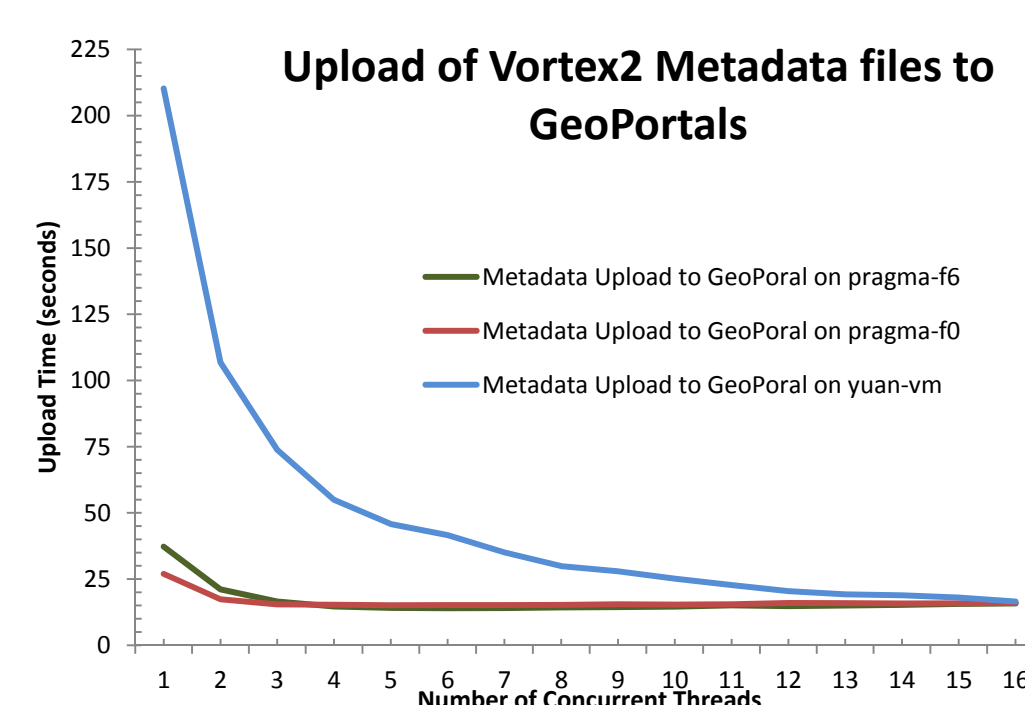
1) Measures upload time of set of metadata files using multi-threads clients. Result shows that multi-thread upload of Vortex2 metadata advances the performance than single-thread upload.

2) Stress test of uploading metadata onto server by increasing number of clients. Results show average turnaround time for each upload increases in linear manner.
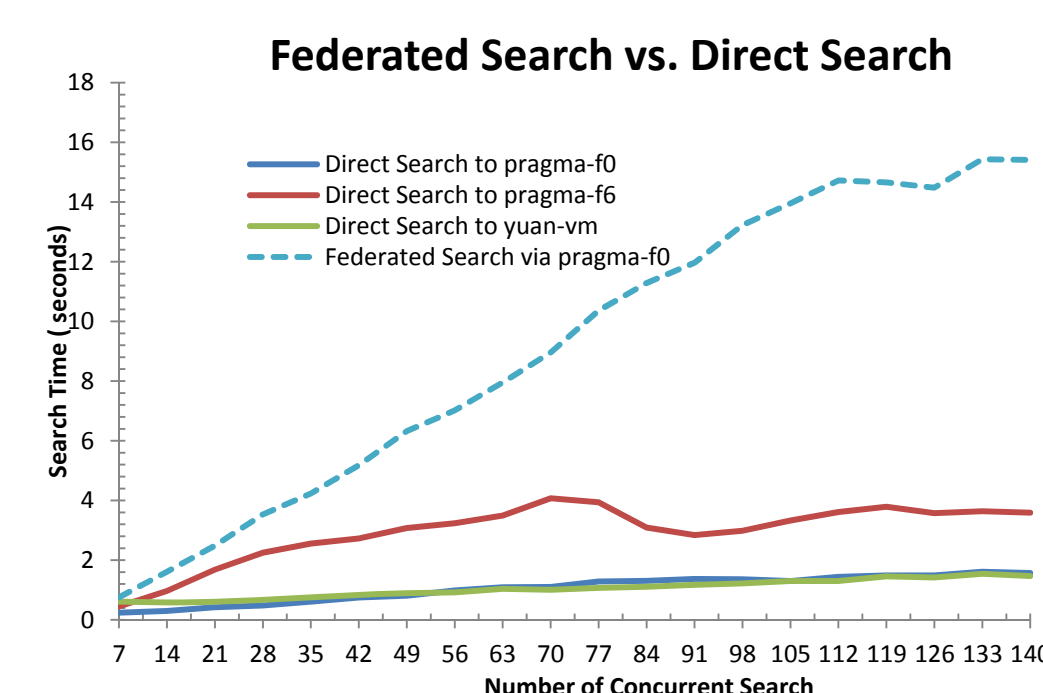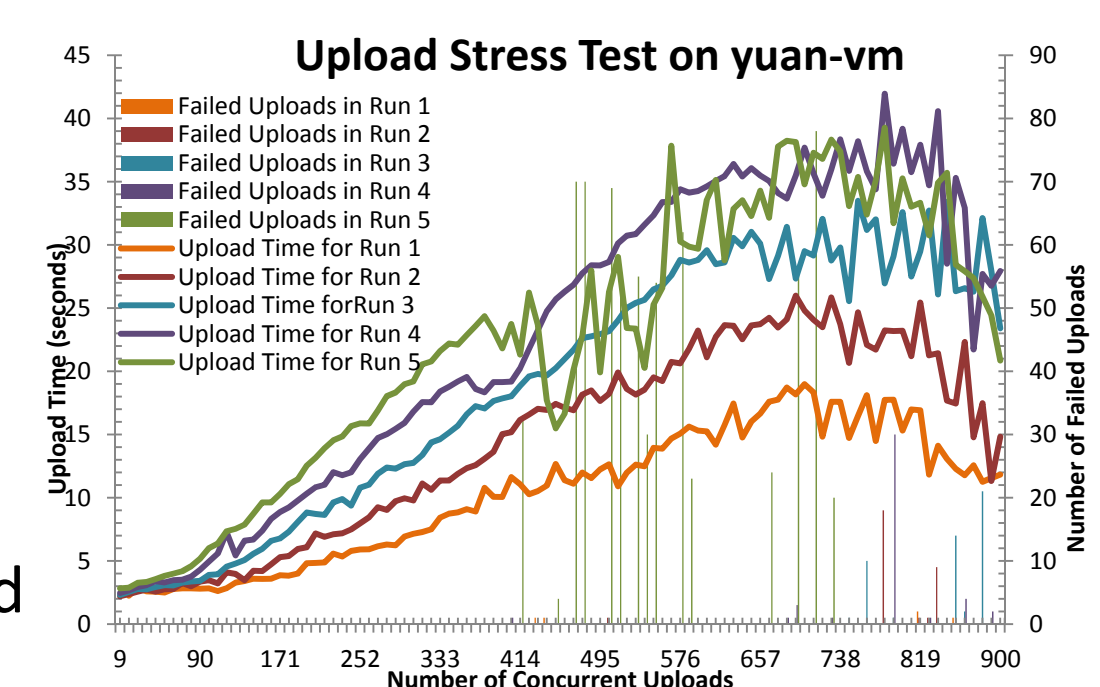
3) Comparison of turnaround time of federated search and direct search. Results show that direct search in which more work has to be done at the client side performs better than federated search in which more work is done on the main portal.