Efficient and reliable data access using distributed and coordinated cache system

Sonia Taneja
INFN-CNAF

Personal info…

- Post-doctoral research fellow at INFN-CNAF, Bologna Italy
- User support - contact person for CMS
- Research and Development division (was part of INDIGO-DataCloud project)
- Started this fellowship - April 2018

Project and research interest

- Theme - Innovative Workflow and Data Management solutions for Large Scale science: large datasets, large workloads, heterogeneous platforms.
- Project- Distributed and coordinated cache system
  - Architecture - based on pool of distributed caches (provided by well connected WLCG sites), which are loosely coordinated by a central orchestrator to create an effective larger cache which will scale to better accommodate LHC needs for an efficient data access
  - Reduce latencies / Improve efficiency on remote data access
  - Reduced operational cost
- Present status:
  - Cache for http/WebDAV and StoRM (Nginx)
  - Collaborating with INFN-Perugia to converge on a generic cache solution

Future activities

- Automated deployment
- Exploring available cache technologies
- Customise the cache algorithms to match experiment requirements (Predictive analysis)
- To investigate AuthN/Z policies
- Implement federated cache
- Eventually test on commercial clouds