



Aalborg Universitet

AALBORG UNIVERSITY
DENMARK

ATLAS DDM integration in ARC

Behrmann, Gerd; Cameron, David; Ellert, Mattias; Kleist, Josva; Tage, Adrian

Publication date:
2007

Document Version
Publisher's PDF, also known as Version of record

[Link to publication from Aalborg University](#)

Citation for published version (APA):

Behrmann, G., Cameron, D., Ellert, M., Kleist, J., & Tage, A. (2007). *ATLAS DDM integration in ARC*. Poster presented at Computation in High Energy Physics - CHEP'07, Victoria, Canada.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- ? Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- ? You may not further distribute the material or use it for any profit-making activity or commercial gain
- ? You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.

ATLAS DDM Integration in ARC

Josva Kleist¹, David Cameron¹, Adrian Taga², Gerd Behrmann¹, Mattias Ellert¹

¹Nordic Data Grid Facility

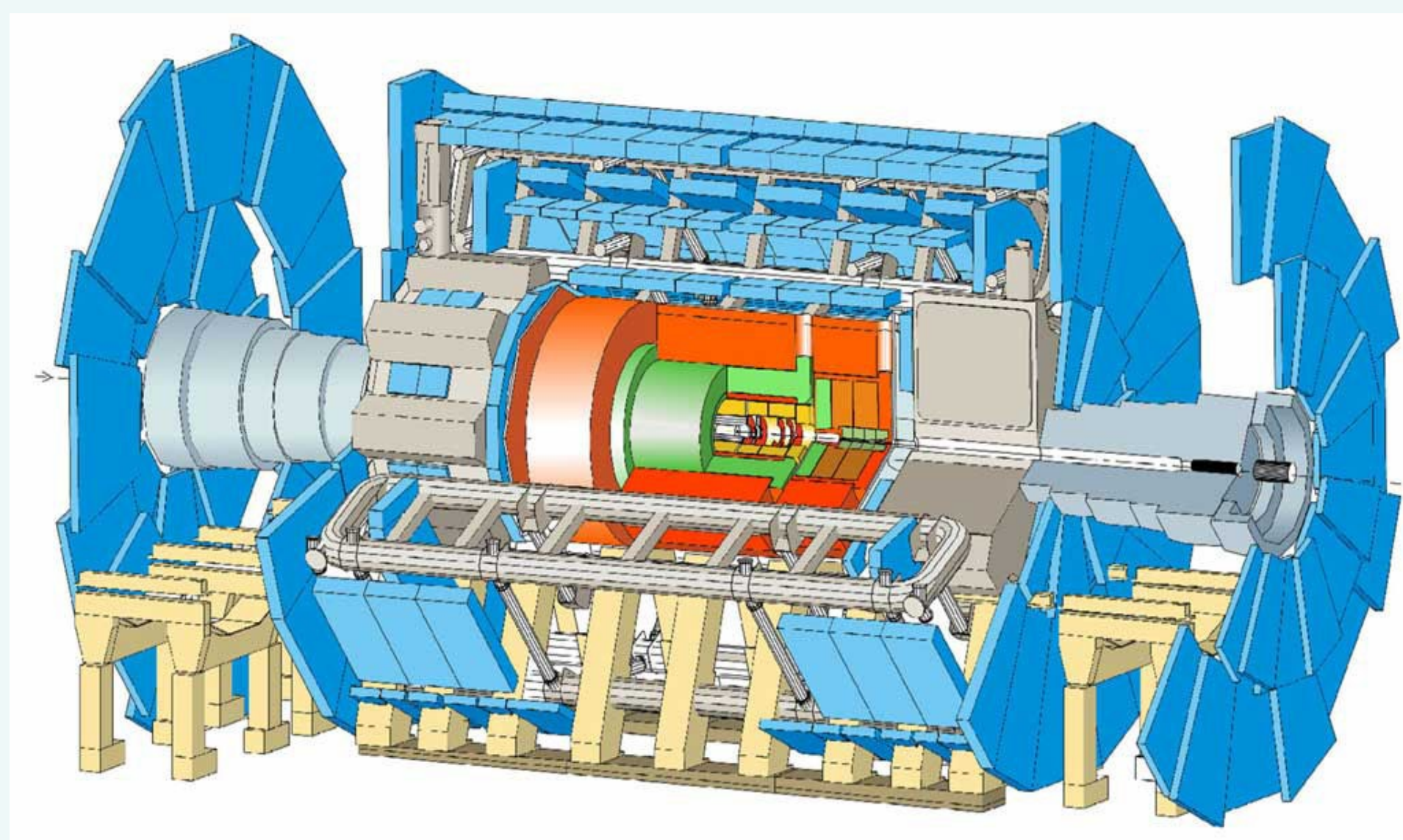
² University of Oslo, 0316 Oslo, Norway



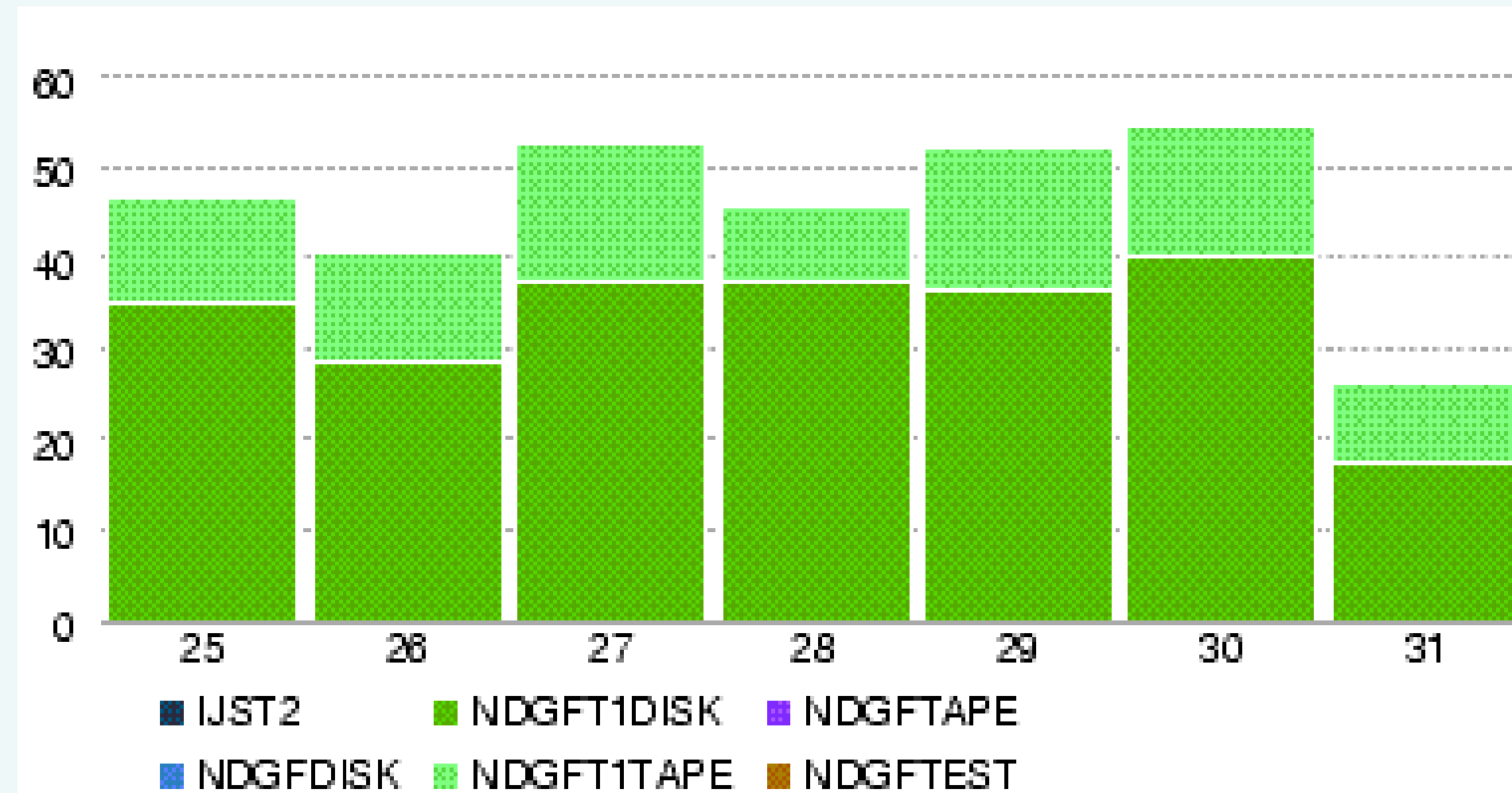
The Nordic Data Grid Facility (NDGF) consists of Grid resources running ARC middleware in Scandinavia and other countries. These resources serve many virtual organisations and contribute a large fraction of total worldwide resources for the ATLAS experiment, whose data is distributed and managed by the DQ2 software. Managing ATLAS data within NDGF and data distribution between NDGF and other Grids used by ATLAS (the LHC Computing Grid and the Open Science Grid) presents a unique challenge for several reasons. Firstly, the entry point for data, the Tier 1 centre, is physically distributed among heterogeneous resources in several countries and yet must present a single access point for all data stored within the centre. The middleware framework used in NDGF differs significantly from other Grids, specifically in the way that all data movement and registration is performed by services outside the worker node environment. Also, the service used for cataloging the location of data files is different from other Grids but must still be useable by DQ2 and ATLAS users to locate data within NDGF. This poster presents in detail how we solve these issues to allow seamless access worldwide to data within NDGF.



Data Export from CERN



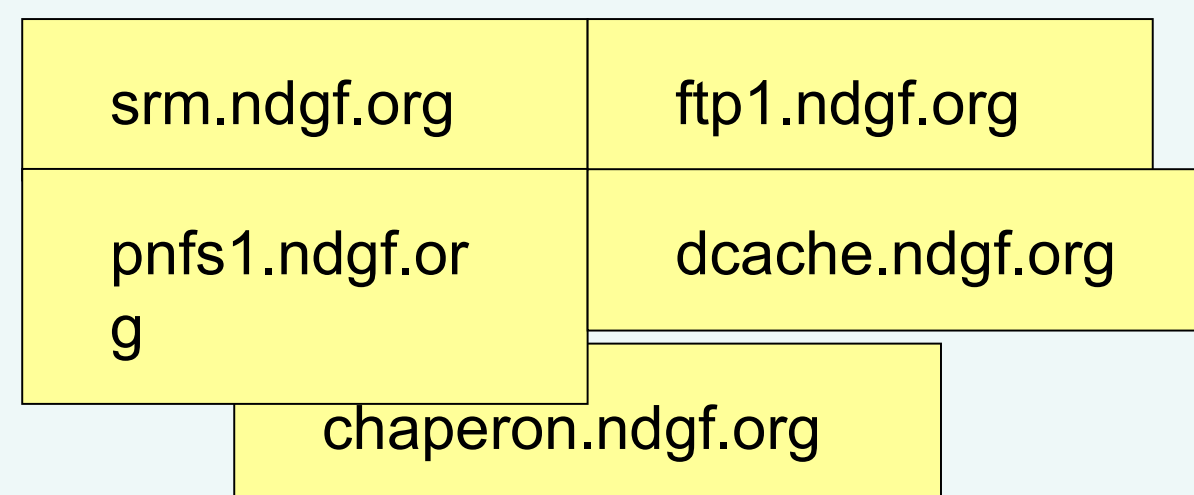
Throughput (MB/s) per day to NDGF T1



We have already shown that we can accept data from CERN at the nominal rate expected at full detector running.

Data streaming at 50 MB/s

Tier 1 Endpoint (Copenhagen)

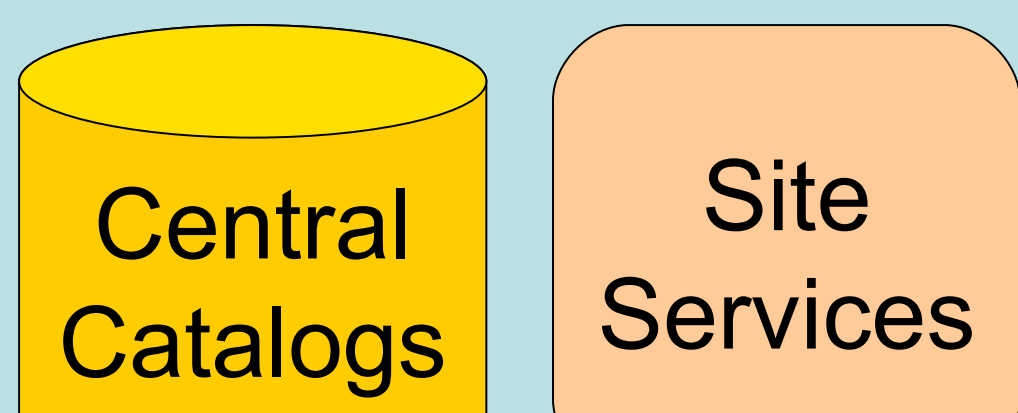


For more information on the distributed Tier 1 see Talk #147, "A Distributed Tier-1 for WLCG".

Tier 1 dCache pools



DQ2



DQ2 is the ATLAS distributed data management system. It sends data to the Tier 1 endpoint at srm.ndgf.org. The data is physically distributed around the Tier 1 dCache pools.

File Replica Catalog

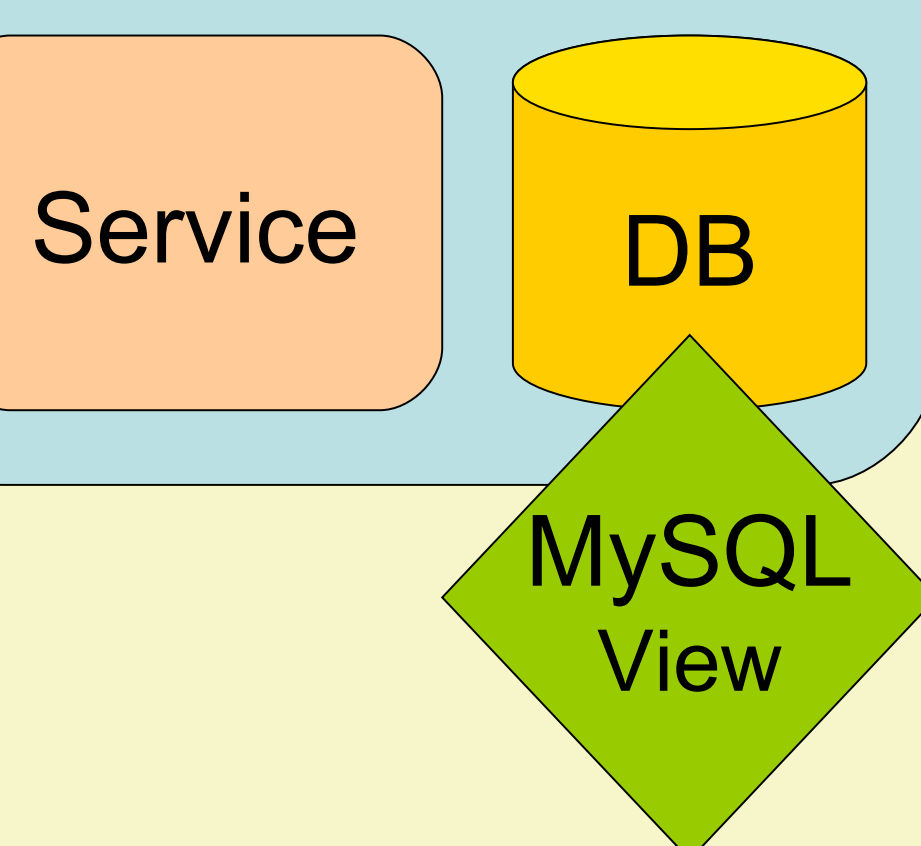
ATLAS file replica information in NDGF is stored in the Globus Replica Location Service (RLS). This data needs to be accessible to the rest of the world without dependencies on NDGF specific software. There are two ways to do this.

HTTP web service

globus-rls-cli

The web service approach involves deploying a simple service which interfaces to RLS and can be queried by any HTTP client (eg curl). The web service queries RLS using SQL directly on the RLS DB tables, or using the Globus client tool. This tool gives good performance but requires setting up and maintaining an extra service.

Globus RLS



The MySQL view approach involves creating a special database to expose the data in RLS in a way that looks like a MySQL LRC. These are the replica catalogs used in the Open Science Grid (OSG) in the USA with which DQ2 is already integrated. This MySQL view has the advantage of being easy to set up and requiring no additional resources, however it does not give optimal performance, especially with bulk queries.

ATLAS MC Production

Dulcinea is the executor for the ATLAS production system. It is in charge of executing simulation tasks assigned to NorduGrid.

job

Dulcinea sends jobs to ARC CEs

Dulcinea

Tasks

After successful completion, the data is uploaded to the Tier 1 storage by ARC. Dulcinea registers attributes required by DQ2 (eg GUID and MD5 checksum) in the RLS and registers the new output files in DQ2.