Large scale tailored reprocessing on NAF

S. Fleischmann, R. Prabhu, P. Wienemann
Universität Bonn

LHC-D Meeting, Aachen 26.-27 November 2008

Background / Motivation

- development of novel **tau reconstruction algorithm** for ATLAS based on particle flow concepts

- required input objects (energy flow objects) currently not stored in AOD

- require several **packages in addition** to those available in standard releases

- requires tailored reconstruction on large data samples with access to calorimeter cell deposits

- NAF provides an ideal work ground

Desired data samples (RDO): Total ~ 12 TB

Process		Data samples	Estimated size (GB)
QCD	**J0**	**misal1_csc11.005009.J0_pythia_jetjet.digit.RDO.v12003103**	2020 x 4
	J1	**ideal2_mc12.005010.J1_pythia_jetjet.digit.RDO.v13003004**	
	J2	**ideal2_mc12.005011.J2_pythia_jetjet.digit.RDO.v13003004**	
	J3	**ideal2_mc12.005012.J3_pythia_jetjet.digit.RDO.v13003004**	
SUSY	**SU1**	**misal1_csc11.005401.SU1_jimmy_susy.digit.RDO.v12003107**	600
	SU3	**misal1_csc11.005403.SU3_jimmy_susy.digit.RDO.v12003107**	1500
SM	**Z->tautau**	**misal1_mc12.005179.ZtautauNoEF.digit.RDO.v12000605**	400
	W ->tauNu	**misal1_csc11.005107.pythia_Wtauhad.digit.RDO.v12003103**	1120
Private	**various**		30

Data access & replication of datasets

- Reprocess data primarily using **NAF batch system**
- Currently a subset of all samples (~1.8 TB) copied to **scratch** on NAF (using dq2-get)
- Aim: **replicate** all samples (~180 000 files) to NAF (mainly from BNL via FZK)
  - all request have been processed
- QCD & SU1 already replicated to **DESY-HH / DESY-ZN**
- Remaining samples replicated to **FZK**, waiting for replication to DESY

Summary

- Using NAF infrastructure for the **reprocessing** of large data sets in the context of new developments in tau reconstruction

- **Our experience**: NAF provides a **user-friendly** and **efficient** work bed for reprocessing and analysing large data samples (no major obstacles experienced). Sometimes compiling on NAF login machines is very slow (~hours to compile a few Athena packages)

- Subset of requested datasets currently copied to scratch, but **full datasets** are being **replicated to DESY-HH/DESY-ZN**

- Thanks to **W. Ehrenfeld & NAF-support team** for forthcoming and helpful assistance!