

# A New CDF Model for Data Movement Based on SRM



Manoj K. Jha  
INFN- Bologna

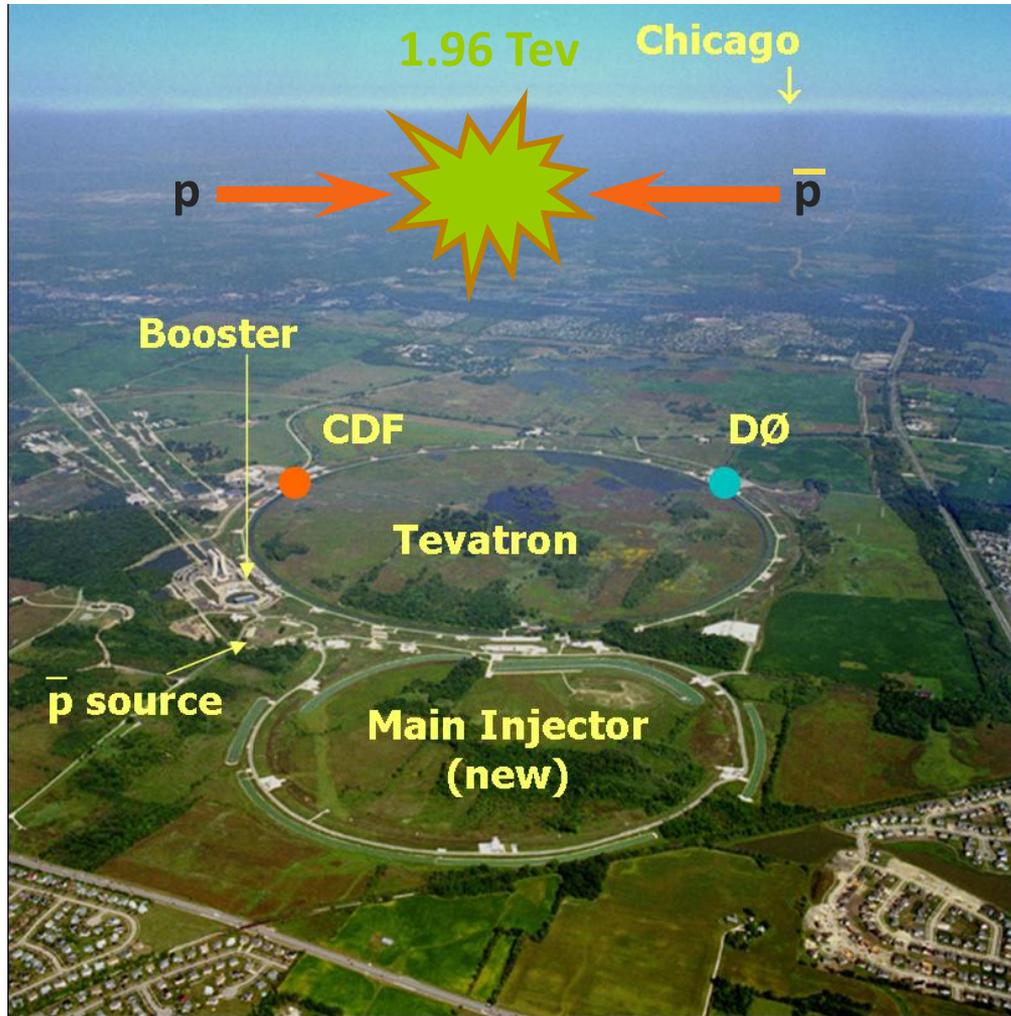
27<sup>th</sup> Feb., 2009  
University of Birmingham, U.K.

# Overview

---

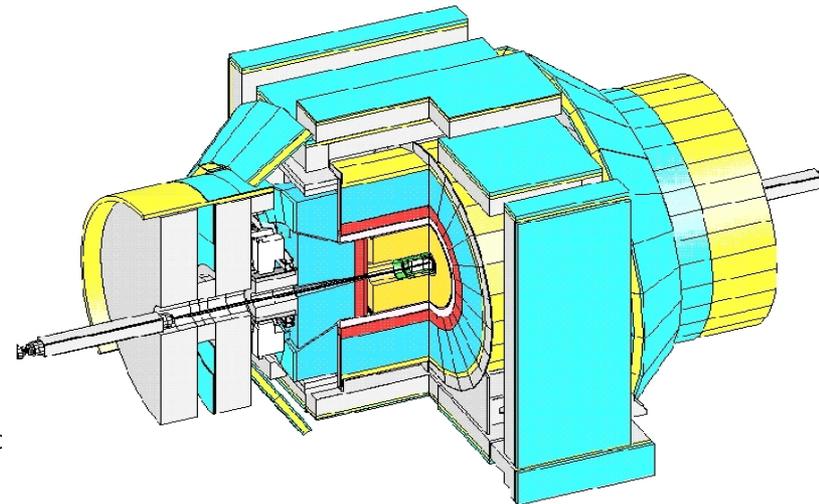
- ❖ Introduction
  - ❖ CDF
  - ❖ Tevatron Performance
- ❖ CDF Analysis Framework (CAF)
  - ❖ Present MC Prod. Model
  - ❖ Data Transfer Model
- ❖ Proposed Model
- ❖ Test Framework
- ❖ Performance Parameters
- ❖ Other Activities

# Collider Detector at Fermilab (CDF)



## ❖ CDF :

- ❖ Large multipurpose Particle Physics Experiment at Fermi National Accelerator Lab (Fermilab)
- ❖ Started collecting data in **1988**
- ❖ Data taking will continue until at least Oct. 2009 (with desire to extend another year) also in 2010

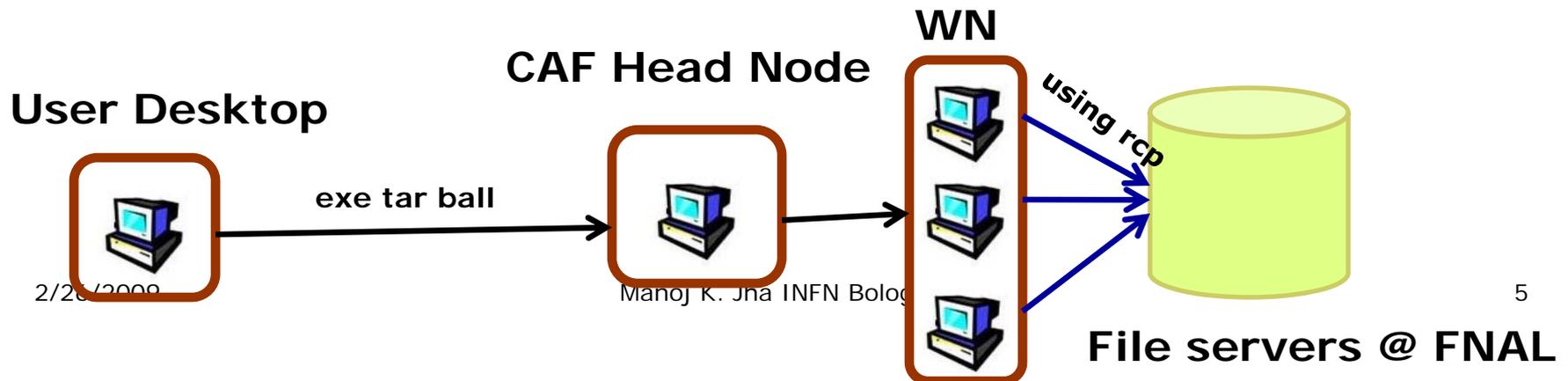




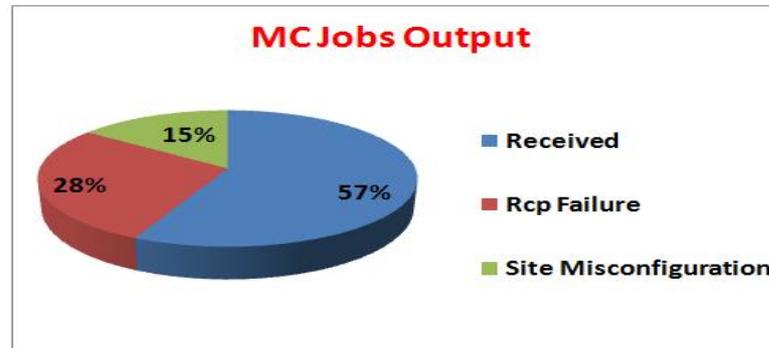
# CDF Analysis Framework(CAF)

## ❖ CAF developed as a portal.

- ❖ A set of daemons (submitter, monitor, mailer) accept requests from the users via kerberized connections.
- ❖ Requests are converted into commands to the underlying batch system.
- ❖ A user job consists of several sections. Each section has a wrapper associated with it.
  - ❖ *Task of wrapper is to setup security envelop & prepare environment before the actual user code starts.*
  - ❖ *When the user code finishes, it will also copy whatever is left to a user specified location using rcp.*



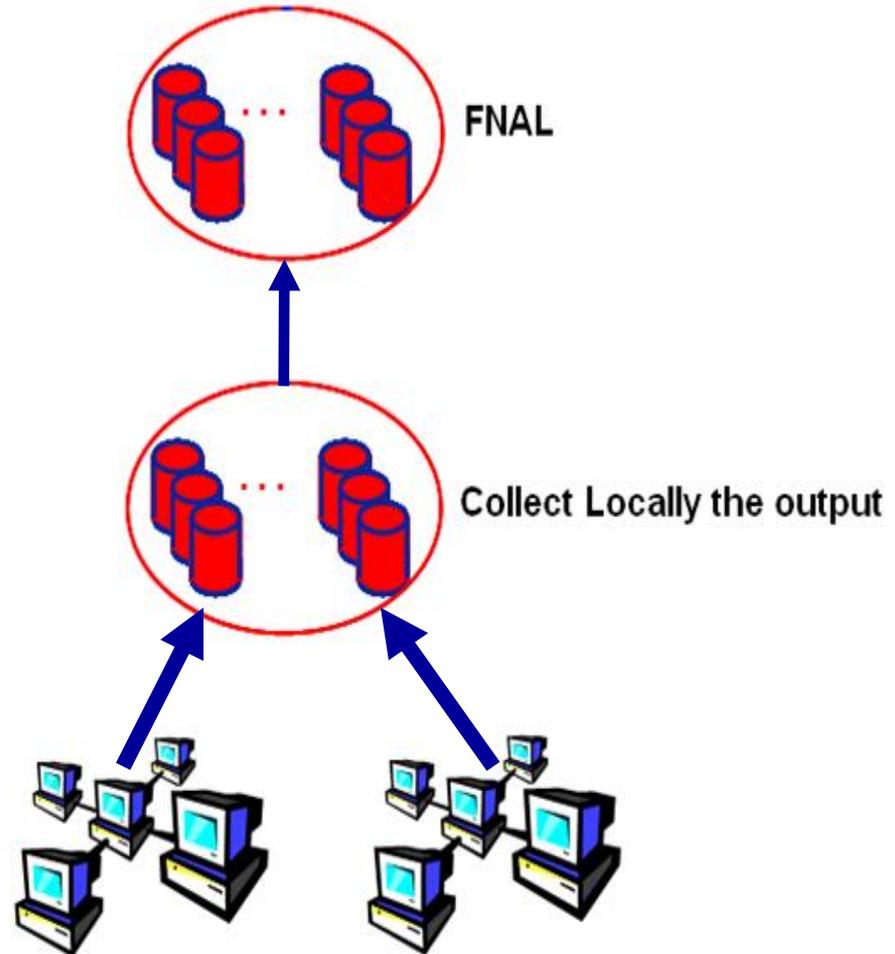
# Present CAF MC Prod. Model: Drawbacks



- ❖ Presently, CDF is relying on rcp tool for transfer of output from Worker Nodes(WN) to destination file servers. It leads to
  - ❖ WN are sitting idle just because another WN is transferring files to the file servers.
    - ❖ *Loss of output from the WN in most of the cases.*
    - ❖ *Wastage of CPU and network resources when CDF is utilizing the concept of opportunistic resources.*
  - ❖ No mechanism to deal with the sudden arrival of output from WN at file servers. Especially, happens during the conference period.
    - ❖ *Overloading of available file servers in most of the cases.*
    - ❖ Users have to resubmit the job.
- ❖ No catalogue of MC produced files on user basis.
- ❖ *CDF needs a robust and reliable framework for data transfer from WN to Storage Element(SE) at Fermilab.*

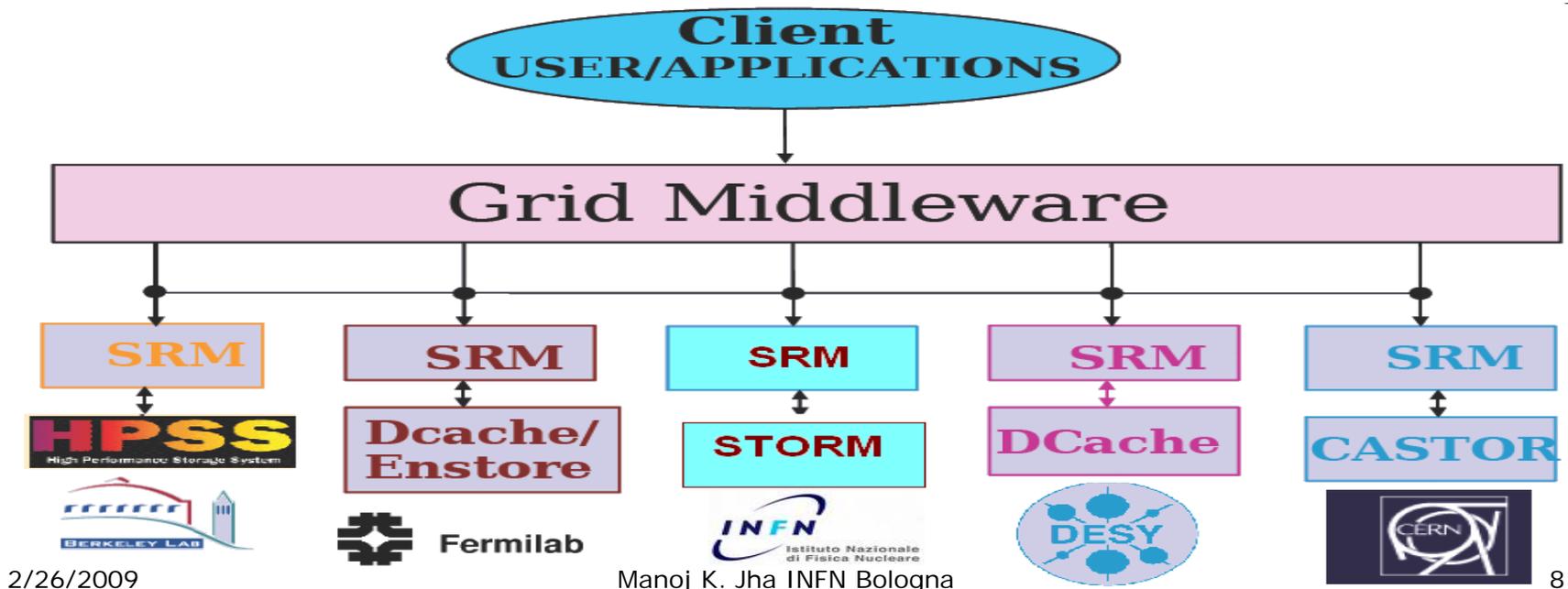
# Date Transfer Model

- ❖ Output from WN will be temporarily collected in the SE closer to WN.
  - ❖ Transfer or wait time for MC output data on WN will be considerably reduced due to large bandwidth between WN and its SE.
- ❖ Transfer sequentially the collected output to the destination SE.
  - ❖ How to manage the files on SE ?
    - ❖ *Use SRM managed SE*
  - ❖ How to transfer file b/w SEs ?
    - ❖ *Use Sequential Data Access via Meta-Data (SAM) features for transfer of files between its stations.*
- ❖ Model uses the SAM SRM interface for data transfer.



# Storage Resource Manager(SRM)

- ❖ Uniform Grid interface to heterogeneous storage
  - ❖ Negotiate protocols
  - ❖ Eliminate specialized code
- ❖ Dynamic space allocation and file management on shared storage components on Grid
- ❖ Interface specification v1 is in field, v2 is latest



# SAM Data Management System

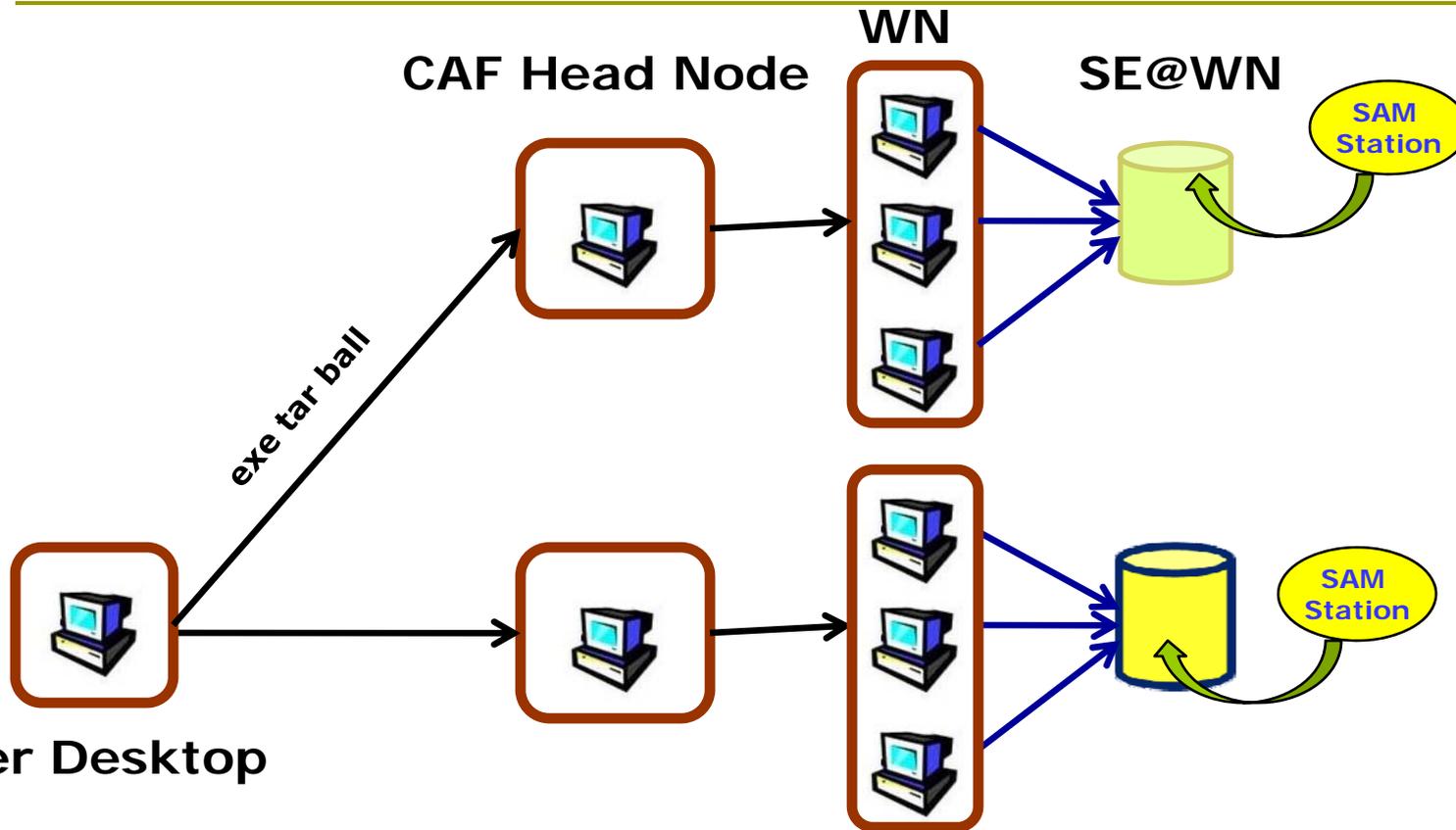
- ❖ SAM is Sequential data Access via Meta-data
- ❖ Est. 1997. <http://dodb.fnal.gov/sam>
  - ❖ Flexible and scalable distributed model
  - ❖ Reliable and Fault Tolerant
  - ❖ Adapters for many batch systems: LSF, PBS, Condor, FBS
  - ❖ Adapters for mass storage systems: Enstore, (HPSS, dCache, and SRM managed SE.
  - ❖ Adapters for Transfer Protocols: cp, rcp, scp, encp, GridFTP.
  - ❖ Useful in many cluster computing environments: Desktop, private network (PN), NFS shared disk,...
  - ❖ Ubiquitous for CDF users



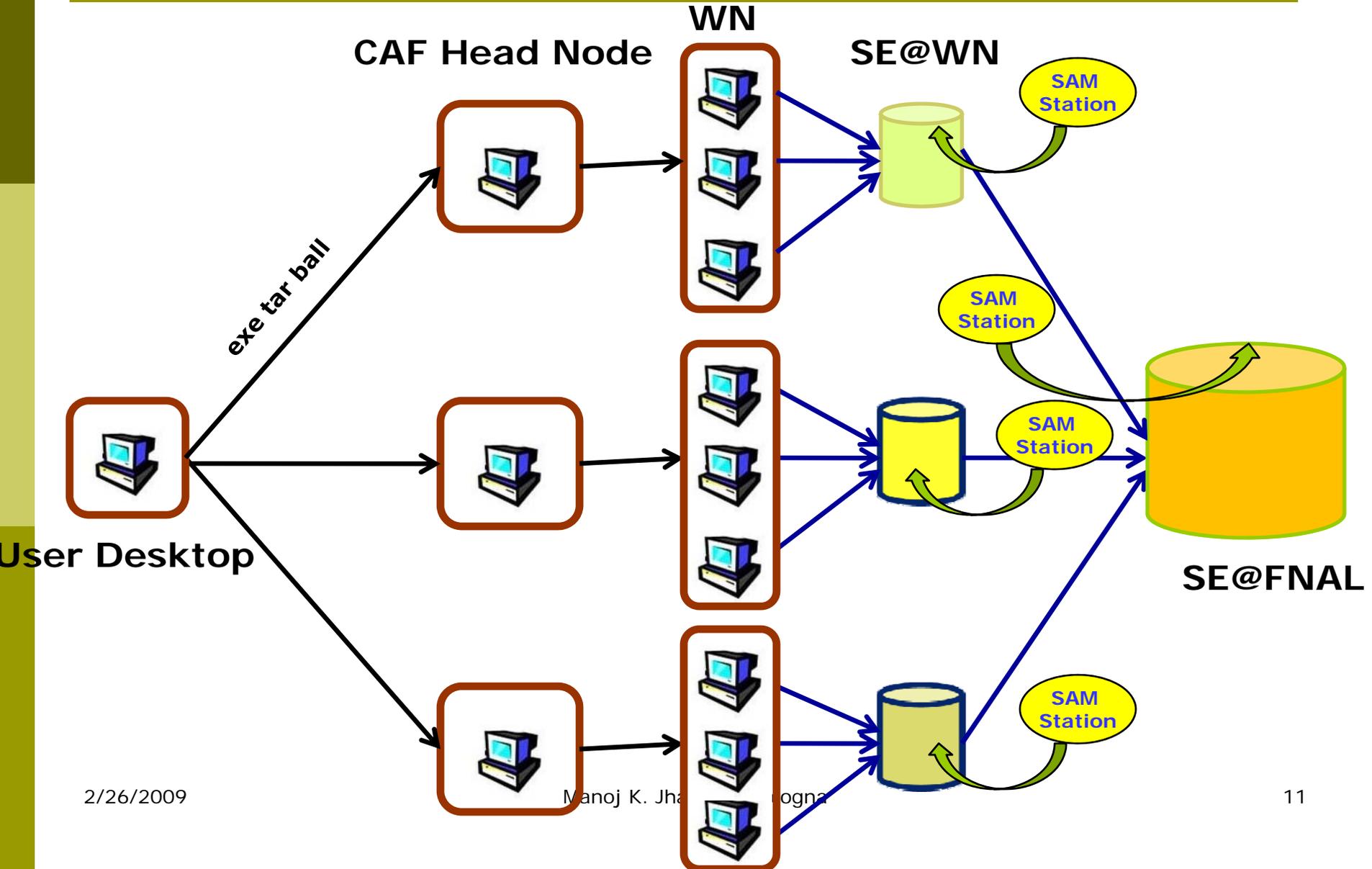
## SAM Station:

- ❖ Collection of SAM servers which manage data delivery and caching for a node or cluster.
- ❖ The node or cluster hardware itself.

# Proposed Model



# Proposed Model



# Test Framework

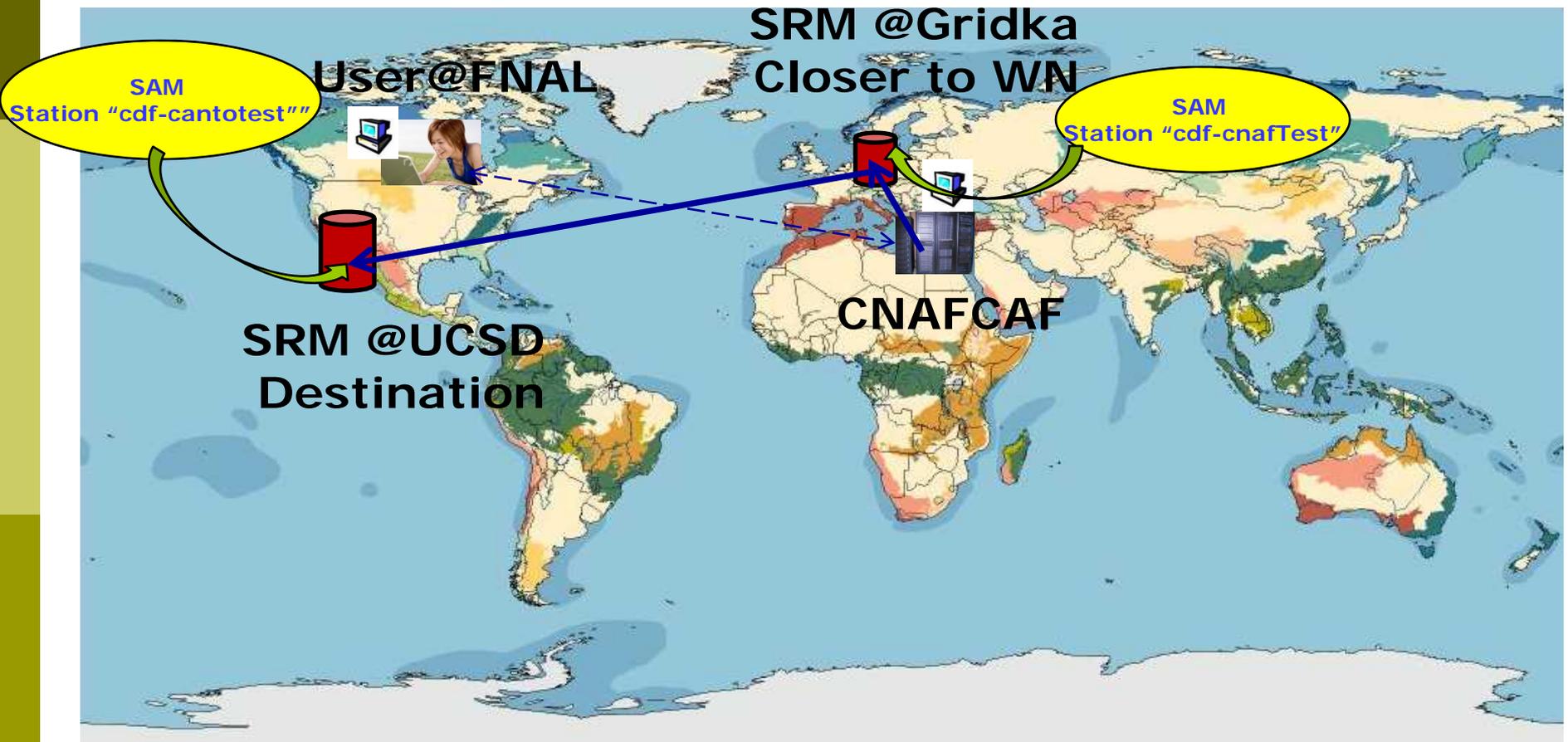
## ❖ Hardware:

- ❖ CAF: CNAF Tier -I
- ❖ SE: dCache managed SRMs
  - ❖ *500 GB of space at Gridka, Karlsruhe, 10 channels per request (closer to WN)*
  - ❖ *1 TB of space at University of California, San Diego(UCSD), 50 channels per request (destination SE)*
- ❖ SAM station:
  - ❖ *station “cdf-cnafTest” at “cdfsam1.cr.cnaf.infn.it” (closer to WN)*
  - ❖ *station “canto-test” at “cdfsam15.fnal.gov” (destination SE)*

## ❖ Setup:

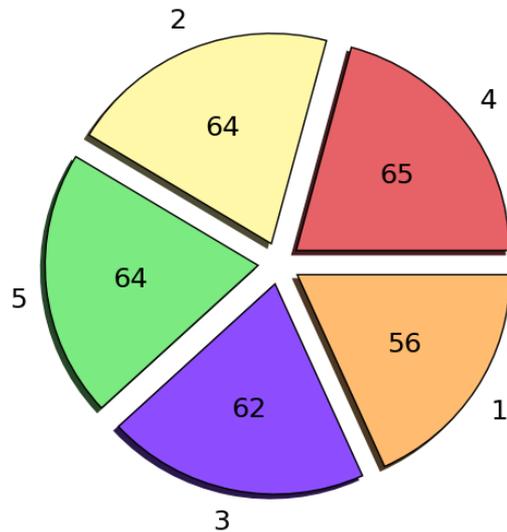
- ❖ A cronjob submits a job of variable no. of segments every 10 minutes at CNAF CAF. Maximum segments per job is 5.
  - ❖ *Each segment creates a dummy file of random size which vary between 10 MB to 5 GB.*
- ❖ A cron process running at station “cdf-cnafTest” for creation of datasets from the files in SRM closer to WN (Gridka).
- ❖ Another cron job running at station “canto-test” for transfer of dataset between two stations.

# Test Framework



# Distribution of no. of job sections

Distribution of job sections (Sum: 312)



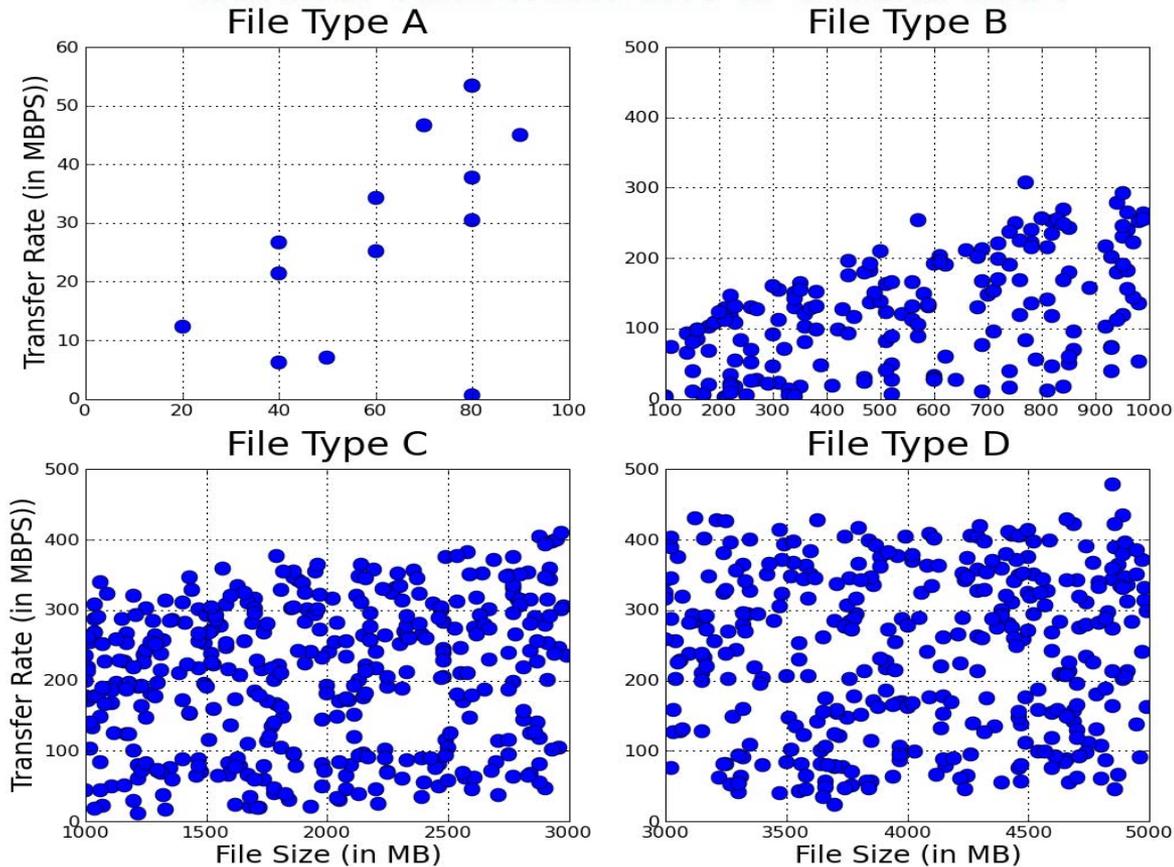
■ 4 (65)   ■ 2 (64)   ■ 5 (64)   ■ 3 (62)   ■ 1 (57)

- ❖ No. outside the circle represents no. of section being submitted in a single job
- ❖ No. inside the circle represents total no. of section being submitted for the test framework.
- ❖ For ex: A job with 1 section had been submitted 56 times, with 2 sections 64 times and so on.

File Type	File Size
A	10 MB < size < 100 MB
B	100 MB < size < 1 GB
C	1GB < size < 3 GB
D	size > 3 GB

# Transfer rate from WN to its closer SE

## Transfer rate from WN to Gridka SRM



**Transfer rate increase with file size for file types A, B & C.**

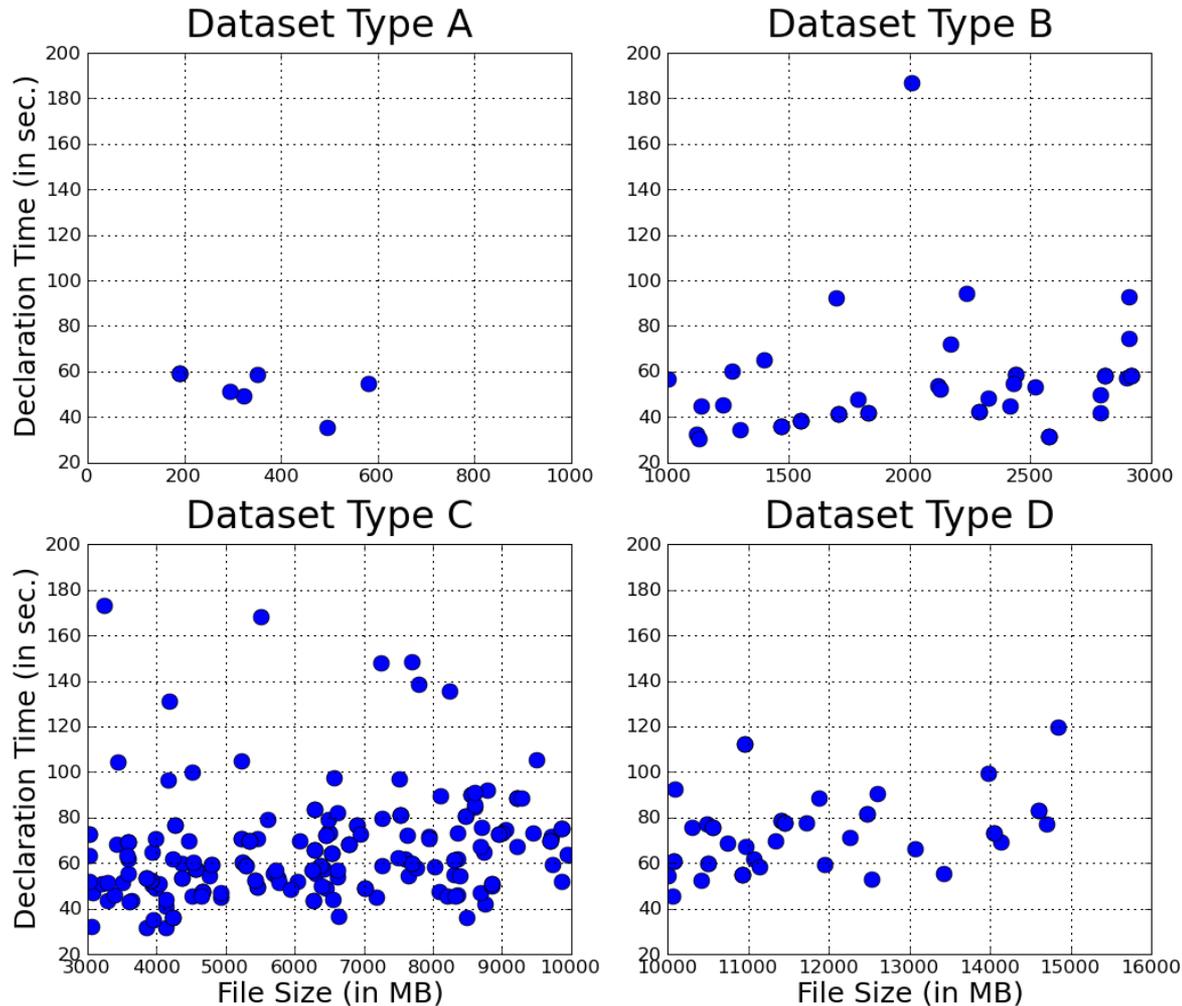
# Dataset Type

- ❖ A CAF job corresponding to JID consists of number of sections.
  - ❖ Each section has its output and log files.
    - ❖ *Size of output file varies from few Mega bytes to several Giga bytes while that of log file are of the order of few Kilo bytes.*
- ❖ Model proposes two types of dataset corresponding to a JID.
  - ❖ Output dataset:
    - ❖ *Collection of all the output files corresponding to a JID*
  - ❖ Log dataset:
    - ❖ *Collection of all the log files corresponding to a JID.*

Dataset Type	Dataset Size
A	10 MB < size < 1 GB
B	1GB < size < 3 GB
C	3 GB < size < 10 GB
D	size > 10 GB

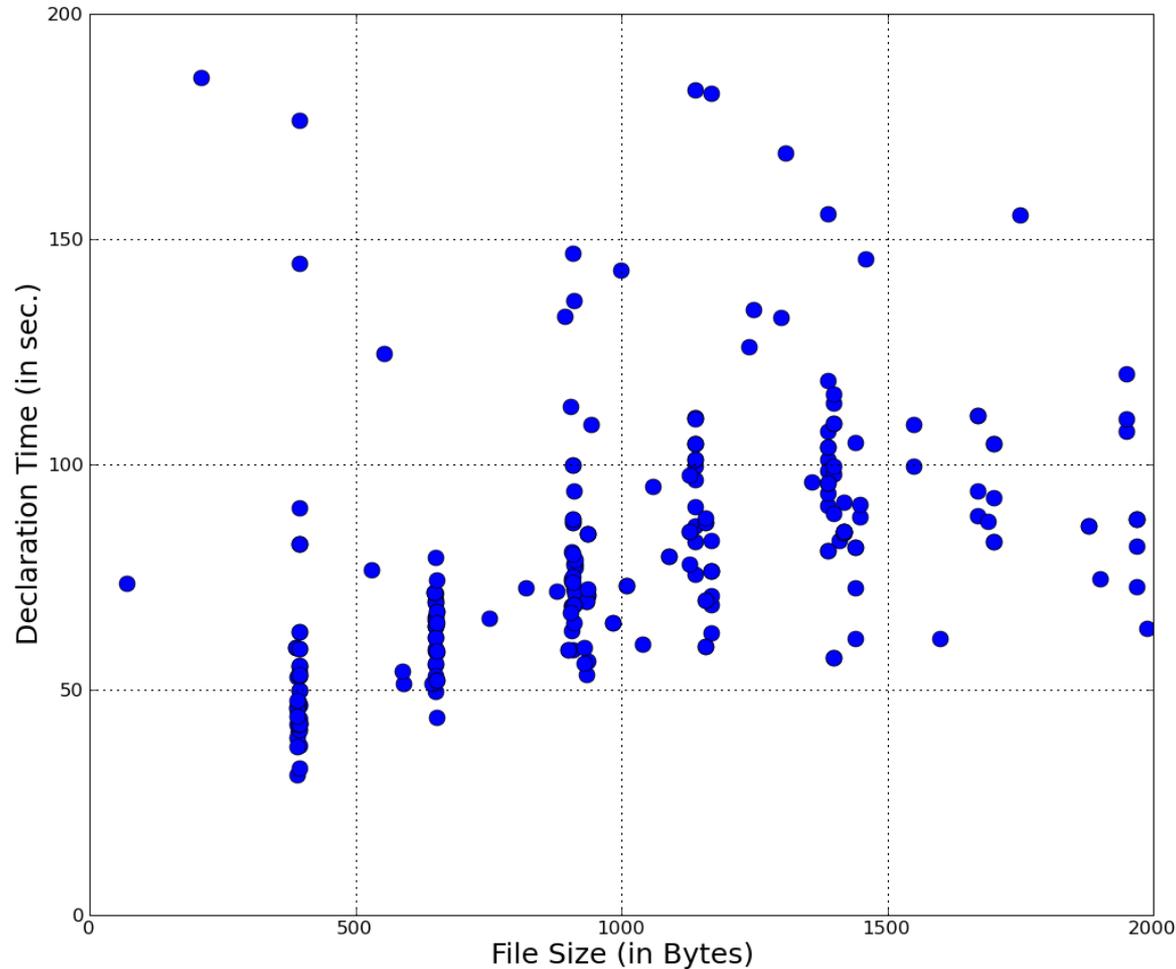
# Output File Dataset Creation Time

Time needed to declare dataset from output files in SAM



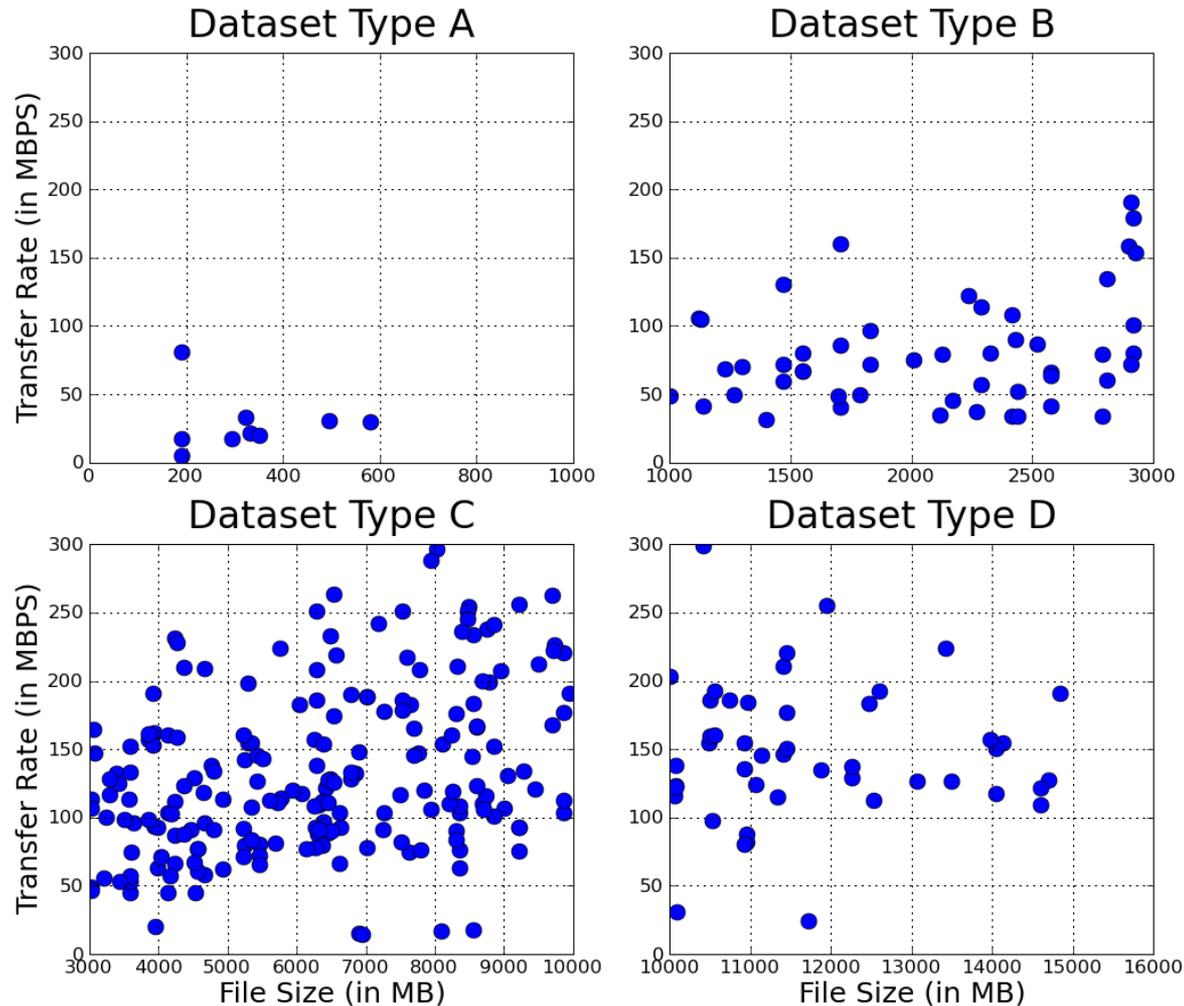
# Log File Dataset Creation Time

Time needed to declare dataset from log files



# Transfer Rate for Diff. Dataset

Transfer rate for diff. output dataset from Gridka to UCSD



# Getting files on User's Desktop

- ❖ Requirements on user's desktop:
  - ❖ A SAM station environment with its service certificate installed .
  - ❖ A SRM client
- ❖ File query:
  - ❖ User's will know the dataset name in advance corresponding to a JID. Since dataset name will follow a fix syntax.
  - ❖ User's can query the list of files which matches dataset using the sam command "sam list files --dim=" <datasetName>
  - ❖ File location:
    - ❖ Command "sam locate <fileName>" tells the location of file name.
- ❖ Getting file on desktop:
  - ❖ Command "sam get dataset -defName=<datasetName> --group=test -downloadPlugin=srmcp "
- ❖ A wrapper script can be written such that above changes will be transparent to users.

**Abstract of this work has been accepted to CHEP 09, Prague**

# Other Activities

## Computing:

- ❖ **Conducted the simulation workshop for CMS in Feb. 04.**
  - ❖ Participants learnt the installation of CMS software and their use in physics analysis
  - ❖ System administrator for University of Delhi, High Energy Physics Group from 2000 – 2006
- ❖ **Member of LHC Physics Center (LPC), Fermilab MC production group**
  - ❖ Generated large number of MC samples for validation and physics analysis of new releases of CMS software.

## Physics Analysis:

- ❖ **CDF:**
  - ❖ Search for Quark Compositeness Using Dijets (Abstract accepted to APS, Denver 2009)
  - ❖ Estimation of Heavy Flavor content in Minimum Bias events
- ❖ **CMS:**
  - ❖ Proposed the jet clustering algorithm for CMS: **CMS AN 2008/02**
  - ❖ CMS Sensitivity to Contact Interactions using Dijet: **J. Phys. G36:015004, 2009**

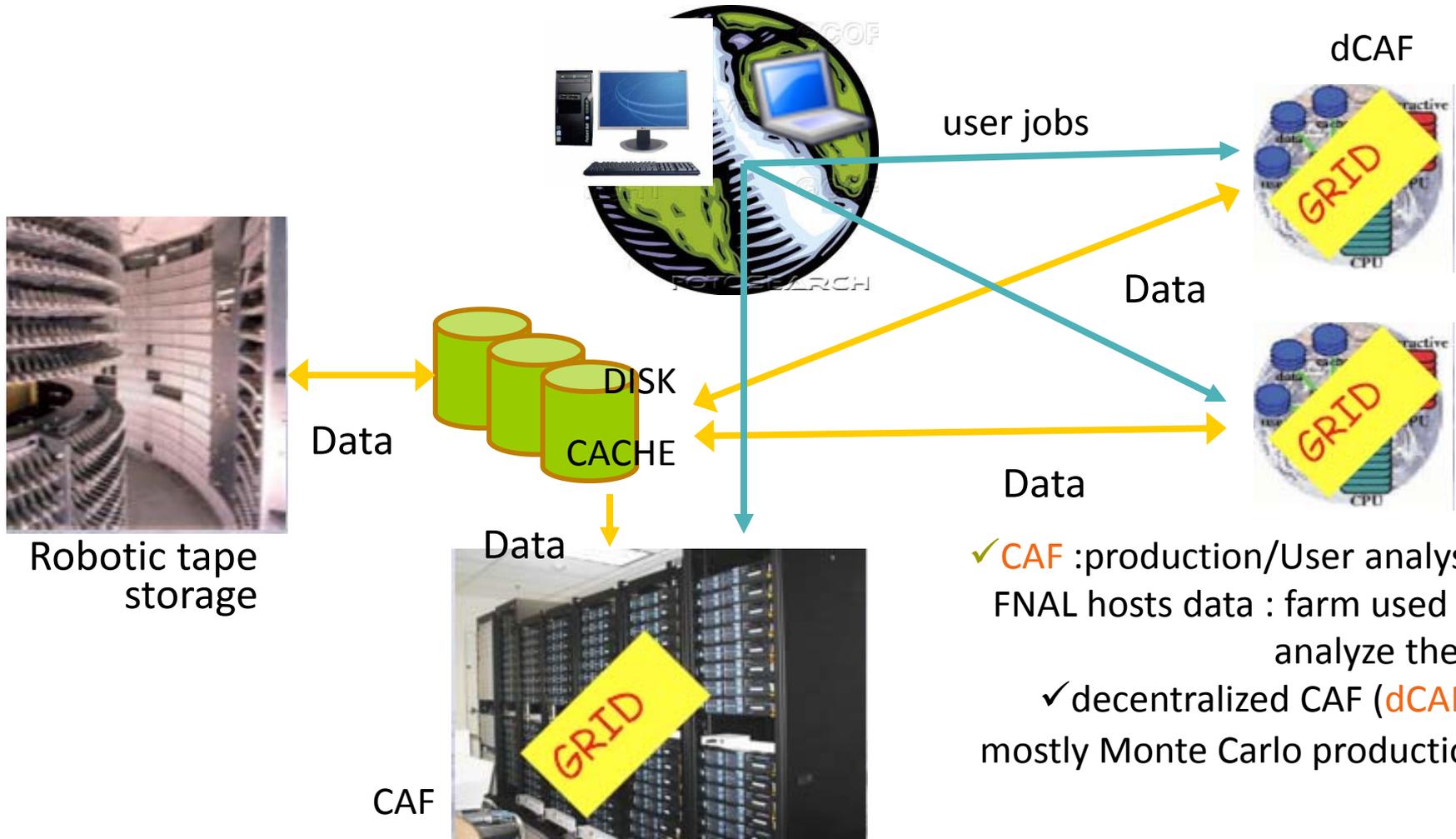
---

**Thank You !**

---

# Backup Slides

# Computing Model

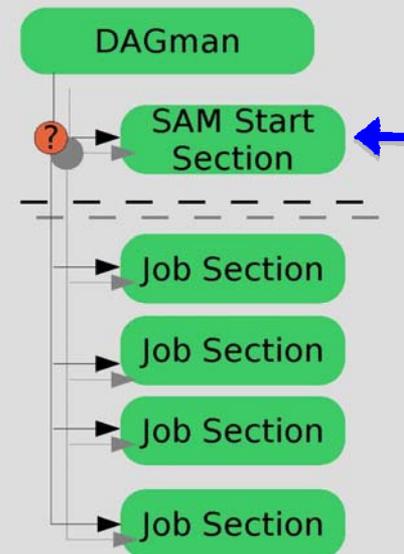


- ✓ CAF :production/User analysis  
FNAL hosts data : farm used to analyze them
- ✓ decentralized CAF (dCAF):  
mostly Monte Carlo production

# Condor's view of Computing job.

- ❖ We think of job as task with the related parallel tasks (Job sections)
- ❖ Condor – each job section – independent Job DAGMan – allows CDF Users – Condor to work together

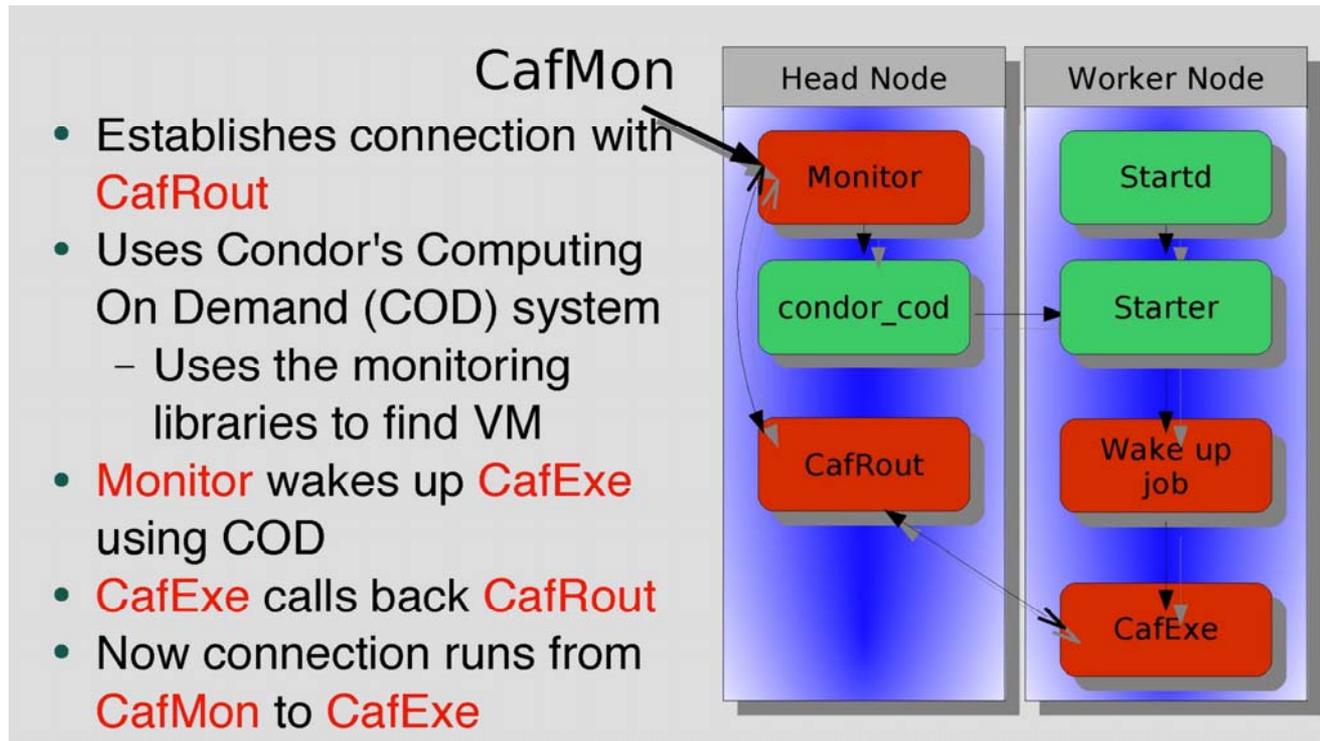
- A Condor “Scheduler Universe” job
  - This is jobs that manage other jobs
  - Runs on head node
- DAGman submits individual job sections automatically
  - Submission can be conditional
    - E.g. SAM start section first, other sections only if start works
- DAGman job finishes when all sections are done



Data Handling Step

# Interactive User Job Monitoring

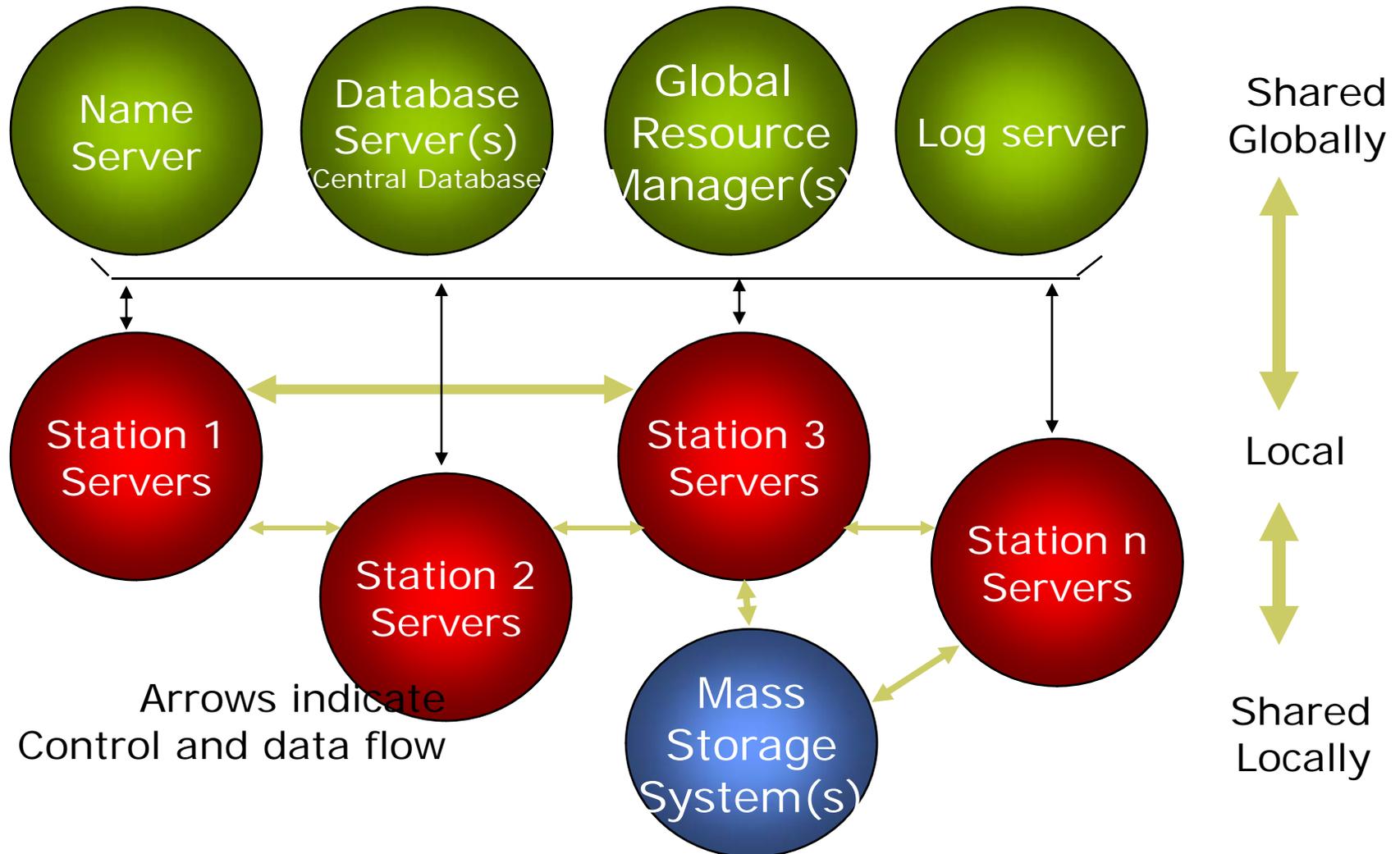
- ❖ Condor Computing on Demand (COD) is used to allow users to monitor/interact with their job



CDF Code and Condor Code together give the users tools needed to get their jobs efficiently

- w/ Condor COD Users can:
- look their working directories and files
  - check their running jobs (debug if needed)

# Overview of Sam



# The SAM Station

---

- ❖ Responsibilities
  - ❖ Cache Management
  - ❖ Project (Job) Management
  - ❖ Movement of data files to/from MSS or other Stations
- ❖ Consists of a set of inter-communicating servers:
  - ❖ Station Master Server,
  - ❖ File Storage Server,
  - ❖ Stager(s),
  - ❖ Project Manager(s)

# Components of a SAM Station

