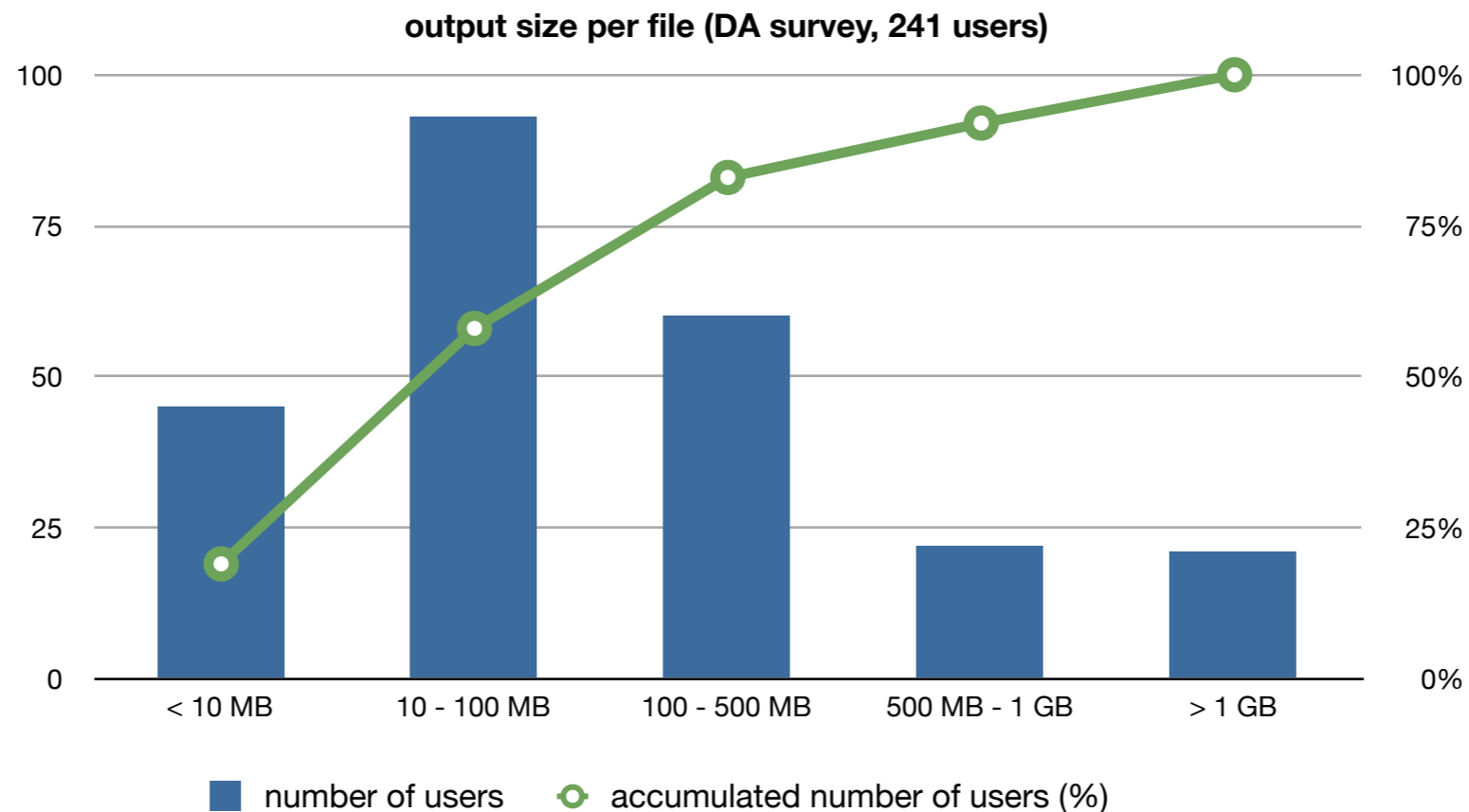




Output Merging for Analysis Jobs

Hurng-Chun Lee

- **user analysis usually produces small files**



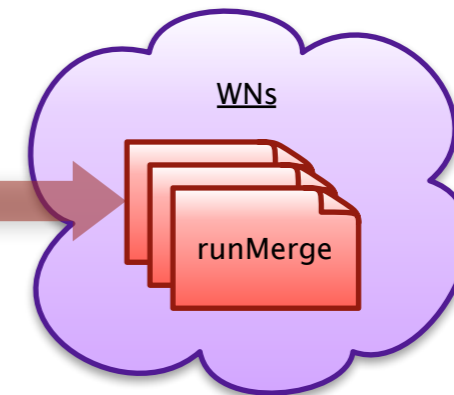
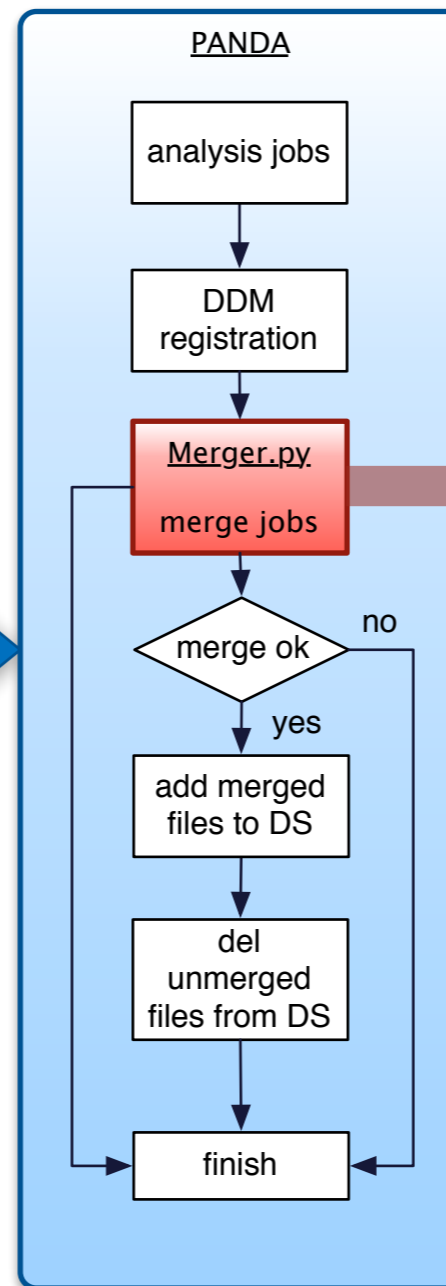
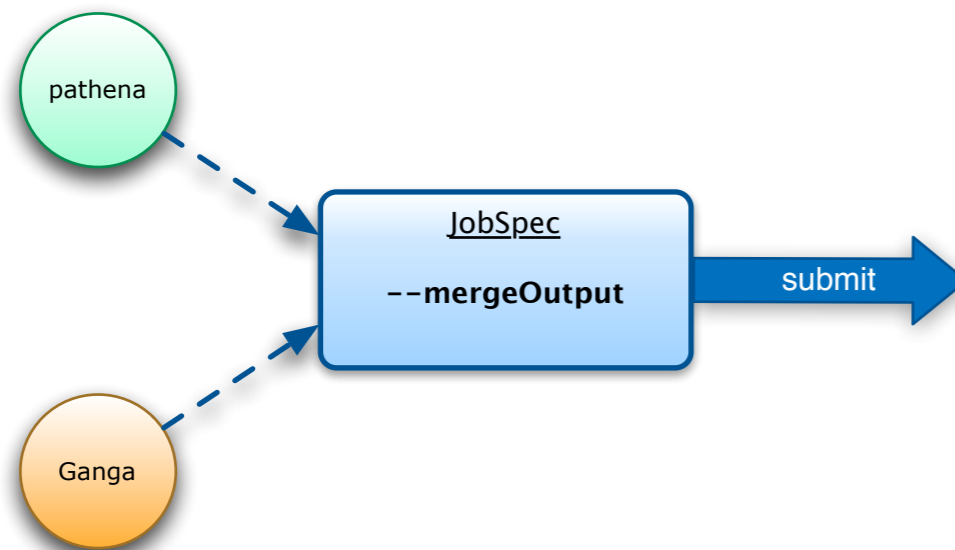
larger file size → smaller transfer overhead

less number of files → faster, easier in data transfer

missing functionality according to the DA survey

2. PANDA creates and chains merging jobs with user analysis jobs

1. user submits analysis jobs with `--mergeOutput` option



3. pilot runs merging transform on WN

4. PANDA deletes unmerged files, place DaTRI requests on merged datasets

- **merging tools**
 - tools/executables doing the merging work
- **merging transform**
 - wrapper for merging tools (similar to runAthena and runGen transforms)
- **job chaining in Panda server**
 - creating merging jobs, cleaning unmerged files, handling the data transfer, etc.
- **job monitoring**
- **client interface**

Challenge: generic merger not possible, merging is content dependent

We start with some simple types

ROOT histogram/NTuple

- **hmerge** (by Johannes Ebke and Peter Waller)
 - python code with extension flexibility
 - supports copy-to-scratch, direct I/O, FileStager

POOL file

- **Merging_trf.py** (from PATJobTransforms)
 - also used by production for “merging” step
 - maintained by PAT group

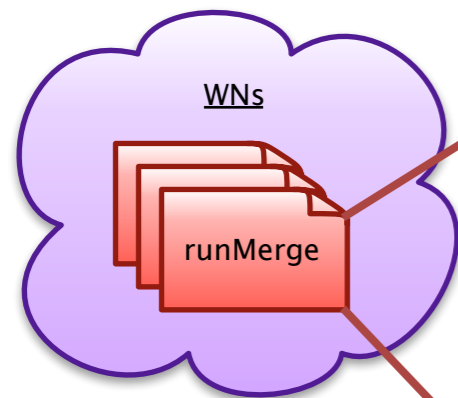
log.tgz / TEXT-based output files

- **tarfile library** (python module)

user type

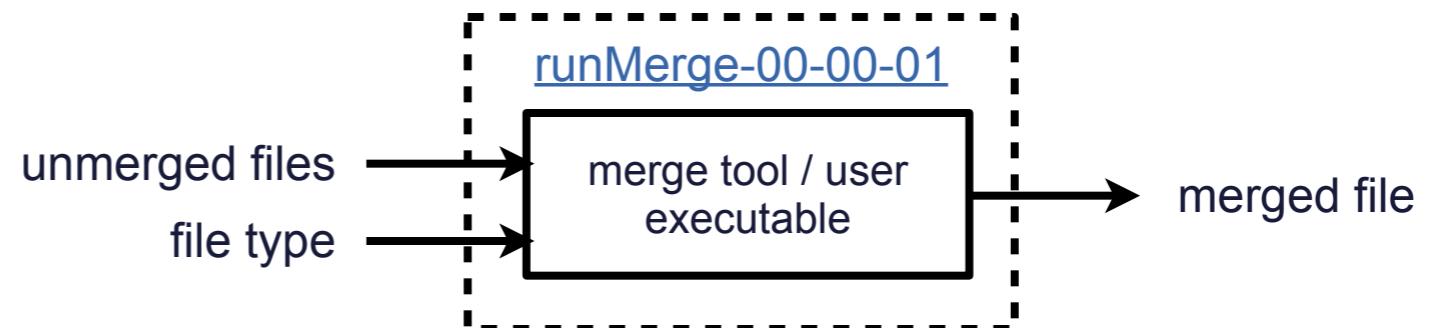
- **user executable** (presented in libDS)

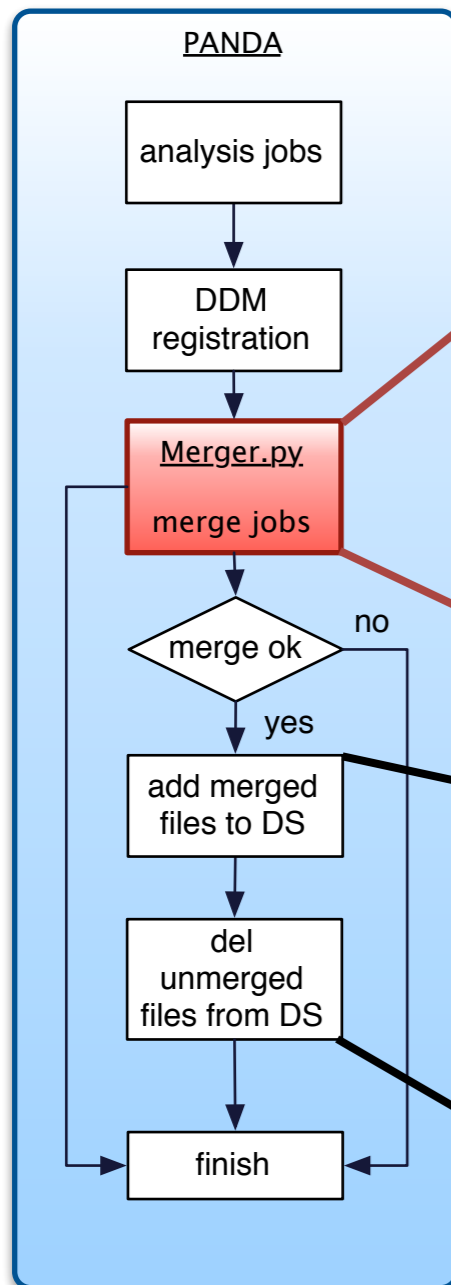
New types? either by extending the tools or by introducing new tools



runMerge-00-00-01

- transform executed by the pilot job
- self-extracting executable (transform + merging tools → one download per job)
- wrapper of merging tools
 - load merge tool according to file type





Merger.py

- merging job creation
 - merging type auto-detection
- merging job splitted by
 1. site (different unmerged dataset location → new merging job)
 2. merging type (different merging type → new merging job)
 3. maxMergedFileSize (> 5 GB → new merging job)
 4. maxNumToBeMerged (> 200 → new merging job)
- merging jobs forced to copy-to-scratch model for inputs
 1. some merging tools don't support other modes
 2. input size is limited

output management

- “*.merge” datasets created (in the same output dataset container) for the merged files
- unmerged files get deleted after they are successfully merged
- automatic DaTRI request issued when first merging job finished
- DaTRI request ONLY on “*.merge” datasets

merging job retry will be added

in case that part of the merging jobs failed ...

```

% dq2-ls -r user.hclee.20110331224155.37/
user.hclee.20110331224155.37.ANALY_TOKYO.merge:
    INCOMPLETE:
    COMPLETE: TOKYO-LCG2_SCRATCHDISK
user.hclee.20110331224155.37.ANALY_TOKYO:
    INCOMPLETE:
    COMPLETE: TOKYO-LCG2_SCRATCHDISK

Container name: user.hclee.20110331224155.37/
Total datasets: 2
Summary:

```

SITE	/ # COMPLETE	/ # INCOMPLETE	/ TOTAL
TOKYO-LCG2_SCRATCHDISK	2	0	2

- dataset contains merged files
- dataset to be replicated to final destination

- dataset contains unmerged files that are not able to be merged

... all outputs expected by user are still available via container

Summary of all jobs for all days, jobsetID 9972 in any state at any site plot Go

Retrieve All

14 jobs. Click job number to see details.
 States: finished:13 failed:1
 Users (1): [Hung-Chun LEE:14](#)
 Releases (1): [Atlas-16.6.2:14](#)
 Processing types (2): [ganga:1](#) [usermerge:2](#)
 Job types (1): [user:14](#)
 Transformations (2): [runAthena-00-00-1:12](#) [runMerge-00-00-01:2](#)
 Working groups (1):
 Sites (1): [FR.ANALY_TOKYO:14](#)
 Regions (1): [FR:14](#)
 Clouds (1): [FR:14](#)

Showing 1 jobsets modified from 2011-03-31 21:01 to 2011-03-31 21:22

User:jobID	Created	Latest	Jobs	Pre-run	Running	Holding	Finished	Failed	Cancelled	Merging	buildJob	Site
Hung-Chun LEE:9972	2011-03-31 20:22	2011-03-31 20:22	14		end		13	1		2		ANALY_TOKYO
In: data10_7TeV_00152409.physics_MinBias.merge.AOD.f238_m427 Out: user.hclee.20110331222236.35/												

processingType=usermerge

Showing 2 jobs modified from 2011-03-31 21:01 to 2011-03-31 21:25

PandaID, Owner, Working group	Job	Status	Created	Time to start	Duration	Ended/ Modified	Cloud/Site, Type	Pr
1214819115 Hung-Chun LEE	jobsetID= 9972 runMerge-00-00-01	failed	2011-03-31 21:01	0:00:24	0:03:38	03-31 21:05	FR/FR.ANALY_TOKYO , analysis-run	99
Error details: pilot: Put error: lfc_setfilesize failed for lfn: /grid/atlas/users/pathena/user.hclee/user.hclee.20110331222236.35.ANALY_TOKYO.merge_sub018247522/user.hclee.009972_1214812932.merge.log.rc -1, err_num 22, errstr Invalid argument								
In: data10_7TeV_00152409.physics_MinBias.merge.AOD.f238_m427 Out: user.hclee.20110331222236.35/								
1214819114 Hung-Chun LEE	jobsetID= 9972 runMerge-00-00-01	finished	2011-03-31 21:01	0:00:24	0:03:39	03-31 21:05	FR/FR.ANALY_TOKYO , analysis-run	99
In: data10_7TeV_00152409.physics_MinBias.merge.AOD.f238_m427 Out: user.hclee.20110331222236.35/								

Run jobs in this job set: [1214812932](#) [1214812933](#) [1214812934](#) [1214812935](#) [1214812936](#) [1214812937](#) [1214812938](#) [1214812939](#) [1214812940](#) [1214812941](#) [1214812942](#) [1214812943](#)

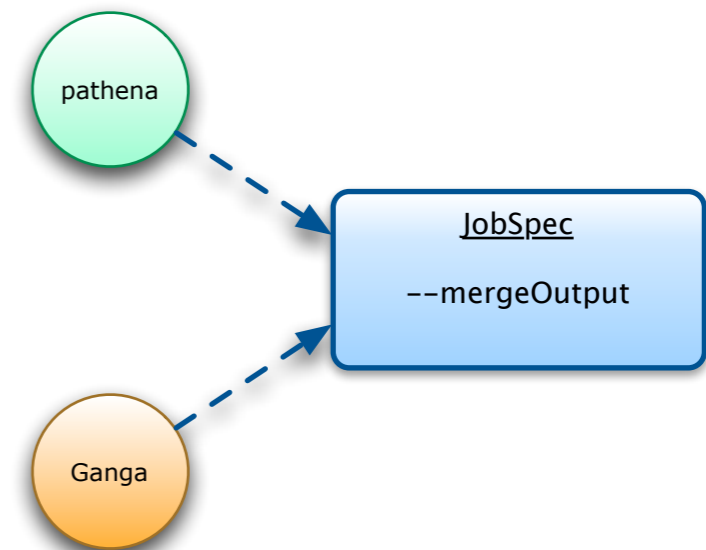
Merging jobs in this job set: [1214819115](#) [1214819114](#)

- **Ganga**

- `PandaRequirements.enableMerge = True`
- to be available in next major Ganga release 5.6.0

- **pathena**

- `pathena --mergeOutput ...`
- available



Still experimental functionality, merging is by default disabled

functional tests

- job chaining
- situation with part of merging jobs failed
- situation with multi-site job splitting
- situation with remote destinationSE

stress tests (on-going)

- test template in HammerCloud
- different cases with merging types
- tests in different clouds

- **analysis jobs produce small output files → output merging can save cost in transferring and managing small files**
- **output merging mechanism is ready in Panda and client tools → stress tests with HammerCloud to gain confidence for commissioning**
- **merging is type specific → improvement by feedback → better merging tools are welcome**



Acknowledgement

- **Tadashi Maeno**
- **Paul Nilsson**
- **Valeri Fine**
- **Johannes Ebke**
- **Peter Waller**
- **Mark Slater**
- **Daniel van der Ster**
- **Johannes Elmsheuser**
- **Alexander Lincoln Read**