MC production in ILD

ILC-JP end-of-year physics and detector meeting

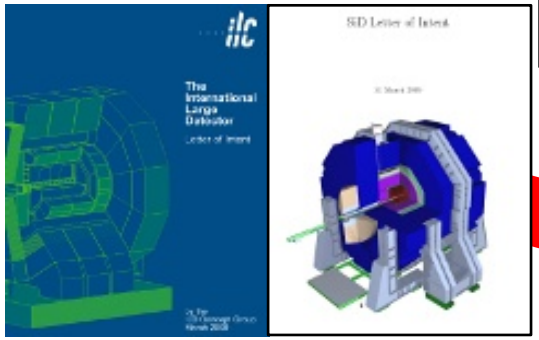
@KEK

Hiroaki Ono (NDU), Akiya Miyamoto (KEK)

MC sample production for document preparation

Large MC sample production has organized for detector benchmarking analysis and physics study (Shared by KEK, DESY, SLAC previously)
→ Now its produced by ILD MC production group for ILD

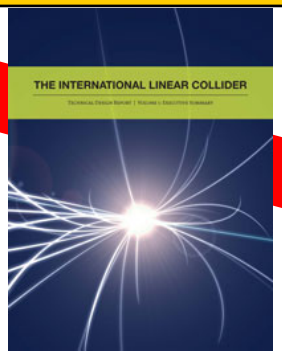
LOI 2010



Detector concept study

LOI sample (ILD, SiD)
Full sim SM sample
250, 350, 500 GeV

DBD/TDR 2013



Physics performance with full simulation 250 – 1000 GeV

DBD sample (ILD, SiD)
Full Sim SM smaple
250, 350, 500, 1000 GeV

Snowmass P5 2015

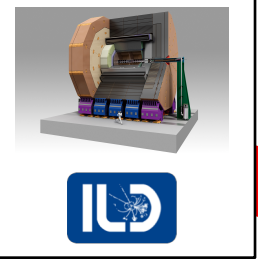
Physics performance from LHC results

Higgs, top, NP
250, 350, 500 GeV

ILD MC production group

IDR

On going



ILD detector benchmarking

Two detector model
L5 and S5 (250, 500 GeV)
Full detector simulation

MC production and Data share LCG GRID

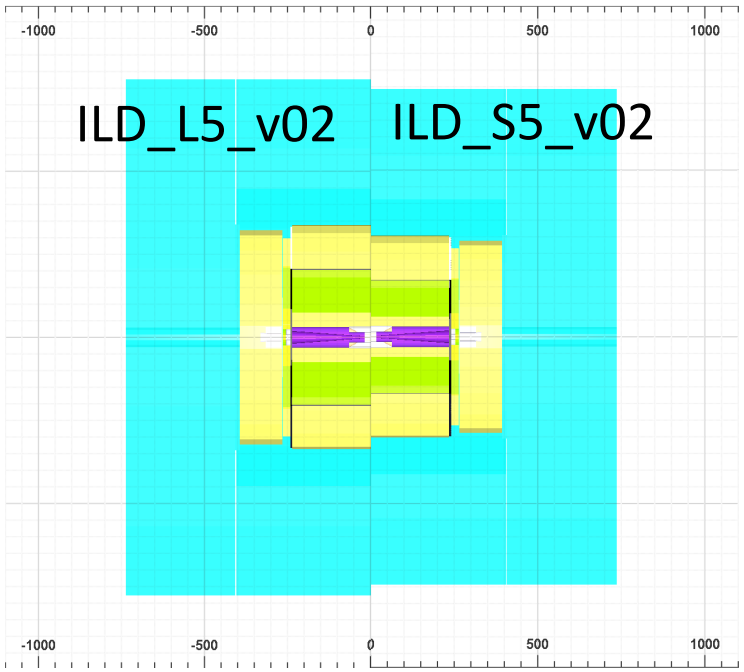
ILCDirac usr job

ILCDirac production job

Detector benchmarking study

ILD detector optimization by two detector models with large MC samples

- **Large(L5) and small(S5)** size with detailed detector description (DD4HEP)
- Hybrid ECAL with Analog/Semi-Digital HCAL (SiW+AHCAL for o1 option)
- Use latest reconstruction software codes (iLCSoft, LCFIPlus, Tracking etc)



Planned MC sample Production

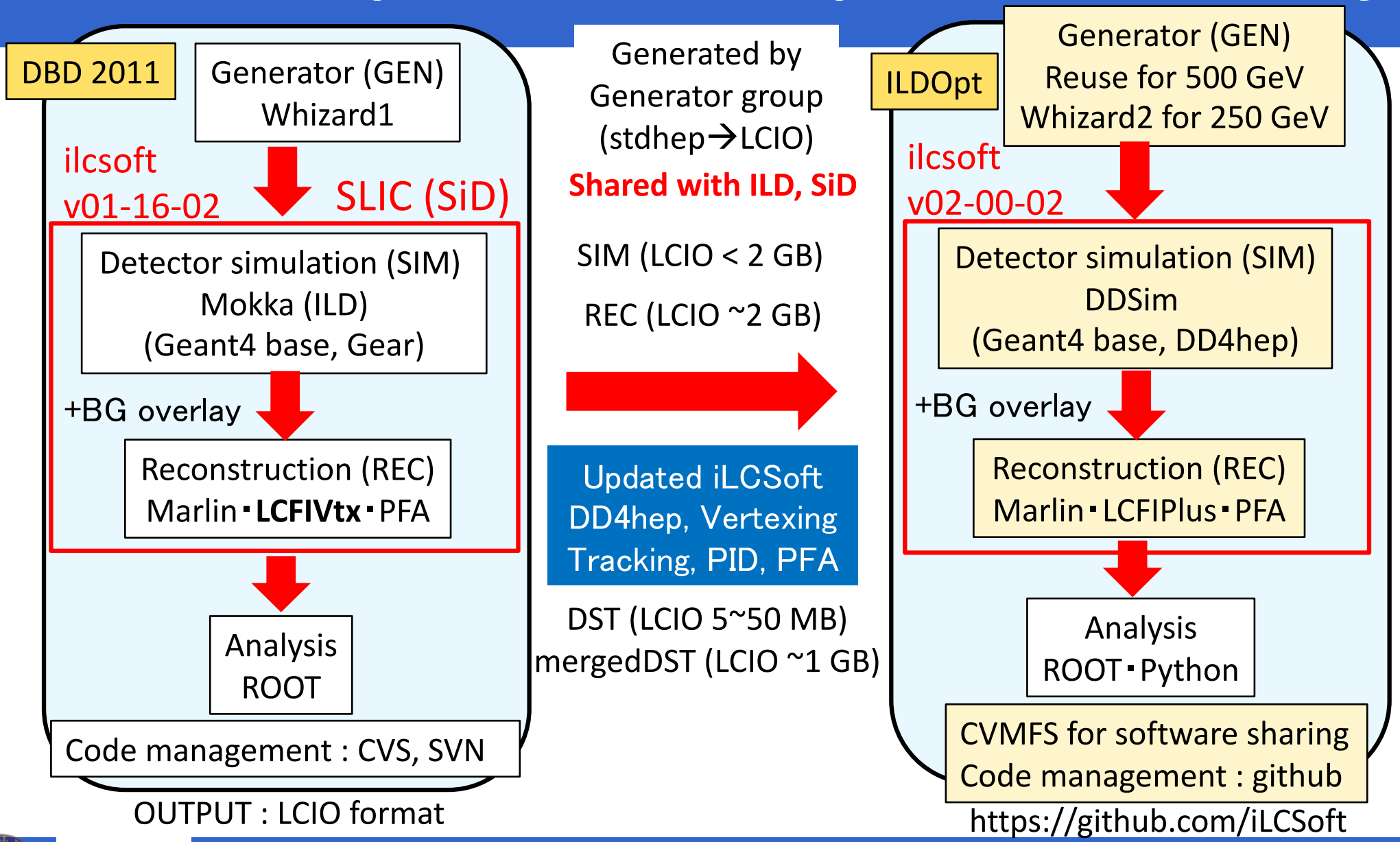
1. 500 GeV samples produced with DBD GEN samples

- **SM** and **calibration sample** (single, uds, flavortag)
- Same or larger statistics with DBD
- Overlay aa_lowpt + pair BG
- Correct IP smearing and vertex offset

2. 250 GeV samples with whizard 2

- Will be used new beam parameter

Software update on ILD optimization study



DIRAC and ILCDirac for large scale production

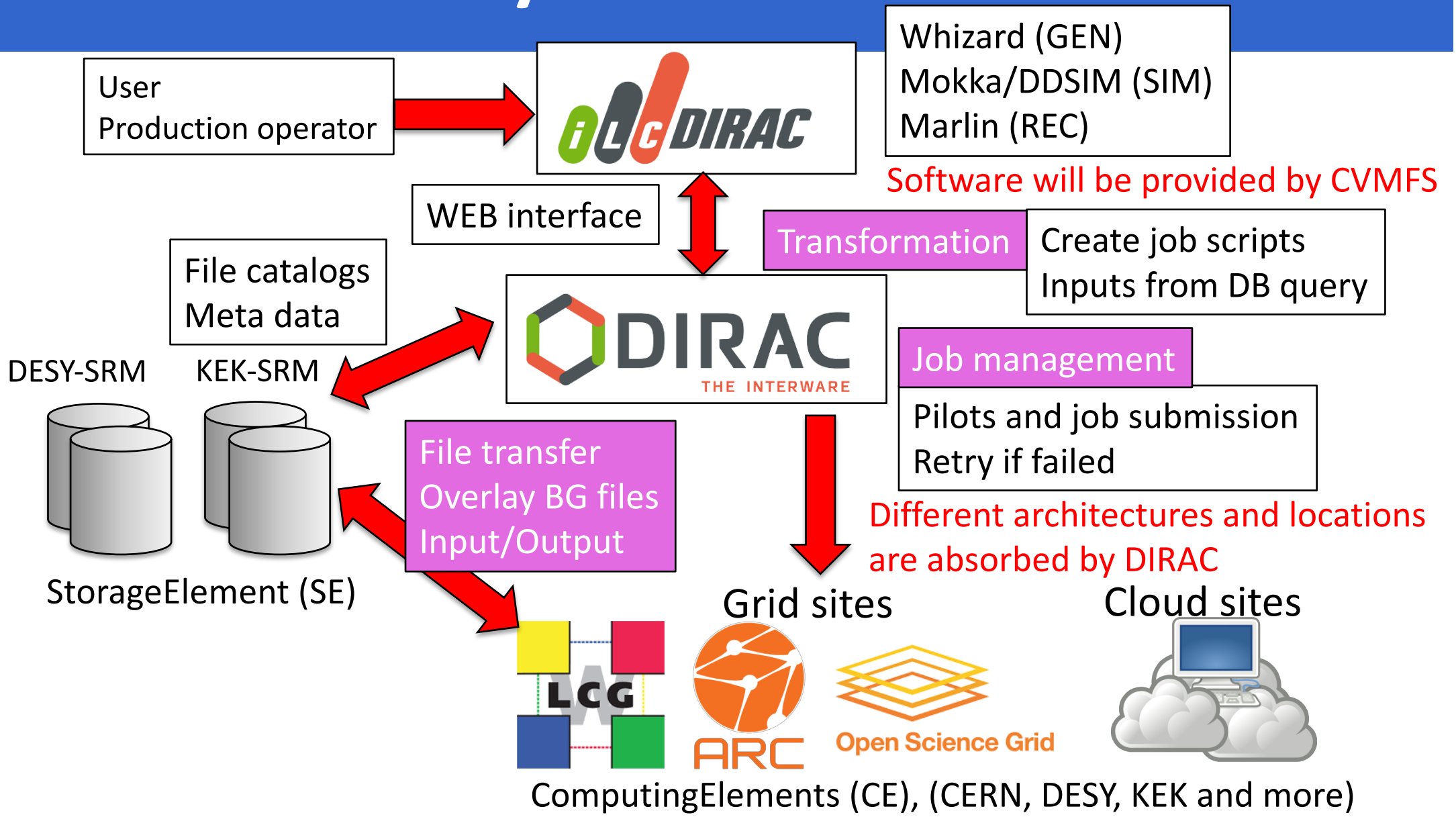
DIRAC (**D**istributed **I**nfrastructure with **R**emote **A**gent **C**ontrol)

- Initially produced for LHCb but now used with extension for each collaboration
  →ILCDirac for ILC and BelleDIRAC for Belle II collaboration
- Software framework and interface to access distributed resources by user
  - Job managements
  - File catalog
  - Transformation system for large scale productions
  - Written in Python 2
  - Web interface

ILCDirac : Extension of DIRAC for ILC

- ILCDirac is developed and maintained by CLICdp group since 2010.
  - Production jobs with iLCSoft SIM/REC (whizard/Mokka/DDSim/Marlin)
  - CERN (CLICdp) hosts and manages DIRAC servers and DBs
- **ILD start to use ILCDirac after the DBD production**
  - Used for large scale of MC production with efficient resource usage

DIRAC system and ILCDIRAC



ILD specific feature on ILD production

File naming rule on ILD

ex: /ilc/prod/ilc/mc-opt-3/ild/rec/500-TDR_ws/2f_Z_leptonic/ILD_o1_v05/v01-16-p05_500/rv01-16-p05_500.sv01-14-01-p00.mILD_o1_v05.E500-TDR_ws.I250108….

- Actually file catalog has meta info, thus not need to put all software info in file name but from the naming rule in ILD, prepared utility to follow filename convention on ILCDirac
- Name limitation less than 128 characters from DB restriction

Generator samples

Generated by ILD generator group with validation (Not processed on DIRAC by MCProd group)

- Use DBD generator sample for 500 GeV production but make new sample for 250 GeV

Merged DST file

Keep order of DST file serial number in merged file serial number

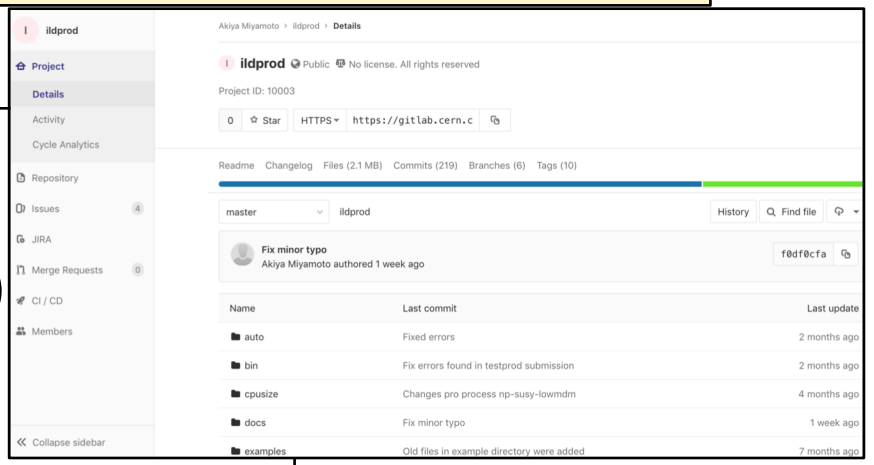
Logging

All log should be placed on accessible location (DIRAC stored all logs on DB)

- Retrieve all production logs and upload to DESY-SRM
- Use ELOG system for summary (hosted on NDU https://ild.ngt.ndu.ac.jp/elog)

ILD MC production tool (ILDProd)

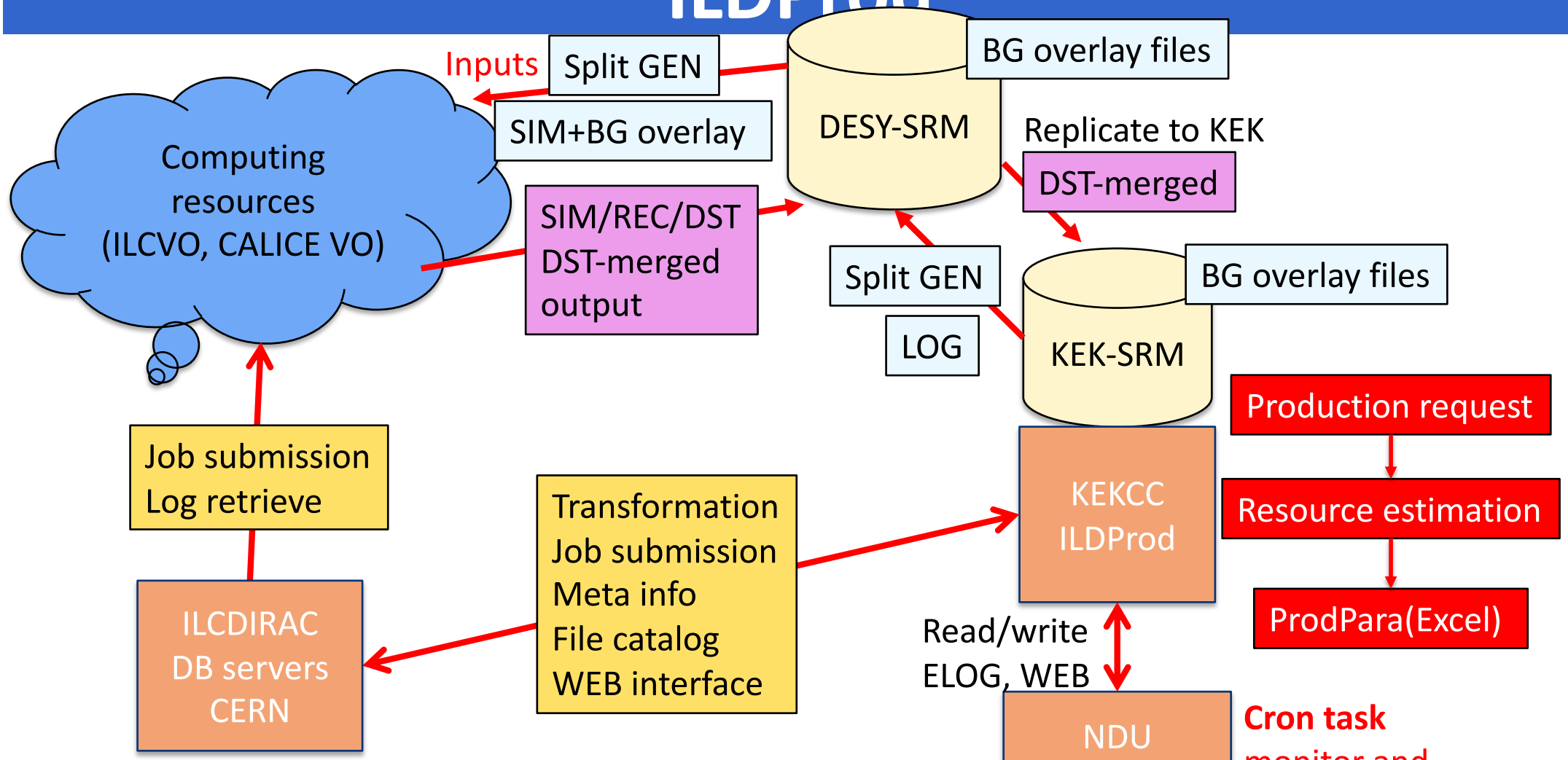
ILDProd tool is developed and updated by Miyamoto-san for ILD MC production
https://gitlab.cern.ch/amiyamot/ildprod/ (v02-00-02)



- Python and shell script base
- Compilation of production utilities
  - Generator sample splitter (stdhep splitting)
  - Resource estimation tool (CPU/disk size)
  - Utility of read/write meta information
  - Helper utility of preparing DIRAC transformation script
  - ELOG modifier with python
  - DST merge utility with user job
- Automatize many manual handling process (Cron base)
- Automatic logging to ELOG and WEB summary
- Production manual documents

Efficient production with removing many baby sitting work

Rough scheme of MC production with ILDProd



Resource estimation for ILD opt study before production (500 GeV sample)

Resource estimation have performed before launching production jobs (Performed at KEKCC batch system)

				CPU days		Data size(GB)		
Process	Nb.Procs	k Evts	Nb. Jobs	SIM	REC	SIM	REC	DST
uds	12	120	430	30	13	385	379	6
single	94	1,520	589	37	36	382	307	4
higgs	32	951	6,189	322	243	4,367	4,417	72
2f	8	3,780	16,475	1,198	706	14,386	13,740	215
4f	40	11,289	48,219	3,108	1,726	33,369	34,721	475
5f	200	2,029	7,627	520	342	6,175	6,442	116
6f	188	6,907	530,029	2,900	2,564	38,456	40,506	725
aa_4f	80	816	2,344	158	115	1,794	2,042	41
flavortag	5	535	5,329	290	262	3,943	4,142	76
Sum	659	27,947	617,231	8,563	6,007	103,257	106,696	1,732

2 detector model with calibrations (~23 HS06 at KEKCC one CPU)

Total for 500 GeV	CPU(HS06 year)	Storage (TB)
2 Size x 1 Cal	1,882	423

Disk arrangement for MC-Opt production

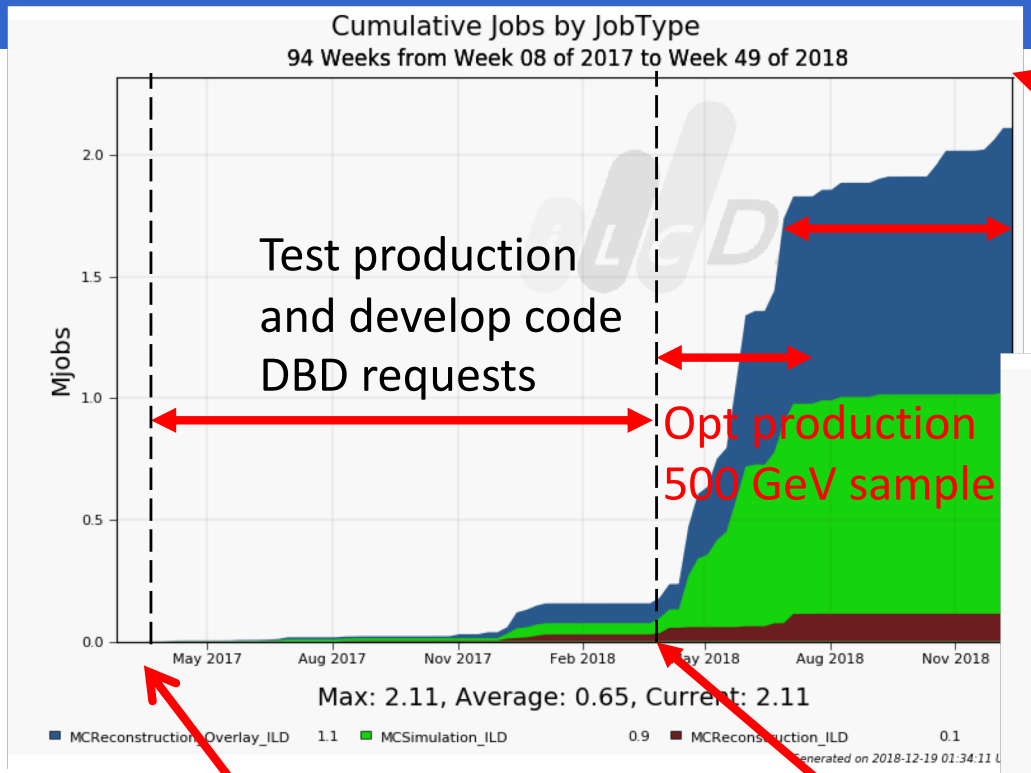
Disk arrangement policy

- Generator files used for production should locate to KEK-SRM for preparation
- All produced files (SIM/REC/DST/LOG) are written to **DESY-SRM**
- **All DST-Merged files** are replicated to **KEK-SRM** for analysis, redundancy
- **Split GEN/SIM/REC/DST files will be removed for effective space usage**

Types	Location	Actual media location
Split stdhep	/ilc/prod/**mc-opt.dsk**/ild/gensplit	DISK on DESY
PROD	/ilc/prod/**mc-opt-3**/ild/|sim|rec|dst-merged]	TAPE
LOG	/ilc/prod/**mc-opt.dsk**/generated /ilc/prod/**mc-opt.dsk**/ild/[simlog/reclog/dstmlog]	DISK on DESY
GEN	/ilc/prod/**mc-dbd**|**mc-opt-3**/generated	TAPE
DST	/ilc/prod/**mc-opt.dsk**/ild/dst	DISK on DESY

/ilc/prod/ilc/mc-dbd is also used for analysis using DBD tools reconstruction (iLCSoft v01-16-02)

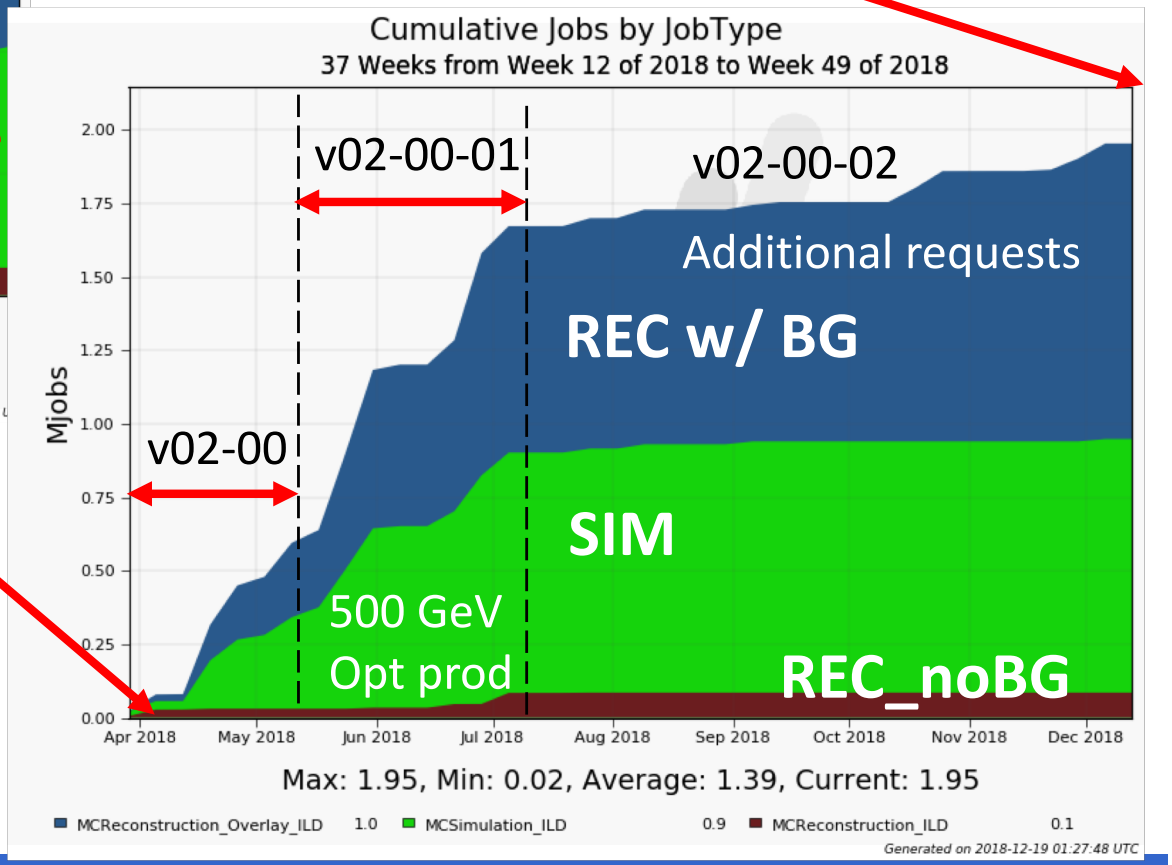
Cumulative jobs for ILD Opt production



Test production
and develop code
DBD requests

Opt production
500 GeV sample

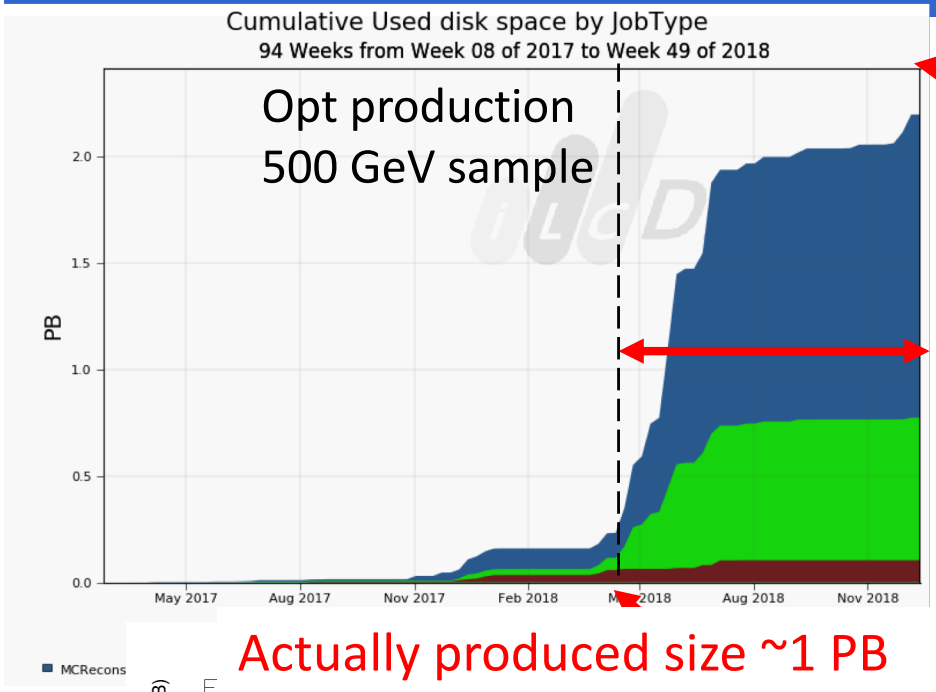
Avg. 2k jobs/hour run

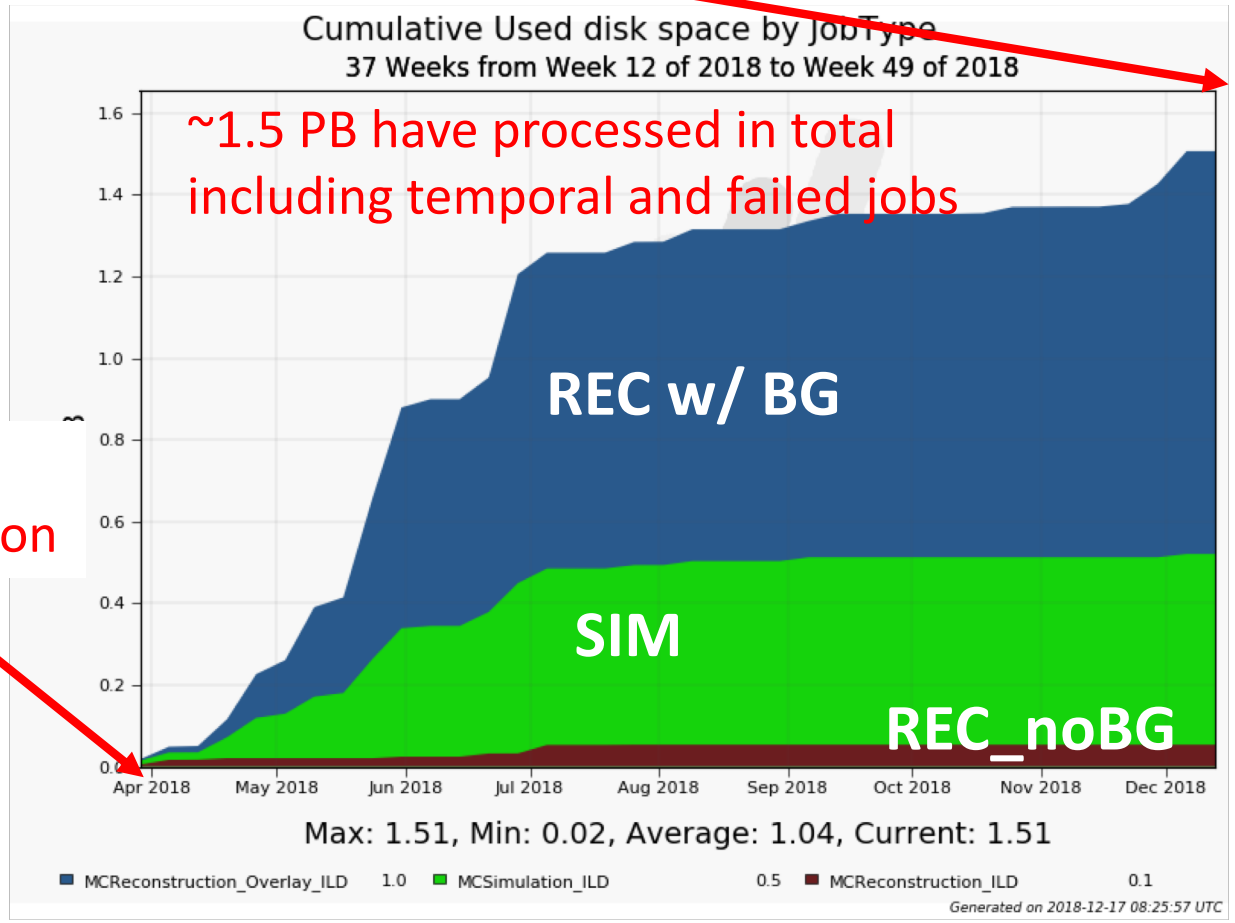


MC production group has organized
and start to use ILCDirac and

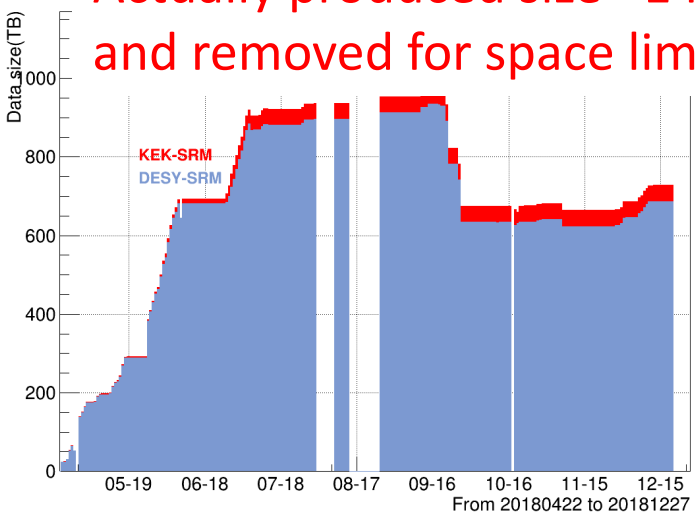
~2 Mjobs have processed
for ILD production in this year

Cumulative Disk usage for ILD opt production





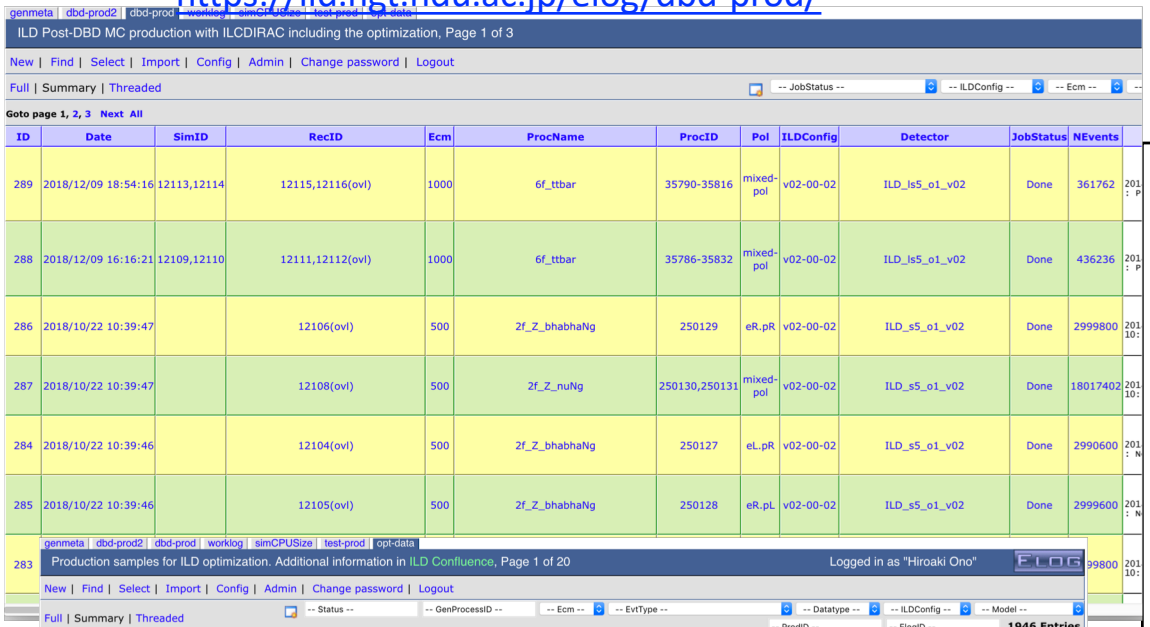
Actually produced size ~1 PB
and removed for space limitation



Production summary on ELOG, WEB

Production information is summarized on ELOG and WEB

https://ild.ngt.ndu.ac.jp/elog/dbd-prod/



https://ild.ngt.ndu.ac.jp/elog/opt-data/

Generator/SIM/REC/DST-merged relation summarized with production logging in ELOG

Summary of samples of elogid from 183 to 289

Total events: 161,034,736 planned/ 160,819,315 produced/ 176,502,366 available (**processed 99.9 %**)
- summarized based on the elog data for dst-merged events of ILDConfig v02-00-01,v02-00-02
- default detector model is ILD_l5/s5_o1_v02. Other model is indecated in () after elog id
- produced/processed does not count events produced more than plan.
-">" in the field "Processed" indicates that the nb. of produced events is larger than the plan in some processes.

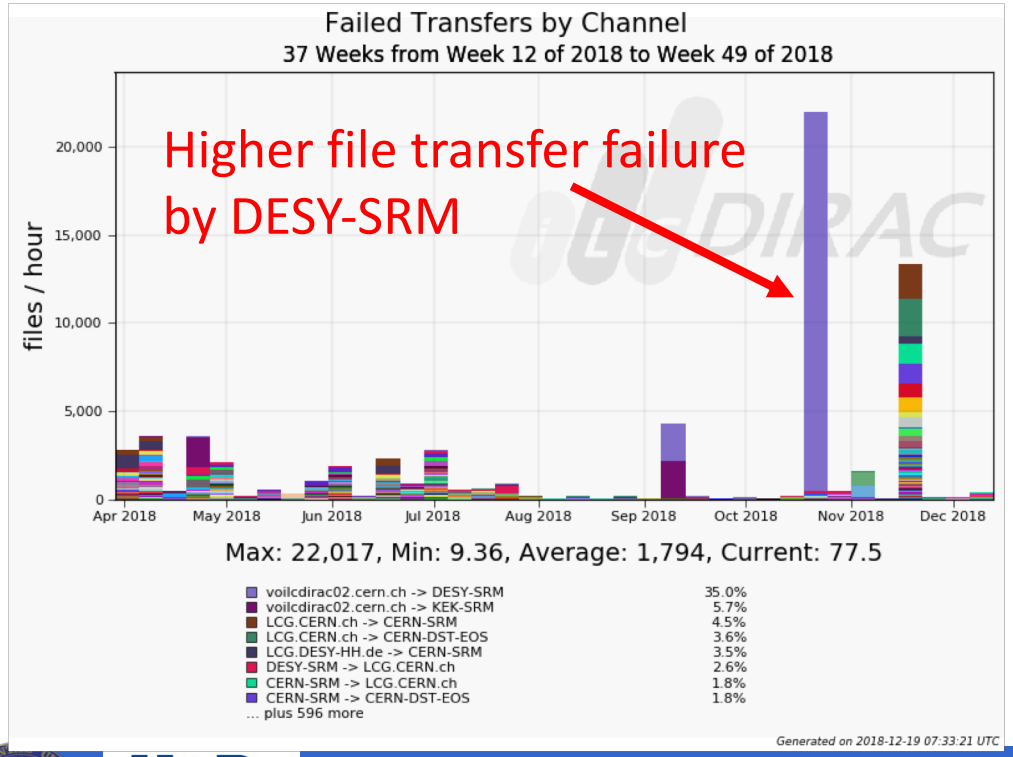
(created on 2018/12/17 16:30:02)

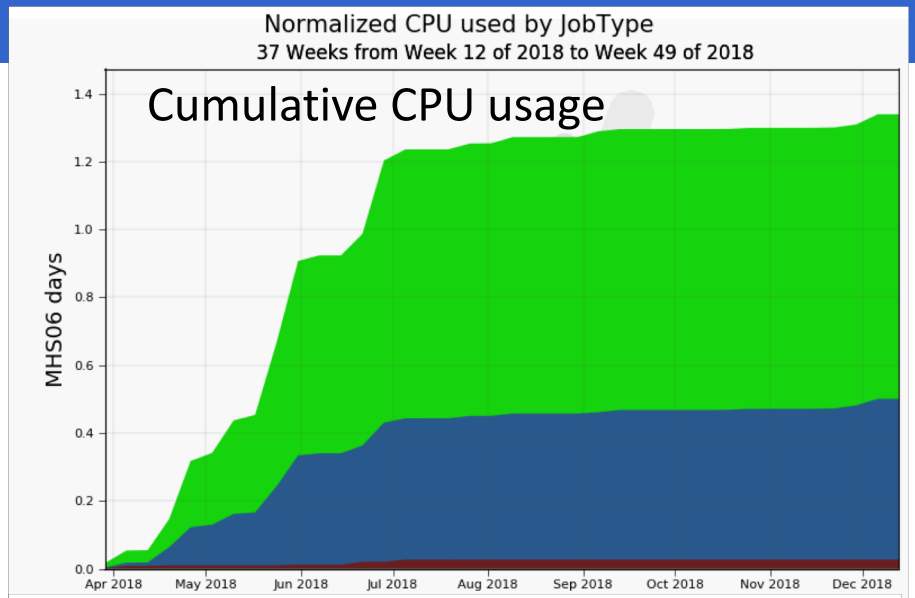
E1-calib : EvtType	Nev plan	Nev prod.	Processed(%)	ElogIDs
resonance	200,000	200,000	100.0	264(nobg),268(o2,nobg)
single	6,080,000	21,731,663	99.8	259(nobg),261(o2,nobg)
uds	480,000	480,000	100.0	260(nobg),262(o2,nobg)
E91-nobeam : EvtType	**Nev plan**	**Nev prod.**	**Processed(%)**	**ElogIDs**
flavortag91	693,126	690,626	99.6	258(nobg)
E500-TDR_ws : EvtType	**Nev plan**	**Nev prod.**	**Processed(%)**	**ElogIDs**
2f_Z_bhabhaNg	36,000,108	35,987,471	100.0	241,242,243,246,283,284,285,286
2f_Z_bhabhag	1,362,964	1,362,964	100.0	208
2f_Z_hadronic	4,766,778	4,703,534	98.7	219
2f_Z_leptonic	1,431,166	1,430,566	100.0	183
2f_Z_nuNg	54,055,374	54,052,206	100.0	240,287
4f_WW_hadronic	785,534	785,434	100.0	216
4f_WW_leptonic	307,226	303,226	98.7	193
4f_WW_semileptonic	999,820	999,820	100.0	196

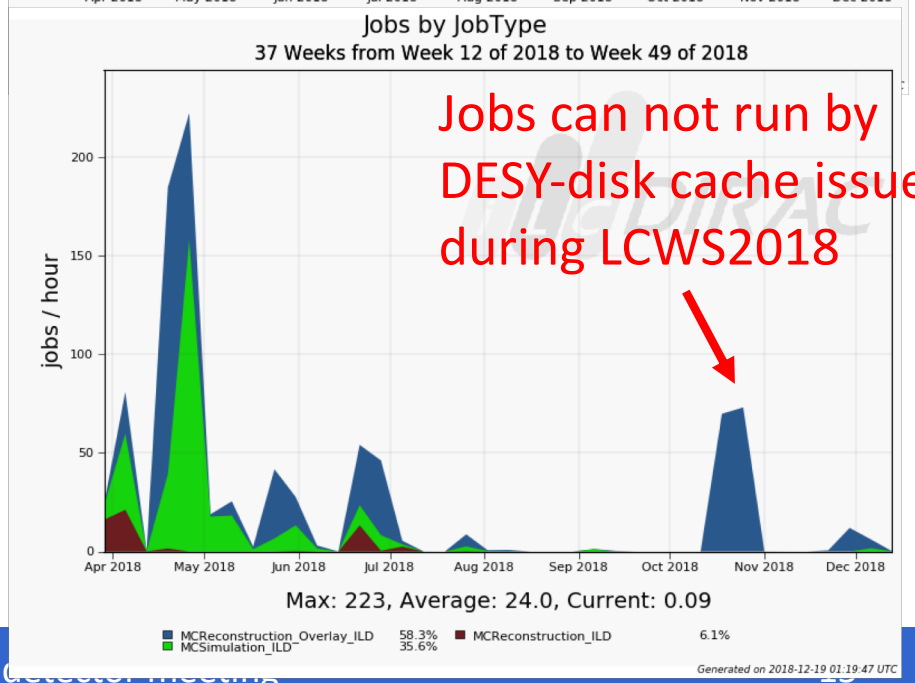
https://ild.ngt.ndu.ac.jp/mc-prod/prodmon/summary-by-evttype.html

Failed job rate in production period

Failure job rate during
the production period from Apr. 2018







DESY disk system access issue

Frequent access time out or disk full failure has happened at read/write stage to DESY-SRM

DESY-SRM (HSM system same as KEK)

Stage-in to cache disk if we access old files

ILD production : all output write to DESY disk
→Jobs have frequently failed at first Input or final upload stage due to cache disk full or access timeout

Read

Access from jobs

Write

Cache disk

~100 TB

Stage-in

Much slow

Stage-out

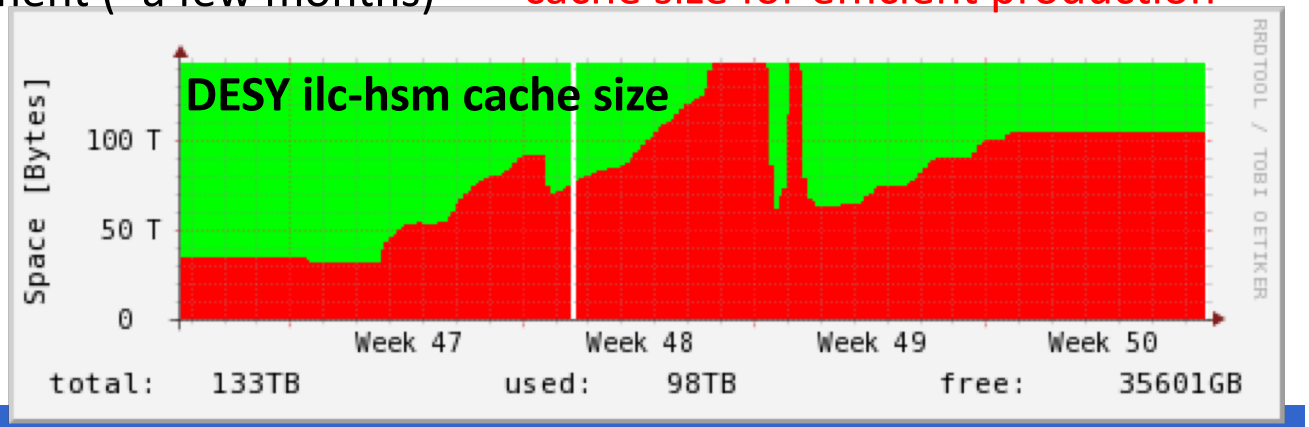
TAPE backend

Stage-out from oldest files if cache disk becomes full (Or exist both by timing)

Re-reconstruction request has failed because of **staging timeout of large amount of SIM files**
→Separate (Chunk) "Prestaging" should request to DESY admin… (No automatic request mechanism)

Took much longer time for disk arrangement (~a few months) than production job itself (only a week to finish all jobs)

Need to monitor and care cache size for efficient production



At KEK, there is a mechanism to request automatic stage-in by user (Useful for analyzer too)

DESY-SRM d-cache situation monitor

So far, we need to monitor the DESY d-cache situation before launching production jobs

DESY admin provided following d-cache pool monitor (Only accessible from DESY internal)

PoolGroup	Total Space/MB	Free Space/MB	Precious Space/MB	Layout (precious/cached/free)
afs-hsm	140364785	18268	10136360	
all	51931498	7710794	44220031	
belle	380094536	68722586	301006755	
belle-1	197548933	5079146	192469786	
belle-hsm	5515042	3194858	0	
belle-local	122046441	97399606	24646834	
calice	92330932	16911581	75392076	
calice-hsm	55189132	13138	22282	
dcache-build	6122589	3703316	2388462	
default-pools	33863431	45097	0	
dphep	715664826	13736666	701215892	
dphep-hsm	12197816	8906	0	
dphep-scratch	24715361	636944	24072761	
eltest	12197816	12197816	0	
ilc	116204459	16935281	99194520	
ilc-hsm	136436668	36455050	4365429	
lhcb	2040863	2040863	0	
linac-test	327315859	10865	0	
olympus-hsm	10643005	4240536	0	
test-hsm	18252799	20188	10122953	
testbeam	79932598	28668320	51264278	

Request
pre-staging
to DESY d-cache admin
before using TAPE files
or write from production

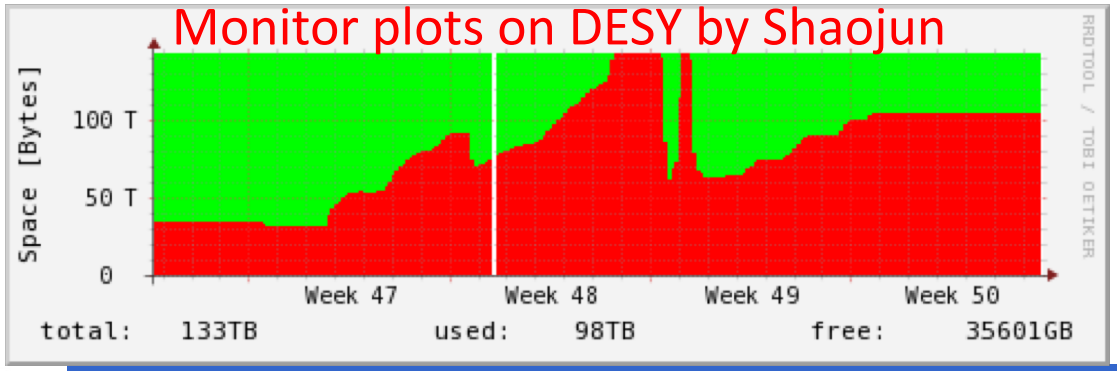
Monitor on NDU

dCache space monitor

Last 20 records monitored by PoolManager

Date	Time	Free space(MB)	Precious space(MB)
20181219	081805	36455050	4365429
20181219	071804	36455050	4365429
20181219	061804	36455050	4365429
20181219	051805	36455050	4365429
20181219	041804	36455050	4365429
20181219	031805	36455050	4365429
20181219	021804	36455050	4365429
20181219	011805	36455050	4365429
20181219	001805	36455050	4365429
20181218	231804	36455050	4365429
20181218	221804	36455050	4365429
20181218	211805	36455050	4365429

Monitor plots on DESY by Shaojun

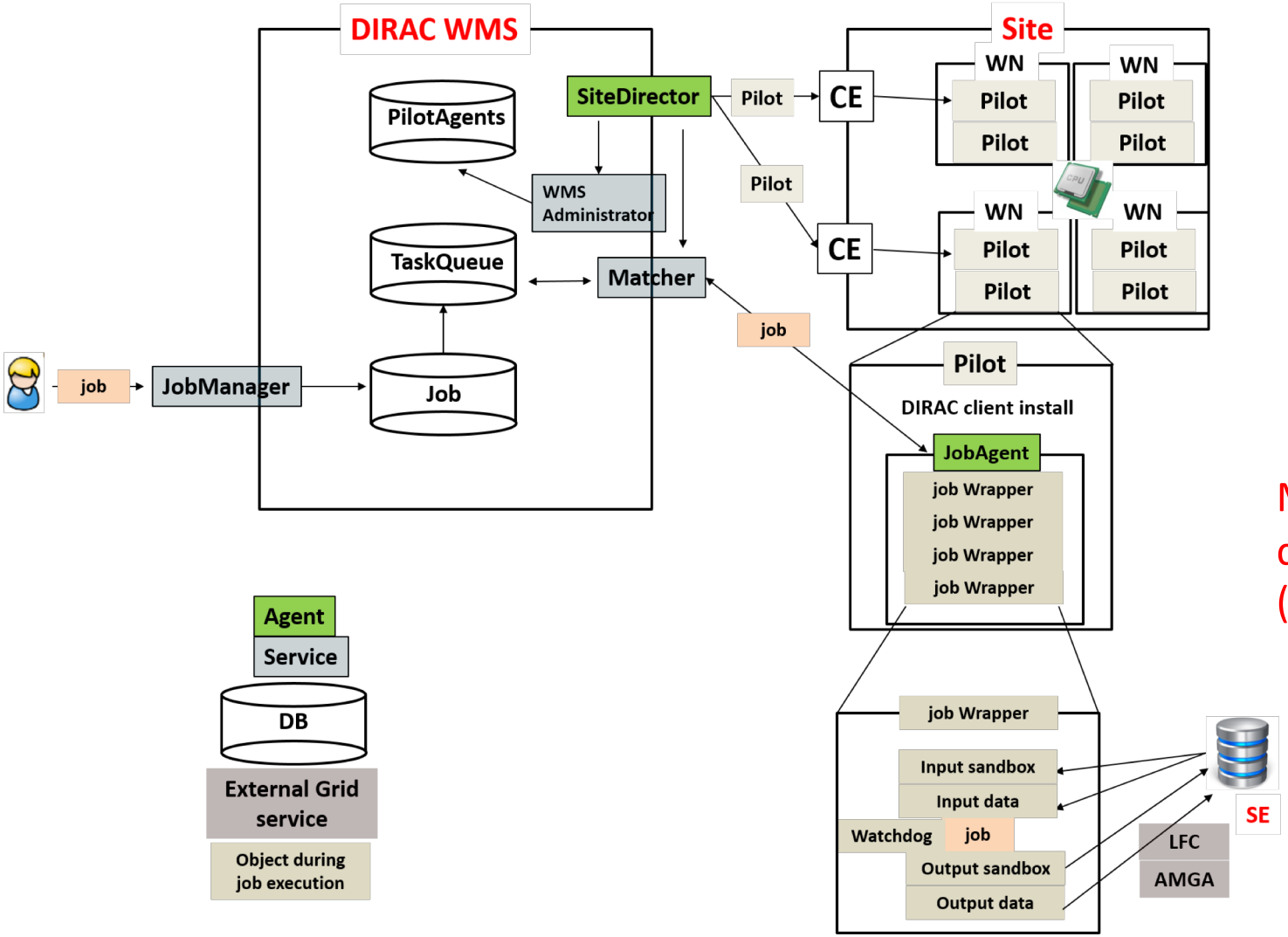


Summary

- ILD MC production is launched since 2018 Apr for the detector benchmarking study
- 500 GeV common sample production has finished
- Additional 1 TeV benchmarking process has also processed
- Preparation for 250 GeV sample production with updated ILD production tools.

BACKUP

DIRAC job submission to sites



Pilot jobs identified the site/WN is suitable to run that jobs

Monitor the job/WN situation during running jobs (DiskIO/Mem size/disk size)

Finally upload outputs to StorageEelment (SE)

Made by Belle II DC group