

# Evolution of production tools

---

Luca Tomassetti

University of Ferrara & INFN

SuperB Distributed Computing Group



# Summary

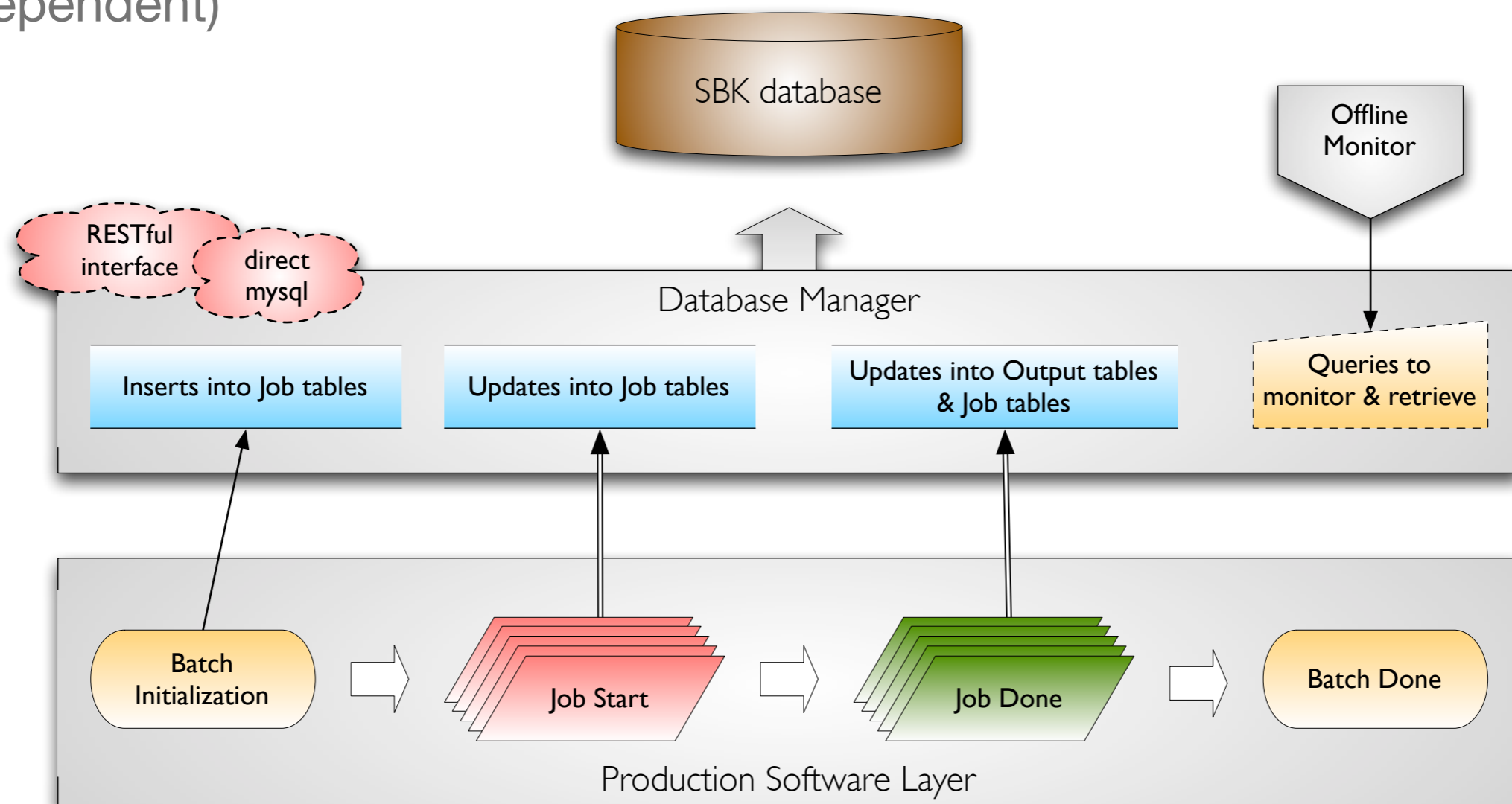
---

- Production Tools:
  - WebUI
  - Bookkeeping Database
  - Job scripts / Submission scripts
  - Monitoring jobs and production
- February Production (issues)
  - Sites
  - Submission
  - Configuration

People involved:

E. Luppi	Ferrara
L. Tomassetti	Ferrara
A. Fella	CNAF
G. Fontana	Ferrara
M. Ronzano	Ferrara

- **Web based User Interface**
- PHP, javascript, JQuery framework for AJAX functionalities
- database classes for communication with bookkeeping database (RDBMS independent)



# Production tools

WebUI

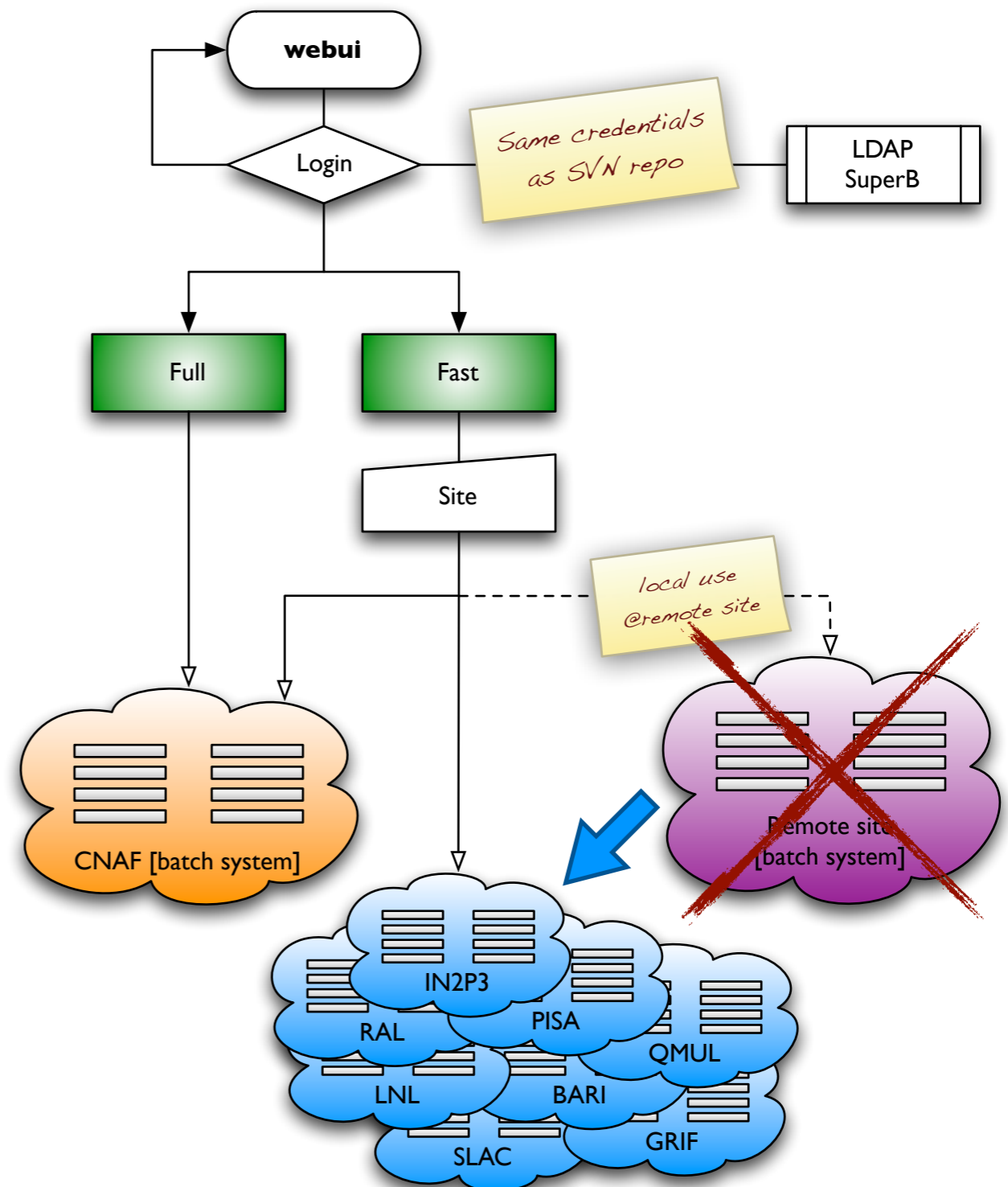
- **Production Submission Workflow**
- basic authentication and authorization based on common LDAP directory service

- FullSim / FastSim differentiation

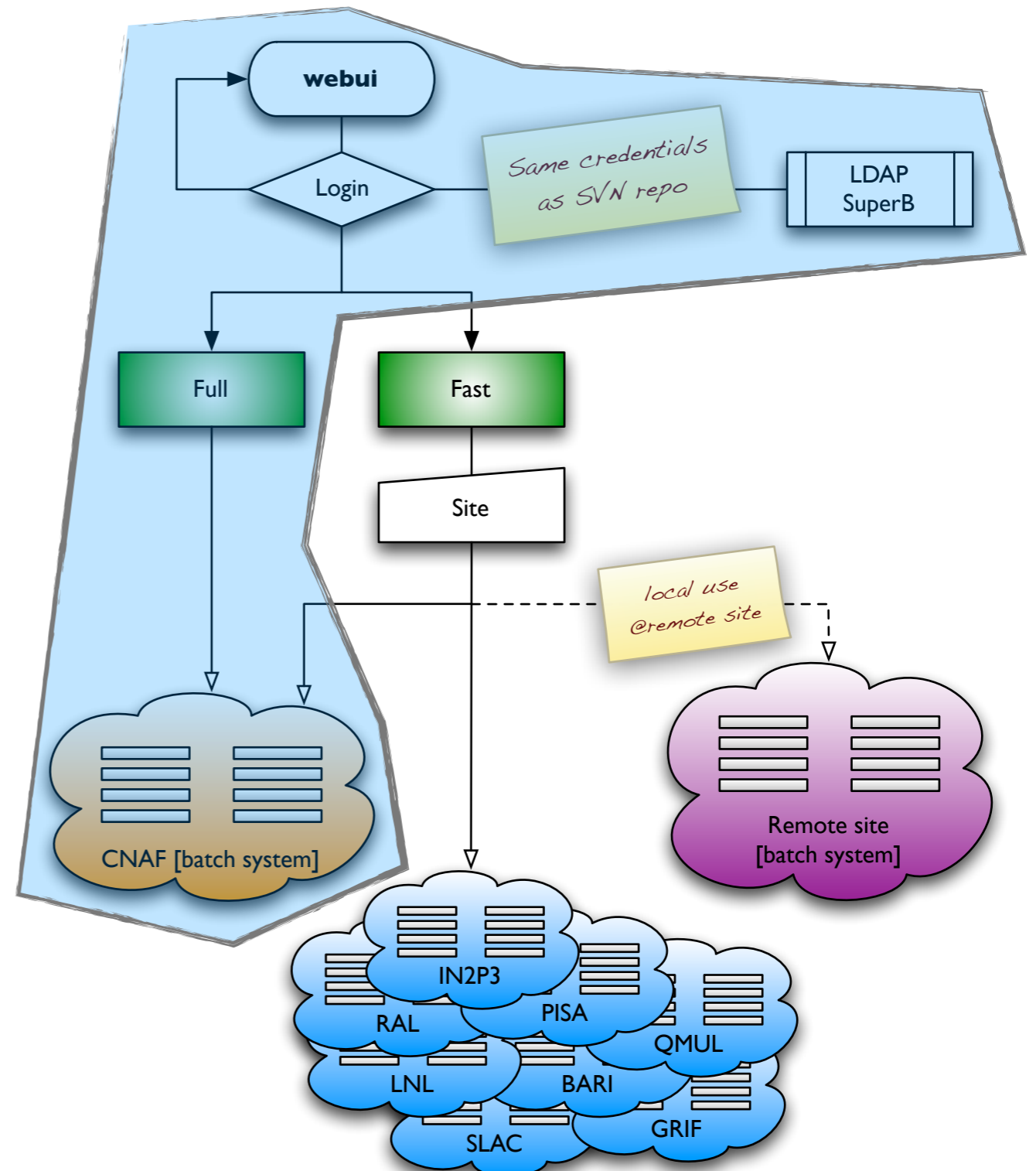
- Batch System / Distributed (Grid) submission

August milestone!  
Already done!!

- Possibility to 'export' the WebUI to remote sites for local usage (on batch system, requires a dedicated machine, ...)



- **FullSim**
- LSF Batch System submission @CNAF only
- User (production manager) choices:
  - Geometry
  - Generator (and parameters)
  - Physics List
  - type (full / bgframes)
  - Number of events per job
  - Number of jobs (+ batches)



- **FullSim**

- generator and geometry: e.g. RadBhaBha.Prod.mac, SuperB.Prod.mac
- workdir must be accessible and readable by apache.apache
- Production root (where output files will be written)  $\Rightarrow$  permissions

The screenshot displays the 'FULL PRODUCTION SERIES' configuration page. At the top, the 'Production Series' is set to '2010\_02\_full\_HP'. Below this, a table lists key parameters: Production Series (2010\_02\_full\_HP), Production root (/storage/gpfs\_babar6/sb/prod/), TAG (V00-01-12), ARCH (Linux26SL4\_i386\_gcc346), Release working Directory (/home/BABAR/disimone/Bruno-V00-01-12/), and LDAP uid (LTomassetti). To the right, three dropdown menus are shown for selecting the Generator (RadBhaBha singleparticle), Geometry (SuperB\_Wolf\_shielded), and Physics (QGSP\_BERT). The 'Job Details' section at the bottom includes input fields for '# of Runs' (500), '# of Events' (250), and 'Min. Delta E' (0.1), along with dropdowns for 'Generator' (RadBhaBha), 'Geometry' (SuperB\_Wolf), 'Type' (fullsim), and 'Physics' (QGSP\_BERT).

Parameter	Value
Production Series	2010_02_full_HP
Production root	/storage/gpfs_babar6/sb/prod/
TAG	V00-01-12
ARCH	Linux26SL4_i386_gcc346
Release working Directory	/home/BABAR/disimone/Bruno-V00-01-12/
LDAP uid	LTomassetti

Job Details

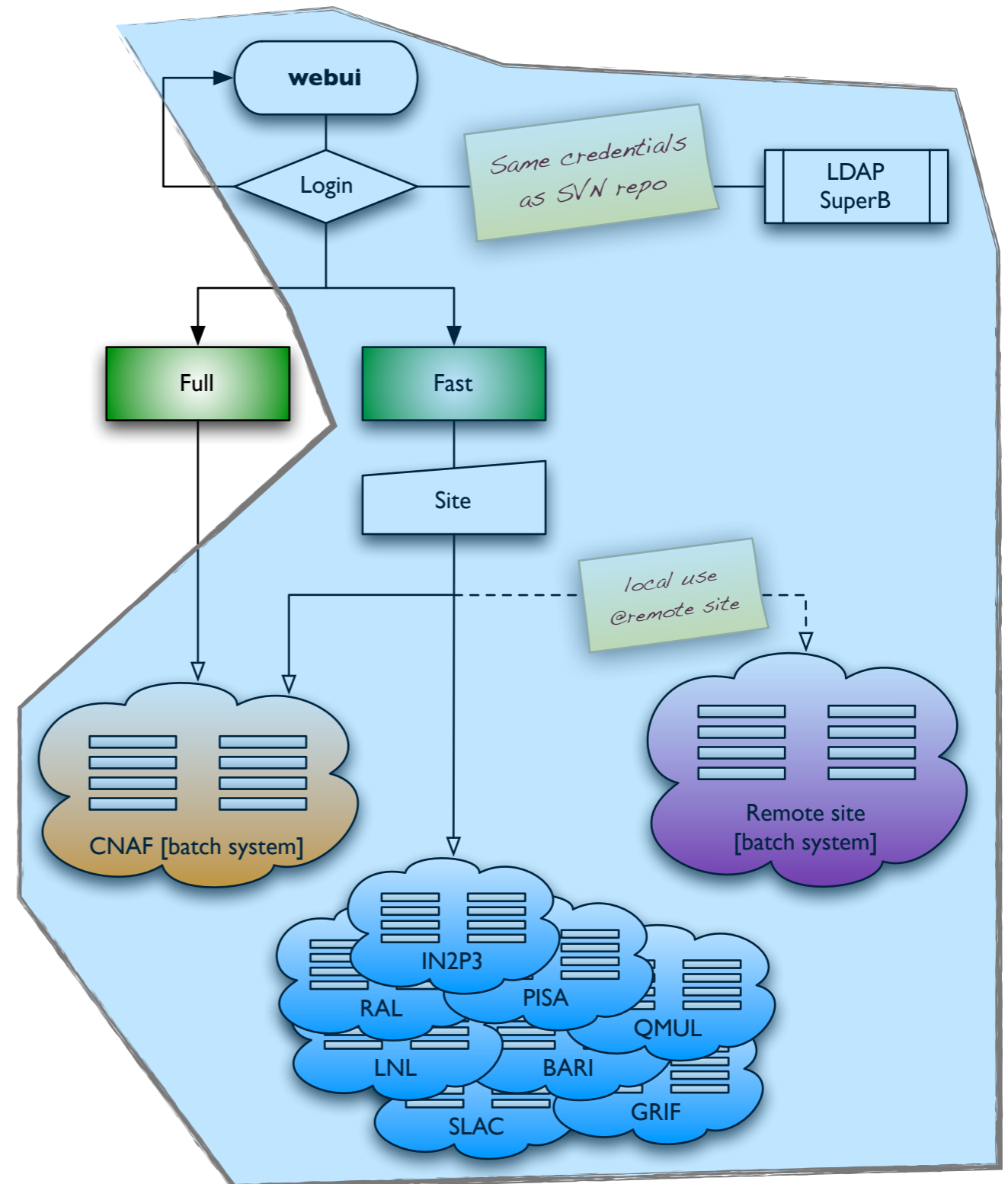
# of Runs: 500 # of Events: 250 Generator: RadBhaBha Geometry: SuperB\_Wolf Type: fullsim Physics: QGSP\_BERT

Min. Delta E: 0.1 [BRUNOBBBMINDE]

# Production tools

WebUI

- **FastSim**
- LSF Batch System submission @CNAF + Distributed submission
- User (production manager) choices:
  - Geometry
  - Generator
  - Background mixing (tcl)
  - Number of events per job
  - Number of jobs (+ batches)
  - Site



- **FastSim**

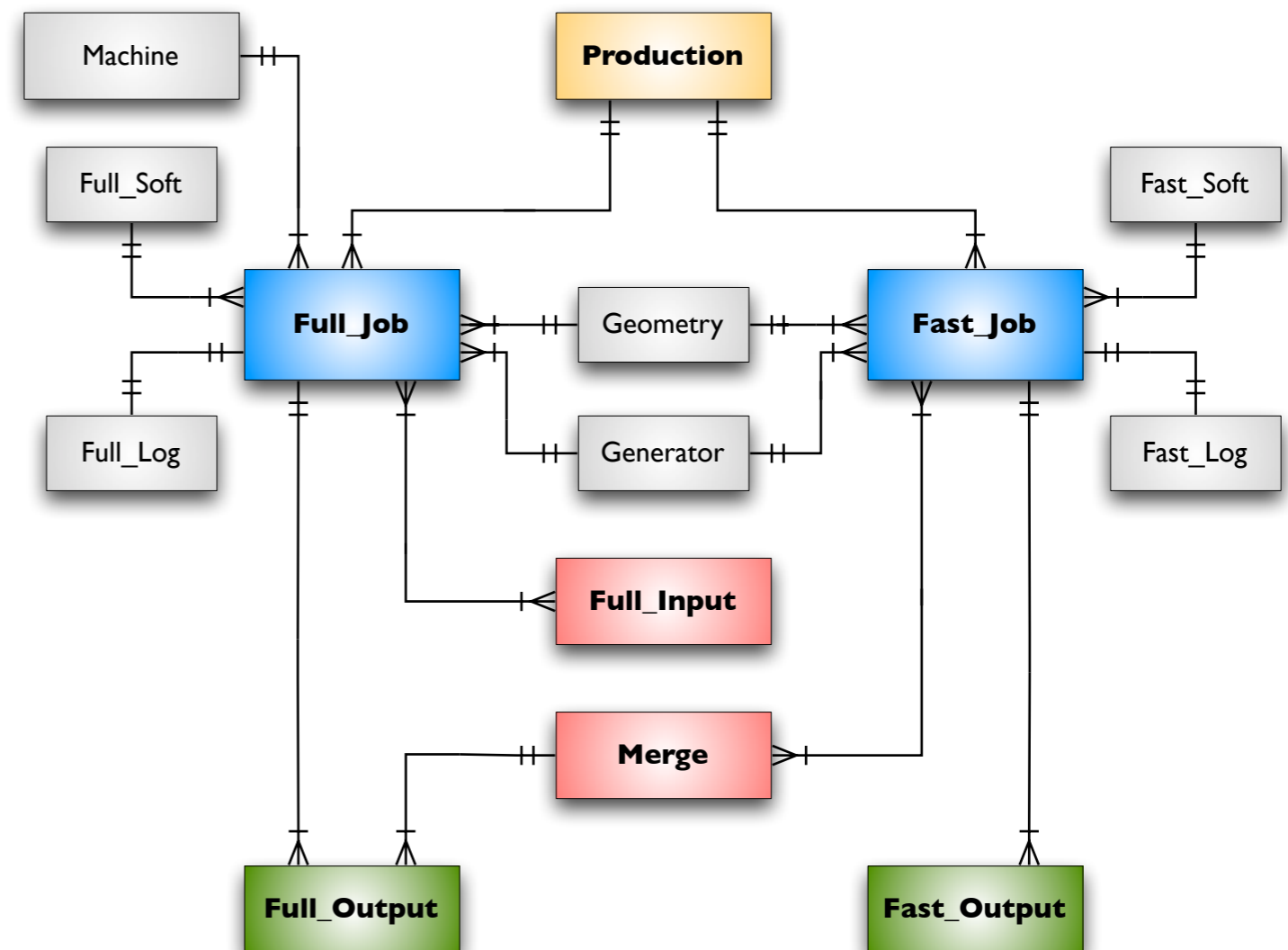
- generator, geometry, bkg mixing files (tcl, dec, ...) must be in the software
- Production root  $\Rightarrow$  permissions
  - even worse than FullSim:  
local user + 'grid' user

The screenshot displays the FastSim WebUI interface. At the top, the 'Submission mode' section shows 'Submission Type' set to 'Grid' and 'Submission Site' set to 'IN2P3-CC'. Below this, the 'Job Details' section includes input fields for '# of Runs' (200) and '# of Events (per job)' (10000). The 'Generator' is set to 'BOB0bar\_gen', 'Geometry' to 'DG\_3', and 'Bkg' to 'All Bkg, No P'. Three dropdown menus are open, showing available options: the first lists various BOB0bar generators with 'BOB0bar\_generic' selected; the second lists DG geometries with 'DG\_3' selected; the third lists background categories with 'All Bkg, No Pair' selected. A fourth dropdown menu shows submission sites with 'IN2P3-CC' selected.



- **Bookkeeping database schema**

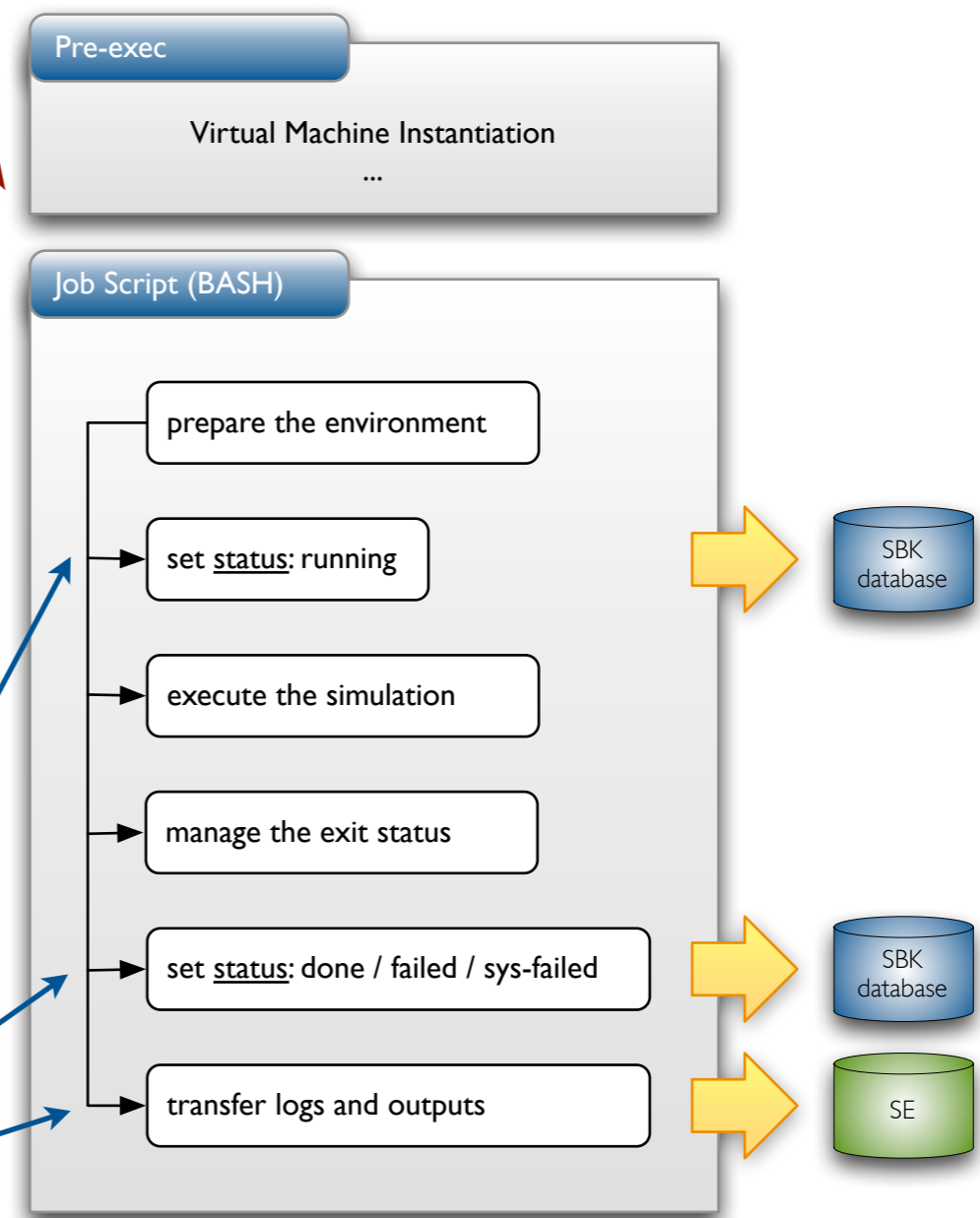
- Relevant metadata: geometry, generator, # of events, ...
- Status: done, prepared, submitted, running, failed, sys-failed
- wct, timestamp
- log reference, output reference, filesize



# Production tools

## Job scripts

- **bash script** which executes the simulation application
- prepared with proper parameters (from WebUI)
- base script (one per batch) with runnum as parameter
- environment + db interaction + exit status management + log and output transfer
- differentiation: LSF / Grid / SLAC



- DB interface:

- local jobs  
direct mysql connection
- distributed jobs  
RESTfull interface



CURL  
only update

- File transfer:

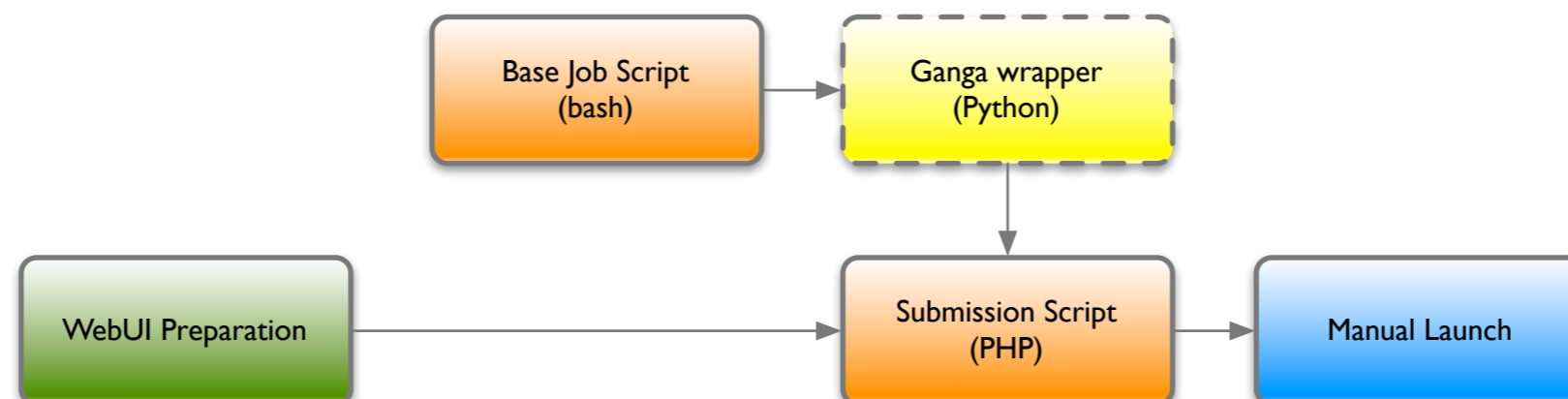
- local jobs  
cp
- distributed jobs  
lcg-cr, lcg-cp, grid-ftp



site dependent  
failover policy

Replica on  
remote site

- Wrappers / launchers around the job scripts
- PHP *or* PHP + GANGA (Python) *or* PHP + low-level Grid commands  
The one that the production user launches from CNAF (bbr-serv09)
- Allows multi-job / multi-set submissions
- Provides initialization (inserts) of the bookkeeping database
- LSF Batch system (PHP) script provides resubmit functionalities

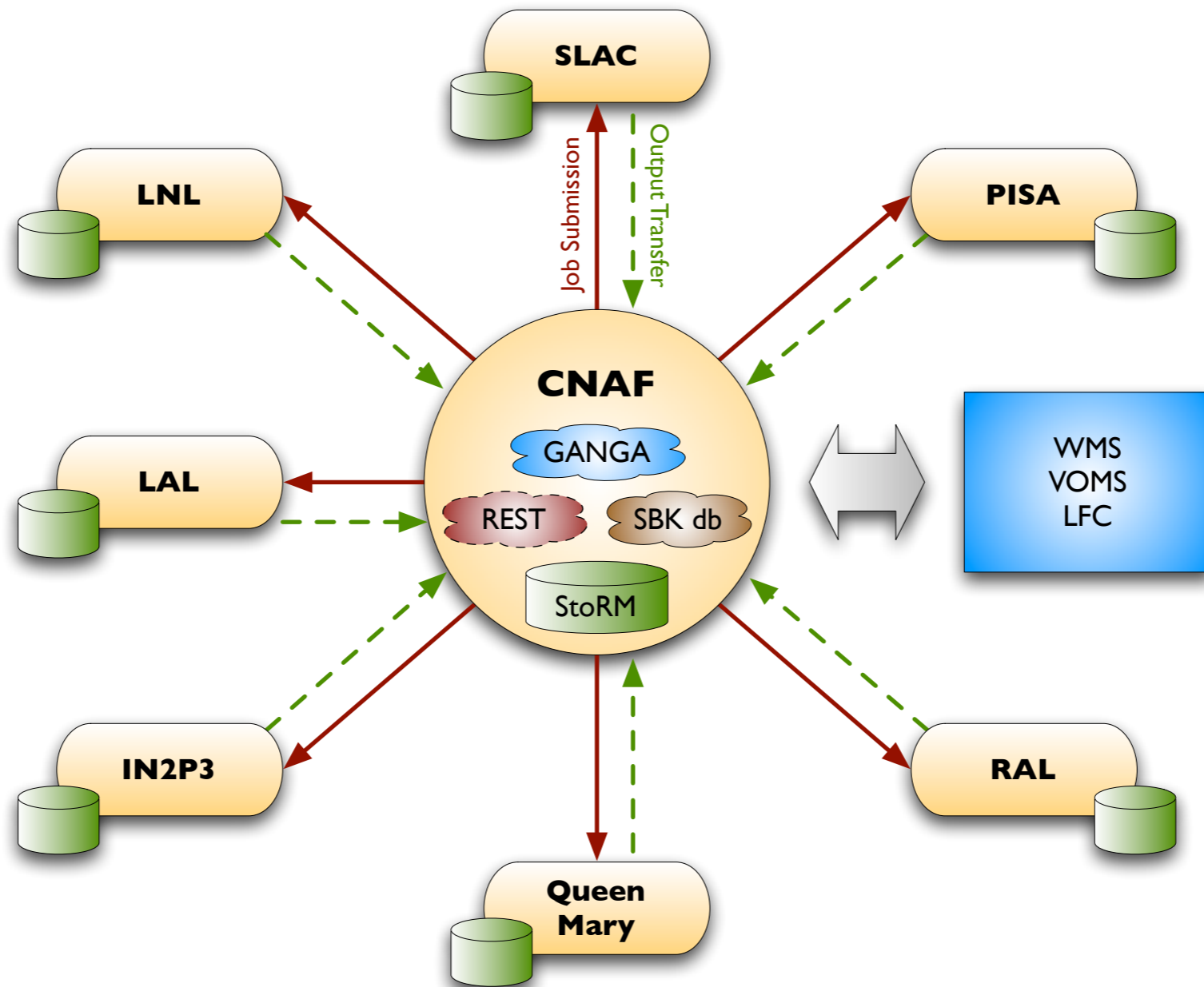


- Bulk Submission
- Site from WebUI  
⇒ partial use of WMS
- Site Requirements / CE(s)  
embedded in the scripts  
⇒ partial use of WMS
- GridJobID management
- Access to the Dashboard  
included



# Production tools

- **FastSim Distributed**



# Production tools

# Monitor&Report

- Basic Monitor based on bookkeeping database

- Parametric

- Detailed

- Log file a

- Check or

### Job's Output

channel_type	location_path	file_name	size	
DstD0ToKspipi	/storage/gpfs_superb/prod/2010_February_Generics/FastSim/DG_3/B0B0bar_generic/1003011099	DstD0ToKspipi.root	1.453 mb	✓
HadRecoil	/storage/gpfs_superb/prod/2010_February_Generics/FastSim/DG_3/B0B0bar_generic/1003011099	HadRecoil.root	4.690 mb	✓
S2b	/storage/gpfs_superb/prod/2010_February_Generics/FastSim/DG_3/B0B0bar_generic/1003011099	S2b.root	0.596 mb	✓
SemiLepKplusNuNu	/storage/gpfs_superb/prod/2010_February_Generics/FastSim/DG_3/B0B0bar_generic/1003011099	SemiLepKplusNuNu.root	0.062 mb	✓
TwoBody	/storage/gpfs_superb/prod/2010_February_Generics/FastSim/DG_3/B0B0bar_generic/1003011099	TwoBody.root	0.081 mb	✓

Runnum	# o	# e
start		ve
end		
100200		
100301		

events
1 576 440 000
35 545 000
96 265 000
1 708 250 000

number: 1002002000  
number: 1003013629

996 s (18.758 yr)

wct (s)

uperbBk

show output graph

259 jobs found matching the search criteria.

1003013599	10000	B0B0bar_generic	DG_3	V0.2.1 221	done	1072757	INFN-T1	MixSuperbBkg_NoPair.tcl	19213.24
1003013598	10000	B0B0bar_generic	DG_3	V0.2.1 221	done	1072756	INFN-T1	MixSuperbBkg_NoPair.tcl	18830.00
1003013597	10000	B0B0bar_generic	DG_3	V0.2.1 221	done	1072755	INFN-T1	MixSuperbBkg_NoPair.tcl	19634.58
1003013596	10000	B0B0bar_generic	DG_3	V0.2.1 221	done	1072754	INFN-T1	MixSuperbBkg_NoPair.tcl	23934.02
1003013595	10000	B0B0bar_generic	DG_3	V0.2.1 221	done	1072753	INFN-T1	MixSuperbBkg_NoPair.tcl	19002.69



# Production tools

# Monitor&Report

- Reports on sites, channels and last status changes (only FastSim)

GRIF			IN2P3-CC			INFN-BARI		
Status	# of jobs	events	Status	# of jobs	events	Status	# of jobs	events
done	<b>done JOBS</b>							2 850 000
failed								3 350 000
sys-failed								31 300 000
							500 000	
<b>List of the last 50 done JOBS</b>								
Timestamp	runnum	Site	DG	Generator	tcl	status		
2010-03-10 09:10:24	1003010682	UKI-LT2-QMUL	DG_4	B+B-_generic	MixSuperbBkg_NoPair.tcl	done		
2010-03-10 09:09:46	1003010609	UKI-LT2-QMUL	DG_4	B+B-_generic	MixSuperbBkg_NoPair.tcl	done		
2010-03-10 09:09:46	1003010637	UKI-LT2-QMUL	DG_4	B+B-_generic	MixSuperbBkg_NoPair.tcl	done		
2010-03-10 09:07:11	1002016717	UKI-SOUTHGRID-RALPP	DG_3	uds	PacProduction.tcl	done		
2010-03-10 09:06:13	1002014818	UKI-SOUTHGRID-RALPP	DG_4	ccbar	PacProduction.tcl	done		
2010-03-09 11:16:53	1002013408	UKI-SOUTHGRID-RALPP	DG_4	uds	PacProduction.tcl	done	660 000	
2010-03-09 11:14:47	1002012521	UKI-SOUTHGRID-RALPP	DG_4	ccbar	PacProduction.tcl	done	540 000	
<b>Total</b>							<b>247 200 000</b>	
DG_4	ccbar	MixSuperbBkg_NoPair.tcl	<b>330</b>	<b>6 600 000</b>	<b>3 976 354</b>			
DG_4	uds	MixSuperbBkg_NoPair.tcl	<b>252</b>	<b>10 080 000</b>	<b>3 679 825</b>			
DG_3	B+B-_generic	PacProduction.tcl	<b>2 092</b>	<b>104 600 000</b>	<b>79 818 019</b>			
DG_3	B0B0bar_generic	PacProduction.tcl	<b>2 060</b>	<b>103 000 000</b>	<b>79 992 260</b>			
DG_3	ccbar	PacProduction.tcl	<b>1 129</b>	<b>112 900 000</b>	<b>24 731 284</b>			
DG_3	uds	PacProduction.tcl	<b>1 556</b>	<b>311 200 000</b>	<b>51 541 216</b>			
DG_4	B+B-_generic	PacProduction.tcl	<b>2 085</b>	<b>104 250 000</b>	<b>80 612 976</b>			
DG_4	B0B0bar_generic	PacProduction.tcl	<b>2 025</b>	<b>101 250 000</b>	<b>84 969 268</b>	vents		
DG_4	ccbar	PacProduction.tcl	<b>1 101</b>	<b>110 100 000</b>	<b>27 916 869</b>	<b>94 125 000</b>		
DG_4	uds	PacProduction.tcl	<b>2 624</b>	<b>524 800 000</b>	<b>89 399 621</b>	<b>28 375 000</b>		
<b>Total</b>	<b>Total</b>		<b>19 180</b>	<b>1 576 440 000</b>	<b>591 541 996</b>	<b>122 500 000</b>		

# February Production

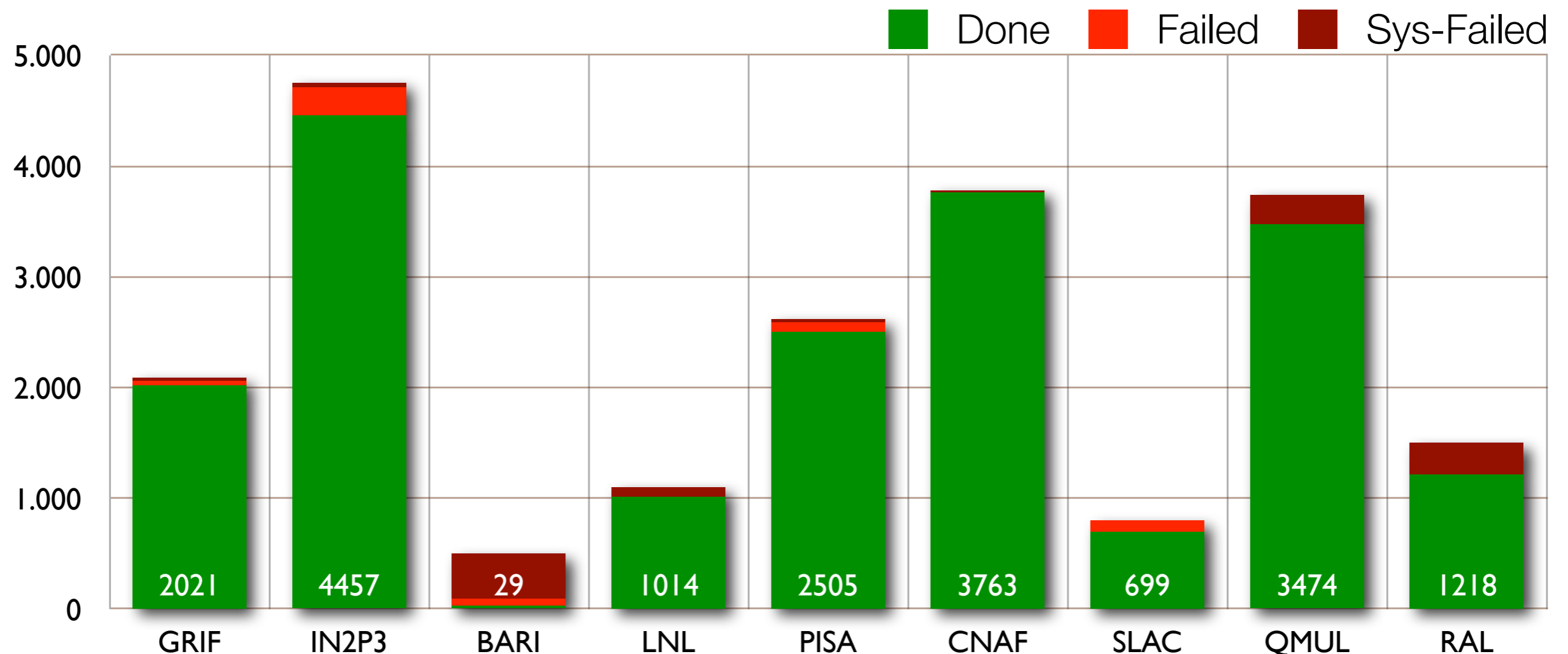
## Remote Sites

- 2010\_February\_Generics

integrated wct = 591541996s (18.8yr)

20876 submitted jobs:            19180 done            544 failed            1152 sys-failed

⇒ 92% success rate (failures are mostly due to tests, site overload, config)



# February Production

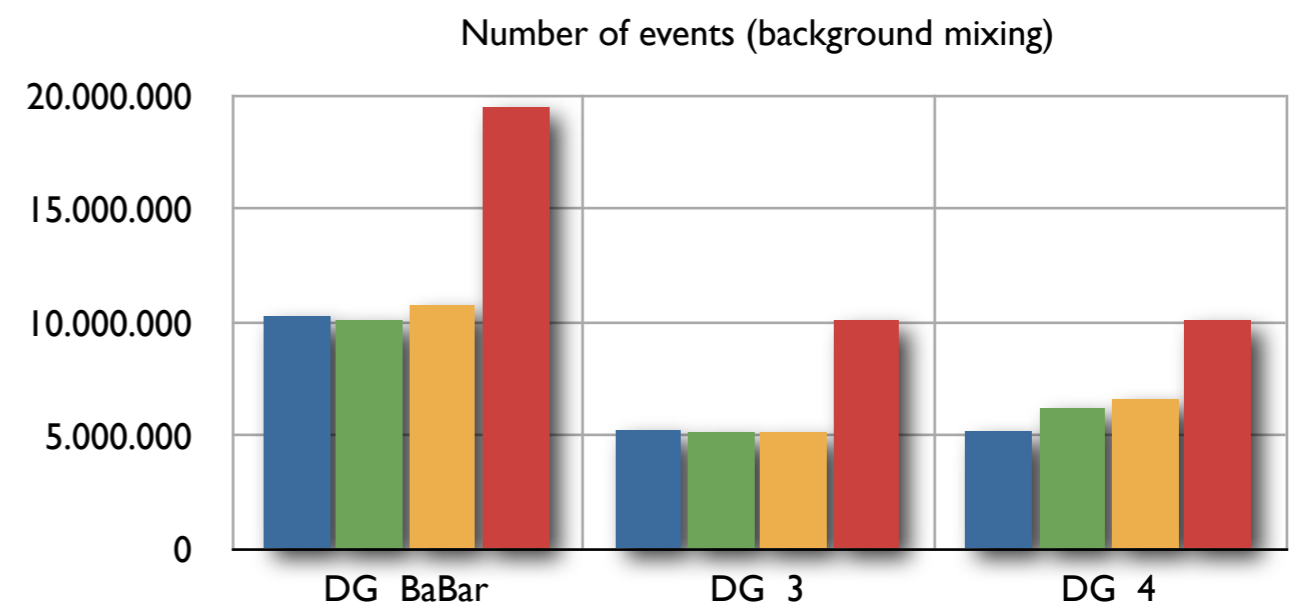
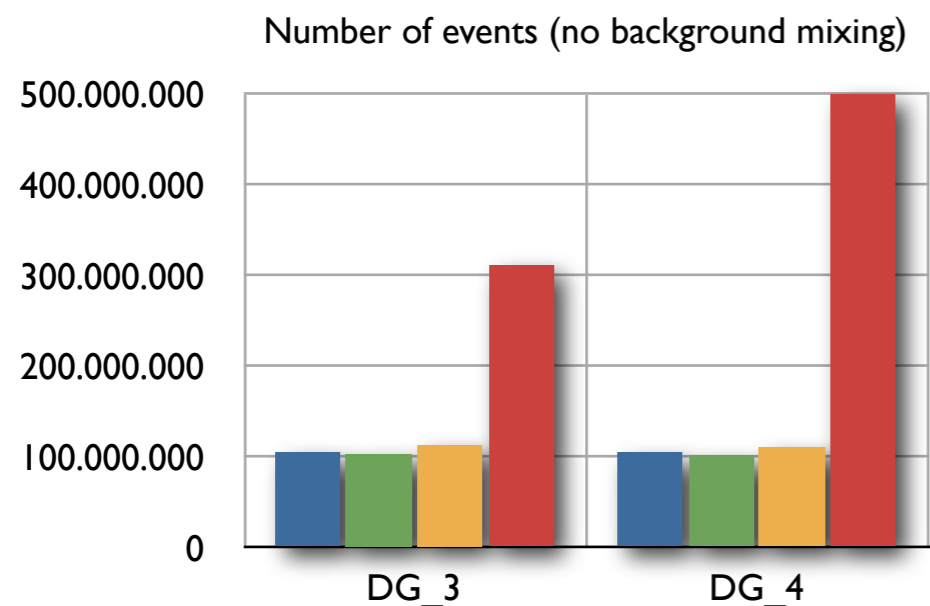
## Failures

- 
- 2010\_February\_Generics
  - Access to input files and/or software @remote sites (site overloads) 2.6% - (544)
  - Tests (@BARI, @SLAC) 2.4% - (504)
  - Proxy expiration (under investigation) 2.0% - (422)
  - Output and Log transfers 3.0% - (629)

simply  $N / \text{Total Number of submitted jobs}$   
[overlap of failure causes]

# February Production

- FullSim background frames:  $1 \cdot 10^6$  events  
background studies:  $8 \cdot 10^5$  events
- FastSim generics:  $\sim 1.5 \cdot 10^9$  events +  $10^8$  events ( $\geq 20\%$  of requests)  
signal mode:  $\sim 8 \cdot 10^7$  events (100% of requests)



■ B+B-generic ■ B0B0bar-generic ■ ccbar ■ uds

- Full Simulation (two flavours)
- Fast Simulation (two flavours)
- Not taking into account the Prod. Tools development/testing time  
3 people: **A. Fella**, M. Ronzano, **L. Tomassetti**  
~~24/24 with some kind of shifts~~  
3 exhausting weeks!
- Some issues:
  - software build and distribution
  - configuration
  - interactions with others users
  - scheduling

} see A. Fella talk

- Manual check of site load – both local (LSF) and remote  
⇒ ~~collisions with user jobs~~, collaboration with site contacts extremely helpful
- Manual adjustment of job parameters (with interaction with experts)  
⇒ running time as a function of numbers of events, geometry, generator, ...
- Manual check and management of SE (central repository) load  
⇒ distribution in time of file transfers (job duration, scheduling, ...)  
⇒ collisions with analysis and user accesses
- Write permissions on SE  
⇒ local user, storm user

- Organization of software deployment
  - Common location for simulation programs (work dir and test releases)
  - Agreement on configuration file and structure (mac, gdml, tcl, ...)⇒ independence of executables from configuration files
  - Better separation of site requirements/parameters and job script
  - Automatic generation and distribution of test releases (fastsim)

Baseline to involve more people...

# Conclusions

---

- Overview of the Production Tools
- FullSim & FastSim WebUI
- Use of the distributed infrastructure to enhance our computing power
- Very brief report on last productions
- Necessity to generalize the Tools and improve the inter-independence with software
- Involve more people



# Open Questions

---

- Data accessibility
- Analysis of production performances
- Grid submission of FullSim jobs?
- Evolution of simulation executables (new features  $\Rightarrow$  upgrade to Prod. Tools)
- Next Production: July – September?