

Aggiornamento sul sistema di gestione dei flussi computazionali del CRMA e data provenance

Seminario interno ARPA FVG

Palmanova, 10 maggio 2018
Dario B. Giaiotti

ARPA FVG – CRMA
Centro Regionale di Modellistica Ambientale
crma@arpa.fvg.it

Motivo del seminario e sommario

Perché questo seminario?

1. Ci sono stati importanti cambiamenti nei flussi di calcolo operativi del CRMA.
2. Sono maturate le condizioni per un aggiornamento dei processi in qualità ISO9001, di competenza del CRMA
3. Condivisione con i colleghi di parti del lavoro svolto negli ultimi mesi

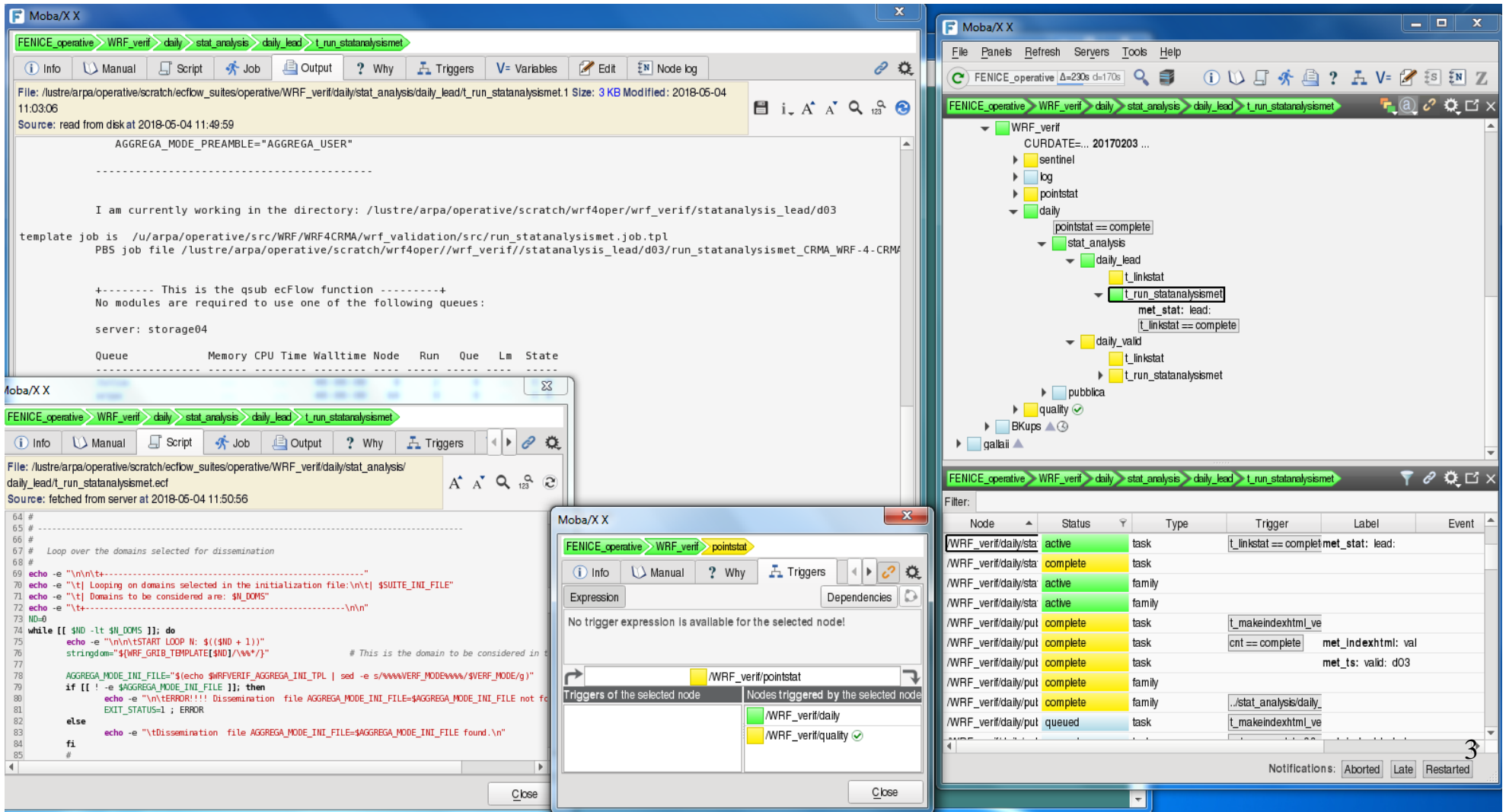
Argomenti trattati

1. Aggiornamento del software ecFlow per la gestione dei flussi operativi
2. Nuovo sistema di gestione delle code di calcolo su FENICE
3. Data provenance dei flussi operativi
4. Prospettive sulla modifica dei processi in qualità

Flussi di calcolo: nuova versione del workflow manager ecFlow

Siamo passati da ecFlow 4.0.7 (Mar 2015) a ecFlow 4.7.1 (Nov 2017)

Stiamo usando la nuova interfaccia grafica di controllo ed esecuzione dei flussi: **ecflow_ui**



The image displays the ecFlow user interface (ecflow_ui) running in a Moba/X X terminal window. It shows the workflow execution process for a WRF model verification task.

Workflow Execution Details (Left Panel):

```

File: /lustre/arpa/operative/scratch/ecflow_suites/operative/WRF_verif/daily/stat_analysis/daily_lead/t_run_statanalysimet.1 Size: 3 KB Modified: 2018-05-04 11:03:06
Source: read from disk at 2018-05-04 11:49:59

AGGREGA_MODE_PREAMBLE="AGGREGA_USER"

-----

I am currently working in the directory: /lustre/arpa/operative/scratch/wrf4oper/wrf_verif/statanalysis_lead/d03

template job is /u/arpa/operative/src/WRF/WRF4CRMA/wrf_validation/src/run_statanalysimet.job.tpl
PBS job file /lustre/arpa/operative/scratch/wrf4oper//wrf_verif//statanalysis_lead/d03/run_statanalysimet_CRMA_WRF-4-CRMA

+----- This is the qsub ecFlow function -----+
No modules are required to use one of the following queues:

server: storage04

Queue      Memory CPU Time Walltime Node  Run  Que  Lm  State
-----

```

Task Tree (Right Panel):

- WRF_verif
 - sentinel
 - log
 - pointstat
 - daily
 - pointstat == complete
 - stat_analysis
 - daily_lead
 - t_linkstat
 - t_run_statanalysimet
 - met_stat: lead:
 - t_linkstat == complete
 - daily_valid
 - t_linkstat
 - t_run_statanalysimet
 - pubblica
 - quality ✓
 - BKups
 - qalib

Task Status Table (Bottom Right Panel):

Node	Status	Type	Trigger	Label	Event
/WRF_verif/daily/sta	active	task	t_linkstat == complet	met_stat: lead:	
/WRF_verif/daily/sta	complete	task			
/WRF_verif/daily/sta	active	family			
/WRF_verif/daily/sta	active	family			
/WRF_verif/daily/put	complete	task	t_makeindexhtml_ve		
/WRF_verif/daily/put	complete	task	cnt == complete	met_indexhtml: val	
/WRF_verif/daily/put	complete	task		met_ts: valid: d03	
/WRF_verif/daily/put	complete	family			
/WRF_verif/daily/put	complete	family			
/WRF_verif/daily/put	complete	family			
/WRF_verif/daily/put	complete	family			
/WRF_verif/daily/put	queued	task			
/WRF_verif/daily/put	queued	task			

Task Configuration Dialog (Bottom Center Panel):

Expression: No trigger expression is available for the selected node!

Triggers of the selected node: /WRF_verif/pointstat

Nodes triggered by the selected node: /WRF_verif/daily, /WRF_verif/quality ✓

Come utilizzare la nuova versione di ecFlow: caricare il modulo

La nuova versione e la vecchia sono entrambe disponibili, basta caricare il modulo ambientale

```
[operative@access tmp]$ module avail
----- /u/arpa/operative/oper_mods -----
devel_ecflow/4.0.7      oper_ecflow/4.0.7      user_ecflow/4.0.7
devel_ecflow/4.7.1(default) oper_ecflow/4.7.1(default) user_ecflow/4.7.1(default)
----- /opt/Modules/versions -----
3.2.10
----- /opt/Modules/3.2.10/compilers -----
gnu/4.7.2  gnu/4.9.4  perl/4.7.2  perl/5.26.1  pgi/15.7
----- /opt/Modules/3.2.10/libraries -----
acml/5.3.1/open64          grib_api/1.26.1/pgi          netcdf/4.3.3.1/pgi
armadillo/7.800.2/gnu/4.7.2  gsl/1.11/pgi                 netcdf/4.3.3.1/pgi-enable-v2
boost/1.63.0               gsl/1.16/gnu                 netcdf/4.4.0/pgi
Eigen/3.3.3                gstreamer/1.10.4/gnu/4.7.2  netcdf/4.4.1.1/gnu
emoslib/4.4.7/gnu         hdf5/1.10.0-patch1/gnu      netcdf-cxx4/4.3.0/gnu/4.7.2
fftw/2.1.5/gnu             hdf5/1.10.0-patch1/gnu      netcdf-cxx-LEGACY/4.2/gnu/4.7.2
fftw/2.1.5/open64         hdf5/1.8.11/gnu             NumPy/1.9.2/Python/2.7.10/gnu
fftw/3.3.3/gnu            hdf5/1.8.15/pgi             odb_api/0.16.2/gnu/4.7.2
fftw/3.3.3/open64        jasper/1.701.0/pgi          openblas/0.2.18/gnu
fftw/3.3.6/gnu            lapack/3.6.0/gnu            qt/5.10.1/gnu/4.9.4
gdal/2.1.2/gnu/4.7.2      libpng/1.2.12/pgi           qt/5.1.1/gnu/4.9.4
glib/2.45.8/gnu/4.7.2    libpng/1.6.24/pgi
grib_api/1.10.4/gnu       netcdf/3.6.3/pgi
grib_api/1.14.4/gnu       netcdf/4.2.1.1/gnu
grib_api/1.14.4/pgi       netcdf/4.2.1.1/Python/2
grib_api/1.19.0/pgi/15.7  netcdf/4.2.1.1-static/g
----- /opt/Modules/3.2.10/mpi -----
openmpi/1.10.0/pgi/15.7    openmpi/1.6.4/gnu/4.4.7    openmpi/2.0.4/gnu/4.7.2
openmpi/1.10.0/pgi-pbspro/15.7 openmpi/1.6.5/gnu/4.7.2
----- /opt/Modules/3.2.10/applications -----
cdo/1.6.1/gnu              ECC/2.0.2/gnu/4.7.2        gnuplot/4.6.3              ncl/6.1.2                  octave/3.6.4               R/3.3.1
cdo/1.8.2/pgi/15.7        ECC/2.0.2/pgi/15.7        gnuplot/5.0.2              ncl/6.4.0                  proj/4.9.3/gnu/4.7.2      tau/2.22.2/gnu
cmake/3.7.1                gmt/4.5.13                 grads/2.0.2                 nco/4.3.4                  Python/2.7.10/gnu
conda/4.0.5                gmt/4.5.9                  grads/2.1.1.b0(default)    nco/4.7.4                  R/3.1.3
degrib                      gmt/5.3.1                  grid-tools                  ncvview/2.1.2              R/3.2.2
----- /opt/Modules/3.2.10/contribs -----
arpa
----- /opt/arpa/modules -----
aria_reg/1.1.0             aria_reg/1.4.0             aria_reg/1.7.0             shufem/6.1.62(default)
aria_reg/1.2.0             aria_reg/1.5.0             aria_reg/1.8.0(default)   user_ecflow/4.0.7
aria_reg/1.3.0             aria_reg/1.6.0             aria_reg/1.9.0             user_ecflow/4.7.1(default)
```

← User: operative

Tutti gli utenti

```
[operative@access tmp]$ module load arpa
module path /opt/arpa/modules added, type 'module avail'
[operative@access tmp]$ module avail
```

Come utilizzare la nuova versione di ecFlow: avviare il server

Bisogna avviare il server ecFlow

```
+-----+
|HOW TO CHECK WHETHER THE SERVER IS RUNNING AND HOW TO START THE SERVER|
+-----+
```

In order to use ecFlow, first check whether the ecflow_server is running.
Execute one of the following commands:

```
ecflow_client --stats
```

```
ecflow_client --ping --host=$ECF_HOST --port $ECF_PORT
```

```
netstat -lnptv | grep ecflow (only if server started with your user ID)
```

If the message sent to stdout shows errors in connecting to the server and there is no mention to: Status RUNNING, then you need to start the ecflow_server.
On a shared machine multiple users and ecFlow servers can coexist.

```
!!!!!! HOW TO START AND STOP THE SERVER !!!!!!
```

If the server is not running, then start an ecflow_server : by typing the following

```
ecflow-crma_start.sh -d $ECF_HOME -p $ECF_PORT
```

```
ecflow-crma_stop.sh -p $ECF_PORT
```

```
!!!!!! - - - - - !!!!!!
```

Costruire la suite, caricarla sul server e [gestirla tramite ecflow_ui]

Per costruire la suite, usare il manuale ecFlow o clonare le suite operative:

- <https://software.ecmwf.int/wiki/display/ECFLOW>
- Il remote repository di Git è: ssh://git@grid1.mercuriofvg.it/operative_workflows.git
- Il progetto di TRAC associato è: http://grid1.mercuriofvg.it/trac/operative_workflows

```
ecflow_client -load=/La_mia_suite La_mia_suite.def  
ecflow_client -begin=/La_mia_suite
```

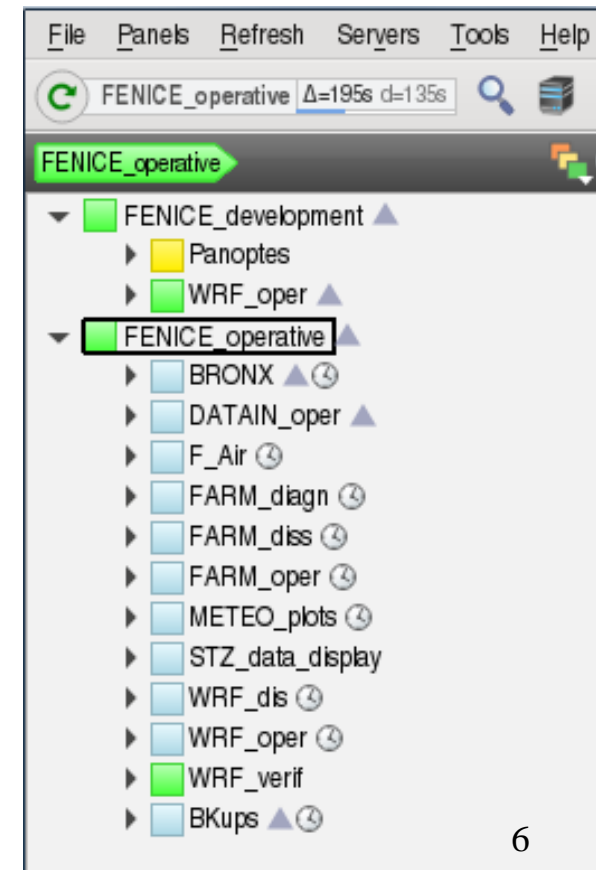
Esiste anche un documento di sintesi

disponibile sia sul web agenziale

(www.arpa.fvg.it/cms/tema/aria/utilita/Documenti_e_presentazioni/tecnico_scientifici.html)

che su WikiCRMA

(http://ms051xarpa.arpa.fvg.it/wiki/index.php/Pagina_principale/FENICE)



Sostituzione del software per la gestione delle code di calcolo

Dal giorno 24/01/2018 il software **PBSPro** è stato reso disponibile sulla FENICE

Dal giorno 15/02/2018 il software **PBSPro** ha ufficialmente sostituito il PBSTorque

Dal giorno 19/04/2018 il software **PBSPro** ha sostituito definitivamente il PBSTorque

Il nuovo software richiede una sintassi leggermente diversa rispetto al precedente, ma fornisce l'opzione **block=True**, che **trattiene l'avanzamento dell'interprete SHELL fino al termine dell'esecuzione del job** sui nodi di calcolo.

Essenziale per i flussi operativi ed evita la pratica di continue richieste sullo stato del job al gestore delle code (**problema emerso nella sua gravità nel corso del 2017**)

La documentazione su come usare PBSPro è disponibile sul WikiCRMA:

http://ms05lxarpa.arpa.fvg.it/wiki/index.php/FENICE_utilizzo_code_calcolo

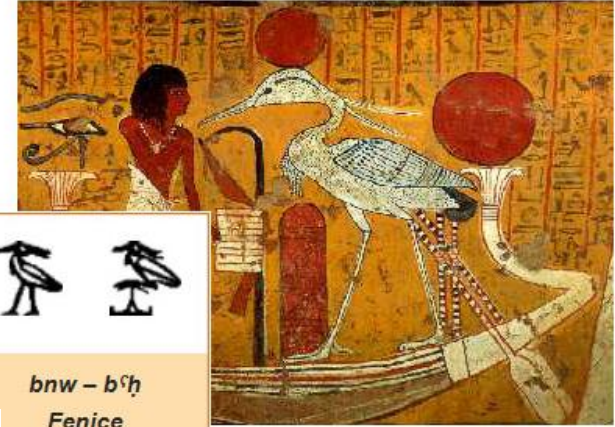
Inoltre è stato realizzato un job template che agevola la realizzazione di job da sottoporre alle code di calcolo. Lo trovate alla pagina WikiCRMA oppure al

- Il remote repository di Git è: ssh://git@grid1.mercuriofvg.it/ambiente_FENICE.git
- Il progetto di TRAC associato è: http://grid1.mercuriofvg.it/trac/ambiente_FENICE

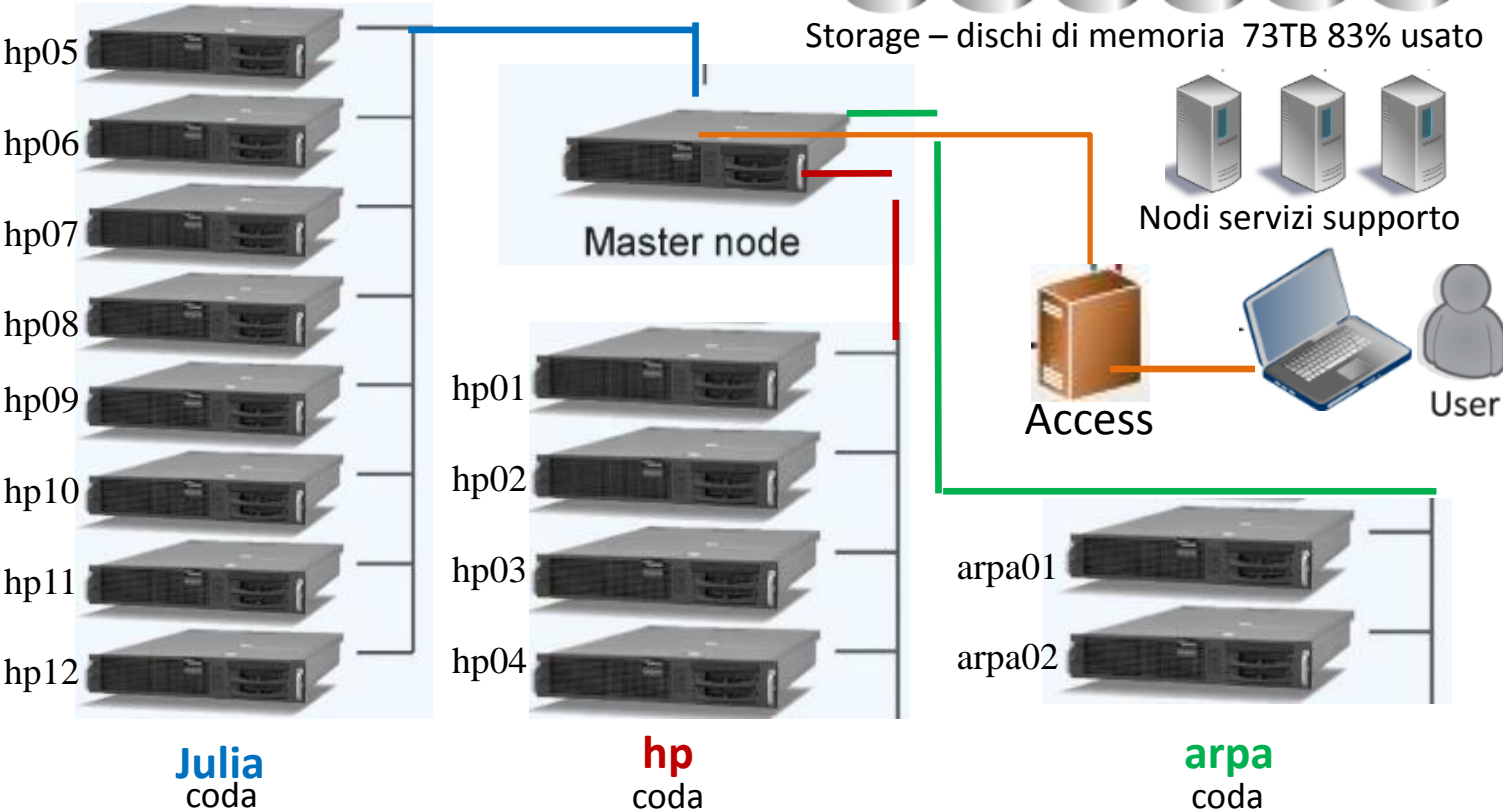
Coda di calcolo? Per chi non usa il cluster FENICE è ...

Il cluster FENICE (**F**vg **EN**hanced **I**nfrastructure and **C**omputational **EN**vironment) è composto da diversi calcolatori, ognuno dei quali svolge una classe di compiti ben definiti

Il software che gestisce le code di calcolo, impartisce gli ordini di esecuzione dei programmi



Storage – dischi di memoria 73TB 83% usato



Coda di calcolo? Per chi usa il cluster FENICE è ...

id_coda	nodi	cores_nodo	cores_tot	rete	hw_freq	hw_ram
arpa	2	64	128	Gigabit	2.10 GHz	63 GB
hp	4	32	128	infiniBand	2.30 GHz	126 GB
Julia	8	40	320	infiniBand	2.10 GHz	126 GB

Totale **14** nodi computazionali per un totale di **576** core ... ma non tutti utilizzabili contemporaneamente dallo stesso codice.

Regole comportamentali per tutti gli utenti della FENICE

Coda arpa

vanno sottoposti codici seriali e codici paralleli che non garantiscono alta scalabilità o i cui risultati non richiedono la massima velocità di esecuzione.

Coda hp

vanno sottoposti codici paralleli che hanno una grande scalabilità o che necessitano l'esecuzione nel minor tempo possibile.

Coda Julia

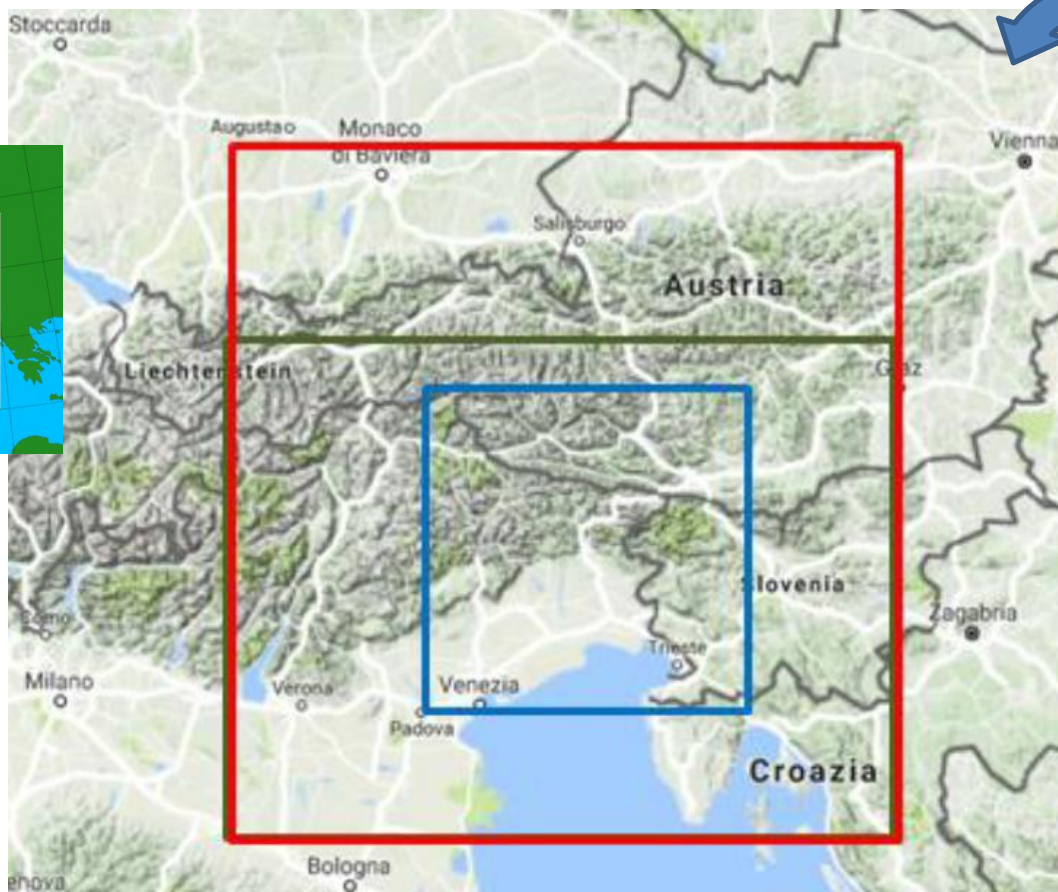
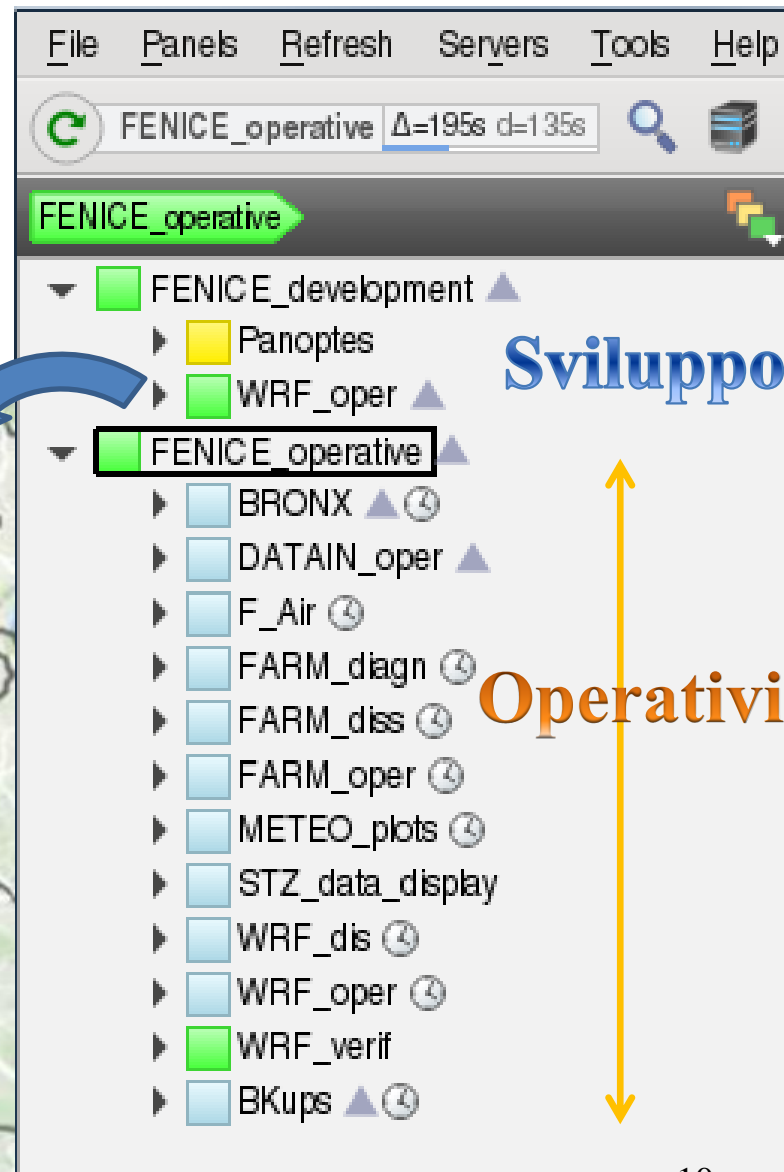
vanno sottoposti codici paralleli che hanno una grande scalabilità o che necessitano l'esecuzione nel minor tempo possibile.

Ad uso prevalente dell'utente **operative**; tutte le mattinate operative ne ha l'uso **esclusivo**

Flussi operativi consolidati e in esecuzione pre operativa al 09/05/2018

12 operativi ed in business continuity + 2 in test

Dal **20/04/2018** viene eseguita anche la previsione meteorologica tramite modello WRF sul dominio AlpeAdria (risoluzione spaziale 2 km x 2 km, campi orari, +120h) [coda Julia 4 nodi x 40 core x 2.5 ore]

File Panels Refresh Servers Tools Help

FENICE_operative Δ=195s d=135s

FENICE_operative

- ▼ FENICE_development ▲
 - ▶ Panoptes
 - ▶ WRF_oper ▲
- ▼ FENICE_operative ▲
 - ▶ BRONX ▲ (4)
 - ▶ DATAIN_oper ▲
 - ▶ F_Air (4)
 - ▶ FARM_diagn (4)
 - ▶ FARM_diss (4)
 - ▶ FARM_oper (4)
 - ▶ METEO_plots (4)
 - ▶ STZ_data_display
 - ▶ WRF_dis (4)
 - ▶ WRF_oper (4)
 - ▶ WRF_verif
 - ▶ BKups ▲ (4)

Sviluppo

Operativi

Data provenance: una necessità inderogabile

Cosa si intende per **data provenance**? (https://www.w3.org/2005/Incubator/prov/wiki/What_Is_Provenance)

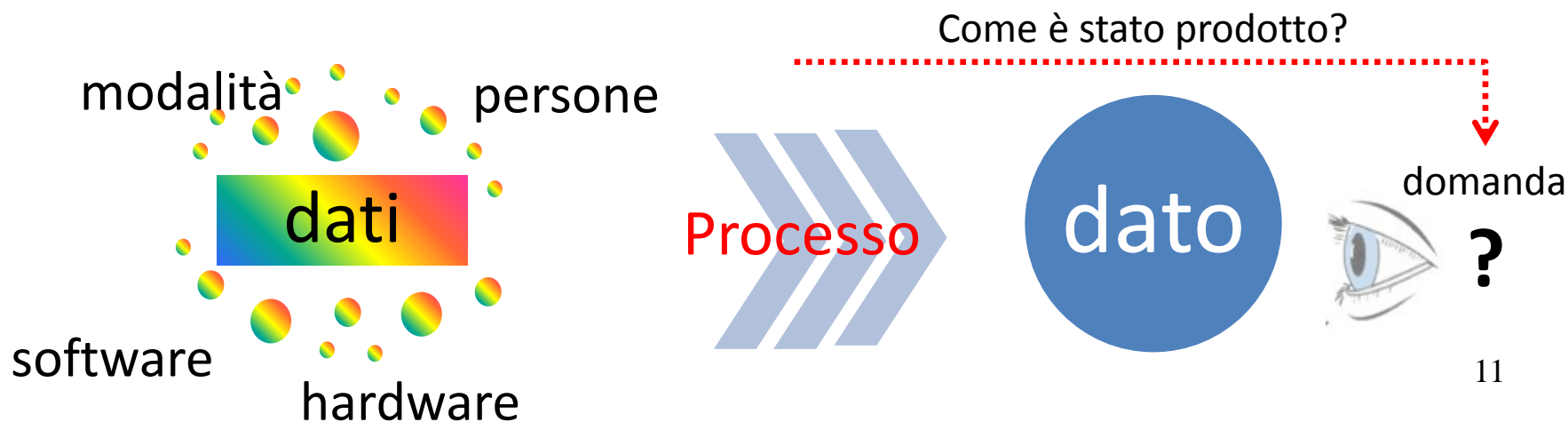
In ambito computazionale, data provenance, o solo Provenance è:

l'insieme delle informazioni necessarie e sufficienti a comprendere come è stato prodotto un dato.

Più in dettaglio:

*l'insieme delle informazioni sui **dati**, i **software**, gli **hardware**, le **persone** e le **modalità** in cui questi attori hanno prodotto una **risposta numerica ad una domanda**.*

Cosa spinge alla definizione e all'applicazione della data provenance?

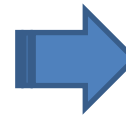


La data Provenance è una necessità dei processi complessi!

Esempio di processo **semplice** che produce il dato utile a dare una risposta

Domanda: qual è il peso medio del dott. DBG stamattina?

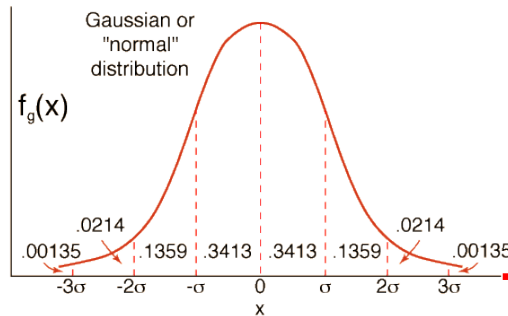
84.5, 84.3, 84,5, 84.6, 84.5, 84.2, 84.6, 84.5, 84.5, ...



$$\frac{1}{n} \sum_{i=1}^n a_i$$

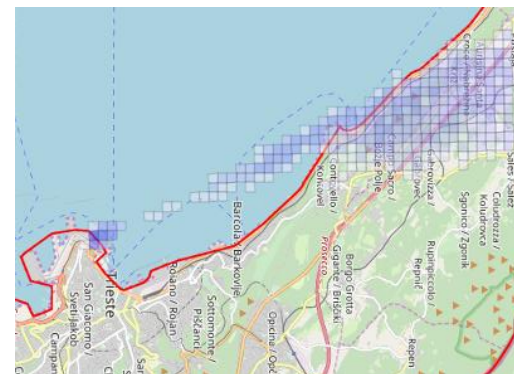
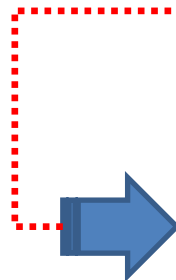
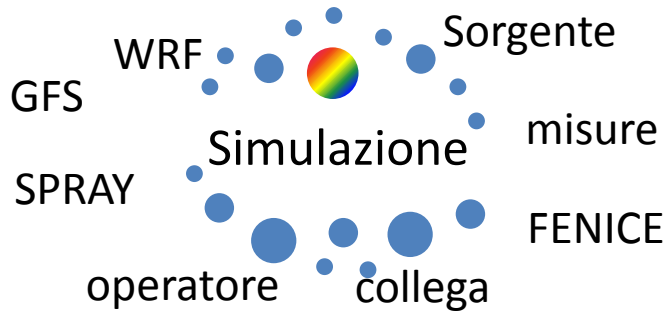


Come?



Esempio di processo **complesso** che produce il dato utile a dare una risposta

Domanda: qual è la direzione lungo la quale saranno dispersi i fumi dell'incendio scoppiato a XXXX stamattina?



Come?



Data provenance: il dato e le informazioni sul processo

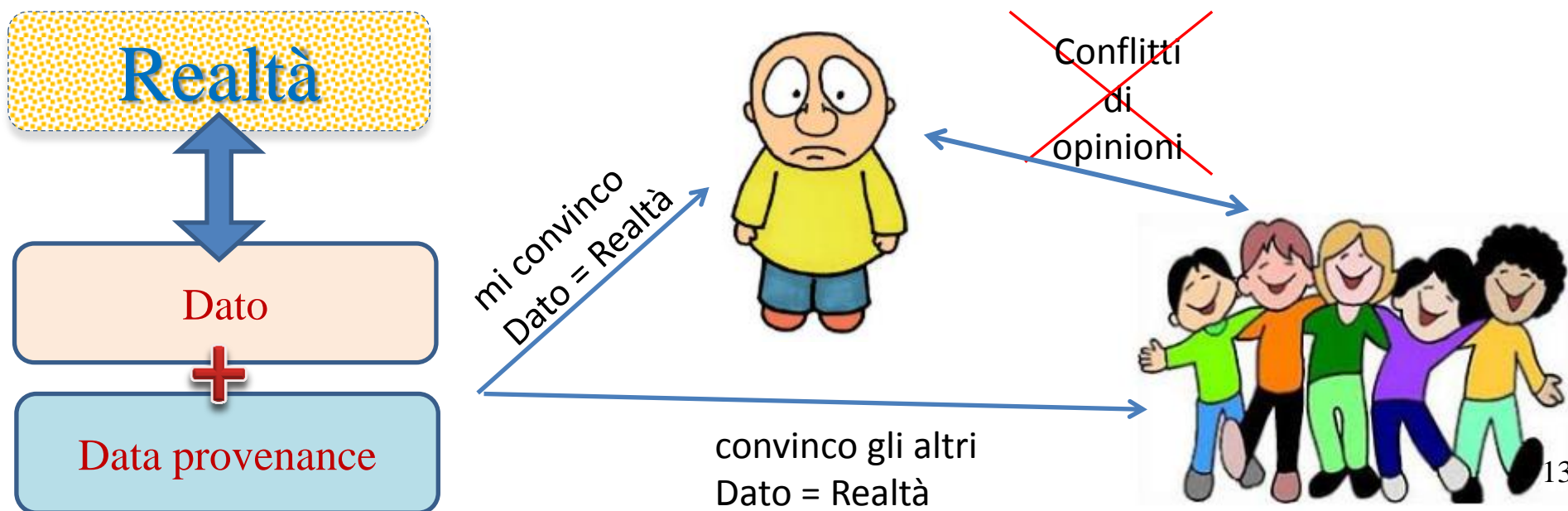
Proposizione «*Un dato è valutabile solo se conosciamo il processo che lo ha prodotto*»

Questa affermazione è valida **per tutti i dati che utilizziamo per descrivere la realtà.**

Misura in laboratorio, Osservazione, Simulazione, Sondaggio d'opinione, Costo della vita, ...

È la base del metodo scientifico, che garantisce (in principio) riproducibilità cioè obiettività, (in principio) escludendo soggettività.

Quindi la data provenance non introduce nessuna novità, ma consegue dal metodo scientifico.



La data provenance è implicitamente contemplata nella gestione della qualità

Ricordando che, per il sistema di gestione della qualità della nostra Agenzia, si definisce

processo: *insieme di attività correlate o interagenti che trasformano elementi in entrata in elementi in uscita*

La data provenance riguarda i processi che generano dati, quindi

Data provenance



Gestione della qualità

Data provenance per i flussi operativi del CRMA

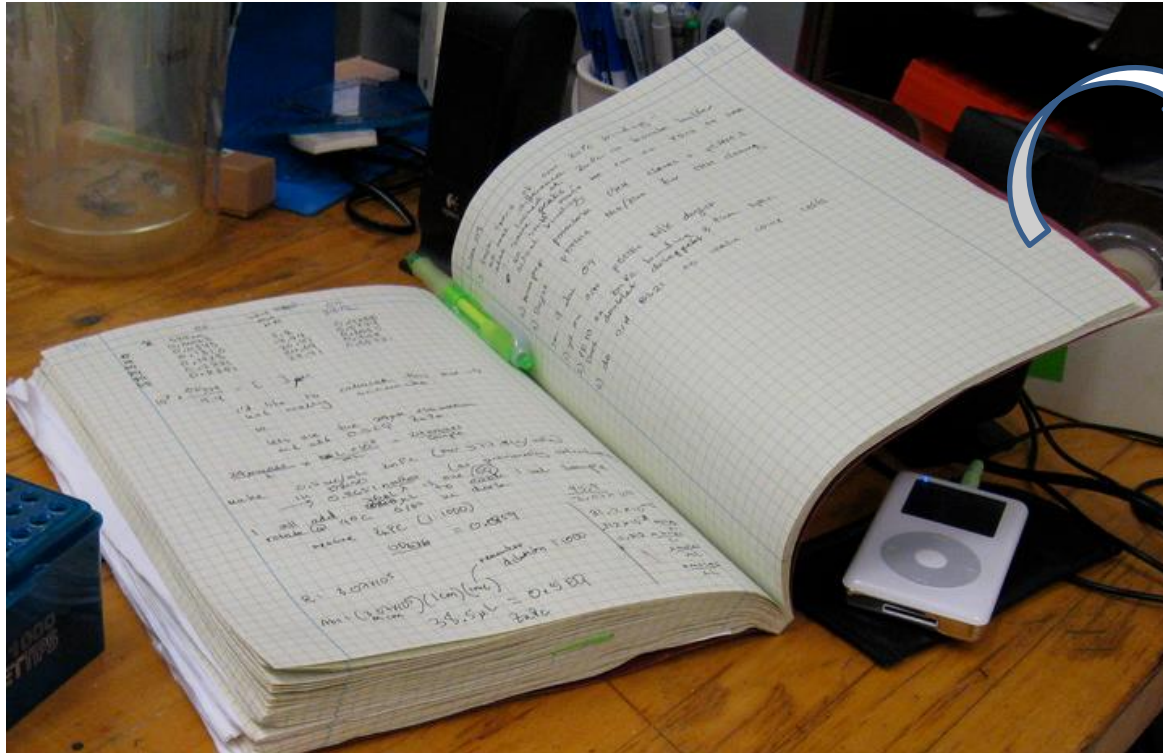
L'idea di fondo è la creazione di una raccolta di file, essenziale per ricostruire ed analizzare il processo che ha prodotto il dato, tramite un flusso di calcolo.

È un **file archivio compresso** in cui sono contenuti **tutti i file sull'esecuzione, il flusso, l'origine dei dati e la destinazione dei dati**, ma ...

NON CI SONO I DATI IN INGRESSO E NEPPURE QUELLI PRODOTTI

```

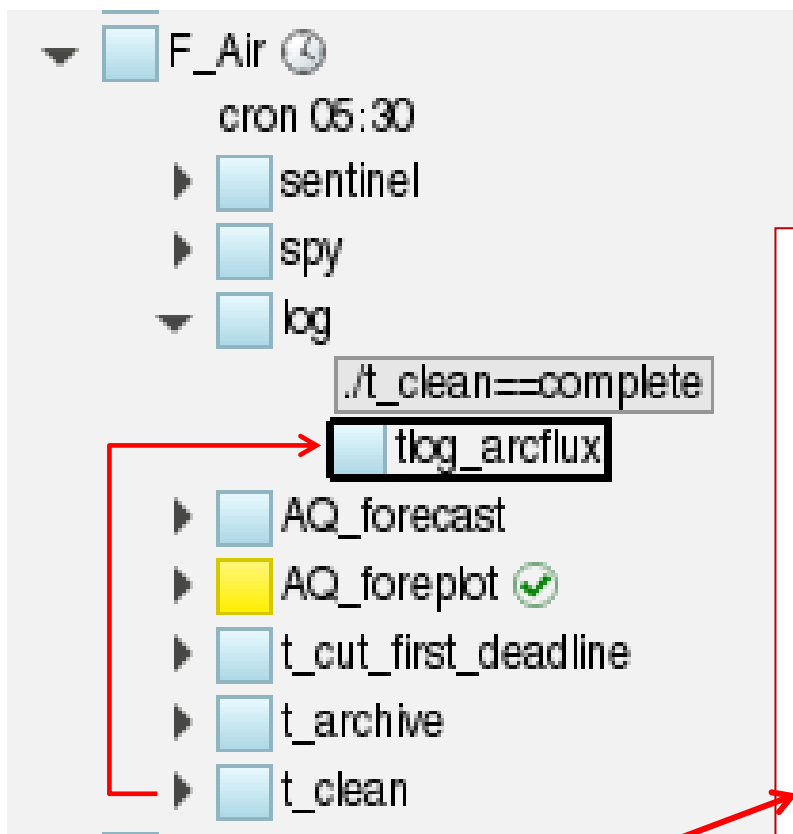
0 2017-01-19 12:59 FARM_oper/
0 2018-04-28 07:32 FARM_oper/farm_run/
0 2017-01-19 12:58 FARM_oper/sentinel/
8661 2018-05-07 07:30 FARM_oper/sentinel/tsen_run_cleaner.job1
16555 2018-05-07 07:30 FARM_oper/sentinel/tsen_run_cleaner.1
0 2017-01-19 12:58 FARM_oper/sentinel/meteo_inputs/
0 2017-01-19 12:57 FARM_oper/sentinel/meteo_inputs/step_048/
22360 2018-05-07 07:30 FARM_oper/sentinel/meteo_inputs/step_048/tsen_meteo.job1
2376 2018-05-07 07:30 FARM_oper/sentinel/meteo_inputs/step_048/tsen_meteo.1
0 2017-01-19 12:58 FARM_oper/sentinel/meteo_inputs/step_120/
22361 2018-05-07 07:30 FARM_oper/sentinel/meteo_inputs/step_120/tsen_meteo.job1
377 2018-05-07 07:30 FARM_oper/sentinel/meteo_inputs/step_120/tsen_meteo.1
0 2017-01-19 12:58 FARM_oper/sentinel/meteo_inputs/step_096/
22360 2018-05-07 07:30 FARM_oper/sentinel/meteo_inputs/step_096/tsen_meteo.job1
2376 2018-05-07 07:30 FARM_oper/sentinel/meteo_inputs/step_096/tsen_meteo.1
0 2018-04-18 10:06 FARM_oper/sentinel/meteo_inputs/step_024/
22360 2018-05-07 07:30 FARM_oper/sentinel/meteo_inputs/step_024/tsen_meteo.job1
2375 2018-05-07 07:30 FARM_oper/sentinel/meteo_inputs/step_024/tsen_meteo.1
0 2017-01-19 12:58 FARM_oper/sentinel/meteo_inputs/step_072/
22360 2018-05-07 07:30 FARM_oper/sentinel/meteo_inputs/step_072/tsen_meteo.job1
2376 2018-05-07 07:30 FARM_oper/sentinel/meteo_inputs/step_072/tsen_meteo.1
0 2018-04-28 07:32 FARM_oper/farm_run/
0 2017-01-19 12:58 FARM_oper/meteo/
0 2018-04-25 07:32 FARM_oper/meteo/step_048/
6599 2018-05-07 07:31 FARM_oper/meteo/step_048/t_surfpro4farm.1
5318 2018-05-07 07:30 FARM_oper/meteo/step_048/t_wrfdaily.1
0 2018-05-07 07:31 FARM_oper/meteo/step_048/jobstdeo/
1000 2018-05-04 07:31 FARM_oper/meteo/step_048/jobstdeo/sur-gap_daily-48.e16832
235955 2018-05-06 10:24 FARM_oper/meteo/step_048/jobstdeo/run-wrf4gap_daily-48.o17421
13039 2018-05-03 07:30 FARM_oper/meteo/step_048/jobstdeo/run-wrf_daily-48.o16430
1232 2018-05-07 07:30 FARM_oper/meteo/step_048/jobstdeo/job_summary-17906.joblog
235955 2018-04-29 07:30 FARM_oper/meteo/step_048/jobstdeo/run-wrf4gap_daily-48.o14969
99030 2018-05-04 07:31 FARM_oper/meteo/step_048/jobstdeo/run-gap_daily-48.o16826
1232 2018-05-05 07:30 FARM_oper/meteo/step_048/jobstdeo/job_summary-17209.joblog
1231 2018-05-01 07:30 FARM_oper/meteo/step_048/jobstdeo/job_summary-15645.joblog
1244 2018-05-06 10:25 FARM_oper/meteo/step_048/jobstdeo/job_summary-17422.joblog
13300 2018-05-01 07:31 FARM_oper/meteo/step_048/jobstdeo/sur-gap_daily-48.o15681
13039 2018-05-07 07:30 FARM_oper/meteo/step_048/jobstdeo/run-wrf_daily-48.o17903
13039 2018-04-29 07:30 FARM_oper/meteo/step_048/jobstdeo/run-wrf_daily-48.o14966
1244 2018-05-07 07:31 FARM_oper/meteo/step_048/jobstdeo/job_summary-17911.joblog
1211 2018-05-06 10:25 FARM_oper/meteo/step_048/jobstdeo/run-gap_daily-48.e17422
1000 2018-04-30 07:31 FARM_oper/meteo/step_048/jobstdeo/sur-gap_daily-48.e15331
99030 2018-05-02 07:31 FARM_oper/meteo/step_048/jobstdeo/run-gap_daily-48.o15928
13300 2018-05-02 07:31 FARM_oper/meteo/step_048/jobstdeo/sur-gap_daily-48.o15332
13039 2018-05-01 07:30 FARM_oper/meteo/step_048/jobstdeo/run-wrf_daily-48.o15345
1248 2018-05-04 07:31 FARM_oper/meteo/step_048/jobstdeo/job_summary-16832.joblog
1000 2018-05-05 07:31 FARM_oper/meteo/step_048/jobstdeo/sur-gap_daily-48.e17221
    
```



Chi produce la data provenance dei flussi operativi del CRMA?

La provenance è generata dal flusso stesso come ultima azione prima della conclusione *(quest'ultima azione è solo parzialmente inclusa nelle Provenance)*

Il progettista del flusso deve ricordarsi **di includere quest'ultima azione e definire alcune variabili nel file di inizializzazione della suite**. Il resto è tutto automatico

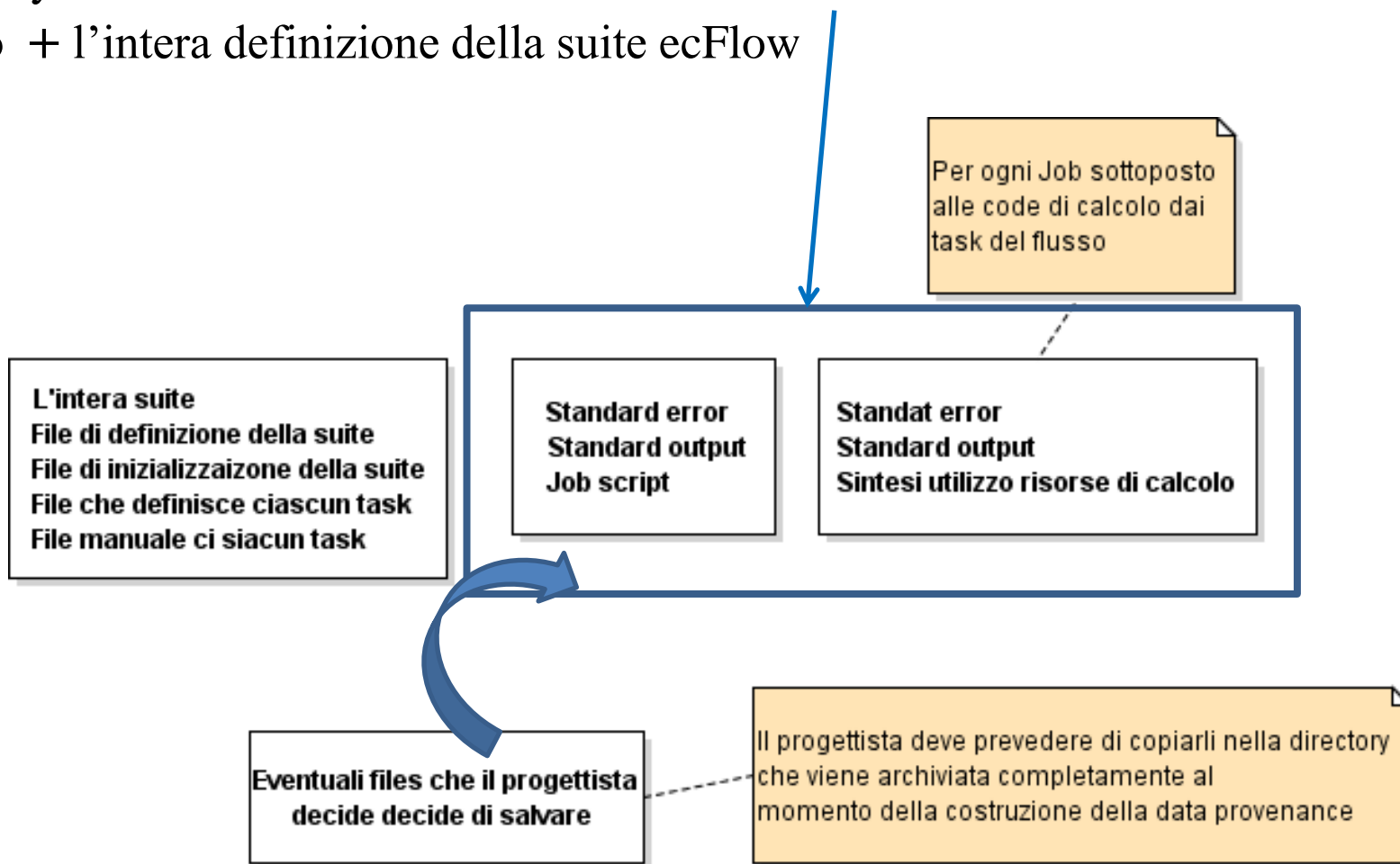


```
#
# /-----/
#
#          LOG          VARIABLES
#
#   All parameters and variables required for the Log family.
#   Preamble name is: PROV_LOG_
#
# /-----/
#
#   The log activity archives all the ecFlow nodes files generating during
#   the flow. The ecflow script is really simple and executes a tar gzipped an
#   starting from the ECF_HOME/%SUITE% directory.
#   Archive is created in a destination directory
#
#
PROV_LOG_ROOT_DIR="#{FAIR_GEN_RUN_HOME}"
PROV_LOG_TAR_DIR="/lustre/arpa/operative/scratch/backups/ecFlow/operative/"
PROV_TIME_BUFFER="10 days"

PROV_TIME_LONGTAIL="10"
```


Quali sono i file raccolti nell'archivio della data provenance?

La directory archiviata è la **%ECF_HOME%\$SUITE** che viene individuata di default dal flusso + l'intera definizione della suite ecFlow



Il file archivio ha il nome composto da **\$SUITE_flux-backup_YYYYMMDD-hhmmss.tgz**

Nome della suite

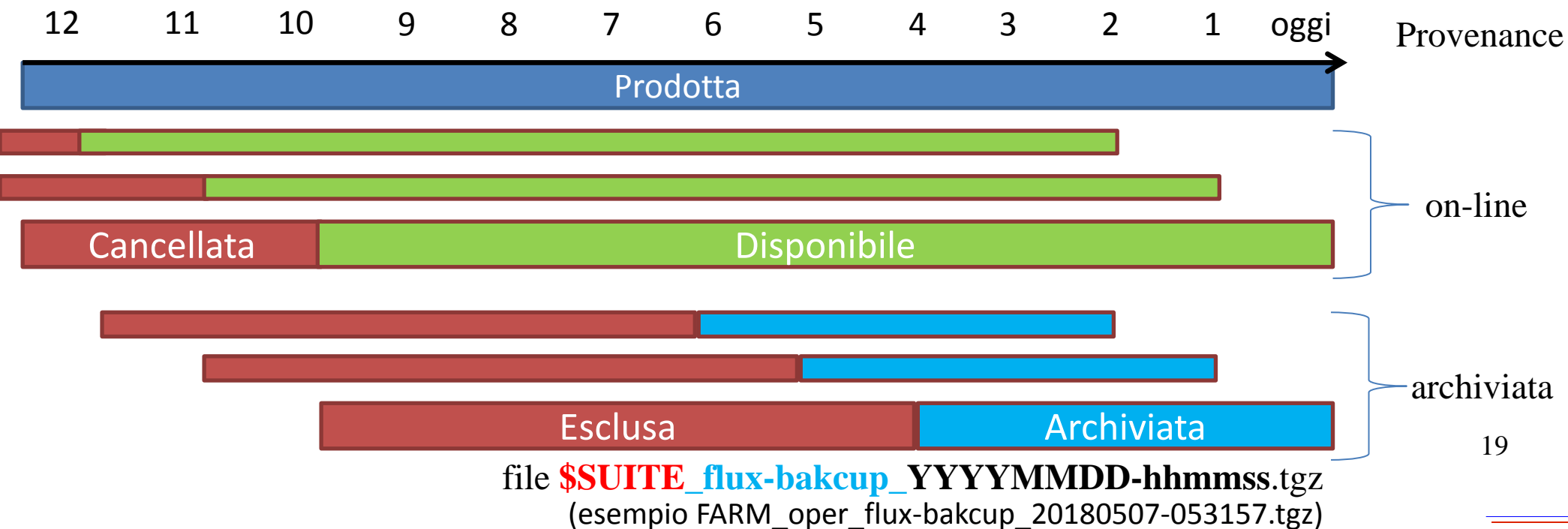
Istante in cui avviene l'archiviazione

Ridondanza delle informazioni sulla data provenance – on line

È possibile definire (nel file di inizializzazione della suite) una finestra temporale, entro la quale tutti i file sulla data provenance vengono conservati nella directory che viene sottoposta ad archivio, così come quelli che sono troppo vecchi per essere archiviati

```

PROV_TIME_BUFFER=" 4 days"
PROV_TIME_LONGTAIL="10"
# Time window for keeping the logs in the archive
# Allowed units are:
#   seconds, minutes, hours, days, months, years
# Time window for keeping the logs in the %ECF_HOME%
# Units are days and integers are mandatory.
# All kind of files older than #PROV_TIME_LONGTAIL
# are removed. If the parameter is not set, then files
# are never removed and they are left there.
  
```



Ridondanza dell'archiviazione permanente di file sulla data provenance

I file archivio **\$SUITE_flux-backup_YYYYMMDD-hhmmss.tgz** sono salvati su supporti esterni alla FENICE periodicamente. Salvataggio sia su **disco rimovibile** che su **DVD**

Il vantaggio nell'uso del DVD è che aumenta la ridondanza essendo dispositivi distinti e, a regime, sempre pieni di file.

Anno esecuzione backup	Tipo supporto utilizzato per il backup	Spazio disco disponibile	Anno completamente archiviato							
			2016	2017	2018	2019	2020	2021	2022	
Sempre	Disco esterno E	500 GB	—————							
2018	Bk_1 DVD_01	4 GB	—————							
2019	Bk_2 DVD_02	8 GB	—————							
2020	Bk_3 DVD_03	8 GB	—————							
2021	Bk_4 DVD_04	8 GB		—————						
2022	Bk_5 DVD_05	8 GB			—————					
2023	Bk_6 DVD_06	8 GB				—————				
Ridondanza			4	5	5	5	

Possibili sviluppi della Provenance a partire dallo stato attuale

Sviluppi tecnici

- Un software per l'esplorazione agevole dei file archivio e l'analisi dei contenuti

Sviluppi gestionali

- Inclusione del concetto e dell'applicazione della data provenance nelle procedure di processo.
- Sostituzione dell'archivio dei software utilizzati nei processi in qualità con il file archivio della data provenance
- Utilizzo degli archivi sulla data provenance per l'analisi di rischio dei processi in qualità