

Workflow manager ecFlow e utente operative della FENICE

Dario Giaiotti¹

Sommario

La gestione operativa dei flussi di calcolo del Centro Regionale di Modellistica Ambientale avviene per mezzo di software che automatizzano un gran numero di processi. Questo approccio garantisce l'efficienza dei processi, minimizza le probabilità di errori umani e consente la documentazione dei flussi di calcolo.

Keywords

workflow manager, cluster FENICE

¹ ARPA FVG - CRMA

*Autore di riferimento: dario.giaiotti@arpa.fvg.it

Indice

1	Introduzione	1
2	Elementi essenziali per la costruzione del modulo che genera l'ambiente per l'uso dell'ecFlow	2
3	Script di attivazione e disattivazione dei server ecFlow	2
4	Appendice - Tabelle	4

1. Introduzione

Il workflow manager utilizzato dal CRMA dell'ARPA FVG per l'esecuzione controllata di applicativi e modelli è ecFlow (<https://software.ecmwf.int/wiki/display/ECFLOW>). Questo software è stato sviluppato al European Center for Medium Range Weather Forecast (ECMWF) ed è un software libero. L'installazione di ecFlow è stata svolta seguendo le istruzioni presenti sul sito dedicato al pacchetto.

I dettagli sono riportati nel file ecFlow-installation-notes_on_FENICE.doc, oppure sulla pagina wikiCRMA (http://ms051-xarpa.arpa.fvg.it/wiki/index.php/Workflow_sw_ecFlow).

I due elementi essenziali del workflow manager sono il server ed il client; l'utente e i suoi applicativi comunicano con il server tramite il client. Il server si occupa dell'esecuzione degli applicativi dell'utente e di molte altri aspetti gestionali. La documentazione completa su come utilizzare il client ed il server si trova aggiornata sul sito web dedicato al software (<https://software.ecmwf.int/wiki/display/ECFLOW/Documentation>).

Per gli scopi operativi, di sviluppo e di ricerca, il CRMA ha definito tre server distinti che sono gestiti dall'utente operative. Ciascun server viene dedicato all'esecuzione di applicativi appartenenti ad una delle seguenti classi:

1. Flussi di programmi operativi che generano prodotti essenziali e consolidati; per esempio previsioni meteorologiche, di qualità dell'aria, elaborazioni dati e distribuzioni dati routinarie, ecc.
2. Flussi di programmi in corso di sviluppo che ambiscono ad diventare operativi, oppure flussi di applicativi eseguiti per scopi di ricerca.
3. Flussi di programmi non appartenenti alle due classi precedenti e che sono considerati per scopi personali dell'utente.

Al CRMA, la gestione dei server ecFlow è competenza dell'utente **operative** e avviene per mezzo di due strumenti:

- i moduli ambientali, che definiscono le variabili ambientali;
- gli script per l'attivazione e la disattivazione controllata dei server.

I moduli sono tre, uno per ciascun server, mentre gli script sono due. Tutti e tre i moduli permettono di accedere ai due script in quanto i moduli li mettono a disposizione nel \$PATH. Anche i moduli sono tutti accessibili all'utente operative, tra quelli del suo archivio personale (/u/arpa/operative/oper_mods), mentre solo il modulo dedicato alle esecuzioni personali di ciascun utente è disponibile a tutti gli altri utenti ed è rintracciabile tra quelli del gruppo arpa. I tre moduli sono identificati da questi nomi:

1. **oper_ecflow** per il server dedicato ai flussi di programmi operativi
2. **devel_ecflow** per il server dedicato ai flussi di programmi in corso di sviluppo o di ricerca.
3. **user_ecflow** per il server dedicato ai flussi di programmi personali di ciascun utente

2. Elementi essenziali per la costruzione del modulo che genera l'ambiente per l'uso dell'ecFlow

La corretta gestione dei server e client ecFlow avviene attraverso dei moduli ambientali e degli script che il CRMA ha sviluppato soprattutto per agevolare i compiti previsti dall'attività operativa. Le variabili d'ambiente più importanti che debbono essere definite sono la porta e l'host che verranno usati dal server e dal client.

Le principali variabili d'ambiente definite per la FENICE sono le seguenti:

- **ECF_NODE**=access.nfs
- **ECF_PORT**=NNNN
- **ECF_HOME**=\$HOME/ecflow_server

In particolare il valore attribuito alla porta è estremamente importante in quanto definisce il tipo di server e conseguentemente l'autorizzazione all'accesso. Il CRMA ha definito i seguenti server con le rispettive porte:

- la porta **2012** viene utilizzata per il server operativo, accessibile solo da utente **operative**.
- la porta **2013** viene utilizzata per il server di sviluppo, accessibile da utente **operative** e dagli **sviluppatori**.
- la porta **2014** ad uso di **tutti gli utenti** di FENICE, quindi il server non è ristretto da nessuna whitelist.

Di default, nel modulo comune a tutti gli utenti, la variabile **ECF_PORT** viene inizializzata con 2014, mentre il modulo locale dell'utente operative la variabile è inizializzata al valore 2012; infine nel modulo dedicato al server per le suite in sviluppo e ricerca, l'inizializzazione è a 2013.

I file di log ed altri ausiliari usati dal server vengono depositati di default in **\$ECF_HOME**, a meno che non si decida di cambiare il default. Da questa scelta discendono le altre variabili che sono: (cfr Tab.1 Tabella - File di log)

L'insieme di variabili ambientali che vengono definite ed esportate all'ambiente tramite modulo sono le seguenti: (cfr Tab.2 Tabella - Elenco delle variabili ambientali)

Tutte queste variabili sono state individuate eseguendo una richiesta di help all'applicativo che implementa le funzioni del server (ecflow_server-help).

Nel caso si voglia usare le api python, vanno inoltre definite le seguenti variabili d'ambiente:

- **PYTHONPATH** è la directory dove il file ecflow.so è stato installato
- **LD_LIBRARY_PATH** deve includere la directory dove libboost_python.so è stata installata.

Il file con la whitelist **\$ECF_LIST** deve essere creato secondo il formato previsto (usare il comando ecflow_client -help reloadwsfile).

I tre moduli appositamente creati sono versionati ed archiviati nel repository Git dedicato all'ambiente di calcolo della FENICE (il cui remote è ssh://git@grid1.mercuriofvg.it/ambiente_FENICE.git) e sono accessibili nella sezione di moduli locali dell'utente operative (oper_mods), mentre il solo modulo per tutti gli utenti è accessibile nella sezione dei moduli condivisi tra gli utenti arpa (contribs/arpa).

3. Script di attivazione e disattivazione dei server ecFlow

Gli script di attivazione e disattivazione del server ecFlow, ecflow_start.sh e ecflow_stop.sh, vengono distribuiti con il codice sorgente ma sono adattati per l'utilizzo nell'ambiente di calcolo del centro europeo (ECMWF). Tenuto conto che, per gli scopi operativi del CRMA si rende necessario scrivere degli script che permettano il riavvio o spegnimento automatico dei server ecFlow in caso di problemi della piattaforma di calcolo, si è deciso di costruire due script BASH di supporto all'utente operative che eseguano l'avvio e lo stop controllato dei server ecFlow in funzione delle variabili ambientali che sono definite nel modulo associato al tipo di server che si intende considerare. Tali script, che sono depositati assieme agli originali nella cartella bin dove è installato ecFlow, sono stati chiamati:

- **ecflow-crma_start.sh** che attiva il server se non già attivo
- **ecflow-crma_stop.sh** che spegne il server se attivo

I due script sono versionati ed archiviati nel repository Git dedicato all'ambiente di calcolo della FENICE (il cui remote è ssh://git@grid1.mercuriofvg.it/ambiente_FENICE.git)



Copyright © ARPA FVG, 2015

This work is released under the terms of the license

Creative Commons Attribution/NonCommercial/ShareAlike.

Information on how to request permission may be found at:

[ARPA FVG-Aria-Elaborati tecnico-scientifici](http://cmsarpa.regione.fvg.it/cms/tema/aria/utilita/Documenti_e_presentazioni/tecnico_scientifici.html)

[\(http://cmsarpa.regione.fvg.it/cms/tema/aria/utilita/Documenti_e_presentazioni/tecnico_scientifici.html\)](http://cmsarpa.regione.fvg.it/cms/tema/aria/utilita/Documenti_e_presentazioni/tecnico_scientifici.html)



[ARPA FVG-Aria-Elaborati tecnico-scientifici](http://cmsarpa.regione.fvg.it/cms/tema/aria/utilita/Documenti_e_presentazioni/tecnico_scientifici.html)

[\(http://cmsarpa.regione.fvg.it/cms/tema/aria/utilita/Documenti_e_presentazioni/tecnico_scientifici.html\)](http://cmsarpa.regione.fvg.it/cms/tema/aria/utilita/Documenti_e_presentazioni/tecnico_scientifici.html)

4. Appendice - Tabelle

Tabella - File di log

Nome file	Utilizzo
ECF_LOG=\$ECF_HOME/\$ECF_NODE.\$ECF_PORT.ecf.log	File di log del server
ECF_CHECK=\$ECF_HOME/\$ECF_NODE.\$ECF_PORT.ecf.check	File dei check
ECF_CHECKOLD=\$ECF_HOME/\$ECF_NODE.\$ECF_PORT.ecf.check.b	File di backup e checkpoint
ECF_LIST=\$ECF_HOME/\$ECF_NODE.\$ECF_PORT.ecf.list	File con elenco utenti abilitati

Tabella 1. Tabella file di log

Tabella - Elenco delle variabili

Variabile	Scopo
ECF_NODE	indica il nodo su cui viene installato il server
ECF_PORT	indica la porta sul nodo su cui viene installato il server
ECF_HOME	indica la directory dove di prassi sono depositati e cercati i file utilizzati dal server
ECF_LOG	indica il nome del log file delle operazioni eseguite dal server
ECF_CHECK	indica il nome del checkpoint file del server
ECF_CHECKOLD	indica il nome del file di backup del checkpoint
ECF_CHECKINTERVAL	indica l'intervallo di tempo ogni quanto eseguire il checkpoint, espresso in secondi
ECF_LISTS	indica il nome del file che specifica quali utenti possono accedere e come al server
ECF_TASK_THRESHOLD	indica il tempo massimo entro il quale deve generare il job, espresso in millisecondi

Tabella 2. Tabella Elenco delle variabili