Centro di Calcolo IASF-Bologna

A. De Rosa M. Malaspina F. Schiavone A.Tacchini
G. Taffoni E. Torresi

Rapporto interno IASF Bo n. 646 2014
Versione 1.0 - novembre 2013

Obiettivi del centro di calcolo

Offrire gli strumenti necessari all'attivita' scientifica e tecnologica dell'istituto, come:

- Posta elettronica
- Sistemi di stampa
- Sistemi di calcolo
- Siti web
- Storage dei dati
- Supporto agli utenti nell'utilizzo delle risorse disponibili

Model philosophy

Per ogni sistema implementato, o da implementare, dopo un confronto con lo stato attuale dell'arte sia nel mondo della ricerca che della produzione, si sono cercati e si cercheranno come requisiti essenziali:

- L'affidabilita'
- I bassi costi d'esercizio
- La scalabilita'
- La semplicita' d'intervento nel caso di malfunzionamenti

Layout funzionale del centro di calcolo

Cluster di calcolo

Task:

- Analisi dati
- Storage
- Sviluppo software

Cluster di macchine virtuali

Task:

- Implementazione dei servizi, come la posta elettronica, ftp, servizio di stampa di rete (cups), server di licenze, dns
- Macchine virtuali relative a progetti e/o task specifici, in gestione al centro di calcolo

Rete d'Istituto

- Rete pubblica 192.167.166.0/24
- Rete privata statica 192.168.166.0/24
- Rete privata servizio 192.168.64.0/24
- Rete privata dhcp 192.168.176.0/24

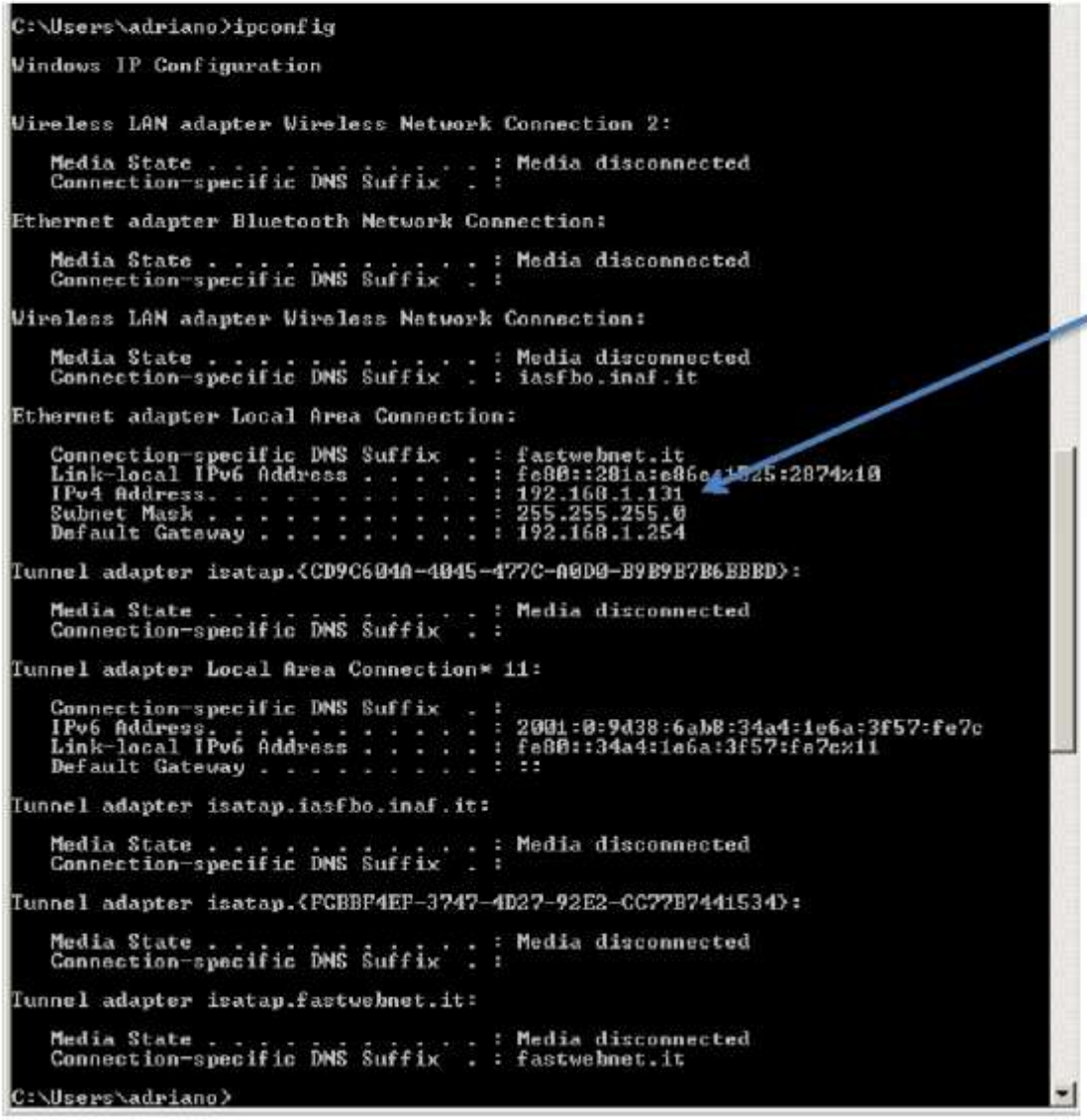
Gli IP di rete pubblica sono in parte su computer non piu' utilizzati, quindi da recuperare.

Mantenere un IP su di un pc non utilizzato per esser sicuri di averlo disponibile un domani, non e' un gesto corretto nei confronti degli altri utenti.

Rete d'Istituto

- Per conoscere il proprio indirizzo di rete e' possibile o navigare fra l pannelli di controllo del proprio sistema operativo, oppure in modo piu' semplice da una shell (interprete comandi).
- In windows per aprire la shell occorre dal menu' start selezionare il prompt dei comandi. Il comando da utilizzare e' "ipconfig /all"
- In linux e macos la shell di solito e' chiamata terminale, ed il comando e' "ifconfig -a"

Rete d'Istituto, il proprio IP - windows



Rete d'Istituto, il proprio IP - Linux

```
[derosa@login01]->ifconfig -a
em1       Link encap:Ethernet  HWaddr 60:EB:69:81:F6:97
          inet addr:192.167.166.66  Bcast:192.167.166.255  Mask:255.255.255.0
          inet6 addr: fe80::62eb:69ff:fe81:f697/64 Scope:Link
          UP BROADCAST RUNNING MULTICAST  MTU:1500  Metric:1
          RX packets:177960260 errors:0 dropped:0 overruns:0 frame:0
          TX packets:54261621 errors:0 dropped:0 overruns:0 carrier:0
          collisions:0 txqueuelen:1000
          RX bytes:21432893289 (19.9 GiB)  TX bytes:32487906078 (30.2 GiB)

em2       Link encap:Ethernet  HWaddr 60:EB:69:81:F6:98
          inet addr:192.168.210.111  Bcast:192.168.210.255  Mask:255.255.255.0
          inet6 addr: fe80::62eb:69ff:fe81:f698/64 Scope:Link
          UP BROADCAST RUNNING MULTICAST  MTU:1500  Metric:1
          RX packets:1430112290 errors:0 dropped:0 overruns:0 frame:0
          TX packets:548900848 errors:0 dropped:0 overruns:0 carrier:0
          collisions:0 txqueuelen:1000
          RX bytes:1734136768458 (1.5 TiB)  TX bytes:103661904841 (96.5 GiB)

Ifconfig uses the ioctl access method to get the full address information, which l
Because Infiniband address has 20 bytes, only the first 8 bytes are displayed corr
Ifconfig is obsolete! For replacement check ip.
ib0       Link encap:InfiniBand  HWaddr 80:00:00:03:FE:80:00:00:00:00:00:00:00:00:
          inet addr:192.168.7.111  Bcast:192.168.7.255  Mask:255.255.255.0
          inet6 addr: fe80::211:7500:70:7488/64 Scope:Link
          UP BROADCAST RUNNING MULTICAST  MTU:2044  Metric:1
          RX packets:3692264 errors:0 dropped:0 overruns:0 frame:0
          TX packets:2910650 errors:0 dropped:5 overruns:0 carrier:0
          collisions:0 txqueuelen:256
          RX bytes:4334320489 (4.0 GiB)  TX bytes:553306747 (527.6 MiB)

lo        Link encap:Local Loopback
          inet addr:127.0.0.1  Mask:255.0.0.0
          inet6 addr: ::1/128 Scope:Host
          UP LOOPBACK RUNNING  MTU:16436  Metric:1
          RX packets:64515708 errors:0 dropped:0 overruns:0 frame:0
          TX packets:64515708 errors:0 dropped:0 overruns:0 carrier:0
          collisions:0 txqueuelen:0
          RX bytes:38023302341 (35.4 GiB)  TX bytes:38023302341 (35.4 GiB)
```

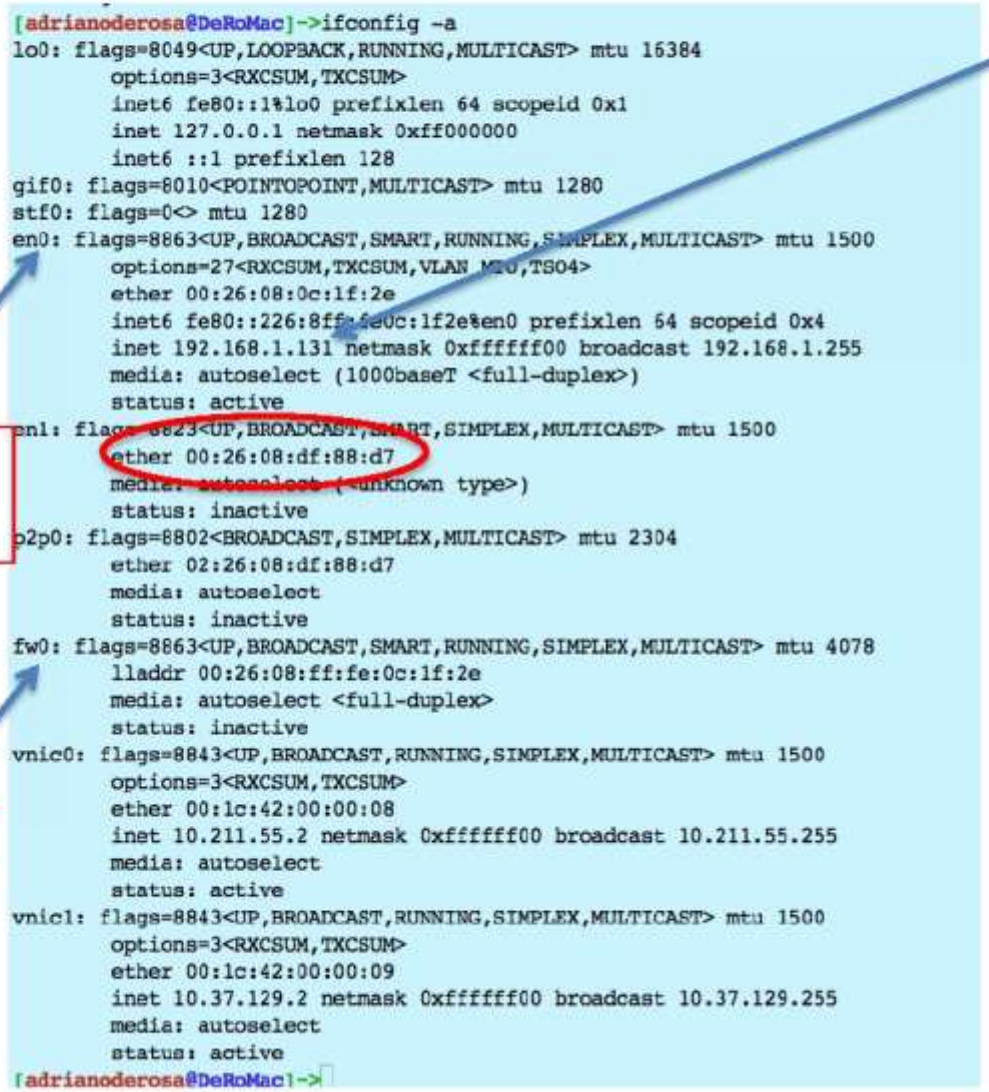
Nome interfaccia

Indirizzo IPv4

Mac address

Generalmente le interfacce in linux possono chiamarsi anche eth0, eth1..

Rete d'Istituto, il proprio IP - MacOS



In MacOS le interfacce di rete ethernet hanno nome en0 en1 (kernel BSD - analogo in AIX).
Nei MacBook en0 e' l'interfaccia con il cavo, en1 e' la wireless.
Il mac-address richiesto a volte per potersi connettere alle reti WiFi e' quello cerchiato

Rete d'Istituto, il proprio IP

Essere in grado di individuare il proprio IP e' importante per I seguenti motivi:

- aiuta in una prima analisi dei problemi di connessione di rete; se ad esempio l'ip non dovesse avere i primi tre numeri uguali a quelli delle reti definite per l'istituto, potete dedurre che: l'ip vi viene fornito dinamicamente, quindi i primi tre numeri dovrebbero essere 192.168.176.xxx, il problema potrebbe essere o nel vostro pc (eseguite un riavvio del sistema) o se persiste nella rete dell'istituto.
- Permette agli utenti di effettuare gli accessi remoti agli altri pc, o meglio a quei pc che appartengono alle reti statiche (192.167.166.0/24 e 192.168.166.0/24)

Cluster di Calcolo: architettura generale

- 4 nodi di login:
  1. bitonno.iasfbo.inaf.it
  2. tonno.iasfbo.inaf.it
  3. login01.iasfbo.inaf.it
  4. login02.iasfbo.inaf.it
  5. login03.iasfbo.inaf.it (in preparazione.. Analogo a tonno)
- 4 nodi di calcolo con 64 core ciascuno (gestiti da uno scheduler "IBM loadleveler")
- 4 nodi di storage

Cluster di Calcolo: nodi di login

I nodi di login si differenziano fra di loro per utilizzo, in modo da poter distribuire il carico di lavoro:

- **bitonno**: lettura mail, desktop remoto, accesso ai filesystem
- **tonno**: lettura mail, desktop remoto, accesso ai filesystem, compilazione, sottomissione job sui nodi di calcolo
- **login01 - login 02**: sessioni grafiche (X su ssh), accesso ai filesystem, compilazione, sottomissione job sui nodi di calcolo

Cluster di Calcolo: nodi di login

Caratteristiche tecniche:

1. tonno - bitonno: 4 core intel xeon, 16GB ram
2. login01 - login02 : 8 core amd, 32GB ram

Sistemi operativi linux (RedHat like):

1. bitonno - CentOS 5.5
2. tonno, login01, login02 - CentOS 6.4

Cluster di Calcolo: nodi di login

Modalita' di accesso:

1. SSH (senza X forwarding)

- ssh username@tonno.iasfbo.inaf.it
- ssh username@bitonno.iasfbo.inaf.it
- ssh username@login01.iasfbo.inaf.it
- ssh username@login02.iasfbo.inaf.it

Cluster di Calcolo: nodi di login

Modalita' di accesso:

2. SSH (con X forwarding)

- ssh -X username@tonno.iasfbo.inaf.it
- ssh -X username@bitonno.iasfbo.inaf.it
- ssh -X username@login01.iasfbo.inaf.it
- ssh -X username@login02.iasfbo.inaf.it

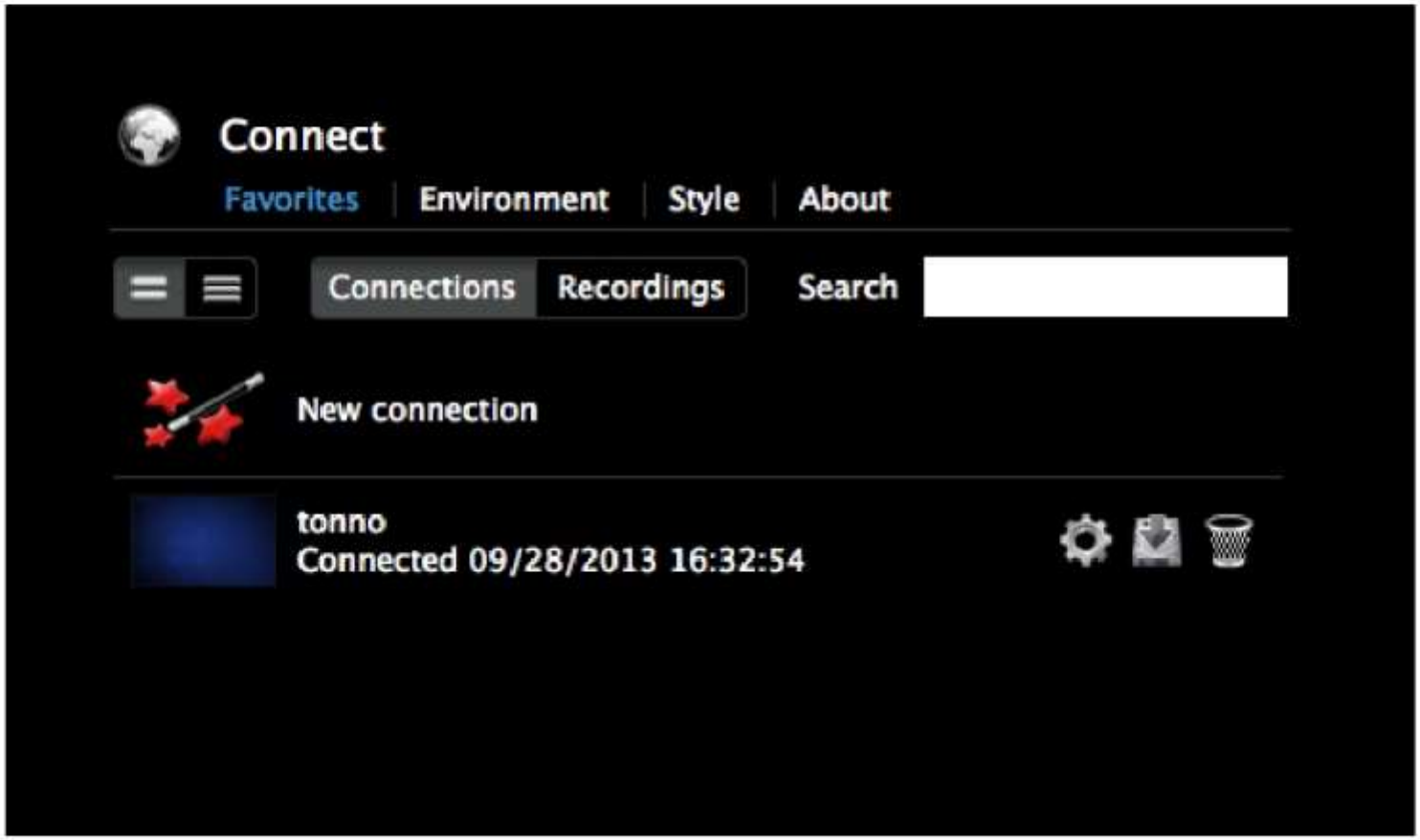
Cluster di Calcolo: nodi di login

3. Desktop Remoto (solo su tonno e bitonno)

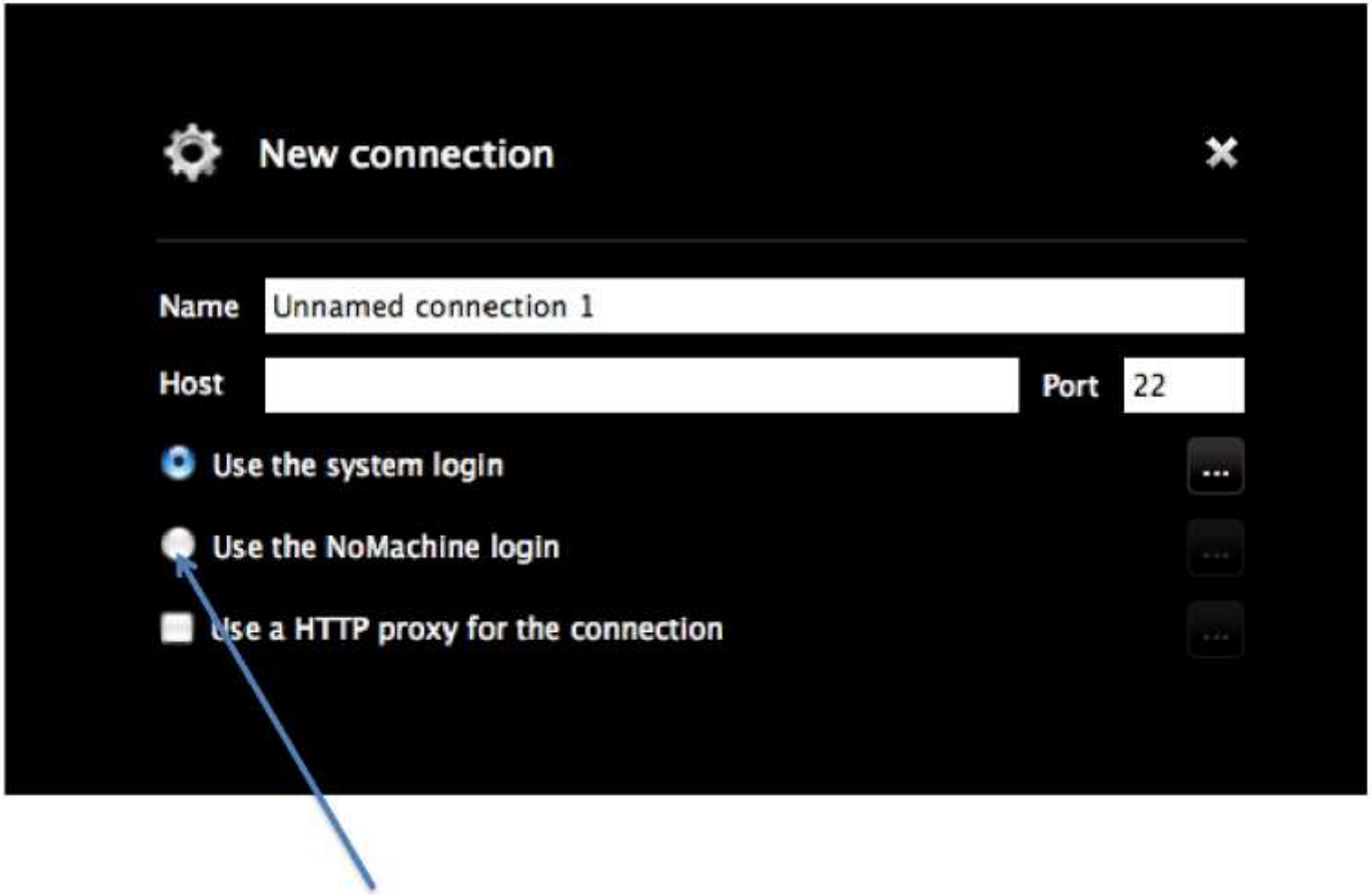
Basato su FreeNX, connessione su porta 22 (ssh)

- Scaricare il Player per il proprio sistema operativo
  http://www.nomachine.com/download
- Installarlo sul proprio pc
- Eseguire NoMachine Player

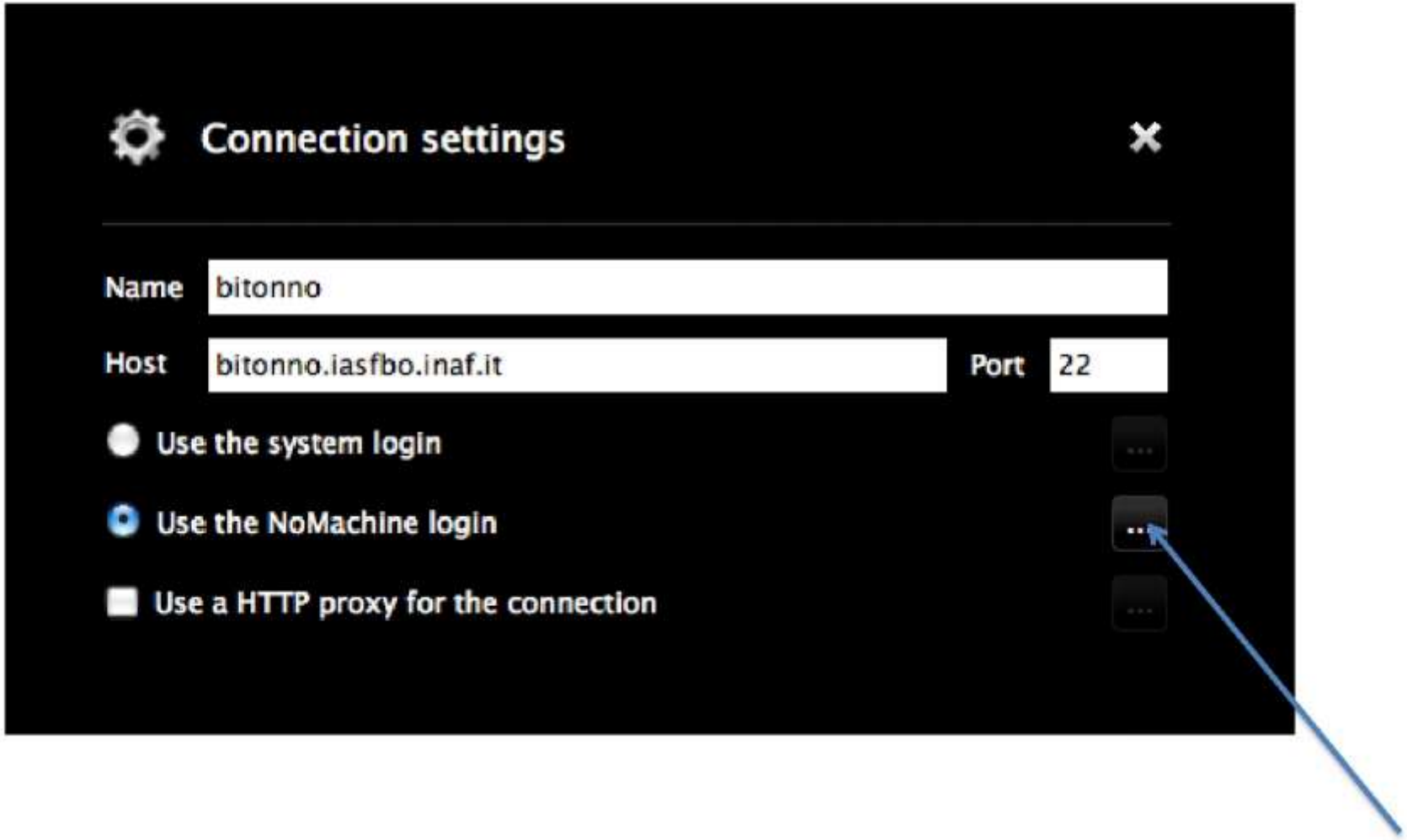
Cluster di Calcolo: nodi di login - FreeNX



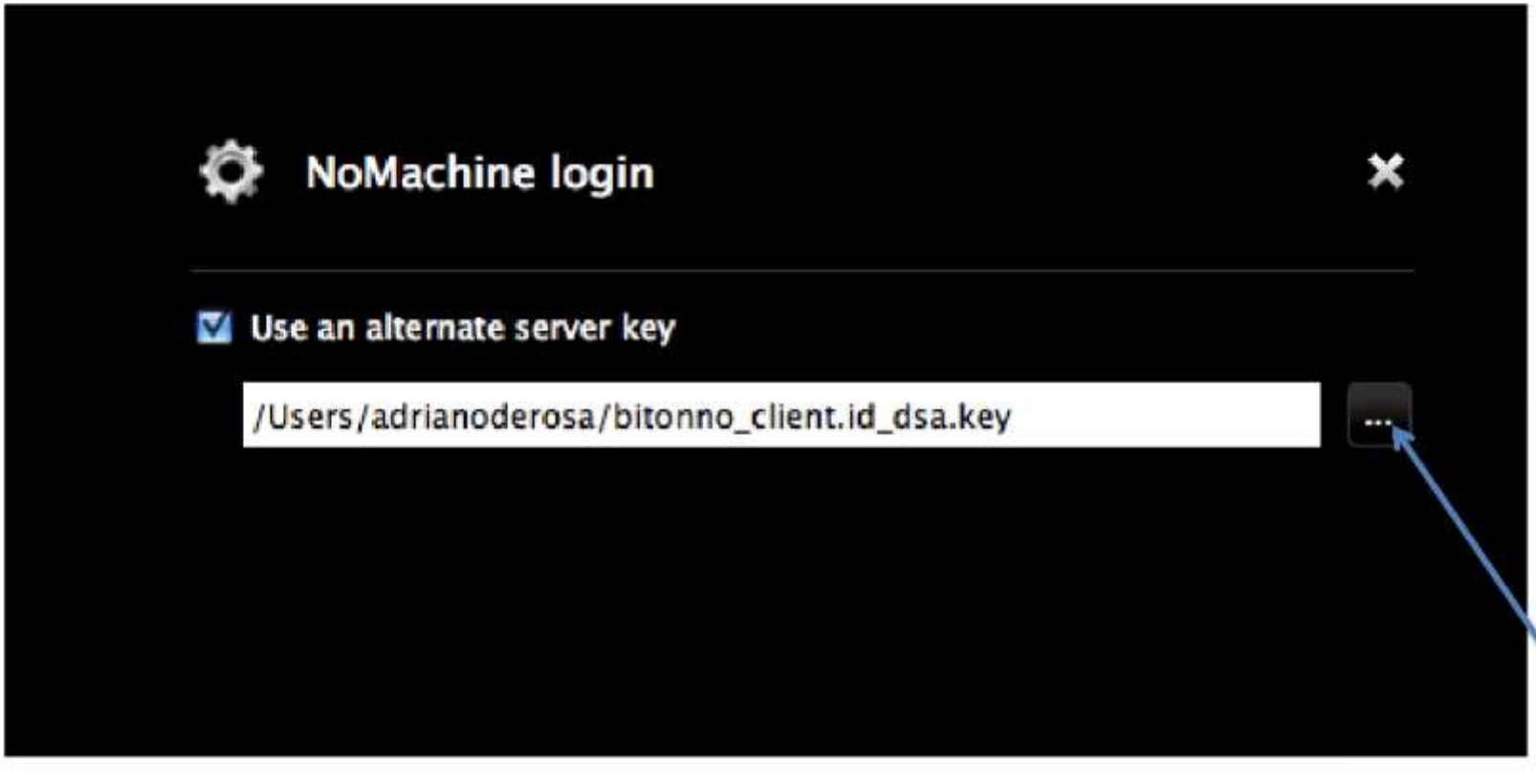
Cluster di Calcolo: nodi di login - FreeNX



Cluster di Calcolo: nodi di login - FreeNX



Cluster di Calcolo: nodi di login - FreeNX

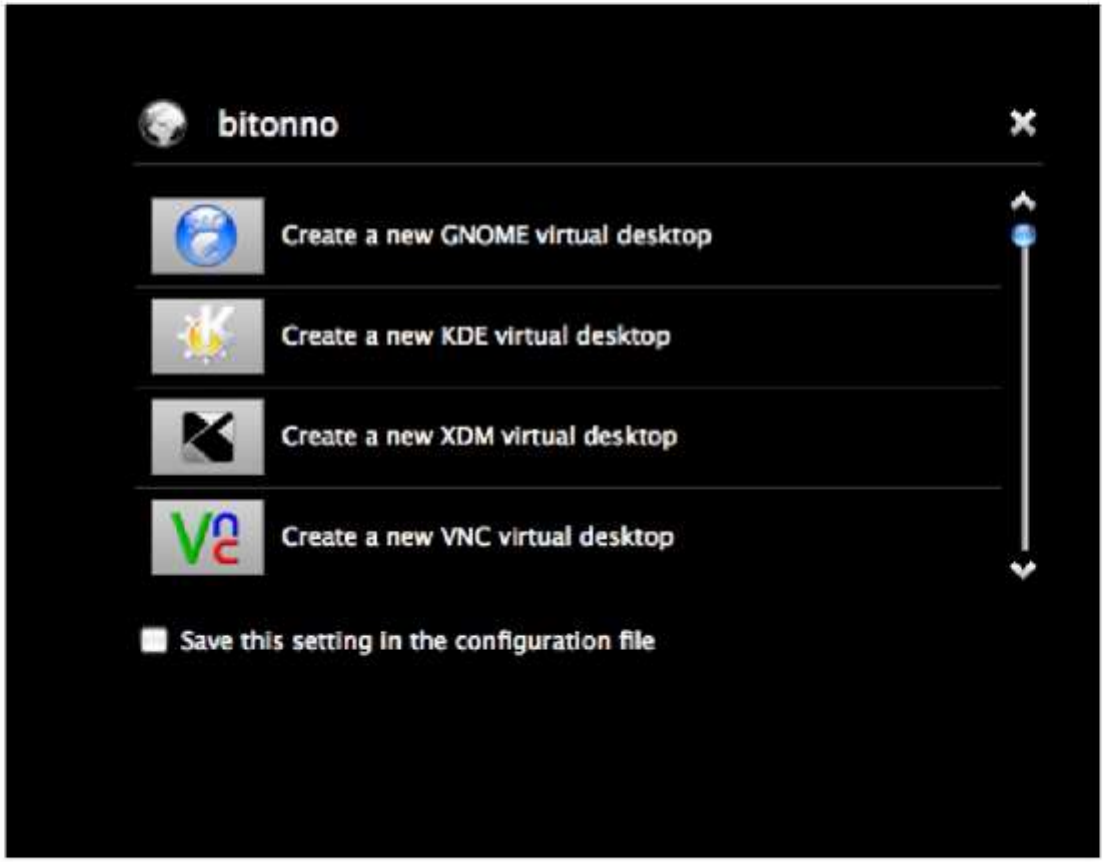


Le chiavi dei server sono nella cartella /prod_iasfbo/FREENX_KEY

Occorre quindi collegarsi PRIMA al server per scaricare in locale la chiave

Cluster di Calcolo: nodi di login - FreeNX

Cluster di Calcolo: nodi di login - FreeNX



Utilizzare preferibilmente GNOME perche' molto piu' leggera di KDE

Cluster di Calcolo: ambiente di sviluppo

- Il cluster ha un ambiente di sviluppo completamente modulare, ovvero, per evitare conflitti fra i programmi installati, occorre "caricare" l'ambiente opportuno (programma e versione) per i propri task.
- La gestione dell'ambiente viene effettuata mediante il comando **module**

SOFTWARE SCIENTIFICO @ CLUSTER DI ISTITUTO

Gli utenti che volessero analizzare dati provenienti da *diversi satelliti* senza dover installare il software appropriato sul proprio computer, **possono farlo attraverso il cluster di Istituto**, a cui possono accedere semplicemente facendo il login sui server front-end tonno o bitonno con il proprio username e password.

Sul cluster sono installati librerie e software scientifici pubblici: oltre ai compilatori come GCC, librerie MPI o di calcolo scientifico o di accesso ai dati scientifici (Cfitsio e HEALPIX) sul cluster sono installati i software scientifici per la riduzione dei dati relativi a diverse missioni spaziali. Ogni pacchetto scientifico di riduzione dati e' corredato dai relativi file di calibrazione che vengono aggiornati periodicamente.

Sul cluster d'Istituto possono esser presenti anche **codici di calcolo e dati 'proprietari'** delle missioni spaziali, come ad esempio viene fatto per le missioni Planck e Agile, garantendo la riservatezza dei dati e dei software.

SOFTWARE DI SVILUPPO			
	Tipo	**Descrizione**	**Ultima versione su tonno**
Compilatori e librerie MPI	GCC	linux distrution default gnu collection compiler	4.4.6
	Scalasca	opensource profiler	1.3.2
	Intel	intel linux compiler suite	13.0
	openMPI	opensource MPI library	1.6.3
Librerie di calcolo scientifico	Scalapack	MPI based dense matrix computaional library	2.0.2
	Lapack	dense matrix computational library	
	MKL	intel Lapack version	13.0
	ACML	amd Lapack version	5.3.0
	NAG	numerical software	mark22
	GSL	numerical software	1.15
Librerie per accesso ai dati astrofisici (I/O)	HEALPIX	map visualization and analisys	3.3
	Cfitsio	scientific file format I/O library	3.32
	HDF5	hiercal data format I/O library	1.8.1

SOFTWARE DI SVILUPPO			
	Tipo	**Descrizione**	**Ultima versione su tonno**
Software di visualizzazione dati scientifici	Paraview	visualizationand postprocessing suite	3.98
	Paraview MPI		3.98
	DS9	DS9 is an astronomical imaging and data visualization application. It supports FITS images, binary tables, multiple frame buffers, etc.	DS9 v. 7.1 with funtools
	gnuplot	gnuplot is a portable command-line driven graphing utility for Linux, OSX and many other platforms. It allows to visualize	4.4.2
	Grace, Xmgr	Grace is a plotting tool based on X11 and Motif used to produce 2D plots and for graphical selection. It runs on any version of Unix-like OS. Grace is a descendant of Xmgr.	5.1.22
	Smongo	SM is an interactive plotting package for creating graphs that mostly works with vectors.	SM v. 2.4

SOFTWARE DI SVILUPPO

	Tipo	Descrizione	Ultima versione su tonno
Suzaku, Swift →	HEASoft	HEASoft is the generalized term for the full set of software distributed from the HEASARC. In particular, it is a unified release of the FTOOLS and XANADU packages.	HEASoft v. 6.14
FERMI →	Fermi Science Tools	Fermi-LAT data analysis tools	ScienceTools v9r27p1
Chandra →	CIAO (Chandra Interactive Analysis of Observations)	CIAO is a data analysis system written for the needs of users of the Chandra X-ray observatory	CIAO v. 4.5 caldb v. 4.5.7
XMM-Newton →	SAS (Scientific Analysis System)	SAS is a collection of tasks, scripts and libraries, specifically designed to reduce and analyze data collected by the XMM-Newton observatory.	SAS v. 13.0.0
Software per analisi dati missioni spaziali	SPEX	SPEX is a software package optimized for the analysis and interpretation of high-resolution X-ray spectra, in particular obtained by current X-ray observatories like XMM-Newton, Chandra, Suzaku.	SPEX v. 2.01.05
Integral →	OSA (Integral Off-Line Scientific Analysis)	The OSA software package contains all tools to perform INTEGRAL data analysis of any of the four instruments onboard.	OSA v. 9.0
	IRAF (Image Reduction and Analysis Facility)	IRAF is a general purpose software system for the reduction and analysis of astronomical data. The main IRAF distribution includes a good selection of programs for general image processing and graphics, plus a large number of programs for the reduction and analysis of optical and IR astronomy data (the noao packages).	IRAF v. 2.15
	pyRAF	pyRAF is a command language for running IRAF tasks that is based on the Python scripting language.	pyRAF v. 2.0

SOFTWARE DI SVILUPPO			
	Tipo	**Descrizione**	**Ultima versione su tonno**
Software di calcolo e simulazioni scientifiche	IDL	The Interactive Data Language (IDL) is a programming language used for data analysis.	IDL-8.2
	MPIDL	MPI library for IDL	v.2.4
	FASTIDL	Task farming library for IDL	v.2.4
	MATHEMATICA	Computational software program used in scientific and mathematical fields for example.	MATHEMATICA-6.0
	Geant	GEometry ANd Tracking (GEANT) is a "toolkit for the simulation of the passage of particles through matter" as reported in the official website. It is a platform for simulations in high energy, nuclear and accelerator physics as well as medical and space science. It includes facilities for handling geometry, tracking, detector response, visualization and user interface.	Geant4

Cluster di Calcolo: comando module

Il comando module permette di gestire l'ambiente, ovvero impostare le variabili d'ambiente necessarie all'utilizzo dei software.

Per una lista completa delle opzioni:

[derosa@login01]~>man module

Oppure:

[derosa@login01]~>module

Possono sorgere dei conflitti per quegli utenti che hanno personalizzato, di propria scelta, l'ambiente al login

Cluster di Calcolo: comando module

Comandi fondamentali:

- Lista dei moduli dispononibili sul sistema:

[derosa@login01]~>module av

- Per caricare un modulo (ad esempio IDL-8.2):

[derosa@login01]~>module load IDL-8.2

- Per conoscere i moduli caricati:

[derosa@login01]~>module list

- Per ripulire l'ambiente:

[derosa@login01]~>module purge

- Per togliere un modulo da quelli caricati:

[derosa@login01]~>module unload IDL-8.2

Cluster di Calcolo: comando module

- NOTE PRATICHE:

1. E' possibile utilizzare il tasto <TAB> per il completamento del nome del modulo o per avere un lista dei moduli che iniziano con le lettere digitate (bash)

[derosa@login01]~>module load IDL-

IDL-6.4 IDL-7.0 IDL-7.0-astron IDL-7.1 IDL-7.1-astron IDL-8.0 IDL-8.0-healpix IDL-8.2

Cluster di Calcolo: comando module

2. Nel caso in cui oltre il comando module sia necessario inizializzare l'ambiente con il comando linux "source", o utilizzando comandi come heainit, ciaoinit…

NON E' SUFFICIENTE "module purge" per ripristinare l'ambiente di default, ma e' necessario aprire una nuova shell

Cluster di Calcolo: i moduli presenti

- I moduli sono suddivisi in **macrogruppi**:

/prod_iasfbo/modules/AGILE
/prod_iasfbo/modules/FERMI
/prod_iasfbo/modules/GNU
/prod_iasfbo/modules/IBM
/prod_iasfbo/modules/IDL
/prod_iasfbo/modules/INTEGRAL
/prod_iasfbo/modules/INTEL
/prod_iasfbo/modules/JAVA
/prod_iasfbo/modules/PLANCK
/prod_iasfbo/modules/SSDG
/prod_iasfbo/modules/SYSTEM
/prod_iasfbo/modules/WOLFRAM

I NOMI DEI MODULI SONO UNIVOCI A PRESCINDERE DAI MACROGRUPPI, OVVERO NON POSSONO ESSERCI COMUNQUE MODULI CON LO STESSO NOME IN MACROGRUPPI DIFFERENTI.
I GRUPPI FONDAMENTALI SONO INTEL, GNU, IBM

Cluster di Calcolo: i moduli presenti

- **INTEL**: contiene diverse versioni del compilatore e le librerie compatibili con la suite intel di compilazione. (Eccezione per le Gnu Scientific Libraries che compilate con gnu hanno un'interfaccia compatibile con intel)
- **GNU**: contiene le librerie compatibili con i compilatori gnu 4.4.7. Sono presenti anche versioni successive del compilatore per test
- **IBM**: ambiente per il calcolo parallelo di IBM, nelle versioni GNU e INTEL

Cluster di Calcolo: moduli convenzioni e makefile

- Ogni modulo inerente ad una libreria genera due variabili d'ambiente fondamentali per facilitare la scrittura dei Makefiles

1. NOMELIBRERIA_INCDIR: dove sono contenuti i files .h e/o .mod
2. NOMELIBRERIA_LIBDIR: dove sono contenuti i files .a e/o .so

Ad esempio i moduli di healpix generano delle variabili d'ambiente: HEALPIX_INCDIR e HEALPIX_LIBDIR

In questo modo e' possibile cambiare versione della libreria senza dover modificare i makefiles.

Cluster di Calcolo: utilizzo

- **Nodi di login**: pre/post processing di dati di piccole dimensioni, compilazione, accesso ai filesystem. VI E' CONCORRENZA FRA I PROCESSI, QUINDI FRA GLI UTENTI (**fair play**)
- **Nodi di calcolo**: sessioni interattive e batch per pre/post processing, analisi dati, simulazioni. NON VI E' CONCORRENZA FRA I PROCESSI SE NON PER LA RETE E L'ACCESSO AI FILESYSTEMS. Occorre DICHIARARE le risorse (ram e cpu) necessarie al momento della richiesta!

Cluster di Calcolo: accesso ai nodi

L'accesso ai nodi viene regolamentato da uno scheduler (gestore delle risorse). La scelta e' ricaduta su Loadleveler di IBM in quanto unico prodotto di livello enterprise fornito alle strutture educational senza costi (allo stato attuale).

Piu' complesso di quello prima impiegato (PBS) ma con feature non presenti nelle versioni opensource, come la gestione "rigida" delle risorse richieste dai job, la memory e processor affinity.

Cluster di Calcolo: Loadleveler

- Premesse:

Gli scheduler gestiscono le risorse del sistema permettendo di definire le priorita' e le risorse disponibili per diverse tipologie di job, introducendo il concetto di code.

Ad ogni coda puo' quindi esser assegnato un tempo massimo di esecuzione, un numero massimo di processori utilizzabili, una priorita'.

Puo' esser definito inoltre il numero massimo di job che ogni utente puo' sottomettere contemporaneamente in ogni coda.

Allo stato attuale non sono imposti limiti al numero dei job che ogni utente puo' sottomettere nelle code. **Tenete presente che quello che interessa al sistema (teoria delle code) non e' tanto il numero di job che un utente sottomette ma la DURATA del job. Job molto lunghi portano la coda a congestionarsi.**

Cluster di Calcolo: Loadleveler

Tipologie di job:

1. **Interattivo**: del tutto analogo ad avere una shell aperta sul nodo con forwarding dell'X

1. **Batch**: sono job che non richiedono interazione con l'utente nella loro esecuzione o con il server X (i job in batch non possono aprire finestre grafiche)

Cluster di Calcolo: Loadleveler - code definite (@class)

- Visibili con il comando llclass da console dei nodi di login

1. **interactive**: interactive job queue 6hr
2. **medium**: normal job queue 6hr
3. **large**: large job queue 240hr

Allo stato attuale non vi sono limitazioni sul numero di core che possono esser impegnati dalla coda large… Se sara' **proprio** necessario verranno definite!

Cluster di Calcolo: Loadleveler - principi di funzionamento

- Lo scheduler richiede che siano definiti dei parametri per ogni job.
- I parametri vengono inseriti in delle keyword predefinite dello scheduler.
- Le keyword vanno inserite in un file, detto FILE DI SOTTOMISSIONE, insieme ai comandi che devono essere eseguiti.
- Le keyword hanno la forma #@nome_keyword.
- Iniziando con # qualunque interprete comandi saltera' tale riga.
- Alcuni possibili template utilizzabili possono essere trovati nella cartella /RossiFumi/SCRIPT_TEMPLATE
- L'estensione del file di sottomissione utilizzata, ".ll", e' una mia convenzione. Puo' esser utilizzata qualunque estensione!

Cluster di Calcolo: Loadleveler - comandi

- **Sottomissione di job interattivi:**

llrun -f job_file.ll

- **Sottomissione di job batch:**

llsubmit job_file.ll

```
[derosa@login01]TESTING>llsubmit job1.ll
llsubmit: The job "login01.443" has been submitted.
```

- **Visualizzazione di tutti i job:**

llq

```
[derosa@login01]TESTING>llq
Id                  Owner Submitted ST PRI Class        Running On
------------------- --------- ---------- -- --- ---------- ----------
login01.443.0          derosa 10/17 21:06 ST 50 large        node01

5 job step(s) in queue, 0 waiting, 1 pending, 4 running, 0 held, 0 preempted
```

Cluster di Calcolo: Loadleveler - comandi

Informazioni aggiuntive di llq con llq -s job_number

[derosa@login01]TESTING>llq -s 443

===== EVALUATIONS FOR JOB STEP login01.443.0 =====

Step state : Starting
Since job step status is not Idle, Not Queued, or Deferred, no attempt has been made to determine why this job step has not been started.

- **Cancellazione di un job:**

llcancel job_number

[derosa@login01]TESTING>llcancel login01.443.0
llcancel: Cancel command has been sent to the central manager.

Cluster di Calcolo: Loadleveler - comandi

- lliasf - permette di vedere lo stato di tutti l nodi
- llsubmit - permette di sottomettere job in batch
- llrun - permette di sottomettere job interattivi
- llq - mostra lo stato delle code
- llq -s job_name - permette di vedere l dettagli di un job in coda
- llhold job_name - permette di rilasciare un job in stato di hold (trattenuto dal sistema)
- llcancel job_name - permette di rimuovere un job dalla coda

Cluster di Calcolo: Loadleveler - stato di un job (1)

- Lo stato di un job e' visibile con llq

Job status	llq	
Canceled	CA	The job has been canceled as by the llcancel command.
Completed	C	The job has completed.
Complete Pending	CP	The job is completed. Some tasks are finished.
Deferred	D	The job will not be assigned until a specified date. The start date may have been specified by the user in the Job Command file or it may have been set by LoadLeveler because a parallel job could not obtain enough machines to run the job.
Idle	I	The job is being considered to run on a machine though no machine has been selected yet.
NotQueued	NQ	The job is not being considered to run. A job may enter this state due to an error in the command file or because LoadLeveler can not obtain information that it needs to act on the request.
Not Run	NR	The job will never run because a stated dependency in the Job Command file evaluated to be false.
Pending	P	The job is in the process of starting on one or more machines. The request to start the job has been sent but has not yet been acknowledged.
Rejected	X	The job did not start because there was a mismatch or requirements for your job and the resources on the target machine or because the user does not have a valid ID on the target machine.
Reject Pending	XP	The job is in the process of being rejected.

stato di un job (2)

Removed	RM	The job was canceled by either LoadLeveler or the owner of the job.
Remove Pending	RP	The job is in the process of being removed.
Running	R	The job is running.
Starting	ST	The job is starting.
Submission Error	S	The job can not start due to a submission error. Please notify the Bluedawg administration team if you encounter this error.
System Hold	X	The job has been put in hold by a system administrator.
System User Hold	S	Both the user and a system administrator has put the job on hold.
Terminated	HS	The job was terminated, presumably by means beyond LoadLeveler's control. Please notify the Bluedawg administration team if you encounter this error.
User Hold	TX	The job has been put on hold by the owner.
Vacated	H	The started job did not complete. The job will be scheduled again provided that the job may be reschellued.
Vacate Pending	V	The job is in the process of vacating.
Checkpointing	VP CK	Indicates that a checkpoint has been initiated.

Cluster di Calcolo: Loadleveler - interactive - file di sottomissione

```
#!/bin/bash
#
# SCRIPT PER UTILIZZARE SESSIONI GRAFICHE INTERATTIVE
# PARTITO IL JOB CARICARE IL MODULO INTERACTIVE APPENA APERTA
#LA SESSIONE SUL NODO —!!!!!! Module load INTERACTIVE

#@ job_name = test #@
job_type = serial
#@ environment= $PATH; $DISPLAY; COPY_ALL
#@ class = interactive
#@ wall_clock_limit = 2:00:00
# numero di cpu richieste e memoria richiesta
#@ resources = ConsumableCpus(1) ConsumableMemory(2000Mb)
#@ notify_user = username@iasfbo.inaf.it
#@ queue
```

Cluster di Calcolo: Loadleveler - interactive - sottomissione del job - comando llrun

llrun -f job_file.ll

Sintassi completa di llrun:

llrun -N <nodes> -n <mpi_processes> -t <threads> executable

Usage :
llrun [<options>] <exe> [<user_or_poe_args>] <options>:
-N: number of nodes (Default: 1)
-p: total number of processes, same as
-n (Default: 1) -n: total number of processes, same as -p (Default: 1)
-P: Task per node
-t: number of threads per process (Default: 1)
-s: use SMT/Hyperthreading (oversubscribe nodes)
-A: No automatic adjustment of number of nodes (Default: automatic adj.) -f: Pin processes and Threads to Fixed physical cores (IBM MPI only) -b: submit batch job -c: submit to class (default: test) -m: email -w: submit batch job with wallclock limit (Default: 00:15:00) -i: include content of file before parallel execution -I: include content of file after parallel execution -o: do not run/submit job but save to file -h: help (this message) -v: verbose (Default) -V: NO verbose

Cluster di Calcolo: Loadleveler - batch

Esistono tre tipologie di job:

- **serial** : job che utilizzano 1 processo; sono quindi compresi i job multithreaded, cioe' con 1 processo e n thread (n<=64 per i nostri nodi).
- **MPICH**: job che utilizzano n processi (n<=256) utilizzando le librerie MPI non IBM, come OpenMPI. La somma dei processi e dei thread per processo deve comunque non superare 256.
- **parallel:** job che utilizzano n processi (n<=256) utilizzando le librerie MPI IBM. La somma dei processi e dei thread per processo deve comunque non superare 256. La differenza sostanziale fra MPICH e parallel e' nella stretta integrazione dello scheduler (loadleveler) con l'ambiente MPI di IBM, che permette di definire keyword particolari al momento della sottomissione.

NOTA IMPORTANTE MPICH-parallel: gli eseguibili compilati per MPICH non sono compatibili con parallel e viceversa! Per usare parallel occorre compilare i sorgenti con i moduli IBM. Per utilizzare MPICH occorre compilare con OpenMPI!

Cluster di Calcolo: Loadleveler - serial

```
#!/bin/sh
#
# SCRIPT DI SOTTOMISSIONE DI JOB SERIALI
#

#@ shell = /bin/bash #@
job_name = test #@ job_type =
serial
#@ environment= COPY_ALL
#@ class = large
# E' POSSIBILE CHIEDERE PIU' DI UNA CPU per I job multithreded
#@ resources = ConsumableCpus(1) ConsumableMemory(4000Mb)
#@ wall_clock_limit = 3:00:00
#@ error = job.$(jobid).err
#@ output = job.$(jobid).out
#@ notify_user = username@iasfbo.inaf.it
#@ queue

  module load MODULI NECESSARI

./EXECUTABLE
```

Cluster di Calcolo: Loadleveler - MPICH - single thread

```
#!/bin/bash
#
# Script per sottomettere un job openmpi
##

#@ shell = /bin/bash #@ job_name
= test
#@ job_type = MPICH
#@ environment= COPY_ALL
#@ class = large
#@ wall_clock_limit = 12:00:00

#RISORSE PER OGNI TASK esempio per processo single thread
#@ resources = ConsumableCpus(1) ConsumableMemory(1000Mb)
# CON PIU' DI UN NODO OCCORRE CARICARE IL MODULO OPENMPI PRIMA DI SOTTOMETTERE
#@ node = 1
#@ tasks_per_node = 2
#@ error = job1.$(jobid).err
#@ output = job1.$(jobid).out
#@ notify_user = username@iasfbo.inaf.it
#@ queue

  module load

  mpirun executable <executable-parameters>
```

Cluster di Calcolo: Loadleveler - MPICH - multithreaded

```
#!/bin/bash
#
# Script per sottomettere un job openmpi
##

#@ shell = /bin/bash #@ job_name
= test
#@ job_type = MPICH
#@ environment= COPY_ALL
#@ class = large
#@ wall_clock_limit = 12:00:00

#RISORSE PER OGNI TASK esempio per processo multithreaded
#@ resources = ConsumableCpus(4) ConsumableMemory(4000Mb)
#CON PIU' DI UN NODO OCCORRE CARICARE IL MODULO OPENMPI PRIMA DI SOTTOMETTERE
#@ node = 1
#@ tasks_per_node = 2
#@ error = job1.$(jobid).err
#@ output = job1.$(jobid).out
#@ notify_user = username@iasfbo.inaf.it
#@ queue

 module load

 export OMP_NUM_THREADS=4

 mpirun executable <executable-parameters>
```

Cluster di Calcolo: Loadleveler - MPICH - multithreaded

- Occorre tenere presente che in tutto ci sono 64 cores di calcolo su ogni nodo.
- I thread non possono andare su un nodo diverso dal processo che lo ha generato
- Se volessi 32 thread per un processo, ne deriva che su un nodo possono girare solo 2 processi
- Se volessi 64 thread per un processo, ne deriva che su un nodo puo' girare solo 1 processo
- Se volessi 128 thread per un processo, ne deriva che NON PARTIRA' FINO A QUANDO NON COMPREREMO E NON METTEREMO NEL CLUSTER UN SERVER CON 128 CORES

Cluster di Calcolo: Loadleveler - MPICH - multithreaded

Se impostassi:

#@ resources = ConsumableCpus(4) ConsumableMemory(4000Mb)

e

export OMP_NUM_THREADS=8

Il numero di cores di calcolo riservati dal sistema e' 4. Gli 8 threads generati dall'eseguibile possono comunque utilizzare solo i 4 cores riservati dal sistema condividendoli. Non e' escluso a priori che alcuni codici possano trarre beneficio da un overthreading sui cores di calcolo, sicuramente non quelli che eseguono operazioni di calcolo algebrico. L'overthreading non e' l'hyper-threading! Le CPU logiche ottenute via hardware con l'hyperthreading sarebbero eventualmente gia' considerate dallo scheduler, ma i nostri nodi non hanno la tecnologia hyperthreading! Solo Intel e IBM la possiedono, i nostri nodi sono AMD.

A parte IBM (architettura POWER), Intel offre al massimo 48 cores con 2 threads per core, ad un costo quasi triplo dei 64 cores AMD. Allo stato attuale quindi AMD offre la soluzione costi/prestazioni migliore sul mercato per nodi HPC.

Cluster di Calcolo: Loadleveler - parallel - single thread

```
#!/bin/bash
#
# Script per sottomettere un job parallel
##

#@ shell = /bin/bash #@ job_name =
test
#@ job_type = parallel
#@ environment= COPY_ALL
#@ class = large
#@ wall_clock_limit = 12:00:00
#RISORSE PER OGNI TASK
#@ resources = ConsumableCpus(1) ConsumableMemory(100Mb)
##@ node = 1
##@ tasks_per_node = 2
# in alternativa a task_per_node
#@ total_tasks = 4
#@ error = job1.$(jobid).err
#@ output = job1.$(jobid).out
#@ notify_user = derosa@iasfbo.inaf.it
#@ queue

 date

 module load IBM-ENV-BASE-GNU

 poe ./test

 date
#
```

Cluster di Calcolo: Loadleveler - parallel

- L'eseguibile viene lanciato attraverso il comando poe.
- Non e' necessario caricare l'ambiente mpi nel caso si utilizzino piu' nodi prima della sottomissione.
- Permette l'utilizzo di keyword aggiuntive come ad esempio:

```
#@ task_affinity = Core(4)
#@ parallel_threads = 4
```

La keyword task_affinity dice allo scheduler che vogliamo che i 4 cores richiesti dal processo siano sullo stesso socket hardware. Per la nostra architettura, di tipo NUMA (non uniform memory access), ha importanza per ottimizzare le prestazioni nell'accesso alla memoria. Per come e' realizzato un nodo AMD ha un impatto elevato fino a 16 core.

A prescindere dall'impostazione particolare di questa keyword, lo scheduler, per l'impostazione generale data, cerchera' sempre di allocare le risorse per il processo tenendo conto di questi aspetti, ma non e' mandatorio!

Cluster di Calcolo: Loadleveler - parallel - multithreaded jobs

```
#!/bin/bash
#
# Script per sottomettere un job parallel
##

#@ shell = /bin/bash #@ job_name =
test
#@ job_type = parallel
#@ environment= COPY_ALL
#@ class = large
#@ wall_clock_limit = 12:00:00

#RISORSE PER OGNI TASK

#@ resources = ConsumableCpus(4) ConsumableMemory(1000Mb)

#@ task_affinity = Core(4)
#@ parallel_threads = 4

##@ node = 1
##@ tasks_per_node = 2
# in alternativa a task_per_node
#@ total_tasks = 4
#@ error = job1.$(jobid).err
#@ output = job1.$(jobid).out
#@ notify_user = derosa@iasfbo.inaf.it
#@ queue

 date

 module load IBM-ENV-BASE-GNU

 poe ./test

 date
#
```

Cluster di Calcolo: Loadleveler - multistep

- Loadleveler permette la definizione di job multistep, ovvero di job che sono costituiti da piu' step di analisi (simultanei o sequenziali), permettendo l'implementazione di pipeline di analisi dati, con un flusso logico fra i programmi da eseguire: esistono quindi keyword con gli operatori logici applicabili agli step.
- Inoltre offre la possibilita' di richiedere risorse differenti per ogni step, ottimizzando la gestione delle risorse del cluster. (Se un utente richiede 64 cores per poi utilizzarne 20 per l'80% del tempo…)

Cluster di Calcolo: Loadleveler - multistep

```
#! /bin/bash
# @ job_type=parallel
# @ step_name = step_0
# @ output parallel.$(jobid).$(stepid).out # @ error =
parallel.$(jobid).$(stepid).err
# @ resources = ConsumableCpus(1) ConsumableMemory(100Mb)
# @ class = medium
## @ node = 1
# @ total_tasks = 2
# @ queue

poe ./test0


# step 1
# @ step_name = step_1
#controllo sull'esito dello step 0 per l'esecuzione dello step1
# @ dependency = (step_0 == 0)
# @ output = parallel.$(stepid).out
# @ error = parallel.$(jobid).$(stepid).err
# @ resources = ConsumableCpus(1) ConsumableMemory(100Mb)
# @ class = medium
## @ node = 1
# @ total_tasks = 4
# @ queue

poe ./test1
```

Cluster di Calcolo: Loadleveler - autonomous

- Lo scopo e' quello di utilizzare lo scheduler per eseguire le linee di comnado contenute in un file.

Esempio di file contenente linee di comando:

autonomous.cmdfile

```
echo "Command 1 - stdout: Hello world from `hostname` task $LL_SAMPLE_TASK_ID"; echo "Command 1 - stderr: Hello world from `hostname` task $LL_SAMPL
E_TASK_ID">&2
echo "Command 2 - stdout: Hello world from `hostname` task $LL_SAMPLE_TASK_ID"; echo "Command 2 - stderr: Hello world from `hostname` task $LL_SAMPL
E_TASK_ID">&2
echo "Command 3 - stdout: Hello world from `hostname` task $LL_SAMPLE_TASK_ID"; echo "Command 3 - stderr: Hello world from `hostname` task $LL_SAMPL
E_TASK_ID">&2
echo "Command 4 - stdout: Hello world from `hostname` task $LL_SAMPLE_TASK_ID"; echo "Command 4 - stderr: Hello world from `hostname` task $LL_SAMPL
E_TASK_ID">&2
echo "Command 5 - stdout: Hello world from `hostname` task $LL_SAMPLE_TASK_ID"; echo "Command 5 - stderr: Hello world from `hostname` task $LL_SAMPL
E_TASK_ID">&2
```

Autonomous e' composto da un insieme di scripts foriti da IBM. Per ragioni storiche IBM utilizza la shell ksh (default shell di AIX).

Possono quindi sorgere dei problemi con eseguibili contenuti in pacchetti che richiedano una configurazione particolare dell'ambiente, soprattutto se I file di inizializzazione dell'ambiente sono disponibili solo per bash o csh.

Cluster di Calcolo: Loadleveler - autonomous

```
#! /bin/ksh
# @ job_type=MPICH
# @ step_name = step_0
# @ output = autonomous.$(host).$(jobid).$(stepid).out
# @ error = autonomous.$(host).$(jobid).$(stepid).err
# @ resources = ConsumableCpus(1) ConsumableMemory(100Mb)
# @ class = medium
## @ node = 1
# @ total_tasks = 2
# @ queue

/opt/ibmll/LoadL/resmgr/full/samples/autonomous/autonomous_master.ksh -f autonomous.cmdfile


# step 1
# @ step_name = step_1
# @ dependency = (step_0 == 0)
# @ output = autonomous.$(host).$(jobid).$(stepid).out
# @ error = autonomous.$(host).$(jobid).$(stepid).err
# @ resources = ConsumableCpus(1) ConsumableMemory(100Mb)
# @ class = medium
## @ node = 1
# @ total_tasks = 4
# @ queue

/opt/ibmll/LoadL/resmgr/full/samples/autonomous/autonomous_master.ksh -f autonomous.cmdfile1
```

Cluster di Calcolo: Loadleveler - embarrassing parallel jobs

```
#!/bin/bash
#
# Script per sottomettere un job embarrassing parallel basato su un unico file di input che contiene le linee di comando da eseguire
##

#@ shell = /bin/bash #@ job_name =
test
#@ job_type = parallel
#@ environment= COPY_ALL
#@ class = large
#RISORSE PER OGNI TASK
#@ resources = ConsumableCpus(1) ConsumableMemory(100Mb)
##@ node = 1
##@ tasks_per_node = 2
#@ total_tasks = 4
#@ error = job1.$(jobid).err
#@ output = job1.$(jobid).out
#@ notify_user = derosa@iasfbo.inaf.it
#@ queue

  date

  module load IBM-ENV-BASE-GNU

   poe ./test1.cmd

#soluzione alternativa, vengono pero' eseguite solo le prime n righe del file di input, con n pari al numero di processi mpi che sono stati richiesti
# poe -cmdfile test1.cmd


  date
#
```

Cluster di Calcolo: Loadleveler - keywords

#@ node = <min,max>

The scheduler attempts to get max nodes to run the job step, but will start the job step on min nodes if necessary.

#@ node = <number>

The scheduler will find number nodes on which to run the job step.

#@ tasks_per_node = <number>

Used in conjunction with #@ node, each node is assigned number tasks. tasks_per_node must be less or equal 64.

#@ total_tasks = <number>

Rather than specifying the number of tasks to start on each node, the total number of tasks in the job step across all nodes can be specified.

Cluster di Calcolo: Loadleveler - advanced keywords

#@ restart = yes | no

Specifies whether LoadLeveler considers a job to be restartable.

If restart=yes (default), and the job is vacated (e.g. in case of system errors) from its executing machine before completing, the central manager requeues the job. It can start running again when a machine on which it can run becomes available.

If restart=no, a vacated job is canceled rather than requeued.

#@ dependency = step_name operator value

value is usually a number that specifies the job return code to which the step_name is set.

It can also be one of the following LoadLeveler defined job step return codes:

CC_NOTRUN: The return code set by LoadLeveler for a job step which is not run because the dependency is not met. The value of CC_NOTRUN is 1002.

CC_REMOVED: The return code set by LoadLeveler for a job step which is removed from the system (because, for example, llcancel was issued against he job step). The value of CC_REMOVED is 1001.

Operators include ==, !=, <=, >=, <, >, &&, ||

A step can have dependencies on more than one step, like

#@ dependency = (step1 == 0) && (step2 >= 0)

Cluster di Calcolo: Loadleveler - variabili utilizzabili in alcune keywords

Several variables are available for use in job command files.

$(domain): The domain of the host from which the job was submitted.
$(home): The home directory for the user on the cluster selected to run the job.
$(user): The user name that will be used to run the job. This might be a different user name.
$(host): The hostname of the machine from which the job was submitted.
$(jobid): The sequential number assigned to this job by the schedd daemon.
$(stepid): The sequential number assigned to this job step when multiple queue statements are used with the job command file.
Some variables are set from other keywords defined in the job command file:
$(executable)
$(class)
$(comment)
$(job_name)
$(step_name)
$(base_executable): Automatically set from the executable keyword; consists of the executable file name without the directory component (basename).

Example: #@ output = $(home)/$(job_name)/$(step_name).$(schedd_host).$(jobid).$(stepid).out

Cluster di Calcolo: Loadleveler - variabili d'ambiente interne agli script

MP_CMDFILE

Determines the name of a POE commands file used to load the nodes of your partition. If set, POE will read the commands file rather than STDIN. Valid values are any file specifier. The value of this environment variable can be overridden using the -cmdfile flag.

Cluster di Calcolo: Loadleveler - variabili d'ambiente interne agli script

MP_PGMMODEL

Determines the programming model you are using. Valid values are spmd or mpmd. If not set, the default is spmd. The value of this environment variable can be overridden using the -pgmmodel flag.

Cluster di Calcolo: Loadleveler - variabili d'ambiente interne agli script

$MP_CHILD
utilizzo negli script:

```
cat<<EOF>job.sh
#!/bin/bash
echo " executing process no. \$MP_CHILD" > output.\$MP_CHILD
./myprog < input.\$MP_CHILD >> output.\$MP_CHILD
EOF
chmod u+x ./job.sh
poe ./job.sh
```

- This results in a parallel execution of four independent instances of the myprog code (different tasks):

```
> ./myprog < input.0 > output.0 (task 0) >
./myprog < input.1 > output.1 (task 1) >
./myprog < input.2 > output.2 (task 2) >
./myprog < input.3 > output.3 (task 3)
```

Cluster di Calcolo: Loadleveler - checkpoint

Keyword	Explanation
# @ notification	Specify when the user is sent e-mail
# @ notify_user	User to whom e-mails are sent
# @ checkpoint	Indicate if a job can be checkpointed
# @ ckpt_dir	Directory which contains the ckpt file
# @ ckpt_file	Name of the ckpt file

Cluster di Calcolo: Loadleveler - pipeline di analisi

- Utilizzando quindi gli script di sottomissione, e' possibile creare pipeline complesse di analisi dati.

Cluster di Calcolo: Filesystems

- /home - contiene le home degli utenti, e' soggetta a quota (20GB per utente) e viene eseguito un backup giornaliero (ext3)
- /prod_iasfbo - contiene i programmi e le librerie installate sul cluster, le immagini iso dei cd/dvd dei prodotti licenziati (/prod_iasfbo/iso) (ext3)
- /RossiFumi - spazio dei dati senza quota (al momento). Ogni utente ha una sua cartella sotto /RossiFumi/users/username nel quale poter mettere i propri dati. Non e' un vero spazio di scratch, in quanto i dati non vengono cancellati periodicamente in automatico. (GPFS)

Cluster di Calcolo: Filesystem RossiFumi

Il filesystem e' stato profondamente rivisto. Oltre ad essere l'unico a sfruttare la rete a 40Gb/s al meglio delle sue possibilita', e' stato modificato per permettere l'alta disponibilita' del filesystem.

Rimodernando server gia' in possesso all'istituto (sostituendo motherboard, controller..) ed acquistando un po' di dischi, il filesystem risiede su 4 server (il quarto e' in fase di ultimazione e privo di dischi).

Si ha cosi' una copia doppia dei dati e dei metadati del filesystem permettendo quindi che la rottura o la manutenzione di un intero server non comprometta la funzionalita' del sistema.

Tale scelta ha un costo elevato, poiche' ogni file occupera' uno spazio doppio. Se guardo la dimensione del filesystem con df, ad esempio, vedo la dimensione "grezza", cioe' senza tener conto della replicazione dei dati. Quindi dei 30 TB oggi disponibili in realta' solo 15TB sono occupabili.

Il fatto di avere una ridondanza dei dati e dei metadati, porta ad un incremento delle prestazioni del filesystem.

Per questo motivo e' fra le prime priorita' un ulteriore ampliamento dello spazio disponibile e soprattutto la volonta' di ridurre i filesystem "riservati" a vantaggio di una struttura piu' performante e sicura per tutto l'istituto.

Cluster di Calcolo: alta disponibilita' e backup

- Allo stato attuale solo dello spazio Home viene costantemente eseguito un backup.
- Lo spazio Home non e' in alta disponibilita'.
- Lo spazio del cluster e' in alta disponibilita' ed e' ora anche "backupabile", per meglio dire, le sue prestazioni in lettura e scrittura sono tali da permettere oltre alle normali operazioni effettuate dagli utenti anche un processo di backup, senza che il sistema subisca dei rallentamenti tali da renderlo inutilizzabile.
- Se ritenuto necessario e' possibile implementare un sistema di backup su nastro per lo spazio disco RossiFumi .

Cluster di Calcolo: roadmap

- Infiniband QDR, installata e disponibile per il cluster di calcolo. Essendo poche le porte disponibili e' strettamente riservata ai server del cluster in completo sharing con l'istituto. I nodi di calcolo che accedono ad infiniband possono anche esser di progetto, purche' possano esser utilizzati anche dagli altri utenti (lasciando una priorita' maggiore al progetto). I server di storage devono condividere tutti i loro dischi con il filesystem del cluster /RossiFumi non possono essere esportati fuori dall'ambiente del cluster.

- Aumento dello storage.

- Aumento dei nodi di calcolo (ottimale sarebbe 384 core)

- Introduzione di acceleratori, quali gpu e xeon phi

Cluster di macchine virtuali

- Basato sull'infrastruttura Oracle Virtual Machine (OracleVM), prodotto opensource di livello enterprise per la gestione delle infrastrutture virtuali
- Basato su Xen
- OracleVM e' fondato su RedHat linux
- Solo il supporto diretto di Oracle e' a pagamento

- Allo stato attuale ci sono 7 server dedicati a questa infrastruttura:

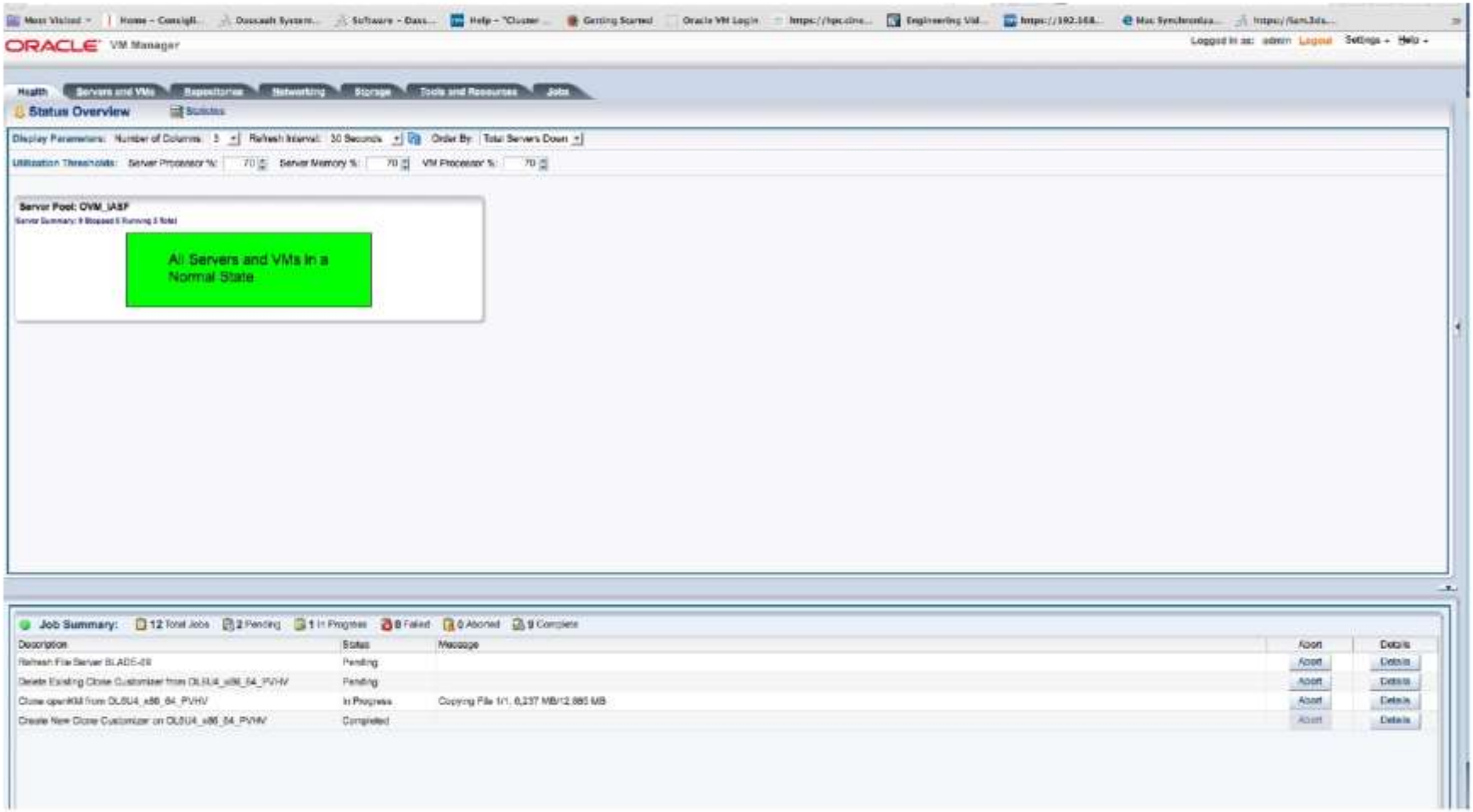
1. 1 server di management dell'ambiente con 16GB di ram e 8 cores
2. 2 server con 32GB di ram e 8 cores per VM 3. 3

server con 16GB di ram e 8 cores per VM

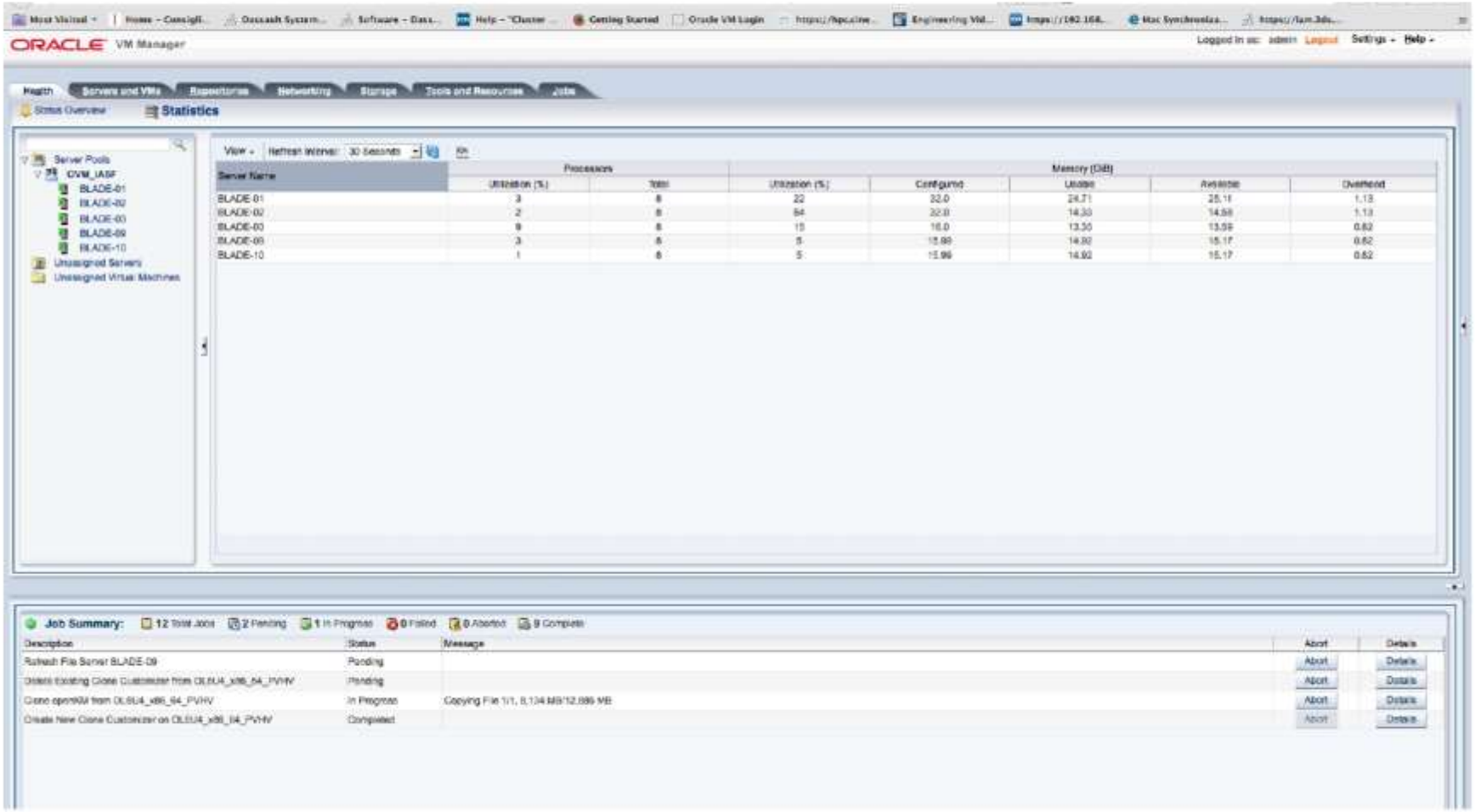
4. 1 nas server connesso via iscsi per lo spazio disco di alcune macchine virtuali (zimbra)

Complessivamente ci sono 40 cores e 110 GB di RAM per ospitare le macchine virtuali.

Cluster di macchine virtuali



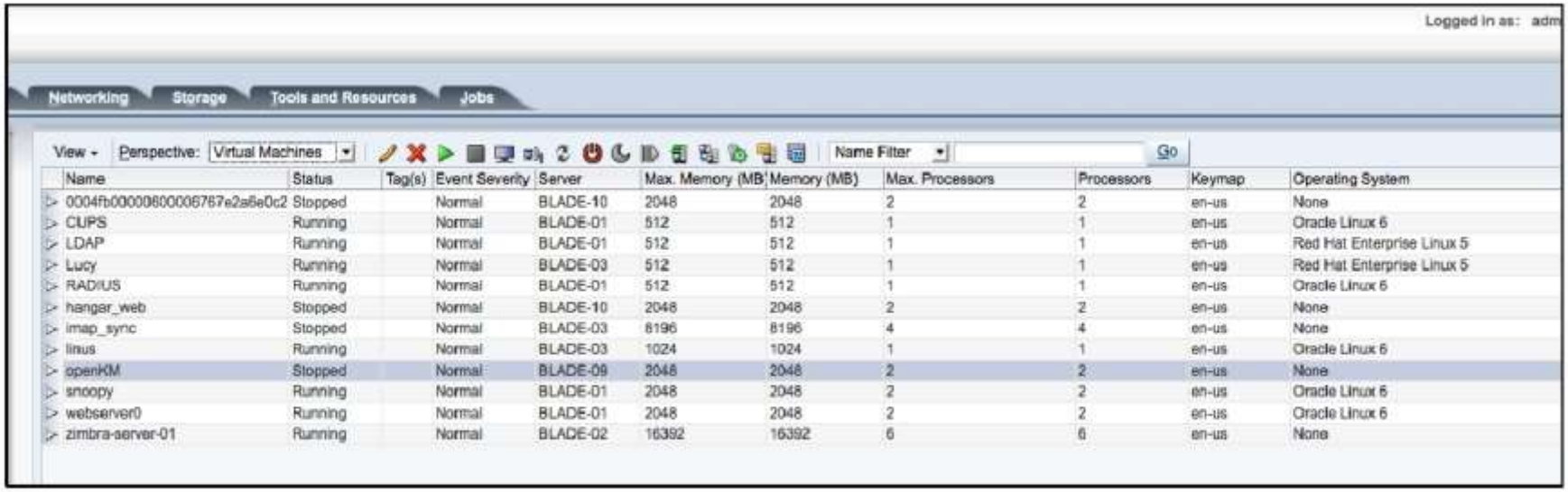
Cluster di macchine virtuali



Cluster di macchine virtuali

Tutti i servizi sono gia' offerti tramite l'utilizzo di macchine virtuali OracleVM:

1. Sistema di stampa
2. Server di autenticazione LDAP
3. Server di autenticazione RADIUS
4. Server ftp
5. Server DNS
6. Server per operazioni occasionali (imapsync)
7. Server squirrelmail - webmail
8. Server lucy per operazioni di processamento della posta per l'invio di sms



Name	Status	Tag(s)	Event Severity	Server	Max. Memory (MB)	Memory (MB)	Max. Processors	Processors	Keymap	Operating System
0004fb0000060000067e67e2a6e0c2	Stopped		Normal	BLADE-10	2048	2048	2	2	en-us	None
CUPS	Running		Normal	BLADE-01	512	512	1	1	en-us	Oracle Linux 6
LDAP	Running		Normal	BLADE-01	512	512	1	1	en-us	Red Hat Enterprise Linux 5
Lucy	Running		Normal	BLADE-03	512	512	1	1	en-us	Red Hat Enterprise Linux 5
RADIUS	Running		Normal	BLADE-01	512	512	1	1	en-us	Oracle Linux 6
hanger_web	Stopped		Normal	BLADE-10	2048	2048	2	2	en-us	None
imap_sync	Stopped		Normal	BLADE-03	8196	8196	4	4	en-us	None
linus	Running		Normal	BLADE-03	1024	1024	1	1	en-us	Oracle Linux 6
openKM	Stopped		Normal	BLADE-09	2048	2048	2	2	en-us	None
snoopy	Running		Normal	BLADE-01	2048	2048	2	2	en-us	Oracle Linux 6
webserver0	Running		Normal	BLADE-01	2048	2048	2	2	en-us	Oracle Linux 6
zimbra-server-01	Running		Normal	BLADE-02	16392	16392	6	6	en-us	None

Cluster di macchine virtuali: policy di utilizzo

- Il sistema di macchine virtuali e' in gestione al solo centro di calcolo.
- Il sistema puo' ospitare server virtuali di progetto, fornendo l'accesso come root al server virtuale ma non al sistema di macchine virtuali.
- Nel caso in cui dei progetti necessitino di spazio disco superiore ai 12GB di base di un server virtuale, occorrera' valutare come ottenere questo spazio, SEMPRE CON UNA POLITICA DI CONDIVISIONE DELL'INFRASTRUTTURA.
- Non sono ammessi server fisici dedicati.

Servizio di posta: Zimbra

- Soluzione di classe enterprise per la gestione di sistemi di posta, dopo Microsoft exchange e IBM Lotus domino e' il piu' venduto al mondo.
- Permette anche la soluzione di clustering di posta, anche se non implementata visto l'esiguo numero di caselle.
- Basato su strumenti opensource come postfix.
- Si differenzia di molto dalla versione opensource di zimbra, soprattutto per la gestione dei backup.
- Utilizza un database interno mysql per l'indicizzazione delle mail e degli allegati, permettendo di effettuare query non solo sul contenuto delle mail ma anche degli allegati.
- Il filesystem delle mail non e' accessibile agli utenti, facendo sparire ogni link simbolico a posizioni di filesystem non dedicati alla posta.

Zimbra: accesso alle caselle di posta

Accesso al Web Client Zimbra

E' possibile accedere al web client Zimbra dall'indirizzo https://zimbra.iasfbo.inaf.it inserendo la propria mail o il proprio username

La webmail permette di: Leggere e-mail, comporle, rispondere ad e-mail ricevute e gestire i propri contatti di posta. Oltre a questo Zimbra offre una serie di funzioni avanzate che andremo a vedere.

Esistono tre versioni della webmail Zimbra:

Standard: Per gli utenti che utilizzano connessioni lente con poche risorse HTML).

Avanzata: Per gli utenti con connessioni veloci (AJAX).

Mobile: Per telefoni cellulari e dispositivi mobili con connessioni 3G

E' possibile scegliere, in caso di necessità, una delle due diverse versioni principali cliccando, nel menù

Preferenze -> Generali -> Opzioni di accesso

(I dispositivi mobili saranno automaticamente riconosciuti al momento dell'accesso)

Zimbra: accesso alle caselle di posta

https://zimbra.iasfbo.inaf.it

Zimbra: accesso alle caselle di posta con zimbra desktop

Dal link presente sulla pagina di login

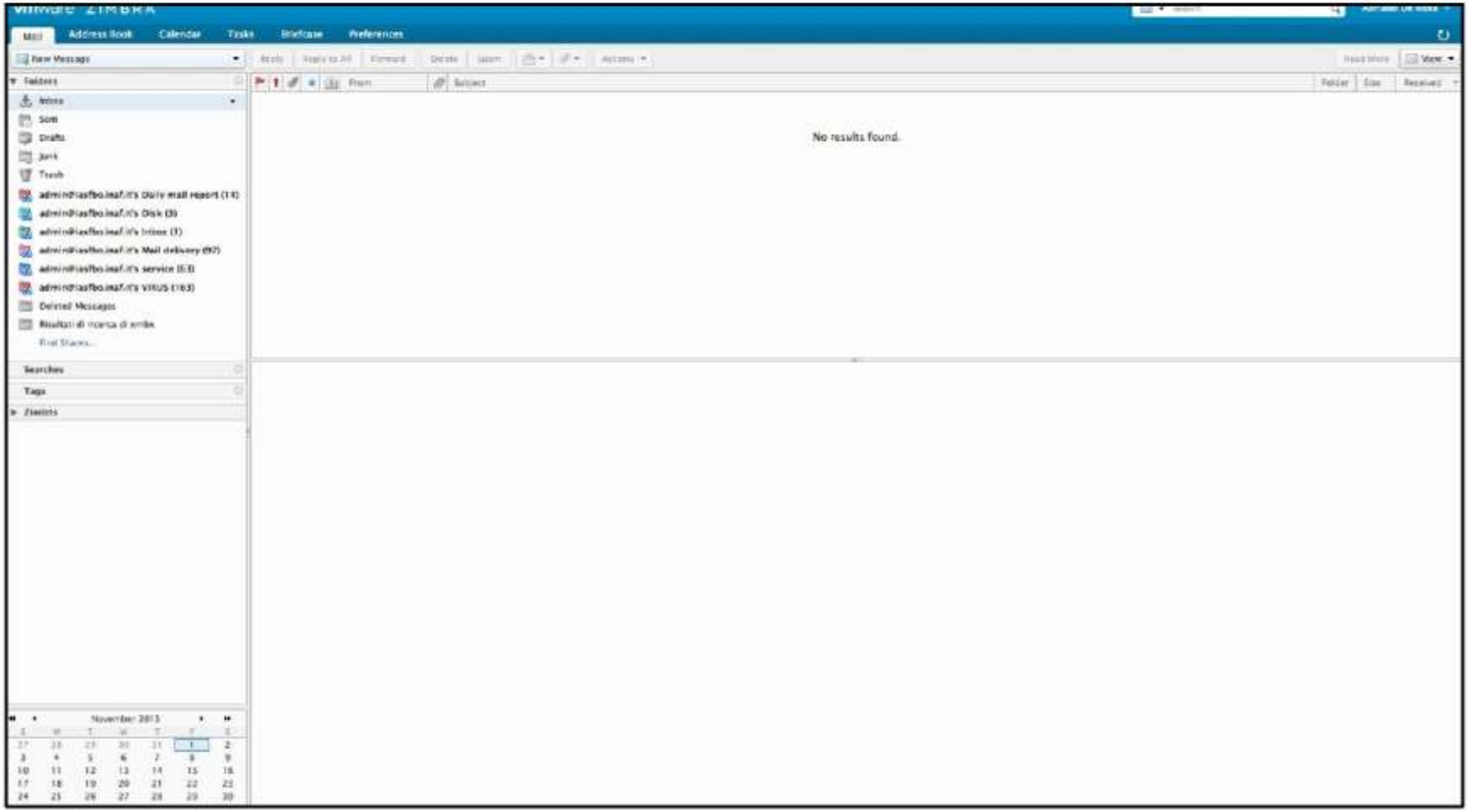
https://zimbra.iasfbo.inaf.it

e' possibile raggiungere la pagina web del client zimbra-desktop

http://www.zimbra.com/products/desktop.html

Da cui scaricare l'ultima versione del client e gli aggiornamenti.

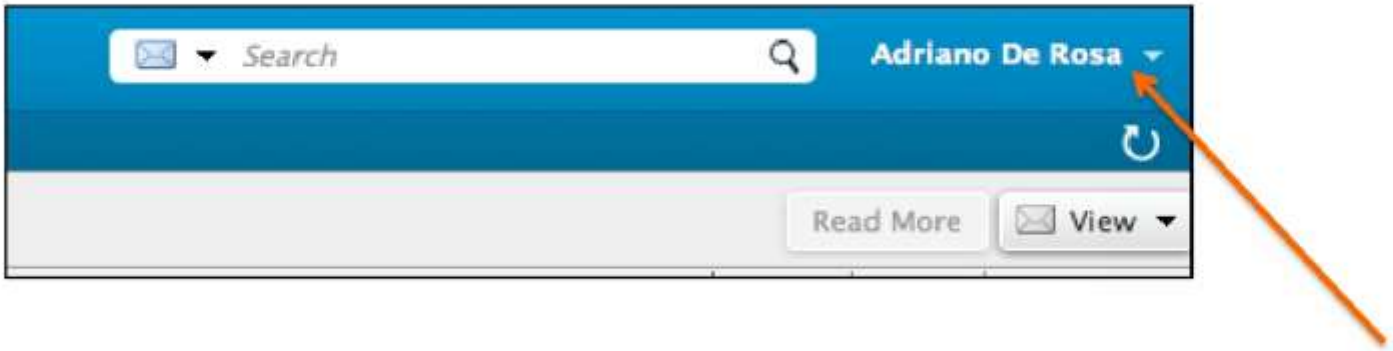
Zimbra: aspetto pagina web



Zimbra: Guida on line

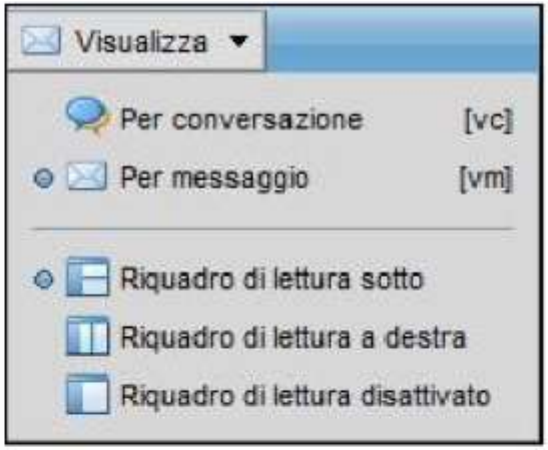
Guida Online

E' disponibile, all'apertura di zimbra, sull'angolo superiore destro dello schermo, il pulsante dello username, dove e' presente il link alla guida ufficiale sull'utilizzo di Zimbra.



Zimbra: modalita' di visualizzazione

- Modifica della visualizzazione

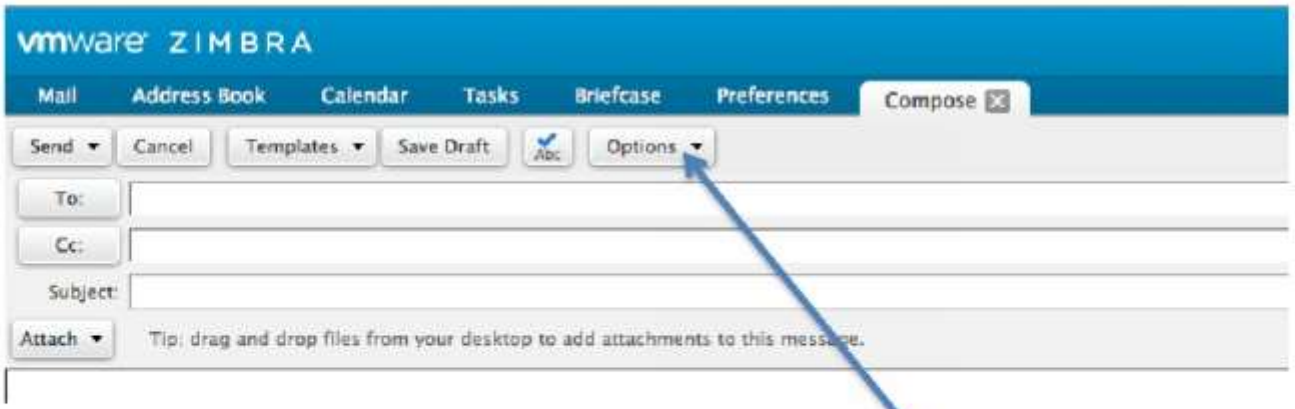


Zimbra: richiesta conferma lettura

- **Richiesta di conferma lettura**

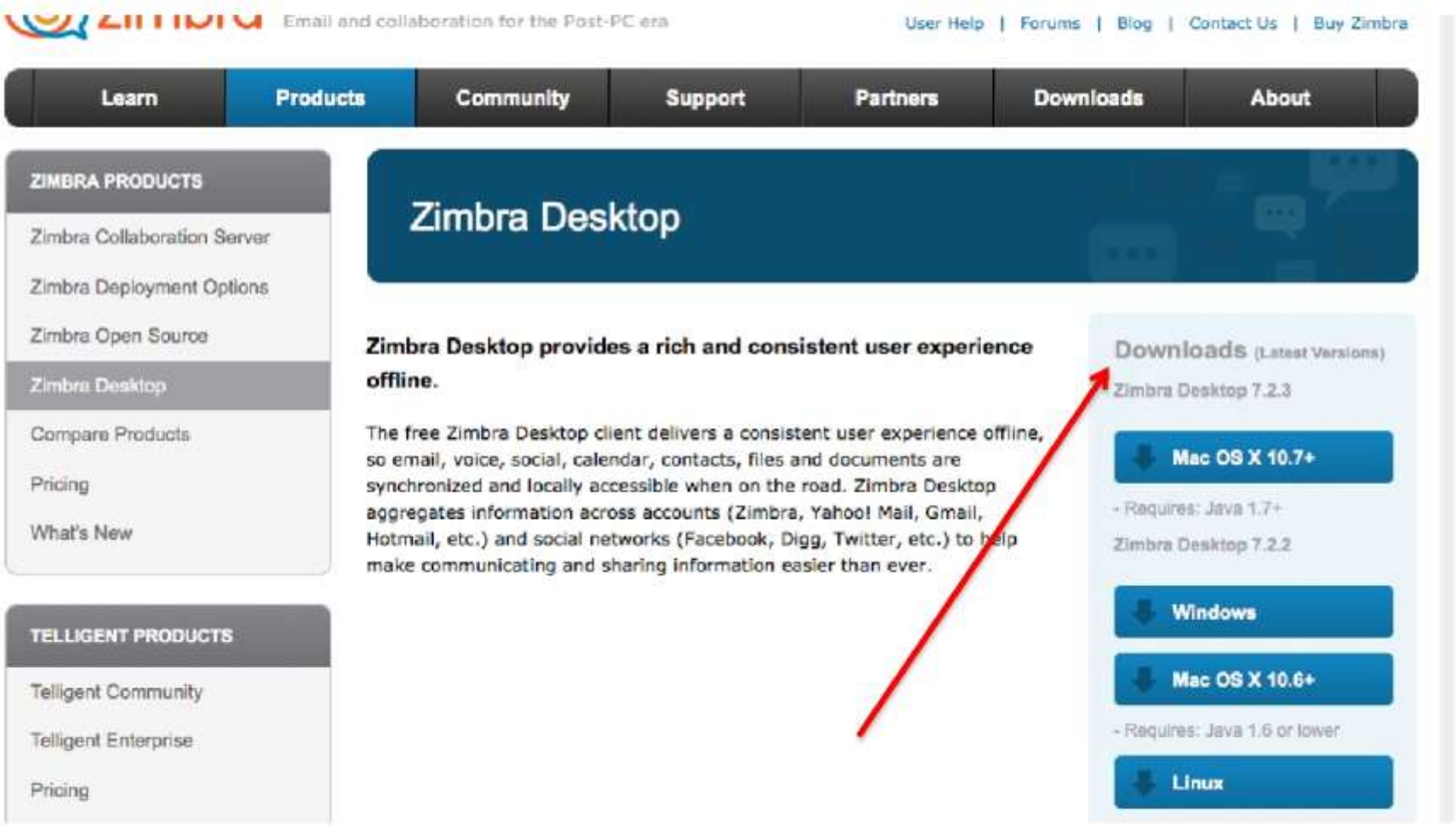
La nuova versione di Zimbra permette di richiedere la ricevuta di lettura delle e-mail spedite. Il destinatario potrà notificare al mittente l'avvenuta lettura del messaggio. Il mittente quindi conoscerà l'istante esatto in

cui il messaggio sarà stato letto.



Dal menu' opzioni e' possibile richiedere conferma di lettura delle mail

Zimbra: zimbra desktop



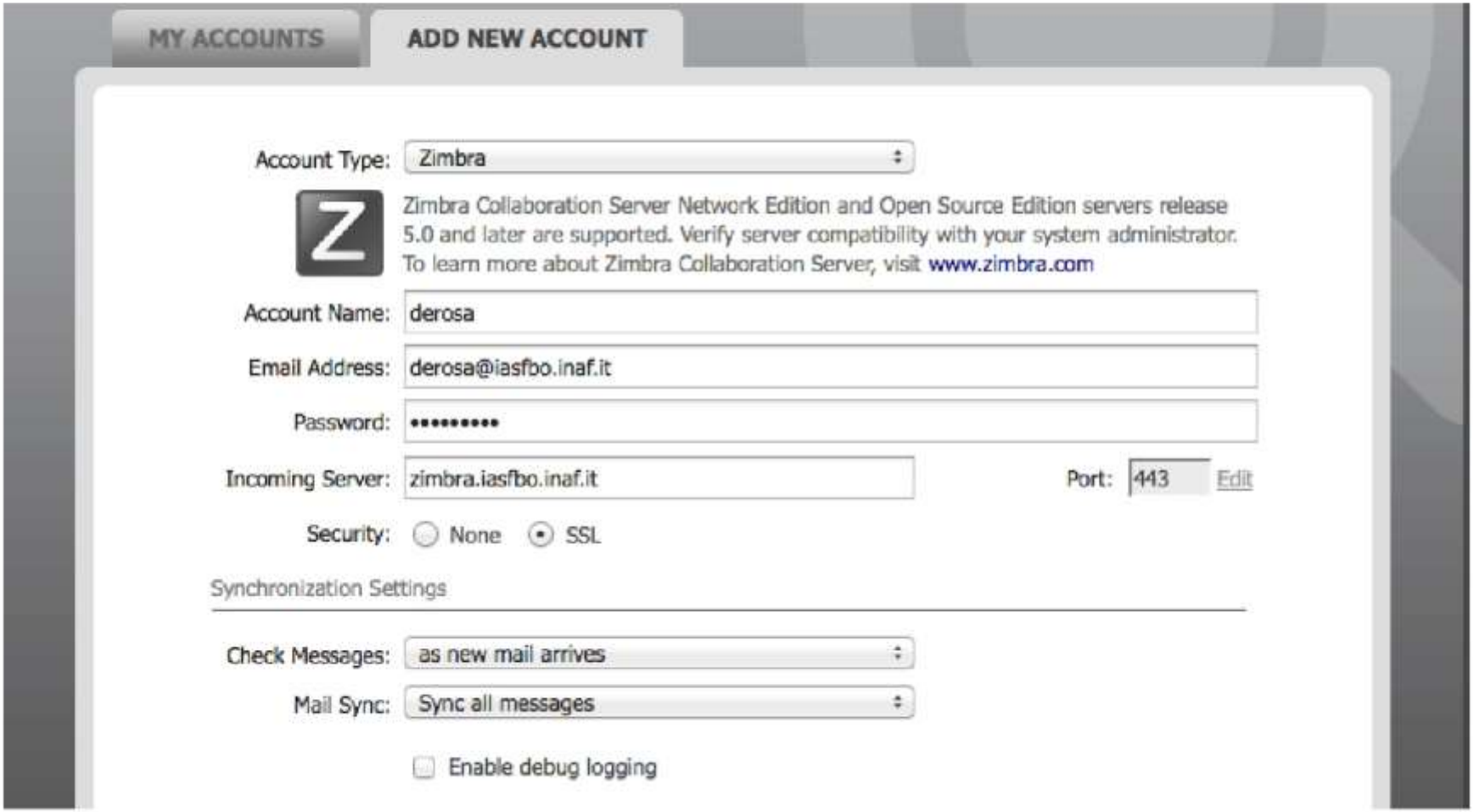
Zimbra desktop: installazione

Zimbra desktop: installazione

Aprire il menu a tendina

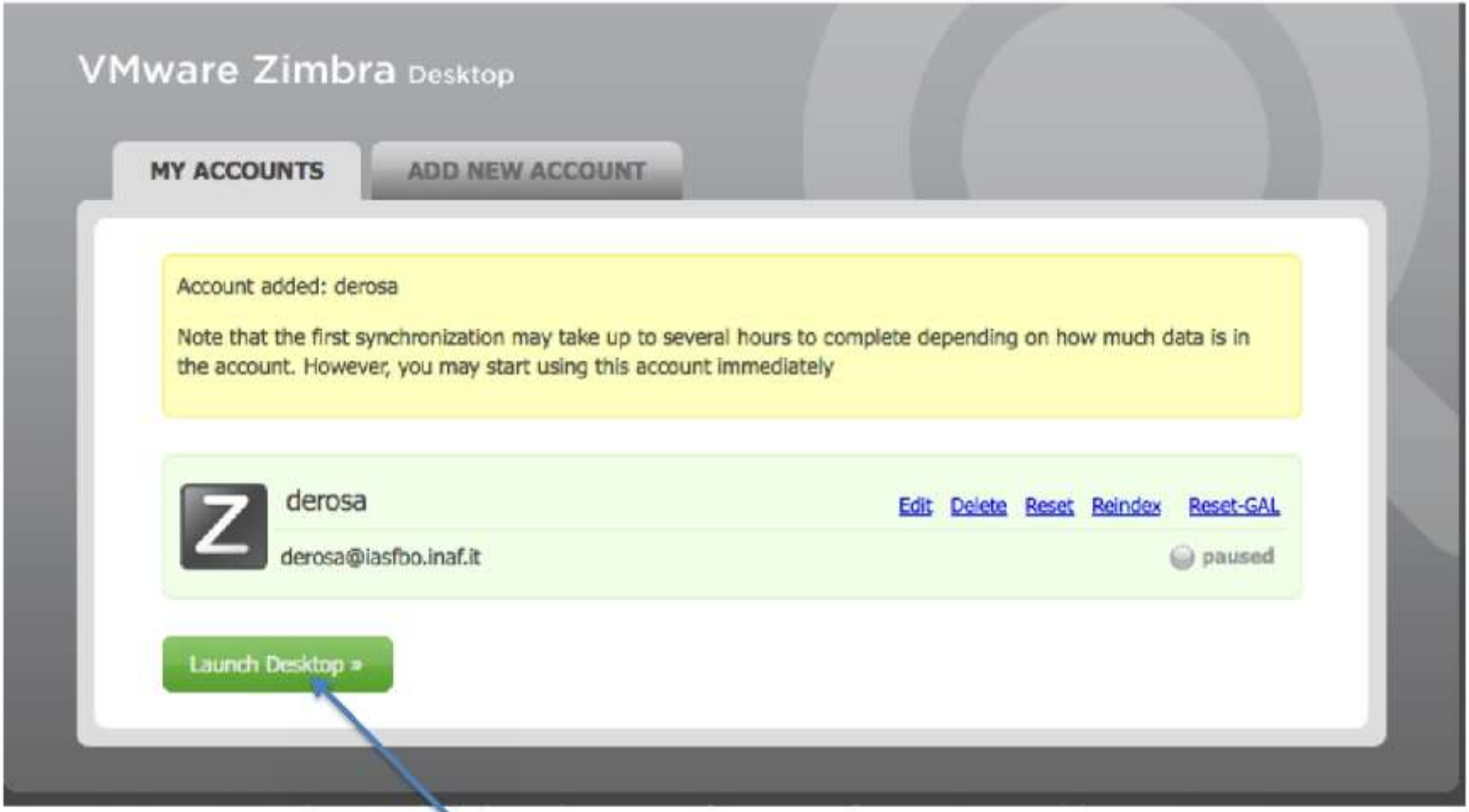
Selezionare ZIMBRA

Zimbra desktop: installazione



Riempire i campi richiesti (username, email address, password)

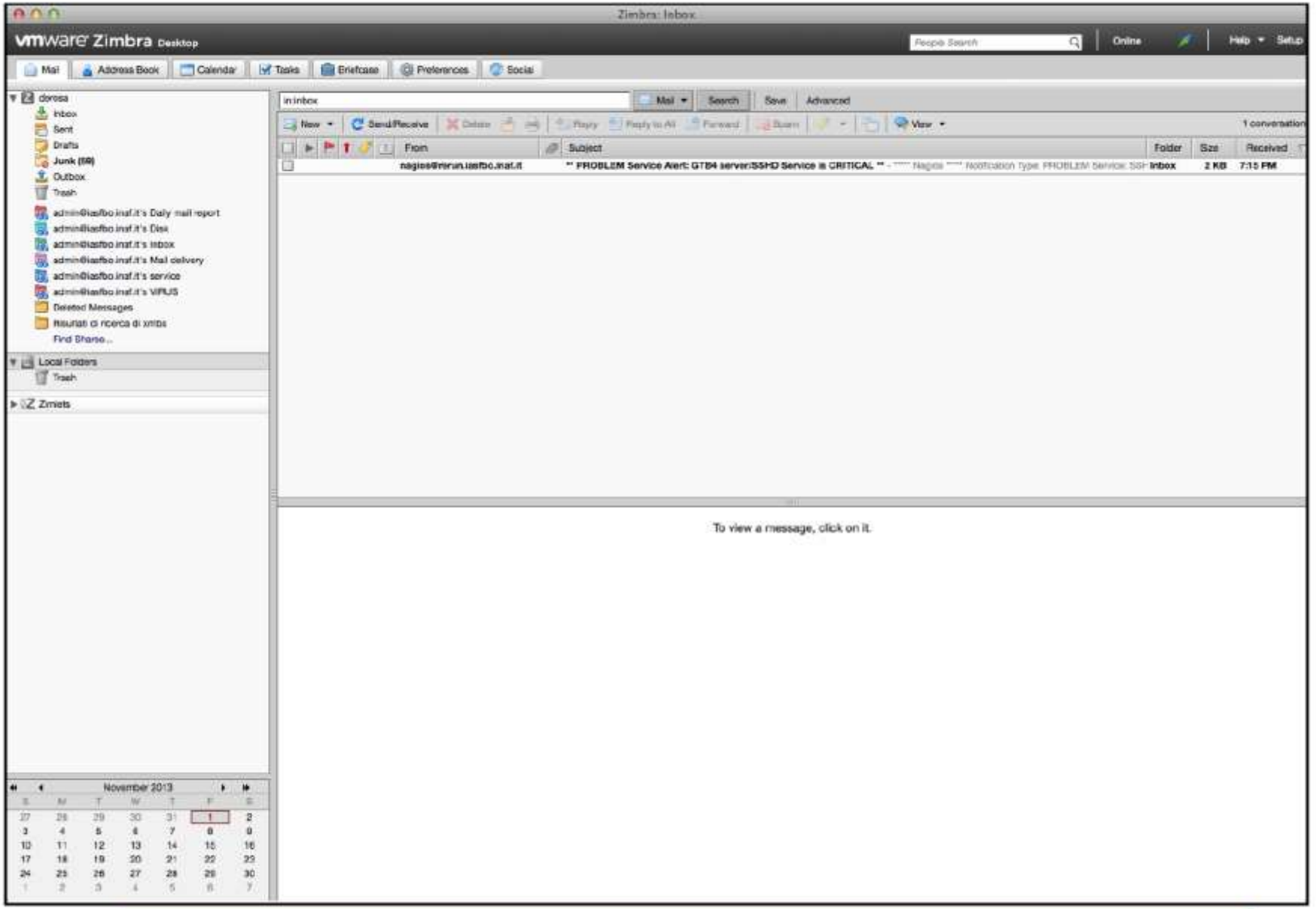
Zimbra desktop: installazione



Potete lanciare Zimbra Desktop

Se avete impostato bene… al posto di derosa troverete il vostro username

Zimbra desktop: aspetto



Zimbra desktop ricalca nell'aspetto e nelle funzioni la pagina web di zimbra

Zimbra: POP, IMAP, SMTP

- POP

pop.iasfbo.inaf.it
porta 995
SSL - abilitato
autenticazione password (a volte può esser richiesto da certi client)

- IMAP

imap.iasfbo.inaf.it
porta 993
SSL - abilitato
autenticazione password (a volte può esser richiesto da certi client)

- SMTP

smtp.iasfbo.inaf.it
porta 25
SSL - abilitato
autenticazione password (a volte può esser richiesto da certi client)

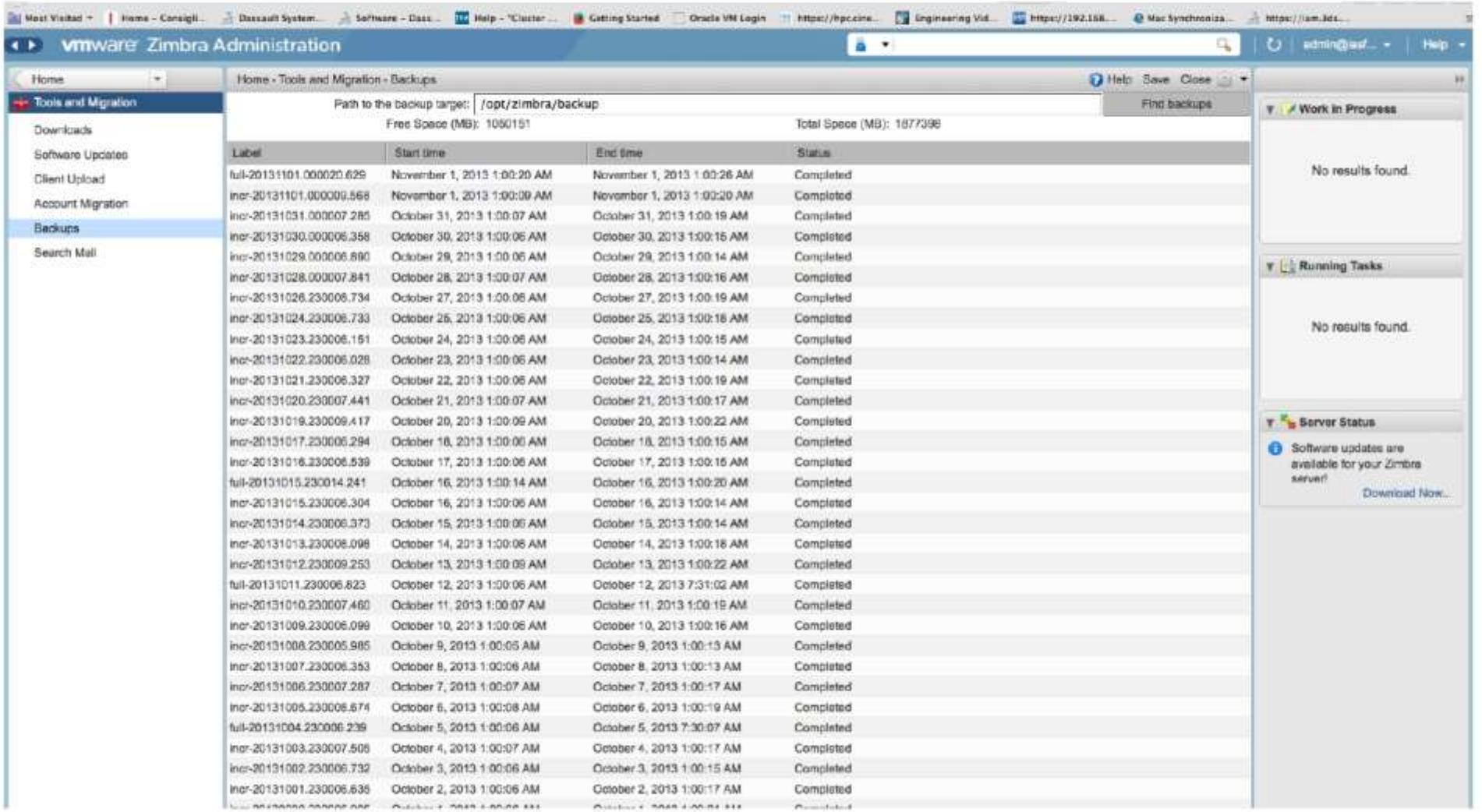
Per la nostra rete interna AL MOMENTO e' possibile anche disabilitare SSL

Zimbra: livelli di backup implementati

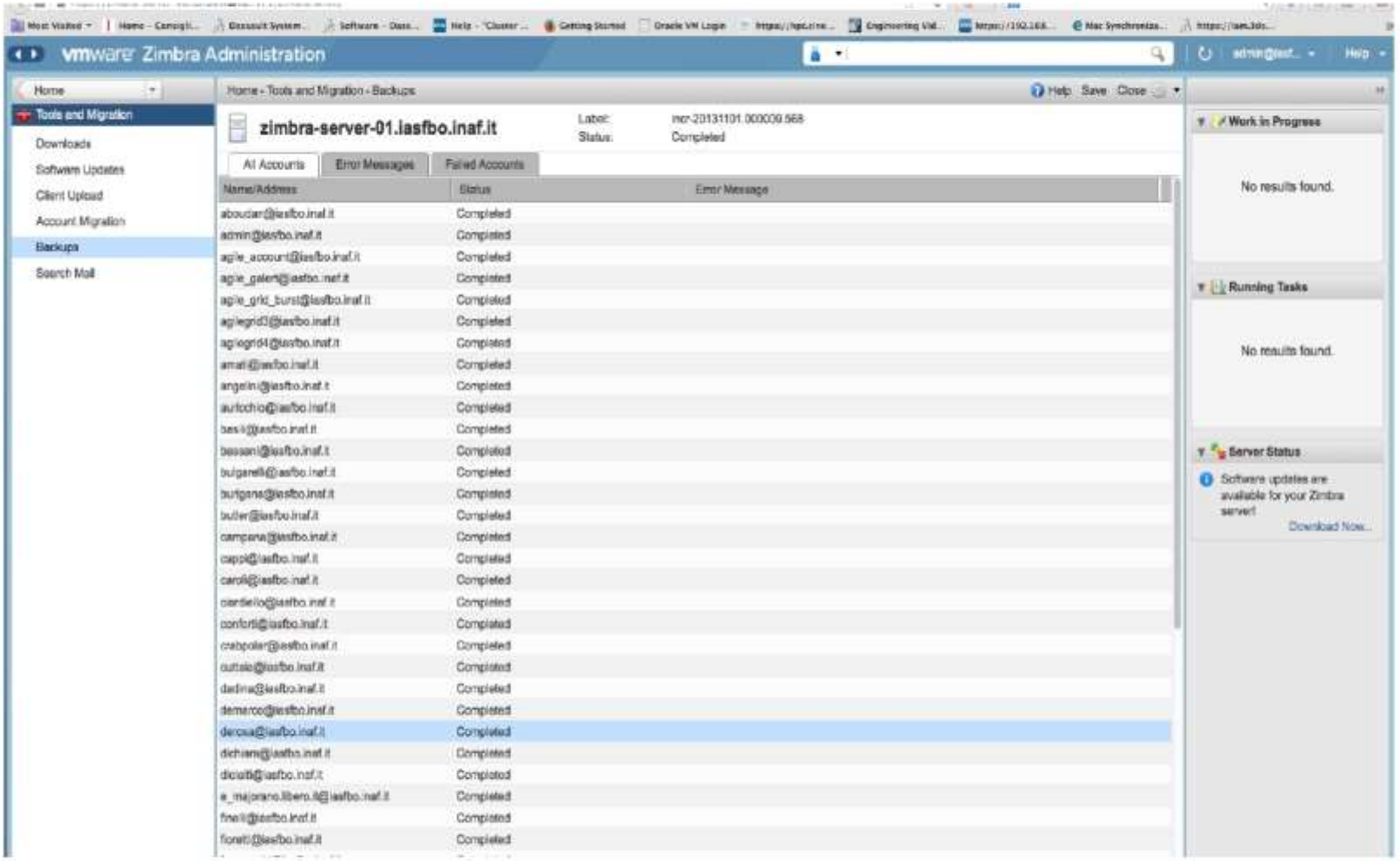
- Zimbra possiede 3 livelli di backup

- Il primo livello e' direttamente gestibile dall'utente. Le mail cancellate ed eliminate dal cestino restano presenti sul sistema e l'utente puo' recuperarle direttamente dall'interfaccia web di zimbra
- Il secondo livello di backup e' gestibile dall'amministratore che puo' ripristinare le caselle dai backup creati da zimbra, che restano costantemente online. I backup sono giornalieri, ovvero ogni notte zimbra effettua un consolidamento di tutte le operazioni effettuate sul database e sui dati. Vengono mantenuti costantemente online solo gli ultimi 30 giorni di backup.
- Il terzo livello di backup consiste in una copia dei backup consolidati di zimbra su un server esterno al sistema.

Zimbra: backup online livello 2

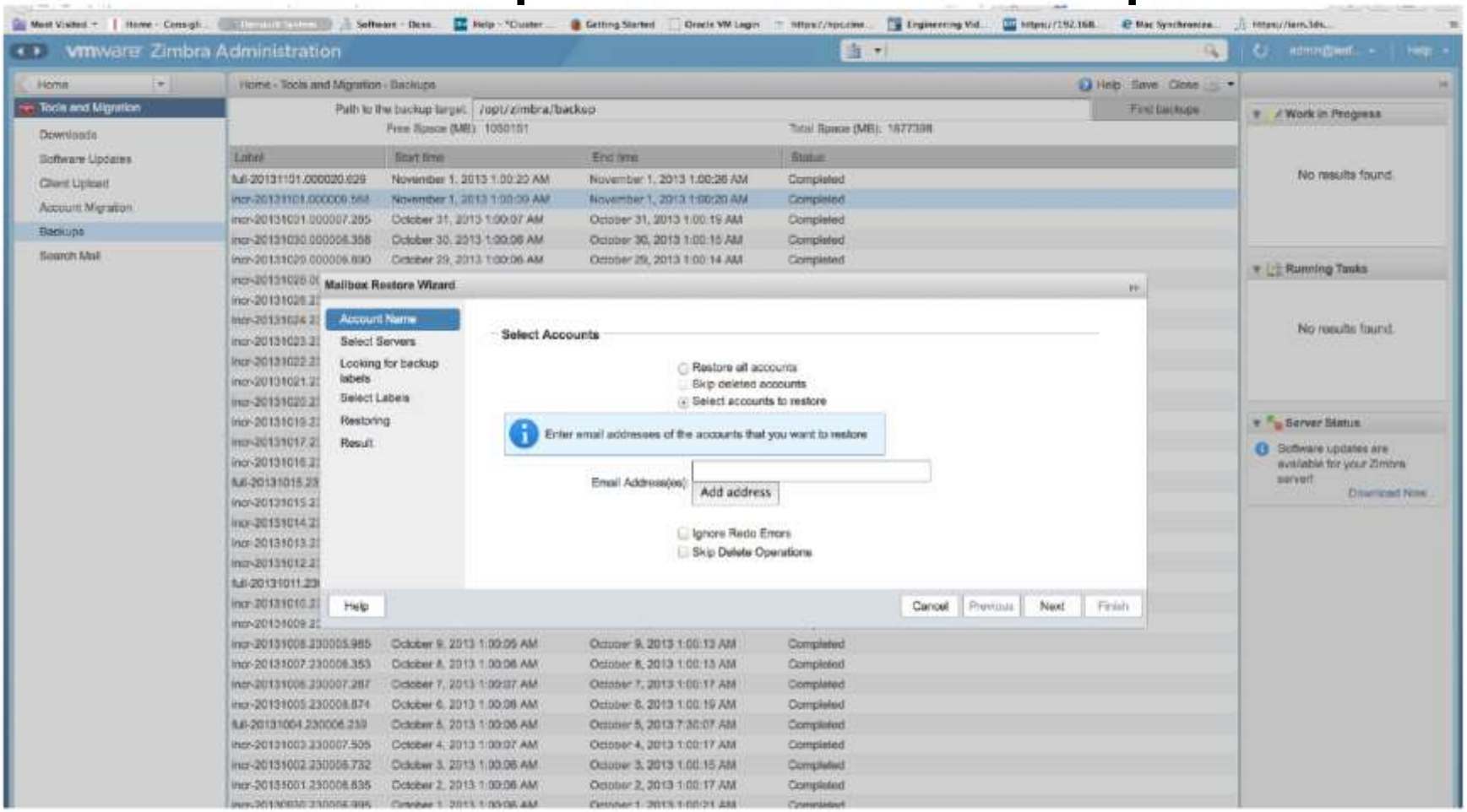


Zimbra: backup online livello 2



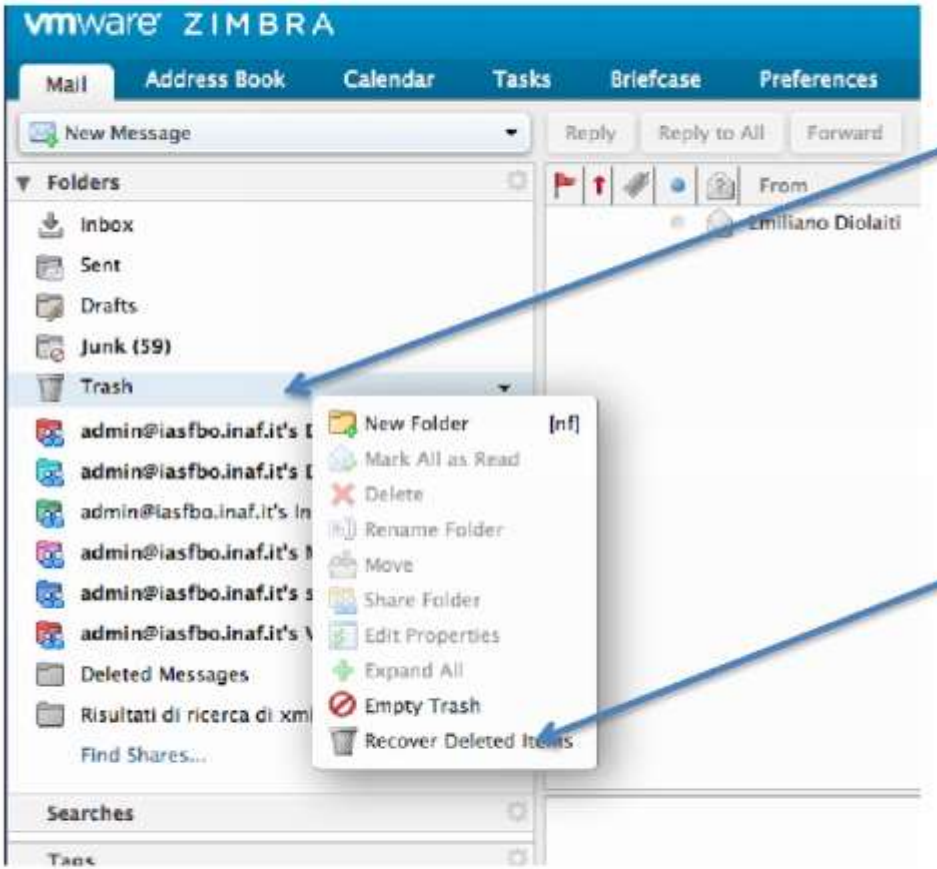
Il backup viene effettuato in modo organico per utente. Il backup contiene sia inbox che tutte le cartelle di posta.

Zimbra: backup online livello 2 - ripristino



Il ripristino puo' esser fatto per tutto o alcuni utenti, selezionando il punto di ripristino, scegliendo se creare account temporanei, se effettuare un merging delle caselle recuperate con quelle esistenti.

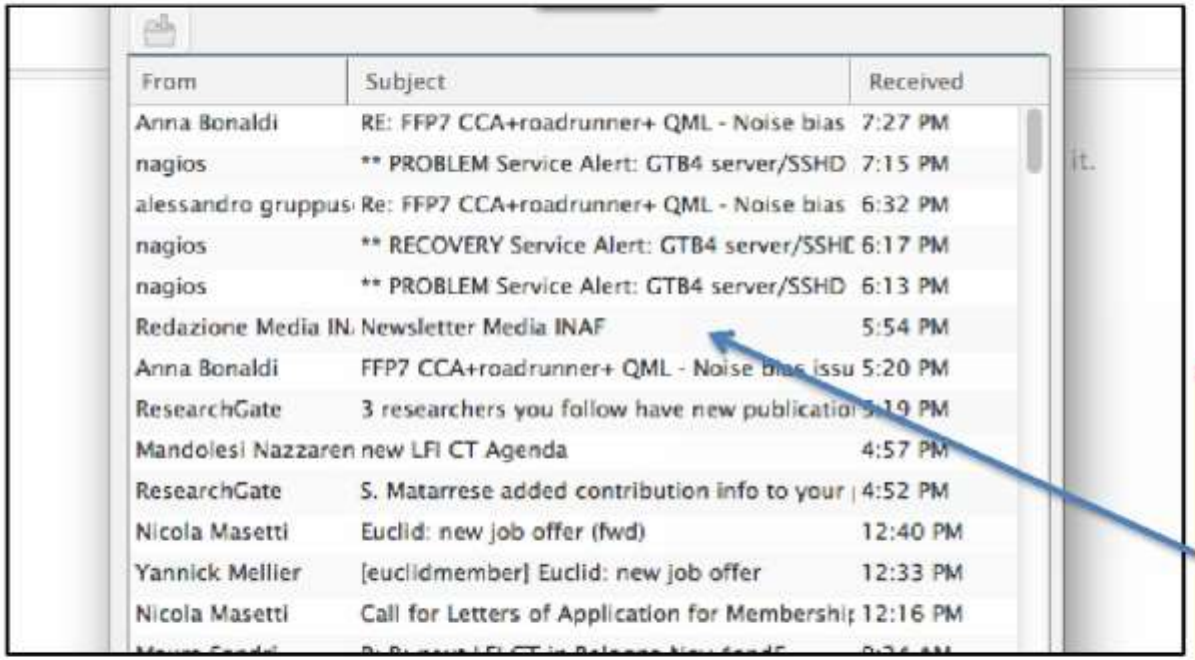
Zimbra: backup online livello 1



1- cliccando col tasto destro si apre il menu' a tendina

2 - cliccare (tasto sinistro) su recover Deleted Items

Zimbra: backup online livello 1

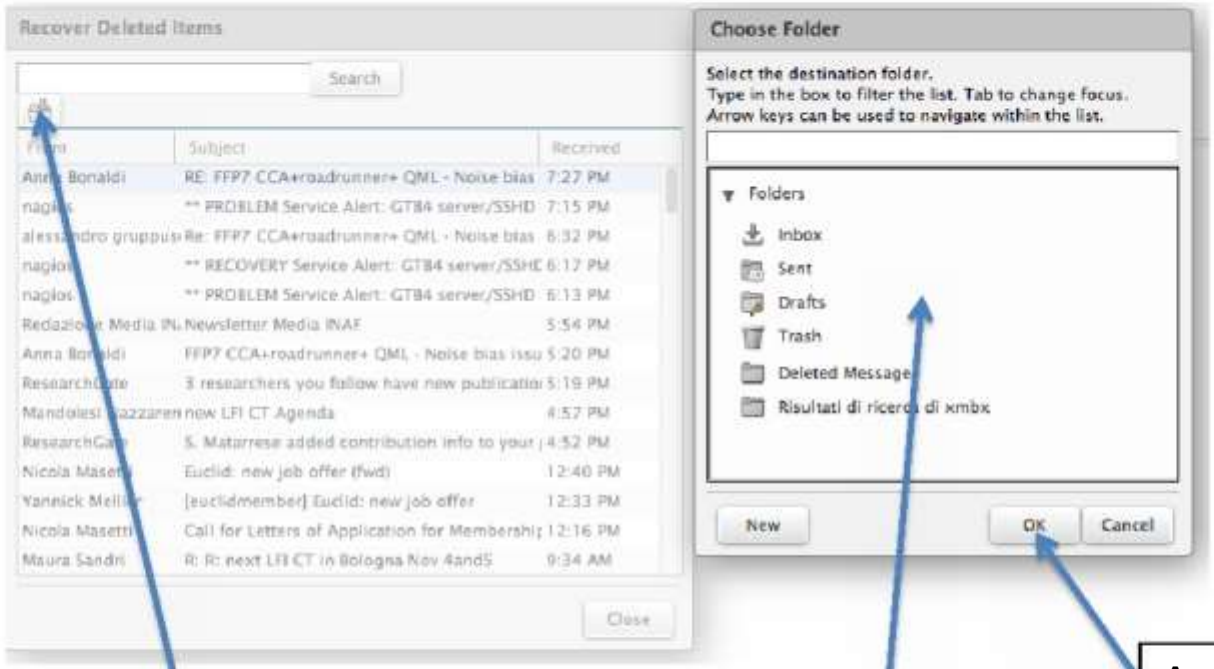


From	Subject	Received
Anna Bonaldi	RE: FFP7 CCA+roadrunner+ QML - Noise bias	7:27 PM
nagios	** PROBLEM Service Alert: GTB4 server/SSHD	7:15 PM
alessandro gruppus	Re: FFP7 CCA+roadrunner+ QML - Noise bias	6:32 PM
nagios	** RECOVERY Service Alert: GTB4 server/SSHD	6:17 PM
nagios	** PROBLEM Service Alert: GTB4 server/SSHD	6:13 PM
Redazione Media IN	Newsletter Media INAF	5:54 PM
Anna Bonaldi	FFP7 CCA+roadrunner+ QML - Noise bias issu	5:20 PM
ResearchGate	3 researchers you follow have new publicatio	5:19 PM
Mandolesi Nazzaren	new LFI CT Agenda	4:57 PM
ResearchGate	S. Matarrese added contribution info to your	4:52 PM
Nicola Masetti	Euclid: new job offer (fwd)	12:40 PM
Yannick Mellier	[euclidmember] Euclid: new job offer	12:33 PM
Nicola Masetti	Call for Letters of Application for Membershi	12:16 PM

Mail cancellate

DA NOTARE come i client impostati come POP (vedete il mio) spostino nel cestino tutte le mail ricevute in inbox

Zimbra: backup online livello 1

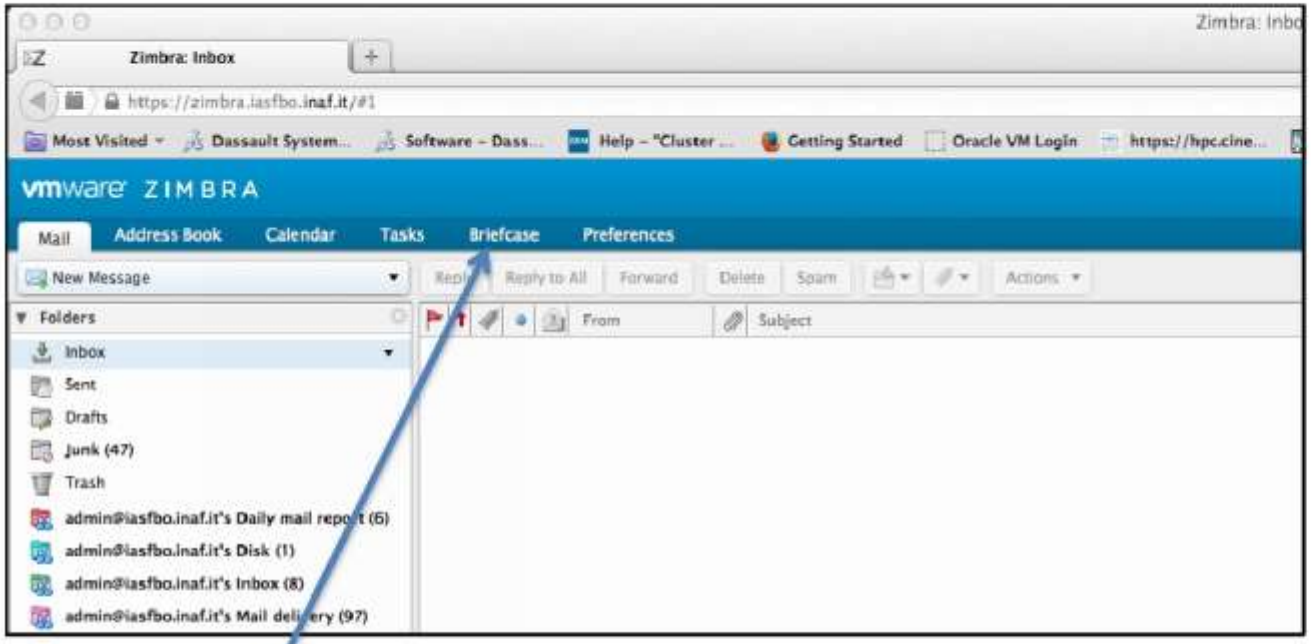


1 -Selezionare le mail poi cliccare sull'icona folder di destinazione

2-Selezionare il folder

Avviare il ripristino

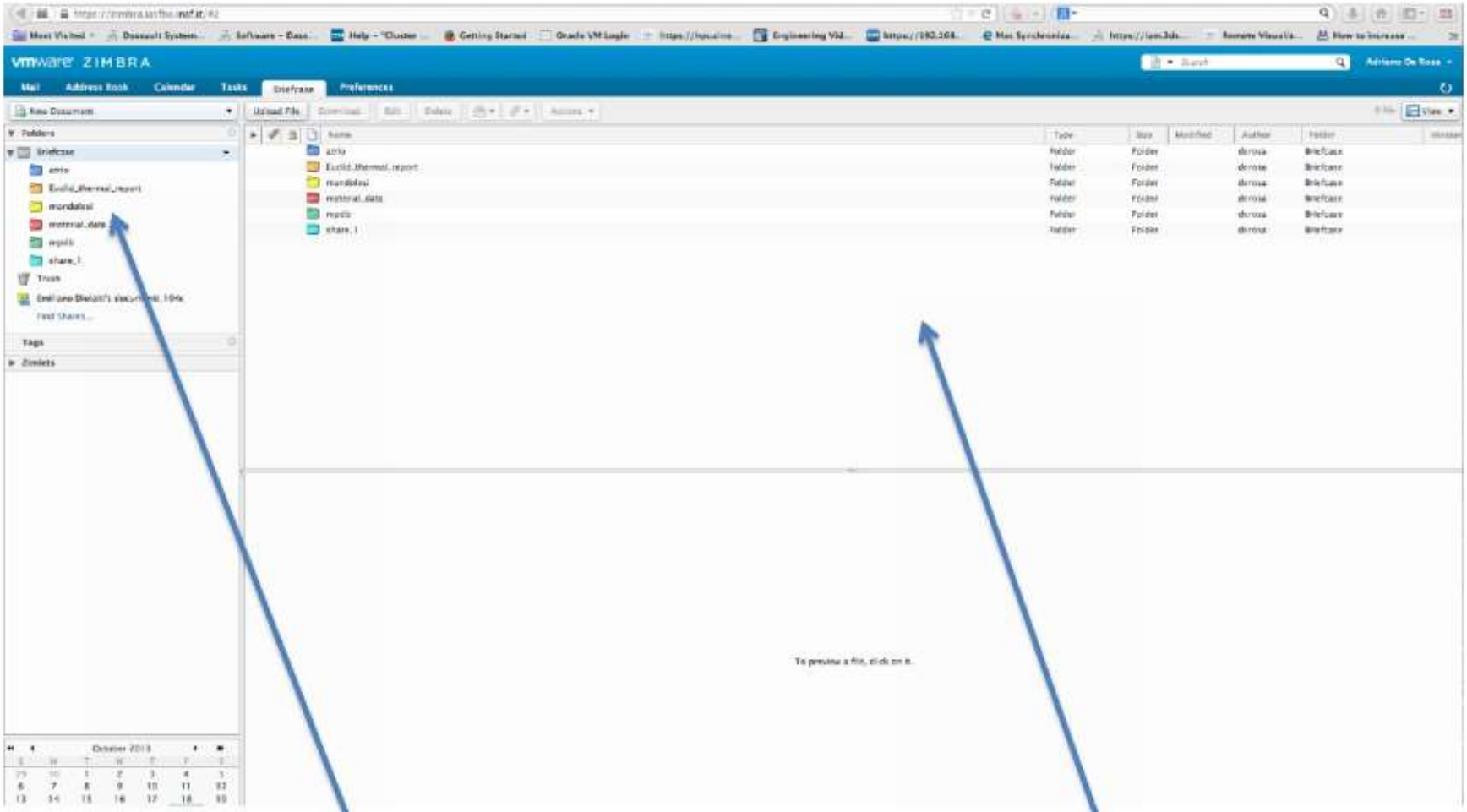
Zimbra: valigetta



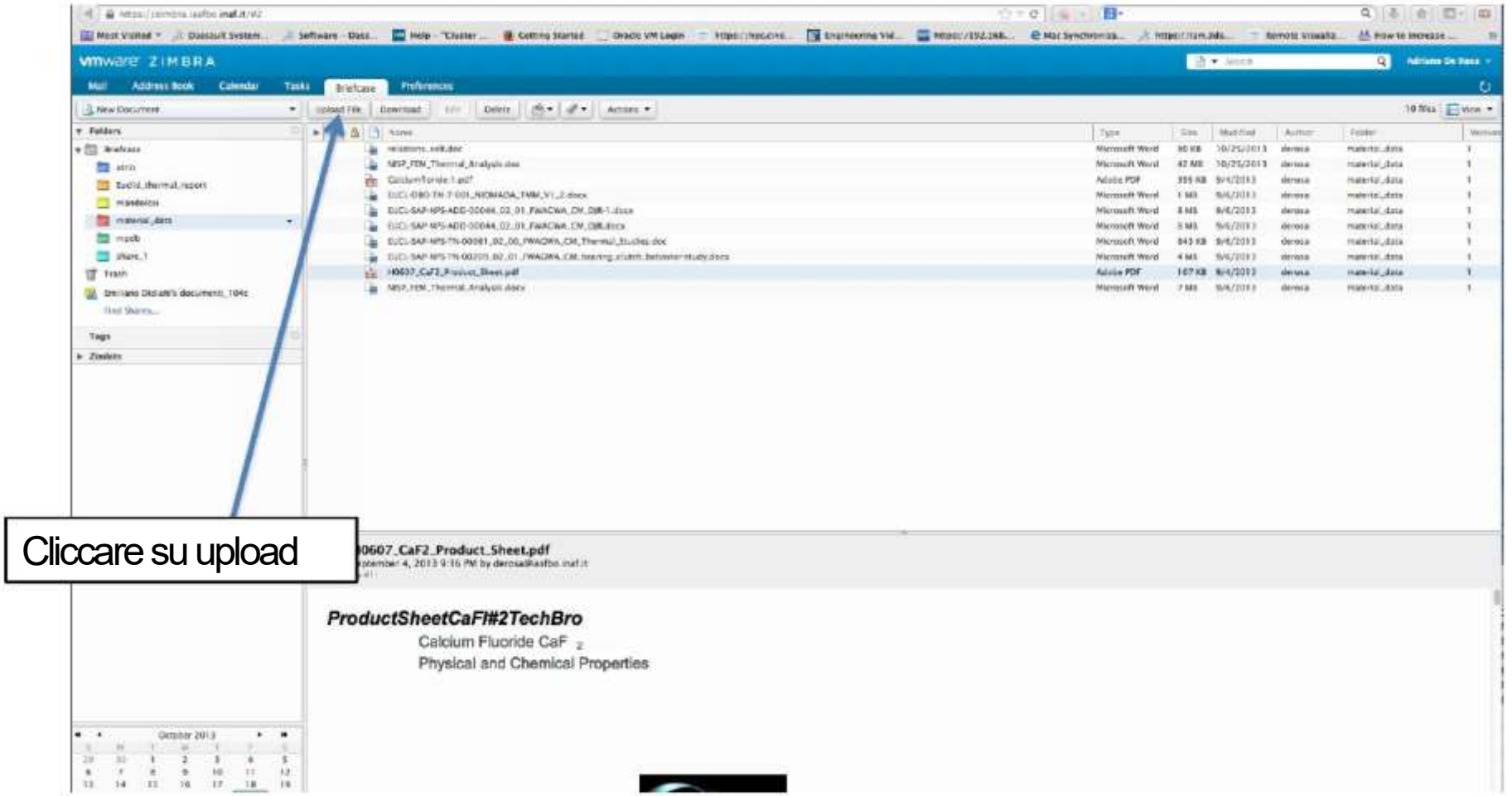
Cliccare su Briefcase o valigetta (in base alla lingua impostata)

Accessibile solo via web o zimbra desktop

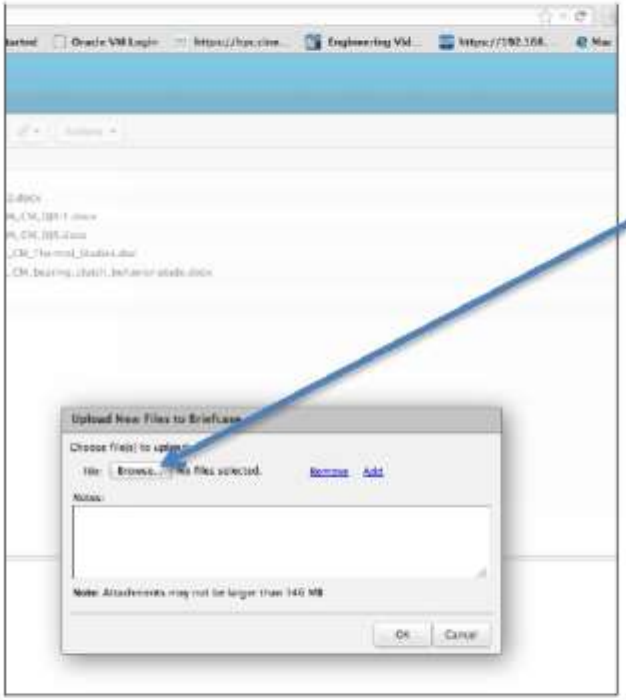
Zimbra: valigetta



Zimbra: valigetta - caricare un file

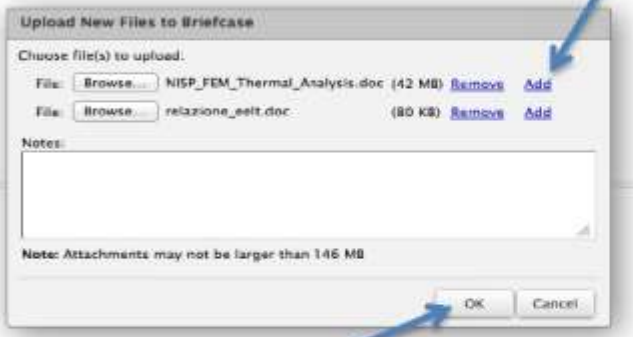


Zimbra: valigetta - caricare un file



Selezionare il e/o i files da caricare (windows internet explorer permette la selezione di un solo file alla volta)

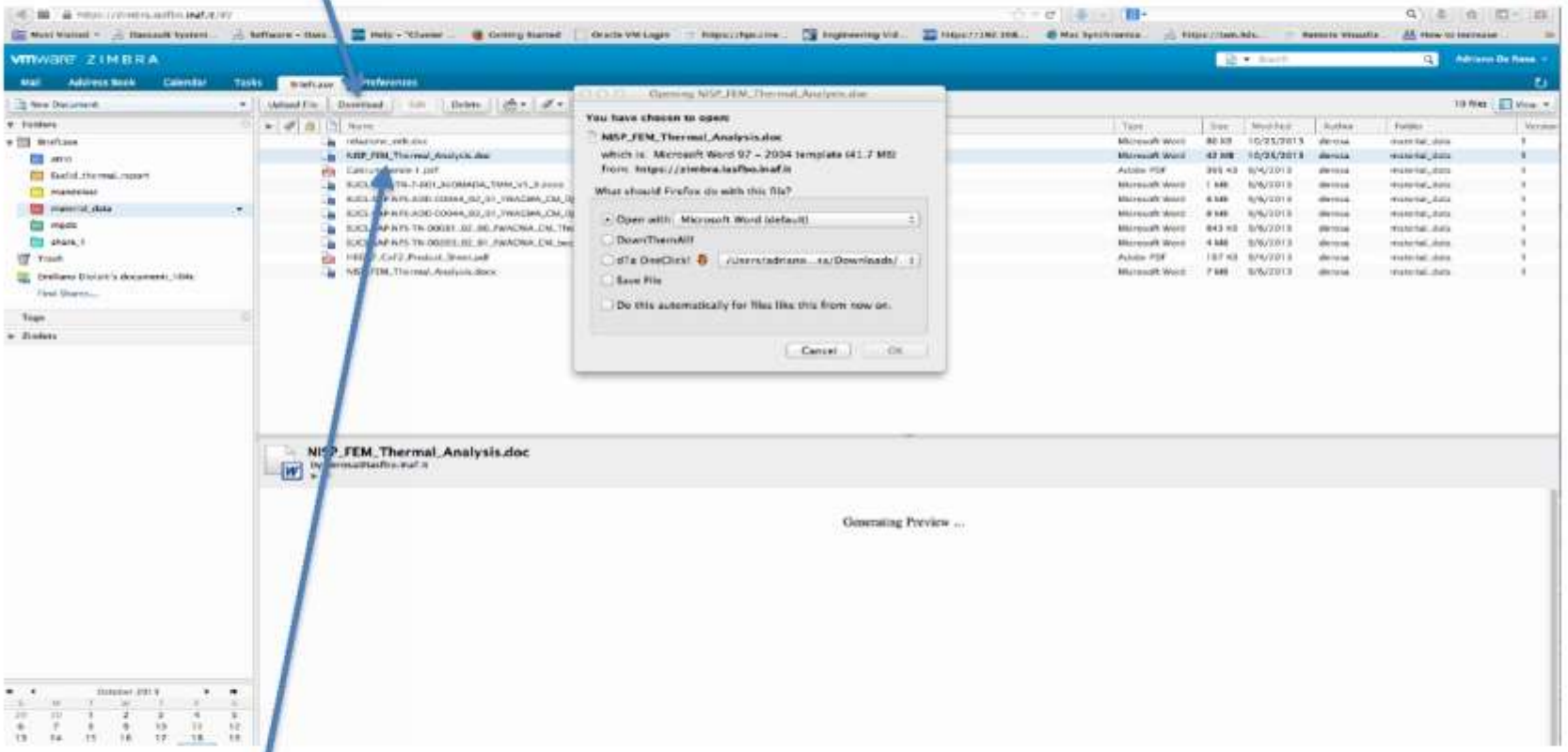
Selezionare piu' files



Iniziare l'upload

Zimbra: valigetta - download di uno o piu' files

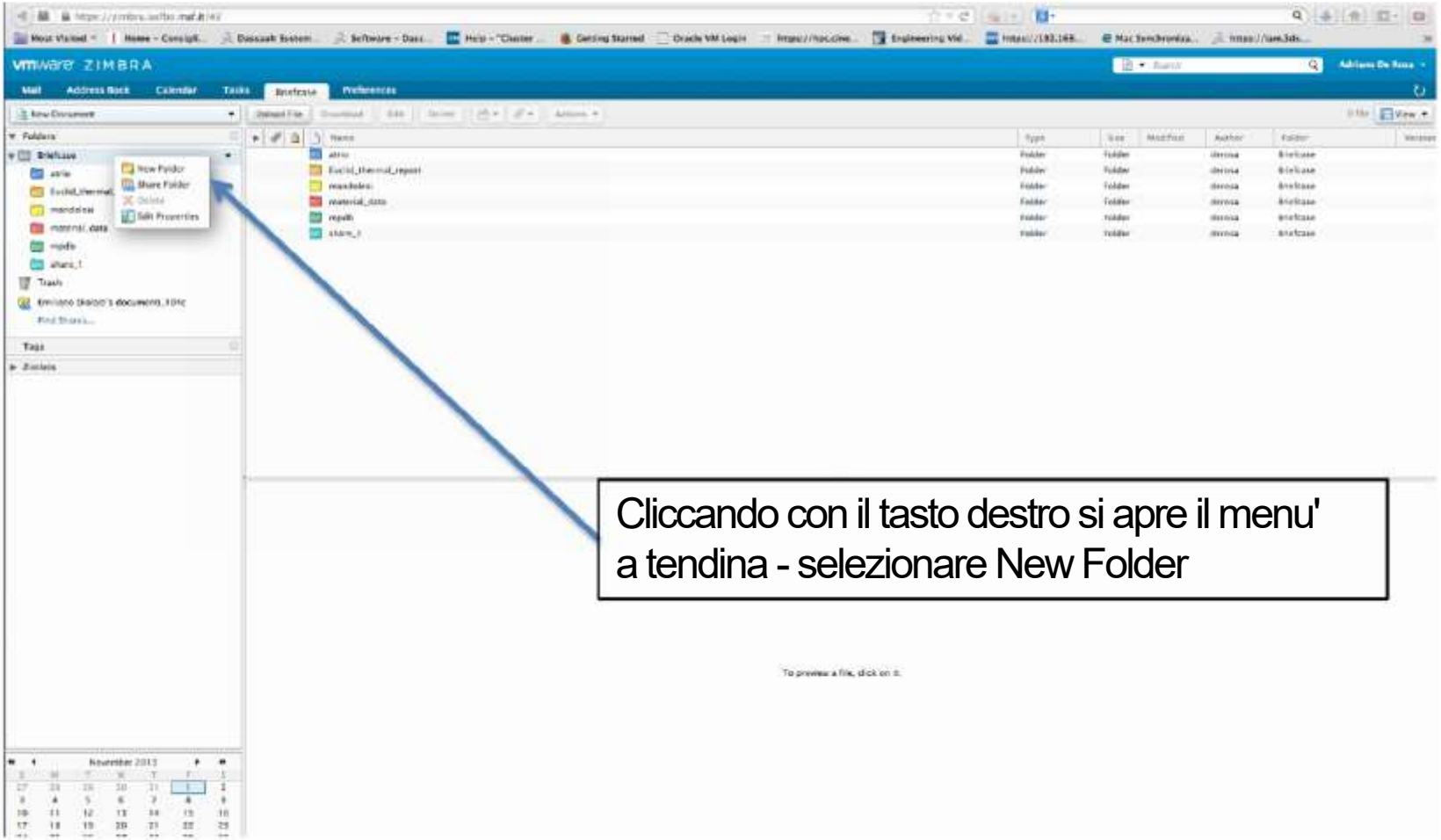
Cliccare su download



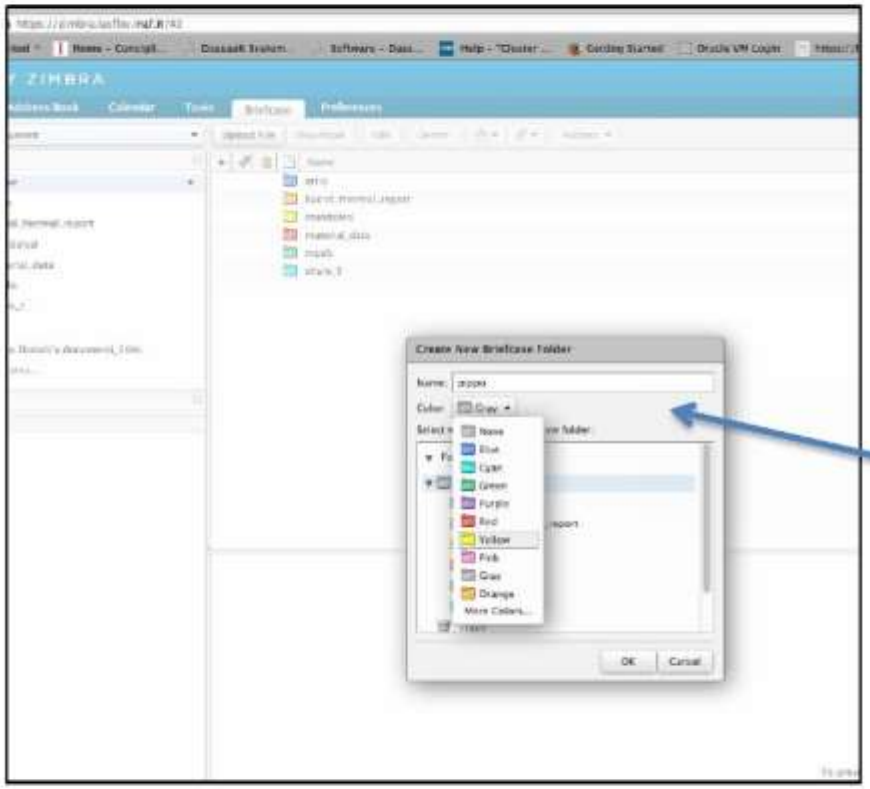
Selezionare il file o i files da scaricare localmente

Nel caso in cui i files da scaricare siano piu' di uno, zimbra scarichera' un file compresso in formato zip

Zimbra: valigetta - nuova cartella

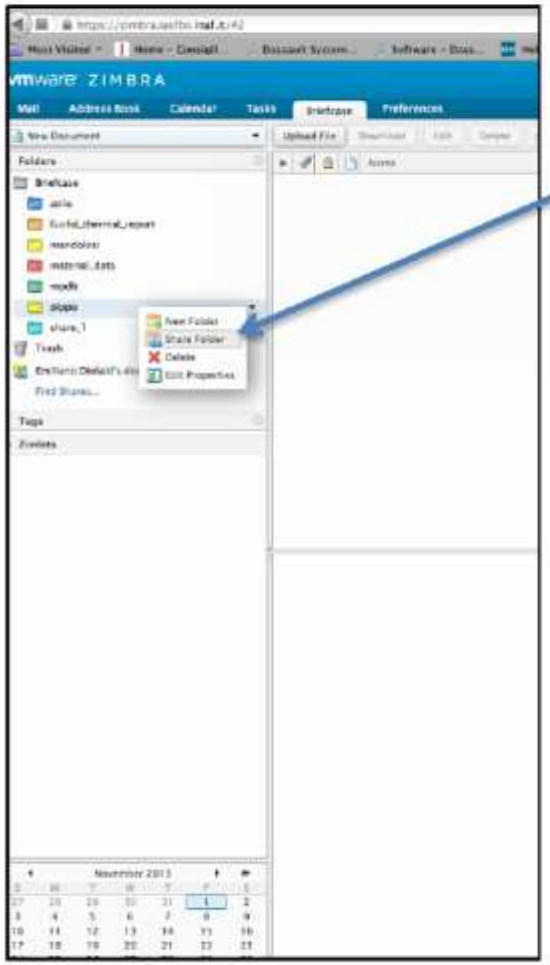


Zimbra: valigetta - nuova cartella



Mettere il nome della nuova cartella, selezionare il colore, posizionare nell'albero la nuova cartella

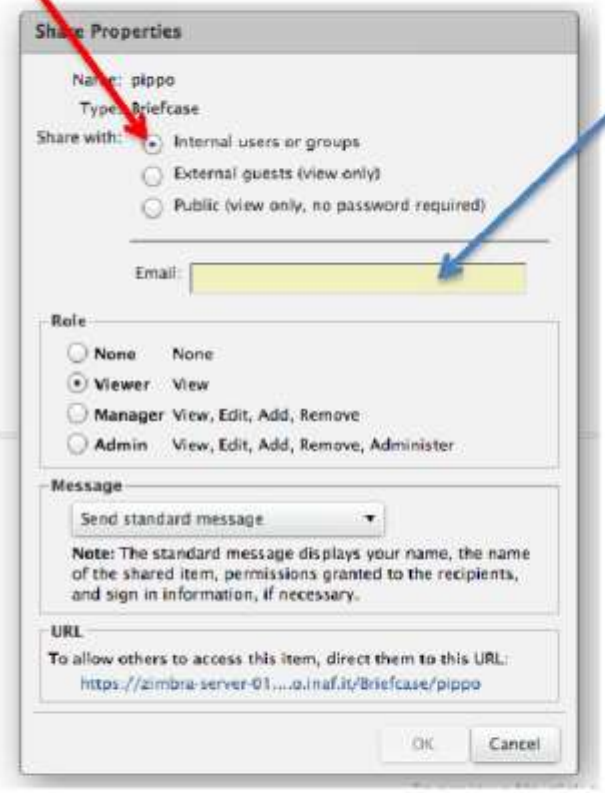
Zimbra: valigetta - condivisioni



Cliccando con il tasto destro, si apre il menu' a tendina, selezionare Share Folder

Le condivisioni di cartelle di posta fra account interni all'istituto (al server) hanno la medesima procedura. Utile per account di servizio, dove e' possibile eliminare il forwarding di mail (quindi una loro copia sul server), con la condivisione delle cartelle stesse.

Zimbra: valigetta - condivisioni

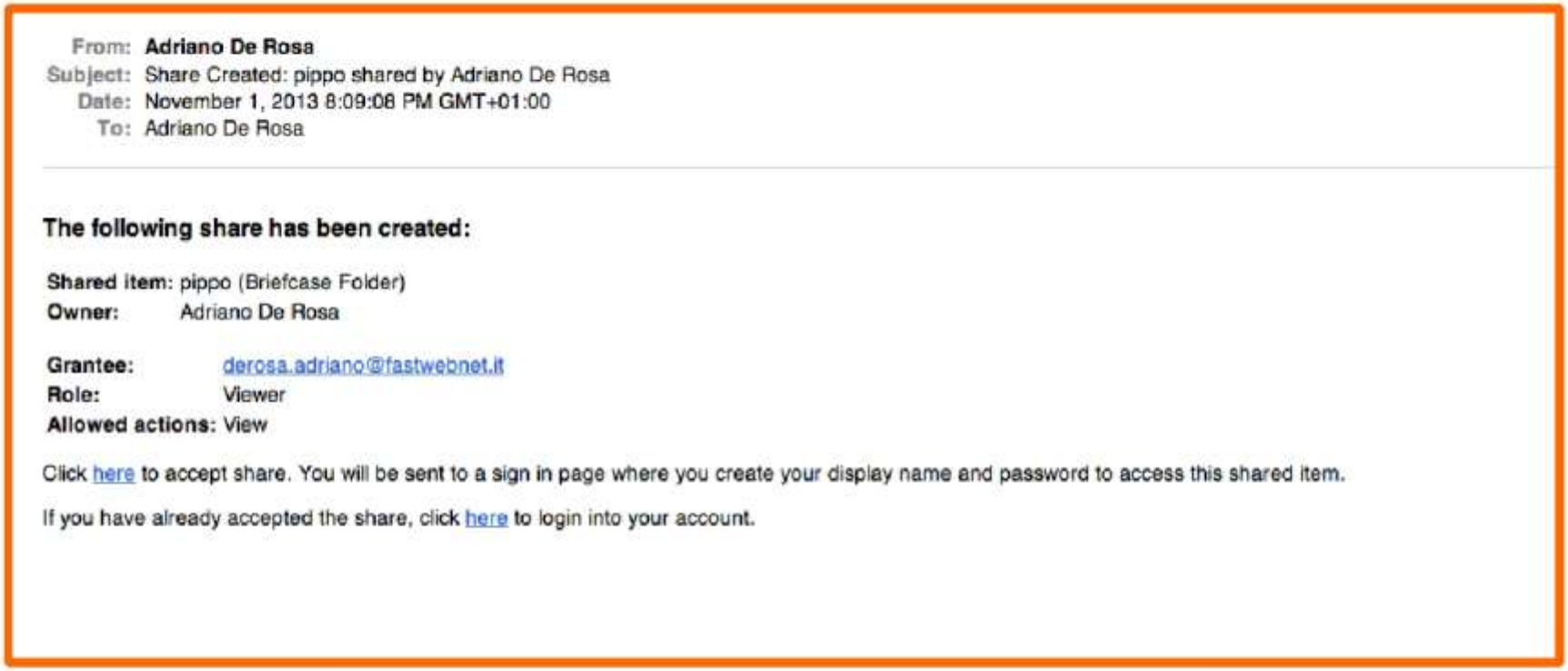


Lista degli indirizzi mail degli utenti con cui condividere la cartella

Condivisione con utenti interni.
La differenza fra Manager e Admin e' nella possibilita' di Admin di condividere a sua volta la cartella con altri

La condivisione con utenti esterni puo' avvenire solo in modalita' di lettura, con registrazione sul server

Zimbra: valigetta - condivisioni



From: **Adriano De Rosa**
Subject: Share Created: pippo shared by Adriano De Rosa
Date: November 1, 2013 8:09:08 PM GMT+01:00
To: Adriano De Rosa

The following share has been created:

Shared item: pippo (Briefcase Folder)
Owner: Adriano De Rosa

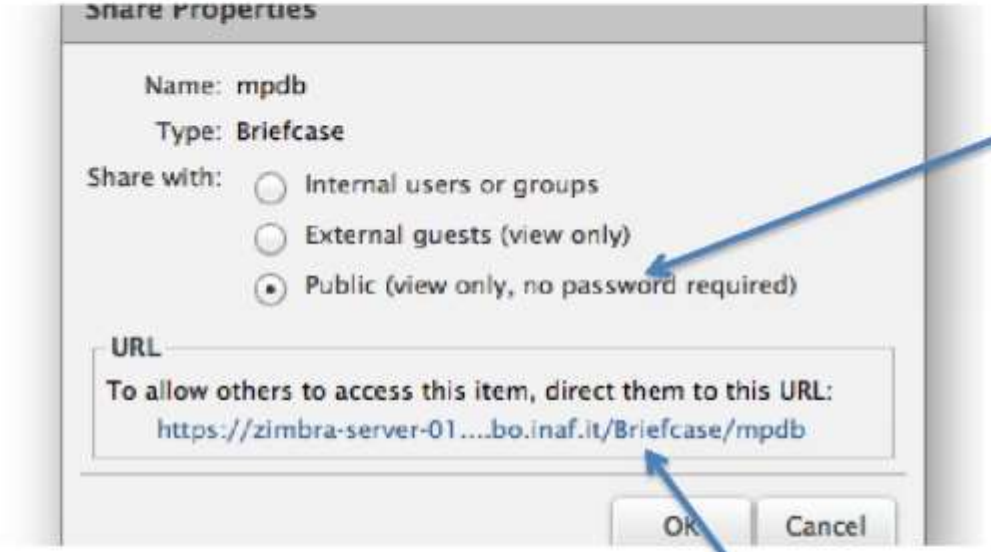
Grantee: derosa.adriano@fastwebnet.it
Role: Viewer
Allowed actions: View

Click here to accept share. You will be sent to a sign in page where you create your display name and password to access this shared item.

If you have already accepted the share, click here to login into your account.

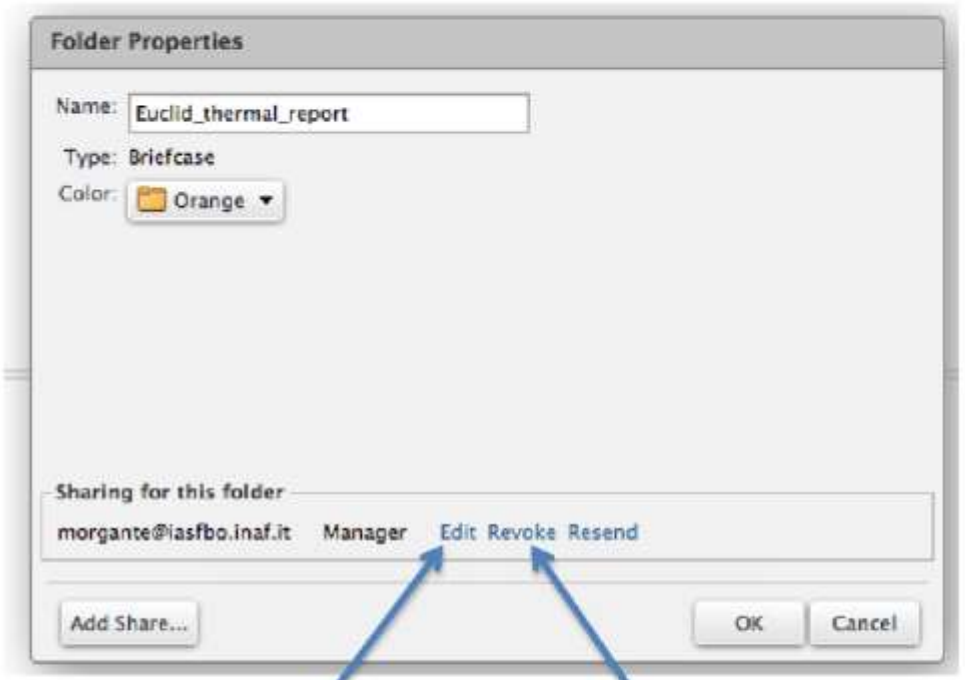
Mail che viene ricevuta dall'utente esterno al sistema, con i link necessari per registrarsi e accedere allo sharing

Zimbra: valigetta - condivisioni



Indirizzo web dal quale si possono scaricare i files che sono nella cartella, senza password. Equivale a quello che si puo' fare mettendo i files nella propria cartella HTML ed inviando i link ai destinatari

Zimbra: valigetta - eliminare le condivisioni



Selezionare Edit Properties

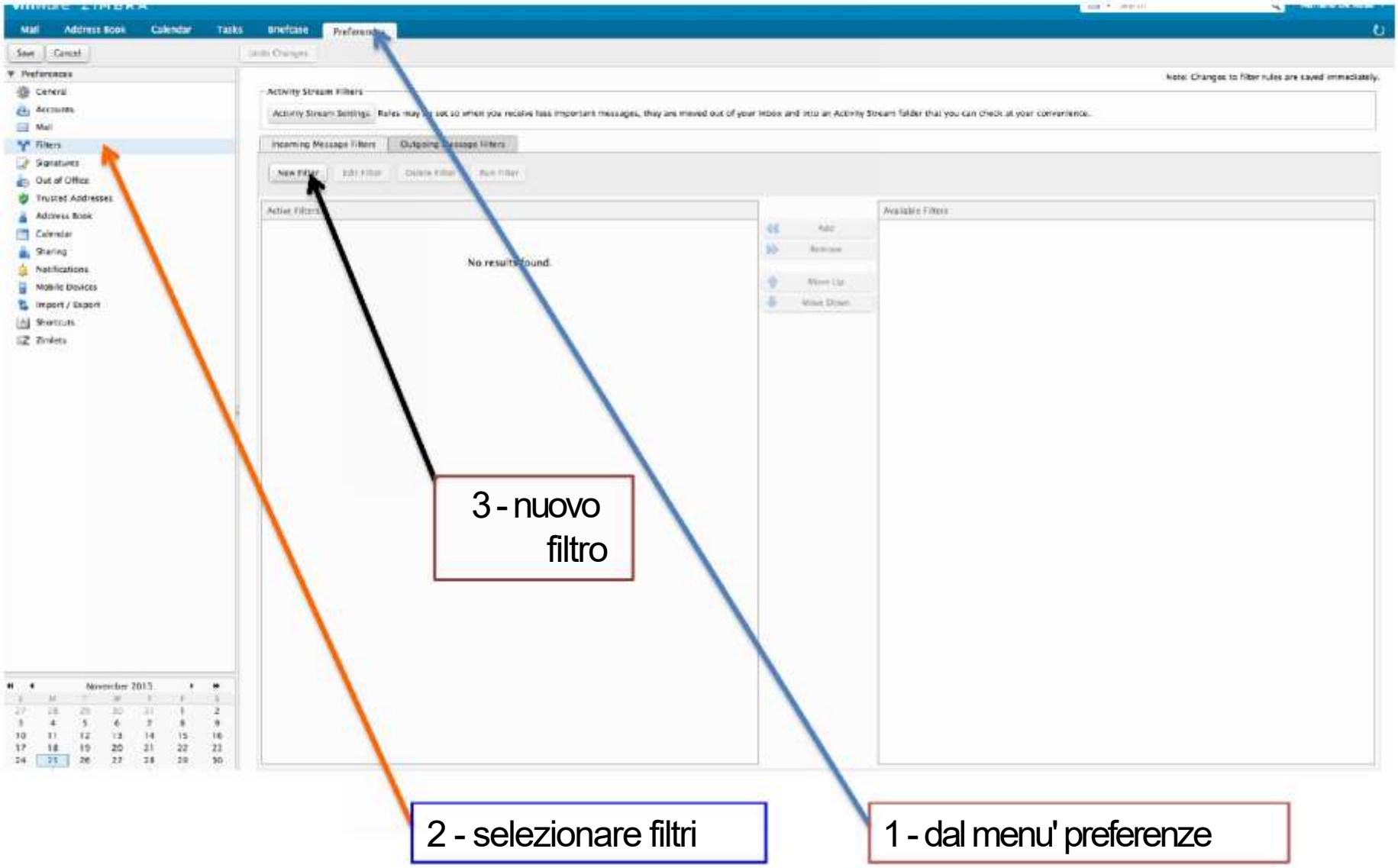
Per cambiare i diritti (Manager, Admin…)

Per eliminare la condivisione

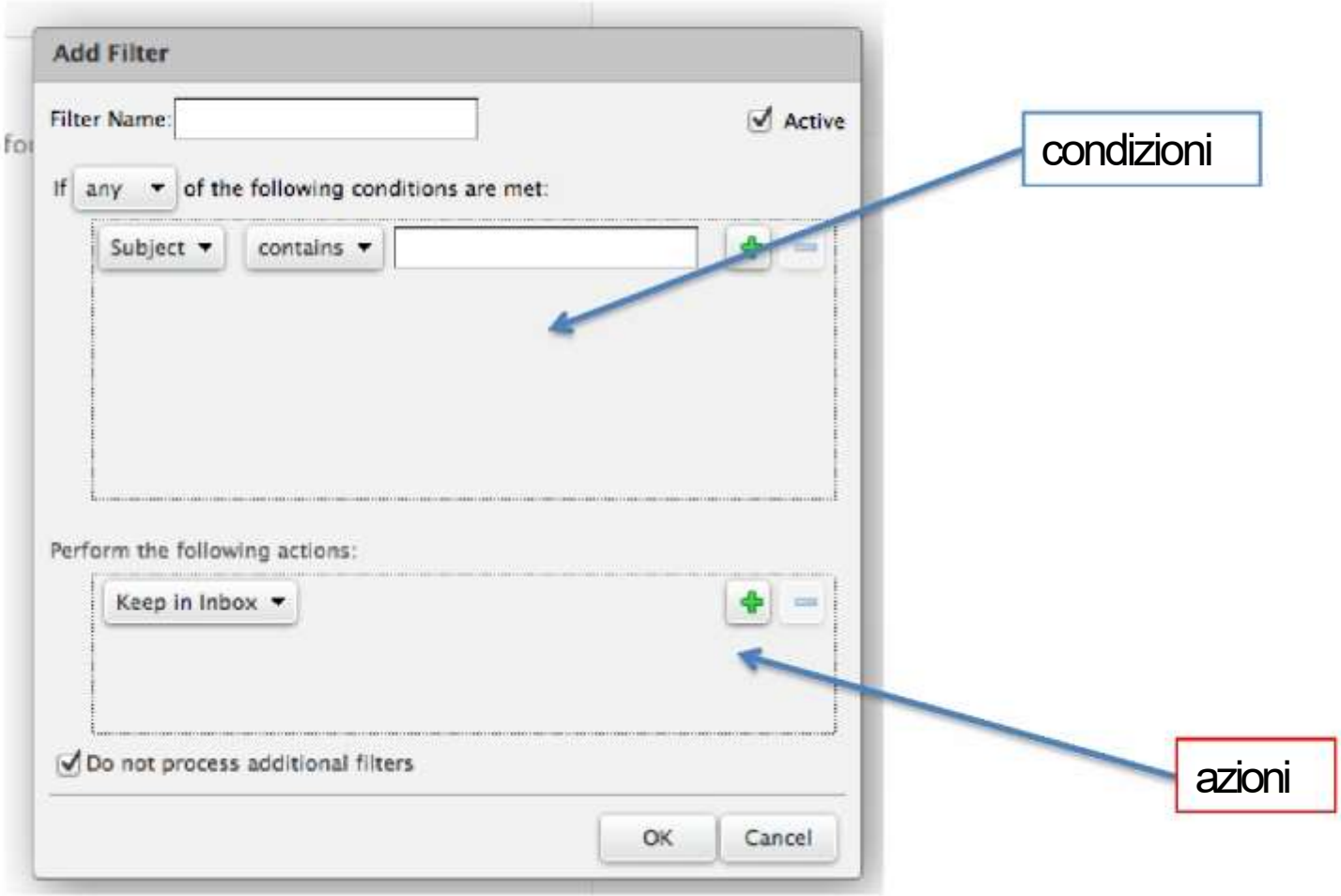
Zimbra: jumbo mail

- La condivisione dei files con utenti esterni offre un servizio del tutto analogo ai jumbo mail offerti da molti providers.
- La condivisione per gli utenti esterni e' solo in download
- L'upload e' possibile solo agli utenti interni dando l privilegi di manager nella condivisione

Zimbra: utilizzo dei filtri di posta



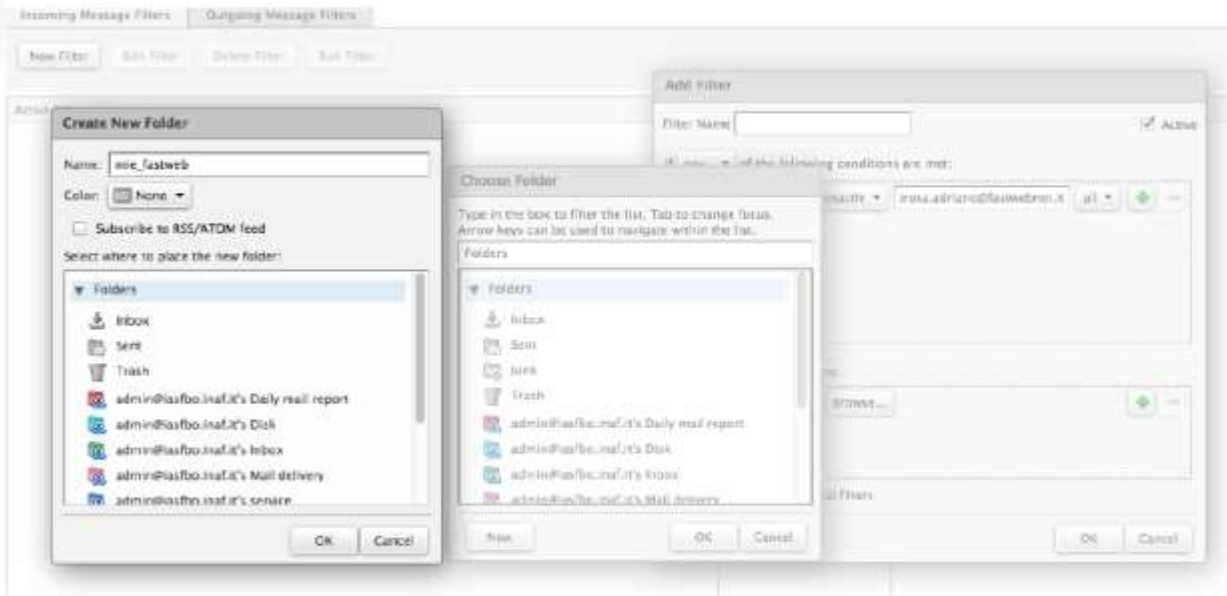
Zimbra: utilizzo dei filtri di posta



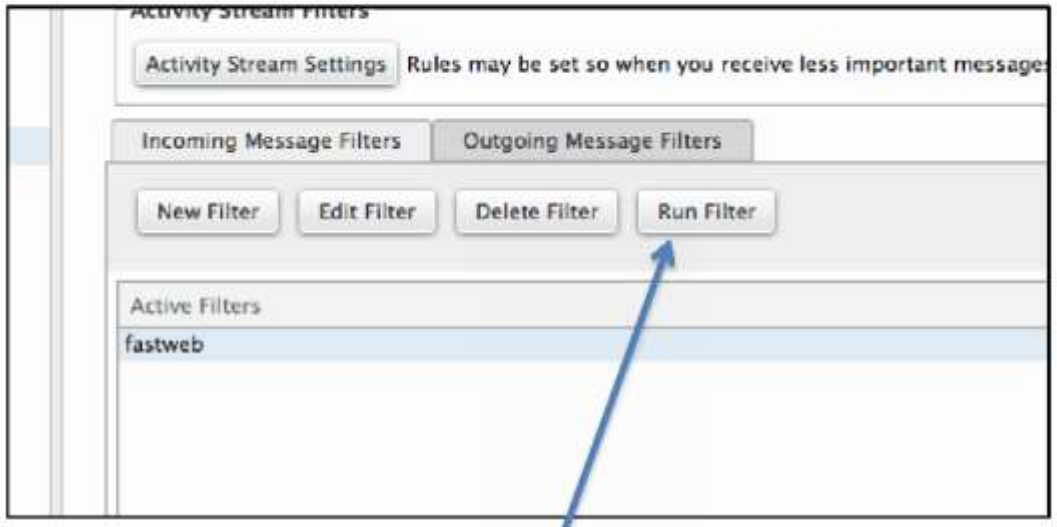
Zimbra: utilizzo dei filtri di posta

Spostare nella cartella

Azioni possibili

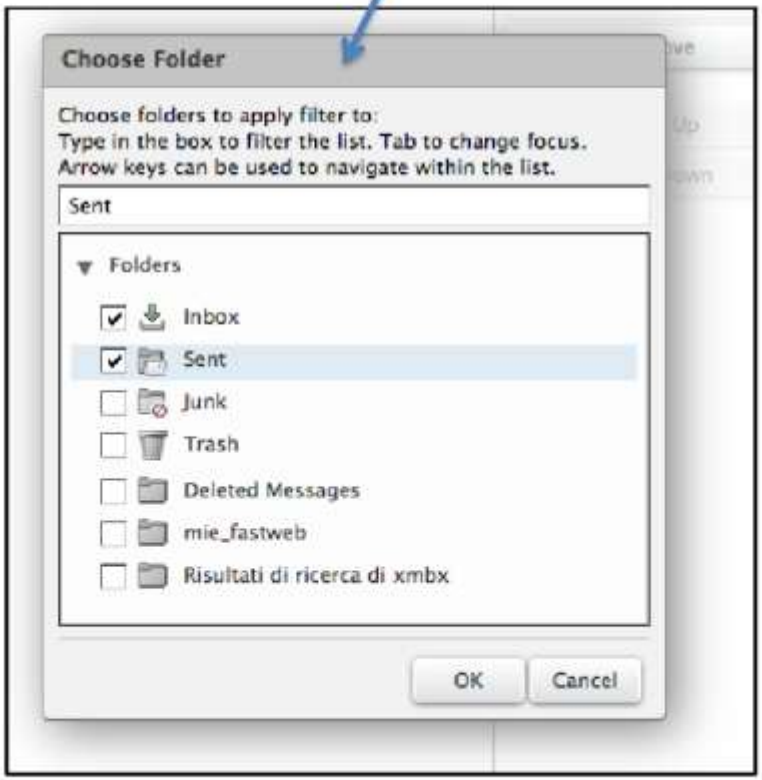


Zimbra: utilizzo dei filtri di posta

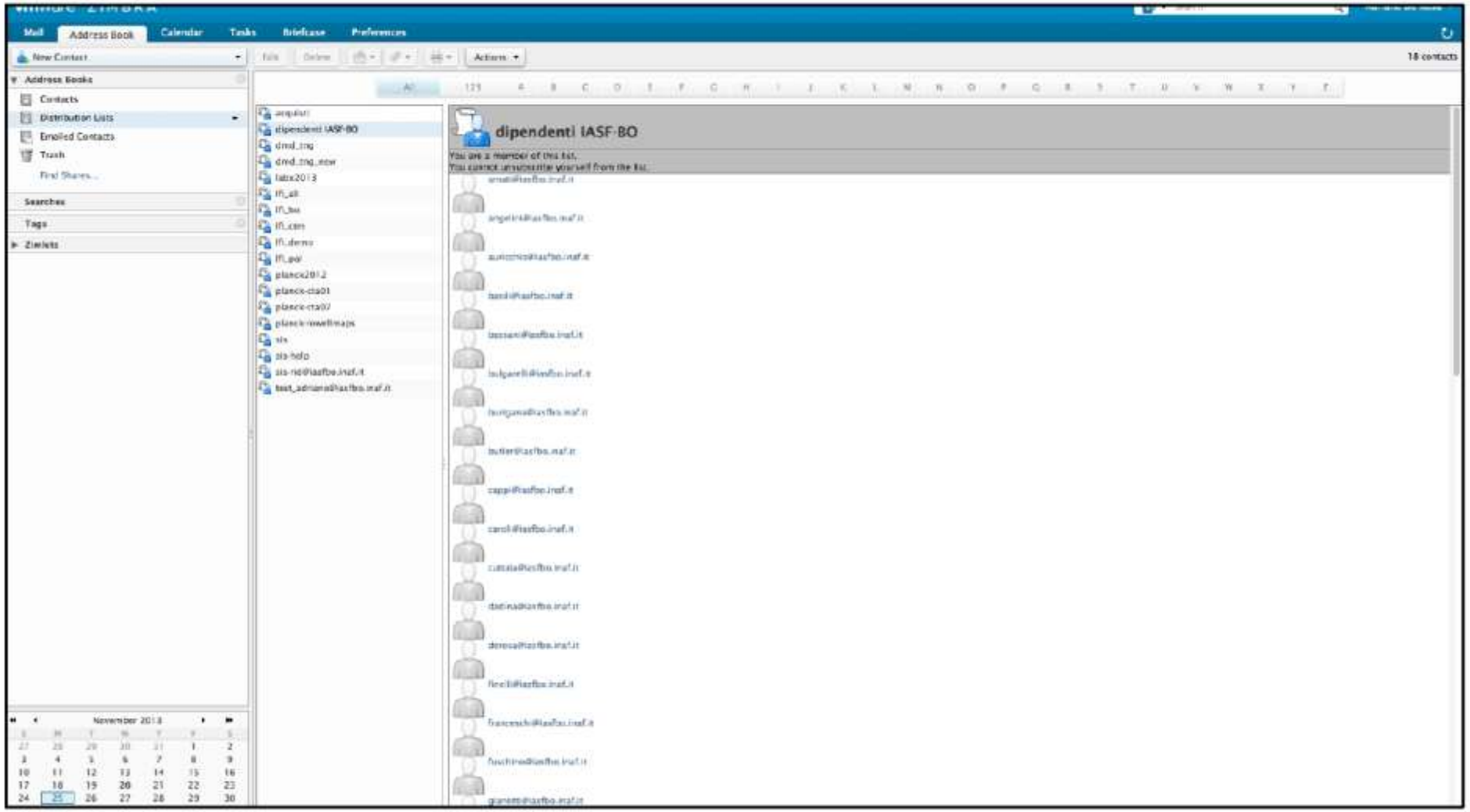


Applicazione del filtro

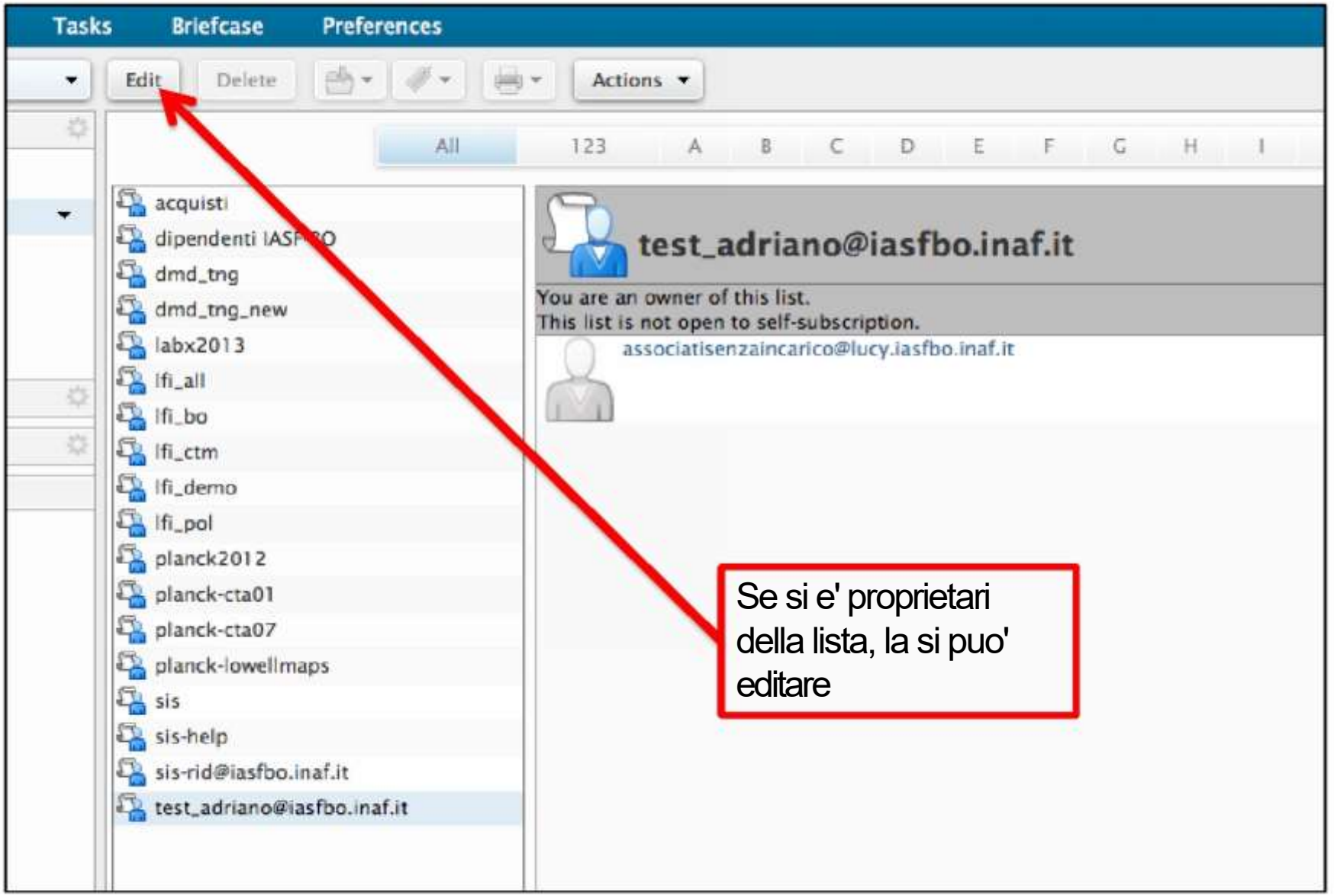
Selezione delle cartelle su cui applicare il filtro



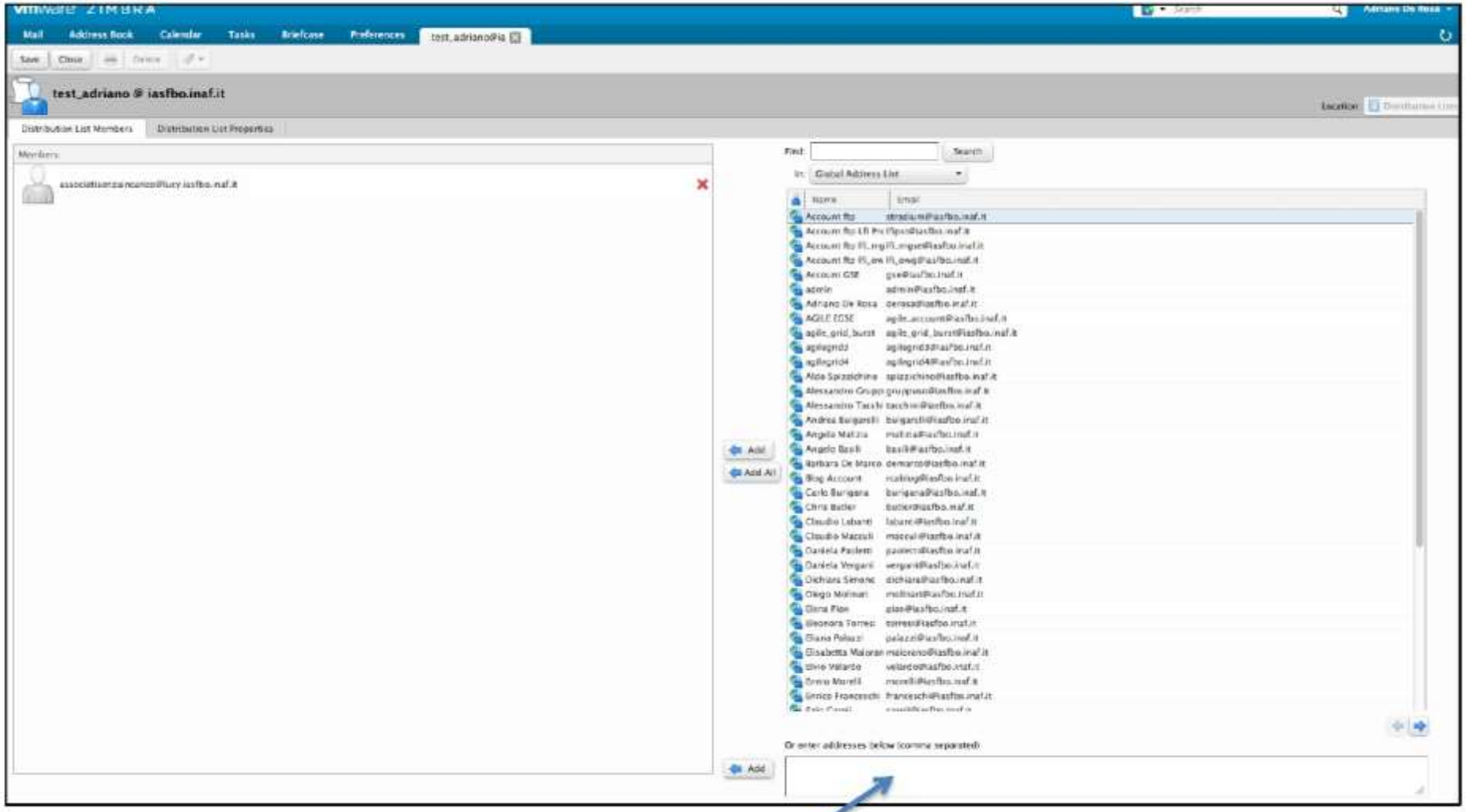
Zimbra: gestione delle liste



Zimbra: gestione delle liste

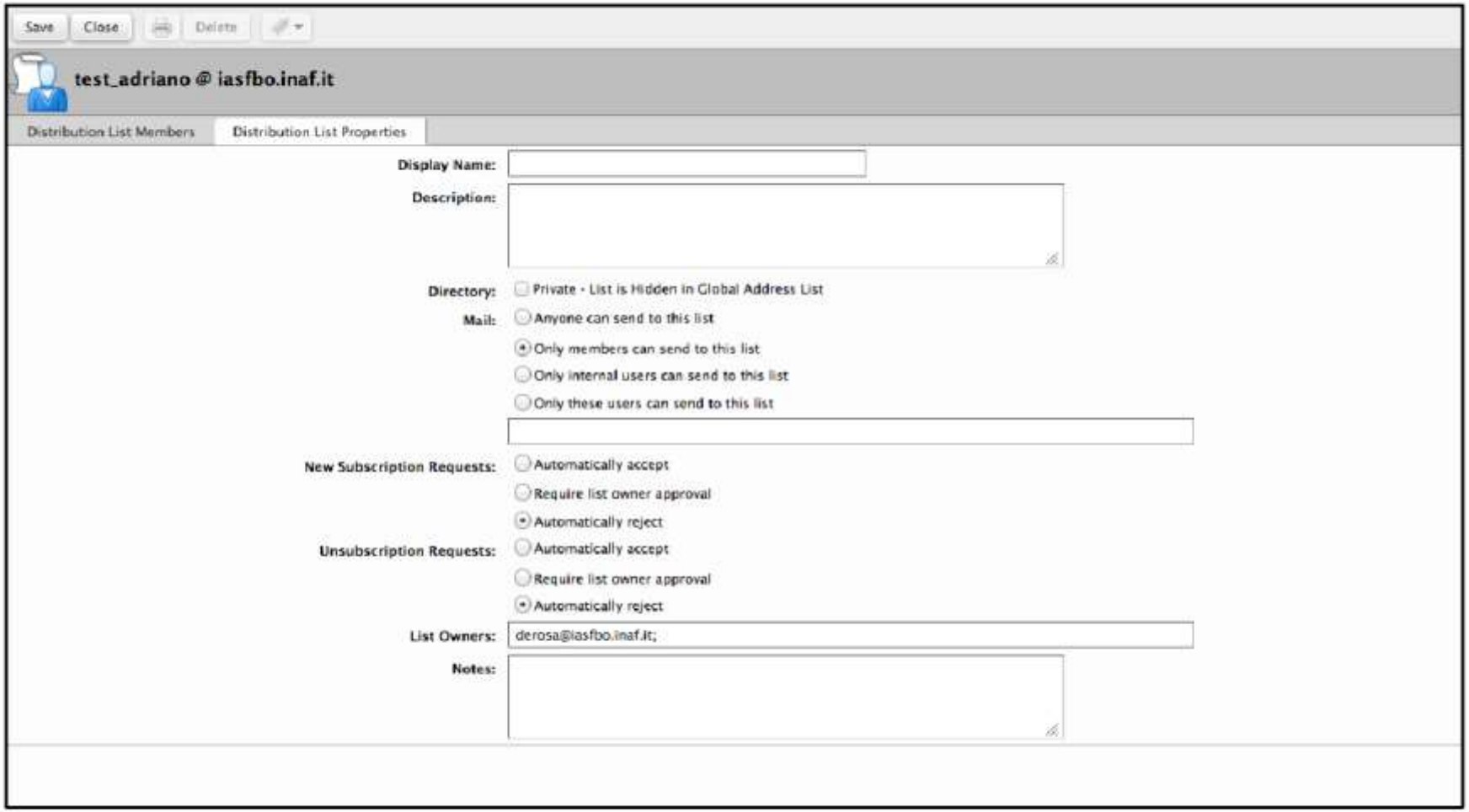


Zimbra: gestione delle liste



Spazio in cui incollare gli indirizzi, separati da virgole

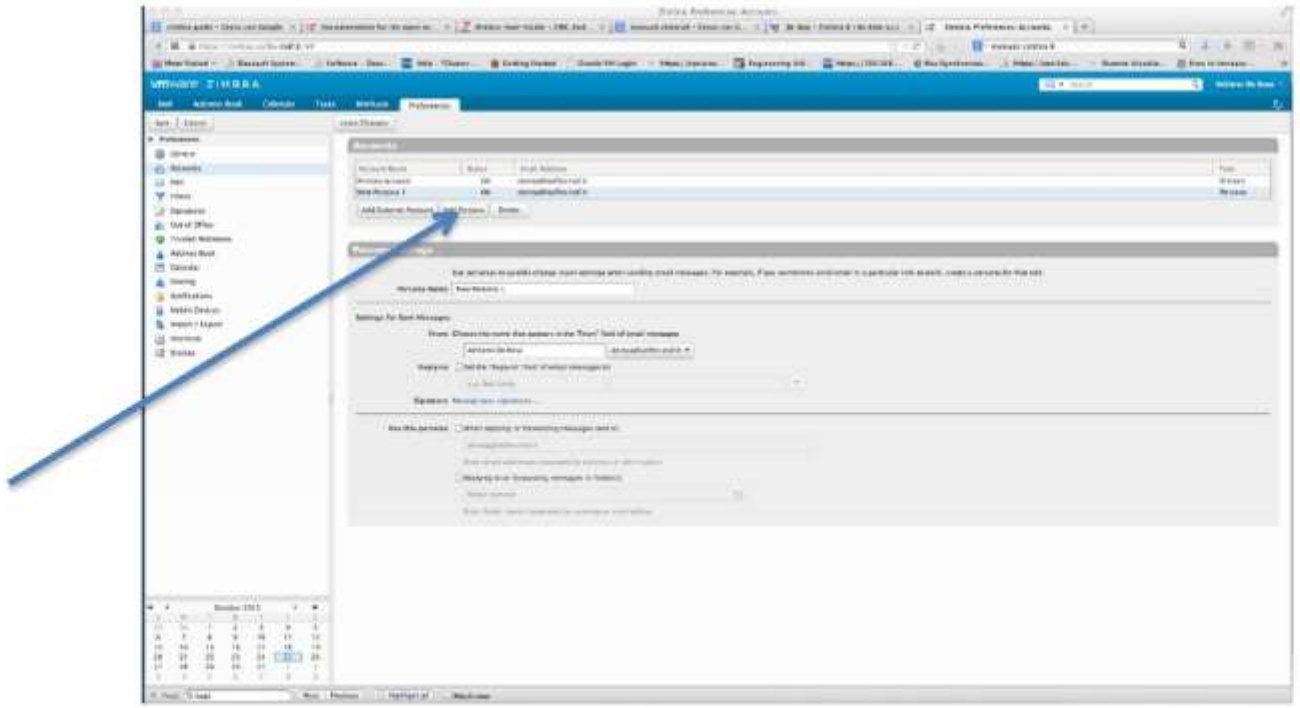
Zimbra: gestione delle liste



Zimbra: Gestione di diverse "personalita' "

Creazione ed utilizzo di "personalità" (alias)

Se un utente dispone di una personalità (l'utente può crearne una in autonomia dal menù Preferenze ->E-Mail -> Account -> Aggiungi Personalità) è possibile, in fase di composizione del messaggio, scegliere con quale "personalità" inviare l'e-mail



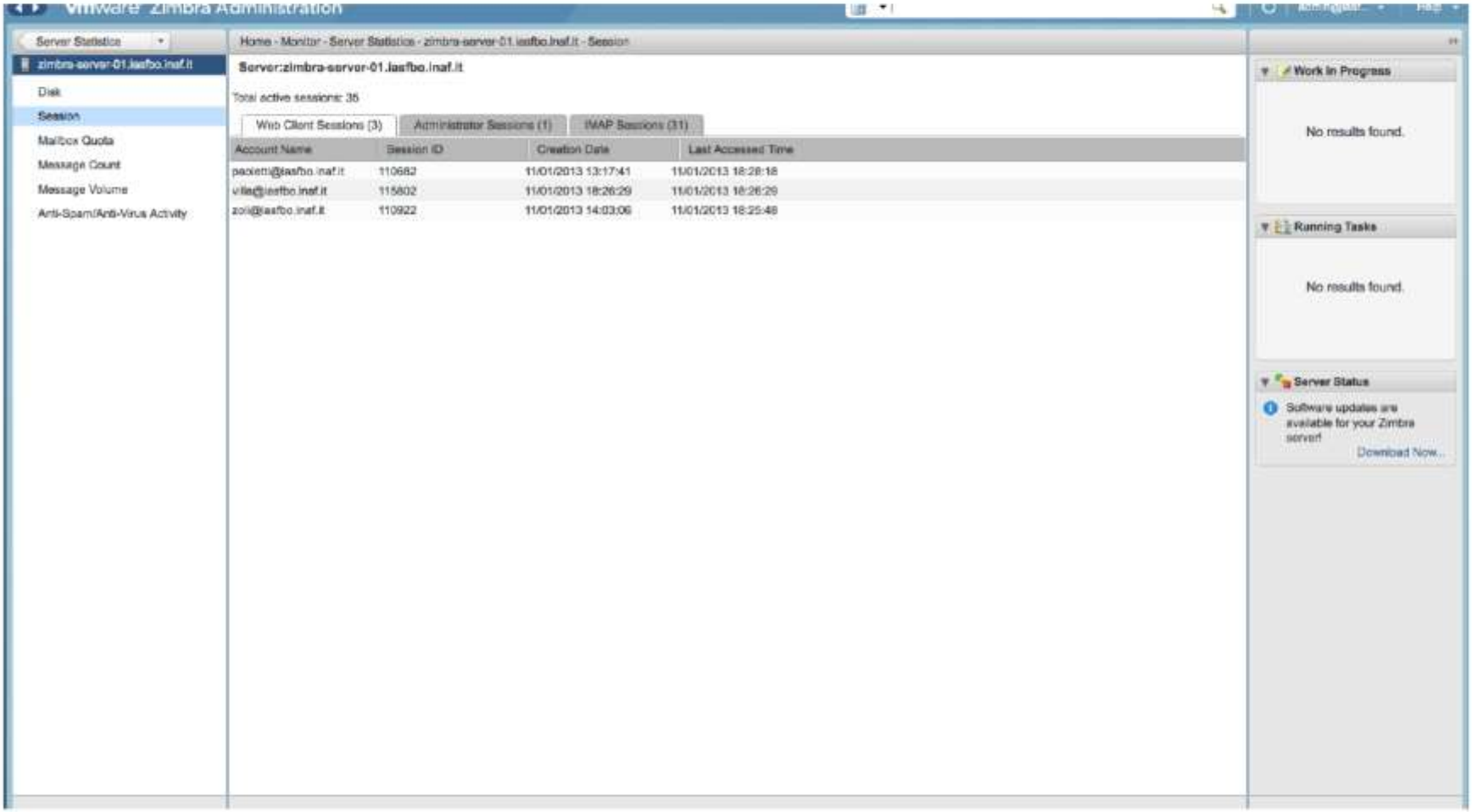
Zimbra: sessioni attive

- webclient: una sessione per ogni utente connesso alla pagina web di zimbra

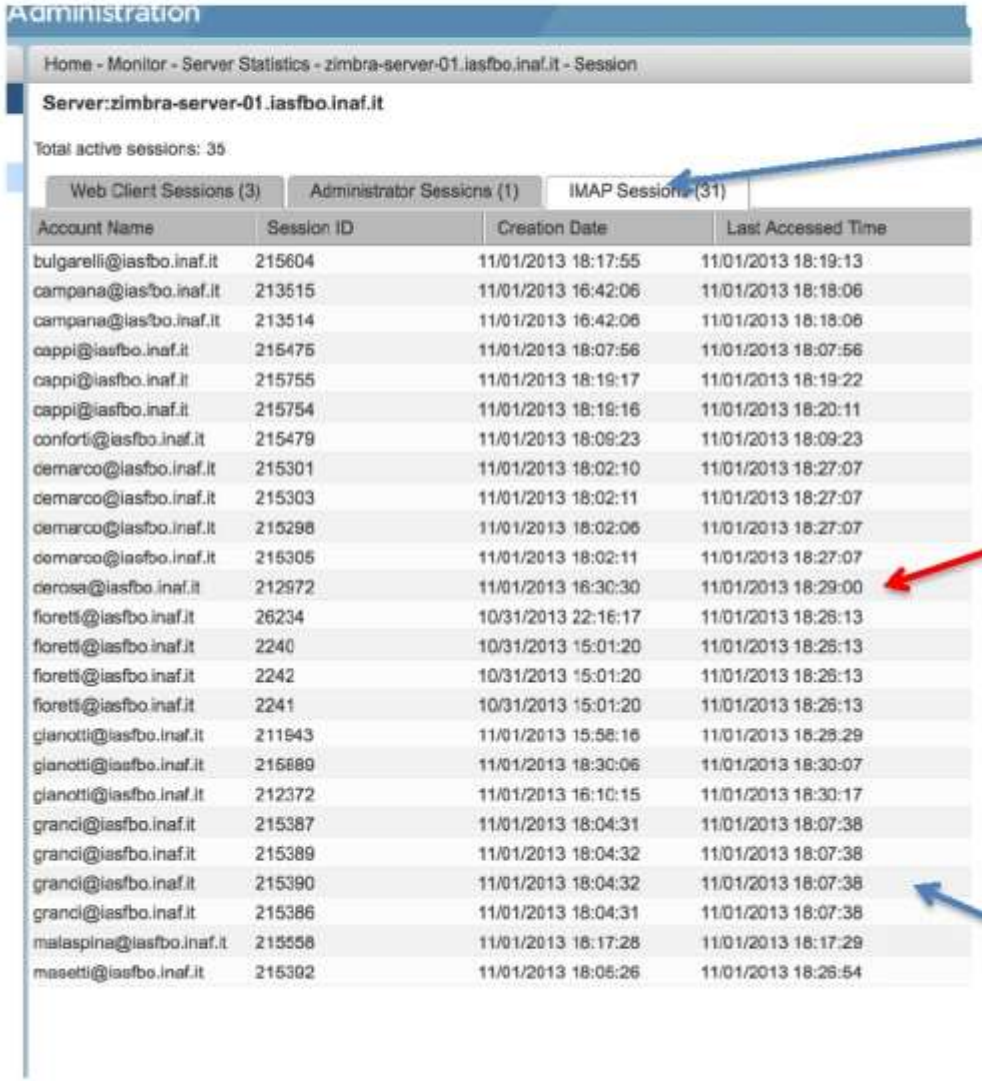
  https://zimbra.iasfbo.inaf.it

  o con zimbra desktop
- IMAP: una o piu' sessioni per ogni utente connesso in base al client di posta utilizzato. Squirrelmail (quello che chiamiamo webmail) apre piu' sessioni per ogni utente, come anche pine. Sono prodotti ormai superati!

Zimbra: sessioni attive



Zimbra: sessioni attive IMAP



Account Name	Session ID	Creation Date	Last Accessed Time
bulgarelli@iasfbo.inaf.it	215604	11/01/2013 18:17:55	11/01/2013 18:19:13
campana@iasfbo.inaf.it	213515	11/01/2013 16:42:06	11/01/2013 18:18:06
campana@iasfbo.inaf.it	213514	11/01/2013 16:42:06	11/01/2013 18:18:06
cappi@iasfbo.inaf.it	215475	11/01/2013 18:07:56	11/01/2013 18:07:56
cappi@iasfbo.inaf.it	215755	11/01/2013 18:19:17	11/01/2013 18:19:22
cappi@iasfbo.inaf.it	215754	11/01/2013 18:19:16	11/01/2013 18:20:11
conforti@iasfbo.inaf.it	215479	11/01/2013 18:09:23	11/01/2013 18:09:23
cemarco@iasfbo.inaf.it	215301	11/01/2013 18:02:10	11/01/2013 18:27:07
cemarco@iasfbo.inaf.it	215303	11/01/2013 18:02:11	11/01/2013 18:27:07
cemarco@iasfbo.inaf.it	215298	11/01/2013 18:02:06	11/01/2013 18:27:07
cemarco@iasfbo.inaf.it	215305	11/01/2013 18:02:11	11/01/2013 18:27:07
derosa@iasfbo.inaf.it	212972	11/01/2013 16:30:30	11/01/2013 18:29:00
fioretti@iasfbo.inaf.it	26234	10/31/2013 22:16:17	11/01/2013 18:26:13
fioretti@iasfbo.inaf.it	2240	10/31/2013 15:01:20	11/01/2013 18:26:13
fioretti@iasfbo.inaf.it	2242	10/31/2013 15:01:20	11/01/2013 18:26:13
fioretti@iasfbo.inaf.it	2241	10/31/2013 15:01:20	11/01/2013 18:26:13
gianotti@iasfbo.inaf.it	211943	11/01/2013 15:58:16	11/01/2013 18:28:29
gianotti@iasfbo.inaf.it	215689	11/01/2013 18:30:06	11/01/2013 18:30:07
gianotti@iasfbo.inaf.it	212372	11/01/2013 16:10:15	11/01/2013 18:30:17
granci@iasfbo.inaf.it	215387	11/01/2013 18:04:31	11/01/2013 18:07:38
granci@iasfbo.inaf.it	215389	11/01/2013 18:04:32	11/01/2013 18:07:38
granci@iasfbo.inaf.it	215390	11/01/2013 18:04:32	11/01/2013 18:07:38
granci@iasfbo.inaf.it	215386	11/01/2013 18:04:31	11/01/2013 18:07:38
malaspina@iasfbo.inaf.it	215558	11/01/2013 18:17:28	11/01/2013 18:17:29
masetti@iasfbo.inaf.it	215302	11/01/2013 18:05:26	11/01/2013 18:26:54

Apple mail

https://webmail.iasfbo.inaf.it

Zimbra: pine e webmail

- Squirrelmail e pine NON SUPPORTANO DI FATTO I PROTOCOLLI DI AUTENTICAZIONE RICHIESTI DA ZIMBRA (esperienza fatta durante la migrazione)
- GENERANO UNA FALLA DI SICUREZZA NEL SISTEMA
- Squirrelmail e pine generano un sovraccarico inutile di sessioni IMAP.
- Sarebbe auspicabile che gli utenti di loro spontanea volonta' utilizzassero gli altri client di posta offerti!

Sistemi di stampa

- I piano:
  - Xerox colorqube 8870 - colori cera (192.168.166.191)
  - Xerox pro35 fotocopiatrice b/n xilografica (192.168.166.11)
  - Hp T610 plotter (192.167.166.15)
- II piano:
  - Xerox colorqube 8870 - colori cera (192.168.166.192) - Xerox workcentre 5775 fotocopiatrice b/n xilografica (192.168.166.12)
- III piano:
  - Xerox phaser 8860 - colori cera (192.168.166.193)

Costi di Stampa / contratti attivi

- Le fotocopiatrici hanno un numero di copie annue compreso nel contratto di noleggio:
  - I piano 40000 copie/anno; - II piano 80000 copie/anno;
- Le stampanti NON hanno un numero di copie compreso nel noleggio o nell'assistenza, ma si paga a consumo (in base alla percentuale di colore):
  I livello (<1.2%) / II livello (1.2% - 8%) / III livello (> 8%)
  - I e II piano 0.0085/0.0365/0.0850 - III piano 0.0096/0.0412/0.0962

I costi indicati sono iva esclusa.

Installazione delle stampanti

La soluzione migliore e' quella di scaricare al momento dell'installazione il driver piu' recente per il proprio sistema operativo dal sito web di xerox. Facilmente trovabile con una ricerca su internet, del tipo "xerox nome_stampante driver".

La scelta del protocollo di stampa in windows e' limitata a PS o PCL6 (si consiglia PS). In windows la ricerca della stampante sulla rete e' automatica, come anche la creazione della porta.

Su linux e mac e' necessario creare prima una porta HP-JETDIRECT o IPP, associata all'indirizzo ip della stampante da installare.
In alternativa all'installazione della stampante e' possibile utilizzare il server cups (spike.iasfbo.inaf.it).
https://spike.iasfbo.inaf.it:631/printers/

Sarebbe opportuno prestare attenzione al NON condividere la stampante installata localmente sul proprio pc sulla rete. A volte il sistema di default imposta la stampante come condivisa (ubuntu).

I files .ppd per linux delle stampanti sono nella cartella /prod_iasfbo/XEROX

Da Joomla a Wordpress

come eravamo

come siamo

come saremo…

Cosa è rimasto uguale?

- Architettura: Joomla e Wordpress sono entrambi CMS
- La maggior parte delle voci di menù
- Le news in primo piano, ora arricchite
- Il *widget* sui *forthcoming events*

Cosa c'è

- Milena/Alessandro/Eleonora
- Lo sfondo che richiama il sito istituzionale
- Il banner con i satelliti
- Doppia lingua
- L'automatismo per le news IASF-BO da Media INAF
- **L'accesso al *backend* per tutti (*dashboard*)**
- La pagina con le presenze accessibile dall'esterno
- L'elenco delle pubblicazioni da CRIS (e, volendo, il CV)
- Il *tool* per le prenotazioni
- Versione mobile
- Facebook

Cosa c'è

- Contenuti consolidati:
  - Presentazione più ricca
  - Attività di Ricerca e Tecnologica
  - Didattica e alta formazione
  - Public outreach
  - Facilities
- Spostato su nuovo server
- Documentazione istituzionale IASF-BO
- Alcune voci di menù (arriveranno)
- Job opportunities (ora nelle news)
- La photo-gallery (transiterà su Flickr)
- I file corrotti dall'hacker

Come crescerà?

- Album fotografico su Flickr
- Pagine personali?
- Strumenti amministrativi?
- Quello che chiederete…

Pagine Web personali

Le pagine web personali sono accessibili dall'esterno all'indirizzo:

http://www.iasfbo.inaf.it/~nome_utente

Il contenuto della propria pagina web risiede nella cartella HTML della home dell'utente.

Essendo parte integrante della home, la cartella HTML contribuisce al raggiungimento della quota disponibile per ogni utente sul filesystem HOME e ne viene eseguito un backup giornaliero.

Siti FTP/SFTP con accessi condivisi

- E' proibito dalle normative di accesso alla rete GARR, di cui facciamo parte, il fatto di non poter identificare gli utenti che accedono ai sistemi in modo univoco.
- Gli account di sharing sono quindi da eliminare dai sistemi.
- Questa regola vale per qualunque sistema connesso alla rete, anche per quelli dei singoli ricercatori.

Licenze software

- IDL
- MICROSOFT
- ACROBAT PRO
- LABVIEW

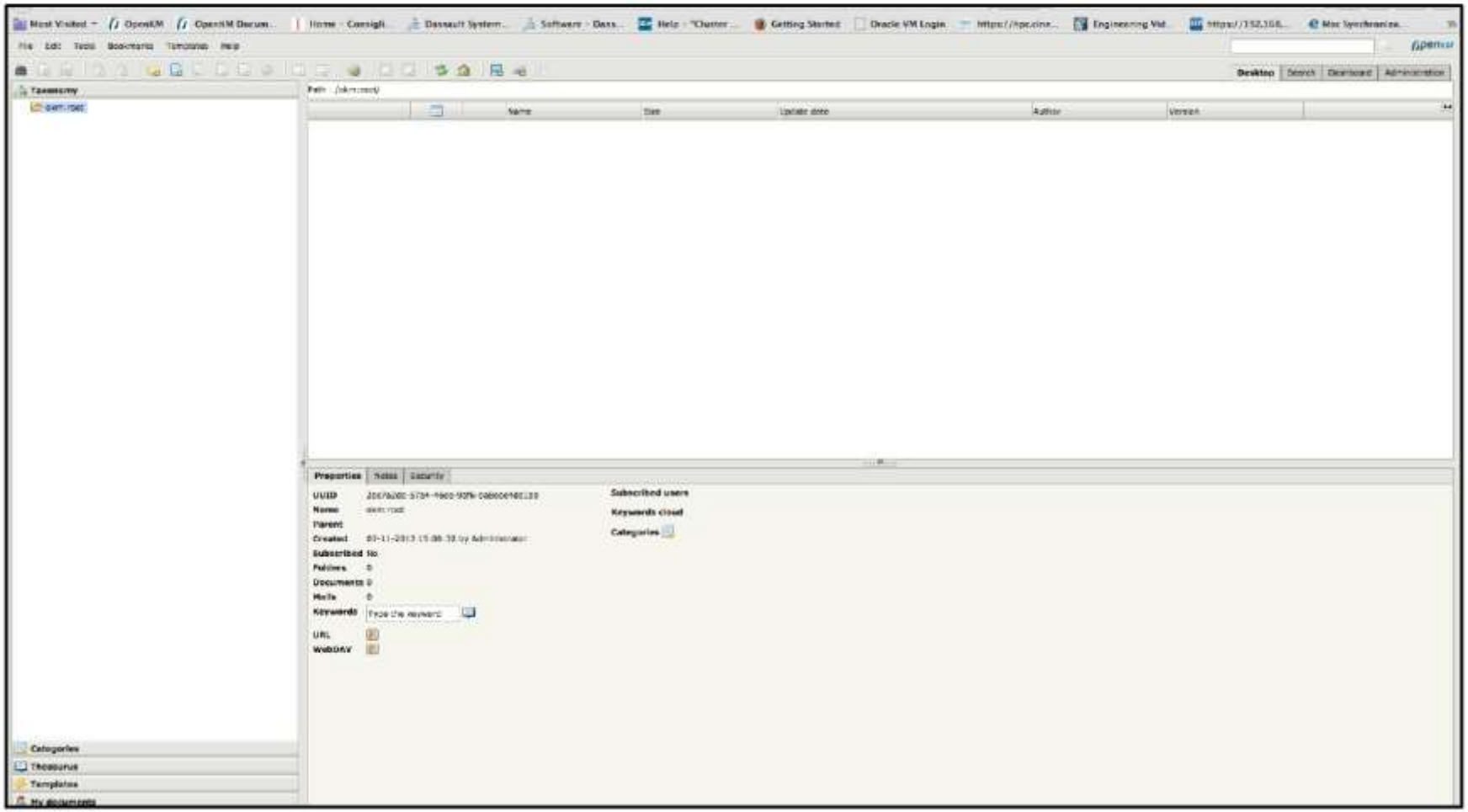
Quelle che acquisirei…

- Compilatore intel
- Gpfs
- NAG fortran library smp - gpu

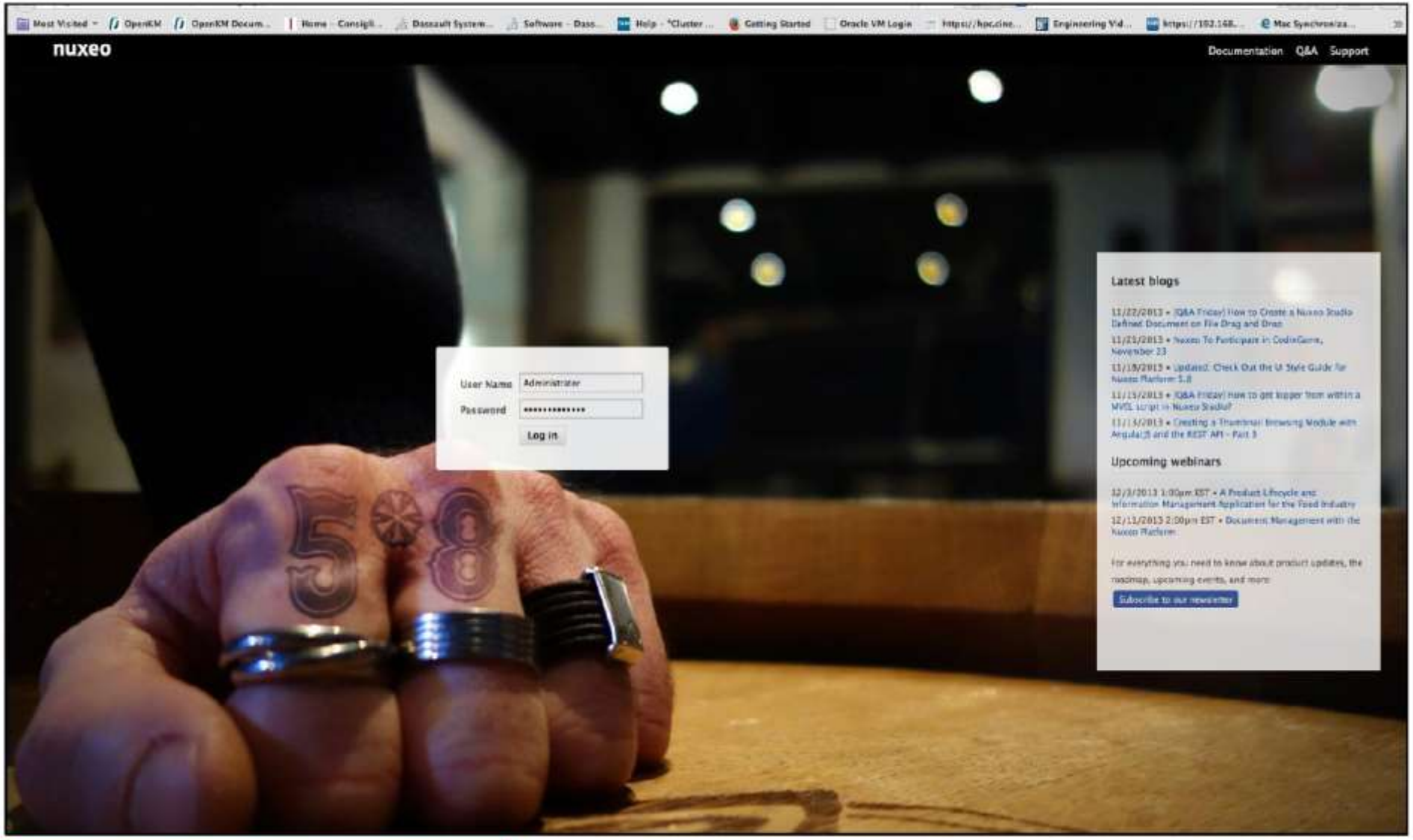
openKM

http://openkm.iasfbo.inaf.it:8080

openKM

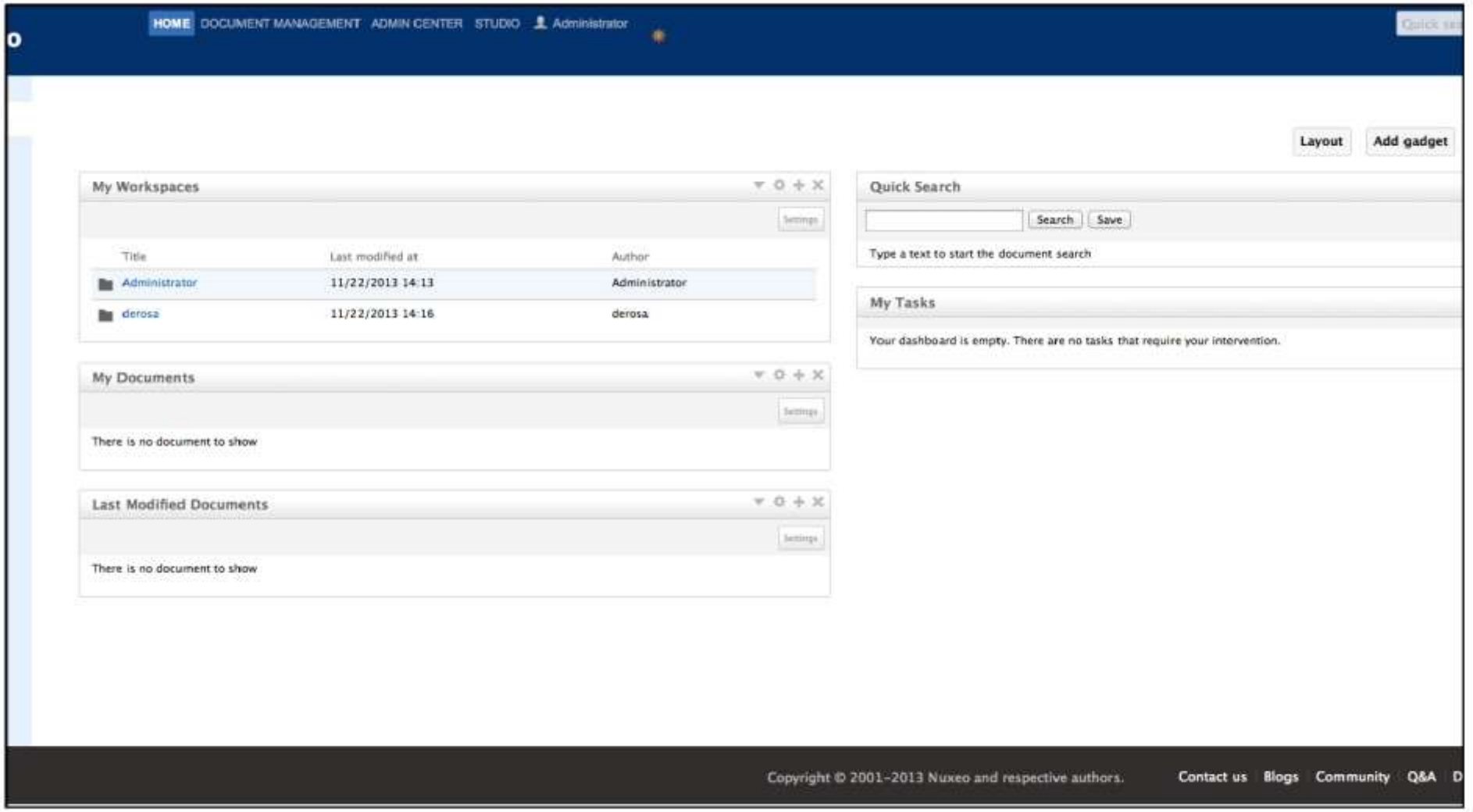

Taxonomy
Path
Name
Size
Update date
Author
Version
Desktop
Search
Dashboard
Administration
Properties
Notes
Security
UUID
Name
Parent
Created
Subscribed
Folders
Documents
Mails
Keywords
URL
WebDAV
Subscribed users
Keywords cloud
Categories
Categories
Thesaurus
Templates

Nuxeo

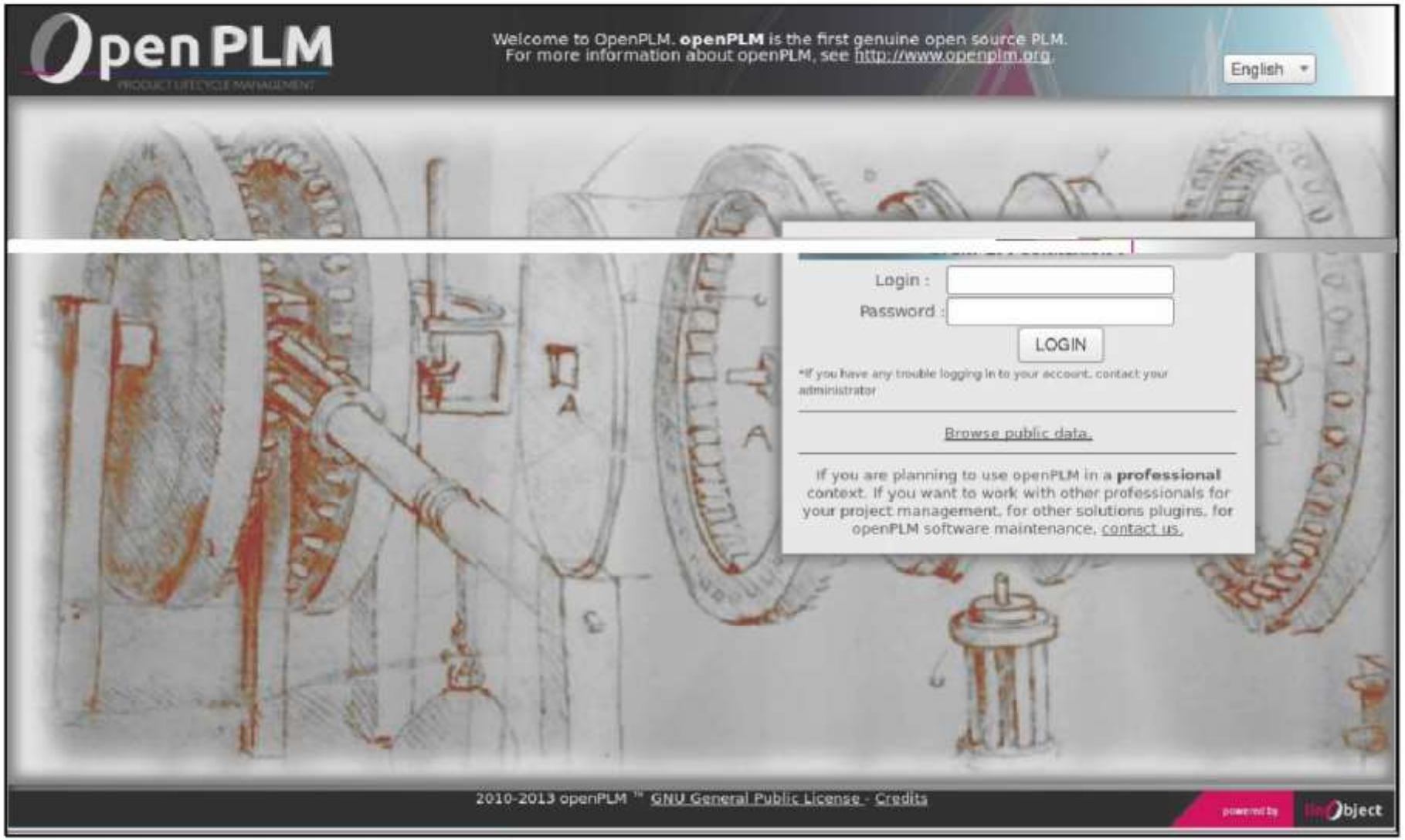


http://nuxeo.iasfbo.inaf.it:8080

Nuxeo


HOME
DOCUMENT MANAGEMENT
ADMIN CENTER
STUDIO
Administrator
Layout
Add gadget
My Workspaces
Title
Last modified at
Author
Administrator
11/22/2013 14:13
Administrator
derosa
11/22/2013 14:16
derosa
Quick Search
Search
Save
Type a text to start the document search
My Documents
There is no document to show
My Tasks
Your dashboard is empty. There are no tasks that require your intervention.
Last Modified Documents
There is no document to show
Copyright © 2001-2013 Nuxeo and respective authors.
Contact us
Blogs
Community
Q&A

OpenPLM



http://openplm.iasfbo.inaf.it

OWNCLOUD

http://owncloud.iasfbo.inaf.it/owncloud