Centralizzazione del servizio di posta per l'INFN

A cura del gruppo Mailing[1]

Premessa

La possibilità di centralizzare il servizio di posta per l'INFN viene ridiscussa periodicamente, con l'intenzione di ridurre il carico di lavoro del personale dei servizi di calcolo impiegato in servizi di base e liberare risorse da destinare a progetti più strategici.

Questo documento, redatto da personale coinvolto direttamente nella gestione del servizio di posta in alcune sezioni dell'ente, cerca di illustrare i diversi aspetti di un possibile servizio centralizzato, mettendo in evidenza vantaggi e svantaggi delle diverse soluzioni. Destinataria del documento è la Commissione Calcolo e Reti, che potrà valutare se l'investimento previsto in termini di personale e risorse economiche possa risultare vantaggioso per l'Ente.

Il documento è formato da quattro parti. La prima parte descrive il servizio di posta elettronica dell'INFN, il livello di servizi offerto e l'impegno attuale dell'Ente in termini di risorse umane, di hardware e software; la seconda parte descrive le tecnologie opensource e di tipo proprietario disponibili oggi sul mercato, ritenute capaci di fornire un servizio di posta centralizzata di livello paragonabile o superiore a quello aggregato attuale; in una terza parte sono schematizzate alcune soluzioni adottate da enti o organizzazioni con esigenze simili a quelle dell'INFN; l'ultima parte riassume e ricombina le soluzioni descritte in una serie di schemi, allo scopo di mostrare in modo sintetico quali siano le possibili alternative che l'Ente ha oggi di fronte.

[1] Testi di Mirko Corosu, Fulvia Costa, Ombretta Pinazza, Alessandro Spanu, Riccardo Veraldi, Giulia Vita Finzi

Prima parte

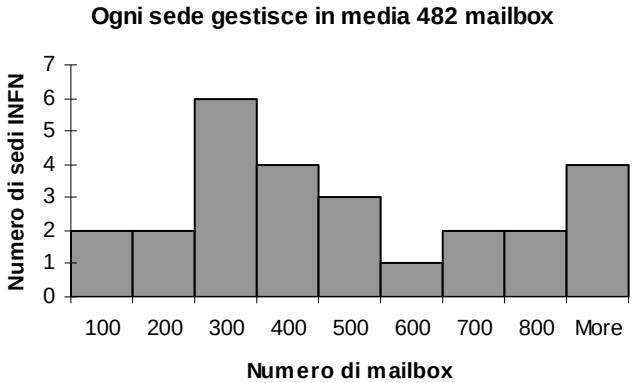
Il servizio attuale

Nel maggio 2007 il gruppo Mailing ha completato un censimento fra le sedi INFN per capire quale livello di servizio è fornito nelle diverse sedi, con quali tecnologie hardware e software e quali sono i problemi principali segnalati dal personale che gestisce il servizio.

Hanno partecipato alla raccolta dati le 20 Sezioni, i 4 Laboratori, il CNAF e il gruppo collegato di Parma che gestisce un servizio paragonabile a quello di molte Sezioni. I valori totali riportati sono quindi riferiti a 26 unità.

La prima tabella aiuta a comprendere le dimensioni del servizio. Le caselle di posta elettronica gestite dai Servizi Calcolo dell'Ente non sono solo utilizzate dal personale dipendente o associato (che comunque preferisce mantenere anche un account di posta presso le sedi INFN, oltre all'account fornito dalle Università) ma anche da un grande numero di utenti che non hanno rapporti ufficiali con l'INFN: studenti non associati, collaboratori, ecc.
Quattro sedi (Roma1, Milano, Napoli e LNF) si ritrovano addirittura a gestire circa mille utenti ciascuna, la metà dei dipendenti INFN.

Utenza	Totale	Sezione media
Numero di mailbox	**12532**	**482**
- dipendenti	~2000	105
- associati	3355	129
- altri	7177	258



Il personale

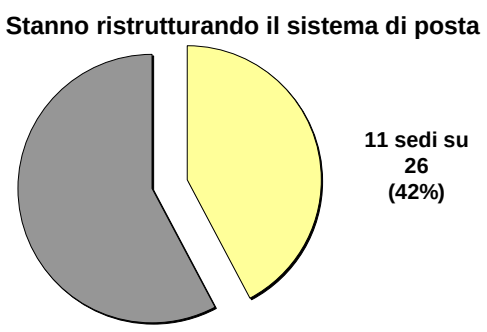
La seconda tabella riporta il numero di dipendenti facenti parte dei Servizi Calcolo e Reti[2]
Non è facile stimare quale frazione del personale possa essere associata al servizio di posta elettronica. E' parere del gruppo Mailing che, con il servizio a regime, in ogni sede un servizio di helpdesk composto da due FTE sia impegnato per il 10% del proprio tempo a risolvere problemi legati alla posta; ciò porterebbe a calcolare 0.2 FTE per sede e 5.2 FTE per tutto l'Ente.

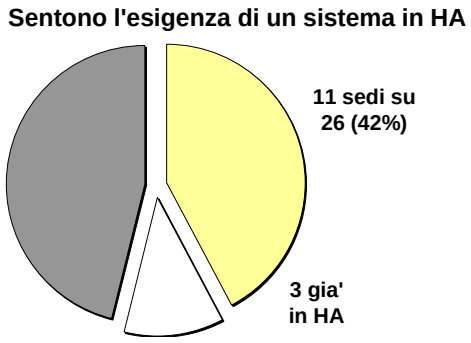
Servizi calcolo - Personale	Totale		Sezione media	
Tecnologi/ricercatori	37		1.4	
Tecnici (FTE)	78	(62.3 FTE)	3.0	(2.4 FTE)
Totale	**115**	**(99.3 FTE)**	**4.4**	**(3.8 FTE)**
Personale impegnato nella gestione della posta (stima)	**5.2 FTE**		**0.2 FTE**	

Le fasi di installazione, aggiornamento, riconfigurazione risultano particolarmente impegnative per tutto il servizio, ma non sono proporzionali al numero di utenti.

Dalla raccolta di informazioni presso le sedi è risultato anche che quasi la metà delle sedi sta lavorando per aggiornare, rinnovare o migliorare il proprio sistema di posta.

Tutte le strutture si dicono in generale soddisfatte del livello di servizio fornito, ma la metà manifesta la proccupazione di garantire un servizio ad alta affidabilità, che al momento risulta disponibile soltanto in tre sedi[3].





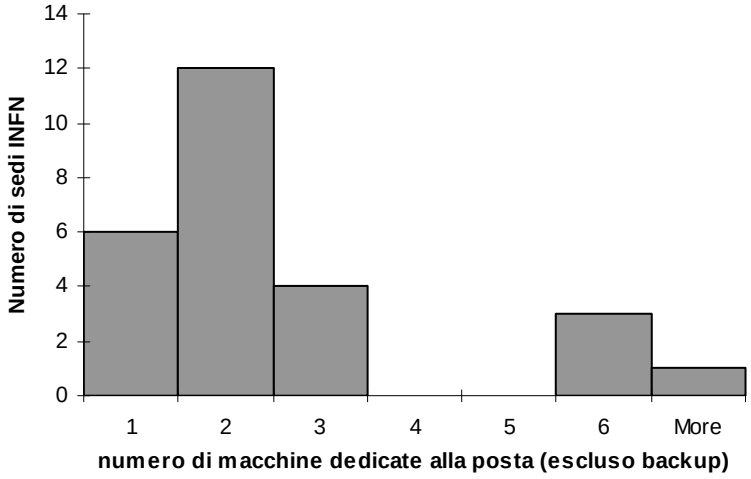
[2] Dati ricavati dal database Preventivi 2008
[3] A. Tirel, C. Strizzolo, Servizio di Posta Elettronica ad Alta Affidabilità, INFN/TC-07/01
A. Brunengo, M. Corosu, Configurazione dei Servizi di Posta Elettronica per la Sezione INFN di Genova, INFN/TC-07-2

L'hardware

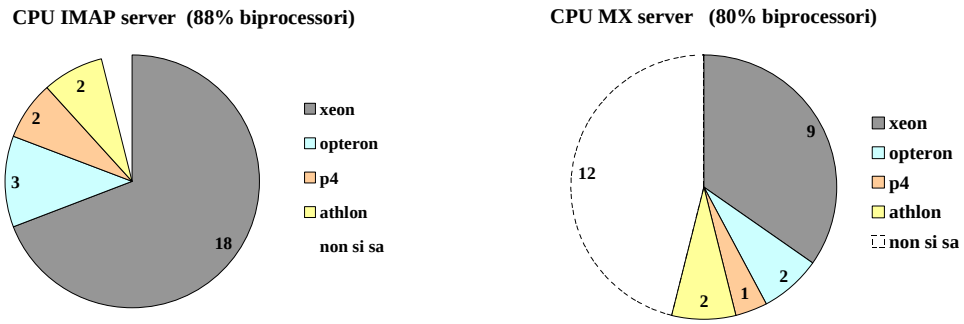
La maggior parte delle sedi ha scelto di dedicare al servizio di posta elettronica almeno due macchine, separando quasi sempre il servizio di ricezione/spedizione dei messaggi dalla gestione delle caselle di posta.

Hardware impegnato	Totale	Sezione media
Macchine	70	2.7
Spazio disco allocate (stima)	16 TB	700 GB
Spazio disco utilizzato (stima)	3 TB	125 GB

La quantità effettiva totale di spazio disco non è nota: per stimarla è stata presa come campione una Sezione (Bologna) che ha un numero di utenti e di macchine molto vicini ai valori della Sezione media.



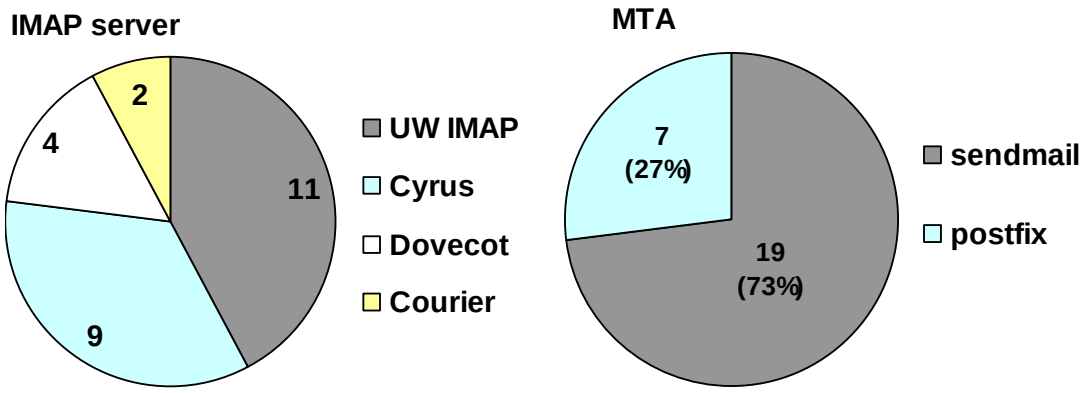
Ogni sede utilizza in media 2.7 macchine dedicate a garantire il solo servizio di posta, senza contare il backup.



Le macchine preferite per il ruolo di IMAP server e server SMTP sono biprocessori Xeon.

Il software

Tutte le sedi INFN utilizzano software opensource su piattaforme linux o BSD. Alcune sedi mantengono ancora macchine Digital Unix, ma stanno provvedendo alla loro sostituzione.



Per il servizio di ricezione/spedizione della posta il software più utilizzato è sendmail, anche se per i sistemi più recenti si tende a preferire postfix, grazie soprattutto alla sua semplicità di configurazione e alla buona integrazione con il sistema antispam/antivirus adottato dall'Ente.

IMAP (Internet Message Access Protocol) è il protocollo di comunicazione per la ricezione di e-mail e permette ad un client (o a più client simultaneamente) di accedere, leggere, cancellare le e-mail da un server, in modalità online o offline.
Le caselle di posta (mailbox) possono essere di tre tipi:

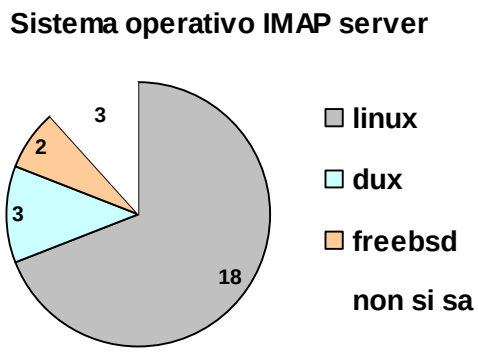
1. ogni mailbox è costituita da un unico file, nel quale i messaggi sono separati da un carattere o una stringa
2. ogni mailbox è una directory (o struttura di directory) e ogni messaggio è un file singolo
3. una mailbox è un database

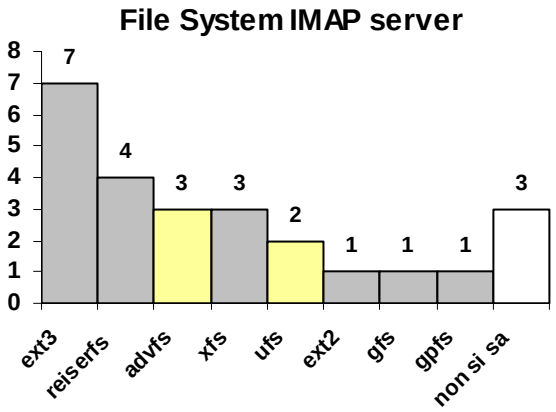
I tipi di server IMAP adottati dalle diverse sedi sono quattro: UW-IMAP, Cyrus, Dovecot e Courier.
UW IMAP è il più diffuso, anche se poco performante e poco scalabile con di caselle di posta di grandi dimensioni; nel formato tradizionale (MBOX, di tipo 1) ogni casella di posta è un unico file, e i diversi messaggi sono concatenati e separati da una stringa.
Un solo processo può accedere al file MBOX in modo read/write ed è attivo un meccanismo di locking. Il pregio di queste caselle di posta è che sono self-consistenti, quindi MBOX è il formato adatto per l'archiviazione delle cartelle.
Non è però adatto a filesystem distribuiti o su NFS, presenta problemi di scalabilità e difficoltà nella gestione di messaggi voluminosi.
Nelle versioni più recenti UW IMAP implementa versioni più avanzate di mailbox, come ad esempio MIX e MBX (sempre di tipo 1, cioè con mailbox costituita da un unico file,

ma con accesso ottimizzato per la ricerca di messaggi), ma solo pochissime sedi hanno preso in considerazione la migrazione ai nuovi formati.

Secondo per diffusione è Cyrus IMAP, che promette ottime prestazioni al prezzo di una gestione non sempre semplicissima. Alcune sedi hanno inoltre segnalato difficoltà di conversione di mailbox MBOX nel formato di cyrus, all'atto della migrazione.

Un altro aspetto importante nella scelta del server IMAP è il filesystem: per aumentare le prestazioni, è infatti raccomandata la scelta di filesystem indicizzati. I due grafici che seguono illustrano la scelta delle sedi INFN.

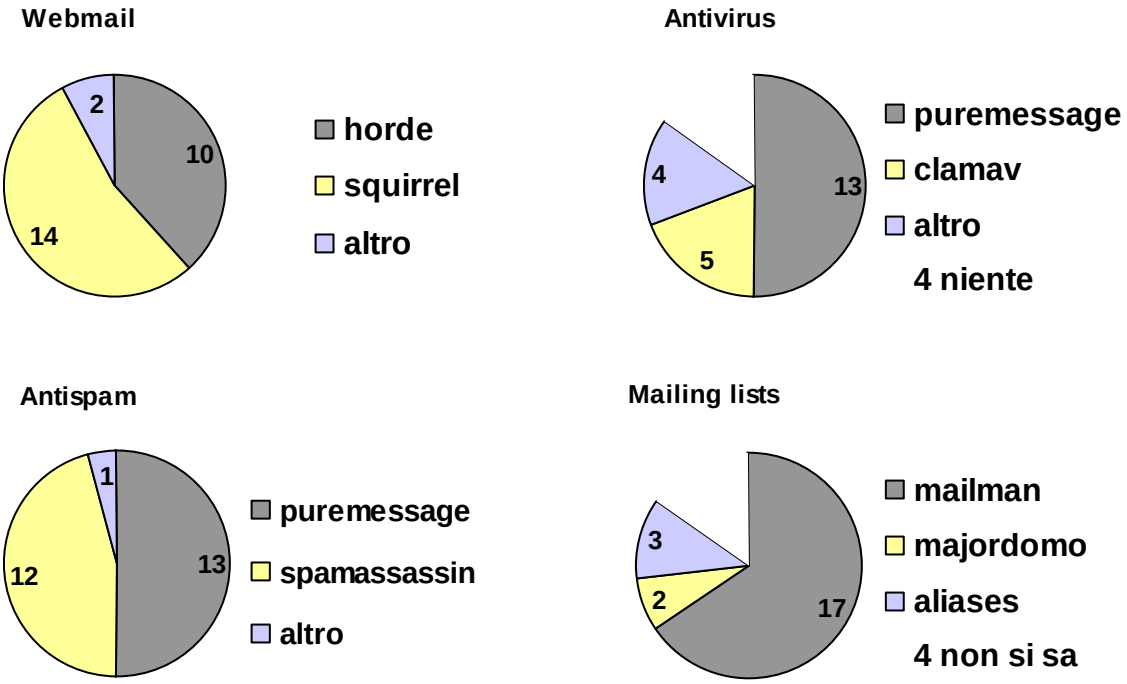




Servizi accessori

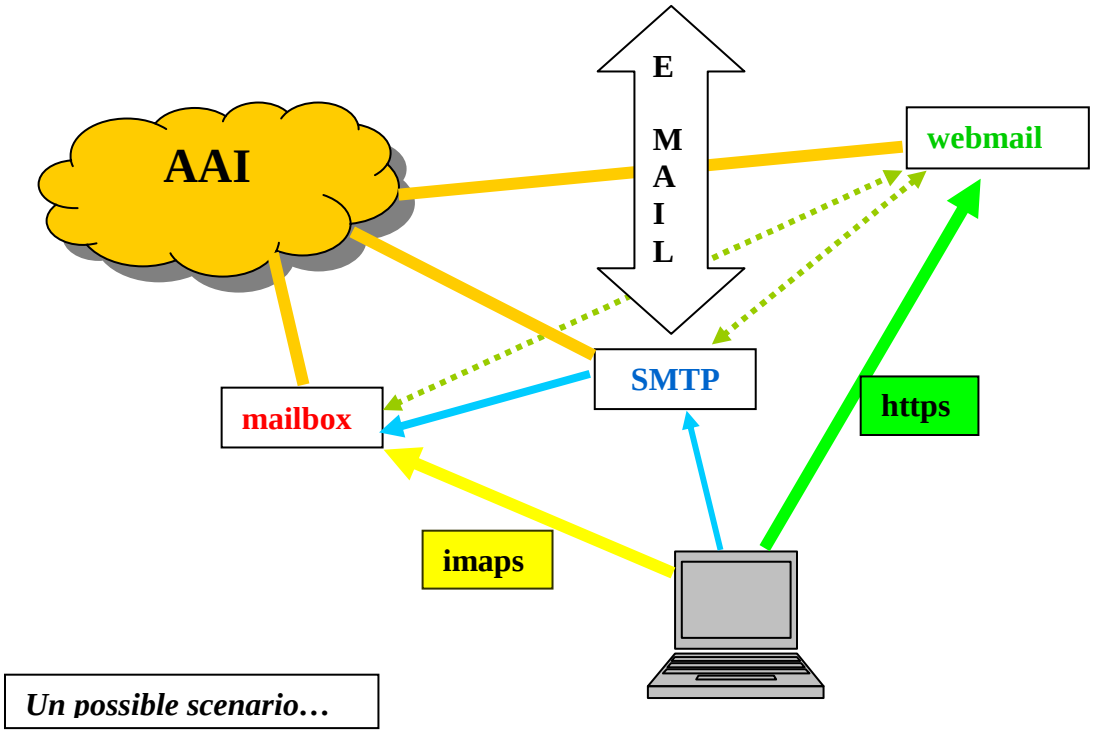
Tutte le sedi forniscono accesso alla posta elettronica tramite web, in aggiunta all'accesso tramite client IMAP e POP.

Tutti hanno implementato da tempo sistemi antispam, tanto che la ricezione di spam non è più considerata il problema prioritario del servizio di posta. I sistemi antispam preferiti sono spamassassin (opensource) e Sophos Puremessage. Quasi tutte le sedi inoltre hanno installato un antivirus sul server di posta.



Molte sedi gestiscono delle mailing list locali, quasi sempre tramite mailman. I dati rilevati sono però precedenti alla pubblicazione del servizio di mailing list Sympa del CNAF.

Autenticazione



Nella gestione quotidiana di un servizio informatizzato rivolto all'utenza in generale è necessario accedere ad una serie di informazioni che sono legate sia all'utenza stessa che alla configurazione del sistema.

In particolare, per quanto riguarda un servizio di posta elettronica, ci sono due aspetti differenti da considerare:

- l'accesso alle caselle di posta elettronica (POP/IMAP)
- l'accesso al servizio di spedizione e consegna della posta (SMTP/LMTP)

Entrambi i sottosistemi devono poter accedere ad informazioni che spesso sono distribuite su più database e server .

Ad esempio, l'accesso alle caselle di posta elettronica, effettuato sia attraverso il protocollo POP che IMAP, può essere garantito solo dopo che l'utente è stato Autenticato (tipicamente attraverso coppia di username e password) e comunque solo per le caselle per le quali l'utente è Autorizzato (tipicamente attraverso un identificativo di utente o gruppo, o anche attraverso la valutazione di Access Control Lists).

Per l'uso di un servizio di spedizione della posta (SMTP) normalmente è sufficiente solo il primo tipo di verifica, ossia l'Autenticazione o SMTP-Auth. Questa può richiedere semplicemente una lista di reti o domini autorizzati ad usare il servizio, ovvero una

Autenticazione più restrittiva, attraverso la verifica di credenziali quali username/password o il possesso di un certificato digitale X.509 valido.

Il servizio di consegna della posta (SMTP/LMTP), d'altro canto, deve poter accedere sia a database in cui sono definiti gli indirizzi "fisici" (maildrop) delle caselle di posta dei vari utenti, sia ai diversi modi con cui uno stesso utente può essere raggiunto (mailname, aliases).

Come già evidenziato, tutti questi database possono essere istallati (e normalmente lo sono) su server differenti e possono essere messi a disposizione dei servizi con modalità differenti, fra le quali mappe NIS o database locali. Questa configurazione presenta notevoli svantaggi, come ad esempio l'accessibilità non omogenea ai dati e la proliferazione dei database, database che devono essere replicati su più server in caso di architetture ridondanti realizzate per garantire alti livelli di affidabilità.

Per ovviare a tali svantaggi è necessario implementare una Infrastruttura di Autenticazione ed Autorizzazione (AAI) che facilita la gestione di tutte le informazioni, garantendone la coerenza e l'aggiornabilità, rendendo accessibili TUTTE le informazioni necessarie a TUTTI i servizi che ne hanno diritto, attraverso l'uso di UN UNICO protocollo.

Se a livello di singola sede si può sopperire alla mancanza di una AAI attraverso un controllo ed una amministrazione puntuale di tutti i database che contengono le informazioni, un sistema che ambisca ad offrire a tutte le sedi un servizio di posta elettronica deve necessariamente avere a disposizione una AAI nazionale.

Un gruppo di lavoro CCR (AAI-WG) è impegnato sull'ipotesi di realizzare una struttura nazionale, che armonizzi tutte le AAI locali e offra, al contempo, un servizio fruibile da tutte le sedi e da tutti i servizi di interesse nazionale, sia da quelli già in produzione (ad esempio la gestione mailing-list via Sympa, gli strumenti collaborativi via web, gli strumenti di gestione scientifico-amministrativa per la gestione dei preventivi) che da tutti gli altri che man mano saranno disponibili.

Help Desk

In questa parte si descrive l'aspetto della gestione del servizio mail che riguarda i rapporti diretti con l'utente. Cercando tra le varie richieste che un servizio calcolo riceve, si possono individuare tre categorie di attività:

1. affrontare i problemi segnalati dagli utenti,
2. rispondere alle richieste degli utenti
3. operare per la gestione e manutenzione del servizio.

Nel caso di un servizio di mail centralizzato si può cercare di distinguere quello che dovrebbe essere fatto sul sito centrale da quello che invece dovrebbe continuare ad essere fatto nelle sedi. A tal fine si parlerà rispettivamente di servizio centrale e di servizio locale.

1. Problemi segnalati dagli utenti:

1. *Problemi di configurazione:* rimangono a carico del servizio locale nel caso in cui l'utente abbia modificato erroneamente la configurazione, se però la configurazione è apparentemente corretta potrebbe essere necessario l'accesso al server centrale per la verifica.
2. *Spam falsi positivi:* dipende dalla tecnologia scelta per l'individuazione degli spam; il servizio locale può o meno capire perché un mail è stato considerato spam, ma l'eventuale correzione del problema è a carico del servizio centrale.
3. *Tracciabilità dei mail:* una delle richieste più frequenti consiste nel sapere che fine ha fatto un certo mail, se è partito o è fermo in qualche coda o se è arrivato e non è stato consegnato, se ci sono problemi tra i mail server coinvolti, ecc. Questa attività può essere svolta solo sul server centrale.
4. *Mail che rimbalzano:* spesso gli utenti hanno difficoltà nell'interpretare i messaggi di errore, il servizio locale può aiutare nell'interpretazione, può fare meno per risolvere il problema.
5. *Mail sospetti:* gli utenti spesso segnalano mail di phishing, o mail che tornano indietro senza che loro abbiano spedito niente, o cose simili; il servizio locale può in genere far fronte a questo problema e anche verificare l'eventuale presenza di virus sul computer dell'utente.
6. *Spam:* a seconda della politica che verrà adottata sul mail server centrale gli spam saranno recapitati o meno, ci saranno segnalazioni di falsi negativi o di filtri poco o troppo severi, il servizio locale non potrà fare niente.
7. *Quote:* la gestione delle quote è fatta sul server centrale.
8. *Recupero dal backup:* nel caso di mail cancellati inavvertitamente, l'operazione di recupero deve essere fatta nel sito centrale.

2. Richieste degli utenti:

1. *Configurazione del programma di lettura:* rimane a carico del servizio locale.
2. *Forward e filtri:* poiché forward e filtri sono configurati sul mailserver, queste richieste devono essere soddisfatte dal servizio centrale.

3. *Creazione e gestione di mailing lists:* con l'impostazione di un servizio centralizzato, anche le mailing lists dovranno risiedere sul server centrale.

3. Gestione del servizio:

1. *Creazione e cancellazione degli account di posta:* dipenderà dal modello di centralizzazione scelto.
2. *Controllo:* compiti di controllo quali la verifica dello stato delle code e dei mail al postmaster rappresentano un servizio che viene reso all'utente, e dovranno essere svolti dal servizio centrale.

Deduzioni:

Per poter mantenere il servizio ad un'efficienza pari a quella di oggi, si ritiene ci siano alcuni aspetti che continueranno a gravare sui servizi locali, e che richiederanno un metodo sicuro attraverso il quale accedere al server centrale:

- possibilità di abilitare e disabilitare delle caselle di posta
- accesso ai log in lettura
- accesso alle code in lettura
- possibilità di modifica le quote
- accesso al backup per fare un restore
- accesso ai programmi di filtraggio per l'utente (es. .procmailrc o script sieve), se non delegato direttamente all'utente
- gestione delle mailing lists
- segnalazione di spam falsi positivi o falsi negativi
- accesso diretto a qualcuno che risponda alle domande se non se ne viene fuori da soli.

Da questo punto di vista il carico di lavoro per un servizio di helpdesk nelle Sezioni cambia poco, cambiano invece le procedure.

Nel sito centrale, oltre a mantenere e gestire i servizi implementati nei server centrali, i compiti dell'help desk saranno:

- rispondere ai problemi che le sedi non sanno o non possono risolvere; si potrà decidere se dare la possibilità all'utente di contattare direttamente il servizio centrale o se può essere più efficiente che sia il servizio locale a dialogare con il servizio centrale. È anche importante che chi risponde possa intervenire direttamente sul problema. Potrebbe essere necessario implementare un servizio a più livelli di gravità: per esempio problemi non urgenti o semplici notifiche o richieste di informazioni da fare via mail o via web e problemi seri che richiedono una risposta telefonica.
- rispondere ai problemi delle procedure di gestione o controllo messe a disposizione delle Sezioni. Questa è una cosa nuova perché si è aggiunto un gradino intermedio.
- fornire una documentazione aggiornata perché spesso è necessario soddisfare anche le curiosità degli utenti.
- fornire delle statistiche.

Parte 2.

Tecnologie software e hardware

In questa seconda parte saranno descritti alcuni sistemi commerciali e opensource che potrebbero essere adottati per la gestione di un sistema di posta costituito da un grande numero di caselle con quote rilevanti.

I sistemi completi di posta elettronica in commercio presi in considerazione sono:

- Oracle Collaboration Suite
- Google Apps
- Apple Mail Server

Sono tutti sistemi che integrano i servizi di posta elettronica principali (IMAP, POP, SMTP) e una serie di differenti servizi accessori.

Altri sistemi di cui non sono state approfondite le caratteristiche sono Sun Java Messaging Server, IBM Lotus Domino e Microsoft Exchange.

Oracle Collaboration Suite

La suite di collaborazione è composta da un backend Oracle DB e una serie di application server modulari che comprendono:

- Posta elettronica
- Instant messaging
- Voice Mail e Fax
- Conferencing
- Content management distribuito

L'application server che gestisce la posta elettronica è accessibile via IMAP, IMAPS, POP3, POP3S, webmail. Contiene un antispam e un antivirus proprietario ma può essere integrato con Norton Antivirus.
Permette la configurazione personalizzata di forward e archiviazione automatica ed è integrabile con LDAP.

Tutto il sistema Oracle viene installato in HA e load balancing. Anche il backup viene eseguito attraverso i servizi nativi di Oracle.

Costo:
Ogni casella di posta costa 33 €/anno, più 8 € per l’assistenza.
Il costo totale per 12500 mailbox supera i 500.000 € l’anno, ma potrebbero essere contrattati degli sconti.

Google Apps

L'offerta di Google[4] comprende:

- Gmail
- Google Talk
- Google Calendar
- Google Documenti
- Page Creator (per pagine web)

Il servizio Google Apps è offerto in tre versioni[5]:

Google Apps Educational	gratuito, per scuole e ONLUS
Google Apps Standard	gratuito, con annunci pubblicitari, con alcune limitazioni
Google Apps Premier	40 €/account/anno

Per le scuole il servizio è gratuito, ogni account (posta e documenti) ha una quota di 5 GB, non sono inseriti messaggi pubblicitari nelle pagine web.

Per le aziende, è disponibile l'opzione Premier, che garantisce una continuità di servizio del 99.9%, una quota di 25 GB e la possibilità di recuperare i messaggi (attiva da novembre 2007).

Un account Google Apps è diverso dagli account gratuiti Google (es. utente@gmail.com): mentre il primo consente l'accesso a servizi gestiti da un'amministrazione cliente di Google, il secondo permette l'accesso ai servizi gratuiti gestiti da Google.

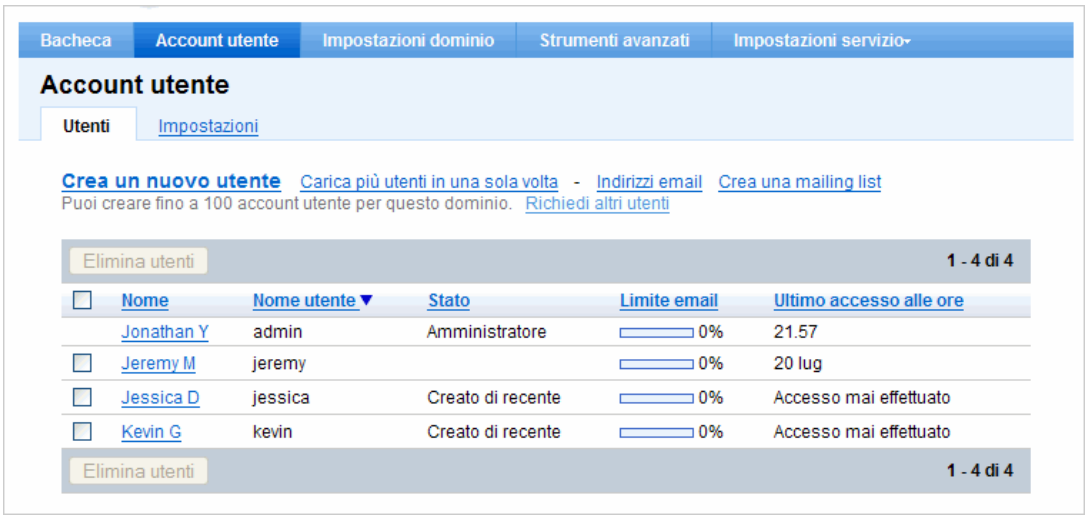
Informazioni tecniche.
I server di Gmail sono contattabili attraverso POP, IMAP e webmail. Tramite un meccanismo di API è anche disponibile l'accesso con Single Sign On.

Antivirus ed antispam sono presenti ma non personalizzabili. E' possibile però la configurazione di whitelist; sono previste funzioni di forward e archiviazione automatica.
Google mette a disposizione uno strumento di migrazione da server Cyrus IMAP che consente di trasferire in modo controllato i messaggi dalle caselle attuali.
E' possibile impostare un servizio di posta mantenendo anche il dominio già in possesso dell'utente, o acquistandone uno nuovo.

Il servizio di assistenza risponde solo agli amministratori e solo per problemi di gestione (es. non risponde di messaggi non ricevuti).

I clienti di Google Apps amministrano il proprio dominio di posta tramite interfaccia web:

[4] http://www.google.com/a/help/intl/it/admins/editions_spe.html
[5] http://www.google.com/a/help/intl/it/admins/editions.html



Costi
Il costo di ogni mailbox è di 40-50 €/anno, secondo l'offerta del momento.
Il prezzo totale per le attuali 12.500 mailbox sarebbe quindi di 500.000 €/anno.
Non si conosce la possibilità di eventuali sconti o della possibilità di aderire alle offerte educational.

Sistemi opensource

Linux

Esistono alcuni sistemi operativi e applicativi opensource che potrebbero riuscire a fornire un servizio di posta elettronica completo, ospitando le caselle di posta di 10.000/20.000 utenti garantendo il livello di servizio attuale. Sistemi di questo tipo sono adottati da alcune Università italiane e da altre nel mondo.

Il sistema operativo più utilizzato è linux, mentre per la scelta del software SMTP/MX quasi tutti si orientano su sendmail.
Per la gestione delle caselle di posta l'unico sistema opensource che è dichiaratamente scalabile, grazie anche alla architettura gerarchica introdotta con la nuova versione, è Cyrus IMAP, della CMU University. I server di frontend gestiscono le comunicazioni fra i client e i server di backend, dove sono fisicamente localizzate le mailbox, mentre le informazioni sulle diverse mailbox e sulle credenziali d'accesso sono mantenute da un terzo server (o sue repliche in una gerarchia master/slave).

Per una migliore efficienza di accesso alle caselle di posta, inoltre, è raccomandata la scelta di un filesystem di tipo indicizzato.

Il mondo opensource offre una vasta scelta di applicazioni anche per i servizi accessori. Fra le più utilizzate ci sono proprio le suite che le sedi INFN meglio conoscono:

- Webmail: Horde, squirrelmail
- Antispam: spamassassin
- Mailing-list: majordomo, mailman

Per quanto riguarda gli antivirus, le poche versioni opensource di antivirus per mailserver sembrano invece inferiori in qualità e prestazioni rispetto ai prodotti commerciali.

Mentre la scelta degli applicativi software sembra quasi obbligata, la progettazione di un sistema hardware affidabile e ridondato risulta più complessa, così come la successiva gestione sistemistica, la manutenzione e il servizio di assistenza agli utenti che richiederanno personale dedicato e preparato.

A fronte di un investimento unico legato all'acquisto iniziale delle macchine, un sistema opensource richiede quindi un forte impegno da parte dell'organizzazione che lo sceglie per destinare e formare personale competente alla gestione di questo servizio.

Apple Mail Server

Apple MAC OS X server 10.4 include un set completo di applicazioni per la gestione della posta elettronica. Tutte il software utilizzato è opensource (Cyrus, Postfix, SquirrelMail, Mailman, Spamassassin, ClamaV). Dispone di interfacce grafiche di configurazione proprietarie.
Il sistema è certificato per oltre 100.000 mailbox.

Il sistema comprende tutti i servizi principali (IMAP, IMAPS, POP3, POP3S, SMTP), mailing list, webmail, antispam e antivirus. E' in teoria possibile integrare le applicazioni installate con altre visto che tutto l'ambiente è opensorce.

Il costo della licenza client unlimited di MAC OS X Server 10.4 è di 999 €, a cui bisogna poi aggiungere il costo relativo alle macchine che dovranno essere installate ma che non dovrebbe superare i 10.000 €.

La gestione sistemistica del sistema Apple Mail Server e la manutenzione del servizio sarebbero a carico di un servizio centrale dell'INFN.

PARTE III

Soluzioni adottate da altri enti e Università

CSITA (Centro Servizi Informatici e Telematici di Ateneo, Università di Genova)

Rappresenta un esempio di sistema "ibrido", che utilizza strumenti commerciali e e di pubblico dominio. Il numero di caselle di posta gestite è di 4000 per i dipendenti e 15000 per gli studenti.
L'accesso alle caselle può avvenire tramite IMAP e POP, e raggiunge i 15000 contatti ogni ora; per confronto si pensi che gli accessi al sistema di posta della Sezione INFN di Genova è di 271 accessi/ora.

Software:

SMTP/MTA: Postfix + LDAP (installazione eseguita da Sophos)
Filtro antivirus/antispam: Sophos Pure Message in configurazione "server group" per il load balancing.
Mailbox Server: Cyrus IMAP + IMAP Proxy (Perdition) + LDAP
Webmail: HORDE/IMP

Hardware

4 Macchine per Sophos/Postfix
3 Macchine per Cyrus+LDAP
1 Storage Clarion FC (120 GB per personale + 140 GB per gli studenti)

Descrizione del sistema

Due macchine sono dedicate al SMTP server e due all'MX server. Sugli MX è installato Sophos Pure Message+Postfix in configurazione di ridondanza e load balancing (server group). I messaggi in ingresso vengono passati al proxy IMAP, installato per ridondanza su tutte e tre le macchine che contengono le mailbox. Il proxy, attraverso un backend LDAP, contatta il server che contiene la casella di posta a cui recapitare il messaggio.
La grande maggioranza degli utenti contatta il server attraverso il protocollo POP3 anche se IMAP è disponibile.
Ciò è probabilmente dovuto alla esigua quota di storage dedicata ad ogni casella (100 MB).

Load balancing

Il load balancing è assicurato dai meccanismi di round robin del DNS sui due MX record e dal proxy IMAP

Alta affidabilità

Per quanto riguarda SMTP ed MX l'alta affidabilità è garantita dai meccanismi del protocollo (anche se va detto che in caso di rottura di uno dei server la posta in uscita andrebbe "a singhiozzo") mentre se uno degli IMAP server dovesse rendersi indisponibile sarebbe necessario l'intervento umano per riconfigurare una delle macchine rimanenti.

Backup

Non sono previsti meccanismi di backup. Ogni utente deve provvedere da se'.

Università di Napoli - CSI (Centro Servizi Informatici)

Il Centri Servizi Informatici dell'Università di Napoli gestisce centralmente cirrca 20000 caselle di posta.
Descrizione del servizio:

- autenticazione tramite un server LDAP (un biprocessore Xeon dedicato)
- 3 server biprocessori Xeon per l'MX
- Gestione delle mailbox tramite un biprocessore connesso in SAN (FC su macchine HP, vari TB di disco SATA II in box RAID 6 connesso con ridondanza a FC);
- software Cyrus IMAP;
- backup su nastro, con cadenza settimanale e dati conservati per poche rotazioni del set di nastri (cioè per poche settimane).
- Tutti i server sono HP Proliant, con Red Hat licenziato
- filtro anti virus e antispam Pure Message di Sophos sui 3 server MX.

IN2P3

Il Computing Centre a Lyon ospita le mailbox per parecchia gente che lavora nei 19 laboratori IN2P3; la maggior parte dei laboratory usa il servizio antispam e antivirus fornito dal Computing Centre, indipendentemente dalla dislocazione fisica delle mailbox.

CESIA, Università di Bologna

Il Cesia fornisce un servizio centralizzato costituito da 16000 mailbox per i soli dipendenti dell'Università. Esistono comunque ancora parecchi mailserver nei vari istituti. Il traffico generato su queste mailbox è ridotto e corrisponde circa ad una metà reale.
Il servizio di mail per gli studenti è invece fornito dal Cineca.

L'infrastruttura è basata su Microsoft Exchange, estremamente funzionale come CMS solo se utilizzato con altri prodotti di Microsoft come Outlook ecc. Permette la condivisione dell' agenda, prenotazioni di sale, ecc.

Le caselle di posta sono ripartite su 5 dischi e i messaggi sono mantenuti per 15 giorni; il Cesia segnala grossi problemi per la deframmentazione.

Il sistema complessivo risulta molto oneroso: licenza Exchange, CAL per ognuna delle mailbox, hardware infrastrutturale.
Un altro costo necessario è dato dalla consulenza di tecnici Microsoft per la pianificazione e l'implementazione della struttura.
Il Cesia disponde di personale dedicato per la gestione del servizio e l'assistenza agli utenti.

CERN

Il servizio di posta del CERN[6] serve più di 18.000 caselle di posta, ognuna di 80 MB in media, con una quota massima di 3 GB
L'infrastruttura è basata su Microsoft Windows Server 2003 server e Microsoft Exchange Server 2003. Il software antivirus è Symantec Antivirus per Exchange.

Per il servizio IMAP sono installati 14 server: Elonex rack-mounted 4U servers, Dual Xeon 2.0Ghz, Hyperthreading on, 3-4 GB memory, 2 SRCU32 Intel RAID controllers ognuno con 1xRAID1 (2x70GB/SCI), 1xRAID5 (3x120GB/SCSI), 1Gbit/s network card

Come SMTP gateway sono installati 6 server, sui quali sono installati Windows Servers 2003, Exchange 2003, Windows Load Balancing, Symantec Antivirus for Exchange e CERN made C# Protocol Event Sink. L'hardware è costituito da server Elonex rack-mounted 2U, Dual Xeon 2.0Ghz, Hyperthreading on, memoria 2GB, RAID controller 3Ware 7506 series con 2xRAID1 (2x40GB/IDE + 2x120GB/IDE), 1Gbit/s network card.

Altri 4 server dello stesso tipo implementano il software CERN SpamKiller (CERN made C# Windows Service) su Windows Servers 2003.

Infine, 4 server fungono da Frontend HTTP e LDAP (Microsoft ADAM) e forniscono accesso IMAP, POP e HTTP alle caselle di posta. L'hardware è costituito da server Elonex rack-mounted 2U, Dual Xeon 2.0Ghz, Hyperthreading on, memoria 2GB, RAID controller 3Ware 7506 series con 2xRAID1 (2x40GB/IDE + 2x120GB/IDE), 1Gbit/s network card.

[6] https://mmmservices.web.cern.ch/mmmservices/

Il traffico di posta elettronica è 90.000 mail ham su 2.900.000 messaggi in ingresso ogni giorno (il 3%), e 37.000 messaggi in uscita. I metodi d'accesso più utilizzati sono IMAP, Outlook Web Access, Outlook XP/2003 MAPI e POP (in dismissione).

Carnegie Mellon University e University of Michigan

La Carnegie Mellon University è autore del software cyrus IMAP e tutto il servizio di posta che fornisce a dipendenti e studenti è basato su cyrus.
Il numero di caselle di posta gestite è circa 20.000 (inbox), mentre il numero di mailbox è dieci volte superiore. La quota massima di ogni utente è di 2 GB. La University of Michigan gestisce un numero di caselle ancora superiore utilizzando le medesime tecnologie.

L'architettura del sistema della CMU è la seguente[7]:
- 5 server di frontend che gestiscono gli accessi alle caselle di posta; l'hardware è tutto Sun con sistema operativo Solaris, tre server sono Sun Ultra 80s biprocessori UltraSparcII, due sono SunFire 280Rs biprocessori UltraSparcIII
- 5 server di Backend che ospitano fisicamente le caselle di posta; quattro server Sun 220R con storage su JetStor II-LVD SCSI RAID arrays e un Sun 280R con storage JetStor III U160 SCSI RAID array
- Un server Dell 2450 per il ruolo di M.Update master
- 8 server Dell 2650 per il ruolo di SMTP/MX suddivisi fra tre domini principali: andrew.cmu.edu, smtp.andrew.cmu.edu e cmu.edu, gestiti da sendmail
- Un server Dell 2650 per gestire le mailing list tramite majordomo
- 3 server per webmail, di tipo Dell Optiplex GX260

Tutti i server Sun montano il sistema operativo Solaris, mentre il server Dell montano Linux.

Un recente aggiornamento ha portato il numero di server di backend a 10, il numero di MX server a 9 e il numero di webmail server a 12. E' stato inoltre installato un sistema di filtri antispam/antivirus su 8 macchine (7 macchine filtro, più un database). Tutto lo spazio mailbox è dotato di backup.

L'accesso alle cartelle di posta avviene principalmente tramite IMAP o via web: i client supportati sono Microsoft Outlook, Entourage (per piattaforme Mac), pine e Andrew Webmail, sviluppato dalla stessa CMU.
Per la gestione delle mailing list è stato introdotto Andrew mailman, una versione di mailman personalizzata dalla CMU, mentre per invii di posta massicci è implementato MassMail.
Il servizio di supporto agli utenti è disponibile dalle 9 alle 17 (via telefono fino alle 19) e impiega anche studenti part-time.

[7] http://cyrusimap.web.cmu.edu/configuration.html#hardware

Quarta parte.

Gli scenari possibili

Partendo dalle informazioni riportate nelle parti precedenti, in questa quarta parte si cercherà di descrivere qualche esempio di configurazione che permetta di fornire all'utenza INFN un servizio di posta paragonabile a quello attuale, oggi erogato da ogni sede con modalità simili.
Il servizio ideale dovrà quindi permettere l'accesso IMAP/IMAPS e l'accesso via web, dovrà prevedere la possibilità di autenticazione degli utenti tramite l'infrastruttura nazionale AAI, l'implementazione e la personalizzazione di filtri antispam e antivirus, la gestione delle quote, la conservazione dei domini infn.it e *sede*.infn.it, per un numero di caselle di posta pari a 5000 (se si considerano i soli dipendenti e associati) o 20000.

1. Outsourcing esterno (Tipo Google Apps)
Tutto il sistema di posta è ospitato e gestito da un fornitore di servizi, in casa non abbiamo più niente.

Vantaggi:
- E' possibile mantenere il dominio, i sottodomini, gli attuali indirizzi di posta
- Non abbiamo più hardware nostro
- Niente spese di corrente, raffreddamento, manutenzione
- Il personale non si occupa della gestione sistemistica ma solo della gestione degli account (ad es. tramite interfaccia web)
- Supporto utenti ridotto alla configurazione del client o all'intermediazione con il supporto del fornitore di servizi.

Svantaggi:
- Interazione con una gestione esterna in caso di problemi, fatta direttamente dall'utente (non prevista da Google) o da un amministratore (uno per tutto l'Ente, o uno per sede)
- Inevitabili ritardi nella soluzione dei problemi degli utenti, soprattutto se ci sono più persone coinvolte a vari livelli.
- Informazioni di importanza minore (è partito il mio mail?) difficilmente reperibili.
- Costo per casella a cui fare fronte ogni anno con qualsiasi Finanziaria.
- Problema degli ospiti (dipartimento, studenti, non associati)
- Il sistema risiede su una rete diversa dalla nostra: ogni messaggio, anche destinato al vicino d'ufficio, è soggetto ai problemi della rete geografica

Non sappiamo se sono possibili:
- Pianificazione dell'autenticazione, vogliamo mantenere i nostri database di autenticazione e i certificati
- Gestione delle mailing list, vogliamo continuare ad averle
- Rispetto delle norme sulla privacy: non sappiamo se il fornitore del servizio è soggetto alla normativa italiana.

2. Outsourcing interno (Tipo Cern)

L'hardware è ospitato in una sede, quindi è interno alla nostra rete, ma viene gestito da personale esterno. Non si pone il problema dei domini, dovrebbe essere possibile l'integrazione con un servizio di autenticazione nazionale o con il meccanismo dei certificati, non si vedono particolari problemi per le mailing lists.

Vantaggi:
- Sistema garantito per hardware e software
- Niente personale per la gestione
- Supporto utenti ridotto alla configurazione del client

Svantaggi:
- Gli stessi del punto 1

3. Gestione interna prodotto proprietario (Tipo Oracle Collaboration Suite)

Si compera il software e si installa in una sede. Il software deve ovviamente offrire i servizi che riteniamo utili.

Vantaggi:
- Software garantito
- Sezioni sollevate dalla gestione del mail server

Svantaggi:
- Una sede dovrà mettere a disposizione più personale di quello che le servirebbe per gestire il proprio mailserver. L'amministrazione fatta remotamente da persone di altre sedi può essere presa in considerazione, ma può comportare seri problemi di organizzazione e di sincronizzazione.
- Potrebbe servire anche una sede di backup, ma non è di facile implementazione.
- Costo del software, dell'hardware e delle manutenzioni, ecc.
- Formazione di personale specializzato
- Istituzione di un helpdesk a carico di personale INFN
- Ritardi nella soluzione dei problemi del singolo utente
- Costo per casella a cui fare fronte ogni anno con qualsiasi Finanziaria.
- Problema degli ospiti (dipartimento, studenti, non associati)

4. Gestione interna open source

In questo scenario si immagina di organizzare un servizio centrale prendendo a modello i sistemi implementati nelle sezioni.
Un problema importante sarà la corretta progettazione hardware.

Vantaggi:

- La gestione è più aderente alle politiche dell'Ente; in particolare si potrà decidere autonomamente la politica di servizio da concedere a dipendenti, associati e ospiti
- Uniformità nelle politiche antivirus, antispam, liste bianche - liste grigie
- Il sistema potrà essere direttamente integrato con l'infrastruttura di autenticazione
- Il software scelto è in gran parte gratuito
- La maggior parte delle Sezioni sarà sollevata dalla gestione del servizio di posta
- Nessun costo fisso per le mailbox
- Parziale precedente conoscenza dei prodotti opensource,
- Possibilità di sviluppare strumenti per suddividere l'attività di helpdesk tra servizio centrale e sedi

Svantaggi:
- Una sede dovrà mettere a disposizione più personale di quello che le servirebbe per gestire il proprio mail server, e uno spazio opportunamente servito da impianti elettrici e di condizionamento;
- Sarebbe opportuno prevedere anche una sede di backup
- Costo dell'hardware e della sua manutenzione
- Manutenzione del software impegnativa, trattandosi di open source che non offre molto supporto
- Istituzione di un helpdesk a carico del personale INFN
- Ritardi nella soluzione dei problemi del singolo utente
- Il personale delle sedi deve comunque poter intervenire sui server centrali per un certo numero di operazioni, ma non gestendo direttamente il sistema rischia di perdere le competenze tecniche necessarie.

Un esempio di sistema opensource gestibile centralmente dall'Ente.

Configurazione hardware

Un sistema basato su Scientific Linux 5.x, configurazione hardware minima per poter gestire qualche migliaio di connessioni contemporanee:
- almeno 2 server di frontend di ultima generazione (dual quad core a 3.0 GHz e 16GB di RAM)
- almeno 2 server per storage di backend per scrivere su disco (dual quad core a 2.4 GHz e 16GB di RAM) con:
  - Storage fiber channel EMC2 CX380 (285 dischi da 500GB, circa 120TB)
  - Switch FC Brocade 4Gbit + scheda FC bicanale QLogic 4Gbit (40.000 €)
- almeno 2 server per la gestione di webmail (dual quad core a 2.4 GHz e 16GB RAM).

Il costo indicativo di un server è 6.000 €. Sarà anche importante attivare contratti di manutenzione per tutto l'hardware, con intervento e risoluzione dei guasti entro 4 ore.

Configurazione software:
- S.O. Scientific Linux 5.x
- Sendmail o Postfix SMTP server
- cyrus imapd+cyrus sasl+cyrus murder per l'implementazione di IMAP
- Horde webmail suite o squirrelmail

- sophos pure message o amavis+sophos antivirus come sistema di antivirus
- Spamassassin o Sophos per l'antispam
- Vale la pena di valutare GPFS come file system per le mailbox su FC

Risorse umane necessarie alla gestione del servizio:
Un sistema centralizzato gestito interamente dall'INFN con software opensource dovrebbe prevedere almeno queste figure professionali:

1) N. 1 esperto dei sistemi di mailing, ma anche di storage e rete che faccia da coordinatore e responsabile del servizio – 1 FTE
2) N.3 esperti di mailing su linux che si occupano dell'installazione, configurazione, mantenimento ed aggiornamento del sistema (mailserver, antivirus, antispam, gestione mailbox, etc.) – 3 FTE
3) N. 4 addetti all'help desk e gestione di base del servizio – 4 FTE

In totale potrebbero servire **almeno** 8 FTE per avere un servizio attivo per 5 giorni la settimana, 12 ore al giorno, con interventi di reperibilità per guasti gravi durante i fine settimana.

5. Gestione locale.

Questa è la situazione attuale in cui ogni sezione possiede e gestisce un proprio servizio di posta elettronica.

Vantaggi:
- Gestione più aderente alle esigenze della Sezione.
- Rapida soluzione dei problemi
- Uso di hardware più economico
- Sviluppo e mantenimento di conoscenze tecniche acquisite
- Minimizzazione del traffico nel caso di allarmistica tramite e-mail

Svantaggi:
- Più persone impegnate in lavori simili nelle varie sedi
- Implementazioni diverse delle politiche antispam

Conclusione

Una soluzione completamente in outsourcing è decisamente onerosa ma può offrire all'utente una qualità di servizio paragonabile a quella attuale. Non è ancora chiaro se tutte le richieste possano essere soddisfatte dalle soluzioni commerciali prese in esame: nessun fornitore ha descritto in dettaglio le tecnologie di autenticazione implementate né la possibilità di integrazione con sistemi di autenticazione diversi.
Alcune fra le soluzioni descritte libererebbero però completamente i Servizi dalla gestione e dalla manutenzione dell'infrastruttura e dalla gestione sistemistica.

Le soluzioni proprietarie installate in sede INFN sarebbero altrettanto costose (almeno 500.000 € annui) e richiederebbero non solo la formazione di competenze, ma anche la sottoscrizione di contratti di consulenza specialistica sul prodotto proprietario acquistato (es Microsoft Exchange). Fa eccezione Apple con un costo di licenza forfettario, al quale si aggiunge l'acquisto di hardware Apple.

La soluzione centralizzata con software opensource, progettata e gestita dall'INFN, sarebbe in pratica l'estensione del servizio fornito da una delle attuali sedi, implementata su di un sistema ad alta affidabilità e ridondato; il modello attuale di servizio così come è implementato in ogni sede è però solo parzialmente valido e riutilizzabile: mentre dal punto di vista software alcuni fra i sistemi potrebbero scalare fino alle dimensioni richieste, il modello hardware deve essere completamente riprogettato, così come deve essere ridisegnato il servizio di helpdesk.

Un altro aspetto da considerare è che buona parte del traffico legato alla posta elettronica è in questo momento *locale*, cioè avviene fra utenti e sistemi della stessa sede; alcuni sensori, ad esempio, cominicano le proprie transizioni di stato e gli allarmi via email. Oggi questo tipo di traffico traffico è molto veloce anche perché è agevolato da filtri più blandi (o dalla completa assenza di filtri) rispetto a quelli applicati a messaggi provenienti dall'esterno, non è soggetto a greylisting e non risente dei problemi di rete su distanza geografica. Caratteristiche analoghe ha il sistema di allarmistica via email utilizzata in alcune delle sedi che ospitano servizi in produzione di rilevanza internazionale come il TIER1 e i TIER2.
Un servizio centralizzato potrà garantire l'inoltro rapido ed efficiente di messaggi di questo tipo?

Scopo di questo studio era verificare la possibilità si alleggerire il carico di lavoro richiesto dalla gestione di un servizio di posta replicato in ogni sede. Alcune sedi hanno infatti manifestato difficoltà nel reperire le risorse umane necessarie a gestire un servizio tanto critico; non tutte le sedi però si sono mostrate insteressate a delegare la posta ad un servizio centrale.
Si potrebbe quindi prendere in considerazione anche un servizio solo parzialmente centralizzato, dedicato inizialmente alle sedi in difficoltà, che potrebbe costituire il banco di prova per l'eventuale riorganizzazione.

Frascati, 11 Dicembre 2007