ISTITUTO NAZIONALE DI FISICA NUCLEARE

INFN-13-02/CCR
20th march 2013

ANALISI PRELIMINARE SULL'EVOLUZIONE DEL SISTEMA DI MAILING DELL'INFN

Riccardo Veraldi[1], Alessandro Brunengo[2], Alberto Gianoli[3], Michele Michelotto[4], Marco Serra[5], Alessandro Tirel[6]

[1)]INFN-CNAF, Viale Berti Pichat 6/2, I-40127 Bologna, Italy
[2)]INFN-Sezione di Genova, Via Dodecaneso 33, I-16146 Genova, Italy
[3)]INFN-Sezione di Ferrara, via Saragat 1, I-44122 Ferrara, Italy
[4)]INFN-Sezione di Padova, Via F. Marzolo, 8, I-35131 Padova, Italy
[5)]INFN-Sezione di Roma, P.le A. Moro 5, I-00185 Roma, Italy
[6)]INFN-Sezione di Trieste, Via A. Valerio 2, I-34127 Trieste, Italy

Abstract

Questo documento fotografa l'attuale situazione del servizio di posta elettronica presente nelle sedi INFN e analizza dei possibili nuovi scenari che possano ottimizzare l'utilizzo delle risorse, sia in termini di hardware impiegato sia per il numero di addetti (FTE).

Commissione Calcolo e Reti

CCR-44/2012/P

*Published by **SIDS–Pubblicazioni***
Laboratori Nazionali di Frascati

CONTENUTI

1 INTRODUZIONE

Questo studio è stato richiesto al gruppo Mailing della Commissione Calcolo e Reti al fine di rivedere e proporre eventuali scenari alternativi all'attuale struttura decentrata nelle singole sedi INFN. L'intento è di perseguire, se possibile, un'ottimizzazione delle risorse umane e infrastrutturali per ottenere una riduzione della spesa. Se da un lato uno scenario completamente centralizzato del servizio di posta elettronica potrebbe alleviare il carico di lavoro ai servizi calcolo delle singole Sezioni, di contro ci sono alcuni svantaggi che vanno considerati.

Il manpower va diviso tra la gestione della parte server (centralizzata o esternalizzata), il supporto agli utenti (helpdesk) e la configurazione dei client (Thunderbird, Outlook,...), compito quest'ultimo che rimarrà sempre e comunque a carico dei servizi calcolo locali. Inoltre, un supporto in sede nei confronti degli utenti andrebbe comunque mantenuto e quindi lo sgravio per la gestione del servizio sarebbe soltanto parziale.

Come si evince da questo studio, un numero elevato di utenti non fa parte del personale dipendente né di quello associato. Questo avviene perché molte sedi INFN hanno in essere convenzioni con i Dipartimenti di Fisica, talvolta con servizi calcolo in comune e devono quindi fornire la posta elettronica anche a utenti che non hanno alcun legame contrattuale con l'INFN. Inoltre, ci sono sedi che ospitano collaboratori esterni che pur non avendo un rapporto di lavoro o un incarico di collaborazione con l'Ente, sono parte integrante delle collaborazioni scientifiche. Per questi ultimi un account di posta INFN è funzionale a migliorare lo scambio d'informazioni all'interno della collaborazione.

Per avere, quindi, una traccia definitiva sulla quale lavorare al termine di questo studio preliminare, occorre capire come l'Ente voglia affrontare questi specifici problemi.

In conclusione, il presente documento si prefigge di analizzare il servizio di posta elettronica, globalmente, da un punto di vista prettamente qualitativo, senza entrare nei dettagli tecnici d'implementazione, se non come casi esemplificativi al fine di produrre una valutazione indicativa sui costi comparativi. Un successivo studio sugli aspetti tecnico-progettuali ha bisogno d'indicazioni chiare da parte dell'INFN in merito ai rapporti di collaborazione con altri Enti e all'individuazione degli argomenti meritevoli di approfondimento.

Si deve infine notare che l'esigenza di avere indirizzi del tipo nome.cognome@infn.it è indipendente dalla struttura impiegata sia essa distribuita, centralizzata o esternalizzata.

2 CARATTERISTICHE COMUNI AD UN SERVIZIO DI POSTA

Qualunque sia la scelta implementativa per un servizio di posta elettronica esistono delle caratteristiche comuni che devono essere elementi fondamentali del servizio stesso e che sono irrinunciabili per avere un servizio almeno allo stesso livello di quello attuale.

2.1 Modalità di accesso e protocolli

Il primo argomento da affrontare è la scelta dei protocolli da utilizzare per comunicare con il servizio di posta e quali siano le modalità di accesso al servizio da parte degli utenti. Esistono dei protocolli definiti da standard internazionali per i servizi di posta elettronica. Un servizio di mailing deve supportare i protocolli standard ed eventualmente non supportare protocolli ormai deprecati.

Nel caso della posta elettronica è opportuno differenziare tra servizio di accesso alla casella di posta degli utenti (posta in entrata) e il servizio d'invio della posta (posta in uscita). Gli utenti normalmente si collegheranno alla propria casella di posta utilizzando un opportuno software MUA (Mail User Agent) attraverso il quale potranno effettuare sia le operazioni di lettura che di invio dei messaggi.

2.1.1 Posta in entrata

È necessario supportare i protocolli IMAPv4 e POPv3 che di seguito verranno indicati semplicemente con IMAP e POP. Entrambi i protocolli utilizzano delle porte TCP standard, poiché la comunicazione avviene in chiaro su queste porte, è diventata prassi comune l'utilizzo di SSL per implementare una trasmissione cifrata in modo da rendere inintelligibile la comunicazione fra le parti per chiunque tenti di fare un'analisi del traffico tra client e server. Un servizio di posta quindi dovrà utilizzare IMAPS (IMAP/SSL) e POP3S (POP3/SSL) per avere i minimi standard di sicurezza. Dal 1999 la IETF ha definito una nuova estensione chiamata STARTTLS da applicarsi ai protocolli di comunicazione in chiaro in modo da conferire confidenzialità e cifratura alla comunicazione senza l'utilizzo di SSL su una porta TCP dedicata. Quindi anche questa funzionalità è considerata come fondamentale irrinunciabile.

2.1.2 Posta in uscita

Il protocollo supportato è SMTP. Secondo i servizi offerti e i ruoli che il server esegue, il protocollo deve prevedere delle opportune estensioni. Se il servizio in oggetto è di MTA (Mail Transfer Agent), come un semplice MX di Sezione che riceve soltanto la posta dall'esterno, non ci sono altre modifiche da apportare. Nel caso si tratti di un MSA (Mail Submission Agent) ovvero un servizio SMTP che consente agli utenti di inviare i propri messaggi, è necessario che il sistema implementi l'estensione SMTP-AUTH protetta da STARTTLS. In questo modo gli utenti per inviare i propri messaggi utilizzeranno un canale

cifrato all'interno del quale sarà richiesta loro l'autenticazione. Per questo tipo di servizio è stato deciso da IETF di utilizzare una porta specifica dedicata in particolar modo agli utenti roaming, ovverosia la porta 587 TCP (submission). Benché sia deprecata da IETF, esiste la possibilità di inviare i messaggi utilizzando SMTP con SSL. Questa combinazione utilizza la porta 465 TCP associata al servizio smtps. Poiché diversi software di posta elettronica, dispositivi embedded e sistemi operativi di smartphone supportano questa tipologia di accesso, si considera utile mantenere la compatibilità con questo protocollo.

2.1.3 Accesso tramite interfaccia web

È necessario che un servizio di posta sia utilizzabile anche tramite interfaccia web. In questo caso il servizio dovrà essere fornito su canale cifrato tramite l'impiego del protocollo HTTP su SSL (HTTPS, 443/TCP).

2.1.4 Autenticazione

Nella gestione quotidiana di un servizio informatizzato rivolto all'utenza in generale è necessario accedere a una serie d'informazioni che sono legate sia all'utenza stessa sia 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 a informazioni che spesso sono distribuite su più database e server. Ad esempio, l'accesso alle caselle di posta elettronica, eseguito sia attraverso il protocollo POP sia IMAP, deve essere garantito solo dopo che l'utente sia 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'accesso al 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 a usare il servizio, oppure un'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 installati (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, che devono essere replicati su più server in caso di architetture ridondanti realizzate per garantire

alti livelli di affidabilità. Alcuni meccanismi di autenticazione sono ormai obsoleti (NIS) e presentano qualche pecca nella sicurezza. Sempre più sezioni si sono dotate negli ultimi anni di sistemi di autenticazione centralizzati a livello locale, abbandonando l'autenticazione locale al server di posta. Sono quindi utilizzati spesso LDAP, Kerberos V o entrambi, per chi implementa un sistema di autenticazione/autorizzazione completo. Questo tipo di architettura è analoga o simile alla infrastruttura di AAI nazionale che si auspica possa inglobare le autenticazioni in un unico sistema nazionale. Attraverso l'infrastruttura di autenticazione e autorizzazione (AAI) è facilitata la gestione di tutte le informazioni, garantendone la coerenza e gli aggiornamenti, 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 e un'amministrazione puntuale di tutti i database che contengono le informazioni, un sistema che ambisca a offrire un servizio di posta elettronica centrale a tutte le sedi deve necessariamente essere interfacciato alla AAI nazionale.

2.1.5 Servizi accessori

Un sistema di posta elettronica, centralizzato o distribuito, deve implementare necessariamente per motivi di sicurezza un sistema di filtro antivirus; è opportuno che questo sistema includa la gestione dei messaggi filtrati tramite una quarantena, e metta in condizione ciascun utente di accedere in modo autenticato alla propria porzione di quarantena per un eventuale recupero dei propri messaggi bloccati a causa di un falso positivo.

È certamente necessario configurare un sistema d'identificazione dello SPAM, che permetta agli utenti di individuare in modo semplice la posta indesiderata, ed eventualmente archiviarla o eliminarla direttamente; anche in questo caso è opportuna la realizzazione di un sistema che permetta agli utenti, in modo autenticato, di personalizzare la configurazione del filtro anti-spam, in modo da modularne la selettività e di introdurre white-list differenti per utente. In caso di servizio gestito internamente, è essenziale un sistema di monitoraggio e allarmistica, che permetta agli amministratori di identificare l'insorgere di un problema a qualsiasi elemento del servizio in tempi brevi, in modo da migliorarne la fruibilità. Nella stessa ipotesi, e in caso di servizio centralizzato, si ritiene di grande utilità la realizzazione di un sistema di statistica sul funzionamento delle componenti del servizio e sull'utilizzo delle risorse ad esso dedicate, cosicché disponendo di tutte le informazioni necessarie sia possibile definire una strategia di sviluppo dell'intero sistema.

Indipendentemente dalla soluzione adottata si ritiene indispensabile la realizzazione di una documentazione completa, semplicemente fruibile, il più facilmente comprensibile, che permetta all'utente di configurare il proprio client di posta in modo da accedere al servizio, configurare le personalizzazioni e sapere come prendere contatto con il supporto qualora fosse necessario. Una documentazione adeguata permette di ridurre notevolmente il lavoro dell'helpdesk.

Vi sono naturalmente altre caratteristiche utili che, in base alle soluzioni tecniche adottate, si possono rendere disponibili o che potrebbero rendersi necessarie, come un sistema

di gestione della quota a livello di singolo utente. Per dettagliare queste si rimanda a un'analisi tecnica più approfondita.

In questo paragrafo si è voluto porre l'accento su alcune caratteristiche ritenute indispensabili e che hanno un impatto rilevante sull'intero sistema. In particolare sui costi, in termini di personale nel caso il servizio sia gestito internamente; in termini di spesa, nel caso il servizio sia dato in outsourcing.

2.2 Gestione utenti e interfacciamento ad AAI

Uno dei primi argomenti da sviluppare è la definizione di procedure chiare e unificate per la gestione degli account degli utenti (creazione, chiusura, verifica dello stato, ecc.), qualunque sia l'architettura utilizzata per il sistema di posta (locale, centralizzato, decentrato, ecc.). Esistono criteri assodati sulla base dei quali un utente ottiene e mantiene un'utenza di posta elettronica. La richiesta di un account, ovvero l'autorizzazione o l'assenso a procedere con la creazione dell'account passa attraverso un'interfaccia amministrativa fino a giungere al personale tecnico incaricato. Una procedura analoga va definita anche per la rimozione dell'utenza o il suo mantenimento alla data di scadenza. Questo flusso autorizzativo va ricavato da procedure attualmente già in uso, precise, efficaci e collaudate nel tempo.

Se il sistema di posta sarà gestito da personale INFN, appare utile come requisito, al fine di ottimizzare i tempi, la possibilità di interfacciare la procedura di creazione/cancellazione con i database dell'ente che contengono le informazioni sulle singole persone (in alternativa è necessario uno snello protocollo di comunicazione con l'amministrazione). In particolare questo requisito è irrinunciabile se si vuole evolvere nella direzione di un sistema unico utilizzabile da più sedi. In quest'ambito è auspicabile anche l'interfacciamento all'infrastruttura di autenticazione e autorizzazione dell'ente (AAI) per gestire l'accesso al sistema di posta, analogamente a quanto avviene per altri servizi INFN. Questo potrà accadere soltanto quando il servizio AAI sarà completamente dispiegato secondo le specifiche in tutte le sedi. Alternativamente se il sistema di posta sarà esterno all'INFN, andrà comunque stabilito un protocollo di comunicazione con i gestori del sistema, che garantisca il controllo delle utenze con procedure semplici e attivabili con una singola autorizzazione. Per l'ente potrebbe essere utile richiedere di utilizzare i dati già presenti nei propri database, ma questo potrebbe comportare lavoro aggiuntivo per renderli disponibili in modo sicuro a personale esterno.

2.3 Supporto utenti e tempi d'intervento

Qualunque sia la tipologia del servizio di posta le richieste tipiche di supporto da parte degli utenti si possono catalogare in azioni dei seguenti tipi:

- Recuperare una mail cancellata per errore.
- Verificare perché una mail sia stata considerata spam.
- Tracciare il percorso di una mail per capire se il messaggio sia stato inoltrato/ricevuto, ecc.

Gli utenti INFN ormai considerano consolidato questo tipo di supporto e nel caso il sistema di posta sia gestito dall'INFN è ragionevole pensare che le richieste non si discostino da questo standard ora offerto.

Riguardo ai tempi d'intervento per qualsiasi tipologia e architettura di mailing si prenda in considerazione che il requisito minimo per i nostri utenti è sicuramente quello di avere un supporto 10x5 (una decina di ore al giorno nei giorni lavorativi). E tipicamente, nelle fasce orarie lavorative, l'aspettativa d'intervento per la risoluzione di un problema serio ed urgente, è quella di avere una risposta nel giro di qualche ora.

Tutti gli utenti INFN sono abituati a interagire con i sistemisti come minimo via e-mail, e quando questo non è possibile perché il sistema stesso ad esempio potrebbe avere qualche problema, subentra l'interazione telefonica con il servizio calcolo. Qualora si passi a un sistema centralizzato non nella propria sede, andrà necessariamente previsto un sistema di helpdesk adeguato che offra una comunicazione con gli utenti via e-mail o ticketing via web. È necessario che venga richiesto anche un servizio basato su chiamata telefonica dal momento che se l'utente apre un ticket potrebbe avere un problema proprio nell'invio della posta elettronica.

Non va poi trascurato il supporto che riguarda la configurazione dei client di posta. Molto spesso è richiesto direttamente dagli utenti al personale dei centri di calcolo, che interviene manualmente sui computer degli utenti stessi.

Un altro servizio spesso richiesto è quello di tracciamento della posta inviata o attesa. L'utente chiede se un mail è stato effettivamente consegnato, se un mail è stato ricevuto, se è bloccato da antivirus o antispam o grey list.

2.4 Backup e tempi di ripristino

È opinione condivisa che il backup delle mailbox sia un requisito fondamentale per qualsiasi sistema di posta, sia esso locale o centralizzato.

Le norme ritenute di livello minimo sono:

- frequenza: giornaliera (incrementale) e settimanale (totale).
- conservazione: almeno un mese.

Il media su cui salvare i backup può essere costituito anche da hard disk a patto che siano protetti da una qualche forma di ridondanza (es.:RAID). Un sistema a nastri, in genere, può fornire una maggiore affidabilità del media e tempi di conservazione più elevati, tuttavia, per ottimizzare il flusso dati, dovrebbe essere implementata una policy di staging su disco che permetta nella maggior parte dei casi un rapido restore ed eviti che il sistema di backup utilizzi i nastri in modalità start/stop. Nel caso si impieghino storage appliance di buona qualità o file system "evoluti" è possibile utilizzare la funzione di snapshot per offrire agli utenti un ulteriore livello di protezione del dato. In caso di necessità, l'operazione di restore dovrebbe avvenire in tempi brevi, al massimo, qualche ora, compatibilmente con l'orario di servizio degli addetti.

2.5 Hardware

L'hardware per la gestione del servizio mail non è costituito solo dai server IMAP e dai server per la ricezione e l'invio della posta. Ci sono anche lo storage delle mailbox e l'hardware per il backup. Tuttavia in caso di centralizzazione del servizio mail parte di questo hardware è necessario per altri servizi locali come le home directories degli utenti e le pagine web personali.

2.6 Il dominio di posta infn.it

L'esigenza già sollevata in passato di avere per gli utenti INFN una casella di posta del tipo nome.cognome@infn.it non va confusa con il problema della posta elettronica in generale. Un eventuale servizio di mail centralizzato deve continuare a gestire tutti gli attuali domini di sede (nome.cognome@*sede*.infn.it) poiché gli utenti hanno distribuito questo tipo d'indirizzo ai loro interlocutori; inoltre, i medesimi permettono di autenticare gli utenti presso strutture esterne all'INFN.

Per contro anche con l'attuale gestione distribuita sarebbe possibile con un minimo sforzo dare agli utenti INFN un indirizzo @infn.it semplicemente creando un database degli alias (userdb) popolato per esempio con il database degli utenti AAI associando a ognuno una maildrop locale.

3 IL SERVIZIO ATTUALE

La panoramica sul servizio in essere (settembre 2012) si basa sull'analisi effettuata prendendo in considerazione le diciannove sezioni principali, i quattro laboratori, il centro nazionale CNAF e la Presidenza INFN. La seguente tabella descrive le dimensioni del servizio prendendo in considerazione il numero di mailbox divise per dipendenti, associati e altri. Per altri s'intendono utenti che non hanno rapporti ufficiali con l'INFN: studenti non associati, collaboratori, ecc.

TAB. 1: Distribuzione numerica dell'utenza

Utenza	Numero di mailbox	Media per sede
Dipendenti	2265	87
Associati	3825	147
Altri	3609	138
Totale	10266	394

Nella tabella 2 è rappresentata la situazione delle singole sedi relativamente al numero di account di posta che si riferiscono ad utenti non INFN e non associati e se esiste una convenzione con i dipartimenti per l'erogazione di servizi da parte dell'INFN in generale. In nessun caso esiste una convenzione che esplicitamente citi il servizio di posta.

TAB. 2: Convenzioni

Sede	Mailbox NON INFN	Convenzioni
Bari	SI	SI
Bologna	SI	SI
Cagliari	NO	NO
Catania	SI	SI
CNAF	SI	EXT
Ferrara	SI	SI
Firenze	SI	SI
Genova	SI	NO
LNF	NO	NO
LNGS	NO	EXT
LNL	NO	NO
LNS	SI	NO
Milano	SI	SI
Milano Bicocca	SI	SI
Napoli	SI	SI
Padova	SI	SI
Parma	SI	SI

Pavia	SI	SI
Perugia	SI	SI
Pisa	NO	NO
Presidenza	NO	NO
Roma 1	SI	SI
Roma 2	SI	SI
Roma 3	NO	NO
Torino	SI	NO
Trieste	SI	SI

Nella maggioranza dei casi le sezioni forniscono il servizio di posta elettronica a personale non INFN e non associato all'Ente, di solito, afferente al Dipartimento di Fisica in virtù di convenzioni stipulate tra Sezioni e Università. In alcuni casi, indicati in tabella con la dicitura EXT, il servizio è fornito anche ad altri enti di ricerca o soggetti privati che collaborano con l'INFN.

3.1 Il personale

Dal censimento effettuato, sono impiegati complessivamente all'interno del mailing un numero di FTE uguale a cinque. Il numero viene dalla somma dei singoli contributi dichiarati da ogni Sezione come mostrato in tabella 3:

TAB. 3: FTE impiegati

Sede	FTE
Bari	0.1
Bologna	0.25
Cagliari	0.2
Catania	0.3
CNAF	0.2
Ferrara	0.1
Firenze	0.15
Genova	0.3
LNF	0.3
LNGS	0.3
LNL	0.2
LNS	0.2
Milano	0.2
Milano Bicocca	0.1
Napoli	0.1
Padova	0.2

Parma	0.1
Pavia	0.1
Perugia	0.1
Pisa	0.2
Presidenza	0.1
Roma 1	0.3
Roma 2	0.2
Roma 3	0.4
Torino	0.2
Trieste	0.1

3.2 Le strutture

In media sono utilizzate 4.6 macchine in ogni sede per la gestione del servizio di posta separando, spesso, la funzione di ricezione/spedizione messaggi da quella di accesso alle caselle di posta. La tabella 4 raccoglie le informazioni sull'hardware utilizzato con particolare attenzione allo spazio disco dedicato alla memorizzazione della caselle di posta. Si noti che alcune volte queste macchine assolvono anche altri compiti all'interno dei centri di calcolo, in particolare nel caso di cluster che ospitano macchine virtuali e d'infrastrutture di storage distribuite come le Storage Area Network.

TAB. 4: Risorse hardware

Tipo di risorsa	Totale sedi	Media
Macchine reali	65	2.6
Macchine virtuali	53	2
Macchine totali	118	4.6
Spazio disco utilizzato	9 TB	340 GB

3.3 Il software

Tutte le sedi INFN utilizzano software open source su piattaforme Linux o BSD. Mostriamo nella pagina seguente alcuni dati riguardanti i prodotti software utilizzati e i file system sui quali insistono le mailbox degli utenti. Per il servizio di spedizione/ricezione (SMTP) posta si utilizza sendmail oppure postfix come mostrato in figura 1; alcune Sezioni utilizzano entrambi. Per quanto riguarda il servizio di consultazione della propria casella di posta ci sono quattro prodotti adottati a piacimento nelle diverse sedi: cyrus-imapd, dovecot, uw-imap, courier come mostrato in figura 2. La figura 3 mostra la distribuzione dei diversi tipi di file system utilizzati per la persistenza dei dati delle mailbox utente. La figura 4 mostra la distribuzione dei sistemi di autenticazione adottati nelle varie sedi e la figura 5 la distribuzione delle soluzioni adottate per fornire un'interfaccia web per la lettura e l'invio della posta elettronica. In figura 6 è riportata la situazione relativa all'utilizzo di sistemi di alta

affidabilità per il servizio di mailing. Come si può evincere dai risultati del censimento esiste una notevole eterogeneità di soluzioni implementate nelle varie sedi INFN. Se da un lato questo significa che ci sono molte competenze all'interno dell'ente, di contro abbiamo tante soluzioni diverse per un problema comune e questo può significare una scarsa ottimizzazione delle risorse necessarie al servizio di mailing.

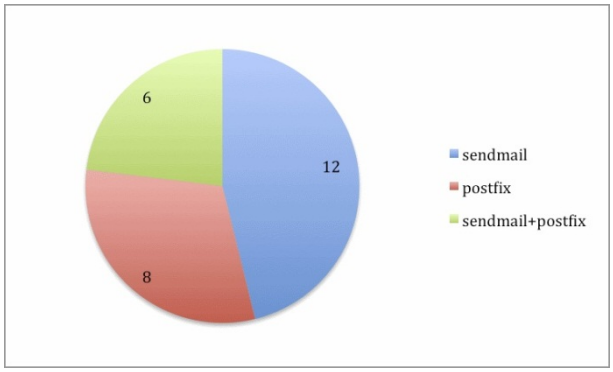

FIG. 1: MTA Server

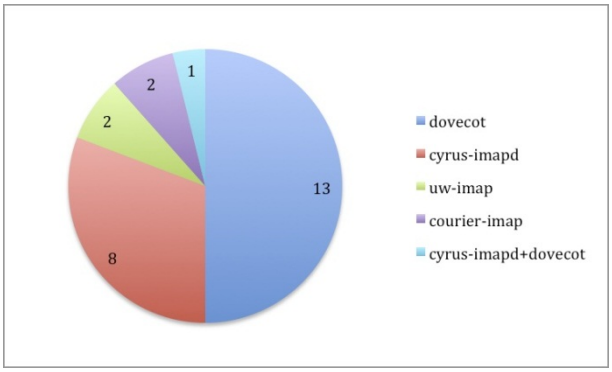

FIG. 2: IMAP Server

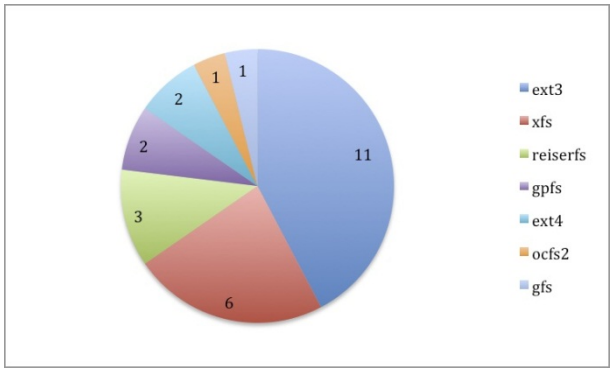

FIG. 3: File System

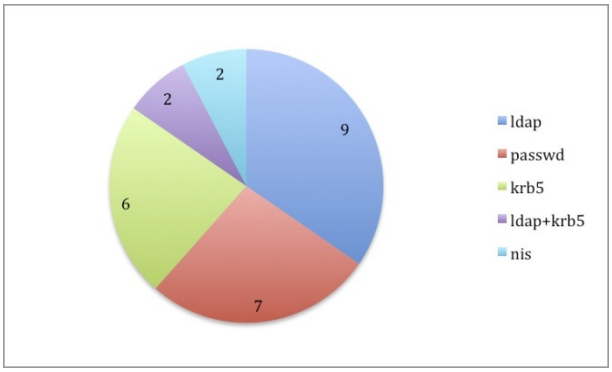

FIG. 4: Autenticazione

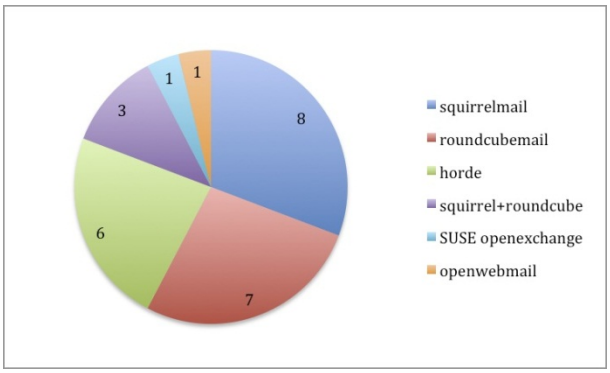

FIG. 5: Webmail

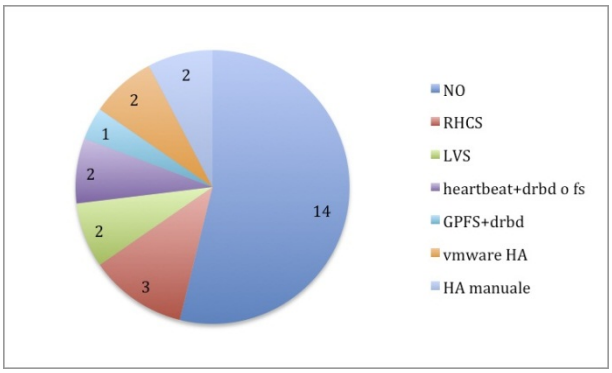

FIG. 6: HA & Cluster

4 POLICY NELLE DIVERSE SEDI E TIPOLOGIE DI UTENTI

Nelle diverse sedi INFN il servizio di mailing è fornito a differenti tipologie di utenti esterni. Nel caso di sedi con forte presenza di collaboratori affiliati ad altre istituzioni, l'assegnazione a questi ultimi di una casella di posta, ancorché prassi comune presso le principali istituzioni internazionali di ricerca, risponde a criteri tesi a rendere il più fluido ed efficace possibile lo scambio di informazioni all'interno degli esperimenti e tra la generalità degli utenti e la struttura in cui le attività scientifiche vengono svolte. L'assegnazione di caselle di posta a collaboratori esterni è quindi, in questo caso, funzionale all'assolvimento delle finalità istituzionali dell'Ente, e difficilmente se ne potrebbe prescindere anche nel caso di esternalizzazione del servizio. Un'altra tipologia molto frequente nel caso di Sezioni ospitate all'interno di strutture universitarie sono gli utenti afferenti ai Dipartimenti di Fisica con i quali l'Ente ha in essere accordi o convenzioni.

Dove l'INFN è il fornitore principale del servizio di posta, gli account utente sono forniti ai dipendenti, agli associati, ma anche a personale temporaneo (laureandi, dottorandi, borsisti, assegnisti) e a personale staff dell'università. A volte sono forniti account di posta anche a utenti di enti presenti all'interno dei dipartimenti che non collaborano direttamente con l'INFN. Questo oltre a comportare un carico di lavoro considerevole (si pensi ad esempio al continuo ricambio di laureandi, assegnisti, borsisti) avrebbe anche implicazioni gestionali non banali qualora si volesse passare ad un servizio centralizzato. Infatti, mentre può essere relativamente semplice gestire utenze esterne all'INFN a livello locale, è sicuramente più complesso gestirle all'interno di un sistema centralizzato, o comunque non in sede locale: anche la sola verifica della posizione amministrativa degli utenti diventa difficoltosa da gestire. Oltre a questo va considerato che la policy applicata alle utenze di posta nelle varie sedi è molto spesso differente e frutto di ottimizzazioni e accordi locali. Sono esempi di questo la durata di un account e le procedure per ottenerlo, la quota sullo spazio disco utilizzato e le forme di backup, la gestione dello SPAM e le modalità di accesso remoto al servizio di posta, ecc. Qualora l'ente decidesse di fornire il servizio di posta solo a dipendenti e associati la transizione per tutti gli altri utenti andrebbe concordata e gestita preventivamente. Viceversa se si volessero continuare a supportare tutti gli utenti adesso gestiti con un sistema diverso da quello ora in funzione, andrebbero prima di tutto individuate delle procedure amministrative con tutti i soggetti interessati e nell'implementazione tecnica del servizio andrebbe prevista ad esempio la gestione di utenti non presenti nei database INFN.

5 SCENARI POSSIBILI PER IL SERVIZIO DI POSTA ELETTRONICA

Vi sono diversi scenari da prendere in considerazione come possibile evoluzione dell'attuale sistema, anch'esso da non sottovalutare rispetto alle altre soluzioni. Qui di seguito abbiamo cercato di descrivere i principali, in modo da poterne confrontare vantaggi e svantaggi e da potere dare una stima, sia pure approssimativa, dei costi sia in termini di hardware sia di software o di FTE da impiegare.

5.1 Mail centralizzato fuori INFN (outsourcing)

Tutto il sistema di posta è ospitato e gestito in outsourcing da un fornitore di servizi esterno all'ente. Occorre prestare attenzione ad alcuni punti fondamentali (secondo cosa si chiede, il costo può lievitare notevolmente):

- Possibilità di usare i nostri sistemi di autenticazione, in modo da mantenere i nostri database di autenticazione (AAI).
- Gestione e SLA dell'helpdesk per quanto riguarda tutto ciò che richiede operazioni sui server (recupero email, verifica stato singolo email, risoluzione eventuali problemi e più in generale i tempi di ripristino).
- Numero di caselle di posta elettronica e quantità di dati (per casella e/o totali).
- Posizionamento dei server (in Italia, in Europa o paesi extra UE).
- Normativa riguardo alla privacy e alla confidenzialità dei dati.
- Network bandwidth.

Vantaggi:

- Il costo è omnicomprensivo (comprende hardware, software, alimentazione elettrica, raffreddamento).
- Secondo lo SLA scelto, è possibile avere un uptime del sistema molto alto.
- Tutta una serie di operazioni sono unificate; dal punto di vista del personale dei servizi di sezione, non si tratta più di occuparsi della gestione sistemistica ma solo della gestione degli account.
- Visto il punto precedente, ci si aspetta un minore impegno di personale tecnico, che però non è immediatamente quantificabile: infatti, se da un lato la manutenzione sistemistica dei server di posta non sarebbe più a carico di personale INFN, il troubleshooting dei client sarebbe ancora a nostro carico.
- Secondo il tipo di prodotto su cui ci si orienta, molto probabilmente insieme alla posta elettronica ci sarà accesso anche ad altri tool collaborativi che possono tornare utili (mailing lists, calendario, instant messaging, rubrica centrale,...). Inoltre il sistema potrebbe essere accessibile da un numero maggiore di piattaforme (es. smartphones).

Svantaggi:

- L'interazione per la soluzione di eventuali problemi deve necessariamente passare

attraverso l'interazione tra tecnici nostri e il servizio di gestione esterna (infatti a priori non pensiamo che la gestione esterna permetta a nostri tecnici l'accesso con privilegi al sistema). Questo comporta:

  - Non avere direttamente una persona disponibile alla quale rivolgersi, che abbia i privilegi per fare qualcosa.
  - Probabili ritardi nell'arrivare a una soluzione in tutti quei casi in cui il problema comporta l'interazione di più persone a più livelli.

- Alcune richieste relativamente semplici alle quali i nostri utenti sono abituati ad avere una risposta in tempi brevi ("È partito il mio email?", "Potete controllare se è arrivato un email particolare?", "Perché un certo email è considerato spam?") richiedono comunque un accesso al sistema con privilegi: in generale finirebbero con l'essere trattate come per la risoluzione di problemi del punto precedente. Inoltre questo tipo d'interazioni potrebbe essere visto non come "risoluzione problemi" ma come attività extra e avere quindi un notevole costo aggiuntivo.
- Il sistema in generale sarà connesso a internet attraverso una rete che non è quella del GARR: eventuali problemi di rete geografica a livello di banda o di congestione possono ripercuotersi sul nostro utilizzo.
- Perdita di know-how da parte del personale tecnico: se dopo alcuni anni di outsourcing si decidesse di ritornare ad avere una gestione interna della posta elettronica, il personale dovrebbe necessariamente re-impadronirsi di competenze tecniche che in questo momento possiede.
- Costi da giustificare: se da un lato è vero che il costo è omnicomprensivo (hw, sw, corrente, raffreddamento, manutenzioni), normalmente viene presentato sotto forma di un costo annuale per la singola casella di posta elettronica (quindi da moltiplicarsi per il numero di caselle), o di un costo annuale totale per un certo numero massimo di caselle di posta. In più, sempre legato a questo costo, normalmente c'è un limite alla dimensione della casella.
- Ospiti: visto il punto precedente (cioè a cosa sono legati i costi di questo tipo di soluzione) e dato il numero di caselle di posta ora presenti presso varie sezioni assegnate a persone che non sono dipendenti/associati/borsisti dell'ente (vedasi tabella 1), occorre necessariamente prendere una decisione riguardo alla domanda "Chi ha diritto alla casella di posta elettronica?". La risposta a questa domanda può avere varie conseguenze. In alcuni casi, infatti, queste persone hanno la casella di posta elettronica in virtù di convenzioni tra le singole sezioni e le università ospitanti: un cambiamento dei termini non può essere preso unilateralmente ma deve necessariamente essere concordato, oppure bisogna aspettare il rinnovo della convenzione per riformularla. Forse questo punto non è uno "svantaggio" nel senso stretto del termine, però bisogna tenerne conto per sapere a cosa si va incontro. È inoltre da considerare la problematica che si riferisce ai collaboratori esterni che svolgono attività presso strutture di ricerca dell'INFN. L'opportunità di fornire il servizio di mailing a questi utenti è legata all'effettivo espletamento delle finalità istituzionali dell'Ente, per cui la scelta sul

continuare o meno a concedere loro una casella di posta INFN anche in caso di esternalizzazione è da fare oggetto di attenta valutazione.

5.1.1 Costi di questa implementazione

I fattori che intervengono nel valutare i costi di questa soluzione sono legati al numero di caselle di posta, allo spazio massimo richiesto per ogni casella di posta e al tipo di SLA (tempi d'intervento, recupero dati, backup, alta affidabilità, …), e sono tipicamente un costo annuale. A questo va aggiunto il personale dell'ente che dovrà fare supporto sul sistema. In più non è per nulla detto che la migrazione iniziale dei mail esistenti dentro il nuovo sistema sia contemplata: questo però dovrebbe essere un problema minore e per il momento non ci soffermeremo su quest'aspetto.

Esaminiamo questi fattori singolarmente:

- Numero di caselle di posta: per questo punto possiamo rifarci alle informazioni presentate nel capitolo 3, in particolare nella tab. 1. Ora abbiamo un totale di poco più di 10000 caselle di posta elettronica, di cui circa 3600 sono di "altri", cioè di persone che non sono né dipendenti né associate. Questi "altri" appartengono a tre distinte categorie: universitari (personale del dipartimento di Fisica), persone che hanno due caselle di posta in sedi diverse (succede tipicamente nei laboratori), esterni (persone che hanno l'esperimento nei laboratori). La prima categoria è già stata oggetto di discussione del cap. 4; la seconda categoria non avrebbe più motivo di esistere; la terza categoria però non è eliminabile così facilmente, per cui cercheremo di tenerne conto. Per questo motivo stimiamo in 9000 il numero di caselle di posta elettronica necessarie.
- Spazio per le caselle di posta: come si può vedere in tab. 3 lo spazio disco totale attualmente a disposizione è di 9 TB. Questo però non tiene conto dell'enorme differenza nella dimensione delle caselle di posta tra utenti diversi. Per esempio, Google Apps Premier fornisce caselle di posta di 25GB. Un'altro esempio è Office 365 on-line che fornisce un servizio del tutto analogo con interfaccia di gestione Exchange, 25GB di quota per casella di posta e una serie di tool collaborativi tra i quali Skydrive Pro, una sorta di disco virtuale che può essere utilizzato come servizio "Dropbox".
- SLA: il tipo di servizio che dobbiamo richiedere deve necessariamente comprendere network bandwidth, tempi d'intervento, alta affidabilità, backup dei dati, interfacciamento con il nostro database di autenticazione (AAI). Eventuali richieste riguardo a privacy, confidenzialità e posizionamento dei server che ospiteranno questi dati (Italia, Europa o extra UE) dovranno essere valutati dall'ufficio legale in accordo con la normativa attualmente vigente. Ad esempio Google Apps e Office 365 hanno clausole legali differenti fra loro. Office 365, ad esempio, offre garanzie sulla localizzazione geografica dei server e sulla privacy dei dati.
- Personale: il personale sarà necessario per la risoluzione ordinaria dei problemi degli utenti (e questo succede già in tutte le sedi per cui in realtà non è un vero costo) e per

fare da filtro verso il supporto di secondo livello fornito dal venditore del servizio. Questo filtro deve essere composto di alcune persone (diciamo 4 o 5) che saranno le uniche autorizzate a prendere contatto con l'assistenza del fornitore del servizio. La stima di carico di lavoro complessivo per questo compito è tra 0.7 e 1 FTE.

Alla luce di quanto detto, per il costo di questa soluzione diamo una stima del costo per singola mailbox, dando un intervallo di valore che tentativamente tiene conto delle varie opzioni. Il costo stimato è tra i 15 e i 50 euro per singola mailbox. Questo si traduce in una spesa annuale tra i 135K e i 450K euro.

5.1.2 Raccomandazioni

Per procedere su questa strada, cioè per prendere contatto con alcuni rivenditori di servizi, bisogna per prima cosa, redigere un documento abbastanza dettagliato sul tipo di SLA richiesto.

5.2 Mail centralizzato con hardware INFN e software commerciale

In questo scenario l'hardware è ospitato presso una sezione o un laboratorio, e il software è un prodotto commerciale. Questo significa che i seguenti costi sono necessariamente a carico nostro: manutenzione hardware, energia elettrica, raffreddamento e licenze software. La manutenzione sistemistica (con questo termine intendiamo la manutenzione del sistema operativo e i suoi aggiornamenti) può in teoria essere affidata in outsourcing. All'interno dell'ente ci siano forti competenze a riguardo, per cui quest'aspetto sembra poco vantaggioso. In compenso la manutenzione del prodotto commerciale usato sarebbe sicuramente data in outsourcing con eventualmente una piccola partecipazione di personale nostro per i compiti "semplici" più frequenti.

Punti su cui fare molta attenzione (secondo le richieste, il costo può lievitare notevolmente):

- Possibilità di usare nostri sistemi di autenticazione, in modo da mantenere i nostri database di autenticazione (AAI).
- SLA sia per quanto riguarda l'hardware che per il software (quest'ultimo è simile al caso dello scenario precedente).
- Numero di caselle di posta elettronica e quantità di dati (per casella e/o totali).

Vantaggi:

- Server posizionati in Italia, collegati direttamente alla nostra rete e sotto nostro diretto controllo.
- Secondo lo SLA si può garantire un uptime molto elevato.
- In caso di problemi "poco gravi" è possibile che il personale interno che segue il sistema possa agire immediatamente poiché hanno alcuni (se non tutti) i privilegi necessari.

- Potrebbe esserci un minore impegno di personale tecnico nelle singole sezioni, che però non è immediatamente quantificabile: anche se in teoria la gestione del software che implementa il servizio di posta potrebbe richiedere meno impegno, il personale deve comunque acquisire un know-how specifico sul software acquistato e applicare comunque gli aggiornamenti al sistema operativo o le patch al software con l'ausilio di eventuali consulenti esperti esterni all'ente. Rimane comunque l'azione di troubleshooting dei client a carico del personale tecnico delle singole sezioni.
- Secondo il tipo di prodotto su cui ci si orienta, molto probabilmente insieme con la posta elettronica ci sarà accesso anche ad altri tool collaborativi che possono tornare utili (mailing lists, calendario, instant messaging, rubrica centrale,...). Inoltre il sistema potrebbe essere accessibile da un numero maggiore di piattaforme (ad es. smartphones).

Svantaggi:

- Costi da giustificare: il problema sollevato nel caso della posta affidata completamente in outsourcing rimane inalterato con una possibile modifica. Cioè anche se in questo caso i costi per hardware (che deve essere comunque certificato per il software che si userà), corrente elettrica, raffreddamento e manutenzione del sistema operativo sono a nostro carico, vi è comunque un costo per il software che viene calcolato sul numero di caselle di posta elettronica. In questo scenario, secondo il prodotto, è possibile acquistare licenze perpetue (cioè non annuali), anche se solitamente il costo di dette licenze è dell'ordine di due o tre annualità e gli aggiornamenti software possono richiedere una contrattazione separata.
- Ospiti: il problema sollevato nel caso della posta data completamente in outsourcing rimane inalterato. Infatti, l'alto numero di attuali caselle di posta elettronica non INFN incide pesantemente sul costo.
- Perdita parziale di know-how da parte del personale tecnico: il rischio è simile a quello sollevato nel caso della posta data completamente in outsourcing.

5.2.1 Costi di questa implementazione

Valgono la maggior parte delle osservazioni fatte nel punto 5.1.1 con i seguenti distinguo: acquisto hardware, personale e sede. Infatti, in questo scenario l'hardware e la sua gestione sarebbero a carico nostro.

La scelta tecnica, cioè il tipo di software utilizzato, influisce pesantemente sulle richieste di hardware e software. Per questo motivo i seguenti numeri devono necessariamente essere considerati più che indicativi, e bisogna aspettarsi variazioni anche molto rilevanti: richieste stringenti sul disaster recovery (ad es. la duplicazione del sito, i tempi di ripristino di servizio, ecc.) possono più che raddoppiare i costi.

L'hardware sarà costituito da server, storage, apparati di rete per la connettività IP e per l'eventuale Storage Area Network. Per i server, dobbiamo aspettarci una spesa media di 3K euro a server, per un totale di 10 server. Lo storage necessario per una stima di 9000 utenti, dedicando una quota media di 5 GB/utente, è pari a 45 TB. A questi vanno aggiunti circa 100

TB di disco per backup. La tipologia di accesso randomico e le prestazioni in termini non solo di banda ma anche di IOps suggeriscono, per ospitare le mail, l'utilizzo di storage SAS a 15 krpm o, meglio ancora, di sistemi intelligenti di storage gerarchico con parti molto performanti (SSD/SAS) e parti meno performanti (NL-SAS a 7.2 krpm), che possono essere oggi quotati a circa 1500 euro/TB. Feature software quali la deduplica dei dati possono essere molto efficienti, quindi desiderabili. Le aree per i backup ed eventuali aree di quarantena per candidati virus e spam possono invece essere ospitate su dischi a basso costo (250-300 euro/TB). Il costo così valutato, circa 100 Keuro, va incrementato del 20% per la ridondanza RAID. Un sistema di questo genere potrebbe richiedere una Storage Area Network, con un costo oggi valutabile in 15 Keuro, oltre ad apparati di rete locale il cui costo è però trascurabile.

Il costo dell'indispensabile manutenzione può essere considerata compresa nella valutazione fornita. Bisogna comunque tenere presente che, in generale, l'hardware andrebbe rinnovato ogni quattro o cinque anni.

L'housing dell'hardware deve essere fatto presso due sedi che abbiano buoni collegamenti di rete e infrastruttura già esistente (ups, condizionamento, gruppo elettrogeno) e probabilmente non influisce sul costo.

Il costo delle licenze dovrebbe variare tra i 20 e i 50 euro per licenza, a seconda che si vogliano licenze annuali o perpetue. Nel primo caso sono compresi upgrade a versioni più nuove, nel secondo non è detto.

Per quanto riguarda il personale, la stima minima è di 4 o 5 FTE preferibilmente concentrati nelle due sedi che ospiterebbero l'hardware. Potrebbe anche essere necessario prevedere un percorso di formazione professionale per un paio di persone. Ad esempio, nel caso si scelga Microsoft Exchange, sarebbe molto consigliabile avere all'interno dell'ente, una o due persone che seguano regolarmente i corsi di certificazione Microsoft.

A questo proposito abbiamo raccolto informazioni sull'implementazione di un servizio basato su Exchange e sui costi che possono risultarne. L'hardware dovrebbe essere opportunamente dimensionato al numero di caselle. I server possono esser virtualizzati ma occorre un'adeguata struttura di virtualizzazione certificata per funzionare con Microsoft. La scelta ricadrebbe su una soluzione VMware che ha un costo di licenza dipendente dal numero di CPU. Lo storage per le mailbox non necessariamente deve essere di tipo SAN con dischi SAS, ma possono essere dei dischi DAS direttamente collegati alle macchine che operano da IMAP server. Questo ridurrebbe i costi per lo storage. La soluzione Exchange ha un costo che è calcolato per CALL. E' necessaria una CALL per ogni FTE dell'ente. Il conto degli FTE non è un'operazione puramente commerciale che va concordata con Microsoft. Alcune ricerche che abbiamo condotto ci hanno portato esempi di costi di licenza attorno ai 10 Euro per FTE, con numero di FTE superiori ai 5000. Fare previsioni economiche in questo caso è veramente difficile; stimando ipoteticamente 5000 FTE per l'INFN, il risultato porta ad un costo complessivo di circa 50K euro all'anno per le licenze client. A questo costo va aggiunto il supporto Microsoft che è indispensabile per risolvere situazioni critiche. Il costo del supporto è di circa 50K euro l'anno. Tralasciando quindi il costo dell'hardware, la spesa

annua per licenze Exchange sarebbe di circa 100K euro. Va aggiunto a questo la parte di design, implementazione e startup che potrebbe avere un arco di sviluppo di diversi mesi. In questa fase si è obbligatoriamente seguiti da due esperti Microsoft. La professionalità di queste persone costa circa 1400 Euro il giorno. E' quindi probabile stimare una spesa iniziale di altri 200K euro per l'implementazione del sistema. Il costo ovviamente è una tantum e non annuale. In definitiva il costo totale si divide in:

- Hardware:
    - Costo dei nodi per erogare il servizio: 16K euro + eventuali 16K euro per una soluzione speculare in una seconda sede INFN
    - Costo dello storage: dai 100K euro ai 150K euro . Il costo raddoppia in caso di voglia ridondare lo storage geograficamente.
        - Nel caso si scelga storage molto economico per utilizzo con Microsoft Exchange la spesa potrebbe scendere a 60K euro.
- Software:
    - Costo di prima implementazione nel caso di Microsoft Exchange: da 50K euro a 200K euro
    - Costo di mantenimento licenze software e supporto Microsoft: circa 100K euro ma la cifra è variabile secondo il tipo di contratto.

In linea di massima il costo di un sistema di questo tipo può variare da un minimo di 120K euro a un massimo di 330K euro come costo d'implementazione per l'hardware (nei sei anni di vita del sistema) secondo le caratteristiche di affidabilità che deve avere, più il costo annuo delle licenze e del supporto che si attesta su un minimo di 100K euro annui. La grande variabilità dei fattori che determinano il costo non consente di avere una stima dei costi più accurata. A questi costi vanno poi aggiunti almeno 4 FTE.

5.2.2 Raccomandazioni

Un sistema di questo tipo non è presente in nessuna sezione né in alcun ente di nostra conoscenza che sia distribuito geograficamente sul territorio come lo è l'INFN. In particolare nessuna sezione ha richieste spinte di disaster recovery o di business continuity e questi aspetti influiscono pesantemente sia sull'aspetto tecnico che su quello economico. Per una soluzione di questo tipo consideriamo necessario lo sviluppo di un progetto pilota.

5.3 Mail centralizzato interno all'INFN con software open source

Questo scenario si basa su una soluzione centralizzata gestita completamente dall'INFN. L'ente sarebbe proprietario del proprio hardware utilizzando un sistema operativo open source, così come i diversi elementi software che compongono il servizio di posta nel suo insieme. L'hardware dovrebbe essere ospitato presso una sezione o un laboratorio. Per avere caratteristiche di resilienza il sistema dovrebbe seguire delle precise specifiche di "disaster recovery" ed essere quindi replicato in almeno un'altra sede INFN. La gestione sia sistemistica che del software utilizzato per la posta elettronica è a carico di un gruppo di

tecnici e tecnologi interni (personale dell'ente).

Vantaggi:

- I costi riguardano solo hardware e personale, mentre il software è gratuito.
- Non vi è un problema particolare legato a costi per numero di caselle di posta elettronica; però, se dovessimo implementare questo scenario ora, dovremmo tenere conto del numero totale di utenti/spazio disco ora usato (anche da utenti non INFN) nel dimensionare il sistema.
- Dovrebbe essere più facile implementare una policy di gestione (ad es. decidere autonomamente il servizio per dipendenti, associati e ospiti).
- Dovrebbe essere più facile l'integrazione con AAI.
- Rimane nell'ente un gruppo di persone altamente specializzate con competenze tecniche nel campo della posta elettronica.

Svantaggi:

- La manutenzione del software può rivelarsi impegnativa: giacché il software è open source non bisogna aspettarsi supporti particolari bensì essere pronti a trovare autonomamente soluzioni a problemi.
- Necessità di istituire un help desk per fornire copertura in orari standard (10x5 nei giorni lavorativi) con la possibilità di avere reperibilità in caso di emergenze.
- Organizzazione della gestione e FTE necessari: è da dimostrare che i membri del gruppo di manutenzione possano effettivamente lavorare efficacemente anche da altre sedi, così come il numero di FTE necessari a coprire il servizio.
- Questo scenario presenta una criticità peculiare ossia l'implementazione di un sistema in alta affidabilità e alta disponibilità. Il fatto di centralizzare completamente il servizio fa sorgere problematiche che normalmente nell'attuale servizio decentrato per sede non s'incontrano. Un disservizio su un server di posta di una qualsiasi sezione non ha impatto sulle altre. In un modello completamente centralizzato un disservizio sul sistema può avere impatto su tutti gli utenti. Per questo è necessario avere un sistema gemello in un'altra sede INFN e renderlo disponibile all'interno di un'opportuna architettura di HA. Da qui nascono diverse complessità da affrontare nel disegno dell'architettura del servizio che normalmente non si riscontrano nei sistemi attuali poiché gestiscono un numero ridotto di utenti e non dipendono l'uno dall'altro.

5.3.1 Scenari possibili d'implementazione

L'implementazione di un servizio di posta centralizzato per un ente fortemente distribuito sul territorio nazionale è una soluzione che fino ad ora non è stata adottata da nessun altro ente che ha caratteristiche analoghe a quelle dell'INFN. I laboratori, centri di calcolo o altri enti di ricerca, anche all'estero, sono identificabili in un'unica unità a livello geografico. Nel nostro caso la diffusione dell'ente su un territorio molto vasto complica molto la situazione. Limitiamoci quindi inizialmente a considerare un sistema di posta centralizzato

che abbia caratteristiche di alta affidabilità locale al sito dove ospitato e di *disaster recovery* a livello geografico ma senza alcuna caratteristica di *business continuity*. Per *disaster recovery* s'intende l'insieme di misure tecnologiche e organizzative e logistiche atte a ripristinare un determinato servizio a fronte di gravi emergenze o eventi catastrofici che rendono il servizio stesso non disponibile. Il ripristino del servizio non ha carattere d'immediatezza ma sarà necessario un determinato tempo all'interno del quale il servizio non sarà disponibile.

Per *business continuity* s'intende la capacità di un servizio di rimanere disponibile senza alcuna interruzione di sorta anche a fronte di eventi avversi o catastrofici che possono colpirlo e renderlo indisponibile nella sede primaria dove è ospitato. Al verificarsi di questi eventi il servizio sarà immediatamente reso disponibile su sede geografica in altro sito.

Vi sono alcune considerazioni indipendenti dal tipo di architettura che si utilizzerà per implementare il servizio di mailing. Considerando l'elevato numero di circa 10000 utenze da servire, il servizio dovrà essere partizionato su un adeguato numero di nodi. Consideriamo separatamente la parte di posta legata all'invio e alla ricezione tramite protocollo SMTP e la parte di accesso alle mailbox da parte degli utenti, governata dal protocollo IMAPv4.

IMAP

La parte del servizio di accesso alle mailbox utente dovrà far fronte a un numero di sessioni di circa 20000 unità nei momenti in cui il servizio è maggiormente utilizzato. Per far fronte a un così elevato carico si può pensare ad una soluzione basata su architettura Cyrus Murder descritta brevemente nel paragrafo 6.2.2. Questa soluzione è stata sviluppata alla Carnegie Mellon University e adottata in diversi altri centri come l'Università di Cambridge, la Columbia University, ma anche da provider commerciali che servono centinaia di migliaia di mailbox come FastMail.fm. Un'architettura di questo tipo necessita al minimo di 8 nodi: 2 front-end, 4 back-end e 1 MUPDATE server ed una sua copia di backup

SMTP

La parte di ricezione ed invio delle mail dovrebbe essere partizionata in un adeguato numero di server dedicati alla posta in entrata e ad un altro gruppo di server dedicati alla posta in uscita esclusivamente previa autenticazione. Un altro gruppo di server dovrebbe essere dedicato alla posta in uscita SMTP non inviata direttamente da un account utente ma ad esempio da nodi di farm di calcolo o componenti del Sistema Informativo che inviano messaggi automatici a diversi destinatari interni all'ente ma anche esterni. Per questo tipo di servizio si prevedono un minimo di 6 nodi. Parte di questi nodi dovrà essere dedicata anche al servizio di antispam e antivirus.

Antivirus/Antispam

E' un servizio fondamentale che richiede un carico di lavoro considerevole. E' ragionevole avere almeno due server separati dedicati a quest'operazione sia per la posta in entrata sia per la posta in uscita.

Soluzioni SMTP/Antivirus/Antispam commerciali

Vale la pena considerare anche soluzioni commerciali che forniscono appliance pronte all'uso quali Cisco IronPort, che si occupano dei servizi sopra citati esclusa la parte IMAP. Questo non solo semplifica l'architettura ma può agevolare la gestione del servizio stesso.

Storage

Una delle parti più complesse per un sistema di questo tipo è l'architettura dello storage da dedicare alla parte IMAP del servizio. Occorrono uno o più dispositivi di storage ad elevate prestazioni con dischi SSD da essere utilizzati come dispositivi di massa cache, e dischi SAS ad elevate prestazioni (15K). Lo storage in questione deve avere adisposizione almeno 100TB di spazio al netto della configurazione RAID che si può scegliere. Lo spazio inoltre deve essere partizionabile per essere reso disponibile ai vari nodi IMAP. Si pensa quindi ad uno storage di tipo FC 8Gbps. I sistemi che implementano SMTP possono utilizzare storage locale ma sempre ad alte prestazioni.

E' naturale in fase progettuale fare riferimento a soluzioni che utilizzino la virtualizzazione per almeno due motivi principali:

- Utilizzo razionale dell'hardware consentendo di sfruttare al massimo le risorse disponibili.
- Maggiore flessibilità in caso di eventi avversi (disaster recovery) con la possibilità di usufruire di snapshot delle macchine virtuali che implementano i servizi.

Prendiamo in esame due modelli principali. Il primo basato su cluster di macchine virtuali RedHat implementato in ambiente open source CentOS e quindi a costo zero. Il secondo commerciale, basato su cluster VMware, costoso ma con elevate caratteristiche di alta affidabilità, gestione delle risorse hardware da assegnare alle macchine virtuali, gestione degli snapshot senza interruzioni di servizio.

5.3.1.1 Soluzione basata su cluster CentOS 6

In quest'architettura l'utilizzo del cluster CentOS serve come infrastruttura software all'interno della quale le singole macchine virtuali che implementano i servizi di mailing sono esse stesse dei servizi per il cluster stesso. Il cluster garantisce l'univocità del servizio gestito e quindi della macchina virtuale (ne sarà garantita l'univocità dell'esecuzione della sua istanza) e il suo ripristino sul primo nodo fisico disponibile in caso di eventuale fallimento hardware del nodo fisico sul quale l'istanza virtuale è in esecuzione in un determinato momento. Il fatto però che un determinato servizio virtualizzato sia in esecuzione e non fallisca deve essere controllato da un agente esterno (ad esempio tramite un server Nagios) che prenda le opportune contromisure in caso di problemi del servizio monitorato. I nodi fisici costituenti il cluster devono essere almeno sei con le seguenti caratteristiche di massima:

- Almeno 2 CPU di ultima generazione con almeno 8 Core per CPU
- Almeno 64 GB di RAM
- Almeno 2 dischi configurabili con diverse opzioni RAID (0,1)

Il numero di nodi può essere inferiore nel caso si adotti una soluzione SMTP su

appliance come descritto sopra. I nodi del cluster dovranno avere accesso alle risorse fisiche dello storage ad esempio tramite protocollo FC e renderle disponibili alle macchine virtuali tramite interfaccia libvirt. Si aggiunge quindi il costo di uno storage condiviso di fascia alta. Valgono in tal senso le considerazioni sui costi analizzate nel punto 5.2.1.

5.3.1.2 Soluzione basata su VMware vSphere

Un esempio di questo tipo di architettura è in produzione presso la Sezione di Trieste. L'hardware impiegato è un sistema blade composto di cinque lame mono e bi-processore, uno stack di quattro switch di rete e una coppia di switch Fibre Channel per il collegamento allo storage rappresentato da un Hitachi AMS2100 mentre la piattaforma di virtualizzazione è VMware vSphere Enterprise versione 5.1, su questa architettura, tra le tante macchine virtuali, viene ospitato il cluster mail costruito sul modello del LVS (Linux Virtual Server). Il numero di macchine virtuali che lo compongono è di nove ognuna con compiti diversi (dall'alto in basso e da sinistra a destra):

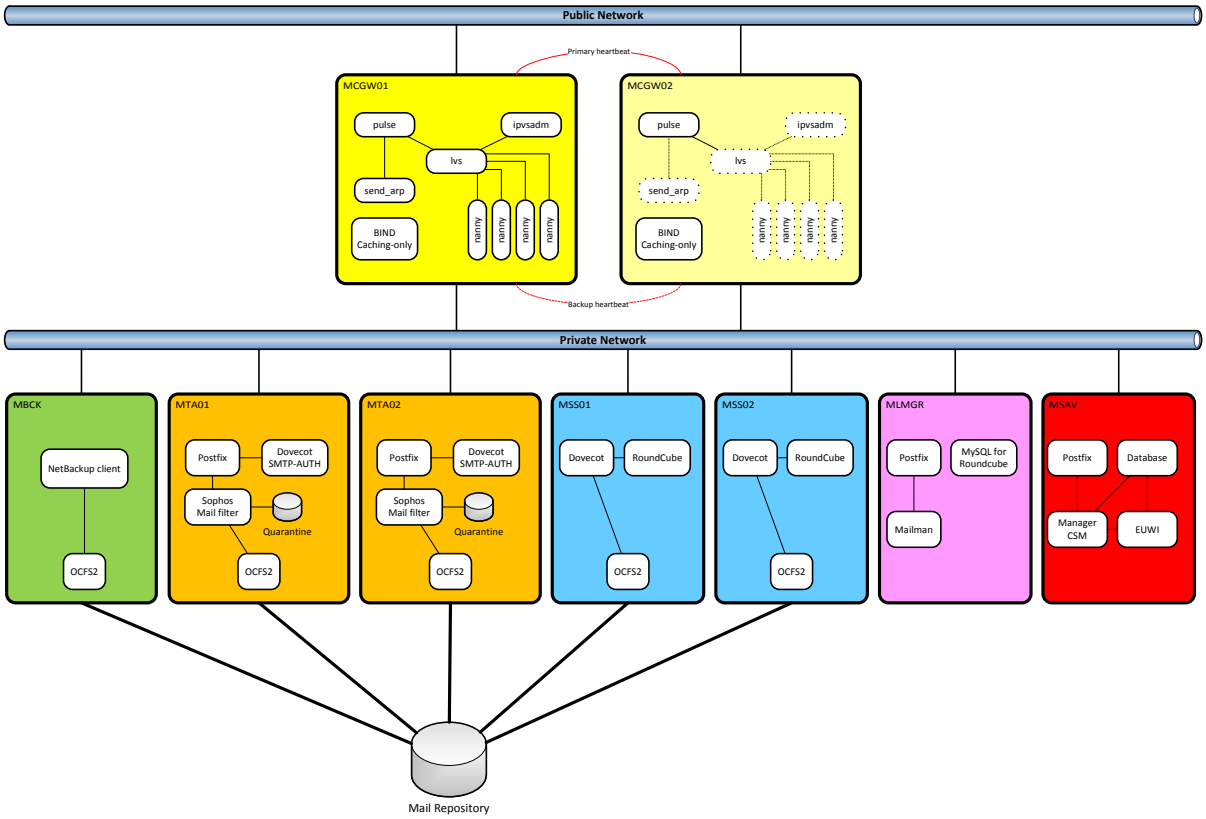

FIG. 7: Mail Cluster virtualizzato

- Router/Load balancer/Process monitor
- Backup Client
- SMTP server
- IMAP server/Web mail
- List manager

- Console centralizzata antispam-antivirus

Le mailbox degli utenti risiedono in un cluster file system OCFS2 che viene montato dai server IMAP, SMTP e dal client di backup. Per sopperire a un carico maggiore di lavoro si possono facilmente aumentare le risorse dedicate ai server SMTP e IMAP (CPU e memoria) e se ciò non fosse sufficiente anche aumentarne il numero.

Le risorse utilizzate complessivamente sono 9 vCPU, 14 GB di memoria ed 1TB di storage per le mailbox. Eccetto lo spazio disco che è dedicato, le altre risorse, CPU e memoria se non utilizzate possono essere messe a disposizione di altre macchine virtuali; infatti, il cluster mail gestisce le 200 mailbox con un carico di lavoro molto modesto.

Questo esempio è solo uno dei possibili scenari, la capacità di vSphere di nascondere le complessità dello strato hardware unita ad un estrema facilità di configurazione oltre che fornire funzionalità di alto livello ne fanno un eccezionale strumento di gestione di sistemi complessi. I costi di questo tipo di architettura dipendono dal livello di funzionalità desiderato, per un'installazione dedicata su singolo sito in termini di licenze software considerando un livello di funzionalità Enterprise sono necessari 3800 euro per il server di gestione (vCenter) al quale vanno aggiunti circa 2200 euro per processore. Se ipotizziamo, alla pari del cluster CentOS citato nel capitolo precedente, una dotazione hardware sei host fisici bi-processore, il totale è di circa 30000 euro. È previsto, dall'anno successivo all'acquisto, un costo annuale di manutenzione che include gli aggiornamenti e il supporto telefonico dal lunedì al venerdì che in questa configurazione si aggira intorno ai 5600 euro. È giusto far notare che l'estrema efficienza degli hypervisor VMware riduce sostanzialmente il numero di host necessari allo scopo che nella configurazione proposta potrebbero essere sovrabbondanti. A questi costi software vanno ovviamente aggiunti i costi per l'hardware riconducibili alle soluzioni già citate.

5.3.1.3 VMware vSphere, vantaggi e svantaggi

Diciamo da subito che l'unico svantaggio della soluzione commerciale è il costo, quello che bisogna valutare è quanto questo costo sia giustificato dalle caratteristiche del prodotto. Sarebbe un errore limitare la valutazione dal punto di vista del solo servizio di mail centralizzato poiché un investimento di questa portata dovrebbe soddisfare le richieste di altri sistemi, non ultimo il sistema informativo. Una virtualizzazione spinta permetterebbe un risparmio in termini di hardware ed eviterebbe in gran parte il problema dell'obsolescenza, infatti, un sistema ben progettato può essere migrato facilmente su host fisici di ultima generazione senza alcuna interruzione del servizio. Per comprendere qual è il livello della piattaforma prodotta da VMware si possono accennare alcuni limiti relativi delle macchine virtuali che si possono gestire:

- Numero massimo di CPU per VM = 64
- Memoria massima per VM = 1TB
- Numero massimo di dischi virtuali per VM = 60
- Grandezza massima di un disco virtuale = 2TB – 512 bytes

A queste caratteristiche basilari vanno aggiunte quelle che sono le features disponibili secondo il tipo di licenza, nella tabella seguente sono descritte le principali:

TAB. 5: VMware features

Nome	Descrizione
Thin Provisioning	Riduce la necessità di spazio disco utilizzando un'allocazione dinamica dello storage senza ridurre le prestazioni della VM.
Update Manager	Coordina le operazioni di patching sia degli hypervisor sia delle applicazioni e i sistemi operativi delle VM.
Data Protection	Fornisce tramite una virtual appliance prodotta da EMC un sistema di backup e ripristino delle VM con funzioni di deduplica.
High Availability	Minimizza in caso di failure di un host fisico il downtime delle VM procedendo ad un restart delle stesse sugli host rimanenti.
vMotion	Rende possibile la migrazione di una VM attiva tra gli host senza downtime.
Hot Add	Per i sistemi operativi che lo supportano è possibile aggiungere CPU, memoria e dispositivi ad una VM attiva senza downtime.
vShield Endpoint	Mette in sicurezza le VM da virus e malware senza installare all'interno delle VM software anti-virus.
Replication	Abilita un efficiente replica delle VM per ottenere un recovery point objective (RPO) intorno ai 15 minuti.
Fault Tolerance	Fornisce un livello zero downtime anche in caso di failure dell'host fisico.
Storage vMotion	Previene il downtime delle VM in caso di manutenzione dello storage (sostituzione, aggiornamenti del firmware,…)
Storage APIs for Array Integration	Incrementa le prestazioni degli array utilizzando istruzioni più efficienti.
Storage APIs for Multipathing	Consente l'utilizzo di software proprietari per il multipathing.
Distributed Resources Scheduler (DRS), Distributed Power Management (DPM)	Bilancia in modo automatico il carico di lavoro delle VM in termini di CPU e memoria tra gli host a disposizione, inoltre permette nelle ore di minor utilizzo di spegnere gli host sotto utilizzati per ottenere un risparmio energetico.
Distributed Switch	Gestione centralizzata delle reti virtuali con monitoring evoluto e possibilità di utilizzare gli switch virtuali Cisco Nexus 1000V e IBM 5000v.
Storage DRS	Bilanciamento automatico delle VM in base al carico prodotto sulle risorse di storage secondo latenza, IOPS,…

L'architettura vSphere è la base della variegata offerta tecnologica di VMware che prevede anche le funzioni di Disaster Recovery tramite l'utilizzo del Site Recovery Manager. È possibile passare da un'installazione vSphere su singolo sito ad una multi sito con SRM senza bisogno di modificare la configurazione delle macchine virtuali.

Ipotizzando una configurazione iniziale su due siti con cinque host bi-processore per sito e la protezione di 25 macchine virtuali questi sono i costi:

- Architettura VMware vSphere Enterprise Plus su due siti = 57500 euro.
- Protezione di 25 VM tramite vCenter Site Recovery Manager Standard (max 75 VM) = 7350 euro.

Nella configurazione si è presunta l'opportunità di impiegare lo Storage DRS (disponibile con il livello di licensing Enterprise Plus) se questa, come auspicato in precedenza, dovesse divenire la struttura di Disaster Recovery dell'Ente, nel qual caso macchine virtuali adibite alla gestione di basi dati potrebbero trarne grande beneficio. Tuttavia limitandosi alla licenza Enterprise che comunque offre ottime funzionalità la spesa per l'architettura vSphere ci sarebbe un risparmio di 10000 euro. Lo storage in questo tipo di struttura riveste un'importanza essenziale giacché devono essere impiegati dispositivi certificati generalmente prodotti dai marchi più noti (EMC, NetApp, EqualLogic, Hitachi,…). Sebbene una precisa quotazione di questo tipo di appliance sia difficile da ottenere poiché dipendente da molti fattori, a puro titolo informativo, è stato richiesto a una società che opera nell'ambito del Disaster Recovery di fornirci un costo approssimativo di una soluzione certificata. La configurazione prevede due appliance, uno per sito, ognuno dotato delle seguenti caratteristiche hardware:

- Doppio controller cluster con IO esteso.
- Protocolli storage supportati: Fibre Channel e NFS.
- Funzione di replica sincrona, semi-sincrona e asincrona.
- Spazio disco raw: 48 dischi SAS 15000 rpm da 600GB + 48 dischi SATA 7200 rpm da 1 TB.
- Flash cache da 1TB.
- Connettività: N.4 NIC GbE rame onboard + 8 aggiuntive, N.4 HBA 4Gb FC onboard + N.4 HBA FC 8Gb aggiuntive.
- Supporto 4 ore onsite 36 mesi.

La soluzione nel suo complesso è stata valutata in 357 Keuro.

5.3.1.4 Disaster recovery a livello geografico

Ci si può appoggiare all'imminente struttura DNS HA che sarà messa in produzione nel corso del 2013. A livello di SMTP server occorre avere una ridondanza dei servizi in una sede secondaria installando macchine gemelle. Queste non hanno bisogno di essere in sincronia con il file system ma solo di essere configurate alla stessa maniera. Per la parte IMAP dipende da quale tipo di implementazione si sceglie. Nella soluzione più economica basata su Cyrus IMAP, nella sede secondaria devono essere installati un numero di IMAP backend uguali a

quelli della sede principale e sincronizzati tramite un meccanismo Cyrus Replica. Occorre poi avere almeno un server di Front End e una copia del server MUPDATE nella sede secondaria. Vi sono però diverse altre possibilità come quella di utilizzare uno storage che preveda la sincronizzazione geografica, ma questo va ad incidere in modo pesante sui costi del sistema.

5.3.2 Costi di questa implementazione

Per l'hardware ci si può rifare alle valutazioni del paragrafo 5.2.1. A differenza della soluzione precedente, non si prevedono costi per licenze e supporto sistemistico. Si dovranno invece considerare i costi per la formazione del personale e per il maggior numero di FTE che questa soluzione richiederebbe (progettazione, realizzazione, mantenimento ed evoluzione del servizio).

In questo caso consideriamo le richieste di disaster recovery, business continuity, ecc. come potenzialmente molto problematiche. Non vi è certezza che l'impiego di solo software open source possa soddisfare i requisiti necessari per fornire un servizio di tale livello. Valgono tutte le osservazioni del punto 5.2.1 con, se è possibile, un grado ancora maggiore d'incertezza.

In termini di FTE, in questo caso dobbiamo distinguere tra gli FTE necessari al normale funzionamento e gli FTE necessari per la progettazione e realizzazione del servizio. I primi sono equivalenti al caso del paragrafo 5.2.1. Per la progettazione e realizzazione la stima minima è di 5 o 6 FTE.

N.B.: il tipo di conoscenze tecniche richieste per la progettazione e realizzazione è completamente diverso da quello necessario alla gestione ordinaria del sistema.

5.3.3 Raccomandazioni

Vale tutto quanto detto al punto 5.2.2: serve un progetto pilota per valutare l'aspetto tecnico.

5.4 Mail decentrato in ogni singola sede

È lo scenario attuale, in funzione da decenni.

Vantaggi:

- I costi riguardano solo hardware e personale, mentre il software è gratuito.
- È facile implementare una policy di gestione (ad es. decidere autonomamente il servizio per dipendenti, associati e ospiti).
- È possibile interfacciarsi alla AAI nazionale anche perché alcune sedi utilizzano già soluzioni tecniche che vengono utilizzate nella AAI (come auth/authz ldap/krb5).
- È un'architettura autoconsistente poiché segue la struttura ad albero del DNS INFN.IT.
- In caso di gravi problemi a uno o più server di posta locali, i messaggi non vengono

persi ma conservati in una macchina presente al CNAF che funziona come MX del dominio INFN.IT e della maggior parte dei sottodomini.

- Qualsiasi tipo di disservizio grave è isolato automaticamente alla sede locale, per com'è creata l'infrastruttura stessa e non ha impatto sul mailing di tutte le altre sezioni. Dovrebbe essere più facile l'integrazione con AAI.
- Rimane nell'ente un gruppo di persone specializzate con competenze tecniche nel campo della posta elettronica.
- I sistemisti risolvono il problema seguendolo sia il lato client sia quello server.

Svantaggi:

- Vanno gestiti i costi per l'hardware che è presente nelle singole sedi.
- Ogni sede ha l'onere di gestire il proprio server di posta e l'help desk per gli utenti.
- Possibile maggiore impegno di personale tecnico nelle singole sezioni.

I costi di mantenimento dell'attuale struttura decentrata sono stati quotati intorno ai 45K euro annui per l'hardware cui vanno aggiunti i 20K euro del contratto Sophos che comprende sia l'antivirus/antispam per i mail server sia la suite di protezione antivirus per i computer dell'Ente. Il personale impiegato come riportato nella tabella 3 ammonta a 5 FTE.

5.5 Soluzione open source parzialmente centralizzata

Tra le varie ipotesi vale la pena considerare uno scenario in cui il servizio di posta sia centralizzato non completamente ma per gruppi di sezioni. In particolare sezioni che dovessero in futuro subire accorpamenti a livello geografico regionale potrebbero usufruire di un unico servizio di posta che le raggruppa in un'unica entità. Rispetto alla soluzione precedente questa avrebbe il vantaggio di avere un minor impatto architetturale e potrebbe esser realizzata ritoccando l'attuale dotazione hardware adeguandola a un maggiore numero di utenti. Questo non necessariamente richiede l'utilizzo di un'alta affidabilità di livello paragonabile a quello di una centralizzazione completa. Questa soluzione implica che nelle sedi, dove risiederebbe il servizio ci sia fin da ora un numero di FTE in servizio adeguati all'aumento del carico di lavoro, e che si creino delle procedure di supporto per le sedi remote (attività nuova rispetto alla situazione attuale). Lo scenario in esame permetterebbe ad alcune sezioni pilota di associarsi per fornire un servizio simile a quello centralizzato senza intaccare i servizi delle sezioni per le quali la migrazione potrebbe essere difficile (a causa delle convenzioni) o non vantaggiosa (per esempio chi ha appena ottenuto una configurazione ad alta affidabilità e con poche necessità di gestione). Il test pilota potrebbe dare delle stime più attendibili su eventuali risparmi economici e di personale per aiutare l'ente a decidere se tornare all'attuale situazione distribuita o andare invece verso una soluzione pienamente centralizzata. Si prevede, tuttavia che il numero di FTE rimarrebbe invariato concentrando il carico di lavoro nelle sedi ospitanti e i costi infrastrutturali potrebbero inizialmente aumentare per effetto della riorganizzazione ma in seguito diminuire di qualche punto percentuale stimabile intorno alla decina.

6 SUPPORTO: INFRASTRUTTURA E HELP DESK

Per valutare i costi del personale coinvolto nel servizio di mailing, dobbiamo distinguere tra i costi per la gestione dell'infrastruttura e i costi per assistere gli utenti.

Per quanto riguarda l'infrastruttura di storage le soluzioni centralizzate dovrebbero portare a un certo risparmio, infatti, gestire una decina di terabyte o un centinaio non cambia gli oneri a carico dei sistemisti supponendo di mettere a confronto infrastrutture con le medesime caratteristiche di affidabilità. Si deve notare comunque che i servizi calcolo dovranno continuare a seguire dei servizi storage locali di alta qualità (home directory, data disks, cluster file systems,…) per cui la gestione centralizzata dello storage per la sola memorizzazione delle mailbox non permette di liberare frazioni significative di manpower nelle sedi locali. Discorso analogo può essere applicato a un'eventuale piattaforma di virtualizzazione, indipendentemente dalla tecnologia impiegata, è molto probabile e addirittura auspicabile che le sezioni e i laboratori continuino a impiegare la virtualizzazione per l'hosting dei servizi di base (ad es.: DNS, WEB, RADIUS,…)

Gli utenti hanno poi necessità di interagire con i servizi calcolo per problemi strettamente legati alla posta: configurazione del client di posta (MUA), folder che 'spariscono', mail o folder cancellati per errore, posta bloccata erroneamente dal filtro antivirus o antispam, posta che ci si attendeva ma non è arrivata, posta inviata che sembra non essere stata consegnata ed infine interpretazione dei messaggi di errore per i messaggi bounced. Ora tutti questi problemi sono risolti dai servizi locali tramite procedure di troubleshooting consolidate dall'esperienza e dalla possibilità di interagire direttamente con l'utente finale.

In un passaggio a servizi centralizzati invece dovremmo pensare a un help desk a due livelli. Un help desk di primo livello a carico dei servizi calcoli locali che risolva i problemi legati alle configurazioni dei client e all'interpretazione degli errori e un help desk di secondo livello, che consiste in circa un FTE distribuito su almeno due o tre persone per risolvere tutti i problemi che richiedono l'accesso ai server centrali, ai file di configurazione e ai log degli stessi.

La stima di un FTE potrebbe essere sottostimata o sovrastimata. Di certo il sistemista dovrebbe rispondere in priorità alle richieste di help desk di secondo livello che dovessero arrivare dai colleghi dell'help desk di primo livello o dagli utenti.

In questo senso il servizio è più prioritario rispetto all'help desk del sistema informativo per il quale se si blocca qualcosa le conseguenze, sono un ritardo nell'aprire, autorizzare o chiudere una missione o nel formalizzare un ordine. Non si può pensare che un utente INFN rimanga senza mail per diverse ore o che un certo mail atteso non sia sbloccato per diverse ore.

7 LINEE GUIDA PER IL SERVIZIO DI POSTA NELLE SEZIONI INFN

Proponiamo una serie di linee guida da seguire per l'installazione di un servizio di posta. Queste riguardano la scelta del tipo di software da utilizzare per fornire il servizio, il tipo di hardware da rapportare ovviamente al numero di utenti da gestire. Differenziamo quindi in base alla dimensione della sezione considerando il numero di caselle di posta da dovere gestire distinguendo in due categorie:

- Sezioni piccole (fino a 200 mailbox)
- Sezioni medio-grandi (oltre 200 utenti)

7.1 Sezioni piccole

Per sezioni di piccole dimensioni si fa riferimento a quelle sezioni che per numero di utenti possono avere un servizio di mailing erogato da un unico server su macchina virtuale che raggruppa al suo interno i principali servizi (SMTP, IMAP, POP3, Webmail, antivirus ecc). Il gruppo mailing rende disponibile una macchina virtuale KVM con una configurazione base che può essere messa in produzione in modo semplice su qualsiasi hypervisor KVM o cluster RedHat. La macchina virtuale dovrà avere assegnate un certo numero di risorse hardware indispensabili per erogare i servizi in maniera ottimale:

- Almeno 16 GB di ram disponibili
- Almeno 2 CPU virtuali
- Almeno 1TB di spazio disponibile per le mailbox

La macchina virtuale è realizzata con i seguenti componenti:

- Sistema Operativo Centos 6.x
- Server SMTP per invio e ricezione della posta: postfix
- Server IMAP per la posta in entrata: dovecot
- Webmail: RoundCube Mail
- Antivirus/Antispam: amavisd-new con Sophos sweep virus engine e SpamAssassin

Ogni utente ha a disposizione 5GB di quota per la propria mailbox e può utilizzare l'interfaccia Web di RoundCube per scrivere i propri filtri sieve per definire regole automatiche di gestione della posta in arrivo

7.2 Sezioni medio-grandi

Per sezioni medio-grandi si fa riferimento a quelle sedi INFN che per numero di utenti non possono essere gestite in modo ottimale da un'unica macchina virtuale che comprende al suo interno tutti gli elementi necessari per un servizio di mailing completo. La struttura di mailing va opportunamente dimensionata con machine reali e/o virtuali secondo il numero di mailbox che devono essere gestite. L'architettura di mailing consigliata ha le seguenti caratteristiche:

- Almeno un server SMTP (definite come MX record nel DNS) per la posta in entrata senza autenticazione.
    - Lo stesso server può erogare il servizio di antivirus e antispam.
- un server imap per ogni 250/300 mailbox. Il delivery della posta tra server SMTP in entrata e server IMAP avviene tramite LMTP.
- Almeno due server SMTP per la posta in uscita eroganti il servizio esclusivamente tramite autenticazione SMTP AUTH su tunnel TLS.
- Almeno un server Webmail.

7.2.1 Server di posta in entrata SMTP

E' necessario predisporre almeno un (ma per motivi di ridondanza sarebbe auspicabile due) server di posta in entrata SMTP. La macchina che fornisce il servizio dovrebbe essere fisica oppure virtuale ma con un adeguato numero di risorse hardware assegnate in modo esclusivo. Tutto il traffico di posta in entrata passa da questa macchina che deve anche eseguire la scansione della posta con antivirus, antispam e implementare meccanismi di greylisting. I software consigliati per realizzare il servizio sono:

- Sendmail per il servizio di posta in entrata SMTP
- Amavisd-new + sophos engine sweep + SpamAssassin per il filtraggio della posta

Come valida alternativa è possibile utilizzare il software Pure Message di Sophos che al suo interno comprende tutte le funzionalità per la gestione di un server SMTP con i servizi di antivirus e antispam

7.2.2 Server IMAP

Per un numero considerevole di utenti l'architettura consigliata da implementarsi nella gestione delle mailbox utenti tramite IMAP è l'utilizzo modello Cyrus Murder. Utilizzando cyrus-imap come software per la gestione delle mailbox è possibile costruire un'architettura ad alta disponibilità. Il modello è basato su un'architettura di aggregazione di server Cyrus IMAP basata su 3 tipologie di server che hanno compiti diversi fra loro:

- IMAP frontend
- IMAP backend
- MUPDATE

I server di backend sono effettivamente quelle macchine che contengono al loro interno i dati delle mailbox e quindi le mail degli utenti e i vari database che indicizzano i messaggi in modo efficiente.

I server di frontend forniscono l'accesso allo spazio imap ma sono effettivamente dataless accedendo ai dati gestiti dai server di backend. Questi server sono fra loro intercambiabili e il malfunzionamento o la non disponibilità di uno di questi server non ha impatto sul funzionamento di tutto il sistema. Il server MUPDATE contiene le informazioni sulla locazione fisica delle mailbox nei server di backend.

L'architettura è molto scalabile, anche se non è una soluzione ad alta affidabilità. Se

opportunamente studiata e dimensionata questa architettura può ridurre al minimo problemi software o hardware che normalmente possono generare un disservizio.

7.2.3 Server di posta in uscita SMTP

Il numero ideale di server di posta in uscita è di due unità, per motivi di ridondanza ed efficienza. Giacché il servizio erogato è di autenticazione e spedizione della posta in uscita previa autenticazione, può essere virtualizzato senza notare un calo significativo delle prestazioni.

7.2.4 Server Webmail

Un unico server Webmail può gestire in modo efficiente anche 500 mailbox, considerando che il servizio non è utilizzato in modo costante da tutti gli utenti, la maggior parte dei quali preferisce un MUA per accedere alla propria casella di posta. Solo per le sezioni con numero di mailbox che superano le 1000 unità, possono richiedere più di un server Webmail. E' auspicabile che questo servizio comprenda anche la possibilità da parte degli utenti di potere scrivere i propri filtri per la gestione automatica della posta in entrata (spostamento dei messaggi in determinati folder, gestione dei tag di SPAM ecc). Il servizio può essere virtualizzato dedicando però in modo esclusivo un'appropriata quantità di RAM.

8 RIASSUNTO DEI COSTI

Nella seguente tabella sono riportati i costi delle possibili soluzioni analizzate, si fa notare che non tutte le soluzioni sono tecnicamente equivalenti e le più costose ovviamente offrono un servizio migliore anche dal punto di vista del disaster recovery.

TAB. 6: Soluzioni e costi

Tipo di soluzione	Costo per lo startup (per almeno un anno)	Numero di FTE per lo startup	Costo di mantenimento annuo	Numero di FTE per il mantenimento	Numero di FTE per l'help desk di 1º livello
Mail Centralizzato fuori INFN (outsourcing)	No	-	135K – 450K	2	1
Mail centralizzato con hardware INFN e software commerciale	50K – 200K	5	20K – 55K	4	1
Mail centralizzato interno all'INFN con software open source	Costo mediato in 6 anni vedere costo mantenimento annuo	6	20K – 60K	4	1
Mail decentrato in ogni singola sede (situazione attuale)	Costo mediato in 6 anni vedere costo mantenimento annuo	-	46K	5	-
Soluzione open source parzialmente centralizzata	Costo mediato in 6 anni vedere costo mantenimento annuo	-	42K - 46K	5	-

9 CONCLUSIONI

Abbiamo illustrato quali devono essere le caratteristiche di un sistema di posta, fotografato la situazione attuale dell'ente ed elencato una serie di scenari possibili con vantaggi e svantaggi. Dai dati raccolti si evince immediatamente che l'INFN ha una situazione anomala da risolvere: quasi un terzo degli utenti di posta gestiti dall'ente riguarda personale non INFN e non associato. Occorre prendere quanto prima una decisione su come gestire questo numero considerevole di mailbox, e interrogarsi se continuare a gestirle. La risposta a questa domanda ha un notevole impatto sui costi e non è sensato portare a termine uno studio tecnico più approfondito per individuare uno scenario ideale se prima non si trova una soluzione al problema degli utenti non INFN, per i quali sono stati evidenziati gli aspetti di opportunità legati sia all'esistenza di convenzioni con i Dipartimenti ospitanti le Sezioni, sia gli aspetti operativi di supporto alle attività istituzionali dell'Ente che sarebbero influenzate in caso di decisioni volte a interrompere questa prassi.

Vale la pena però citare un dato molto importante che è emerso dal censimento. Il numero totale di FTE impegnati nella gestione del mailing è di cinque unità. Il numero è ottenuto dalla somma di tanti piccoli contributi dalle diverse sedi e da questi dati non è evidente per quali sedi la frazione di FTE indicati, anche se piccola, sia un impegno troppo oneroso in termini di personale (alcune sedi hanno evidenziato in passato questo problema).

D'altra parte la centralizzazione o l'esternalizzazione del servizio riguarderebbe solo la parte server mentre il supporto agli utenti per la configurazione, modifica e risoluzione dei problemi lato client rimarrebbero a carico dei servizi calcolo distribuiti.

Anche in un sistema centralizzato con un certo numero di FTE dedicati all'help desk di secondo livello, rimane a carico dei servizi calcolo locali l'help desk di primo livello, come accade ora per il sistema informativo.

Data la cronica mancanza di personale nei servizi calcolo, abbiamo cercato di capire se un sistema di posta centralizzato potesse portare qualche risparmio, ma abbiamo capito che l'eventuale sistema di posta centralizzato gestito dall'ente deve essere replicato ed è molto difficile che possa avvalersi di un numero di FTE inferiore a quello adesso in utilizzo dall'attuale sistema.

Per quanto riguarda le risorse hardware, i risparmi invece si potrebbero trovare ma questi sono in gran parte cancellati dalla necessità di creare un sistema di storage di classe superiore a quelli utilizzati nelle sedi. Nelle sedi comunque rimarrebbero sistemi di storage usati per file locali degli utenti.

L'aspetto più interessante dei sistemi centralizzati è dato dalla possibilità di avere nuove funzionalità quali repliche remote o disaster recovery.

Negli anni scorsi la Commissione Calcolo e Reti ha sempre cercato di centralizzare per rendere più efficienti diversi servizi legati al calcolo (servizi grid, autenticazione e autorizzazione, servizi web e content management system, conferenze, distribuzione di licenze, distribuzione e acquisto di software, contratti centralizzati di manutenzione hardware e software, file system ad accesso geografico, mailing list, audio conferenze e videoconferenze). Il servizio mail invece non si presta alla centralizzazione soprattutto per la

necessità di fornire supporto locale agli utenti lato client e per le strette e variegate relazioni tra le sezioni e i dipartimenti universitari che le ospitano.

10 RINGRAZIAMENTI

All'interno del gruppo mailing vi è stata ampia discussione sulle varie tematiche. Ringraziamo tutti quelli che hanno partecipato rispondendo al sondaggio e Piero Spinnato (LNGS) per il suo contributo.