



UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH



Anàlisi i disseny d'una infraestructura de correu SMTP

**A Degree Thesis
Submitted to the Faculty of the
Escola Tècnica d'Enginyeria de Telecomunicació de Barcelona
Universitat Politècnica de Catalunya
by
Pau Espona Puig**

**In partial fulfilment
of the requirements for the degree in
Telecommunications Technologies and Services
ENGINEERING**

Advisor: José Luis Muñoz Tapia

Barcelona, June 2020

Abstract

The main objective of this thesis is the design and implementation of a mail infrastructure based on the SMTP protocol by a technology company that receives a large number of daily emails which must be filtered without error and stored on their servers. We want to achieve a solution adapted on today, which a very robust architecture and able to operate for many years with high performance.

The development of the project is based on the analysis of the old SMTP system that has already become obsolete, and looking for improvements and alternatives for the new design.

Finally, the proposed reasoning in the report has been implemented and the infrastructure has been simulated with Microsoft Azure, creating the work environments and configuring the various elements.

Resum

L'objectiu principal d'aquesta tesi és el disseny i implementació d'una infraestructura de correu basada en el protocol SMTP per una empresa tecnològica que rep una gran quantitat de mails diaris els quals han de ser filtrats sense cap error i emmagatzemats als seus servidors. Volem aconseguir una solució adaptada a l'actualitat, que disposi d'una arquitectura molt robusta i que sigui capaç de funcionar molts anys amb un alt rendiment.

El desenvolupament del projecte es basa en l'anàlisi de l'antic sistema SMTP que ja s'estava quedant obsolet, i en buscar millores i alternatives per el nou disseny.

Finalment, s'han posat en pràctica els raonaments proposats a la memòria i s'ha implementat la infraestructura amb l'eina Microsoft Azure, simulant els entorns de treball i configurant els diversos elements.

Resumen

El principal objetivo de esta tesis es el diseño e implementación de una infraestructura de correo basada en el protocolo SMTP para una empresa tecnológica que recibe una cantidad muy grande de mails cada día, que tienen que ser filtrados y almacenados a sus servidores. Queremos una solución actualizada, con una arquitectura robusta y que sea capaz de funcionar por muchos años a gran rendimiento.

El desarrollo del proyecto se basa en el análisis del antiguo sistema SMTP que ya está obsoleto, y en buscar mejoras y alternativas por el nuevo diseño.

Finalmente se han puesto en práctica los razonamientos propuestos en la memoria y se ha implementado la infraestructura con la herramienta Microsoft Azure, simulando los entornos de trabajo y configurando los distintos elementos.

Reconeixements

Aquesta tesi no hauria estat possible sense l'ajuda d'en Christian, enginyer de software, qui m'ha ajudat i aconsellat en tot moment a tirar endavant el projecte tot i la meva inexperiència.

Per altre banda, també agrair la confiança al tutor de la universitat, José Luis Muñoz. Finalment, m'agradaria agrair també a la meva família, per ajudar-me en tot moment.

Revision history and approval record

Revision	Date	Purpose
0	01/06/2020	Document creation
1	03/06/2020	Document revision
2	15/06/2020	Document revision
3	23/06/2020	Document revision
4	27/06/2020	Document revision

DOCUMENT DISTRIBUTION LIST

Name	e-mail
Pau Espona Puig	Pau_espu@hotmail.com
Jose Luis Muñoz Tapia	Jose.munoz@entel.upc.edu

Written by:		Reviewed and approved by:	
Date	20/06/2020	Date	29/06/2020
Name	Pau Espona	Name	Jose Luis Muñoz Tapia
Position	Project Author	Position	Project Supervisor

Taula de continguts

Abstract.....	1
Resum	2
Resumen.....	3
Acknowledgements.....	4
Revision history and approval record	5
Table of contents.....	6
List of Figures	8
List of Tables:	9
1. Introduction	10
1.1. Context	10
1.2. Objectius i requeriments.....	11
1.3. Pla de treball, fites i Diagrama de Gantt	12
1.3.1 Tasques.....	12
1.3.2 Fites	16
1.3.3 Diagrama de Gantt	16
1.4. Desviacions i incidències	17
2. Coneixements bàsics.....	18
2.1. Protocol SMTP	18
2.2. Relay SMTP	20
2.3. Balancejador de càrrega	21
2.4. Col·lector de dades.....	21
2.5. Tecnologies emprades	21
3. Desenvolupament del projecte.....	22
3.1. Entorn de desenvolupament	22
3.2. Recuperació de dades	24
3.3. Anàlisi de l'arquitectura de relays actual ECOM	24
3.4. Inventari dels 'sender domains'.....	25
3.5. Estudi del flux de mail	27
3.6. Número de relays	28
4. Disseny	29
4.1. Disseny SMTP en l'entorn productiu	29
4.2. Disseny SMTP en l'entorn de qualitat	30
4.3. Definició regles d'exclusivitat i ACL's.....	30
4.4. Disseny de l'arquitectura del sistema	31
4.5. Anàlisi de riscos	32
5. Implementació	33
5.1. Simulació dels entorns	33
5.2. Afegim i configurem els servidors SMTP.....	34
5.2.1 Muntatge de la infraestructura.....	34
5.2.2 Creació màquina virtual a l'Azure.....	35
5.2.3 Configuració LB.....	35
5.2.4 Configuració servidors SMTP.....	36
5.3. Testejem el mail	38
6. Resultats.....	39
7. Pressupost	40
8. Conclusions i visió de futur	41
Bibliografia	43



Annexos	44
Glossari	49

Llista de figures

Figura 1: La globalització	10
Figura 2: Recorregut d'un correu de principi a fi	18
Figura 3: Funcionament del protocol SMTP	19
Figura 4: Balancejador de càrrega	21
Figura 5: Entorns de treball.....	23
Figura 6: Relays actuals	25
Figura 7: Dos típics distribuïdors de correu.....	26
Figura 8: Disseny de Production.....	29
Figura 9: Disseny de Quality	30
Figura 10: Disseny de l'arquitectura del sistema.....	32
Figura 11: Pàgina principal del portal Azure	33
Figura 12: Creant la màquina virtual	35
Figura 13: Instal.lant IIS 6	37
Figura 14: Visualització del servidor virtual SMTP	37
Figura 15: Panell de 'relay restrictions'	37
Figura 16: Provant el sistema.....	38

Llista de taules

Taula 1: Tasca 1	12
Taula 2: Tasca 2	12
Taula 3: Tasca 3	13
Taula 4: Tasca 4	13
Taula 5: Tasca 5	14
Taula 6: Tasca 6	14
Taula 7: Tasca 7	15
Taula 8: Fites.....	16
Taula 9: Diagrama de Gantt.....	16
Taula 10: Inventari dels ‘server domains’	26
Taula 11: Entorns reservats.....	34
Taula 12: IP’s assignades.....	36
Taula 13: Pressupost	40

1. Introduction

1.1. Context

No fa més de 50 anys, si et volies comunicar amb alguna persona lluny del teu abast, ho havies de fer per carta i esperar unes quantes setmanes per obtenir una resposta.

Què haguessin pensat els nostres avis, que en el futur es podrien comunicar amb qualsevol persona de qualsevol indret amb qüestió de segons? I no només això, sinó que serien capaços de enviar fotografies, música o fins i tot llibres electrònics? Tot això i moltes altres coses han estat possibles gràcies al **correu electrònic**.

L'evolució del correu electrònic ha estat, com en moltes altres tecnologies, de creixement exponencial. Els seus orígens es remunten l'any 1961, quan en una demostració del MIT, es va començar a perfilar un sistema per compartir informació des de dos terminals separats. 10 anys més tard, Raimon Tomlinson va incorporar l'ús de l'arrova, com a divisor entre l'usuari que enviava el correu i la commutadora on es guardava.

Anys després, ja es va normalitzar l'ús del mail i van anar sorgint empreses de distribució de correu, que actualment són gegants de les telecomunicacions.

Avui en dia, s'envien milions de correus cada segon i el flux de dades a la xarxa és gegant, i normalment, la informació que es conté en un mail, és molt important. Per això les empreses necessiten infraestructures robustes que suportin tants correus sense cap pèrdua.

Aquest projecte es basa en la renovació del sistema de mail d'una empresa, analitzant els punts febles i riscos de la vella infraestructura i dissenyant-ne una de nova. S'implementarà una nova xarxa de servidors de correu, basat en el protocol SMTP, per fer front a les necessitats actuals i a l'increment del flux de dades diàries entre l'empresa i l'exterior.

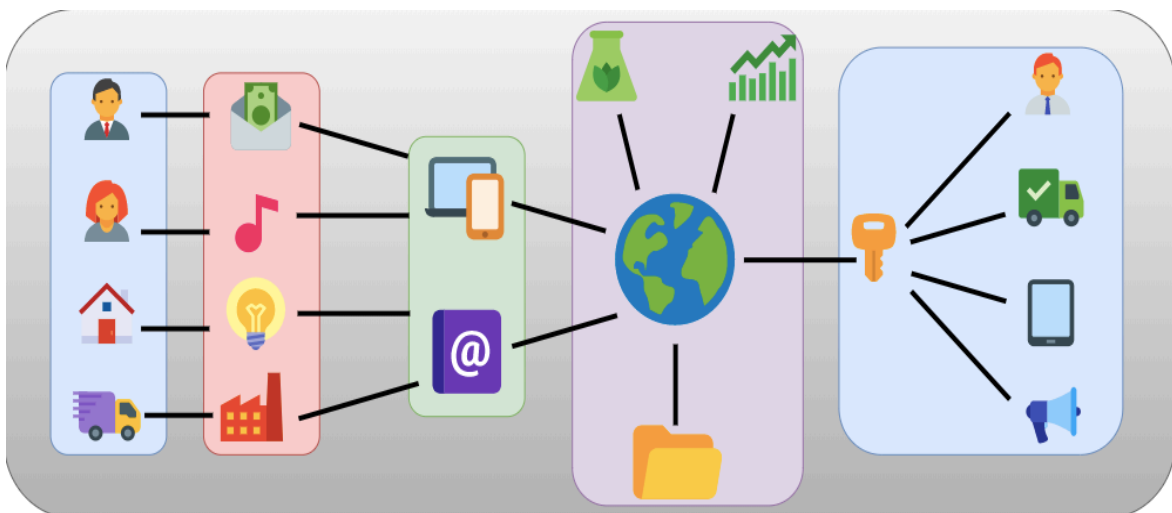


Figura 1: La globalització

1.2. Objectius i requeriments

L'objectiu principal del projecte és proveir un sistema de correu pels entorns productius i de qualitat d'una empresa. S'utilitzarà el protocol SMTP, el qual no en tenia cap coneixement al principi, així que a nivell personal em vaig marcar com a repte aprendre sobre aquest protocol i en conseqüència, aprendre sobre els diferents conceptes i eines que anirien sorgint al llarg del projecte.

Així doncs, els objectius del treball són:

- Millorar l'antiga infraestructura de correu de l'empresa analitzant els aspectes que es volen canviar i donar un nou enfocament a les coses que mínimament ja funcionen. En aquest sentit, ens centrarem sobretot en la neteja de '*sender domains*' no lícits i en un nou disseny òptim de relays.
- Assegurar l'emmagatzematge de les dades de cada mail que arriba, especialment el paràmetre 'FROM:' del domini que envia, 'TO:' del domini a qui va dirigit, i el cos del missatge.
- Aconseguir un balanceig equitatiu de la càrrega de bits entre els diferents servidors SMTP o relays.
- Implementació bàsica del disseny per comprovar que el mail filtra correctament.

Els requeriments per la correcta realització del projecte son els següents:

- Un portàtil amb Windows i navegació a internet.
- Connexió remota per tenir accés als 'server SMTP senders' de l'empresa.
- Subscripció a Microsoft Azure (s'ha fet servir una subscripció gratuïta d'un mes amb accés a les principals característiques del portal. També s'ha realitzat un curs online per aprendre a fer-lo servir).

1.3. Pla de treball, fites i diagrama de Gantt

1.3.1 Tasques

Project: SMTP relay	WP ref: 1	
Major constituent: Bases del projecte	Sheet 1 of 7	
Short description: Definir les bases del projecte, objectius i requeriments. Pensar quines eines hauré de fer servir.	Planned start date: 15/02/2020	Planned end date: 28/02/2020
	Start event: 15/02/2020 End event: 28/02/2020	
Internal task T1: Reunió amb el supervisor i fer el Planning. Internal task T2: Definir les eines pel desenvolupament del projecte.	Deliverables: Work plan	Dates: 28/02/2020

Taula 1: Tasca 1

Project: SMTP relay	WP ref: 2	
Major constituent: Coneixements bàsics i investigació	Sheet 2 of 7	
Short description: Investigar sobre el protocol SMTP i com funciona en qualsevol entorn (Windows servers, LB, SMTP relays, ACL's...). Aprendre a treballar amb aquestes tecnologies i conèixer les diverses aplicacions que poden realitzar.	Planned start date: 15/02/2020	Planned end date: 08/02/2020
	Start event: 15/02/2020 End event: 08/02/2020	

Internal task T1: SMTP protocol	Deliverables:	Dates:
Internal task T2: Aprendre sobre arquitectura de xarxa		

Taula 2: Tasca 2

Project: SMTP relay	WP ref: 3	
Major constituent: inventari ECOM	Sheet 3 of 7	
Short description: Analitzar i inventariar l'antiga infraestructura de relays per tenir una visió dels problemes i començar a pensar les solucions.	Planned start date:01/03/2020	Planned end date:21/03/2020
	Start event:01/03/2020 End event:21/03/2020	
Internal task T1: Entendre com funciona tota l'arquitectura actual	Deliverables:	Dates:

Taula 3: Tasca 3

Project: SMTP relay	WP ref: 4	
Major constituent: Estudi del flux de mail	Sheet 4 of 7	
Short description: Estudi diari del flux de mails a ECOM durant un període de 2 setmanes. Estudiar per cada servidor: clients, dominis que no interessin, spam...)	Planned start date:21/03/2020	Planned end date:05/04/2020

	Start event:21/03/2020 End event:05/04/2020	
Internal task T1: Identificar els 'sender domains' Internal task T2: Estudi de flux del mail (FROM, TO mails, bytes average...).	Deliverables:	Dates:

Taula 4: Tasca 4

Project: SMTP relays	WP ref: 5	
Major constituent: Disseny de l'infraestructura	Sheet 5 of 7	
Short description: Disseny a partir dels estudis anteriors.	Planned start date:10/04/2020 Planned end date:15/05/2020	
	Start event:10/04/2020 End event:15/05/2020	
Internal task T1:Identificar el número de relays necessaris. Internal task T2: Regles d'exclusivitat. Internal task T3: LogCollector. Internal task T4: Disseny complert	Deliverables:	Dates:

Taula 5: Tasca 5

Project: SMTP relay	WP ref: 6	
Major constituent: Implementació a l'Azure	Sheet 6 of 7	

Short description: Simular els relays SMTP a Azure i comprovar que el mail s'envia correctament.	Planned start date:20/04/2020	end
	Planned date:15/05/2020	end
	Start event:01/05/2020 End event:10/06/2020	
Internal task T1: Aprendre sobre Azure (Curs virtual i algunes pràctiques) i sobre el Windows server 2016. Internal task T2: Implementar entorns Q i P Internal task T3: Muntar els diversos elements Internal task T4: Configurar els servidors SMTP Internal task T5: Comprovar que el correu s'envia correctament. ...	Deliverables:	Dates:

Taula 6: Tasca 6

Project: SMTP relay	WP ref: 7	
Major constituent: Documentació	Sheet 7 of 7	
Short description: Escriure la memòria del projecte.	Planned start date:01/06/2020 Planned end date:17/06/2020	
	Start event: End event:	
Internal task T1:Escriure la memòria.	Deliverables:	Dates:

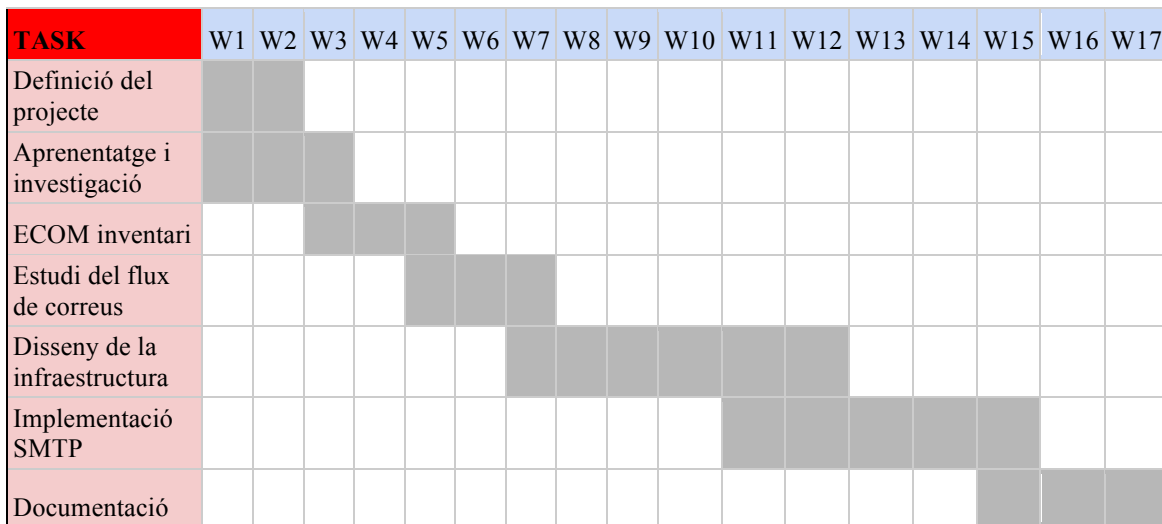
Taula 7: Tasca 7

1.3.2 Fites

WP#	Task#	Short title	Milestone / deliverable	Date (week)
5	5	Disseny	Disseny complet	15/05/2020
6	6	Implementació	Muntatge Azure realitzat	10/06/2020
7	7	Memòria	Memòria final	28/06/2020

Taula 8: Fites

1.3.3 Diagrama de Gantt



Taula 9: Diagrama de Gantt

1.3.4 Desviacions i incidències

Aquest TFG ha estat realitzat a la primavera del 2020, any en que una pandèmia mundial va provocar una greu crisi sanitària i va obligar a la gent a tancar-se a casa. Durant uns mesos, es va paraitzar tota activitat econòmica i laboral que no es pogués realitzar telemàticament. En el meu cas, també m'hi vaig veure afectat, ja que a causa de la forta caiguda de feina i la dificultat de donar material de suport des de casa, l'empresa va prendre la decisió d'apartar temporalment els alumnes en pràctiques. Per tant no vaig acabar el projecte a la feina tal com estava previst.

Per sort, quan va esclatar el COVID-19, gran part del desenvolupament del projecte ja l'havia fet, o almenys la part essencial, la qual havia de fer-la sí o sí a l'empresa, perquè treballava amb dades que havia de buscar dins els seus servidors. Així que en cap moment em vaig quedar sense poder fer alguna tasca imprescindible.

La part del disseny, la vaig realitzar des de casa ja amb tota la informació recollida, i fins i tot vaig mantenir el contacte amb el meu cap, el qual m'anava aconsellant una mica sobre com avançar.

La part del projecte que més s'ha vist afectada, i així ha quedat reflectit a la revisió crítica i en els WP, ha sigut la implementació. La idea original era muntar el disseny a la xarxa de l'empresa com a projecte real, però degut a la situació comentada vaig haver de buscar una alternativa des de casa. Després d'analitzar les necessitats i requeriments que havia de cobrir, vaig trobar l'eina Microsoft Azure la qual s'adaptava perfectament al que jo volia. De fet moltes empreses fan servir aquesta aplicació així que és molt professional. A més, en la seva versió de prova gratuïta, disposava de tots els elements indispensables pel projecte, així que no m'ho vaig pensar més i vaig començar a aprendre sobre Azure.

Analitzant la situació ara que he acabat el projecte, crec que vaig prendre una bona decisió, i tot i que crec que si ho hagués fet tal com estava previst al principi, encara hauria après molt més, estic molt content del resultat final.

2. Coneixements bàsics

Per començar, s'estudiaran els conceptes primordials del projecte per tenir una idea bàsica del que farem. Es vol tractar des de un punt de vista teòric però donant també un enfocament a com es farà servir cada element per la implementació de manera pràctica.

2.1. Protocol SMTP

SMTP es refereix a les sigles '*Simple Mail Transfer Protocol*', que tal com indica el nom, és un protocol molt simple per enviar mails. Quan un usuari envia un correu, aquest es guarda al *sender SMTP server* que és el servidor de l'empresa que distribueix el mail, i mitjançant internet s'envia al *receiver SMTP server* corresponent al distribuïdor del receptor. Aquest, utilitza un protocol anomenat POP o IMAP perquè el mail arribi a la bústia d'entrada de l'usuari remitent i així es tanca el cicle. En aquest projecte, es tractarà el pas intermig, en que l'empresa rep correus dels seus clients.

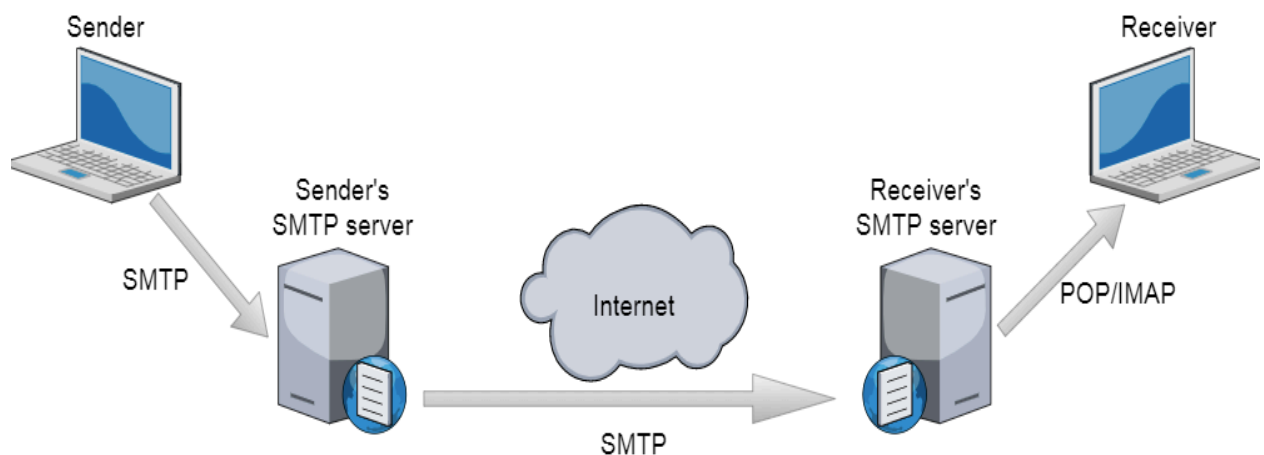


Figura 2: Recorregut d'un correu de principi a fi

El funcionament del protocol és molt senzill, consisteix en un seguit de comandes i paràmetres que donen permís i dirigeixen el tràfic de correus entre el servidor SMTP que envia i el servidor SMTP que remet. En el següent esquema es mostra pas per pas com funciona la connexió '*sender - receiver*':

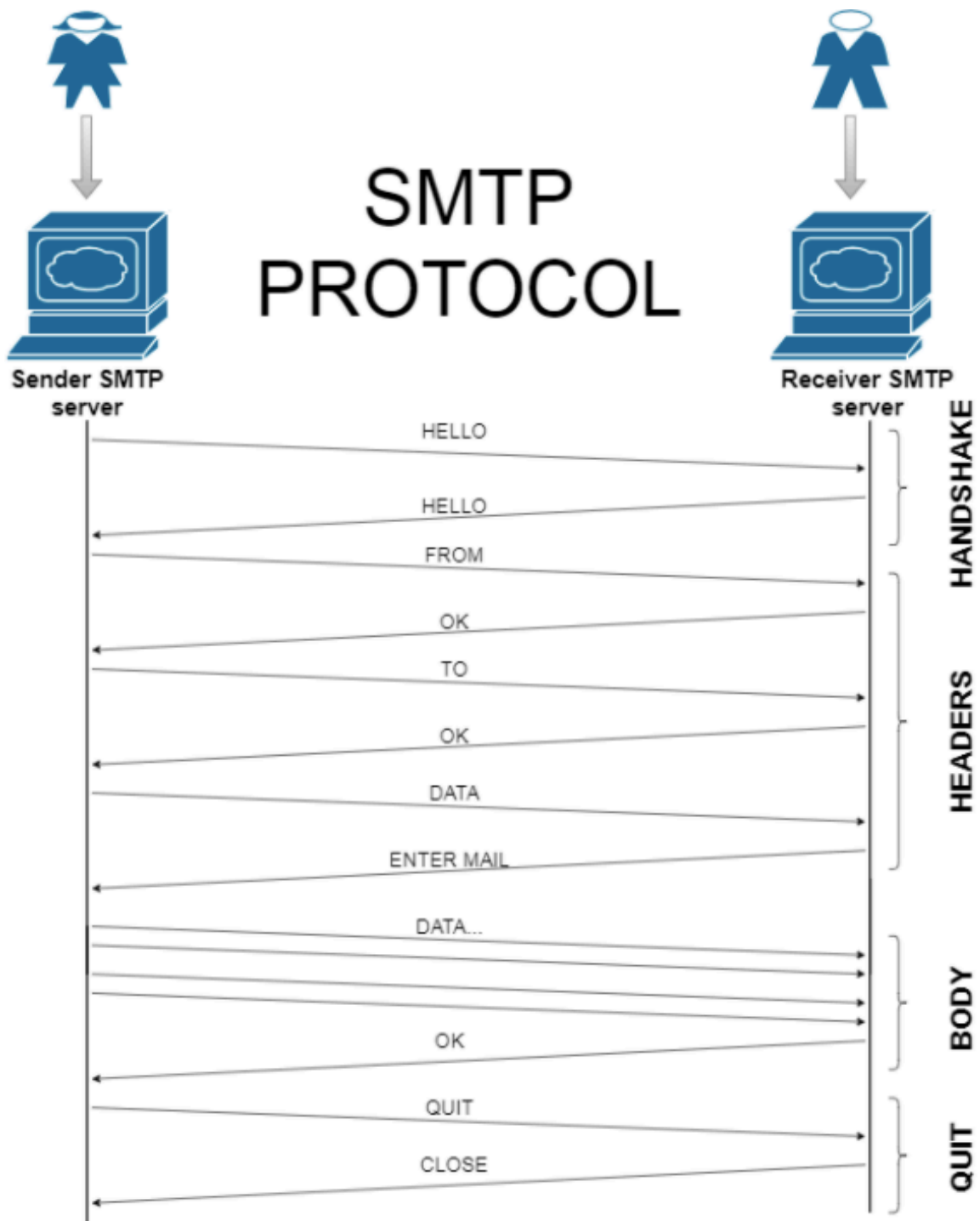


Figura 3: Funcionament del protocol SMTP

Després del *handshake*, un cop s'ha detectat el paquet, s'ha donat per vàlid, i s'ha establert la connexió entre servidors SMTP, l'emissor comunica al receptor qui està enviant el mail i a on a partir del paràmetre 'FROM:' i 'TO:'. Aquesta informació serà de gran valor per la realització del disseny.

El paràmetre DATA també és molt important, ja que avisa del pes del missatge, que normalment es limita a certs KBytes, per no saturar el sistema. Seguidament es procedeix a enviar el cos del mail amb els paquets que calgui. I finalment, es tanca la connexió.

Tota aquesta informació que com s'ha dit és molt important, queda guardada al '*receiver sender*' en forma de **log**, i així es pot recuperar quan sigui necessari. Els logs són arxius de gran capacitat que emmagatzemen aquestes dades en períodes d'un mes.

SMTP fa servir el protocol TCP per la transmissió de dades a través del port 25. En aquest punt, les **ACL** (l·listes de control d'accés) actuen com a filtre de seguretat per paquets d'origen sospitós denegant-los el pas.

TCP garanteix l'entrega del mail assumint que la direcció de destí és correcte. Si per algun motiu, el mail no arribés al destí, s'envia un missatge d'error d'entrega (*mail delivery error*).

2.2. Relay SMTP

Els **relays SMTP** és la forma que d'ara en endavant anomenarem al '*receiver SMTP server*'. Els relays s'encarreguen de recopilar tots els correus que arriben a l'empresa i a partir dels protocols POP o IMAP filtrar-los a on hagin d'anar.

En aquest punt, podem definir dos conceptes que seran molt importants d'ara endavant, i a vegades poden semblar una mica confusos:

- El **domini** d'un correu és la part que segueix de '@hotmail.com', i que en general, cada empresa té personalitzat amb el seu nom en forma de correu corporatiu. En aquest projecte, per reconèixer si el mail prové d'un client o d'un altre, es fa servir el domini.
- El '**distribuïdor de correu**' és l'empresa que prové tota la infraestructura de correu a les empreses que volen usar el seu propi correu corporatiu. Per exemple 'Microsoft Outlook' o 'Gmail' donen servei a partir del seu propi sistema de mail a les empreses o usuaris personals que ho necessitin. Aquest terme és important, perquè els '*sender SMTP servers*' s'organitzen per distribuïdors de mail: així que cada client, independentment del domini que tingui, vindrà d'un '*sender SMTP server*' o un altre, en funció del distribuïdor que faci servir.

Cal afegir que els clients, alhora que els distribuïdors de correu que fan servir, són confidencials i no ens podem referir pel seu nom, en el desenvolupament del projecte s'explica un mètode per referir-se a ells.

Les **regles d'exclusivitat** són unes normes que es poden definir quan configures els relays per tal de crear filtres en els correus d'entrada i així permetre o denegar l'accés al sistema del domini que especifiquis. Aquesta eina serà molt potent per dissenyar els nostres futurs relays.

2.3. Balancejador de càrrega

Aquest concepte es refereix a la distribució de tasques als recursos disponibles, amb l'objectiu de que el sistema treballi de manera eficient. Un bon balanceig de càrrega, pot optimitzar al màxim els nostres relays, evitant que un tingui una sobrecàrrega de feina mentre un altre no té res a processar. Així doncs, en el nostre projecte el farem servir per distribuir el flux dels mails en els diferents servidors SMTP. Els LB fan servir un protocol d'**alta disponibilitat (High Availability)**, que assegura un grau absolut de continuïtat operacional, és a dir, que fins que nosaltres no diguem el contrari, el balancejador serà operatiu sempre i per tant la infraestructura no patirà cap sobrecàrrega.

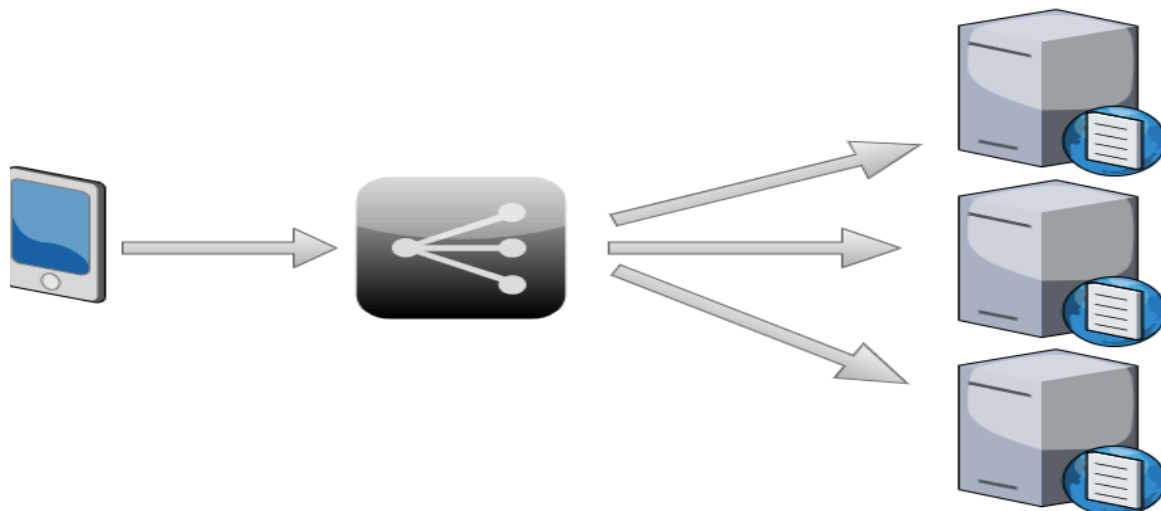


Figura 4: Balancejador de càrrega

2.4. Col·lector de dades

El **LogCollector** serà el nostre centre d'emmagatzematge o *database* on es guarden els **logs** dels correus. Aquest col·lector de logs és de vital importància, ja que els clients poden enviar qualsevol cosa: factures, informes, dades multimèdia, i en cap cas, l'empresa no es pot permetre la pèrdua d'aquestes correus per el motiu que sigui. Per això, caldrà dissenyar aquest sistema per si s'ha de fer una recuperació de dades.

2.5. Tecnologia emprada

Microsoft Azure és una col·lecció de serveis informàtics integrats al núvol (que conté, entre d'altres, eines referents a l'anàlisi, el processament, les bases de dades, les xarxes de gestió d'emmagatzematge, la gestió web...) que farem servir per simular els entorns de treball i posteriorment muntar el disseny complet de la infraestructura SMTP.

És una plataforma de pagament, però disposem d'una subscripció gratuïta d'un mes amb les funcionalitats bàsiques però que ens serà suficient per la nostra implementació.

Per simular els servidors, hem fet servir màquines virtuals operant amb **Windows Server 2016** que és un sistema operatiu desenvolupat i publicat per Microsoft i s'adapta perfectament a les nostres necessitats.

3. Desenvolupament del projecte

Primer de tot, és molt important entendre com funciona l'entorn en el que estem treballant per tal de poder desenvolupar correctament qualsevol part del projecte. Així doncs, també és vital conèixer la vella infraestructura i tots els elements que la conformen (estructura, funcionalitats i arquitectura del software), ja que és d'on extraurem les dades i evidències necessàries per poder fer el **disseny** de la nova.

Per tant, seguidament en farem un anàlisi d'on es vol explicar el perquè s'ha quedat obsoleta i obtindrem informació essencial del que ha de contenir la nova fent un estudi del flux del mail diari en una època amb molta activitat.

3.1. Entorn de desenvolupament

L'empresa s'organitza en dos entorns bàsics i diferenciats. És una empresa gegant i està present a tot el món, però nosaltres farem referència principalment a la nostra regió. Cada entorn conté les seves aplicacions i serveis propis, i per això es necessita una capa forta de correu electrònic que ho comunicui tot.

El primer s'anomena **EMEA** (ja que conté servidors a les regions d'Europa, Orient Mitjà i Àfrica), i s'encarrega de donar el servei als clients. Els servidors emmagatzemen milions de dades, i el flux de mails és gegant, molt superior a l'altre entorn. En aquest cas, el sistema de mail SMTP ja està renovat degut al constant creixement de l'empresa i la necessitat de que aquest progrés sigui conscient i adaptat a l'actualitat de l'empresa.

El segon s'anomena **ECOM** (s'organitza igual a nivell global) i conté tota la informació econòmica de l'empresa. És on s'executen totes les transaccions de diners, actualitzacions de preus en els productes, i on es guarden totes les factures. Aquí és on centrarem els esforços, ja que l'objectiu del projecte és dissenyar un nou model de relays SMTP que permetin una bona comunicació per mail de totes les aplicacions i eines dins d'aquest entorn.

Amb això, no em refereixo a que s'ha d'aconseguir que s'enviïn bé els mails i ja està, sinó que cal investigar, a partir d'aquesta vella infraestructura, com podem millorar la nova. A partir d'això, ens sorgeixen diferents preguntes com:

- Quin disseny té la vella infraestructura? Quants relays té i com estan disposats?
- Quines aplicacions/usuaris són les que envien més correus? N'hi ha que ja no es fan servir però encara estan dins el sistema, i per tant suposen una despesa de recursos?
- El flux de correus és només intern o també arriben mails externs? Cal fer alguna llista negra o desfer-se de spam innecessari?
- Hi ha algun relay que hagi quedat obsolet?
- En general, i per la situació actual de l'empresa, és òptima la infraestructura? És a dir, amb el número de mails diaris, i el recorregut que segueixen abans d'arribar al *receiver SMTP server* passant pel relay que sigui, i tenint en compte els bits que gasta cada mail, està tot ben equilibrat?

Dins de cada un dels dos entorns esmentats, existeixen també dos sub-entorns o ‘espais de treball’. Aquest són el de **QUALITY** i **PRODUCTION**. Això és molt necessari en les grans empreses, perquè hi ha constantment nous projectes que s’han de testear i provar en local i on poder tenir un marge d’error abans de llençar al mercat. O simplement s’han d’afegir noves funcionalitats a un servei ja existent, i aquestes s’han de desenvolupar en un entorn sense riscos.

Així doncs, l’entorn Quality o de qualitat, correspon al conjunt de servidors i aplicacions de l’empresa on es fan totes aquestes proves, i on es desenvolupa qualsevol nova funcionalitat abans de passar a ser operativa. En el nostre cas farem el disseny per Q, adaptat a les necessitats d’aquest entorn, i ens servirà de base per després fer el mateix a P.

Per altre banda, hi ha l’entorn productiu. Aquí és on hi ha l’activitat real del dia a dia (mails amb clients, fluxos de milions de dades, transaccions econòmiques...). Conté moltes més dades que Q, molts més servidors i recursos, i òbviament és el més important, ja que inclou informació transcendental per l’empresa. És per això, que només quan un projecte ha estat testat amb èxit i ha passat uns requisits de qualitat, es desplega a P.

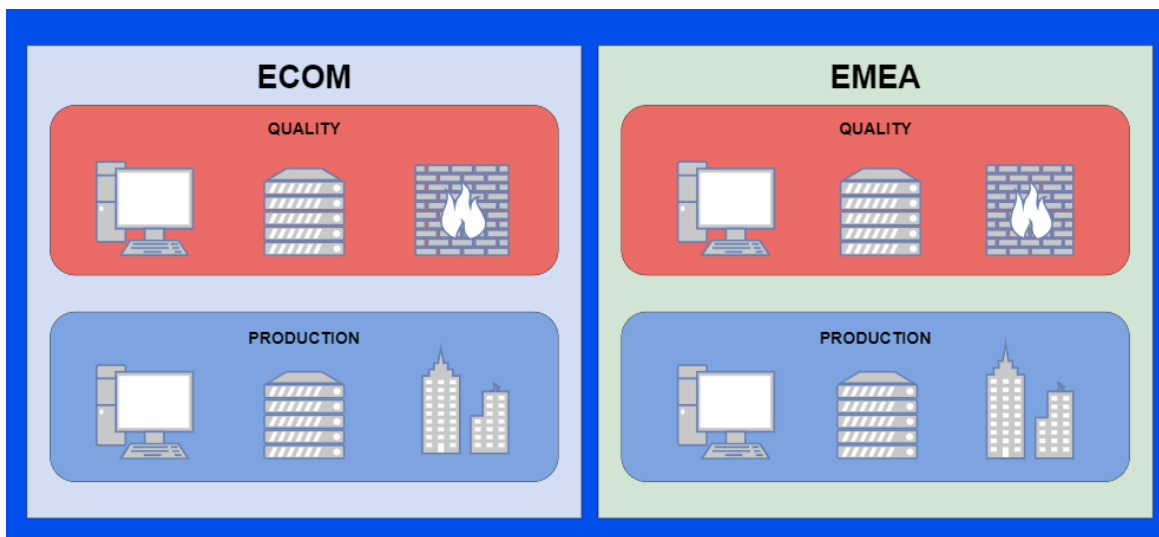


Figura 5: Entorns de treball

3.2. Recuperació de dades

Per tant, com s'ha explicat, podríem dir que l'entorn de Q conté rèpliques de P, i això ens suposa molt interessant a l'hora de fer el disseny, perquè podem fer servir l'entorn de Qualitat com a guardaespalles del sistema. És a dir, si un dia cau l'entorn productiu, i per tant, no s'hi poden guardar les dades, podem fer que aquestes s'emmagatzemin a Q, i així evitar perdre informació important del client (cosa que suposaria un problema important).

De la mateixa manera, es podria fer el procés enviant les dades a EMEA, ja que conté un datacenter exclusiu per aquest tipus d'incidències, i segurament ens costaria menys recursos, ja que no hauríem de desplegar cap col·lector de dades específic. Per altre banda, la comunicació EMEA-ECOM és més complexa i tenim més risc de pèrdua de dades pel camí.

S'haurà d'avaluar quina és la opció més òptima, tot i que per motius de recursos, la segona fa més bona pinta. El que està clar és que hem de tenir un sistema molt efectiu per emmagatzemar totes les dades, ja que és molt probable que si algun mail no arribi al destí per algun motiu, s'hagi de comprovar que realment es va enviar.

3.3. Anàlisi de l'arquitectura actual ECOM

A partir d'aquí, l'objectiu d'aquest apartat és intentar resoldre els dubtes i preguntes que han sorgit i intentar millorar-ho. Volem obtenir totes les especificacions per tal de més endavant realitzar el disseny.

L'arquitectura operativa actualment, i que volem canviar, consta de 5 relays SMTP connectats per un balancejador de càrrega que dirigeix el tràfic. L'empresa ens ha proporcionat aquesta informació:

- 3 dels 5 relays funcionen correctament. Filtren un 85% dels mails totals, però no de manera òptima. És a dir, no tenen una càrrega de treball equitativa i el sistema ha tingut alguna 'caiguda' important a causa de la sobrecàrrega d'alguns d'aquests servidors. Al filtrar tants correus, es creu que hi ha molts *senders* amb 'dominis sospitosos' o que no se sap quina relació tenen amb l'empresa ni perquè s'estan rebent, però que ningú mai s'hi ha preocupat perquè tampoc fan cap mal. Aquest tipus de dominis, els hem d'identificar i eliminar, així s'aconseguirà un estalvi de recursos.
- Un altre relay conté únicament spam. El supòsit és que, en el pas del temps, els enginyers encarregats de mantenir el sistema, han anat re-direccionant aquest tipus de correu per aquest servidor, que s'ha quedat rellevat únicament a aquest ús. També es creu que és un relay que al seu dia no va ésser ben configurat, i ha set víctima de hackers o enviaments massius d'spam. Tots aquests '*sender domains*' els enviarem a la llista negra i els eliminarem del sistema.
- Finalment, l'últim està completament obsolet, i no s'hi registra cap moviment des de fa anys.



Figura 6: Relays actuals

3.4. Inventari dels ‘sender domains’

Seguidament estudiarem els ‘sender domains’ de l’entorn ECOM, amb altres paraules, els emissors dels correus que arriben. Ens basarem en els *logs* o arxius de memòria dels ‘*sender SMTP server*’ durant dues setmanes de forta activitat laboral. Això vol dir que haurem d’accedir dins d’aquets servidors, i recuperar aquesta informació que està emmagatzemada en el seu disc. Cal dir que aquest procés consta d’un gran treball d’investigació i moltes hores d’anàlisi utilitzant tècniques d’Excel. Aquí es mostrarà els resultats de manera resumida i concisa.

Així doncs, obtindrem les principals aplicacions i usuaris (clients) que envien al sistema, tindrem un control dels principals clients i sabrem quins dominis podem eliminar, cosa que ens serà de molta utilitat per el disseny posterior. Cal deixar clar, que al parlar de client, es fa referència a molts usuaris enviant correus des de el mateix domini, ergo corresponen al mateix client.

S’ha realitzat un inventari tant per P com Q, especificant el ‘*sender SMTP server*’ d’on prové i el nom del client.

Com que els clients són confidencials, parlarem del **clientA**, **clientB** i així successivament referint-nos als dominis de mails que usen els clients. Òbviament, també tindrem correus interns de l’empresa, d’aquests en direm ‘**internDomain**’. Els dominis identificats com a fraudolents o d’spam, els anomenarem ‘**spamDomainX**’. El ‘*sender SMTP server*’ també és confidencial, i el definirem en funció del distribuïdor de correu que faci servir (**serverDIST1...serverDIST6**).

Seguidament representem una taula amb l’inventari en qüestió, tasca que com hem dit, es realitza investigant els *logs* dels ‘*sender SMTP server*’. Val a dir que és només una mostra representativa i es té en compte aquells dominis/clients que són, amb diferència, els més actius.

PRODUCTION					
serverDIST1	serverDIST2	serverDIST3	serverDIST4	serverDIST5	serverDIST6
clientA	clientE	clientG	clientK	clientL	spamDomain6
clientB	clientF	clientH	spamDomain3	clientM	spamDomain7
clientC	internDomain	clientI	spamDomain4	clientN	-
clientD	spamDomain1	clientJ	spamDomain5	-	-
spamDomain1	-	-	-	-	-
spamDomain2	-	-	-	-	-
QUALITY					
serverDIST1	serverDIST2	serverDIST3	serverDIST4	serverDIST5	serverDIST6
clientAQ	clientEQ	clientGQ	clientKQ	clientLQ	-
clientBQ	clientFQ	-	-	-	-
clientCQ	internDomain	-	-	-	-

Taula 10: Inventari dels 'sender domains'

D'aquestes taules, se'n poden deduir i concloure varies realitats:

- Cada client fa servir un únic distribuïdor de mail. Per posar un exemple, el clientA fa servir la distribució serverDIST1.
- Tenim spam procedent de moltes vies. És difícil de saber perquè arriba, ja que es pot causar per moltes raons, però almenys el tenim controlat. Serà important a l'hora de definir les ACL's i llistes negres dels nous servidors SMTP.
- Hi ha 'senders SMTP server' que tenen més clients que d'altres, però això no vol dir que enviïn més mails. Simplement pot ser que un client generi molts mails diaris i utilitzi una distribució que no és usual. Aquesta part l'estudiarem a continuació.
- A Quality hi ha flux de mails amb menys clients, i no hi ha spam. Això és degut a que és un entorn on es treballa a nivell intern, i per tant té poc contacte amb 'l'exterior'.



Figura 7: dos típics distribuïdors de correu

3.5. Estudi del flux de mail

A partir d'aquí, ja coneixem quins són els principals clients o '*senders*'. Ara cal anar més enllà i fer un estudi en concret per cada client per tenir una idea de quants mails envia (i sobretot quan ocupen de memòria).

L'objectiu d'aquest apartat és aplicar més endavant els resultats al disseny. Principalment, si sabem el flux d'informació del sistema, sabrem **quants relays** cal posar-hi. Aquesta informació, juntament amb la que ja hem treballat anteriorment, ens donarà una idea ja bastant precisa de com serà l'arquitectura de la nostra infraestructura.

Per realitzar l'anàlisi del flux de mails, també s'han fet servir els *logs* o arxius de memòria dels '*sender SMTP server*', ja que hi queda registrat, per cada mail, quan ocupa en memòria. Així doncs, l'estudi consta en calcular el número de mails diaris de cada client i la memòria total en bytes que aquests ocupen. Agafarem de referència una època de forta activitat de dues setmanes i en farem la mitjana. S'ha decidit així, perquè hi ha dies on la feina és molt superior que d'altres, com per exemple un dimarts i un diumenge. I així es tindrà una visió més general i definida de l'activitat diària.

Un aspecte important a comentar, es que un mail pot ser enviat per una persona, però dirigit a més d'una. En el nostre cas, ens interessa només l'emissor del mail, per tant, ens fixarem en el camp 'FROM:' i no el 'TO:'.

Com en el cas anterior, estudiarem a part els casos per P i Q. Analitzem primer en el cas productiu. Tenim les taules resultants de l'estudi a l'annex.

Així doncs, calculem que tenim aproximadament uns **27000 mails diaris**. Com s'ha comentat, aquests corresponen als clients majoritaris, així que també s'haurà de tenir en compte que hi ha clients més petits que també envien (tot i que a una escala molt menor) i per tant afegirem un **marge de 10000 mails** més. Per anar segurs, i sempre mirant a l'alça, decidim que treballarem el disseny per suportar uns **40000 mails diaris**.

En mitjana, cada mail ocupa 90 bytes, per tant, el sistema haurà resistir un flux de **3,6 MBytes** diaris.

Seguidament analitzem el cas per l'entorn de qualitat. Podem observar que tenim **16317 mails al dia**. Com en el cas anterior, reservem un **marge de 10000 mails** per els clients petits i extraïem que el sistema filtra aproximadament uns **26000 mails diaris**. O sigui, **2,34 MBytes**.

3.6. Número de relays

EMEA té un flux molt major que ECOM. Allà hi han 75000 mails cada dia i es gestionen 6,75 MBytes, que és gairebé el doble del que ha de suportar la nostra plataforma. Sabem que el mail a ECOM funciona perfectament i es filtra equitativament amb 5 relays, on cada un prové una cinquena part del flux total diari (1,35MBytes).

Per tant ens basarem en la mateixa estratègia per P: observem que els 3,6 MBytes diaris que ens ocupa el sistema, es poden gestionar perfectament amb 3 relays, amb els quals es pot omplir fins a 4,05 MBytes (1,35 MBytes x 3). Així doncs, encara sobrarà un petit marge per incidències, o per si l'empresa creix i necessita més memòria.

Amb la mateixa lògica procedim per Q: Utilitzarem 2 relays, que disposen d'un buffer total de 2,7 MBytes. El sistema requereix 2,34 MBytes, així que igualment, es gaudirà d'un petit marge per incidències o futures transmissions.

4. Disseny

En aquest apartat, ens basarem amb tota la informació estudiada anteriorment per fer el disseny de l'estructura SMTP. Com en el llarg del projecte, es diferenciarà els dos entorns Q i P, tot i que aquests estaran en permanent connexió.

Primer de tot, es farà el disseny funcional basat en SMTP. L'objectiu és la millor distribució possible de la càrrega de treball amb els diferents relays, i de manera precisa: a partir de filtres i regles d'exclusivitat en la capa del protocol SMTP.

A partir d'aquí, es realitzarà el disseny de l'arquitectura amb una visió més general del sistema. S'analitzaran les connexions P-Q i EMEA-ECOM, la gestió del *logCollector* i els possibles riscos i com prevenir-los.

4.1. Disseny SMTP en l'entorn productiu

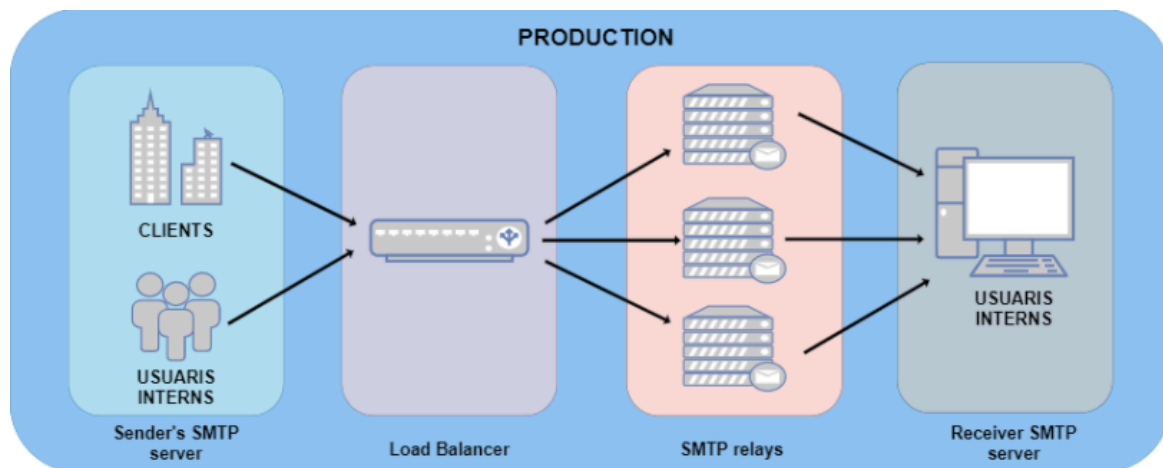


Figura 8: Disseny de Production

Aquest és l'esquema del curs que seguirà un mail des de el servidor SMTP de la part emissora fins arribar al servidor SMTP de la part receptora. En el nostre cas, l'agent emissor serà un client extern o un usuari intern de la pròpia empresa.

Cada mail, abans d'arribar a un relay perquè es dirigeixi al seu destí, ha de passar per un balancejador de càrrega (LB). Aquest conjunt d'alta disponibilitat, permetrà fer aquest balanceig a la xarxa, mantenint una equitat de càrrega per els 3 relays. Així aconseguirem generar una rotació de dades igual per tal de prevenir sobrecàrregues innecessàries.

A més, si un servidor deixa de funcionar per el motiu que sigui, automàticament es redirigirà el flux de dades cap a un dels altres dos.

Seguidament, el mail passa al relay, on es filtra segons les regles d'exclusivitat que hem definit per el protocol SMTP i s'envia a l'usuari (receptor). El mail i la seva informació (*logs*) queda emmagatzemat al servidor SMTP del receptor, o sigui, de l'empresa. D'aquí és on sorgeix la necessitat de tenir un 'col·lector de mails' o *logCollector* que actui com a database. Més endavant es tractarà aquest tema.

4.2. Disseny SMTP en l'entorn de qualitat

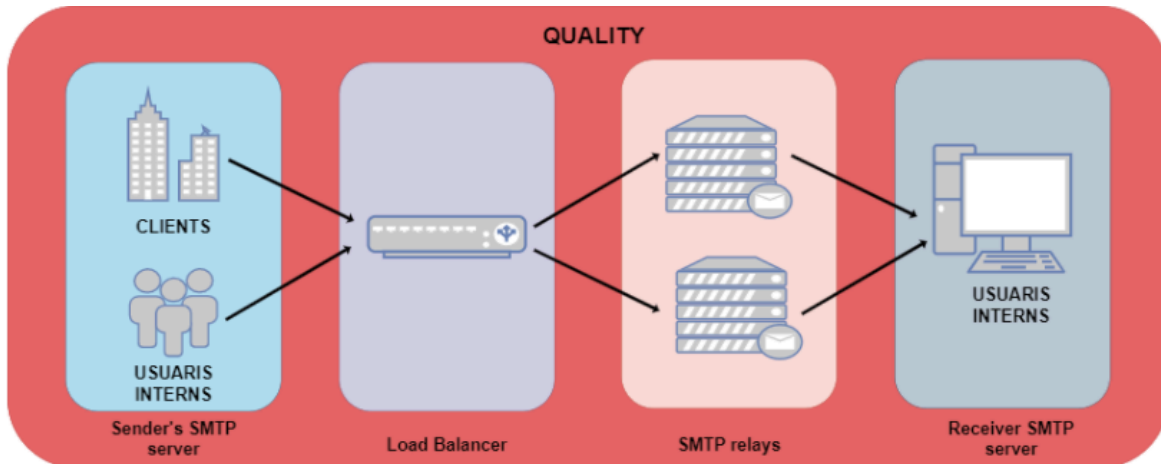


Figura 9: Disseny de Quality

Segueix exactament el mateix procediment que en l'anterior cas, encara que degut al decrement de flux de dades, es redueix de 3 a 2 relays.

4.3. Definició regles d'exclusivitat i ACL's

Una de les característiques més importants d'un bon sistema de mail, és que tingui definides unes bones regles d'exclusivitat. Actualment, s'estima que un **90% dels mails que circulen per tot el món són spam**, dirigit o bé per hackers intentant malmetre el teu sistema o per empreses enviant quantitats ingestes de spam i publicitat a tort i a dret. Òbviament, no ens interessa gens rebre aquest tipus de correus, perquè no volem ser víctimes de cap hacker ni volem gastar tota la memòria disponible amb mails de spam. Així que hem d'estar ben protegits i sempre anar actualitzant una llista negra per si algun cas no desitjat aconseguix perforar el sistema.

Es diferenciaren dos tipus de mails conflictius:

1. **A nivell de protocol SMTP i TCP:** són generalment mails enviats per hackers intentant malmetre el sistema. Quan un paquet de dades entra pel port 25 a partir de la connexió TCP procedent d'internet, es fa una selecció primària de paquets que poden semblar fraudulents i paquets estàndards o normals dins d'un protocol SMTP. Aquesta selecció està definida per les **ACL** (Llistes de Control d'Accés), i el que fan, en trets generals, és denegar el pas als paquets que no semblen lícits i donar permís als que sí ho són. Així doncs, es pot afirmar que les ACL són la primera capa de control de seguretat del tràfic de dades provinent de la xarxa i el primer filtre de seguretat del sistema.
Aquest projecte, no es desenvolupa en aquestes capes tant baixes de protocol, i tampoc estem habilitats per configurar aquests aspectes bàsics i estàndards de l'empresa. Així que s'utilitzarà la configuració ja proposada, la qual ja filtra correctament.

2. A nivell d'empresa : Els mails que s'han anat identificant al llarg del projecte. El nostre objectiu és fer una neteja de tots aquests dominis fraudulents que no interessin, i d'aquesta manera oxigenar notablement el sistema. Per tant caldrà aplicar aquestes **regles d'exclusivitat** quan es faci la implementació i configuració de cada relay. Es disposa de diversos perfils per tractar:

- **SPAM:** Apart del que hem detectat nosaltres en el desenvolupament del projecte, cal demanar si l'empresa conté una *blacklist* o llista negra on es van acumulant els dominis que envien spam. En aquests usuaris se'ls hi denegarà l'entrada a cap relay implicant així que puguin enviar cap més correu.
- **RESTRICTED SENDERS:** També és bastant freqüent que les empreses no mantinguin un bon contacte amb una altre empresa o usuaris particulars. En aquest cas, tenim dominis individuals dels quals no volem tenir cap contacte i n'hem de prohibir la seva entrada.
- **RESTRICTED CLIENTS:** Per altre banda, si es té algun client que es podria considerar VIP o primordial a l'hora de rebre els seus mails, es pot fer que per aquest domini concret es filtri únicament per un relay, evitant així una mínima possibilitat de pèrdua del mail. És a dir, en el cas de que l'usuari no rebí l'esperat mail a la seva bústia d'entrada per algun motiu, es sabrà que aquest mail únicament es pot filtrar per un relay, i per tant, es podrà anar directament a ell a investigar què ha passat.

4.4. Disseny de l'arquitectura del sistema

De moment s'ha realitzat el disseny de manera independent, sense tenir en compte les connexions internes de cada entorn amb cada element i com s'envien les dades al *logCollector* (o base de dades on es guarden els correus).

Aquest apartat doncs, té l'objectiu d'abordar les connexions P-Q i ECOM-EMEA per tal que sigui possible un flux de dades entre ells i així es puguin emmagatzemar al *logCollector*.

Abans de res però, cal analitzar on situarem el col·lector i quines repercussions té. Està clar, que el procediment més senzill seria afegir-ne un a P i un a Q, i oblidar-se del problema. Però un *database* és un cost elevat de recursos per l'empresa, així que només tenim pressupost per un.

Cal tenir en compte, que l'entorn productiu és el que fa guanyar diners a l'empresa i per tant és realment l'important i el que necessita ser cobert, així que s'ha de prioritzar. Seguint aquesta lògica, es procedeix a situar el *logCollector* a P. Els mails que arriben a Q, també han de ser desats, encara que no són prioritaris perquè no suposaria una pèrdua real de diners. Per tant s'ha de crear un pont (*virtual network peering*) entre els dos entorns per aconseguir intercanvi de dades.

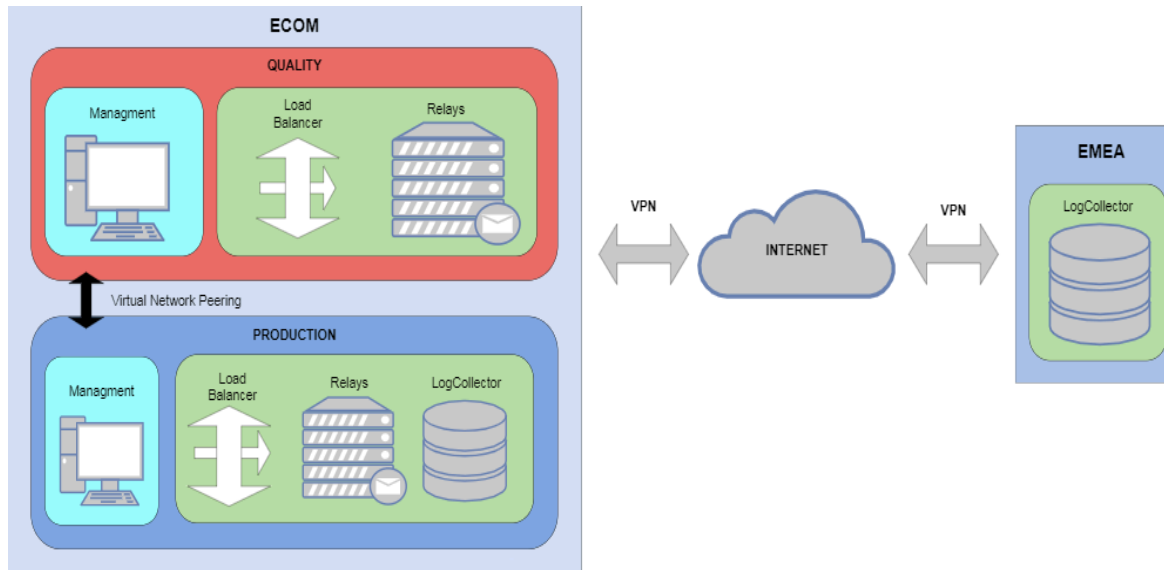


Figura 10: Disseny de l'arquitectura del sistema

4.5. Anàlisi de riscos

Un entorn de producció és habitual que quedi inoperatiu: sempre estan en constant manteniment, i per tant, pot ser que durant unes hores no pugui funcionar com habitualment o simplement hi hagi algun error inesperat que pot desencadenar la caiguda del sistema. Si passa això, les dades no es poden guardar correctament i es perd una informació important per l'empresa que en el pitjor dels casos suposaria grans pèrdues econòmiques.

La solució a aquest problema és tenir el suport del *logCollector* d'EMEA, el qual se'ns ha donat permís per fer servir. S'haurà de traçar una connexió per l'intercanvi de dades a nivell ECOM-EMEA per poder accedir a la *database* quan sigui necessari. Aquest pont es farà per internet amb una VPN.

És molt possible que en el temps de canvi de *database*, que pot durar uns minuts, ja que la informació es transmet per internet, arribin mails que es perdin per complet. És un risc menor que s'ha d'assumir i amb el qual no s'hi pot fer res.

Pel que fa a la possible caiguda de l'entorn de qualitat, les dades ja s'estan guardant a P, així que els mails poden anar arribant sense risc de pèrdua i per tant es poden recuperar més tard.

5. Implementació

Un cop s'ha realitzat el disseny, cal implementar-lo perquè sigui operatiu. Al no haver pogut utilitzar els servidors i els recursos de l'empresa, s'ha hagut de buscar una plataforma alternativa de simulació on programar totes les màquines virtuals i comprovar que tot funciona correctament.

Com ja s'ha comentat al principi de la memòria, s'ha utilitzat el **Microsoft Azure** com a eina de treball, que ens permet crear situacions reals, muntar els elements assignant-los IP's i fent connexions entre sub-xarxes.

5.1. Simulació dels entorns

Primer de tot, es crearan les zones de treball més properes possibles als entorns de l'empresa, és a dir: un entorn ECOM que conté dues subxarxes que funcionaran com P i Q, i un altre entorn EMEA fora de l'àrea operacional de ECOM per poder simular la connexió i intercanvi de dades entre els dos.

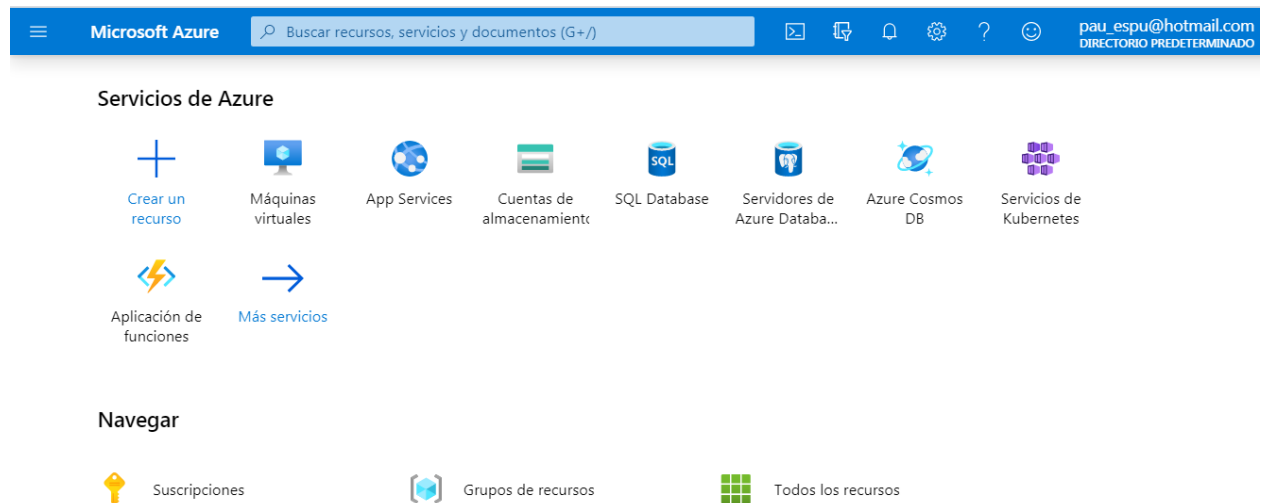


figura 11: Pàgina principal del portal Azure

Azure permet crear recursos dins de zones geogràfiques on té abast. La més propera a nosaltres i que per tant escollirem per situar el nostre entorn principal és la granja de servidors del centre de França. Així doncs, s'afegeix una xarxa virtual anomenada **Production**, que el sistema automàticament assigna a la ip per definició 10.10.0.0, i s'implementen dues subxarxes a dins:

- **ManagmentP:** espai que disposa per definició el sistema per reservar paràmetres de xarxa, administrar serveis futurs com fer *Back-ups*, i on més endavant es desenvoluparà un pont *peering* per connectar amb **Quality**.
- **MailingP:** Sub-xarxa on es procedirà a la implementació dels elements que constituiran la infraestructura. S'assigna a l'adreça 10.10.32.0/24 donant espai suficient per la nostra implementació i futurs projectes.

Es realitza el mateix procediment per l'entorn de Qualitat, dins del centre de França a qui se li assigna la IP 10.20.0.0. I finalment, reservem un espai dins d'un altre àrea de servidors on es situa EMEA (IP: 20.0.0.0). En aquest cas no caldrà fer la diferenciació P/Q perquè només interessa tenir-hi accés per l'intercanvi de dades amb el col·lector.

Resumim la informació de les direccions IP reservades en una taula:

ENTORNS RESERVATS ECOM ADREÇA IP	
ManagmentP	10.10.0.0/24
mailingP	10.10.32.0/24
ManagmentQ	10.20.0.0/24
mailingQ	10.20.32.0/24
EMEA	20.0.0.0

Taula 11: Entorns reservats

5.2. Afegim i configurem els servidors SMTP

A continuació es muntarà l'estructura de **Production**, utilitzarem les llicències del **Windows Server 2016** que van amb la subscripció gratuïta de l'Azure. Cal analitzar quantes màquines virtuals necessitem i amb quins requeriments, i sobre els recursos que disposem:

- 5 màquines virtuals per fer servir com a relays SMTP: **3 per P i 2 per Q**. Són necessaris també 2 balancejadors de càrrega. La plataforma ja ens ofereix alternatives per tal de no usar una màquina virtual exclusiva per cada LB.
- 2 *datacenters* que actuaran de *logCollector* situats un a EMEA-P i l'altre a ECOM. No es configurarà el sistema d'emmagatzematge però sí els punts de connexió.
- Disposem de màquines de 2 GB de Ram i 1 CPU per definició, suficient de moment.
- Per connectar P amb Q a nivell de subxarxa per les adreces de managment, es realitzarà un aparellament de les xarxes (Virtual Network Peering). En canvi, per connectar ECOM i EMEA caldrà fer una connexió VPN ja que el flux serà per internet.

5.2.1. Muntatge de la infraestructura

Seguidament alçarem les màquines virtuals, es programarà el balancejador i es configuraran els servidors SMTP. L'objectiu és aconseguir la implementació bàsica del disseny per tal de demostrar que filtra bé el mail.

Començarem creant les màquines virtuals que fan la funció de relays. És un procés una mica pesat, per tant a la memòria es mostrarà únicament per el primer cas, ja que en els altres és exactament igual. La configuració interna SMTP es realitzarà més endavant.

5.2.2. Creació d'una màquina virtual a l'Azure

És un procés senzill però s'ha de tenir molt clar on es col·locarà cada element. Azure ens obliga a crear un *grup de recursos* per cada regió on treballem, i després de definir especificacions de la màquina, i l'entorn on la volem, ja es pot crear.

Dades principals de la màquina virtual:

- Grup recursos: **servidoresECOM**
- Nom: **FRARELAYP01**
- Tipus SO: **Windows 2016 datacenter**
- Administrador: **pespona**
- Password: **Pesponawin2016**

Interfaz de red

Al crear una màquina virtual, se crea una interfaz de red automàticament.

Red virtual * ⓘ	Production	▼
	Crear nuevo	
Subred * ⓘ	MailingP (10.10.32.0/24)	▼
	Administrar configuración de subred	

Figura 12: Creant la màquina virtual

5.2.3. Configuració LB

Un cop afegits els relays i els dos col·lectors de dades, es pot configurar el **balancejador de càrrega**. El conjunt de relays + LB ha de ser d'alta disponibilitat, per tant Azure requereix crear un conjunt HA dins la sub-xarxa de mailingP i mailingQ on s'hi dipositaran els recursos ja existents. Al crear aquest grup HA, Azure és capaç d'analitzar els elements i a partir de diferents paràmetres que l'usuari designi, es defineix com es distribuirà els paquets de dades.

Entre aquests paràmetres configurem el temps del '*sondeig d'estat*', que correspon a l'interval de temps en que s'està testejant als servidors per saber quan s'haurà de redirigir el servei cap a un altre màquina. L'*Umbral incorrecte*', es refereix al número d'intents seguits que, al no ser escoltat, es redirigeix el trànsit de dades a un altre servidor. Això implica, que si per exemple programem **2 intents x 5 segons**, 10 segons és el temps en que el LB estarà esperant una resposta del servidor per si pot seguir distribuint càrrega abans de canviar cap a un altre relay. S'assigna la direcció 10.10.32.20 al LB de P i 10.20.32.20.

Actualitzant la taula d'IP's amb tots els elements assignats ens queda:

ENTORN	ELEMENT	ADREÇA IP
ManagmentP		10.10.0.0
mailingP	FRARELAYP01	10.10.32.1
mailingP	FRARELAYP02	10.10.32.2
mailingP	FRARELAYP03	10.10.32.3
mailingP	LBPROD	10.10.32.20
mailingP	logCollectorP	10.10.32.10
ManagmentQ		10.20.0.0
mailingQ	FRARELAYP01Q	10.20.32.1
mailingQ	FRARELAYP02Q	10.20.32.2
mailingQ	LBQUAL	10.20.32.20
EMEA		
ManagmentEMEA		20.0.0.0
ManagmentEMEA	logCollectorEMEA	20.0.0.5

Taula 12: IP's assignades

5.2.4. Configuració servidors SMTP

Procedim a la configuració dels relays. Aquest pas és essencial ja que es fixarà el relay amb les especificacions i regles recollides al llarg del treball. Com en el cas anterior, el procediment és igual per cada màquina virtual, així que es mostrarà únicament com s'ha realitzat en el primer cas. El que si cal remarcar, és que com s'ha analitzat en el desenvolupament del projecte, els potencials '*sender domains*' varien en P i Q i per tant les regles d'exclusivitat no seran les mateixes.

Així doncs, primer cal arrancar el servidor i accedir al *server manager* de la màquina. Aquesta aplicació, permet gestionar i administrar diferents característiques i funcionalitats al sistema. En el nostre cas, es busca i s'instal·la el programari IIS 6 (Internet Information Services 6.0), que permet accedir a les prestacions d'un servidor SMTP.

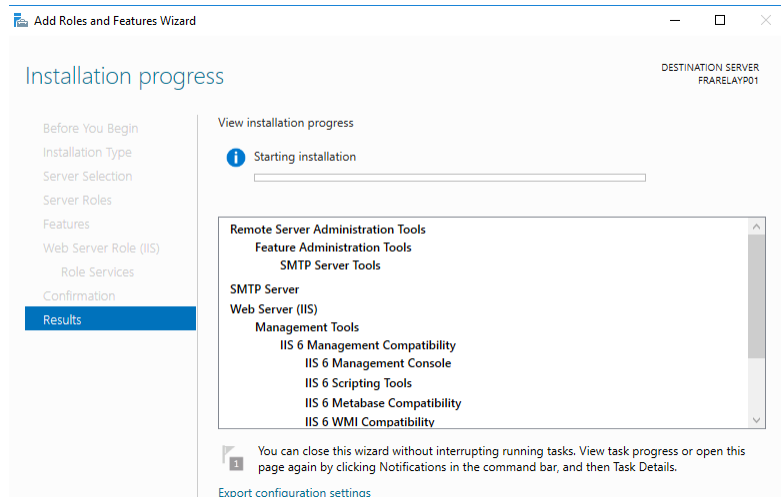


figura 13: Instal·lant IIS 6

Un cop instal·lat, ja es pot arrencar el nostre servidor SMTP virtual, i a falta de fer cap configuració i un cop comprovat que el port 25 està obert, el sistema ja serà operatiu.

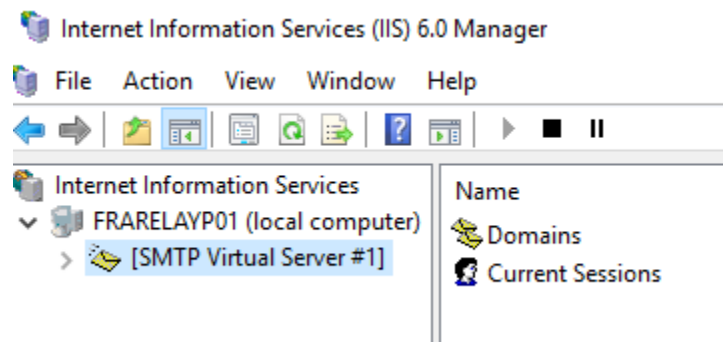


Figura 14: Visualització del servidor virtual SMTP

Per tant, el següent pas serà afegir les nostres especificacions a les propietats al relay. Dins del botó 'Propietats', seleccionem l'opció 'Access': on es defineixen les diferents normes d'accés al relay. Dins de 'relay restrictions' ja podem afegir els dominis que no volem filtrar.

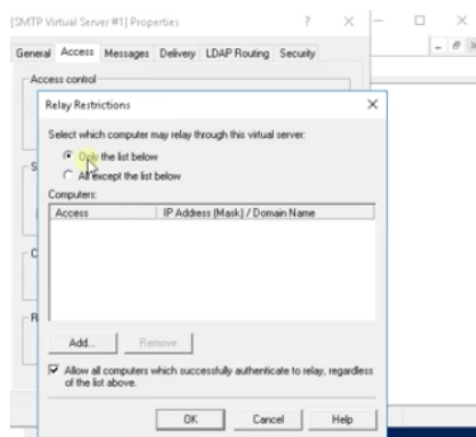


Figura 15: Panell de 'relay restrictions'

Cal afegir un a un cada domini i importar la llista a les següents màquines, diferenciant P i Q. A partir d'aquí, ja hem realitzat la implementació bàsica que ens hem marcat d'objectiu, així que anem a comprovar si funciona.

5.3. Testejem el mail

Realitzarem el test dins del mateix servidor, a partir d'un software descarregat de *github* (<https://github.com/CodeCowboyOrg/SimpleSMTPClient/releases>) que permet simular la funció de client SMTP i així redirigir un mail al servidor en qüestió i connectar a través del port 25. Així doncs, el que volem revisar és que el mail filtra sense problemes per el servidor que hem configurat, comprovant així que la implementació bàsica proposada funciona amb èxit. Posem en pràctica l'execució i observem que el correu s'envia sense problemes.

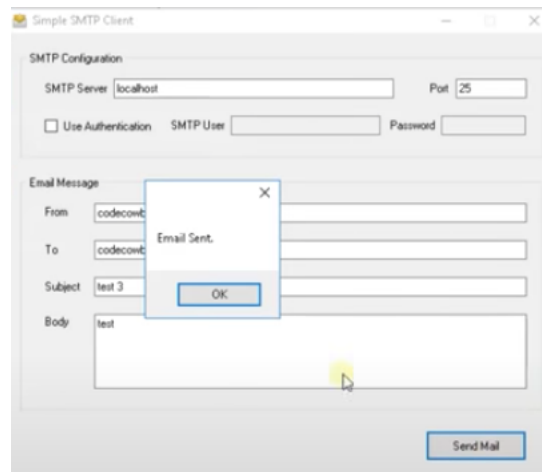


Figura 16: provant el sistema

6. Resultats

En aquest apartat s'avaluaran els resultats obtinguts en base els objectius i especificacions que ens vam marcar al principi. Abans d'entrar en detall, podem afirmar que s'ha aconseguit el principal objectiu proposat, ja que s'ha assolit el disseny de la infraestructura de correu SMTP amb les funcionalitats que volíem, tot i els problemes que han sorgit al llarg del treball, especialment pel COVID 19.

S'ha utilitzat gran part del temps i molts esforços en l'anàlisi de l'antic sistema de mail de l'empresa. Era primordial entendre com funcionava i com s'havia de modificar per tal d'arribar a una solució adaptada a l'actualitat, seguir atenent a les necessitats dels usuaris, i sobretot, que funcioni d'una manera òptima perquè pugui durar molts anys rendint al màxim i sense més modificacions.

Els principals canvis que s'han proposat i executat en el desenvolupament i implementació del projecte són:

- Aconseguir un millor filtratge dels correus que arriben al sistema per tal de no rebre mails innecessaris que suposen una pèrdua de temps i recursos per l'usuari. S'ha realitzat un estudi dels 'sender domains' per tal de reconèixer aquells dominis que simplement no ens interessin o els que ja tenim identificats com spam per tal de denegar la seva entrada als nostres servidors. L'estudi s'ha fet a partir d'un historial de correus de dues setmanes en una 'època pic' a nivell de feina de l'empresa i s'ha implementat al projecte a partir de les regles d'exclusivitat que es configuren en les especificacions pròpies de cada relay.
- S'ha dissenyat una nova arquitectura de relays en els entorns de P i Q ja que l'antiga s'havia quedat obsoleta i tot i que era operativa, estava molt lluny d'un funcionament òptim. S'han analitzat les característiques de milers de mails per conèixer quants s'en reben al dia i així tenir una idea de quants KBytes ha de processar el sistema. A partir d'aquí s'ha decidit quants relays afegiríem. A l'annex s'hi troba l'estudi en qüestió.

Per altre banda, totes les dades contingudes en cada correu, s'han de quedar guardades a la unitat de memòria de la infraestructura, cosa que també hem treballat en el projecte. S'ha optat a usar dos *logCollector* disposats en els dos entorns principals ECOM i EMEA, així només se n'ha hagut de crear un, ja que hem re-utilitzat el d'EMEA que ja estava operatiu, i que ens servirà com a guardaespallles del sistema quan l'altre falli.

Un altre punt important que s'ha tractat, és aconseguir un balanceig equitatiu dels mails entre els relays. Com hem vist, el balancejador de càrrega en un entorn d'alta disponibilitat ens proporciona aquesta funcionalitat. Aquesta distribució de mails és molt important, perquè evitarà una sobrecàrrega innecessària en qualsevol relay que podria suposar la pèrdua de dades, cosa que no ens podem permetre.

Finalment s'ha realitzat una implementació bàsica per fer funcionar els relays i comprovar el correcte funcionament en la recepció d'un correu. S'ha usat l'Azure, que ens ha servit a la perfecció per la simulació dels entorns i la configuració dels diferents elements del disseny.

7. Pressupost

El software utilitzat pel projecte (Microsoft Azure) s'ha usat en la seva subscripció gratuïta, així que no comptabilitzem cap cost en aquest sentit. El PC amb totes les funcionalitats bàsiques operatives ha estat proporcionat per l'empresa, i li ha suposat un cost de 800 €.

La durada del projecte ha estat de 17 setmanes, i si tenim en compte unes 30 hores setmanals, són 510 hores en total. Un enginyer junior cobra 12 € / hora, que afegint la contribució del 30% a la seguretat social serien 15,6 € / hora. Per tant, el salari total a pagar a l'enginyer seria de **15,6 x 510 = 7.956 €**. També hem de comptabilitzar 3 hores setmanals (50 hores en total) de seguiment d'un enginyer sènior que cobra una mica més. En total, el seu salari és: 50 hores x 25 € / hora = 1250 €.

CONCEPTE	PREU
PC portàtil	800€
Salari enginyer junior	7,956€
Salari enginyer sènior	1,250€
TOTAL	10,006€

Taula 13: Pressupost

8. Conclusions i visió de futur

Aquest projecte es tracta d'un exercici d'anàlisi profund d'un sistema de correu SMTP per tal de dissenyar-ne una versió millorada. Així doncs, en bona part de les hores invertides m'he centrat en estudiar i entendre el perquè de cada funcionalitat del sistema antic per tal de millorar cada detall i trobar noves especificacions i procediments per la nova solució.

Un cop acabat el projecte, puc extreure conclusions en 4 punts de vista diferents, 4 punts fonamentals per arribar a la solució final i que òbviament es comuniquen entre ells.

Primerament, a nivell d'investigació, crec que he adquirit molts coneixements que no m'imaginava al principi. Això és degut a les ganes de descobrir coses que tenia quan se'm va proposar el projecte, ja que no sabia res sobre el tema i he hagut de fer un gran esforç per posar-me al nivell bàsic necessari. Per sort, considero molt útil tot el que he tractat, així que m'ha semblat relativament fàcil a l'hora de buscar i conèixer nous elements i eines.

Crec que el protocol SMTP és una eina molt potent i això ho demostra el fet de que fa anys que es fa servir amb molts bons resultats, i és la base de la comunicació per correu, imprescindible avui en dia. Nogensmenys, internet també té el seu cantó negatiu, i tal com hem vist, hi circula molt de spam que a vegades fins i tot pot ésser maliciós per el nostre ordinador. Aïllar l'spam del nostre sistema ha esdevingut un dels nostres principals maldecaps i objectius a completar.

També crec important comentar que apart de la teoria, també ha sigut crucial entendre l'entorn de treball de l'empresa, així com cada departament i les necessitats tant de Quality com de Production.

A nivell d'anàlisi també crec que en puc extreure varies conclusions. Potser no gosaria a dir 'conclusions d'aquest projecte en concret', sinó més aviat amb una visió general, he après com organitzar-me per realitzar una feina d'aquesta envergadura i sobretot com gestionar els milers i milers de dades a les quals m'estava enfrontant. És molt complicat, i una mica desesperant, obrir un Excel amb 500.000 files d'informació. I encara és més difícil saber resumir aquestes dades i arribar a una conclusió que farà que el disseny final sigui d'una manera o altre. Així doncs, crec que aquest projecte m'ha ajudat a descobrir tècniques molt efectives per buscar l'agulla dins el paller (metafòricament parlant), i m'ha portat a ser molt més analític amb qualsevol petit detall.

Pel que fa al disseny, és la part que més he disfrutat al realitzar el projecte. Penso que al partir d'una bona base tant pel que fa al desenvolupament del projecte com en coneixement de servidors i el protocol TCP, no ha sigut massa difícil pensar com havia de ser el sistema. Un cop decidit el nombre de relays, tenia claríssim que per optimitzar-los al màxim necessitàvem equitat de càrrega, cosa que ens proporcionen els balancejadors LB al peu de la lletra.

A més, s'ha dissenyat una estratègia de recuperació de dades imprescindible, ja que en el cas que algun usuari perdi algun correu i s'hagi de buscar a la base de dades, tinguem la total certesa de que es podrà recuperar i no s'haurà perdut per la xarxa.

Finalment, sobre la implementació del disseny, com he comentat, degut al COVID-19 vaig haver de modificar els meus plans al respecte. Al no poder seguir a l'empresa, no vaig poder implementar el sistema en la seva versió real i vaig haver de buscar una alternativa per simular-ho. Azure ha sigut l'eina perfecte en aquest sentit, ja que m'ha permès fer tot el que em vaig proposar, i trobo que és una eina molt potent i intuïtiva que segur faig servir al futur.

Com a visió de futur, òbviament m'agradaria portar a terme el projecte en la seva versió real, però si no és possible, la feina realitzada no ha sigut en va i de ben segur que si no sóc jo, algú la farà servir. Crec que el disseny fet pot durar molts anys de manera òptima sense que cap usuari noti que alguna característica s'està quedant obsoleta. A més, hem previst que els relays tenen memòria de sobres per processar més dades per si en un futur l'empresa pateix un creixement.

A nivell de disseny, i com a implementació de futur, es podria fer un seguiment més individualitzat pel que fa als correus enviats per cada dia de la setmana. Està clar, que un dimarts no es rep la mateixa quantitat de dades que un diumenge, i per tant el sistema s'estressa més o menys. Es podria buscar alguna manera d'optimitzar aquesta observació, que de ben segur seria positiu per el funcionament a llarg plaç.

Bibliografia:

- [1] Smtip mail server – professional SMTP service [Online] Available: <https://serversmtp.com/es/configuracion-servidor-smtp/> [Accessed: 5 febrer 2020]
- [2] What is an SMTP server? SMTP server model and configuration [Online] Available: <https://pepipost.com/blog/what-is-smtp/> [Accessed: 5 febrer 2020]
- [3] Windows server 2016 on SMTP [Online] Available: <https://www.solvetic.com/tutoriales/article/3548-anadir-configurar-servidor-smtp-windows-server-2016/> [Accessed: 10 març 2020]
- [4] Pagina oficial de Microsoft Azure [Online] Available: <https://docs.microsoft.com/es-es/azure/> [Accessed: 10 març 2020]
- [5] SMTP (Simple Mail Transfer Protocol) [Online] Available: <https://ingsistemasunicor.files.wordpress.com/2015/09/protocolo-smtp-simple-mail-transfer-protocol.pdf> [Accessed: 10 març 2020]
- [6] Round Robin in Load Balancing [Online] Available: <https://avinetworks.com/glossary/round-robin-load-balancing/> [Accessed: 15 març 2020]
- [7] Cisco community – ACL assistance [Online] Available <https://community.cisco.com/t5/other-security-subjects/need-acl-assistance-for-smtp/td-p/1826138> [Accessed: 28 març 2020]

Annexos

Annex 1: Estudi del flux de mail per P

Seguidament es mostra l'estudi comentat en l'apartat de desenvolupament. Cada taula representa un client diferent en l'època de 2 setmanes. Els clients són els mateixos de la *taula 10: inventari dels 'sender domains'*. S'han comptat quants mails entren al servidor cada dia d'aquest client concret durant dues setmanes i s'ha n'ha fet la mitjana. Llavors s'han sumat tots els clients i així tenim una idea dels mails diaris.

Com ja s'ha explicat a la memòria, farem servir aquestes dades per determinar el número de relays que posarem al disseny.

serverDIST1		serverDIST1		serverDIST1	
clientB	NUM MAILS	clientC	NUM MAILS	clientD	NUM MAILS
V	8	V	1500	V	30300
S	2	S	750	S	60400
D	0	D	90	D	95000
L	11	L	2100	L	7450
M	2	M	2020	M	9480
X	1	X	4200	X	6200
J	5	J	3300	J	5300
V	1	V	2200	V	3130
S	0	S	900	S	2600
D	0	D	320	D	7500
L	1	L	3000	L	6125
M	7	M	4500	M	6000
X	2	X	2300	X	6500
J	8	J	3800	J	3000
V	4	V	3900	V	2030
APROX 2 WEEKS	3.466666667	APROX 2 WEEKS	2325.333333	APROX 2 WEEKS	19152.69231

serverDIST3		serverDIST3		serverDIST3	
clientH	NUM MAILS	clientI	NUM MAILS	clientJ	NUM MAILS
V	2	V	2400	V	1
S	0	S	140	S	2114
D	0	D	100	D	2311
L	3	L	2500	L	2390
M	3	M	3900	M	2119
X	5	X	1300	X	135
J	4	J	2800	J	164
V	6	V	1400	V	2570
S	3	S	70	S	2278
D	2	D	35	D	2370
L	3	L	1500	L	2185
M	2	M	2000	M	3690
X	2	X	3200	X	45
J	1	J	1200	J	3
V	9	V	1200	V	2719
APROX 2 WEEKS	3	APROX 2 WEEKS	1583	APROX 2 WEEKS	1242.466667

serverDIST2		serverDIST4	
clientF	NUM MAILS	clientK	NUM MAILS
V	6	V	56
S	6	S	70
D	10	D	87
L	12	L	160
M	11	M	127
X	6	X	160
J	4	J	145
V	4	V	130
S	6	S	12
D	3	D	11
L	2	L	54
M	3	M	70
X	12	X	112

J	7	J	108
V	4	V	95
APROX 2 WEEKS	6.4	APROX 2 WEEKS	93.13333333

serverDIST5		serverDIST5	
clientM	NUM MAILS	clientN	NUM MAILS
V	921	V	98
S	569	S	56
D	825	D	67
L	548	L	123
M	160	M	278
X	145	X	212
J	130	J	178
V	125	V	213
S	112	S	67
D	76	D	23
L	345	L	89
M	598	M	130
X	237	X	245
J	457	J	176
V	1790	V	143
APROX 2 WEEKS	469.2	APROX 2 WEEKS	139.8666667

MAILS TOTALS P
26591.15897

Annex 2: Estudi del flux de mail per Q

Per el cas de Quality.

serverDIST1		serverDIST1		serverDIST3	
clientBQ	NUM MAILS	clientCQ	NUM MAILS	clientGQ	NUM MAILS
S	120	S	678	S	3750
D	269	D	455	D	4600
L	321	L	442	L	2000
M	123	M	1233	M	2900
X	277	X	1134	X	2903
J	140	J	890	J	2678
V	122	V	333	V	1234
S	124	S	67	S	1788
D	232	D	321	D	1677
L	122	L	123	L	2344
M	332	M	43	M	2000
X	89	X	322	X	2900
J	98	J	122	J	3455
V	88	V	124	V	1232
S	68	S	4800	S	455
APROX 2 WEEKS	168.3333333	APROX 2 WEEKS	419.1333333	APROX 2 WEEKS	2364.066667

serverDIST2		serverDIST4		serverDIST5	
clientFQ	NUM MAILS	clientKQ	NUM MAILS	clientLQ	NUM MAILS
S	4600	S	2369	S	123
D	3800	D	4554	D	232
L	1234	L	3323	L	235
M	1788	M	4554	M	432
X	1677	X	3443	X	356
J	2900	J	5533	J	765
V	4600	V	2333	V	543
S	4400	S	4600	S	454
D	3750	D	2123	D	234
L	4600	L	2143	L	453



M	2000	M	3122	M	768
X	3432	X	2432	X	577
J	2333	J	2431	J	655
V	3455	V	2687	V	553
S	4800	S	2709	S	567
APROX 2 WEEKS	2971.266667	APROX 2 WEEKS	3043.133333	APROX 2 WEEKS	425.3333333

MAILS TOTALS P

16317.8

Glossary

SMTP: Simple Mail Transfer Protocol

TCP: Transmission Control Protocol

ACL: Acces Control List

IP: Internet Protocol

P: Production

Q: Quality