



Portal de Promoción Turística en Internet. Turespaña. Arquitectura Tecnológica de referencia

Jorge Rubio Navarro.

Subdirector Gral. Adj. Medios de promoción turística. Turespaña

Ana Roman Riechmann.

Jefa de Area Informática. Turespaña

Jesús Quereda Rubio.

Jefe de Servicio Análisis Sectoriales. Instituto de Estudios Turísticos.

1.- INTRODUCCIÓN

El proyecto de la Secretaría General de Turismo nace con la idea de potenciar el turismo en nuestro país a través de un nuevo portal de promoción en Internet que incluya información destinada tanto a profesionales como al usuario final.

En este sentido, se ofrecerán contenidos sobre recursos turísticos, itinerarios, servicios, Patrimonio Histórico Artístico, agenda cultural, gastronomía, comunicaciones, meteorología, etc., orientados básicamente a usuarios, pero sin olvidar también un apartado de amplia información para los profesionales del Sector Turístico, con el desarrollo de una Extranet de carácter restringido con acceso a información exclusiva. Contará, asimismo con una Intranet para informadores turísticos, que permitirá la atención personalizada de las consultas que se reciben en las Oficinas Españolas de turismo en el exterior.



Ayuntamiento de A Coruña





El modelo de Portal se corresponde con el denominado Factoría de Contenidos (Content Factory). De acuerdo a éste modelo los servicios básicos necesarios se agrupan en Servicios de adquisición de contenidos, Servicios de Gestión de contenidos y Servicios de Distribución de contenidos. El Portal dispondrá de contenidos multilingües (hasta en 9 idiomas distintos) con prestaciones de accesibilidad las 24 horas del día y con cualquier tipo de dispositivo (Pc, móvil, Agendas Personales o TV interactiva) además dispondrá de las siguientes prestaciones suplementarias: servicio de proximidad WAP, cálculo de rutas a pie, en coche o transporte público, envío de postales digitales durante la realización del viaje, 754 localidades cartografiadas, 21 visitas virtuales relacionadas con el arte y el turismo cultural.

El Portal de Turespaña se constituye como el principal elemento catalizador o palanca de cambio para la innovación tecnológica del Sector Turístico Español.

El presente documento describe la arquitectura sobre la que se está implantado el Portal de Turismo.

2.- ARQUITECTURA TECNOLÓGICA

En este apartado se describe la arquitectura tecnológica de referencia para el desarrollo de los servicios requeridos en el pliego de prescripciones técnicas para la construcción del portal de promoción turística en internet del Instituto de Turismo de España (Turespaña).

El diseño de la arquitectura se ha realizado en base a unos requerimientos estimados de carga teniendo en cuenta la escalabilidad horizontal y vertical.

El diseño incluye tres entornos de operación, el de producción, pre-producción y el entorno de desarrollo.

La arquitectura tecnológica se ha realizado utilizando como software de base MS Windows 2000 y los productos específicos de la arquitectura .Net de Microsoft.

La Arquitectura de el Portal está diseñada en base a los siguientes criterios:





- Escalabilidad

La asignación de servidores por tipo de servicio hace posible el tratamiento separado de estos, pudiendo adaptar en todo momento cada uno a las necesidades del Portal y a la demanda del mercado.

Aunque la propia arquitectura proporciona tolerancia a fallos, es posible aumentar el número de servidores que proporcionan cada servicio, esto es, escalabilidad horizontal. Esto no solo garantiza la disponibilidad del servicio, sino que permite distribuir la carga entre los servidores que proporcionen el servicio en cuestión.

En el caso de NLB (*Network balancing Service*) este número puede aumentar hasta 32 servidores. En el caso de balanceadores hardware dependerá de cada producto en cuestión el grado de escalabilidad que el mismo sea capaz de alcanzar.

Adicionalmente, también es posible escalar verticalmente aumentando la potencia de los servidores que componen la solución.

- Servicios

Servicios de Certificados: Es posible ampliar la arquitectura mediante los servicios de certificados que la propia plataforma proporciona (Microsoft Windows 2000 Certificate Services). De esta manera se podrán implementar comunicaciones seguras usando IPSec, validación de transacciones, clientes, etc., todo ello mediante la utilización de certificados.

- Seguridad:

- Analizador de Vulnerabilidades.
- Servicio de antivirus





- Alta disponibilidad

Como parte de la solución hardware de alta disponibilidad, se han seguido las siguientes recomendaciones mínimas en los sistemas de la arquitectura propuesta:

- Disponibilidad de almacenamiento RAID en los servidores – Uso de RAID 1 para el sistema operativo y al menos RAID 5 para los datos.
- Sistemas de alimentación redundantes en cada servidor y cabinas de almacenamiento externo.
- Conexiones redundantes a la red, con múltiples rutas de datos que puedan sobrevivir al fallo de un cable, switch o router.

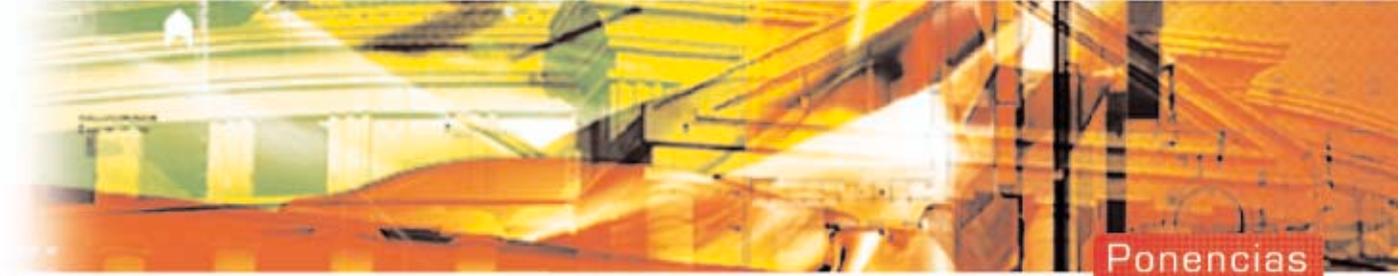
Adicionalmente a nivel de arquitectura se aplican criterios de alta disponibilidad con respecto al agrupamiento de servidores:

- Servidores Web balanceados – Múltiples servidores idénticos con balanceo de carga mediante soluciones de herramientas software como Windows 2000 Network Load Balance Service (NLB) o mediante hardware de balanceo IP como Cisco Local Director o similar.
- Servidores de Aplicaciones balanceados – La mayoría de la lógica de negocio se ejecuta en la capa de aplicación. Varios servidores iguales balanceados utilizando servicios de balanceo de componentes agrupados en un array N+1. Esta agrupación se denomina “Layer 2 Clustering”.
- Servidores de bases de datos en Cluster – Servidores de bases de datos SQL con servicio de cluster instalados sobre el hardware apropiado en agrupaciones de 2 o 4 nodos activos con almacenamiento compartido en cabinas de discos o redes de almacenamiento (SAN). Esta configuración se denomina “Layer 3 Clustering”
- Firewall redundantes – Dos niveles de Firewall redundantes y de diferente fabricante en cada nivel, configurados para realizar balanceo de carga y alta disponibilidad en caso de caída de uno de los nodos.

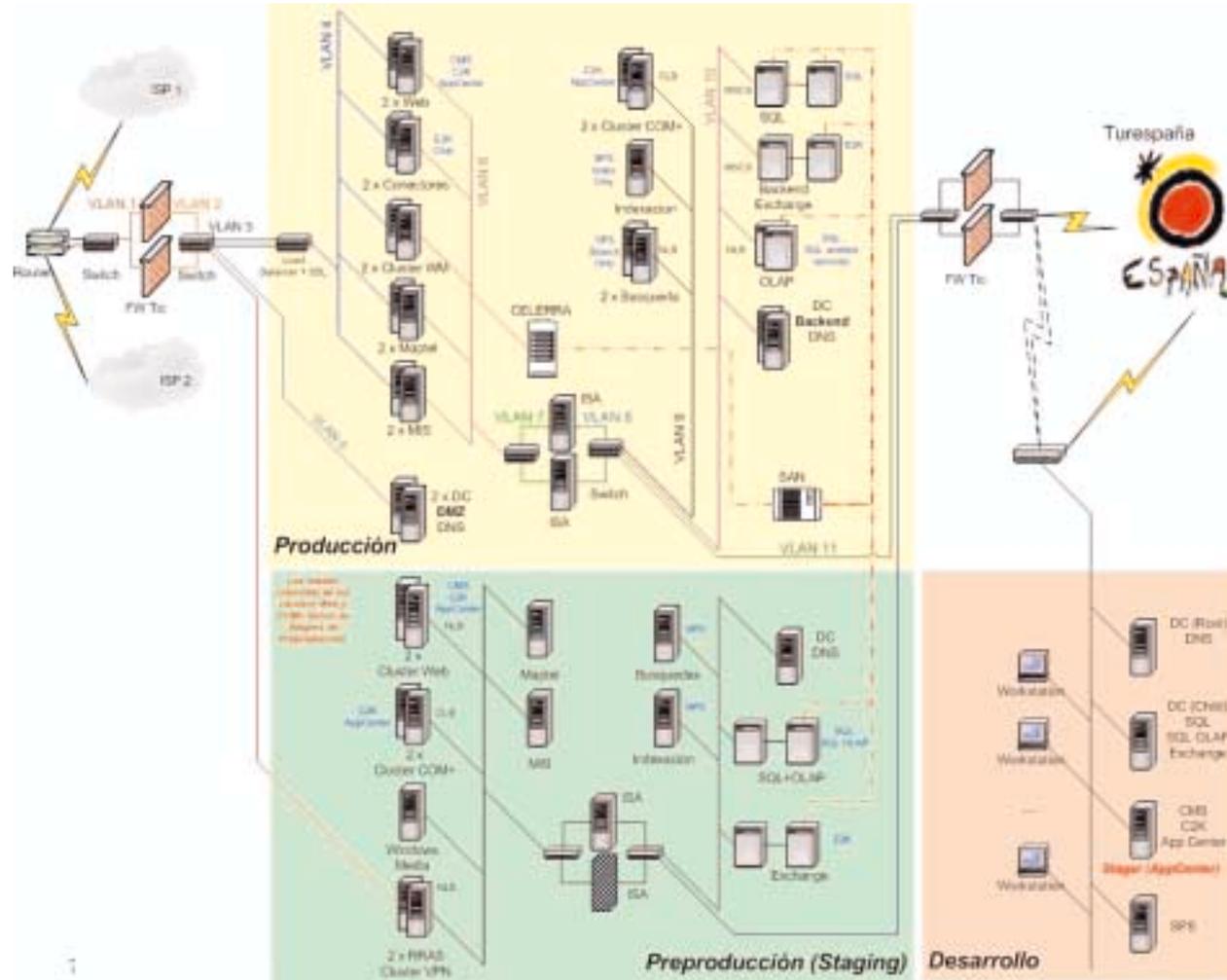


Los requisitos de carga sobre los que se ha diseñado la arquitectura han sido:

Parámetro		Valor	Observaciones
Nº de usuarios que utilizarán el portal	Situación normal	2.000 usuarios/hora	
	En hora punta	20.000 usuarios/hora	
Nº de usuarios registrados		5.000	
Nº de usuarios que usarán los servicios del portal (chat, forums, suscripción contenidos..)	15% de los usuarios registrados		
Nº de transacciones	Situación normal	30.000 Tx/h	
	En hora punta	100.000 Tx/h	
Tipos de transacciones	Complejas	10%	Con validación y accesos a bases de datos
	Simple	90%	
Tamaño medio de transacción		50 Kb	



2.1.- DIAGRAMA DE LA ARQUITECTURA





2.2.- COMPONENTES DE LA ARQUITECTURA Y FUNCIONAMIENTO

La disposición de los servidores se ha realizado en base a la arquitectura DNA, la cual proporciona un modelo en el que se distinguen tres capas:

- Capa de presentación: Formada por los servidores que realizan las funciones de frontales de la solución propuesta.
- Capa de aplicación: donde se desarrolla toda la lógica de negocio.
- Capa de Datos.

2.3.- SERVICIOS

CATÁLOGOS

El servicio de catálogos se encarga del almacenamiento y organización de los elementos objeto de los subsistemas de contenidos. Cada catálogo contiene información básica de los elementos de cada un subsistema, obteniendo así un catálogo para hoteles, otro para balnearios, etc.

Estos catálogos están basados en el sistema de catalogación de omerser Server 2000. Este sistema de catalogación provee de algunos aspectos de funcionalidad tales como:

- Características de búsqueda por categoría, por propiedad, por especificación o por texto libre
- Importación y exportación de catálogos en formato XML
- Informes de actividad
- Personalización de propiedades por grupos de usuarios
- Intercambio de catálogos vía BizTalk Server 2000





El almacenamiento de datos de *Comerse Server* está basado en SQL Server 2000. Además del almacenamiento puro de los catálogos, existe un servicio de Data Warehouse y procedimientos analíticos de información en base a los catálogos de *Comerse Server*. Este servicio está basado en los *Servicios de Proceso Analítico On-line (OLAP)* de SQL Server.

PERFILES DE USUARIO

Los perfiles de usuario funcionan como un almacenamiento de información individualizada. Deben tener la capacidad de identificar a un usuario particular (*autenticación*) y almacenar la información de dicho usuario (*personalización*). Además, esta información de usuarios debe ser fácilmente actualizable.

Esta información (perfil) permitirá una personalización en diversos contextos (por ejemplo: idioma de presentación).

El almacenamiento de los perfiles de usuario está basado en el sistema de perfiles y personalización de *Comerse Server* 2000. Este sistema de perfiles es flexible para analizar datos sobre los usuarios, tales como:

- Análisis de información sobre el acceso de los distintos perfiles a la información del portal
- Crear expresiones de targeting para enviar contenido personalizado
- Crear campañas específicas de anuncios, mailing directo, etc.

Un perfil de usuario tendrá una definición, dentro de la cual existen *propiedades* (valores de introducción requerida por el usuario) y *atributos* (característica sobre el perfil). Cada usuario generará una instancia de la definición de perfil con sus propios datos.

Este sistema de perfiles de *Comerse Server* basa su almacenamiento de datos en SQL Server.

FOROS

Los foros de discusión se implementan en el portal para crear un lugar común de información donde los proveedores puedan ser los propios usuarios del portal, donde intercambian opiniones y se forman conversaciones con varios mensajes.

Estos foros son implementados con un desarrollo personalizado basado. La lógica de aplicación de los foros se realiza mediante componentes COM+ que se integran en la interfaz de usuario. El almacenamiento de los datos relacionados con los foros se realiza a través de SQL Server 2000.



NEWSLETTERS

Las Newsletters son campañas personalizadas de envío periódico de información a los usuarios del portal. Este tipo de campaña enviará un mensaje a un usuario o a un grupo de usuarios dependiendo del desencadenamiento de una condición, ya sea por acción del usuario, suscripción a un determinado servicio de noticias o por decisión del portal.

Las campañas de Newsletters se implementan sobre SQL Server y ommerce Server 2000, utilizando el componente de ommerce Server Direct Mail.

El componente Direct Mail se instala en el servidor SQL Server 2000 como un servicio de Windows y se encarga del mantenimiento, control y gestión de las campañas de e-mailing de ommerce Server.

CHAT

El portal ofrece un servicio de Chat a los usuarios. Este servicio puede tener una serie de canales preestablecidos y canales dinámicos. Los canales preestablecidos confieren una estructura básica organizada, pero utilizando canales dinámicos, los usuarios pueden crear sus propios canales. Eventualmente, si se desean crear eventos especiales, tales como charlas on-line con personajes relevantes en el sector o especialistas concretos, se pueden crear canales en modo Auditorio. En este modo, muchos usuarios concurrentes están escuchando a unos pocos ponentes. Los usuarios sólo ven a los ponentes y los ponentes pueden ver a todo el mundo. Existen más modos de canales que pueden ser establecidos al crearlos.

El servicio de Chat en el portal se implementa a través de Exchange, con el Exchange 2000 Chat Service. Este servicio está basado en el estándar IRC y utiliza las características de directorio activo como base para el mantenimiento, configuración y almacenamiento.

ENCUESTAS

Las encuestas proveen al portal de un sistema de feedback por parte de los usuarios. En un momento dado, el sistema de encuestas del portal pedirá al usuario que provea cierta información sobre un tema concreto.

Las encuestas deben implementarse mediante un desarrollo personalizado basado en componentes COM+ para la lógi-



ca de negocio, páginas ASP para la interfaz de usuario y presentación de la información y SQL Server como sistema de almacenamiento.

CUADERNO DE VIAJE

El Cuaderno de Viaje es un servicio similar a la “Cesta de la compra” que podemos encontrar en un Marketplace. El usuario, cuando encuentre un lugar turístico de su interés en el portal, tendrá la posibilidad de agregarlo a su Cuaderno de Viaje. Al final de la visita, el usuario podrá ver el detalle del Cuaderno de Viaje, con información más detallada de estos lugares. Este cuaderno le servirá al usuario como una guía de viaje personalizada.

El desarrollo del Cuaderno de Viaje debe implementarse como un desarrollo personalizado basado en componentes COM+ para la lógica de negocio, páginas ASP para la interfaz de usuario imbuida en el portal y SQL Server como sistema de almacenamiento final. Además, los componentes COM+ accederán al modelo de objetos de Commerce Server, dado que los elementos susceptibles de formar parte del Cuaderno de Viaje residirán en catálogos de Commerce Server.

BUZONES

Los buzones de correo comportan un servicio de correo electrónico dentro del portal. Este servicio está basado en Exchange Server como servidor de correo electrónico entrante/saliente, y en el servicio SMTP de Windows 2000 para correo saliente.

Dado que no habrá frontend de Exchange, se utilizarán servidores de Exchange no visibles desde Internet para la gestión de los buzones y almacenamiento de datos, y servidores Exchange visibles desde Internet para la redirección de tráfico SMTP/POP3.

FOLLETOS

Determinada información del portal debe presentarse en un formato preparado para ser impreso o descargado al ordenador del usuario. Esta información está organizada en folletos. Los folletos conforman una versión “imprimible” de algunos contenidos del portal. Eventualmente se ofrecerá la posibilidad de agregar o seleccionar los conceptos que irán en un folleto, obteniendo así un “folleto personalizado”.



Los folletos deben realizarse mediante un desarrollo personalizado basado en componentes COM+ para implementar la lógica, páginas ASP para la interfaz de usuario y presentación de la información y SQL Server como sistema de almacenamiento.

STREAMING DE MEDIA

El portal contiene determinados contenidos en formatos multimedia, tales como videos o locuciones. Estos contenidos deben servirse mediante streaming o broadcasting para evitar la descarga completa de archivos de gran tamaño a los ordenadores de los usuarios, lo cual haría impracticable la posibilidad de tener videos o locuciones de calidad aceptable.

Para realizar streaming de media se utilizan los servicios Windows Media. Los archivos objeto del streaming deben estar en un formato compatible con Windows Media (WMV, ASF, WMA, MP3)

CONTENIDO

Los contenidos generales del portal, así como los contenidos asociados a elementos de los catálogos de ommerce Server son gestionados y mantenidos desde Content Management Server 2001.

Content Management Server es además la herramienta principal de creación de páginas ASP para el portal. Es por tanto, el producto principal utilizado tanto por los desarrolladores de las páginas como por los autores, editores y gestores de contenido.

Content Management Server está completamente integrado con ommerce Server para ofrecer una solución conjunta. De esta forma, la gestión de contenidos generales y la gestión de catálogos están integradas en una única solución.

Content Management Server está a su vez basado en SQL Server para el almacenamiento interno, tanto de ficheros, plantilla y páginas como de contenido.

SERVICIOS DE BÚSQUEDA E INDEXACIÓN

Los servicios de búsqueda de contenidos del portal permiten que se busque en el texto de cualquier página o conteni-



do del portal, así como en fuentes de datos externas asociadas al mismo.

La búsqueda del portal se basa SharePoint Portal Server 2001. Para realizar el rol de búsqueda SharePoint tiene el servicio de indexación de metadatos. De esta forma, los servidores destinados a este servicio están separados por funcionalidad, según estén dedicados a generar los índices del servicio o a realizar las tareas búsqueda y devolver resultados al portal.

2.4.- COMPONENTES DE SOFTWARE

El siguiente apartado muestra la disposición de los distintos servicios en los diferentes entornos que conforman el portal.

ENTORNO DE PRODUCCIÓN

2.5.- DISEÑO DEL SERVICIO DE DIRECTORIO

La implementación de los servicios de directorio Active Directory Services y modelo de dominios de Windows 2000 se analiza como alternativa para la definición del entorno de seguridad del portal de promoción turística en Internet.

2.6.- MODELO DE DOMINIOS

Como ya se ha mencionado la arquitectura se basa en arquitectura DNA que proporciona un modelo en el que se distinguen tres capas.

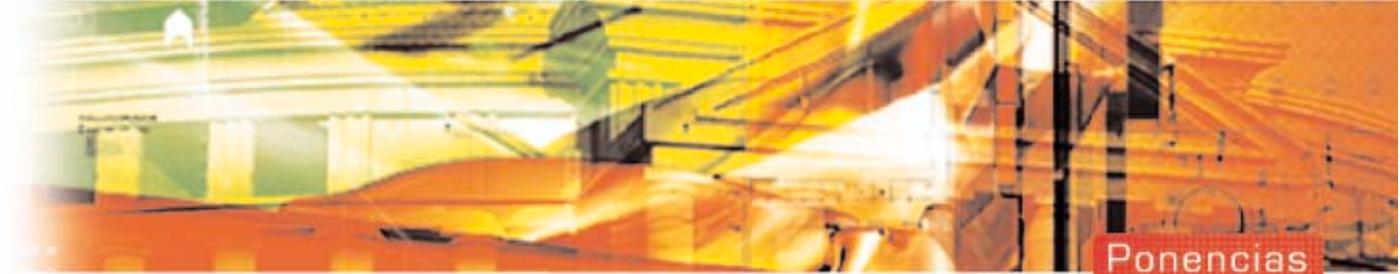
En la arquitectura propuesta, la capa de presentación está físicamente separada de las capas de Aplicación y Datos, por un segundo nivel de Firewall.

Teniendo en cuenta que para la solución deseada la capa de Presentación puede tener un contexto de seguridad independiente del de las capas de Aplicación y Datos, y con el fin de no tener que abrir excesivos puertos de este segundo nivel de Firewall, el diseño de Directorio Activo propuesto es de dos bosques.

Un bosque para los servidores de la capa de presentación que se encuentran en la zona desmilitarizada (DMZ) y que constará de un único árbol y un único dominio, necesario para los servidores de Exchange (conectores y Chat).



Servicio	Producto/Tecnología	Localización
Catálogos	Commerce Server 2000	Cluster de servidores Web
Perfiles de usuario	Commerce Server 2000	Cluster de servidores Web
Newsletters	Commerce Server 2000 Direct Mail, SQL Server 2000	Cluster SQL Server
Contenido web	Content Management Server 2001	Cluster de servidores Web
Cuaderno de viaje	SQL Server 2000	Cluster SQL Server
	Commerce Server 2000	Cluster de servidores Web
	Desarrollo ASP	
	Desarrollo Componentes	
Encuestas	SQL Server 2000	Cluster SQL Server
	Desarrollo ASP	Cluster de servidores Web
	Desarrollo Componentes	Cluster COM+
Folletos	SQL Server 2000	Cluster SQL Server
	Commerce Server 2000	Cluster de servidores Web
	Desarrollo ASP	
	Desarrollo Componentes	
Foros	SQL Server 2000	Cluster SQL Server
	Desarrollo ASP	Cluster de servidores Web
	Desarrollo Componentes	Cluster COM+
Buzones	Desarrollo ASP	Cluster de servidores Web
	Exchange Server 2000	Conectores Exchange
		Cluster Backend Exchange
Chat	Desarrollo ASP	Cluster de servidores Web
	Exchange Server 2000	Conectores Exchange
		Cluster Backend Exchange
Streaming de media	Windows Media Services	Cluster WM
Servicios de búsqueda	SharePoint Portal Server 2001	Cluster de Búsqueda
Servicios de indexado	SharePoint Portal Server 2001	Servidor de Indexado



En la siguiente figura se muestra la distribución de los dominios correspondientes a los entornos de producción y pre-producción:

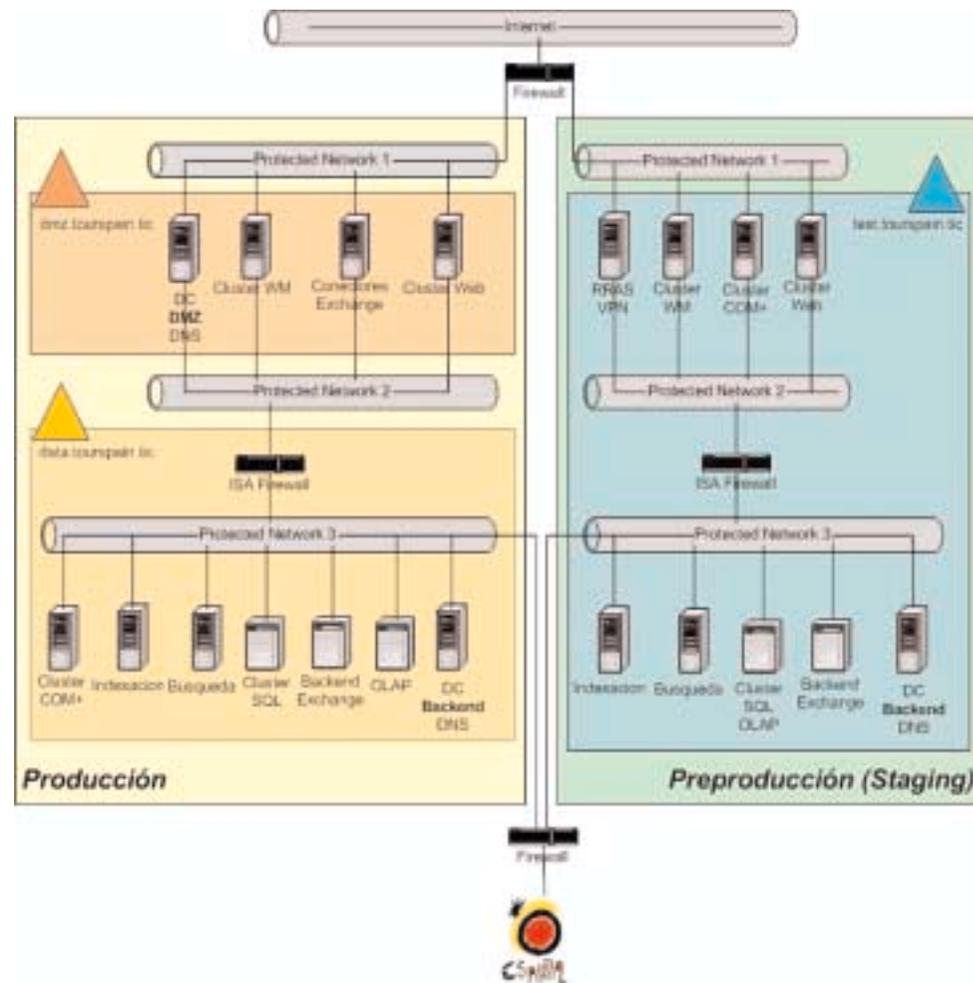


Diagrama de servicios en el entorno de Producción



2.7.- MODO NATIVO

El directorio activo se creará en modo nativo. El modo nativo permitirá mejorar en escalabilidad ya que no va a estar limitado al tamaño de directorio que puedan admitir los controladores de dominio Windows NT. Además desde el punto de vista de Exchange 2000 es ventajoso ya que se podrán utilizar grupos de seguridad universales (Universal Security Groups) y anidamiento de grupos.

El modo nativo o mixto del directorio activo es transparente para los usuarios, que no necesitarán ningún cambio ni configuración para operar en cualquiera de los modos.

2.8.-DNS

La resolución de nombres mediante DNS es especialmente crítica en la implantación de un servicio de mensajería basado en Microsoft Exchange 2000. Por una parte, el servicio de mensajería utilizará la resolución DNS para poder enviar y recibir mensajes con destinatarios o emisores que no pertenecen a la propia organización y para publicar las direcciones de acceso de los clientes. Y por otra parte, de forma interna Exchange 2000 utiliza como servicio de directorio el Directorio Activo que se basa en la resolución de nombres DNS para la localización de los recursos. Este es el motivo por el cual se debe pensar en el servicio DNS en dos entornos distintos:

- Como servicio hacia el exterior
- Como servicio interno.

Esta configuración del sistema de resolución de nombres DNS recibe el nombre de "SPLIT" y consiste en dos subsistemas DNS: uno externo y otro interno. El subsistema DNS externo que resuelve peticiones desde Internet para obtener información de los servidores que están en la DMZ y peticiones desde la DMZ para servidores en la DMZ. Por otra parte, el subsistema interno que resuelve peticiones desde servidores del back-end para obtener información de servidores en el back-end. Peticiones desde el back-end para obtener información de servidores en la DMZ o para servidores en Internet son reenviadas ("forwarding") al sistema DNS externo. El problema de esta configuración es que no protege contra ataques "DNS cache poisoning".



Una consideración importante acerca de los sistemas DNS que se debe añadir es que nunca se deben permitir que las peticiones DNS desde Internet atraviesen la zona DMZ hasta llegar a la red interna para obtener la resolución de nombres, ya que no es necesario que los clientes de Internet realicen peticiones al DNS interno para la resolución.

SISTEMA DNS EXTERNO

Proporciona resolución de nombres a los clientes. En este sistema DNS, sólo estarán registrados aquellos servicios a los que deben acceder los clientes desde Internet.

No es necesario implementar un sistema DNS para clientes del portal, ya que lo proporciona el TIC de Telefónica. El DNS de la DMZ estará dedicado a los equipos de la propia red.

SISTEMA DNS INTERNO

Proporcionan resolución de nombres a los servidores internos, ya sea en las redes DMZ como en las redes internas. Debe conocer las direcciones de los servidores internos y además contener los registros necesarios (registros "SRV") para el directorio activo. Este sistema DNS se realizará utilizando el servicio DNS de Windows 2000 de manera integrada con Directorio Activo. Todos los servidores de la plataforma utilizarán el sistema DNS interno para su resolución de nombres. Los servidores SMTP gateway que envíen mensajes a Internet deberán contar con DNS externos en caso de que no utilicen un smart host.

La integración del servicio DNS con el servicio de Directorio Activo simplifica de forma significativa la administración del espacio interno de nombres DNS. Al mismo tiempo, se podrán administrar las zonas públicas en el DNS externo.

2.9.- ALMACENAMIENTO

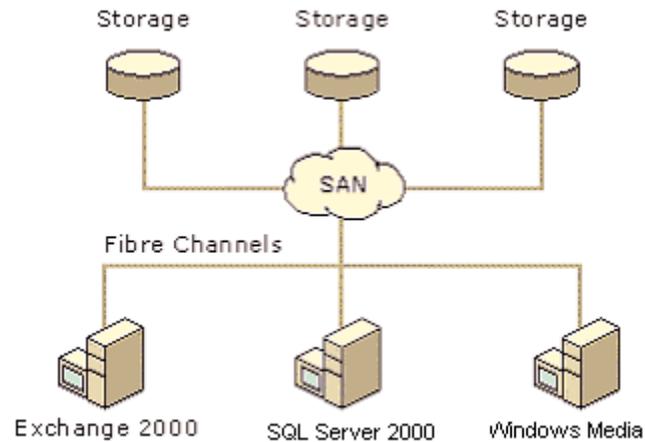
STORAGE AREA NETWORK

La solución de almacenamiento planteada en la arquitectura propone utilizar una SAN *Storage Area Network provista por EMC (sistema Symetrix)*. Esta transfiere la función de almacenamiento de los servidores de uso general a una red de alta velocidad diseñada específicamente para mover una gran cantidad de datos y proporciona gestión centralizada.



Los principales componentes son:

- Fibre Channel switching technology.
- Storage systems.
- Software de gestión de la SAN y del almacenamiento.



En general, las ventajas que aporta utilizar tecnología Fiber Channel son:

- Disminución de las necesidades de proceso por parte de la CPU ya que se asume que la propia SAN tiene capacidades de proceso.
- Posibilidad de mantener a distancias mayores los servidores del almacenamiento.
- Las operaciones de backup pueden realizarse a través del Fiber Channel.



- Posibilidad de utilizar volúmenes mayores.
- Mayor fiabilidad y disponibilidad (se elimina el único punto de error).
- Mayor rendimiento debido al aumento de caché en la controladora de disco.

2.10.- NETWORK LOAD BALANCING

Generalmente para configurar un cluster NLB no es necesario que cada nodo tenga más de un adaptador de red pero para asegurar el máximo rendimiento de la red, se recomienda poner un segundo adaptador.

Network Load Balancing, puede operar en dos modos: unicast y multicast.

En modo *unicast*, se asigna la direcciones MAC del cluster al adaptador de red de la máquina, mientras que la dirección MAC propia de cada adaptador se deshabilita. La configuración de un cluster NLB, en modo unicast con un solo adaptador por nodo, tiene las siguientes limitaciones:

- No es posible la comunicación ordinaria de red entre los nodos del cluster
- El tráfico de red dirigido a otros equipos dentro del cluster, generará sobrecarga de red adicional para todos los equipos del cluster.

En modo *multicast*, se asigna la direcciones MAC del cluster al adaptador de red de la máquina, pero las direcciones propias de cada adaptador se mantienen, la primera para el tráfico entre los clientes y el cluster y la segunda para el tráfico de red específico de la máquina. En este modo, si los clientes acceden al cluster NLB a través de un router, se deben cumplir los siguientes requerimientos:

- Aceptar una respuesta ARP que tiene una dirección MAC en el payload de la estructura ARP, pero parece llegar de una estación con otra dirección MAC, como se identifica en la cabecera Ethernet.
- Aceptar una respuesta ARP que tiene que tiene una dirección MAC multicast en el payload de la estructura ARP.



Esto permitirá al router mapear la dirección IP primaria del cluster y otras direcciones multihomed a la dirección MAC correspondiente. Si el router no cumple estos requerimientos, se puede crear una entrada estática ARP.

3.- CONCLUSIONES

La arquitectura tecnológica de referencia para la construcción del Portal de Turismo presenta las siguientes características:

- solución integrada. Un único proveedor de software lo que está facilitando en gran medida la implantación y la integración entre los distintos módulos de la arquitectura
- plataforma de desarrollo integrada. Los aplicativos y componentes se están desarrollando con herramientas muy aceptadas en la comunidad de desarrolladores (páginas ASP, Visual Basic, java script)
- solución basada en alta disponibilidad
- solución escalable horizontal y verticalmente