

Repaso

November 11, 2016

1 Repaso general

1.1 Términos generales

1.1.1 La Web o WWW (World Wide Web)

El internet es la red que permitió la comunicación de distintas redes de muy diversos protocolos de comunicación física, unificando la comunicación bajo el par de protocolos TCP/IP.

Estos protocolos permiten la transferencia de paquetes por la red, reiterando su envío de ser necesario. Es uno de los más confiables.

Sin embargo el internet era hasta mediados de los 80's, bastante complicado de operar. Se empezaba a dar un boom en la aportación de datos gratuitamente con diversos servicios como el correo electrónico, el ftp para transferir archivos punto-a-punto, las conexiones punto-a-punto con servidores (telnet, ahora ssh), el gopher (antepasado del WWW) para navegar por listados de archivos situados en servidores específicos para ciertos temas, por lo general situados en universidades u oficinas de gobierno norteamericanas) y los boletines de noticias (antepasado de los blogs o foros) donde la gente compartía sus conocimientos, archivos, etc.

El WWW, invento de Tim Berners-Lee (TBL) que trabajaba en ese entonces para el CERN, vino a poner bajo una única plataforma, el navegador, todos estos servicios y a proveer muchísimos más. Mediante el paradigma del *hipertexto*, permitió, junto con los buscadores que comenzaban a ser populares, la comunicación fácil e intuitiva de la información.

TBL por supuesto no realizó todo, empresas como Netscape Communications y Sun Microsystems comenzaron a desarrollar navegadores. Otras gentes desarrollaron protocolos, lenguajes, algoritmos de compresión, bases de datos, esquemas de seguridad, etc. El protocolo que dió un impulso al WWW, y es en el que se basa, es el HTTP (*Hypertext Transfer Protocol*) que permitió la inclusión de tramas de archivos de múltiples tipos (textos, imágenes, sonidos, videos y más hipertextos) juntos para formar una página web.

Además, el espíritu colaborativo de esa época provocó la creación de diversos servidores web gratuitos, el más popular hasta la fecha sigue siendo uno de software libre: Apache, realizado por la fundación del mismo nombre.

1.1.2 Servidor web

El protocolo HTTP permitió la transferencia de hipertextos, junto con sus archivos asociados por la red. Pero lo que permitió almacenar e iniciar la transferencia de la información son los *servidores web*. Estos manejando el esquema *cliente - servidor*, permiten atender la solicitud de cierto hipertexto por parte del cliente (programa *navegador*), enviando el texto base y todos los archivos asociados por la red.

El hardware servidor, es el equipo de cómputo (o máquina virtual corriendo a su vez en un equipo de cómputo), que contiene por lo general programas *servidores*, es decir, que atienden solicitudes provenientes de la red desde los *clientes*.

Un demonio (daemon) es programa que está continuamente corriendo en un hardware servidor y, por lo general, está continuamente monitoreando un servicio, como puede ser un puerto de comunicaciones del internet.

Los *puertos de comunicaciones* son una metáfora de los puertos marítimos que puede tener una región geográfica, en donde llegan las solicitudes de servicios. Los puertos se identifican por un número y hay 1024 puertos reservados por el sistema operativo y otros 64512 puertos disponibles para aplicaciones varias (aunque la tradición ya ha hecho de varios de ellos reservados para aplicaciones específicas).

Cada servicio atendido por los puertos es atendido mediante un *protocolo*. Algunos protocolos interesantes con sus números de puerto son el 22 (ssl) de transmisiones encriptadas, el 21 de ftp, el 23 de telnet, el 25 de smtp (correo electrónico), el 42 de servidor nombres, el 53 de dominios, el 80 de http, el 443 de https (http encriptado), el 525 de conexióna relojes atómicos, el 631 de impresión, el 2049 dns (para compartir sistemas de archivos), etc. (ver archivo /etc/services para un listado de estos puertos y protocolos asociados).

Así, si se desea instalar un servidor web, hay que instalar o crear un programa que atienda el puerto 80 (http), recibiendo conexiones que el sistema operativo canaliza por ahí.

1.1.3 Programación web

Para muchas soluciones, las páginas estáticas en HTML no son suficientes.

En ocasiones se requiere de cierta interactividad con el usuario. Una aplicación de escritorio da eso. Sin embargo, con una aplicación de escritorio muchas veces no es fácil su conexión con bases de datos actualizadas centralmente, o la comunicación con más usuarios, o su utilización en múltiples plataformas con el mismo aspecto general.

Además, si se desea la facilidad de continuar el uso del sistema en otro equipo distinto del original sin requerir instalaciones, una aplicación de escritorio simplemente no lo puede hacer. Para ello existen las aplicaciones web.

1.1.4 ¿Por qué Java?

Actualmente se puede programar en varios lenguajes para web. Ruby-on-rails, Python, PHP, .Net y Java son los más conocidos. De estos, Java es de los más utilizados, tanto en la industria como en el medio educativo.

Su facilidad de uso es media, pero tiene como ventaja su robustez y su facilidad para implementar diversos modelos de diseño.

1.1.5 Arquitecturas web

Actualmente se utiliza la arquitectura separada por *tiers* y por *capas*, afin de que esté claramente indentificadas las tareas de cada parte.

Los *tiers* se refieren a la separación en equipos físicos implementando distintas etapas del sistema.

Las *capas* implementan distintas etapas de la programación dentro del tier, haciendo esta programación *desacoplada* esto es, con muy baja o nula inter-dependencia (pueden funcionar independientemente de otras capas).

Las capas principales son: 1. Presentación (vista) 2. Lógica de negocios (modelo / controlador) 3. Datos (modelo)

Algunas formas comunes de separar una aplicación web son las dos y tres capas:

Dos capas: Se pueden tener dos capas (presentación y lógica) en el tier del `cliente` y, a través de la red, conectarse al tier del `servidor` conteniendo la capa de base de datos.

Tres capas: El primer tier sólo maneja la capa de presentación, el segundo tier contiene una capa adicional de presentación, una de lógica y una de de integración y el último tier sólo maneja la capa de la base de datos.

El saber separar una aplicación en capas y tiers, permite poder implementar dicha aplicación web como aplicación de escritorio o aplicación distribuida, incluso como aplicación para dispositivos móviles, con un mínimo esfuerzo.

1.1.6 MVC

En cuanto a modelos de diseño, uno de ellos es el MVC (Modelo - Vista - Controlador), que como las capas antes vistas, separa el código en porciones específicas para interactuar con el usuario (vista), porciones específicas para la implementación de lógica de negocios, los algoritmos en sí (modelo) y el código que sirve para interconectar la vista y el modelo (controlador).

1.1.7 XML, HTML

XML (Extensible Markup Language), desarrollado por el WWW Consortium, es un lenguaje utilizado para la transmisión a través del internet de información estructurada, aunque también es utilizado como formato de archivo. Por ejemplo, OpenOffice, LibreOffice y las versiones recientes de Microsoft Office, lo utilizan como formato de almacenamiento.

Uno de sus objetivos es ser de fácil lectura tanto por humanos como por máquinas.

Un XML consiste de información marcada con *etiquetas*, por lo general una de inicio y una de fin de marca. Por ejemplo:

```
<nota>
  <titulo negritas="True">Frutas a adquirir</titulo>
  <contenido>
    <lista>
      <elemento>
        <nombre>manzana</nombre>
        <cantidad unidad="kilogramo">2</cantidad>
      </elemento>
      <elemento>
        <nombre>piña</nombre>
        <cantidad unidad="pieza">1</cantidad>
      </elemento>
      <elemento>
        <nombre>naranja</nombre>
        <cantidad unidad="kilogramo">4</cantidad>
      </elemento>
    </lista>
  </contenido>
</nota>
```

Aquí se puede notar que: * Las etiquetas tienen una *jerarquía*, es decir, unas existen *dentro* de otras. * También pueden tener *atributos* (como “negritas” o “unidad”) que tienen a su vez un valor. * La marca para fin de etiqueta es igual a la del inicio con una diagonal luego del signo “<”.

Existen programas para revisar (*parsing*) de archivos XML. El XML tiene usos muy variados.

Todo XML debe tener entre su descripción, una referencia al DTD (Documento Type Definition) en donde se define el lenguaje y sintáxis a utilizar para interpretar el archivo. Por ejemplo:

```
<?xml version="1.0" encoding="UTF-8"?>
```

HTML es un tipo especial de XML con una sintáxis específica para hipertextos. Es decir, se incluyen elementos tipográficos indicando secciones de texto (títulos, listas), elementos para interactividad con el usuario (formularios), elementos para incluir otros archivos en el hipertexto o para referirse a ellos (ligas) en otra parte del sitio web o de otro servidor web. También se incluyen algunas libertades, como etiquetas que no cierran y referencias o URL (Uniform Resource Locator) que es una cadena para referir de manera única a un elemento (por lo general, un archivo de hipertexto, imagen, video etc.) localizable en internet.

(revisar el sitio <http://w3schools.com> para un curso sobre XML y HTML)

1.1.8 Java, Servlets, JSP, Tomcat

El *lenguaje Java*, creado inicialmente por James Gosling que trabajaba para Sun Microsystems, inicialmente fué desarrollado para aplicaciones de escritorio. Al desarrollarse los navegadores, uno de los primeros enteramente escrito en Java, también se utilizó para generar gráficos interactivos en ellos (*Applets*).

Sin embargo los applets eran archivos independientes, cuya interactividad estaba restringida y contenida en una ventana/area de memoria independiente del resto de la página web (modelo *sandbox*, como las cajas de arena donde juegan los niños gringos).

Así Java, tras volverse el monstruo de lenguaje que es actualmente, ya permite que sus clases trabajen tanto en aplicaciones de escritorio como en aplicaciones web. Esta múltiple modalidad se permitió a través de los *servlets* que son clases que permiten recibir y emitir información, tanto desde un HTML como de otros servlets.

Un JSP por su parte, es un tipo de archivo que facilita la generación de vistas, mediante la incorporación a páginas web tradicionales, de elementos que incluyen instrucciones tradicionales de Java y de comunicación con servlets. De hecho un JSP no se ejecuta por si mismo, sino que se tiene que traducir a un servlet, que además de imprimir el código HTML que incluye, incluye las instrucciones java en él.

Al crecer la utilización del lenguaje Java para la creación de sitios web, se ideó una forma de compartir información, mantener y proveer servlets a los clientes con el protocolo HTTP, por lo que la fundación Apache desarrollo el Repositorio de servlets *Tomcat*, que además de ser de software libre, es el más conocido para esta tarea. De esta manera se eliminó la necesidad de contar además de un servidor web tradicional pues el Tomcat además provee esta función.

1.1.9 JDBC

Una *manejador* o *gestor de bases de datos* es un programa que almacena y administra las peticiones a bases de datos. Dichas peticiones provienen de clientes que se conectan a este solicitando información, modificando bases de datos, etc. Algunos ejemplos de manejadores de bases de datos incluyen MySQL, Oracle y PostgreSQL. Este último se utilizó en este curso debido a que es software libre, es de los más populares, cumple con toda la especificación SQL (*Structured Query Language*) y es bastante robusto y rápido.

Una *base de datos* es un conjunto de archivos y métodos optimizados de acceso, conteniendo los datos a trabajar por medio de una aplicación que puede o no ser web.

Una *base de datos relacional* es aquella en la que los datos se almacenan en estructuras tabulares con múltiples *registros* en los renglones, cada uno de ellos con una cierta cantidad y tipo de *campos*. Algunos de estos campos son especialmente designados como *claves* (también llamadas *llaves*) con propiedades como la unicidad, y que representan la información de los otros campos. Además se cuenta con *relaciones* entre tablas, de manera que las claves de unas tablas son campos en otras. A las bases de datos se les ha identificado propiedades como las *formas normales*, que por ejemplo, caracterizan a tablas para las cuales no hay repetición de información, etc.

Desde Java, podemos acceder a un manejador de bases de datos mediante un *Controlador* de la biblioteca JDBC (Java Database Connectivity). Por ejemplo, para acceder al controlador de PostgreSQL utilizamos la instrucción: `Class.forName("org.postgresql.Driver");`

Las principales operaciones a realizar con una base de datos relacional son la consulta (SELECT), la inserción de datos (INSERT), la modificación de datos (ALTER), etc.