



1. INTRODUCCIÓN Y OBJETIVOS

La finalidad de este proyecto fin de carrera es realizar un estudio comparativo entre dos sistemas de gestión de bases de datos bajo entorno Linux.

El Sistema Operativo Linux se ha extendido de forma muy rápida e importante en todo el mundo, superando incluso las expectativas más optimistas. Aunque inicialmente fue concebido como una plataforma barata para PC lo más parecida al profesional sistema UNIX, en la actualidad Linux se adapta perfectamente tanto a los ordenadores de sobremesa como a los grandes servidores UNIX a los cuales quiso inicialmente emular.

La principal característica de Linux es su carácter de software libre, lo que quiere decir que se puede adquirir este sistema operativo de forma gratuita, y no solo eso, sino que también se tiene acceso al código fuente, escrito en su mayoría en lenguaje C. Este código se puede modificar como se desee y se puede compilar libremente. Este hecho ha provocado que en la actualidad existan gran cantidad de distribuciones (algunas de ellas comerciales) de este Sistema Operativo, pero todas bajo un *kernel* común.

Tradicionalmente, una de las desventajas de Linux era la carencia de una interfaz de usuario, puesto que no todos los usuarios se encuentran cómodos con una interfaz basada en líneas de comandos. Pero desde hace bastante tiempo Linux viene evolucionando mucho en lo que a interfaz de usuario se refiere, y como prueba de ello se tienen los diferentes gestores de ventanas que pueden ser utilizados, que proporcionan al usuario un entorno muy parecido al que suele estar acostumbrado (habitualmente Windows).

No obstante, a pesar de su carácter libre, Linux no está muy extendido entre los usuarios “de sobremesa”, que habitualmente prefieren el uso de una plataforma Windows. Pero en el ámbito profesional, cada vez son más los usuarios y empresas que se decantan por éste sistema operativo. En concreto, se puede decir que Linux es uno de los Sistemas Operativos más ampliamente utilizados para proporcionar servicios de Internet, tales como servidores Web, servidores ftp, servidores de correo electrónico, y lo que realmente se abordará en este proyecto, servidores de base de datos.

Hoy en día, Linux es el Sistema Operativo más usado para la instalación de servidores públicos de Web. Según se indica en la publicación “Manual Avanzado de Red Hat Linux 8”, en el año 2002 se calculaba que en torno al 30 % de servidores Web usaban una distribución de Linux.

A continuación se enumeran algunas de las características de Linux que demuestran el por qué de su amplio uso para proporcionar servicios de Internet, y por ende, de sus ventajas para trabajar con gestores de bases de datos:

- Multitarea.
- Multiusuario.
- Multiplataforma.
- Portabilidad de código.
- Mecanismos de protección de memoria entre procesos.
- Soporte nativo del conjunto de protocolos TCP/IP.
- Herramientas de administración de sistema.



- Gestión de memoria compartida.
- Memoria virtual mediante técnicas de planificación de disco.
- Posibilidad de multisesión en un mismo puesto de trabajo.
- Uso de librerías tanto estáticas como dinámicas.
- Gratuito.

En cuanto a los motivos que pueden decantar el uso de Linux para gestores de bases de datos, uno de los más importantes es que Linux está pensado para implementar sin problemas el paradigma cliente/servidor. Además, actualmente existen muchos programas servidores de bases de datos disponibles bajo Linux, así como multitud de herramientas para crear aplicaciones cliente, siendo algunas de estas bases de datos de libre distribución. Algunos de los sistemas gestores de bases de datos bajo entorno Linux son Oracle, MySQL, PostgreSQL, Sybase e Informix. Existen otras, pero quizá no estén tan extendidas como lo pueden estar las anteriormente citadas, o bien no tienen una potencia media-alta en lo que se refiere al manejo de datos. De las bases de datos anteriores, probablemente las más utilizadas sean las tres primeras, según datos de los propios fabricantes. Además, sólo MySQL y PostgreSQL son de libre distribución. Oracle, Sybase e Informix son sistemas gestores de bases de datos comerciales. No obstante, de estas tres últimas se puede disponer de versiones gratuitas o de evaluación. A continuación se muestra una pequeña descripción de estas bases de datos, de entre las cuales, MySQL y PostgreSQL serán objeto de análisis y estudio en este Proyecto Fin de Carrera:

- **Oracle:** Oracle es una base de datos distribuida por la compañía Oracle Corporation. Es una base de datos relacional para entornos cliente/servidor, es decir, que aplica las normas del álgebra relacional (conjuntos, uniones, intersecciones,...) y que utiliza la arquitectura cliente/servidor, donde en un lado de la red está el servidor con los datos y en el otro lado están los clientes que interrogan al servidor. Oracle está disponible para varios sistemas operativos y plataformas (Windows, Linux, Solaris, Mac OS/X, AIX5L, HP-UX, IBM z/OS, HP Tru64 UNIX). Actualmente es una de las bases de datos cliente/servidor más utilizadas, a pesar de su carácter comercial.
- **MySQL:** La base de datos MySQL es uno de los sistemas gestores de bases de datos de código abierto más populares del mundo. Se usa en más de 6 millones de instalaciones. Es multiplataforma, es decir, que es capaz de correr bajo los diferentes Sistemas Operativos más utilizados de forma flexible (Linux, Windows, Mac OS/X, HP-UX, AIX, Netware, Amiga, BSDI, Digital Unix, FreeBSD, NetBSD, Novell NetWare, OpenBSD, OS/2 Warp, Solaris, SunOS). También es una base de datos del tipo cliente/servidor y relacional, y es distribuida por la compañía MySQL AB. Es una base de datos que tiene una gran cantidad de aplicaciones adicionales y que es soportada por una gran cantidad de software actual. Suele venir incluida en las distribuciones de Linux.
- **PostgreSQL:** PostgreSQL es una base de datos relacional, de código abierto y de tipo cliente/servidor. Es soportada por varios sistemas operativos, como son Linux, UNIX (AIX, BSD, HP-UX, SGI IRIX, Mac OS X, Solaris, SunOS, Tru64), BeOS y Windows. Soporta grandes objetos binarios y tiene interfaces de programación nativos para varios lenguajes de programación. Además, posee una gran cantidad de aplicaciones complementarias, como clientes gráficos e interfaces de programación de aplicaciones. Esta distribuida por



PostgreSQL Global Development Group, y al igual que MySQL, suele venir incluida en las distribuciones de Linux.

- Sybase: Sybase proporciona una gran cantidad de productos para la gestión de la información. Entre ellos, se encuentra Sybase IQ, que es una base de datos relacional altamente escalable optimizada para obtener un rendimiento de consulta bastante bueno en grandes “data warehouses”. Posee una gran cantidad de aplicaciones adicionales para el manejo de bases de datos. Según el fabricante, se trata de una base de datos muy rápida, con sofisticados algoritmos de compresión de datos, fácil de usar, escalable y flexible. Quizá la característica que más llama la atención de esta base de datos es lo referente a la compresión, pues según el fabricante puede llegar a reducir el volumen de los datos almacenados en un 70%, lo cual es de gran interés para bases de datos con grandes cantidades de información. Este sistema gestor de bases de datos se encuentra disponible para IBM AIX, HP-UX, Linux, Solaris, Windows y UNIX. Este gestor de bases de datos no es de libre distribución, sino comercial, pero se puede descargar una versión de prueba desde la Web del fabricante.
- Informix: Informix es la solución del fabricante IBM para la gestión de bases de datos relacionales en modelo cliente/servidor. IBM Informix SQL es un sistema de desarrollo de aplicaciones de bases de datos que proporciona una velocidad, potencia y seguridad necesarias para las aplicaciones grandes y pequeñas. Los sistemas operativos y plataformas hardware apropiadas para su uso son, según el fabricante, HP-UX, Linux, Sun-Solaris y Unix. Informix SQL es un paquete que incluye varias herramientas de desarrollo de aplicaciones para bases de datos, es sencilla de manejar, pues con estas aplicaciones permite desarrollar aplicaciones que no requieren ningún lenguaje de programación. También posee funciones de generación de informes sencillas, que permiten analizar datos rápidamente. El inconveniente es que es una aplicación comercial, y gratuitamente IBM sólo distribuye una versión de prueba.

Hoy en día, el software libre se ha consolidado como alternativa, técnicamente viable y económicamente sostenible al software comercial, contrariamente a lo que a menudo se piensa. Por ello, en el presente Proyecto Fin de Carrera se estudiarán dos de las aplicaciones *Open Source* más populares en la actualidad: MySQL y PostgreSQL.

Como objetivo general se tiene el elaborar un documento donde se comparen las características y funcionalidades que ofrece cada uno de los sistemas gestores de bases de datos a estudio. Además, para evaluar el rendimiento de cada sistema, se emplearán una serie de bancos de pruebas que evalúen el comportamiento de cada uno de ellos ante determinadas situaciones, y se evaluará el rendimiento de los mismos cuando se usen en aplicaciones reales. De forma resumida, los objetivos fundamentales de este Proyecto Fin de Carrera son:

- Efectuar una comparativa de las características y funcionalidades de MySQL y PostgreSQL.
- Evaluar el comportamiento de MySQL y PostgreSQL en determinadas situaciones mediante el empleo de bancos de pruebas.
- Evaluar el rendimiento de MySQL y PostgreSQL cuando se emplean como “*backend*” de aplicaciones reales.



Una vez realizados estos estudios, se tratará de determinar en qué situaciones puede resultar más ventajoso el uso de uno u otro sistema.