

¿Qué es Lustre File System?

Es un sistema de archivos distribuido de código abierto, usualmente utilizado en clusters a gran escala. Su curioso nombre proviene realmente de la mezcla entre las palabras Linux y cluster. Este proyecto busca proporcionar un sistema de archivos para clusters gigantescos, de decenas de miles de nodos, con Petabytes de capacidad de almacenamiento, sin comprometer la velocidad o la seguridad, y está completamente disponible en GNU GPL.

Lustre File System tiene diseñadores, desarrolladores y colaboradores de otras compañías y de una gran comunidad de particulares.

La gran mayoría de los supercomputadores más rápidos del mundo son cluster que utilizan el sistema de archivos que provee Lustre como almacenamiento.

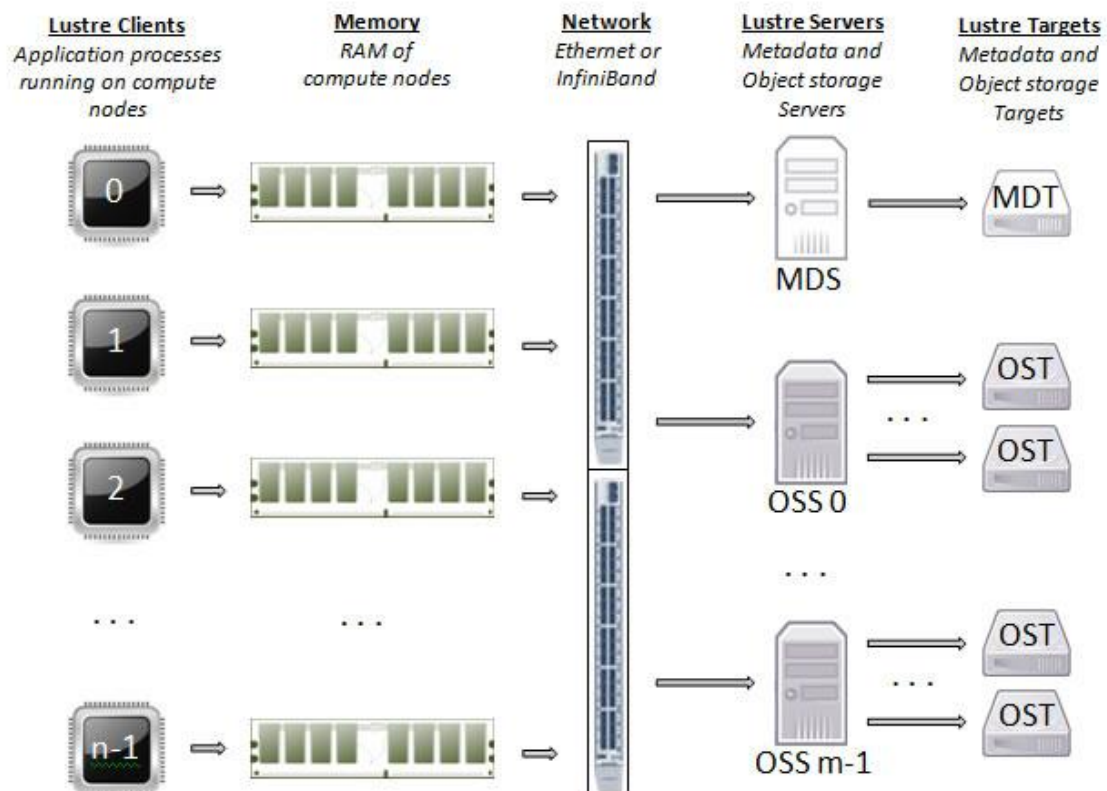
¿Qué ventajas tiene? ¿Qué arquitectura plantea?

Podríamos decir que una de las ventajas de este sistema de archivos es que a cada archivo almacenado en él se le considera un objeto. Le presenta a todos los clientes una semántica POSIX estándar y acceso concurrente de lectura y escritura para objetos compartidos.

Un sistema de archivos Lustre tiene cuatro unidades funcionales que son:

- *Metadata server (MDS)*: Para almacenar los metadatos de archivo.
- *Object Storage Target (OST)*: Para guardar los datos en sí.
- *Object Storage Server (OSS)*: Para manejar los datos reales del OSTs.
- *Clients*: Los usuarios que acceden y utilizan dichos datos.

El esquema básico de ésta arquitectura podría ser:



El MDS, OSS y un OST pueden residir en el mismo nodo o en nodos distintos. Lustre File System no administra directamente los OSTs y delega esta responsabilidad en los OSSs para asegurar la escalabilidad de la red para grandes clusters y supercomputadores de todo el mundo.

En un MPP o Procesador Paralelo Masivo, los procesadores pueden acceder al sistema de archivos Lustre redirigiendo sus peticiones de entrada/salida hacia el nodo con el servicio lanzador de tareas si está configurado como un cliente del sistema de archivos. Debido a que este es uno de los métodos más sencillos, en general proporciona un bajo rendimiento.

Una manera ligeramente más complicada de proporcionar un rendimiento global muy bueno consiste en utilizar la biblioteca *liblustre*. Es una biblioteca de nivel de usuario que permite a los procesadores montar y utilizar el sistema de archivos como un cliente, sorteando la redirección hacia el nodo de servicio. Utilizando esta biblioteca, los procesadores pueden acceder al sistema de archivos incluso si el nodo de servicio en el que se lanzó el trabajo no es un cliente de Lustre. Otra ventaja es que proporciona un mecanismo para mover datos directamente entre el espacio de aplicación y los OSS, sin necesidad de realizar una copia de datos a través del núcleo ligero, logrando así una baja latencia y un gran ancho de banda en el acceso directo de los procesadores al sistema de archivos Lustre. Ésta es la mayor diferencia de diseño entre Lustre en un MPP y los clusters convencionales.