

# Deduplicación: El arte de no repetirse, aplicado al almacenamiento de datos.

*Lortu, como empresa pionera en deduplicación, desarrolló esta tecnología varios años antes de que hubiera incluso un término para definirla.*



## Introducción

A lo largo de los primeros años de la presente década, las empresas empezaron a utilizar de forma masiva dos tecnologías que cambiarían las dimensiones de lo conocido hasta aquel momento, en lo que se refiere a almacenamiento de datos:

- La sustitución del papel por soportes informáticos.
- La interconexión de los sistemas informáticos entre clientes, proveedores y partners gracias Internet.

Adicionalmente al aumento de información comenzaron a utilizarse de forma masiva herramientas como el correo electrónico, bases de datos, etc. que generaban grandes ficheros, que además eran modificados continuamente.

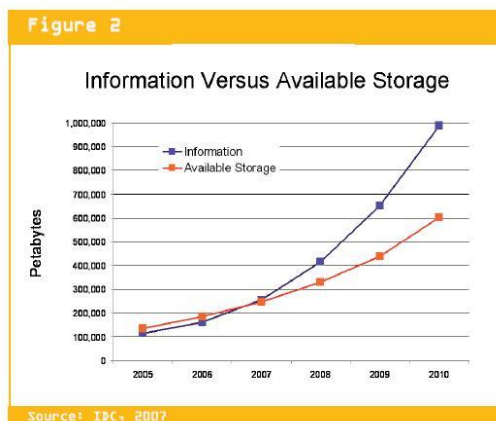
La conjunción de estos dos aspectos provocó que quedara invalidado el método comúnmente utilizado hasta aquel momento para la realización de copias de seguridad: copias diarias incrementales y copias completas el fin de semana, debido a que la información estaba contenida en un solo fichero (correo, bases de datos, etc.) y este cambiaba continuamente. Ello provocaba que las aplicaciones de backup almacenaran todos los datos, todos los días a pesar de tener configurado un backup incremental.

## Innovación = Anticipar una solución, antes de que se produzca el problema.

En febrero de 2003 Lortu Software comenzó a investigar una tecnología que permitiera reducir drásticamente el espacio necesario para almacenar las copias de seguridad de las empresas a partir de la sustitución de datos duplicados por referencias. Tan solo 6 meses después disponíamos de un prototipo con un factor de compactación realmente espectacular. Lortu llamó a esta tecnología Kondar.

Paradójicamente, la industria del Data Storage no concibió un término, comúnmente aceptado, para definir esta tecnología hasta varios años después y no es hasta la segunda mitad del 2006 cuando la tecnología empieza a ser conocida y divulgada por los medios de comunicación especializados, debido en gran medida a la adquisición de dos empresas con tecnología de deduplicación (Avamar y ADIC) por parte de dos de las compañías líderes en Data Storage, como son EMC y Quantum.

Estas dos adquisiciones y otras que se produjeron posteriormente fueron los desencadenantes para que las previsiones de las consultoras líderes como IDC o Gartner, pronosticaran para los años 2007 y 2008 a la deduplicación como tecnología más importante dentro del campo del almacenamiento de datos o Data Storage.



A este hecho hay que añadir que según las previsiones de los analistas en los próximos años el mercado del almacenamiento va a disponer de más datos de los que podrá almacenar.

Otro de los aspectos a reseñar es que más sistemas de almacenamiento equivalen a mayor consumo de energía, por lo que la deduplicación, además de resolver los problemas de almacenamiento de datos, es considerada una **tecnología verde**, debido a su reducido consumo energético.

Otro de los aspectos más importantes que facilita la deduplicación es la capacidad de transferir grandes cantidades de información por un ancho de banda pequeño. Esto posibilita la replicación de datos entre sedes a través de Internet o los servicios de respaldo local + remoto de grandes cantidades de información.

## Lortu Software, la innovación como forma de trabajo.

En Lortu no solamente nos adelantamos varios años en la concepción de la idea y en buscar la solución adecuada a una necesidad futura, además hemos seguido innovando tanto en tecnología como en la forma de comercializarla.

A comienzos del 2006 comenzamos a comercializar nuestra tecnología en forma de componentes a ser integrados en otros productos software o hardware y a lo largo de 18 meses llegamos a contactar con más de 400 fabricantes de cuatro continentes.

Durante este tiempo empresas como: IBM, Fujitsu Siemens, BakBone, FalconStore, ExaGrid entre otras muchas, se han interesado por nuestra tecnología de deduplicación.

Paralelamente se han publicado noticias y artículos sobre nuestra tecnología en medios de comunicación de Europa, América y Asia.

En Abril de 2007 decidimos dar un paso más para innovar no solo en tecnología sino en la forma de comercializarla, convirtiéndonos en el primer fabricante de deduplicación a nivel mundial que ofrece su tecnología como servicio de backup gestionado.

A finales de 2007 terminamos el desarrollo de un appliance de almacenamiento masivo, que permite **almacenar hasta 100 Terabytes**, en el espacio que ocupa un PC y lo que es aún más importante **replicarlo a nuestro DataCenter a través de una línea ADSL convencional**.

A lo largo del primer semestre del 2008 hemos comercializado en Euskadi este appliance como servicio gestionado de backups, de forma que nuestros clientes pueden utilizarlo pagando una pequeña cuota mensual.

Nuestro próximo reto es expandirnos al resto del estado durante el 2008, para lo cual tenemos actualmente varios frentes abiertos y comenzar nuestra expansión internacional el año 2009.

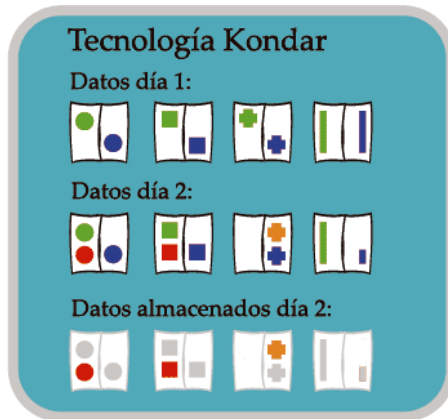
## La tecnología que respalda una trayectoria.

A lo largo de este artículo hemos explicado la historia de la deduplicación y como no podría ser de otra forma, la de Lortu como empresa pionera en la evolución de una tecnología que se encuentra en estos momentos en pleno apogeo y donde todos los grandes del mercado están empezando a trabajar.

Sin embargo, como ocurre en otros campos de la técnica, no todas las tecnologías de deduplicación son iguales.

Una explicación detallada de las diferencias entre las diferentes tecnologías de deduplicación está fuera del alcance de este artículo, pero vamos a explicar brevemente en qué consiste y cuáles son las principales diferencias.

Como decíamos anteriormente, deduplicación es una técnica que permite sustituir datos duplicados por referencias, de forma que ocupen mucho menos espacio.



En el gráfico de la izquierda mostramos un ejemplo simple de cómo funciona la deduplicación. En este caso se trata de comparar los datos de dos días y almacenar solamente los datos del primer día y los datos del segundo día que no se encuentran el primer día.

Para restaurar los datos del día 2 basta con disponer de los datos del día 1 y los datos únicos del día 2.

Esta es una simplificación del proceso para su fácil comprensión, puesto que una tecnología de deduplicación es un proceso más complejo que el mostrado en este ejemplo.

Una de las principales diferencias entre distintas tecnologías de deduplicación es la granularidad utilizada para buscar datos repetidos. En este aspecto nos podemos encontrar los siguientes tipos:

- File level. Esta no es en realidad una tecnología de deduplicación sino el comúnmente conocido como backup incremental o diferencial. Se basa en almacenar los ficheros no modificados en el backup anterior o en el último backup completo.
- Sector level. Esta tecnología almacena solamente los sectores que son modificados. Dos sectores pueden tener el mismo contenido, pero serán almacenados si han sido modificados.
- Block level. Divide la información en bloques y almacena únicamente los bloques únicos.
- Byte level. Es el nivel de deduplicación utilizado por Lortu y el que permite una mayor capacidad de reducción. Solo son almacenados los datos modificados, independientemente de si están alineados a un bloque o no.

Hay otros factores que inciden en la capacidad de reducción y la escalabilidad de las diferentes tecnologías de deduplicación como puede ser:

- Inline o offline
- Agnóstico a los datos o dependiente del formato de los datos
- Basado en comparación o en patrones
- Etc.

La deduplicación, como otras áreas de la técnica, se encuentra en constante evolución y en Lortu, como empresa pionera en este campo estamos trabajando para ofrecer a nuestros clientes las soluciones más innovadoras del mercado y sin complejos ante las grandes multinacionales del sector.

Somos conscientes por nuestra historia que el potencial económico de una empresa no determina necesariamente su capacidad innovadora o de ofrecer las soluciones que nuestros clientes realmente necesitan.

Para más información sobre Lortu Software, sus tecnologías o productos, pueden contactarnos o consultar nuestra página web: <http://www.lortu.es>