

CURSO DE APRENDIZAJE A DISTANCIA
Intercambio de datos interoperables en hidrología

Lección 2.2 **Diferentes soluciones o arquitecturas para el almacenamiento de datos**

- Describir el espectro de arquitecturas disponibles para el almacenamiento de datos
- Determinar cuáles son las ventajas de las distintas soluciones de implementar la arquitectura de almacenamiento
- Describir los sistemas de gestión de bases de datos (SGBD) y las ventajas de utilizarlos
- Explicar qué le aporta el Sistema de gestión de bases de datos meteorológicos, climatológicos e hidrológicos (MCH) a los SMHN

WEATHER CLIMATE WATER

WORLD METEOROLOGICAL ORGANIZATION | WMO HydroHub | UCAR COMMUNITY PROGRAMS COMET | National Research Council of Italy

Bienvenido a la lección 2.2 — Diferentes soluciones o arquitecturas para el almacenamiento de datos. Los objetivos de aprendizaje de esta lección son los siguientes:

- Describir el espectro de arquitecturas disponibles para el almacenamiento de datos.
- Determinar cuáles son las ventajas de las distintas soluciones para implementar una arquitectura de almacenamiento.
- Describir los sistemas de gestión de bases de datos (SGBD) y las ventajas utilizarlos.
- Explicar qué le aporta el SGBD Meteorología, Climatología e Hidrología (MCH) a un Servicio Meteorológico e Hidrológico Nacional.

Diapositiva 2



Aún nos encontramos en la etapa de recopilación de datos de la cadena de valor de los datos, que comprende la gestión de los datos. En esta lección, nos centraremos particularmente en el paso del almacenamiento.

Diapositiva 3



A continuación presentaremos las diferentes arquitecturas de almacenamiento de datos. Como el método de almacenamiento de datos en las diversas arquitecturas es distinto en cada caso, clasificaremos las arquitecturas de la menos estructurada a la más estructurada:

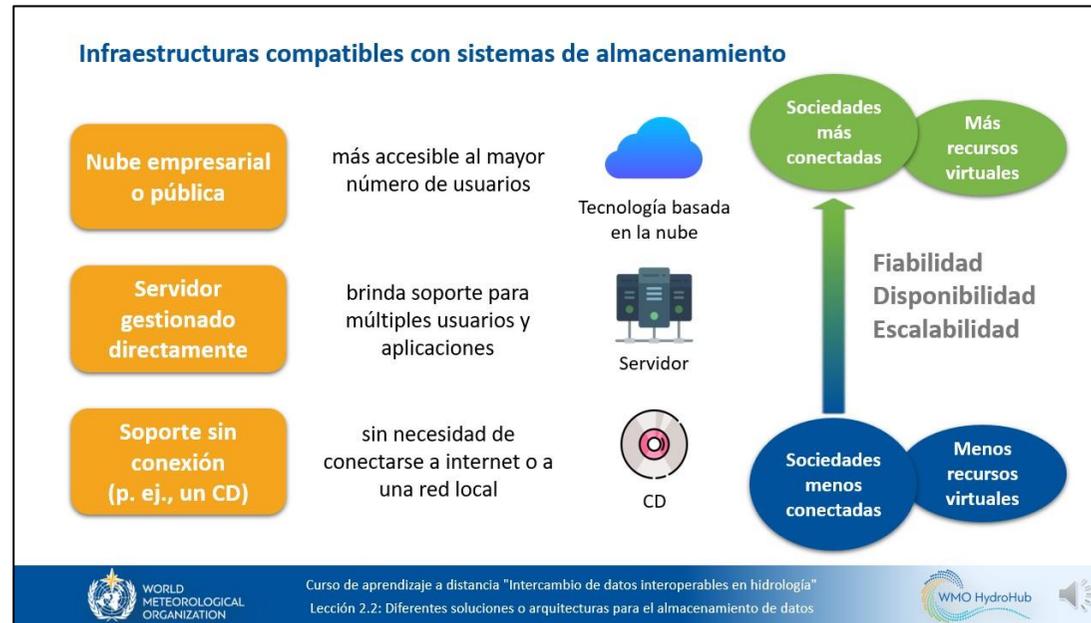
- Sistema de archivos: es una jerarquía de carpetas y archivos, con una estructura mínima en el contenido.
- Almacenamiento de objetos: es una arquitectura en la que los objetos se identifican mediante un ID único que permite localizar y recuperar el contenido asociado a cada objeto.
- Base de datos relacional: se trata de un esquema personalizado con objetos y sus propiedades. Por lo general, todos los datos o la información se almacenan en tablas que se pueden vincular o relacionar en función de los datos que tengan en común.
- Base de datos de series temporales: es una familia de bases de datos específica optimizada para almacenar objetos de series temporales (es decir, cada registro tiene un elemento temporal). Las bases de datos de series temporales también se pueden crear a partir de la arquitectura de bases de datos relacionales.

De izquierda a derecha, los datos y la información se almacenan de forma más estructurada, pero su desarrollo, uso y mantenimiento son más complejos.

Existen otras soluciones avanzadas para el almacenamiento de datos, por ejemplo, las bases de datos NoSQL y los cubos OLAP, que están especialmente diseñadas para el análisis en línea de datos multidimensionales. Estas soluciones avanzadas pueden utilizarse sobre todo para macrodatos. Si desea obtener más información al respecto, consulte la sección de referencias del curso.

Una o varias de estas arquitecturas pueden adaptarse para formar una base de datos hidrológica, lo que ofrece una solución de almacenamiento que facilita la incorporación de observaciones, mediciones y metadatos hidrológicos..

Diapositiva 4



Las arquitecturas de almacenamiento mostradas en la diapositiva anterior pueden tener diferentes implementaciones físicas, desde soportes sin conexión hasta una solución de servidor, pasando por una solución comercial o basada en la nube. Todas ellas tienen sus ventajas:

- El soporte sin conexión (solución independiente) es una solución de almacenamiento en la que los datos se guardan sin necesidad de conectarse a internet o a una red local. Cabe citar, a modo de ejemplo, los datos almacenados y disponibles en un CD. Actualmente, este mecanismo se utiliza principalmente para las copias de seguridad de datos y no para las arquitecturas de datos.
- El servidor gestionado directamente (solución de servidor centralizado) es un mecanismo de almacenamiento en el que los datos se almacenan en el servidor de la organización y están a disposición de los usuarios de la red local o de internet. La ventaja de esta solución es que puede funcionar sin conexión a internet y, a diferencia de la solución anterior, la capacidad de almacenamiento es mayor y es compatible con aplicaciones de múltiples usuarios y admite el acceso desde múltiples usuarios.
- La nube empresarial o pública es un mecanismo de almacenamiento en el que los datos se guardan en una infraestructura distribuida gestionada por una organización u otras entidades. Por ejemplo, una organización utilizaría una nube pública o una nube local gestionada por el departamento técnico de una institución.

La ventaja de esta última solución, siempre que se disponga de una buena conexión a internet, es que es más fiable, incluso al poner datos en común. Asimismo, se garantiza que un mayor número potencial de usuarios pueda acceder a los datos, a la vez que se mejora el desempeño. Gracias a los rápidos avances tecnológicos, las sociedades están cada vez más conectadas, lo que les permite construir programas informáticos que dependen de recursos cada vez más virtualizados. El objetivo de esta virtualización es conseguir que los recursos sean más fiables, extrapolables y que estén más disponibles. Sin embargo, este proceso no avanza al mismo ritmo en todos los países. Esto se debe principalmente a la disponibilidad de conexión a internet y de servicios en la nube.

Diapositiva 5

Infraestructuras compatibles con sistemas de almacenamiento: servicios basados en la nube

- Se amplía la disponibilidad de estos servicios;
- infraestructura práctica para crear aplicaciones fiables y optimizadas;
- solución para que las organizaciones que dispongan de recursos limitados mantengan un servidor local;
- más fiables para que los usuarios dispongan de los datos las 24 horas al día, 7 días a la semana.



WORLD
METEOROLOGICAL
ORGANIZATION

Curso de aprendizaje a distancia "Intercambio de datos interoperables en hidrología"
Lección 2.2: Diferentes soluciones o arquitecturas para el almacenamiento de datos



Cuanto mayor es la disponibilidad de los servicios en la nube, más práctica resulta su infraestructura para crear aplicaciones fiables y optimizadas. Los servicios en la nube también pueden ser una solución para que las organizaciones que dispongan de recursos limitados mantengan un servidor local, ya que ahorran en el costo de mantenimiento. En el caso de un servidor local, se necesitan más recursos informáticos y humanos para el mantenimiento de los equipos físicos.

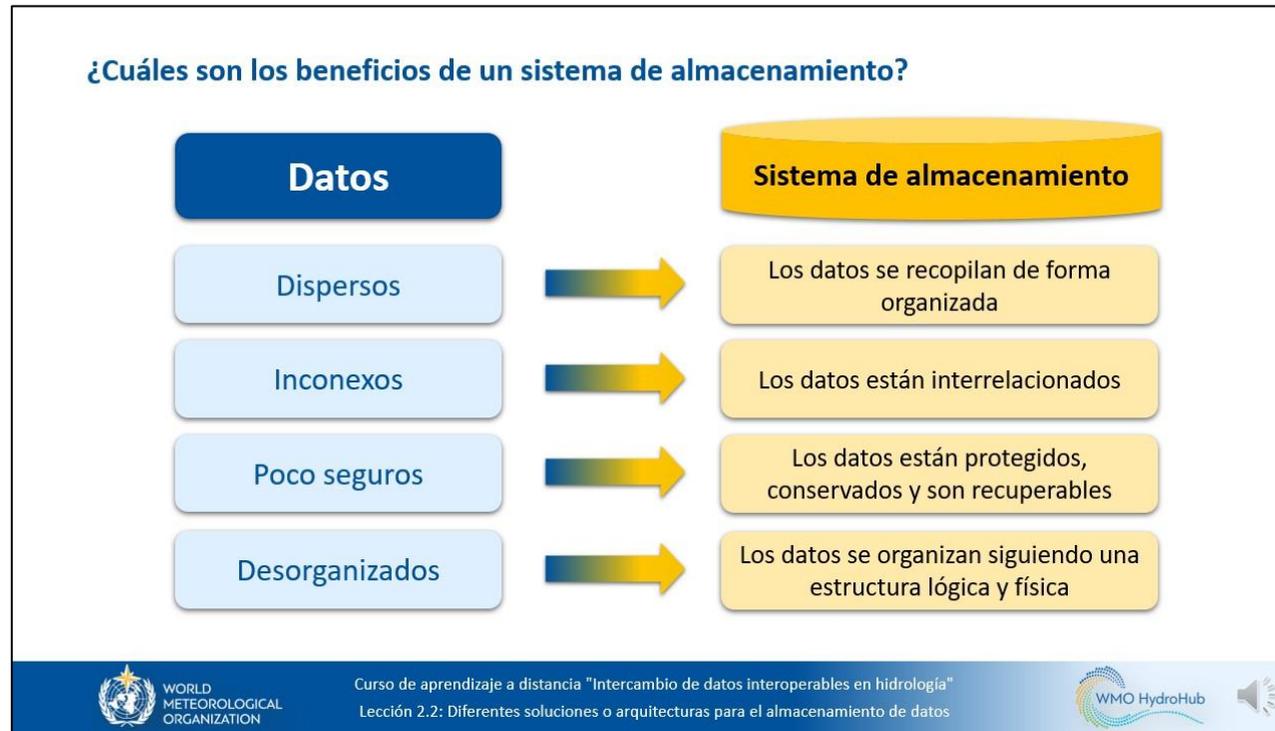
Asimismo, los servicios en la nube son más fiables para que los usuarios dispongan de los datos las 24 horas al día, 7 días a la semana.

Diapositiva 6



Nombremos ahora las conocidas soluciones tecnológicas adoptadas para implementar los distintos sistemas de almacenamiento de datos. En los recuadros de color azul claro, se muestran ejemplos de tecnologías disponibles para arquitecturas de almacenamiento de datos. En los recuadros de la derecha hay ejemplos de algunas soluciones *ad hoc* personalizadas para datos hidrológicos que ya están disponibles y que utilizan los servicios hidrológicos nacionales. Entre ellas figuran CUAHSI HydroServer; Aquarius; WISKI; 52North SOS; Meteorología, Climatología e Hidrología (MCH), o HYDROACCESS.

Diapositiva 7



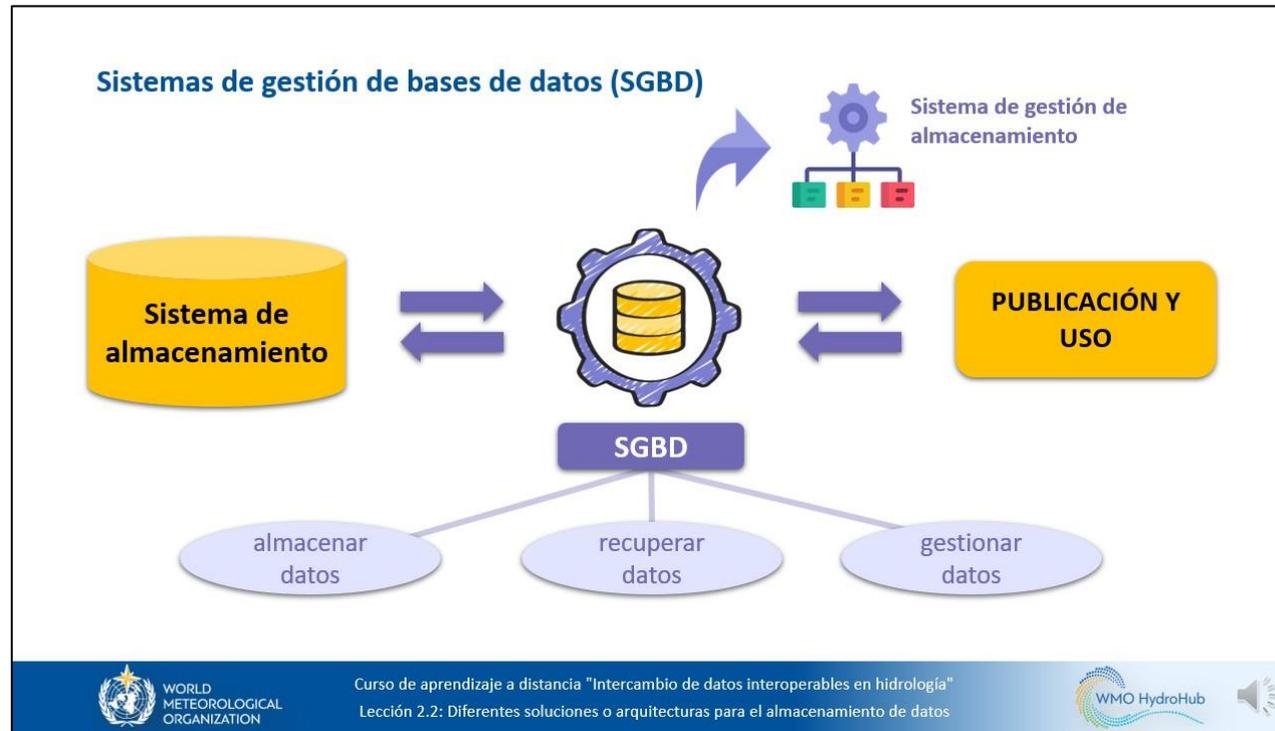
¿Cuáles son los beneficios de un sistema de almacenamiento?

Si no se dispone de un sistema de almacenamiento, los datos estarán dispersos, inconexos, poco protegidos y desorganizados. Los sistemas de almacenamiento contribuyen a que los datos:

- se recopilen de forma organizada;
- estén interrelacionados;
- estén protegidos, sean objeto de mantenimiento y sean recuperables;
- se organicen siguiendo una estructura lógica y física.

Normalmente, para almacenar, administrar y gestionar datos en el sistema de almacenamiento se utiliza un sistema de gestión de bases de datos (SGBD).

Diapositiva 8

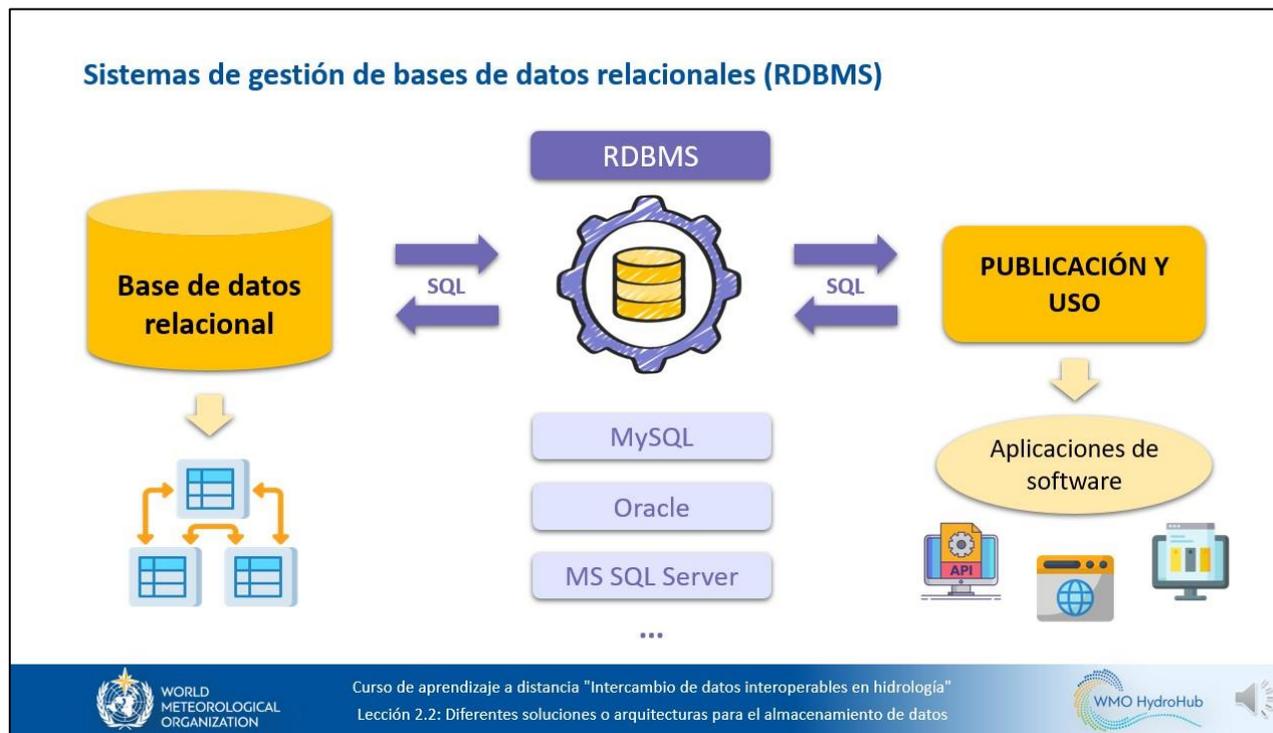


Los SGBD son aplicaciones informáticas utilizadas para crear, gestionar y administrar el sistema de almacenamiento.

El principal objetivo de los SGBD es realizar con eficiencia operaciones como el almacenamiento, la recuperación y la gestión de datos.

A nivel interno, los SGBD pueden estar formados por grupos de programas que llevan a cabo diferentes tareas relacionadas con la gestión del sistema de almacenamiento. Veamos un ejemplo de un SGBD creado específicamente para bases de datos relacionales.

Diapositiva 9



Un sistema de gestión de bases de datos relacionales es un programa que permite crear, actualizar y administrar una base de datos relacional.

Casi todos los sistemas de gestión de bases de datos relacionales utilizan el lenguaje de consulta estructurado (SQL) para ejecutar diversas operaciones en la base de datos, como gestionarla y realizar consultas. Las aplicaciones informáticas (como las aplicaciones web, los sitios web dinámicos y demás) también utilizan el lenguaje SQL para comunicarse con el sistema de gestión de bases de datos. Todas estas operaciones son fundamentales para la recuperación, la publicación y el uso de los datos.

Algunos de los sistemas de gestión de bases de datos relacionales más populares y disponibles a mayor escala son, entre otros, MySQL, Oracle y MS SQL Server.

Se utilizan mucho en todo el mundo y, por tanto, han sido probados y respaldados por los proveedores (en el caso de las soluciones comerciales), por las comunidades de desarrolladores (en el caso de las soluciones de código abierto) y por los usuarios.

Diapositiva 10

Sistemas de gestión de bases de datos hidrológicos

- Adaptados para manejar específicamente tipos habituales de datos hidrológicos;
- permiten gestionar, corregir y representar gráficamente los datos;
- muchas de las herramientas de gestión de datos han sido dotadas de capacidad de análisis.



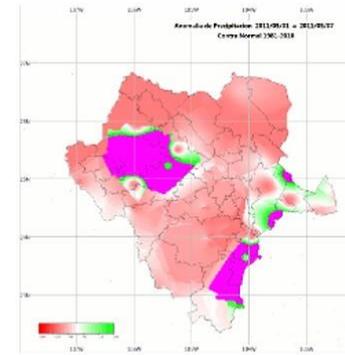
Los sistemas de gestión de bases de datos relacionales y otros sistemas de gestión de bases de datos se pueden adaptar para manejar específicamente tipos habituales de datos hidrológicos. Por ejemplo, es posible que ya se hayan creado tablas de bases de datos y rutinas de acceso que manejan específicamente datos hidrológicos y que almacenan la información descriptiva apropiada y los metadatos.

El sistema suele disponer de un programa informático que permite gestionar, editar y representar gráficamente los datos. Además, muchas de las herramientas de gestión de datos han sido dotadas de capacidad de análisis, por ejemplo, para producir curvas de duración de flujo a partir de datos de flujo, y de métodos estadísticos para ajustar las distribuciones a los picos de crecida.

Diapositiva 11

Sistema de gestión de bases de datos meteorológicos, climatológicos e hidrológicos (MCH)

- La OMM lo ofrece gratuitamente
- Diseñado para:
 - Los SMHN que no dispongan de ningún SGBD u otras soluciones para digitalizar, almacenar y procesar grandes cantidades de datos.
 - Los SMHN que necesiten una solución sencilla, personalizable y gratuita para almacenar y analizar datos, y generar informes sobre grandes volúmenes de datos meteorológicos, climatológicos e hidrológicos.



*MCH example:
Precipitation anomalies
compared to normal*



WORLD
METEOROLOGICAL
ORGANIZATION

Curso de aprendizaje a distancia "Intercambio de datos interoperables en hidrología"
Lección 2.2: Diferentes soluciones o arquitecturas para el almacenamiento de datos



Un ejemplo de SGBD adaptado a los datos hidrológicos es el sistema de gestión de bases de datos Meteorología, Climatología e Hidrología (MCH), que la OMM ofrece gratuitamente a sus Miembros.

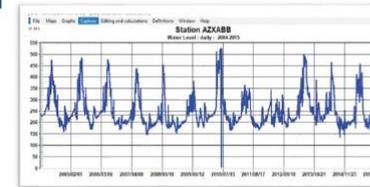
El sistema MCH se ha diseñado para:

- los SMHN que no dispongan de ningún SGBD ni de otras soluciones para digitalizar, almacenar y procesar grandes cantidades de datos;
- los SMHN que necesiten una solución sencilla, personalizable y gratuita para almacenar y analizar datos y para generar informes sobre grandes volúmenes de datos meteorológicos, climatológicos e hidrológicos.

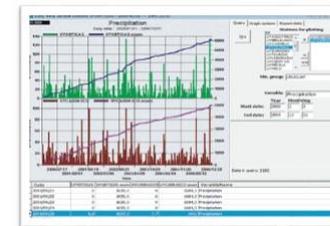
Diapositiva 12

Ventajas del sistema de gestión de bases de datos MCH

- Software de código abierto (no hay que pagar licencias)
- Requisitos mínimos de sistema
- Gestiona datos meteorológicos, climatológicos e hidrológicos
- Sistema configurable; permite conectar módulos externos
- Puede utilizarse en un solo computador o en una configuración de red
- Permite la conexión con estaciones automáticas para recopilar datos en tiempo real
- Es multilingüe; disponible en español, inglés y francés



Example of a graphic of daily water level observations



A comparative graphic for precipitation at two stations



WORLD
METEOROLOGICAL
ORGANIZATION

Curso de aprendizaje a distancia "Intercambio de datos interoperables en hidrología"
Lección 2.2: Diferentes soluciones o arquitecturas para el almacenamiento de datos



A continuación se señalan algunas ventajas del sistema MCH:

- Es un programa informático de código abierto, es decir, no es necesario comprar licencias ni es una "caja negra".
- Los requisitos del sistema son mínimos.
- Ofrece todas las funciones básicas que necesita un sistema de gestión de bases de datos para gestionar datos meteorológicos, climatológicos e hidrológicos.
- Es un sistema configurable (es decir, el usuario establece las variables, estaciones, etc.) y permite conectar módulos externos.
- Puede utilizarse en una sola computadora o en una configuración de red.
- Permite la conexión con estaciones automáticas para recopilar datos en tiempo real.
- Es multilingüe y actualmente está disponible en español, inglés y francés.

Diapositiva 13

Los SGBD desempeñan una función beneficiosa para el intercambio de datos

- Facilitan el cumplimiento de los formatos de datos normalizados internacionalmente;
- los datos pueden exportarse en distintos formatos;
- el proceso de intercambio de datos puede parametrizarse y automatizarse;
- se pueden realizar consultas fácilmente a los datos.



WORLD
METEOROLOGICAL
ORGANIZATION

Curso de aprendizaje a distancia "Intercambio de datos interoperables en hidrología"
Lección 2.2: Diferentes soluciones o arquitecturas para el almacenamiento de datos



Tanto si se utiliza el sistema MCH como cualquier otro, los SGBD son útiles a la hora de poner datos en común. A continuación se indican algunas de sus ventajas:

- Todos los datos se almacenan en una base de datos centralizada de forma estructurada, lo que facilita el cumplimiento de los formatos de datos normalizados internacionalmente.
- Los datos pueden exportarse en distintos formatos.
- El proceso de intercambio de datos puede parametrizarse y automatizarse (de máquina a máquina sin intervención humana).
- Resulta sencillo formular consultas sobre los datos.