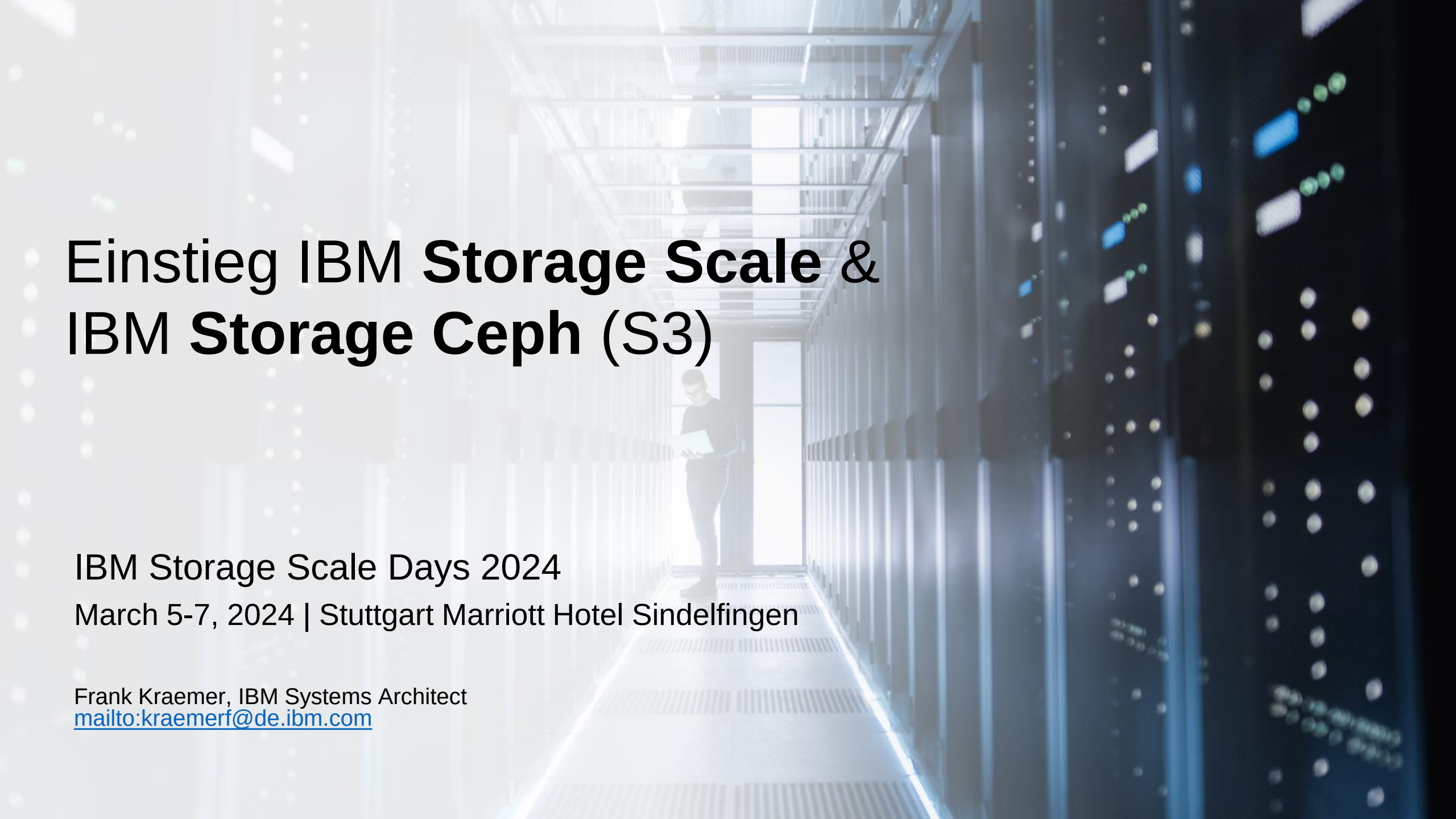


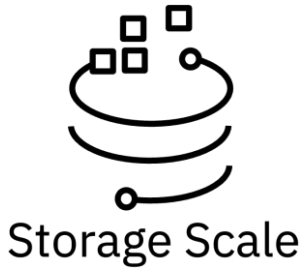
Einstieg IBM **Storage Scale** & IBM **Storage Ceph** (S3)

IBM Storage Scale Days 2024

March 5-7, 2024 | Stuttgart Marriott Hotel Sindelfingen

Frank Kraemer, IBM Systems Architect
mailto:kraemerf@de.ibm.com

Disclaimer



IBM's statements regarding its plans, directions, and intent are subject to change or withdrawal without notice at IBM's sole discretion. Information regarding potential future products is intended to outline our general product direction and it should not be relied on in making a purchasing decision. The information mentioned regarding potential future products is not a commitment, promise, or legal obligation to deliver any material, code, or functionality. The development, release, and timing of any future features or functionality described for our products remains at our sole discretion.

IBM reserves the right to change product specifications and offerings at any time without notice. This publication could include technical inaccuracies or typographical errors. References herein to IBM products and services do not imply that IBM intends to make them available in all countries.

Different Storage Types

File  **Object** **Block**

Serves Applications (Apps) | Serves OS/VM/Pods

- File storage organizes data within a hierarchy in folders.
- Object storage manages data and links to associated metadata.
- Block stages utilize arbitrarily organized storage chunks and evenly sized volumes.

File Storage

File storage uses a hierarchical structure where files are organized by the user in **folders** and **subfolders**, which makes it easier to find and manage files.

To access a file, the user selects or enters the path for the file, which includes the sub-directories and file name.

Real Object Storage

Source: Walter Mier, Wikipedia
Creative Commons CC-BY_SA-3.0

Object Storage

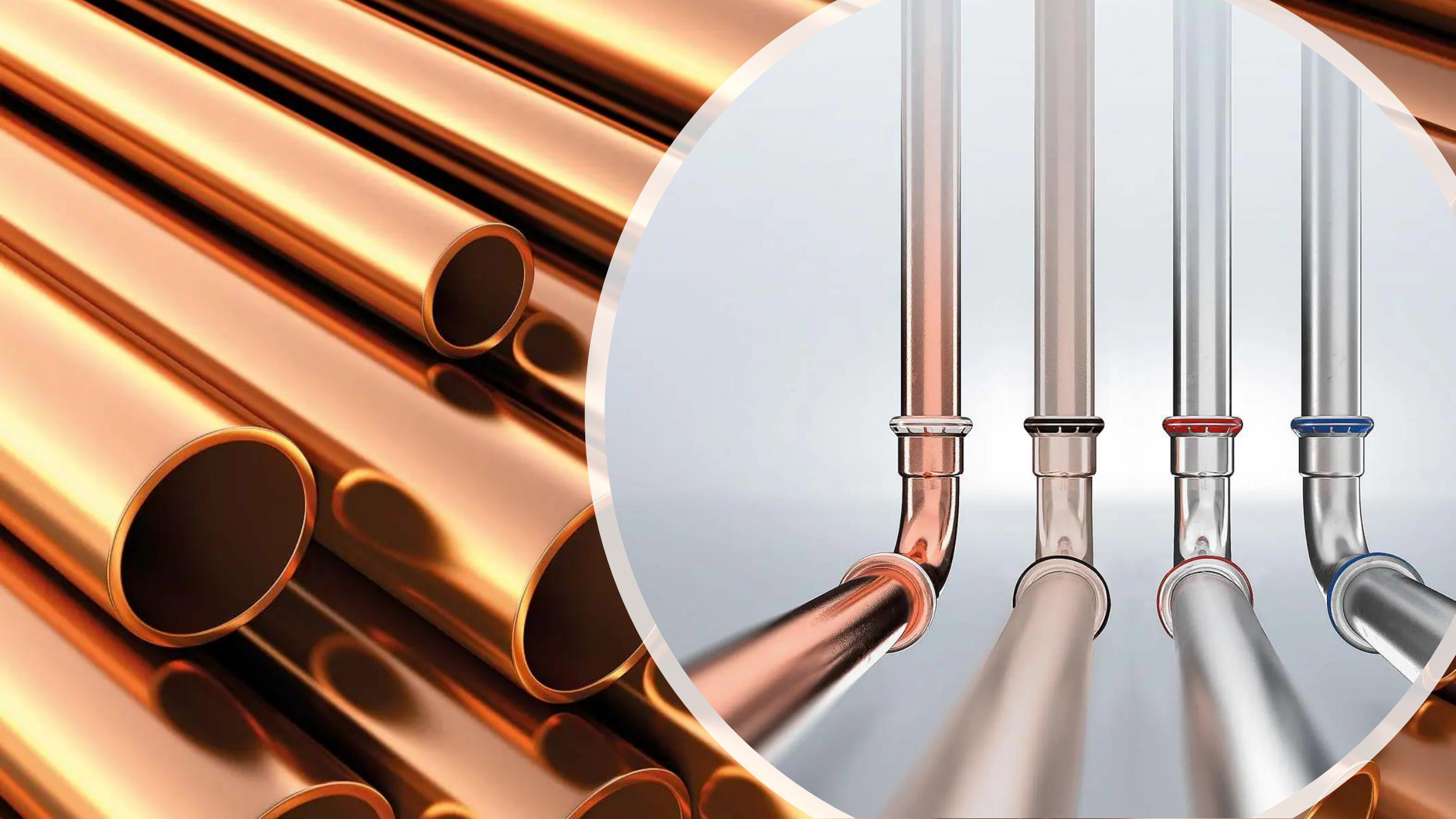
File: **D:\Barbie.mp4** vs **https://AWS-BUCKET.s3.us-west-2.aws.com/barbie.mp4**

Video Editing is a typical FILE based workflow

What is **POSIX**? Why Does it Matter for **File** Storage

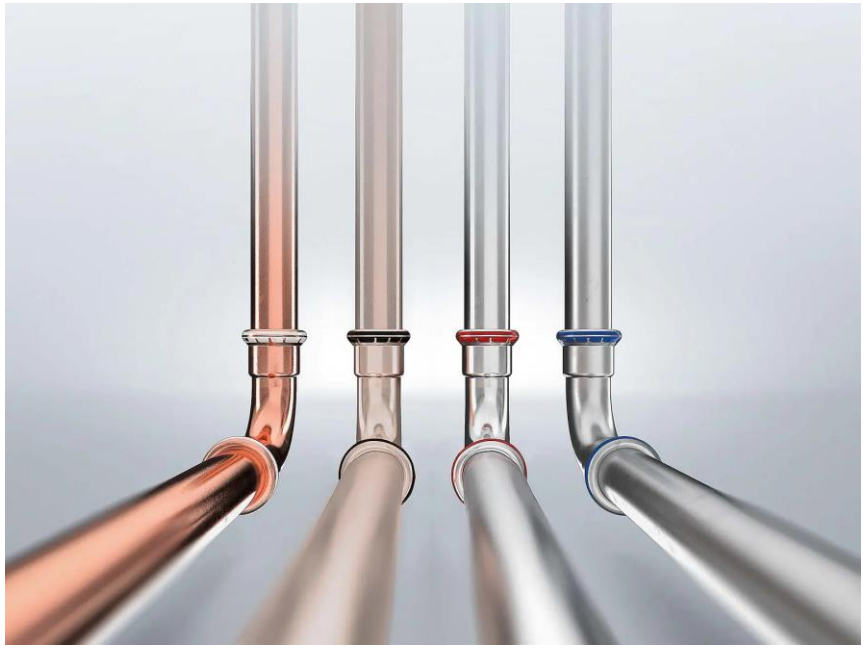
A *file* is a named, ordered stream of bytes.

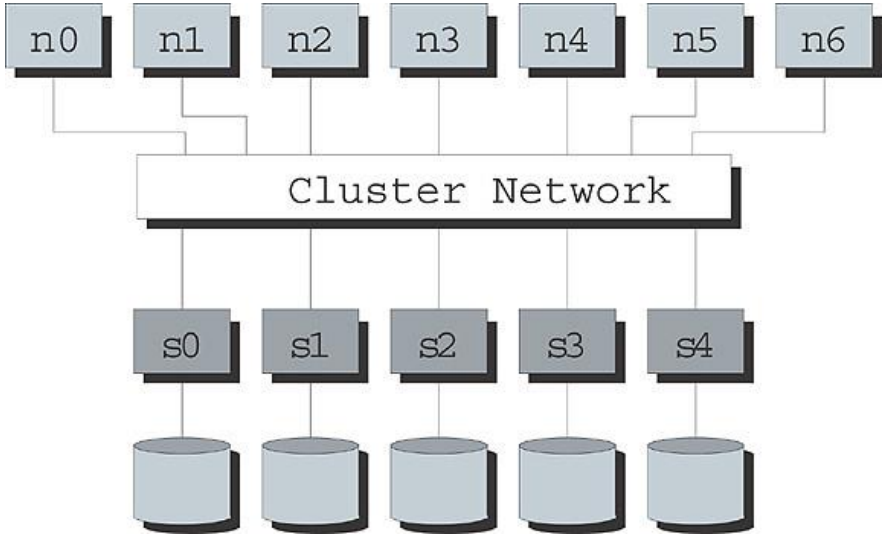
- `open(..)` Open a file for reading or writing. Also allows a file to be locked providing exclusive access.
- `close(..)`
- `read(..)` The read operation is normally *blocking*.
- `write(..)`
- `lseek(..)` Seek to an arbitrary location in a file.
- `ioctl(..)` Send an arbitrary control request (specific to a device). e.g. rewinding a tape drive, resizing a window etc.



A parallel file system is a software component designed to store data across multiple networked servers and to facilitate high-performance access through simultaneous, coordinated input/output operations (IOPS) between clients and storage nodes.

[Source: https://www.linkedin.com/in/briti-gangopadhyay-a988008a/]





Portable Operating System Interface (POSIX) Filesystem
POSIX referred to IEEE Std 1003.1-1988, released in **1988**

IBM Storage Scale

Parallel file systems reach their superior I/O **performance** and decent **scalability** with the help of **striping**.

Instead of storing a file only on one storage server, a parallel filesystem splits up the data of a file into chunks and distributes these chunks across multiple storage servers.

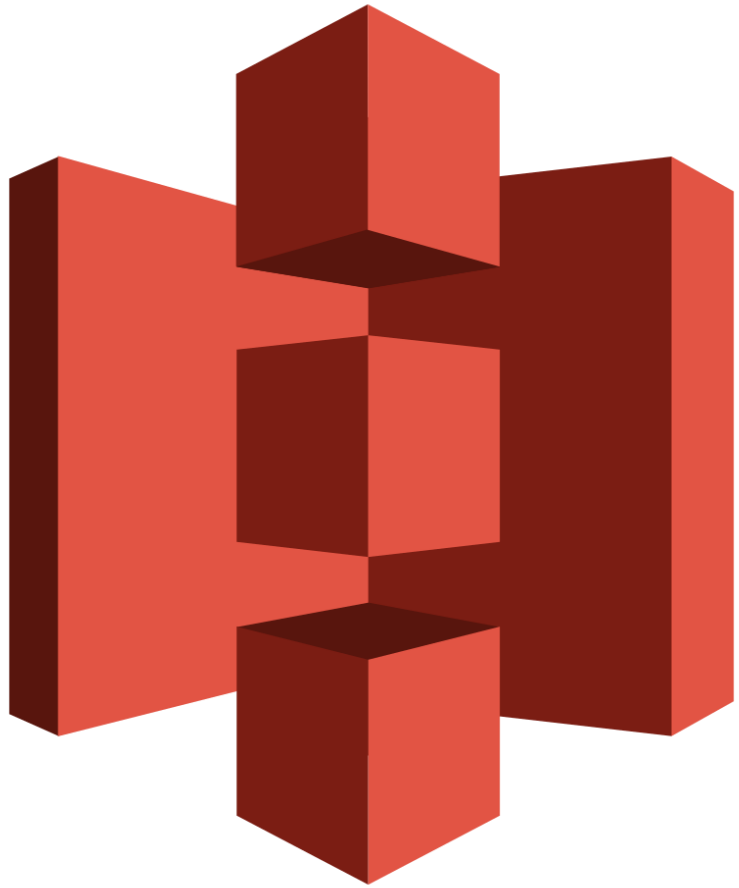
Real World - Logistic Transport Systems

Storage Scale

Tube / Pipe ⟷ EUR-Pallet / Skid

Object Storage

All Robots must have the same color with the right flow rate and pressure to do the best paint job in parallel.



Euro-pallet or EPAL-pallet (European Pallet Association)
Launched in **1961** by Union Internationale des chemins de fer (UIC)

Amazon **S**imple **S**torage **S**ervice (S3)
Launched March 14, **2006**

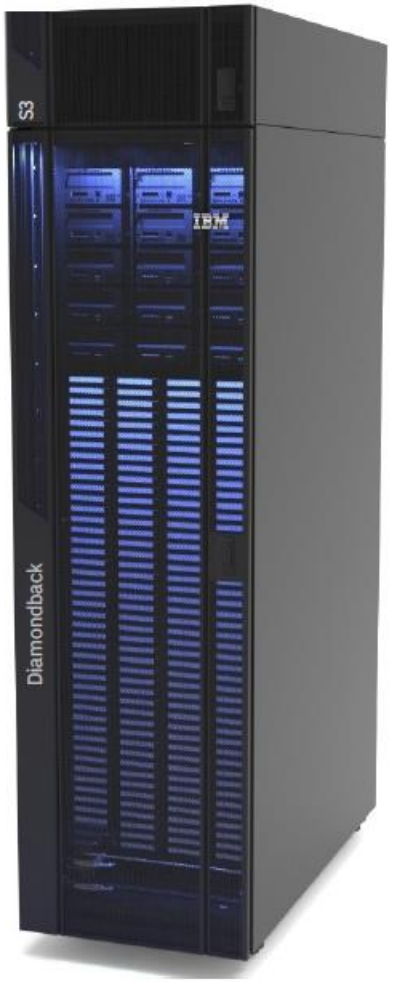
Gitterboxpalette (Gibo)
launched in **1968** by Deutsche Bahn (DB)

2024 “Gitterboxpalette” for Data by IBM

IBM Diamondback S3

For organizations looking for cost-effective long-term data storage in an integrated solution, IBM Diamondback S3 delivers the benefits of tape storage in an easy-to-use, easy-to-integrate solution in the footprint of a single 19" rack with up to 27 PB of capacity.

This solution can be used directly by your existing S3 compatible software to reduce the cost of your cold data by up to 75%, with no data egress fees.




Operational Assurance
... someone will not

A little background…

Traditionally, there have been (2) **primary and separate** data stores used for data analytics within organizations……..

Data Lake

Unstructured Data

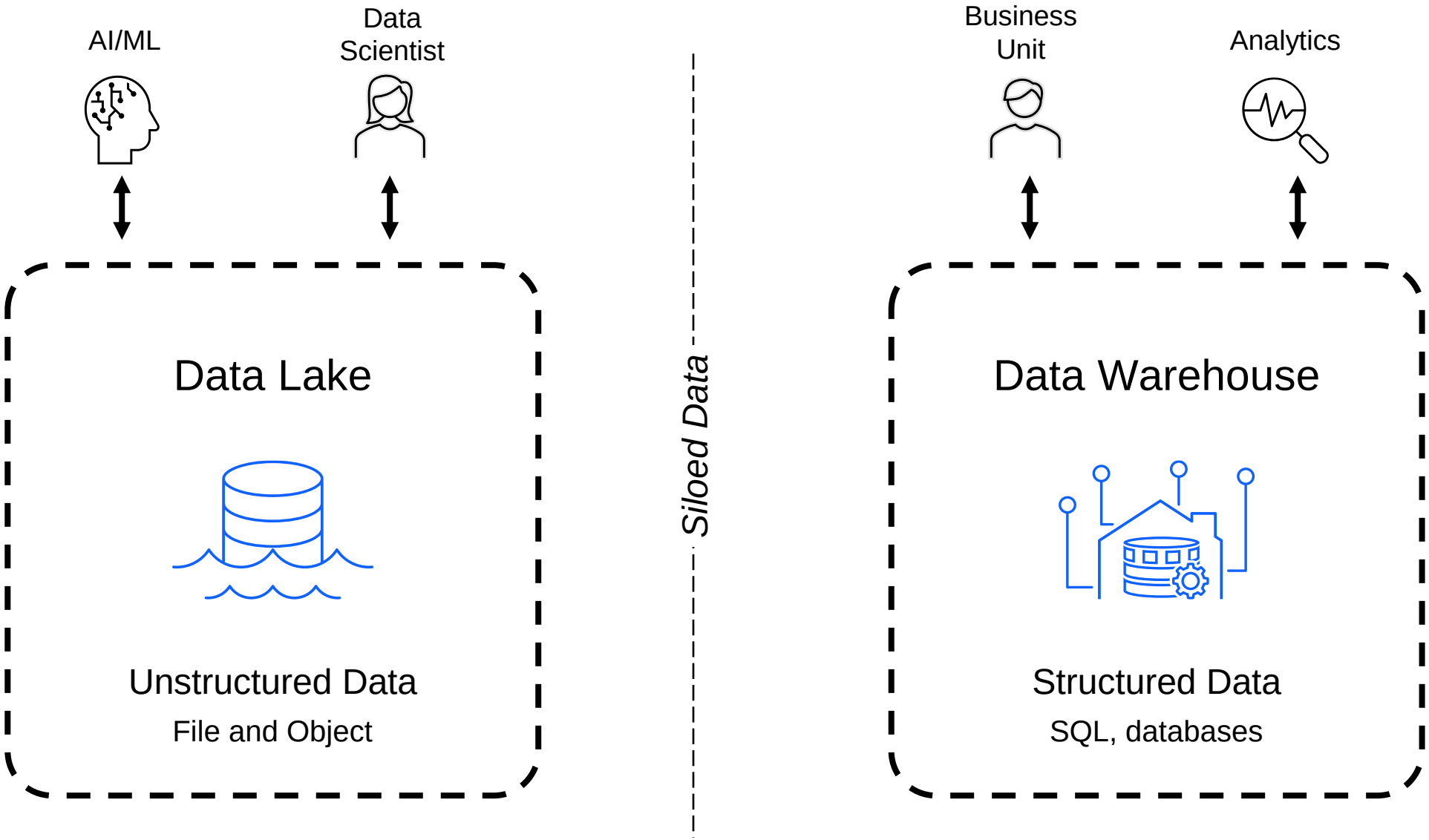
File and Object

Data Warehouse

Structured Data

SQL, databases

The problem with these two separate approaches…



watsonx.data brings it all together with the best of both worlds…

Fusion HCI
and **watsonx**

AI/ML

Data Scientist

Business Unit

Analytics

…in a data Lakehouse

APP APP APP • • • APP

Apache Iceberg

Metadata Management

Data Services

Object Storage

Unstructured | Semi-Structured | Structured

Hybrid Cloud: Data Lakehouse

Data Lake

Unstructured Data
File and Object

Data Warehouse

Structured Data
SQL, Databases

Takeaways for TODAY

GENERATIVE AI IMPACT ON STORAGE IS HUGE

By 2025, Gartner expects generative AI **to account for 10% of all data** produced worldwide, up from less than 1% today

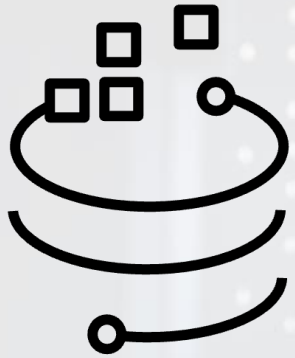
IBM Ceph IS YOUR SCALABLE OPEN SOURCE BASED OBJECT STORAGE FOR DATA & AI READY NODES MAKE IT EASY TO DEPLOY

IBM FUSION HCI IS THE EASIEST AND FASTEST WATSONX ON-PREMISES PLATFORM

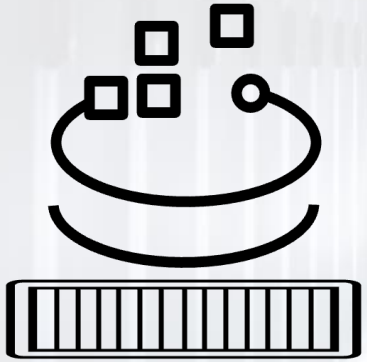
IBM STORAGE SCALE IS THE HIGH-PERFORMANCE AI STORAGE TO MAXIMZE YOUR GPU INVESTMENT

IBM OBJECT TAPE IS THE MOST SUSTAINABLE SOLUTION FOR LONG TERM AI DATA REPOSITORY

Thank you for using



Storage Scale



Storage Scale System