hochschule hof
University of Applied Sciences

UNIVERSIDAD DE MÁLAGA

Overview of advanced storage technologies and storage virtualization

Pablo Pérez Trabado

Dept. of Computer Architecture

University of Malaga (Spain)

Disclaimers

- Some of the figures in the slides are taken from SNIA Education tutorials. To comply with the SNIA conditions of use for these documents, any SNIA tutorial used in the elaboration of these slides has been also provided as a handout
- Many (but not all) figures in the slides have been taken from Internet-available resources, including Wikipedia. Whenever possible by copyright restrictions, the original source has been also provided as a handout along with the slides. Also, whenever possible the original source is quoted. In any case, no ownership claims are made over images in the slides not drawn by the author himself.
-
- VMware does not endorse or make any representations about third party information included in this document, nor does the inclusion of any VMware icon or diagram in this document imply such an endorsement.

hochschule hof
University of Applied Sciences

UNIVERSIDAD DE MÁLAGA

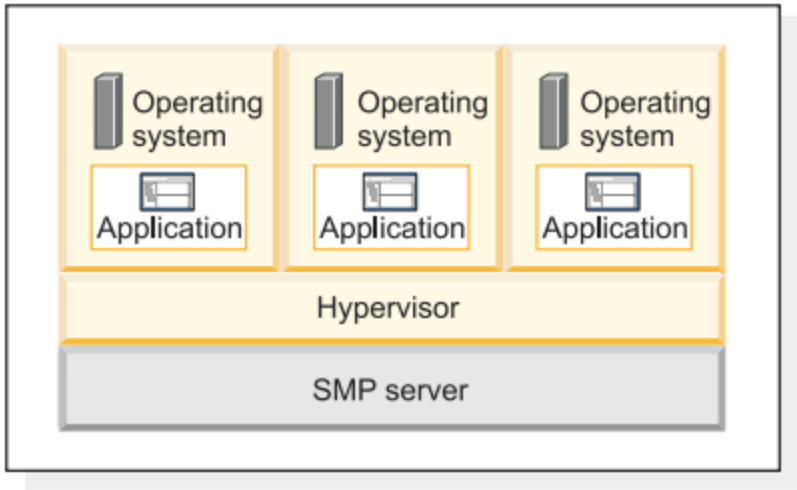
UNIVERSIDAD MÁLAGA

Virtualization and storage

What will we learn?

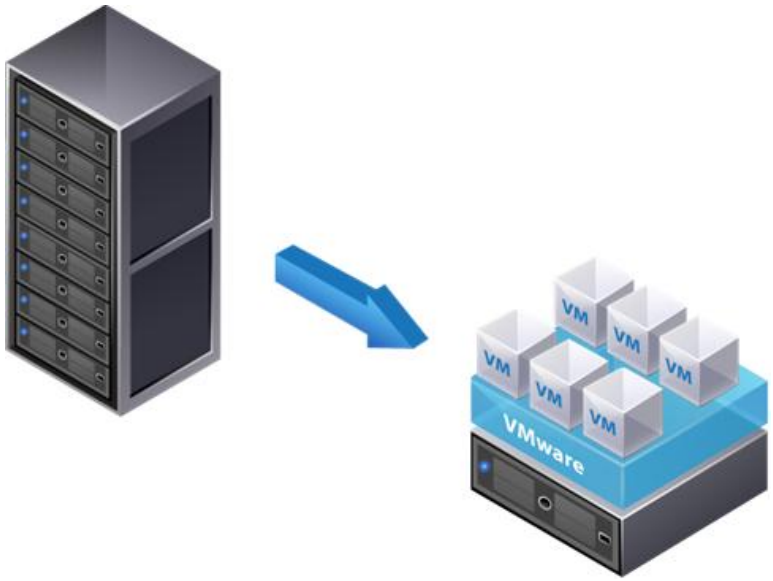
- The concepts of virtualization and virtual storage
- How virtualized storage uses several layers of mappings between the SCSI LBAs seen by the VM and the real physical blocks where data is stored
- How the use of shared storage on the datacenter allows sophisticated techniques for virtual machine migration and high availability
- The use of thin provisioning to reduce storage requirements of virtual machines
- The use of linked clones and deduplication to further decrease the storage needs in datacenters with hundreds of virtual machines
- The pitfalls of storage virtualization; how the use of virtual machines can lead to new types of I/O performance problems

Virtualization



- Virtual Machine (VM) is software implementation of machine, able to execute programs like a physical machine
  - Hypervisor = software that creates VM on the host hardware
  - Hypervisor, thus, is the one with physical access to storage

Virtualization



- Use of VMs allow *Consolidation* for complex applications
  - Substitution of lots of physical machines for just a few multiprocessor servers hosting lots of VMs
  - Nicely suited for non-CPU intensive distributed applications
    - Reduction in purchase and maintenance costs of servers
    - Virtual hardware won't go obsolete (!)

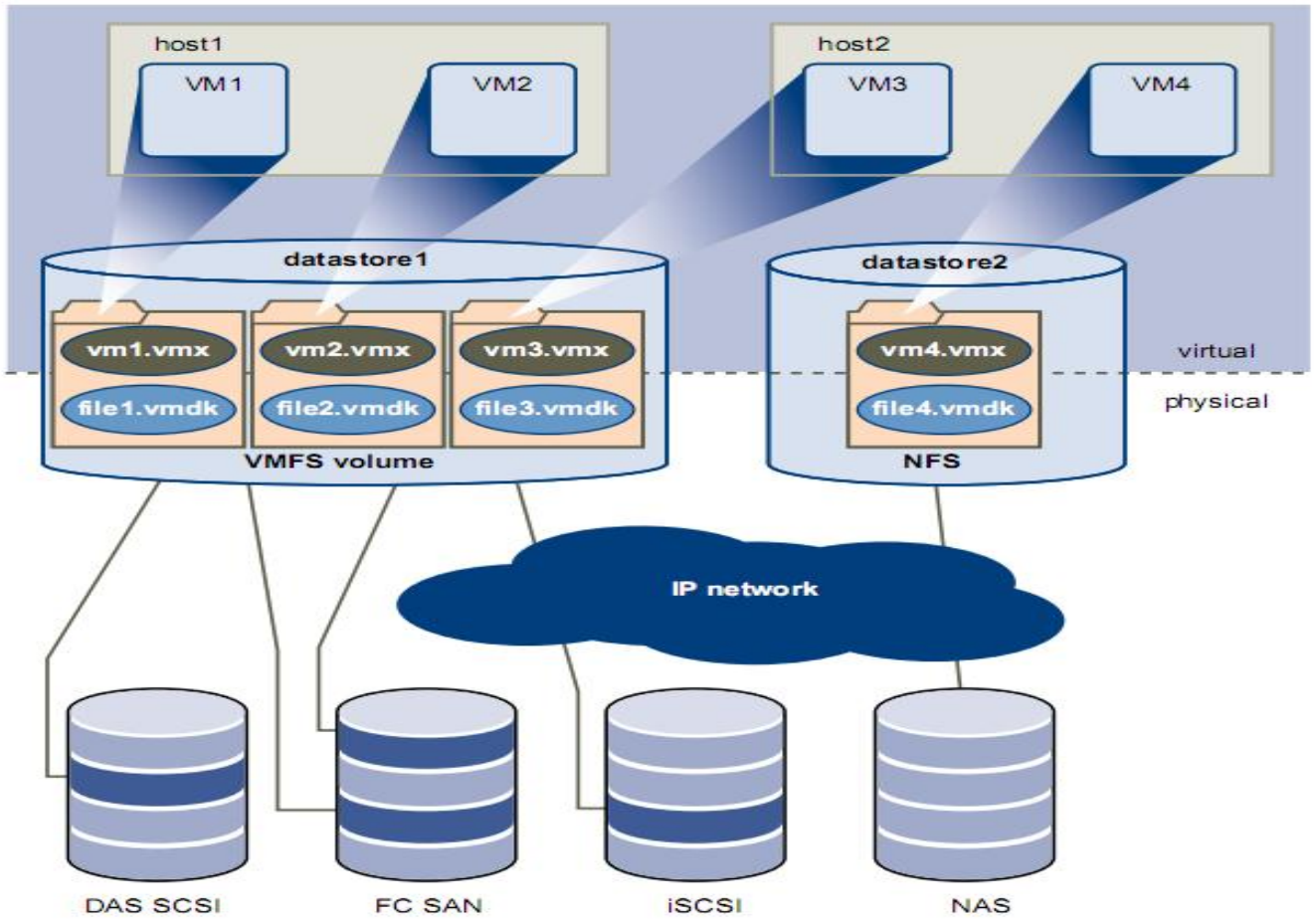
Virtual Storage

- SCSI protocol plays a vital role in storage for VMs

- VM uses a *virtual disk* to store OS, programs and data
  - From hypervisor, virtual disk = large physical file, or set of files, handled like any other file

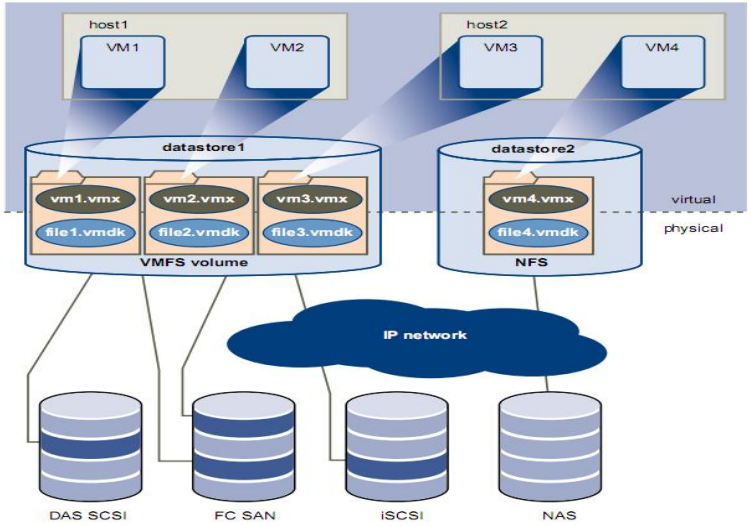
- VM accesses virtual disk through a virtual SCSI controller
  - Virtual disk seen as LUN of virtualized SCSI device, offering array of LBAs
  - VM performs block I/O against virtual disk
  - Underlying physical implementation of storage hidden to VM

Virtual storage



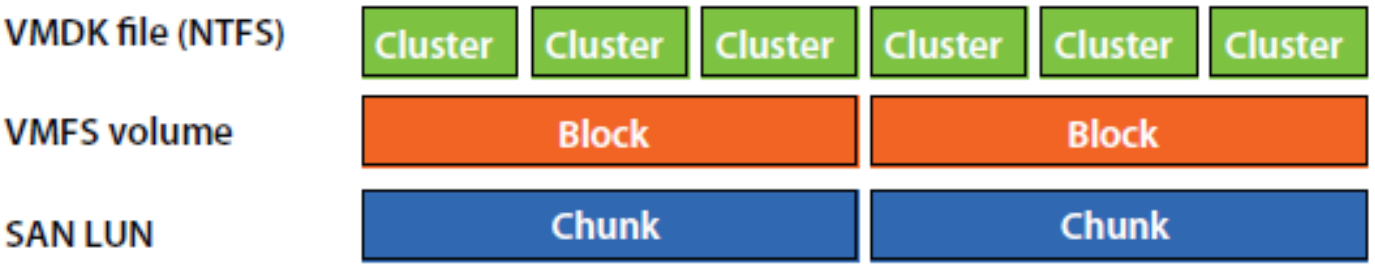
- Virtual Machine files stored in VMFS volumes, called datastores
  - VMFS = Virtual Machine File System

Virtual storage



- Datastore: block-like virtual appliance that represents a pool of physical storage
  - Real storage can be block devices spread across one or multiple hosts
  - Real storage can be also NAS device
- Datastore can be simultaneously accessed from several VMs (or hosts)
  - VMFS is clustering filesystem

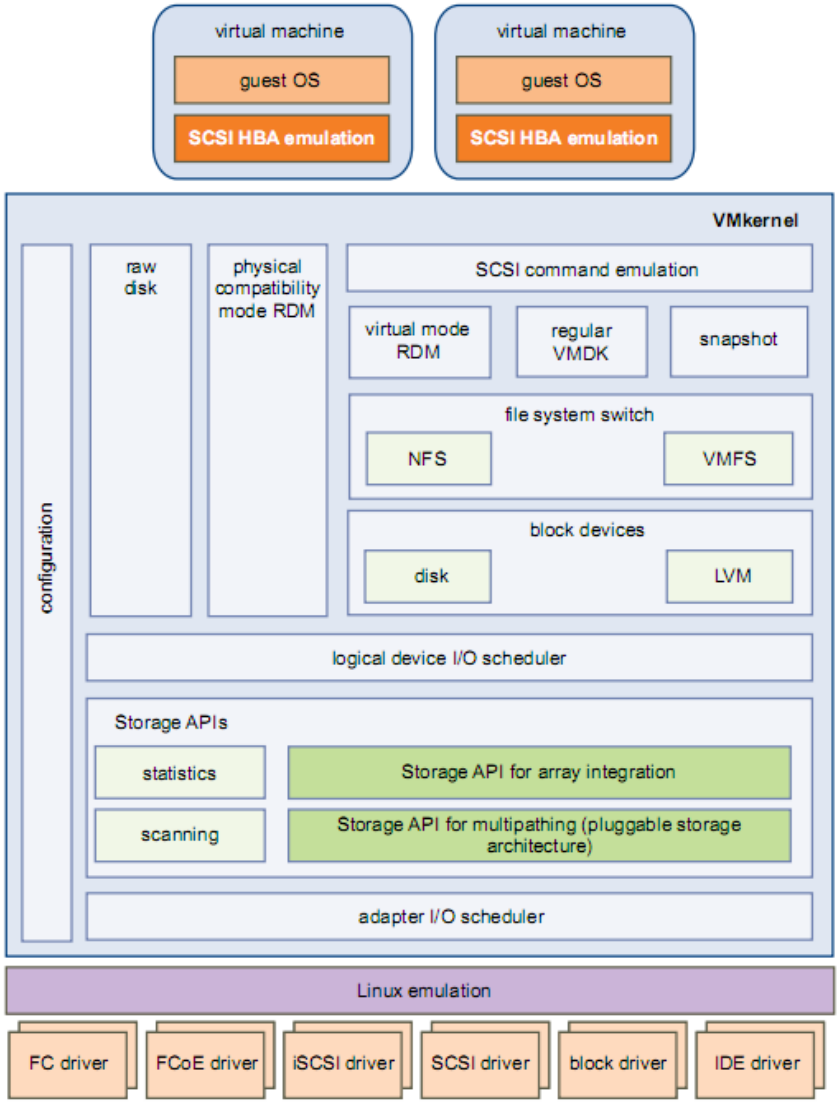
Virtual storage



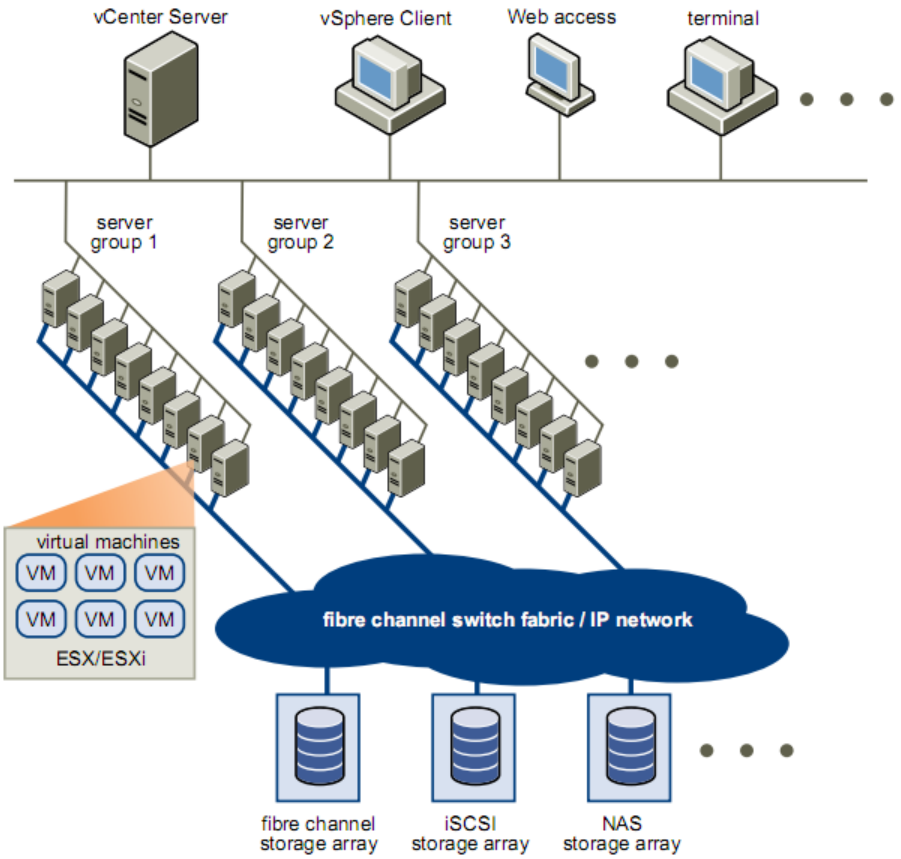
- Datastore allows additional layer of I/O block mappings between the SCSI LUN of the VM and the SCSI LUN of the real storage appliance

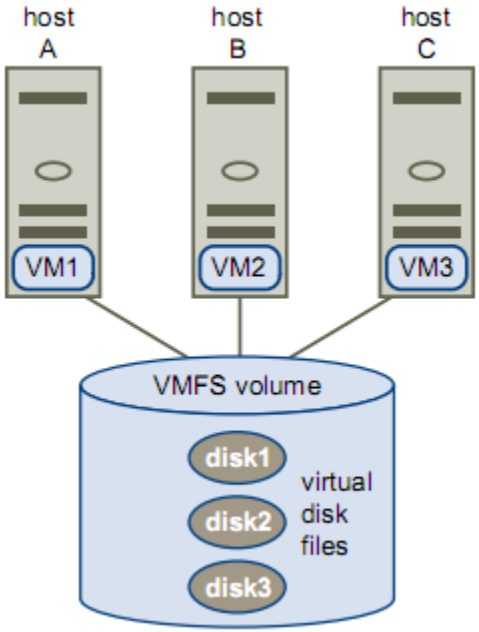
Virtual storage: layered kernel

- Hypervisor implementation maps the emulated SCSI I/O transaction to a real SCSI I/O transaction, down on the host hardware



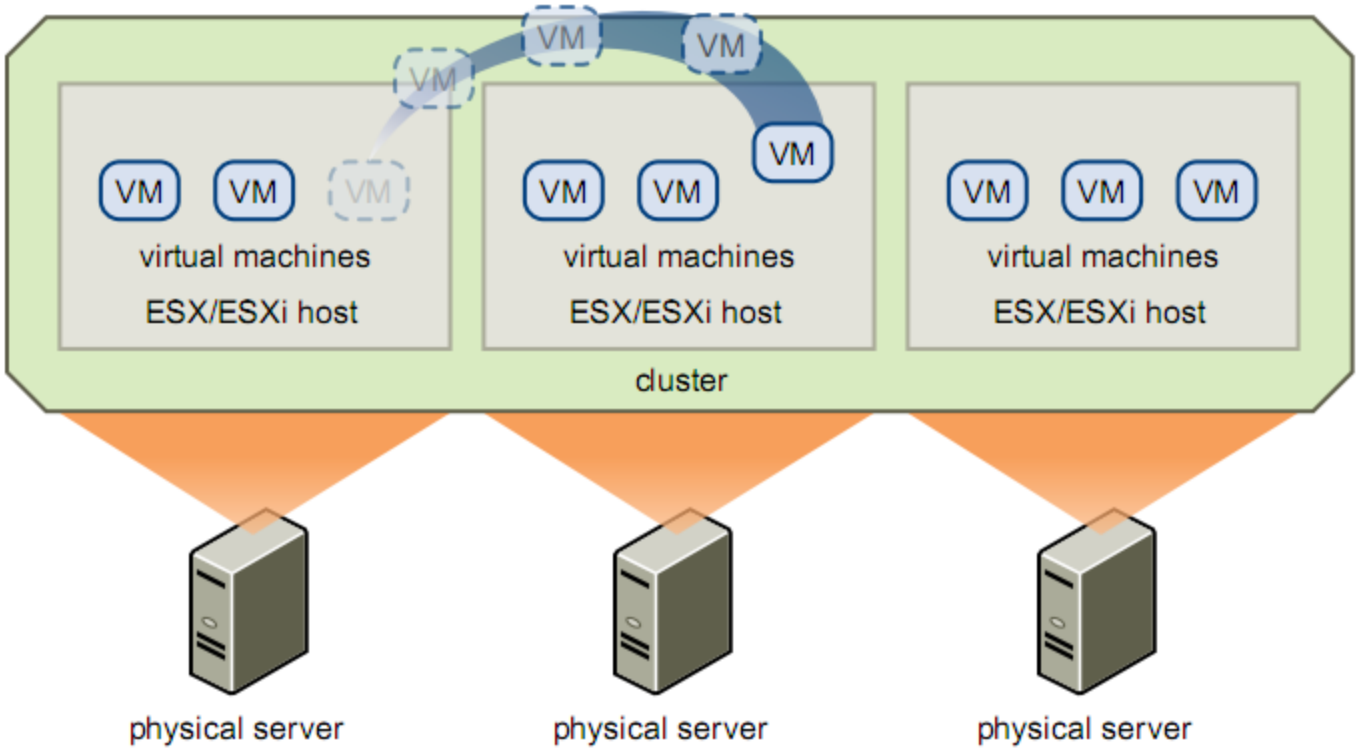
Virtualization Datacenter





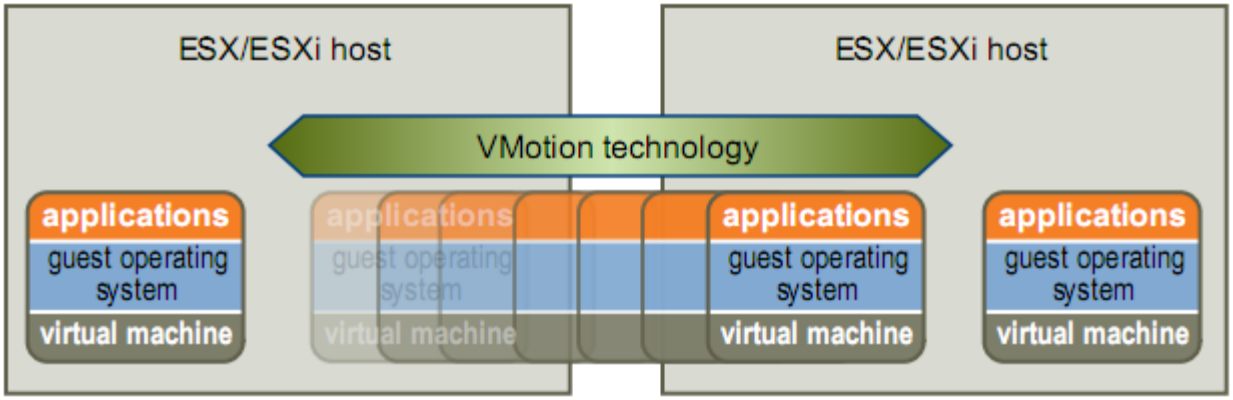
- Datacenter combines:
  - Possibility of running multiple VMs in a single host server
  - If shared storage, ease to access the virtual disk files from any server
    - Several clever "tricks" are made possible by this configuration

VM Migration



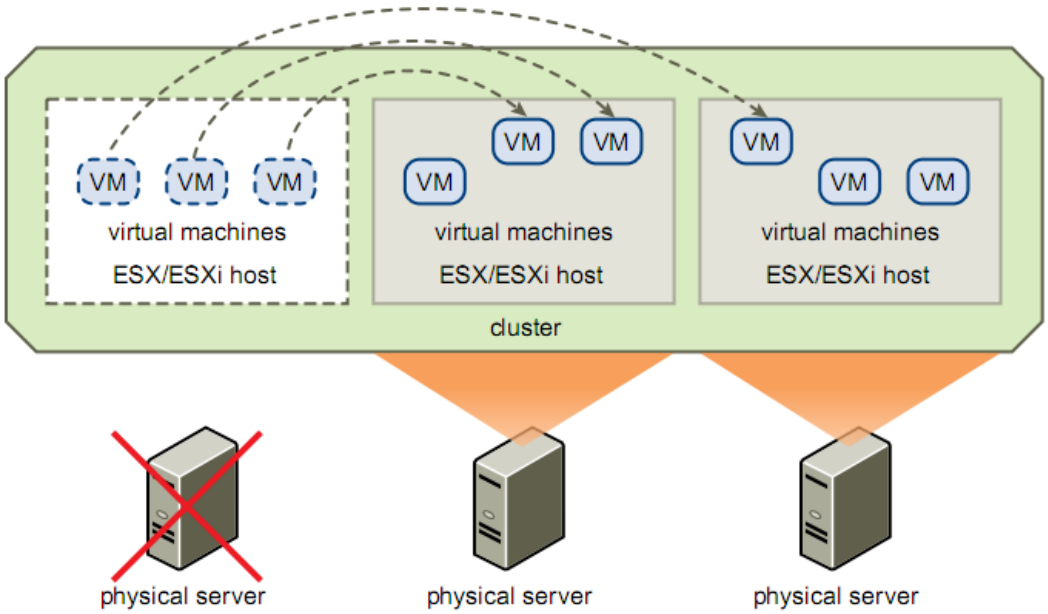
- Virtual machines can be migrated from one server to another to balance host load
  - If both servers share same storage, no need to move physical files

VM Migration



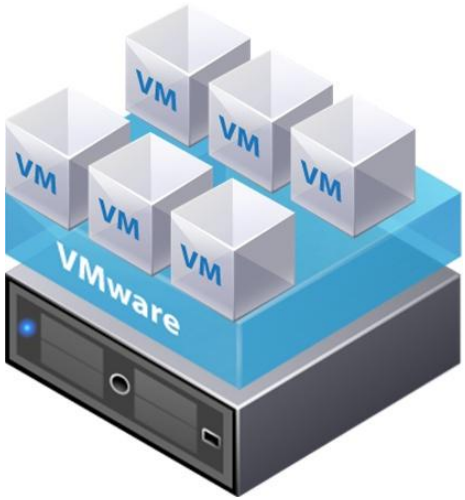
- Virtual machines can also be migrated between servers in different datastores
  - Requires copying files between datastores
  - By keeping track of updated blocks, migration can be performed without shutting down the VM

VM High Availability



- VMs can also be automatically restarted in new host if host server fails
  - No need to copy any files
- If no downtime allowed, two VMs (active/passive) are kept in sync and share virtual disk
  - If active fails, passive turns active
  - Shared virtual storage means instant resume

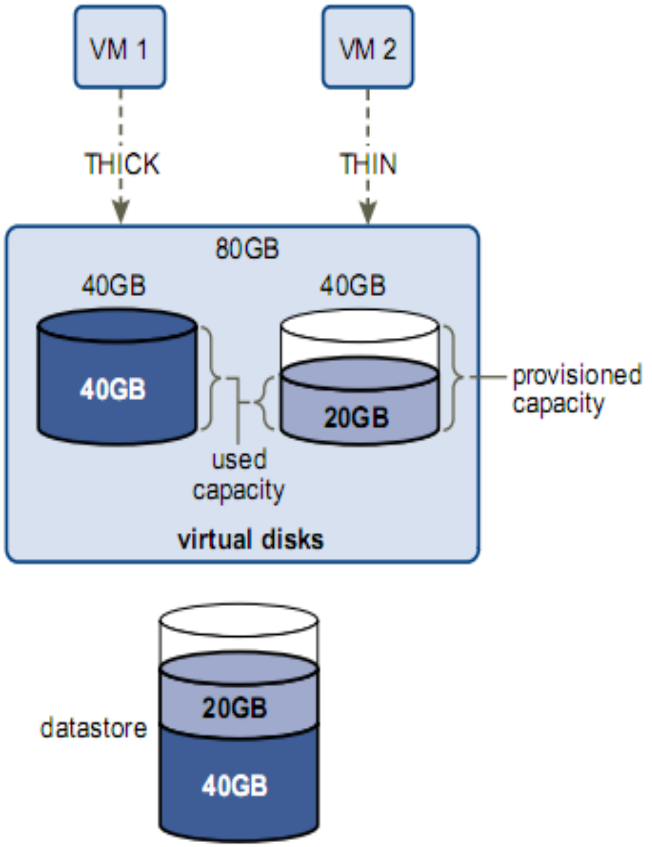
VM Thick provisioning



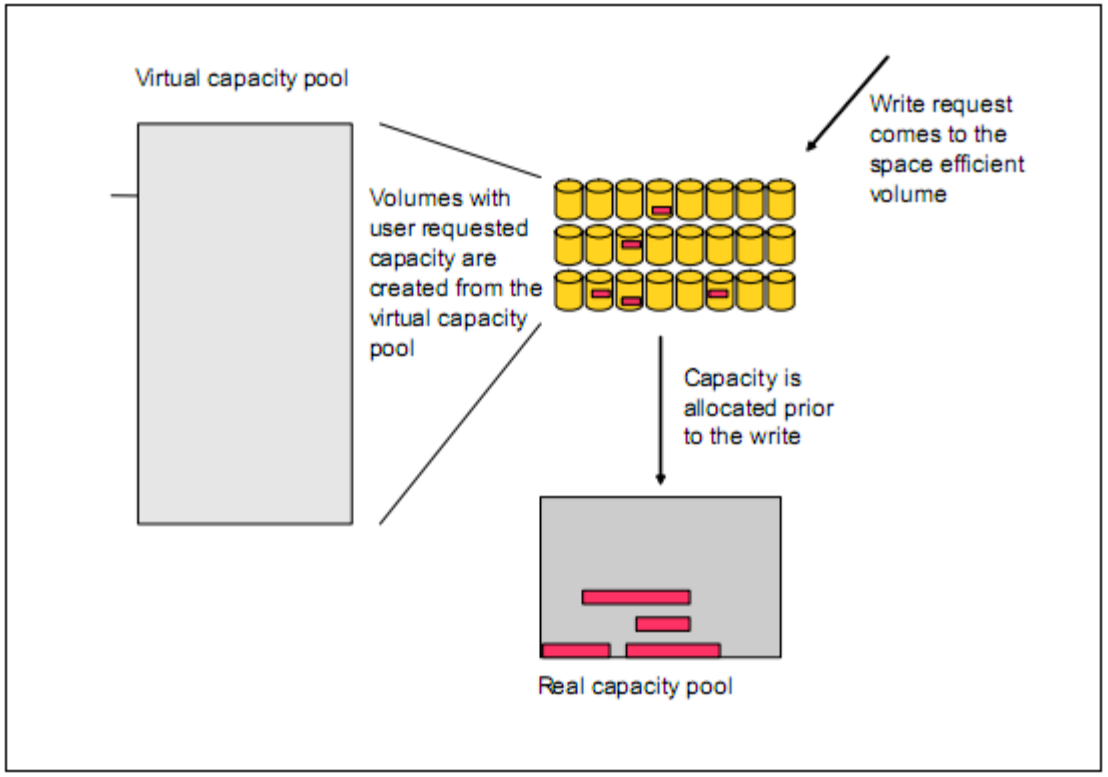
- VM storage provisioning = providing real physical storage for VMs
- *Thick provisioning*: Size of file containing virtual disk matches size of the virtual disk declared for the VM
  - Thick disk immediately occupies the entire provisioned space
- In example above, each VM is thick provisioned a 40 GB disk
  - Total storage space needed by host in datastore: 6 x 40 GB = 240 GB

VM Thin provisioning

- VMs rarely fill their virtual disks with data
  - With thick provisioning, empty disk is wasting storage
- Thin provisioning assigns storage on demand
  - Virtual disk file holds just actual data
    - Size is smaller than declared disk size
  - File grows on demand
    - Hypervisor maps new storage blocks to file

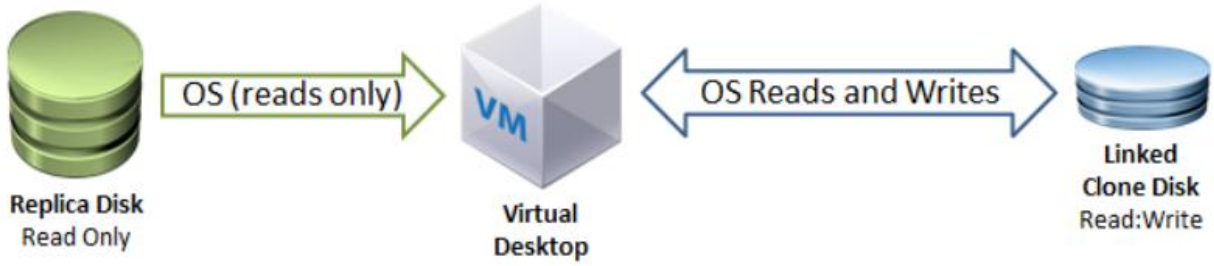


VM Thin provisioning

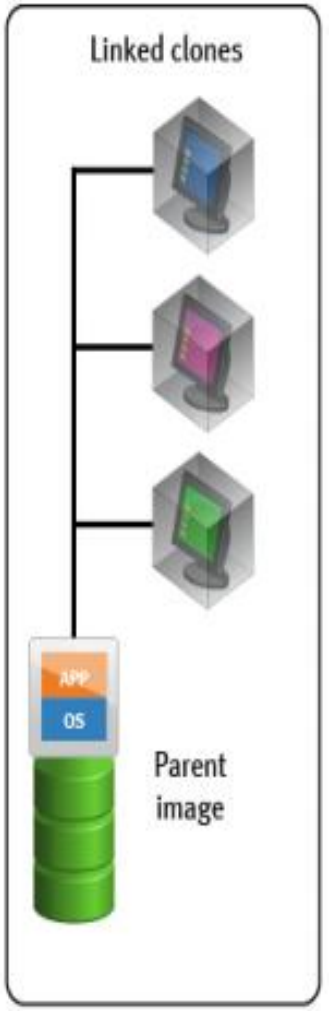


- In example, each VM has a 40 GB disk, but only 10 GB of real data
  - Virtual capacity pool = 6 x 40 GB = 240 GB
  - Real used capacity = 6 x 10 GB = 60 GB (!)
- Over-provisioning: real capacity pool < virtual capacity pool

VM Linked clones and snapshots



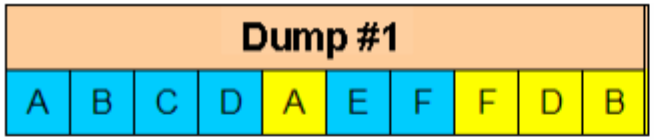
- Storage for linked clone VM composed of two files:
  - Base image, read-only
  - Differential snapshot, read-write
- Multiple linked clones can share the same base image
  - Great for "golden" base system
  - Great for starting VMs on demand (Desktop virtualization)



Deduplication

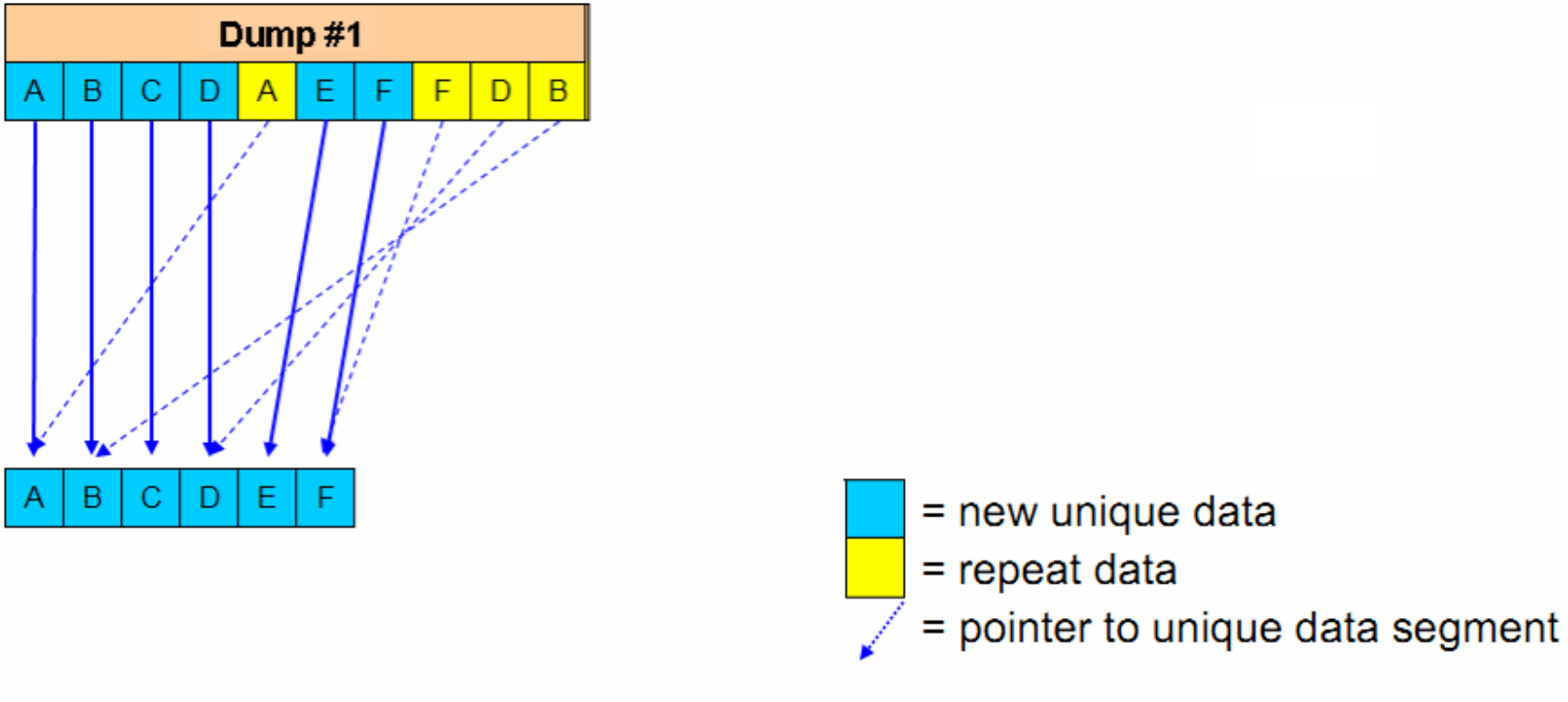
- Increasingly popular method to reduce virtualization and backup storage needs
  - Works at block level
  - Uniquely identifies when two blocks hold exactly same data
    - Hash + metadata comparison
  - Stores a single copy of data
    - Second and further copies are just metadata on database
  - Combined with linked clones can greatly reduce storage for VDI

Deduplication

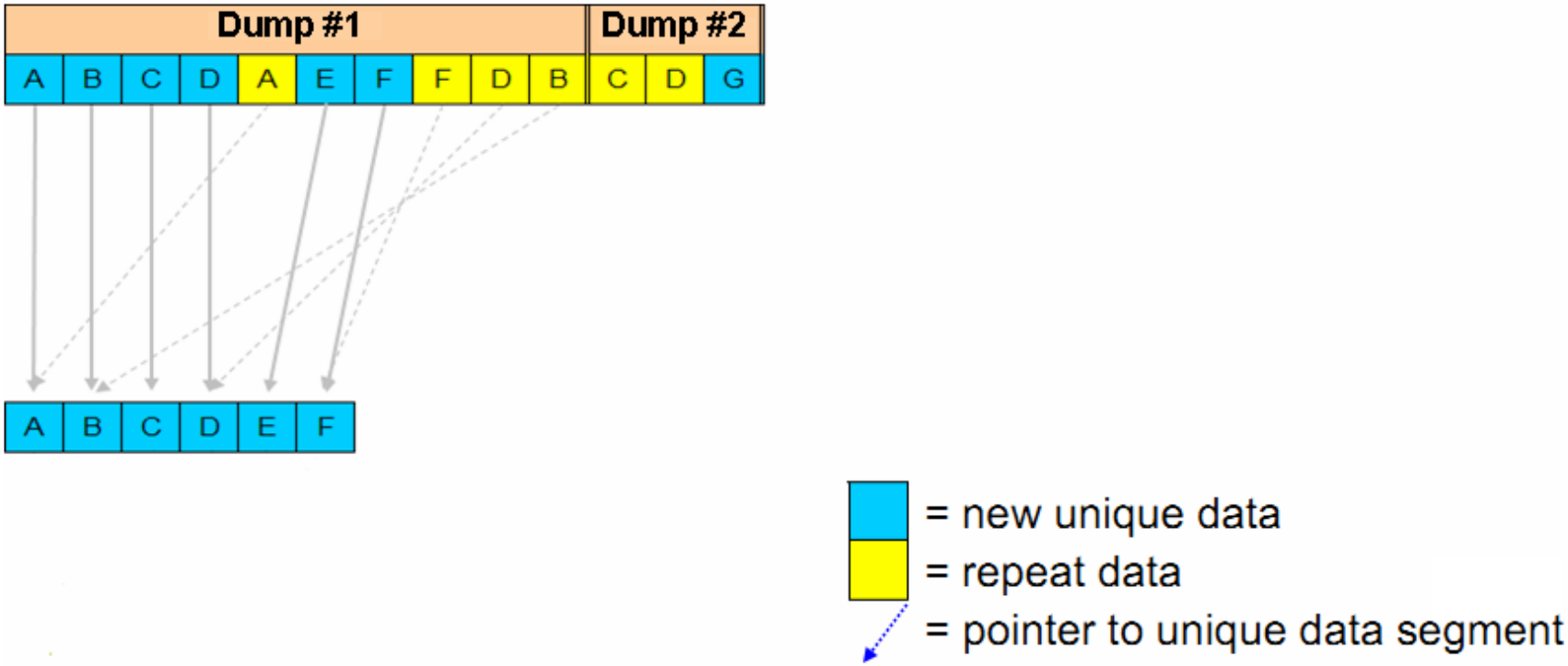


= new unique data
= repeat data

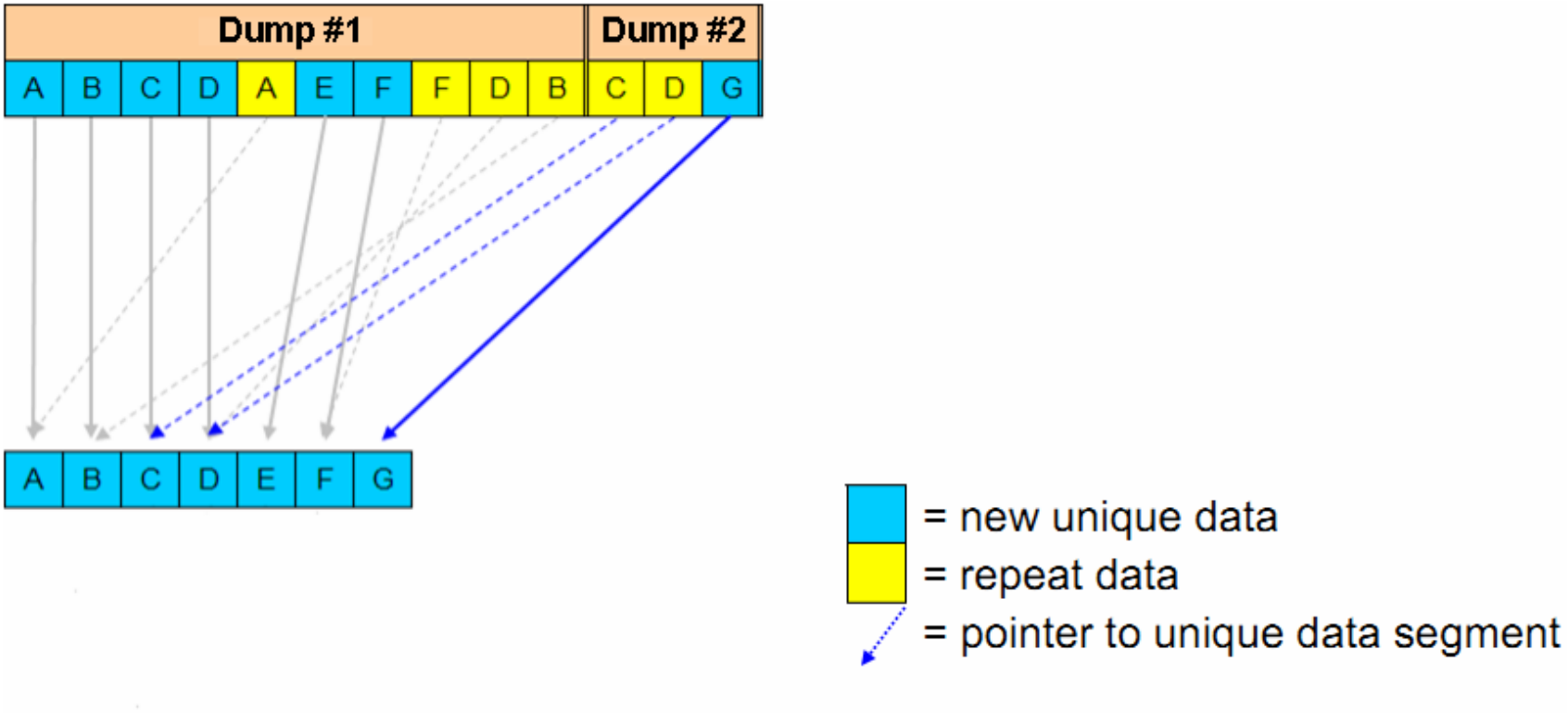
Deduplication



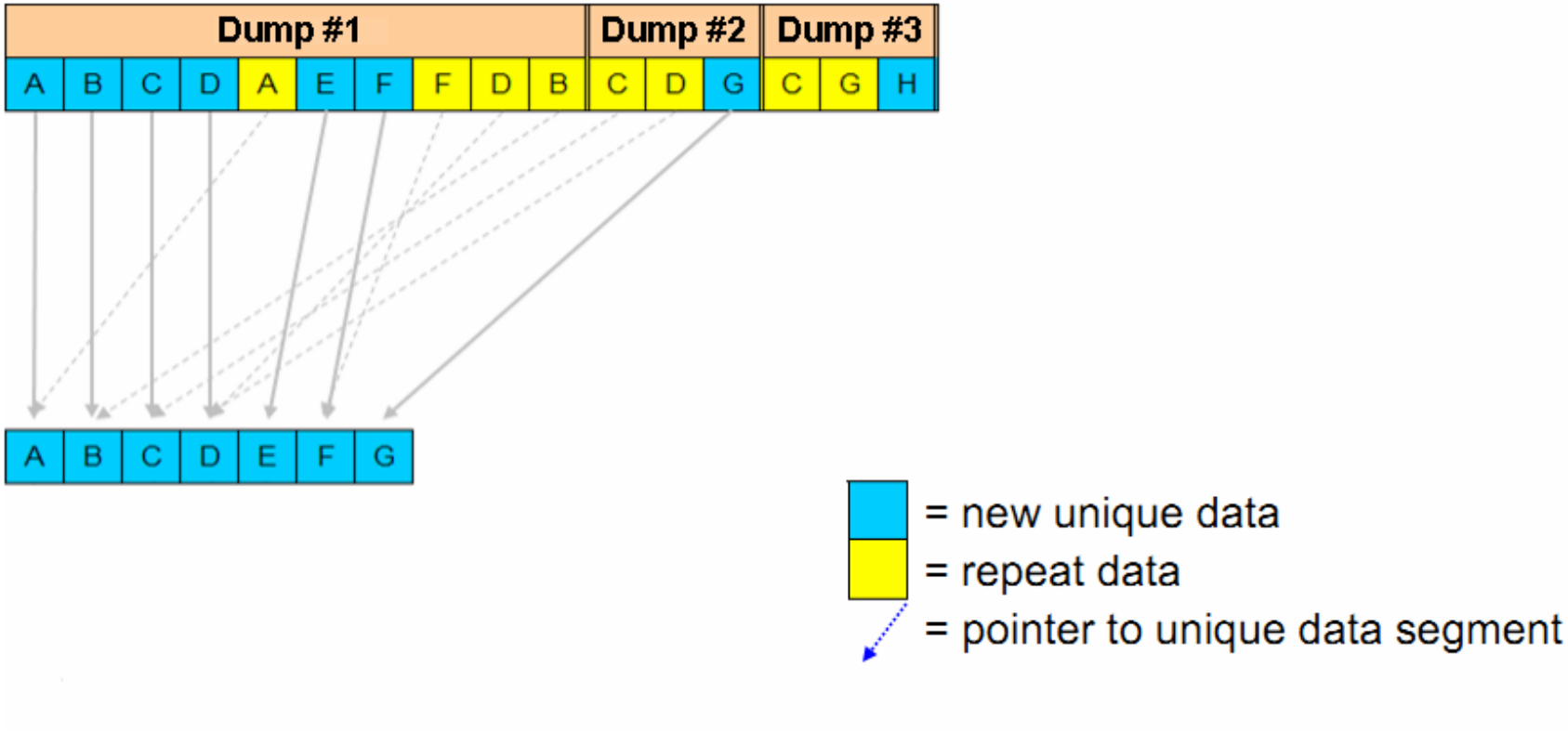
Deduplication



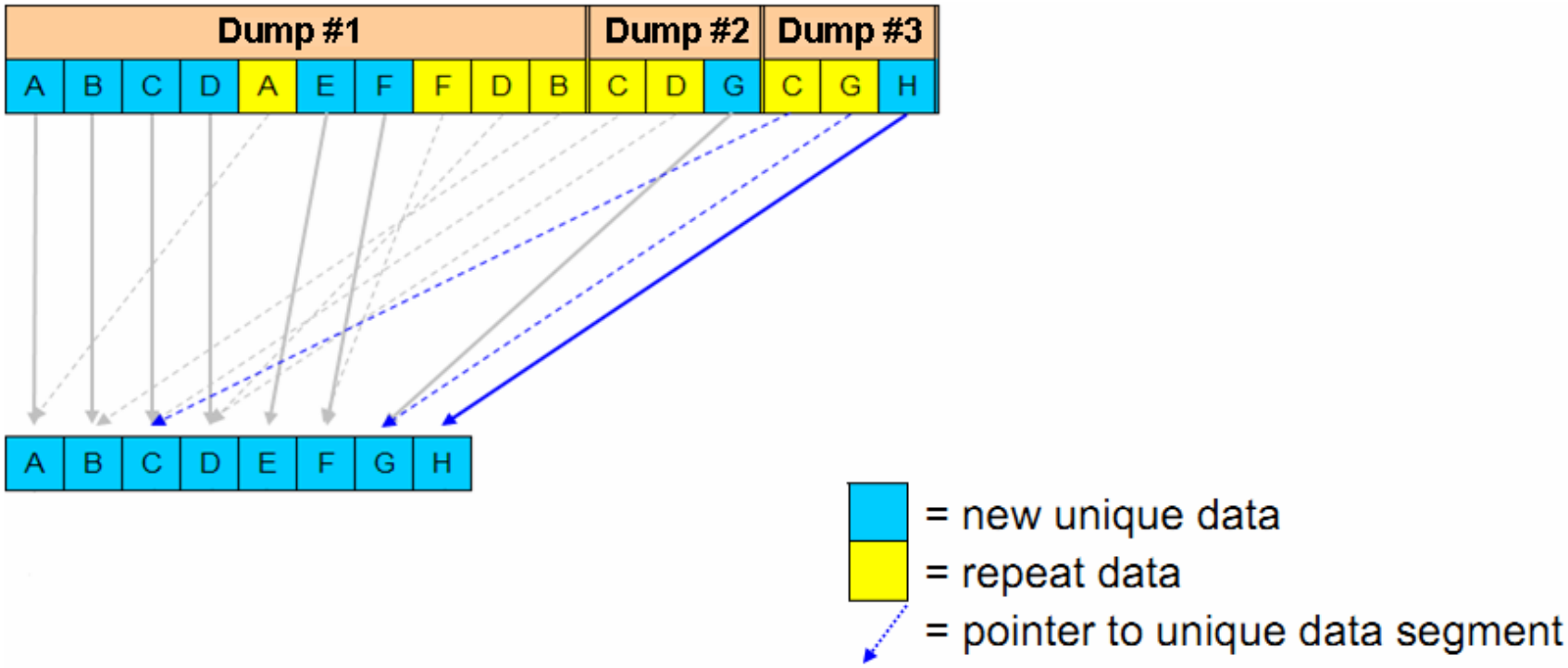
Deduplication



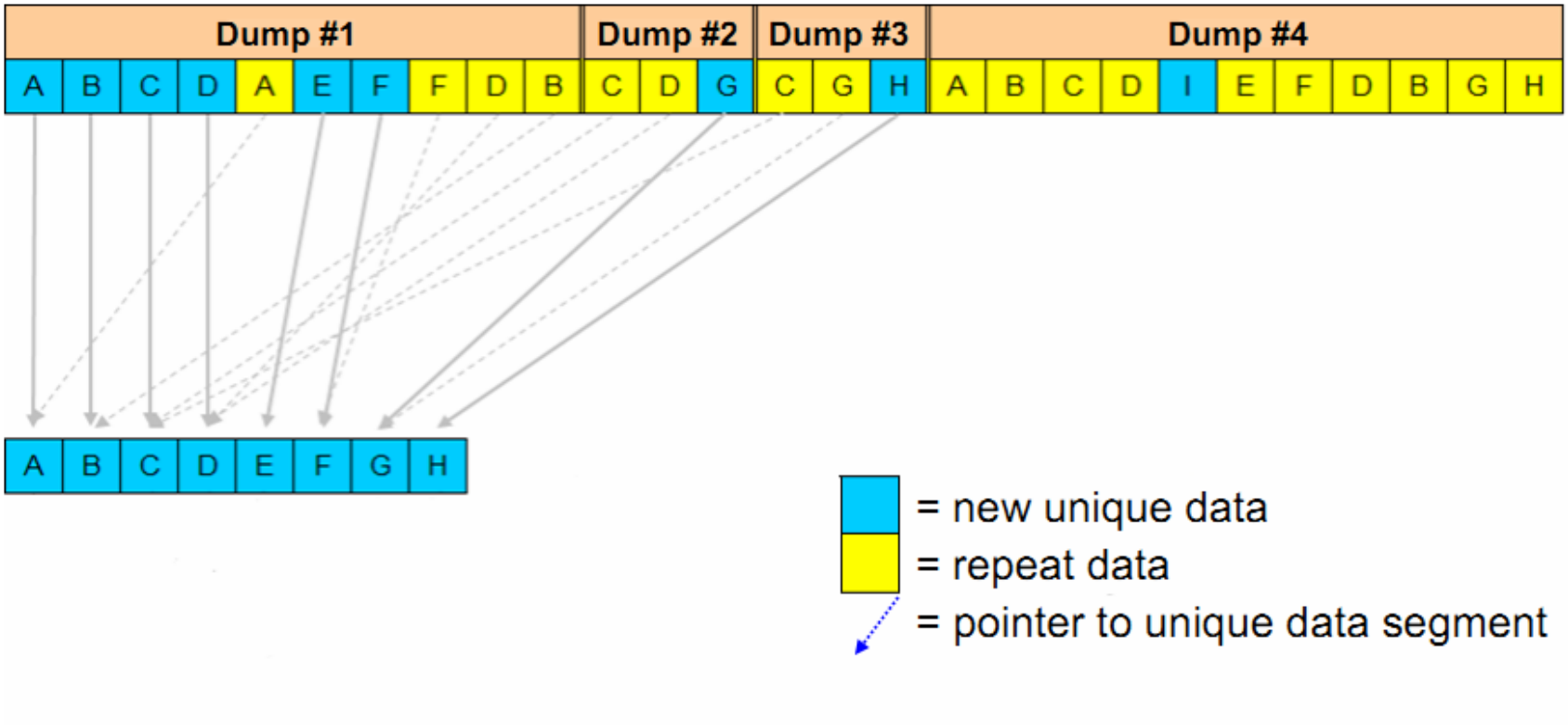
Deduplication



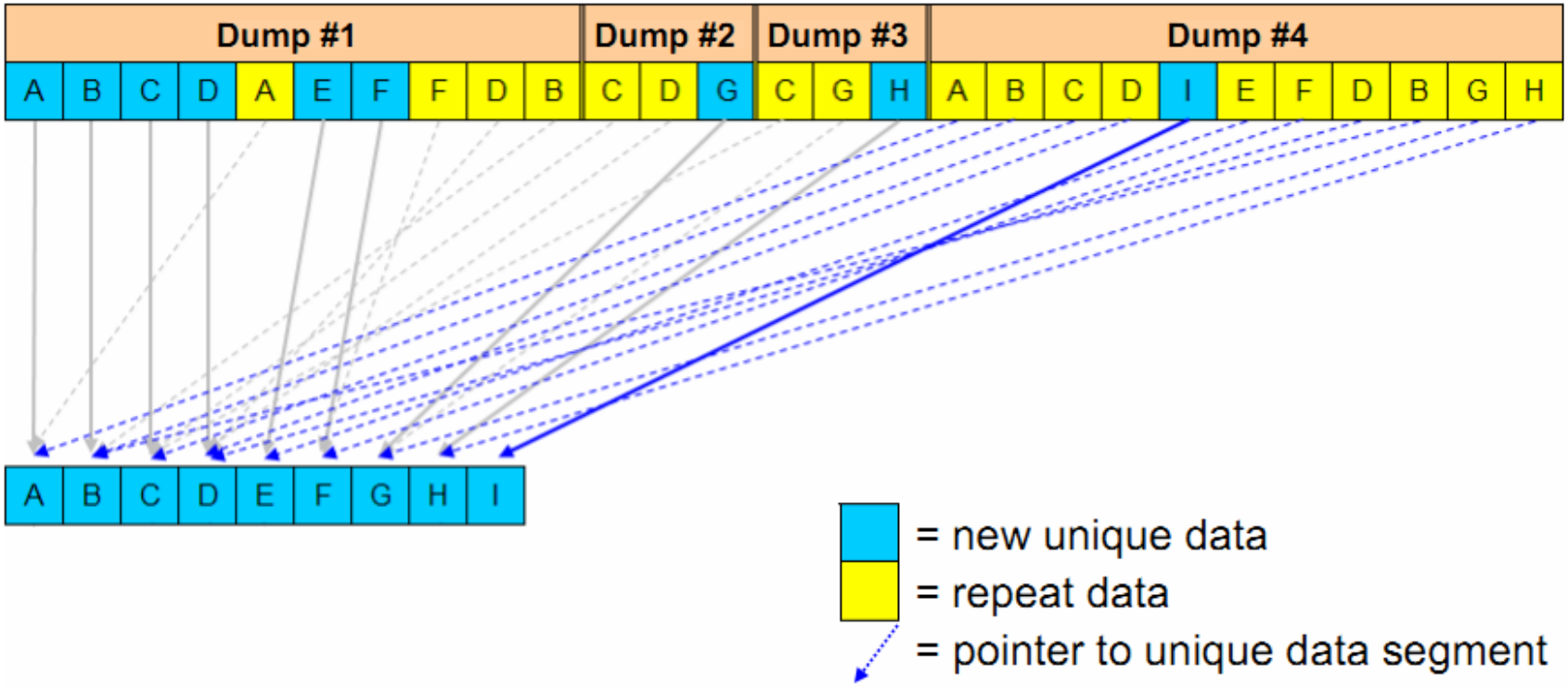
Deduplication



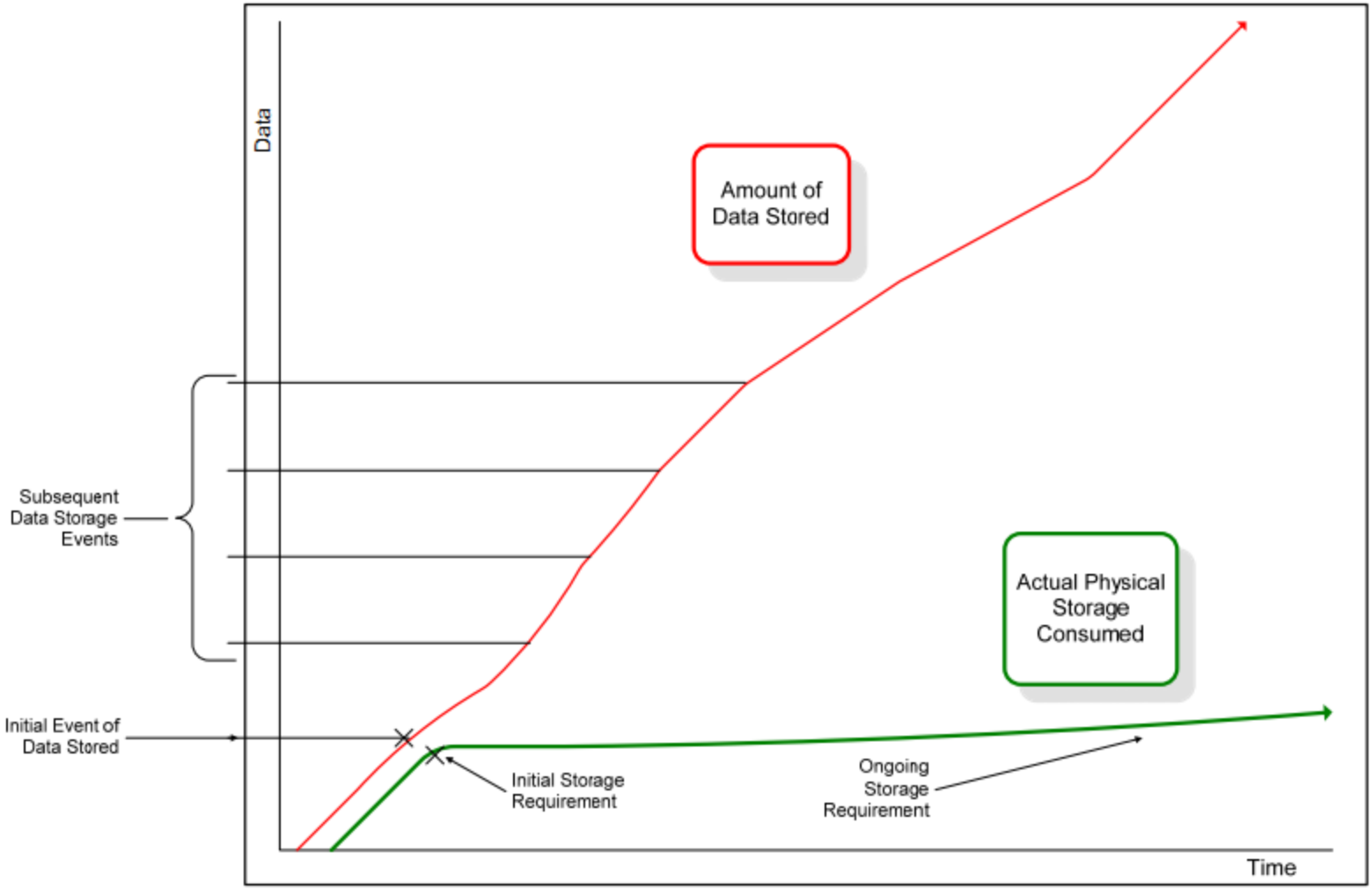
Deduplication



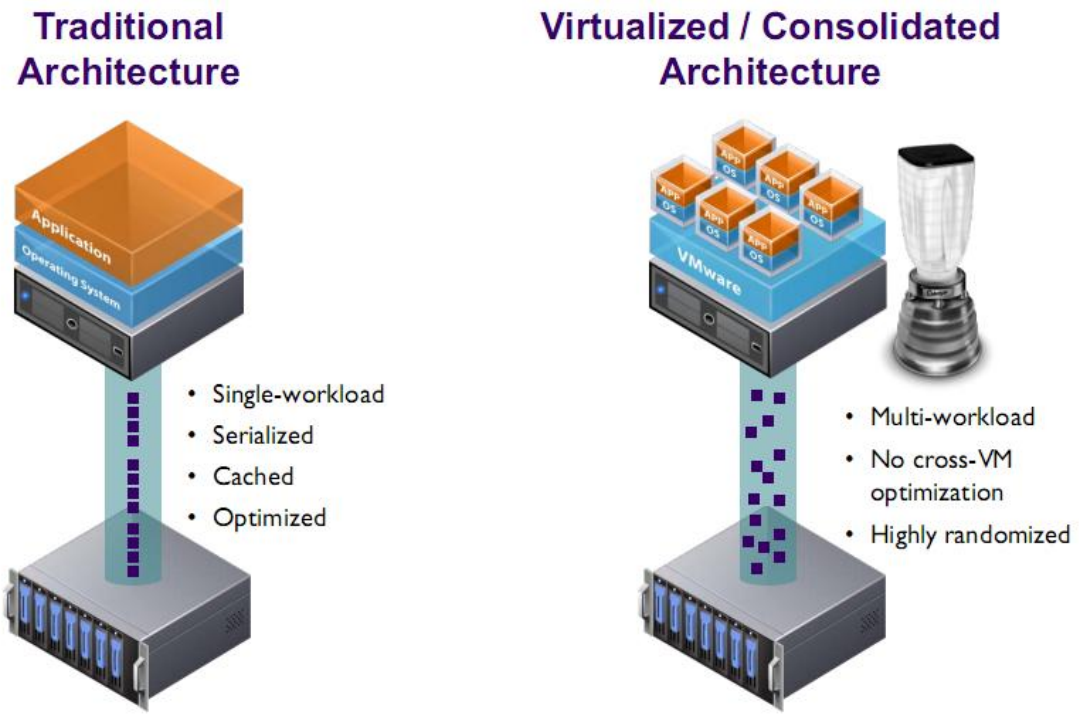
Deduplication



Deduplication

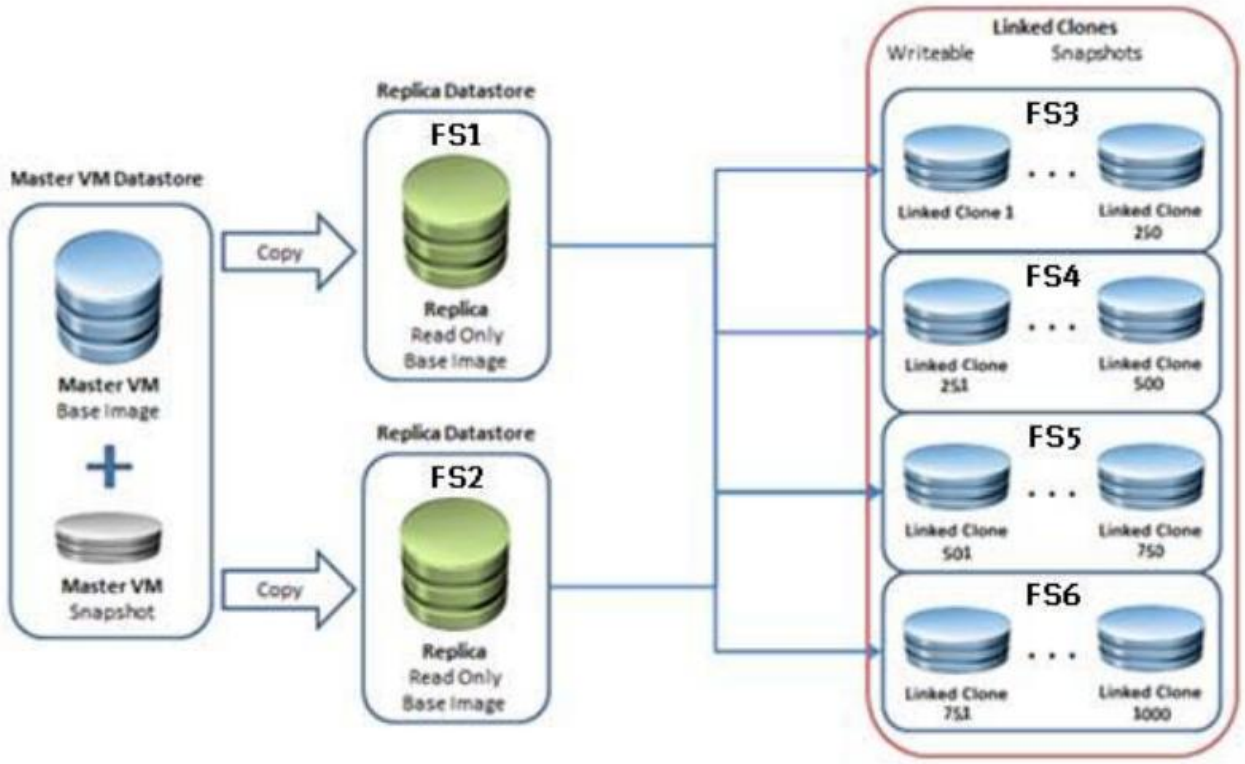


Pitfalls: I/O Blender



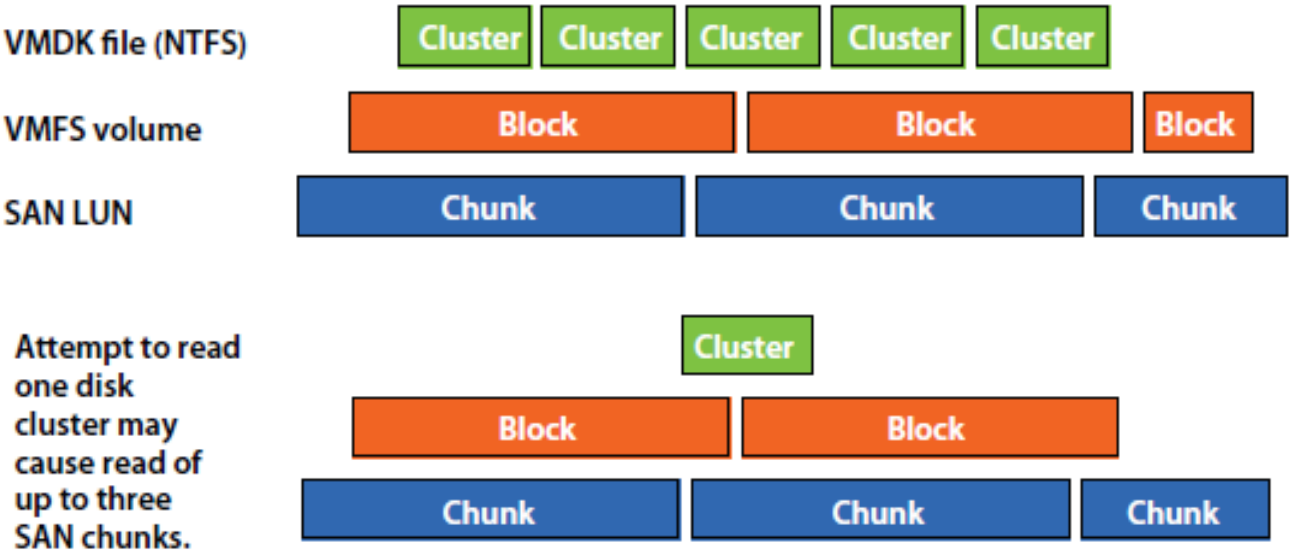
- Multiple simultaneous instances of VM sequential SCSI disk access turns into random access at host
  - Wreaks havoc with IOPS
  - Typical example: Simultaneous boot of multiple virtual desktops (*boot storm)*
  - **Exercise: Solution?**

Pitfalls: Bottlenecks



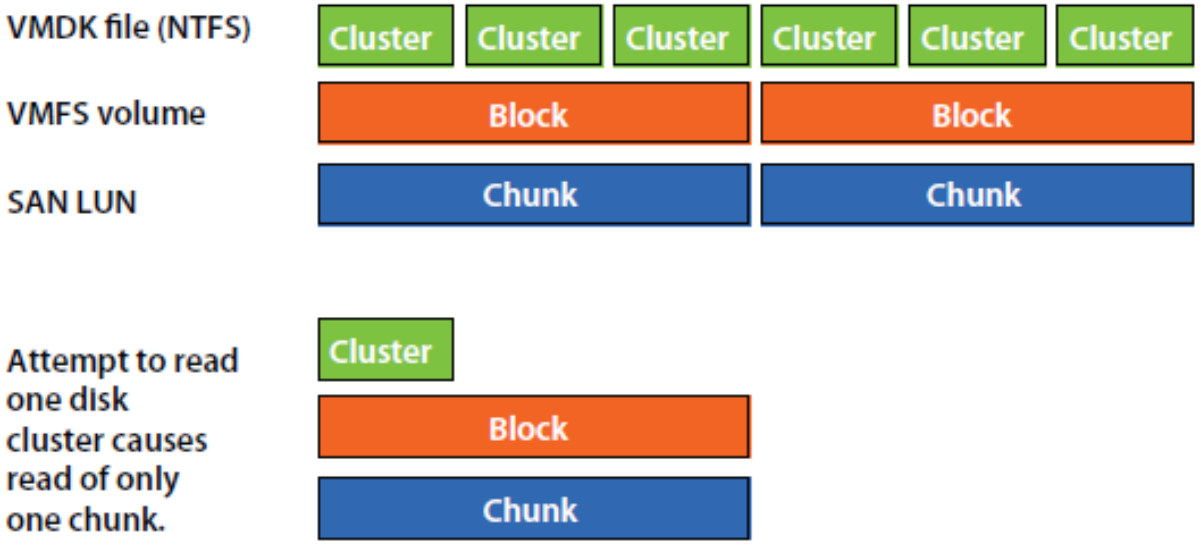
- Linked clones are great for booting hundreds of desktops
  - But: beware the I/O bottleneck accessing the "golden" image
  - **Exercise: Solution?**

Pitfalls: Misalignments



- Virtual storage can also suffer from performance problems if there is misalignment of blocks throughout the three layers of mappings
- In example, read of a VMDK block requires reading *two* blocks of the VMFS volume, which in turn requires reading *three blocks* of the real SAN
- One virtual IOPS would translate then into three real IOPS

Pitfalls: Misalignments



- Correct alignment of block borders throughout the three layers eliminates the problem
- Now, in the example, one virtual IOPS will require just a single IOPS on the real storage

Conclussions

- The SCSI model for storage, with its decoupling of logical blocks from their physical implementation, is today a highly successful tool for the implementation of flexible and sophisticated storage interconnects
- Interaction between servers and storage devices can nowadays be seen just as the transit of SCSI payloads over storage networks, with gateways allowing these payloads to jump between transport protocols, if required
- This concept allows flexibility to choose the physical interfaces at both servers and storage, and the convergence between data and storage networks, as both can today be implemented over the same Ethernet link

Conclusions

- The logical storage model of SCSI has been also critical for the great success of virtualization, as it allows to completely hide to the VM the way how storage is made available to its assigned LUNs and mapped to the LBA belonging to that LUN
- Use of shared storage under these mapping layers allow for VM techniques of migration and CPU load balancing which, simply, would be completely impossible to reproduce (in a cost-effective way) with real servers
- However, a clear understanding of the I/O behavior of applications and servers is still a must, to avoid both the old problems of I/O and the completely new problems that arise with the concurrent operation of hundreds of virtual machines targeting the same storage devices