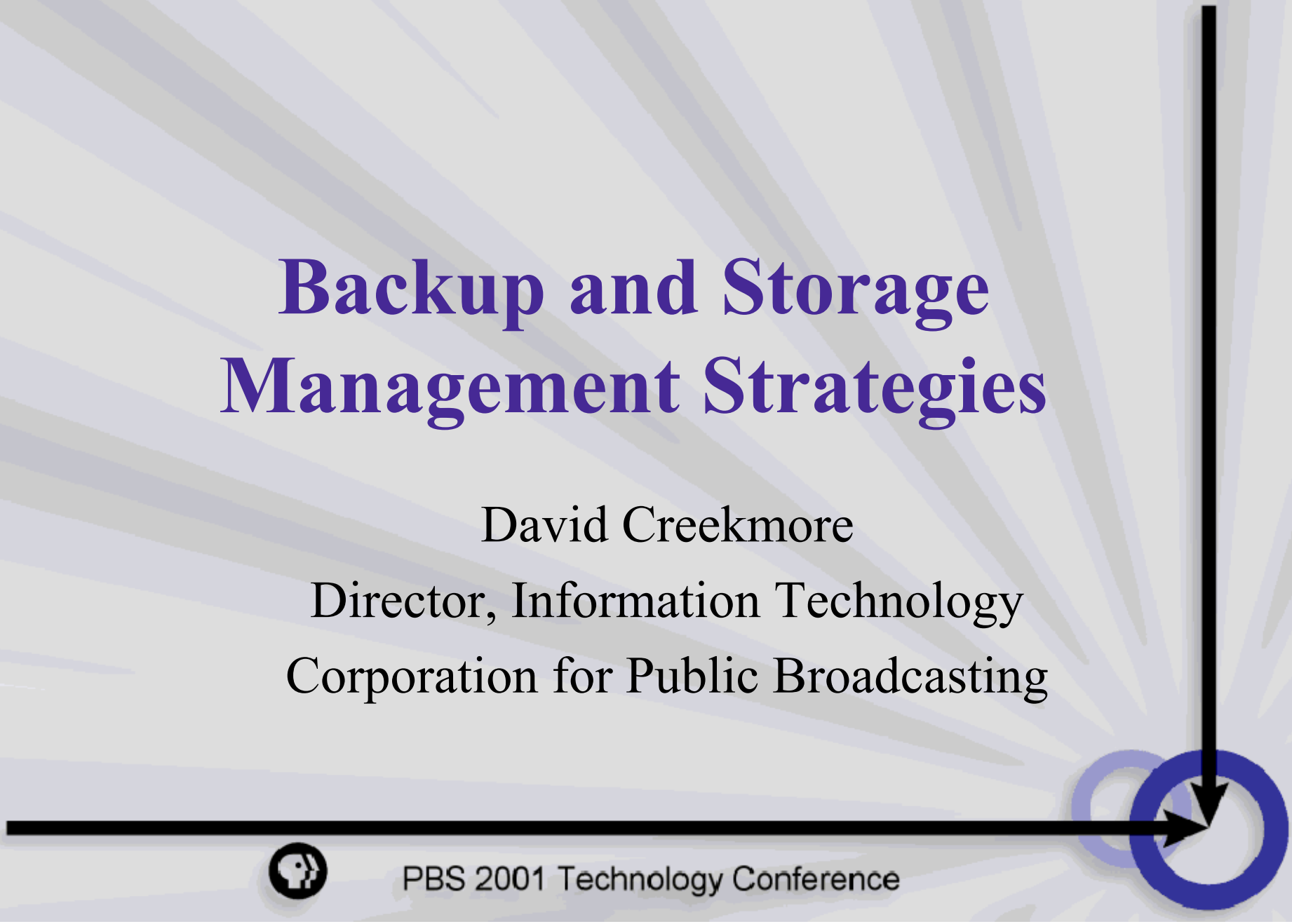
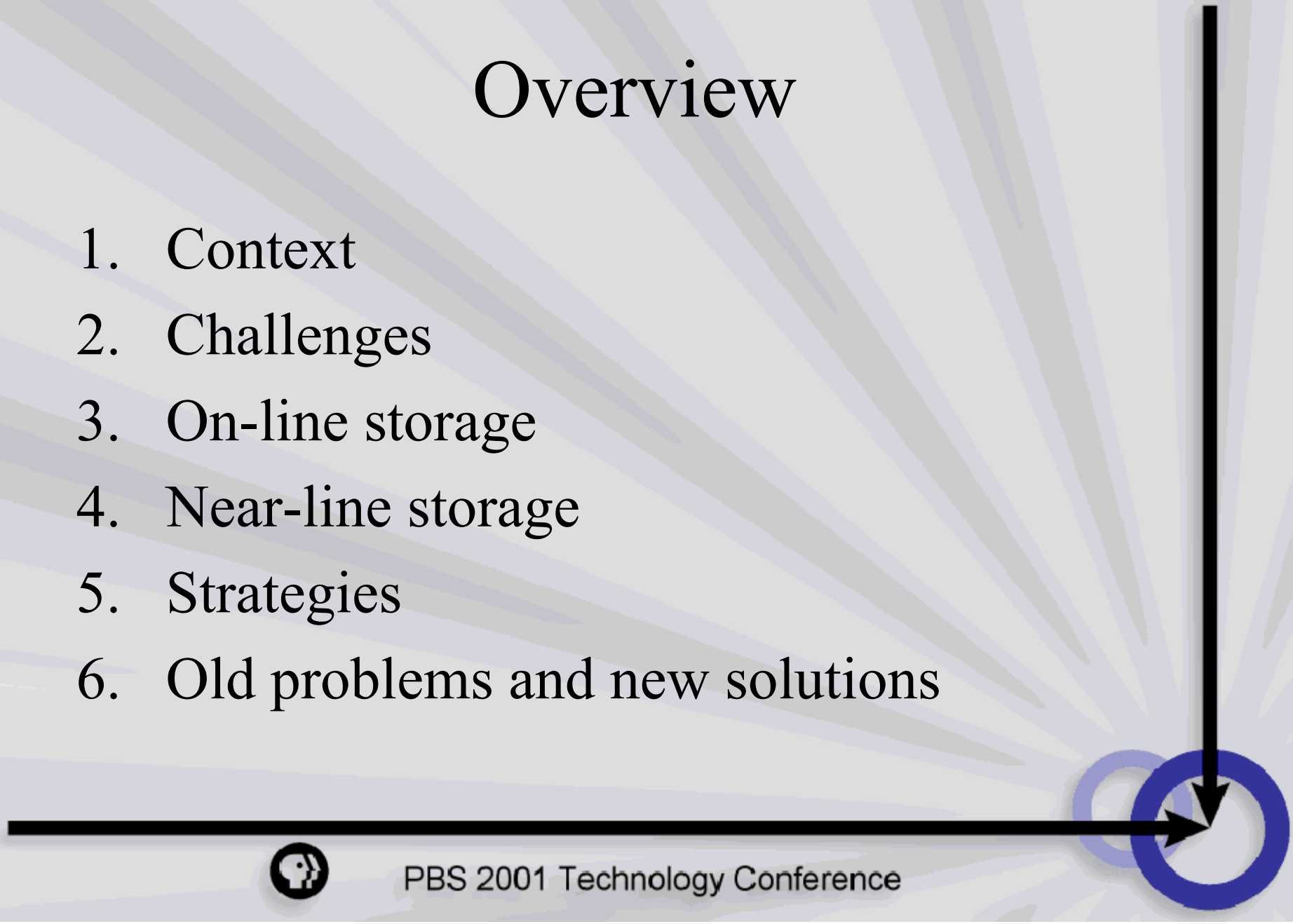
Backup and Storage Management Strategies

David Creekmore
Director, Information Technology
Corporation for Public Broadcasting



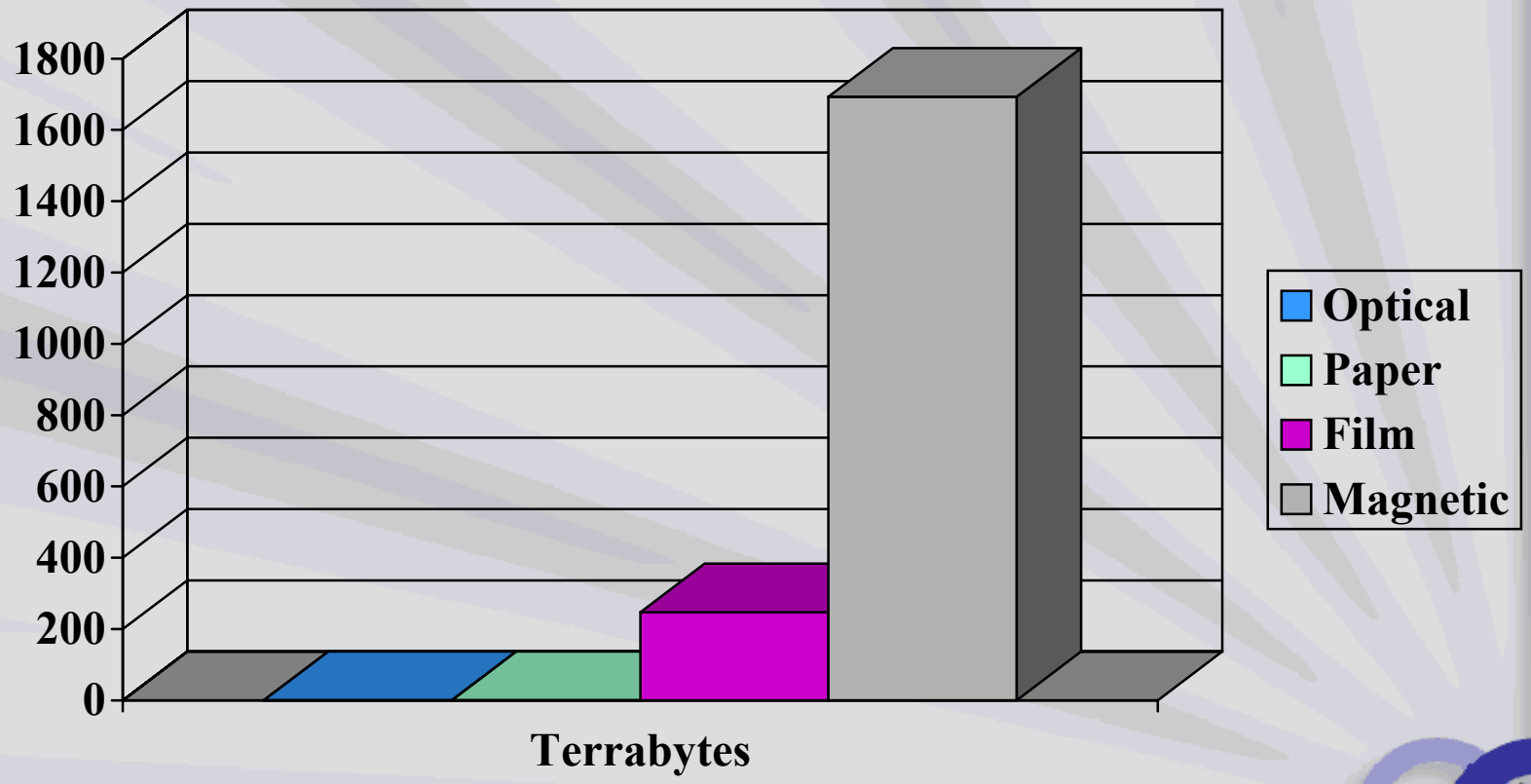
Overview

1. Context
2. Challenges
3. On-line storage
4. Near-line storage
5. Strategies
6. Old problems and new solutions

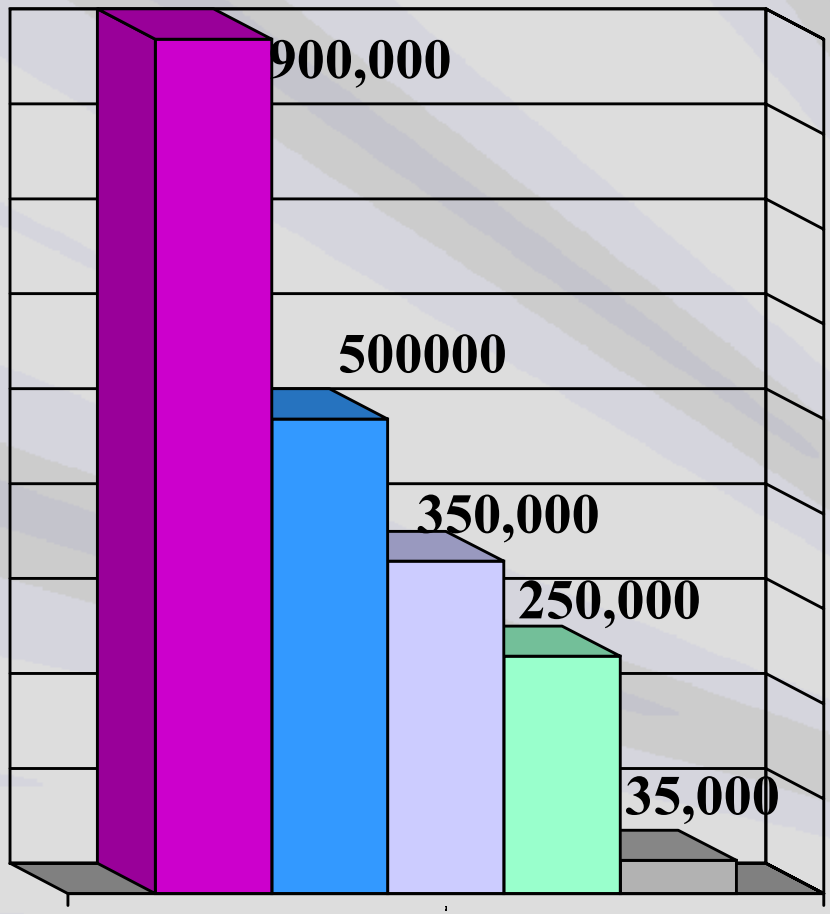


All the information in the world…

(over 2 billion Terabytes)



Where is the data stored?



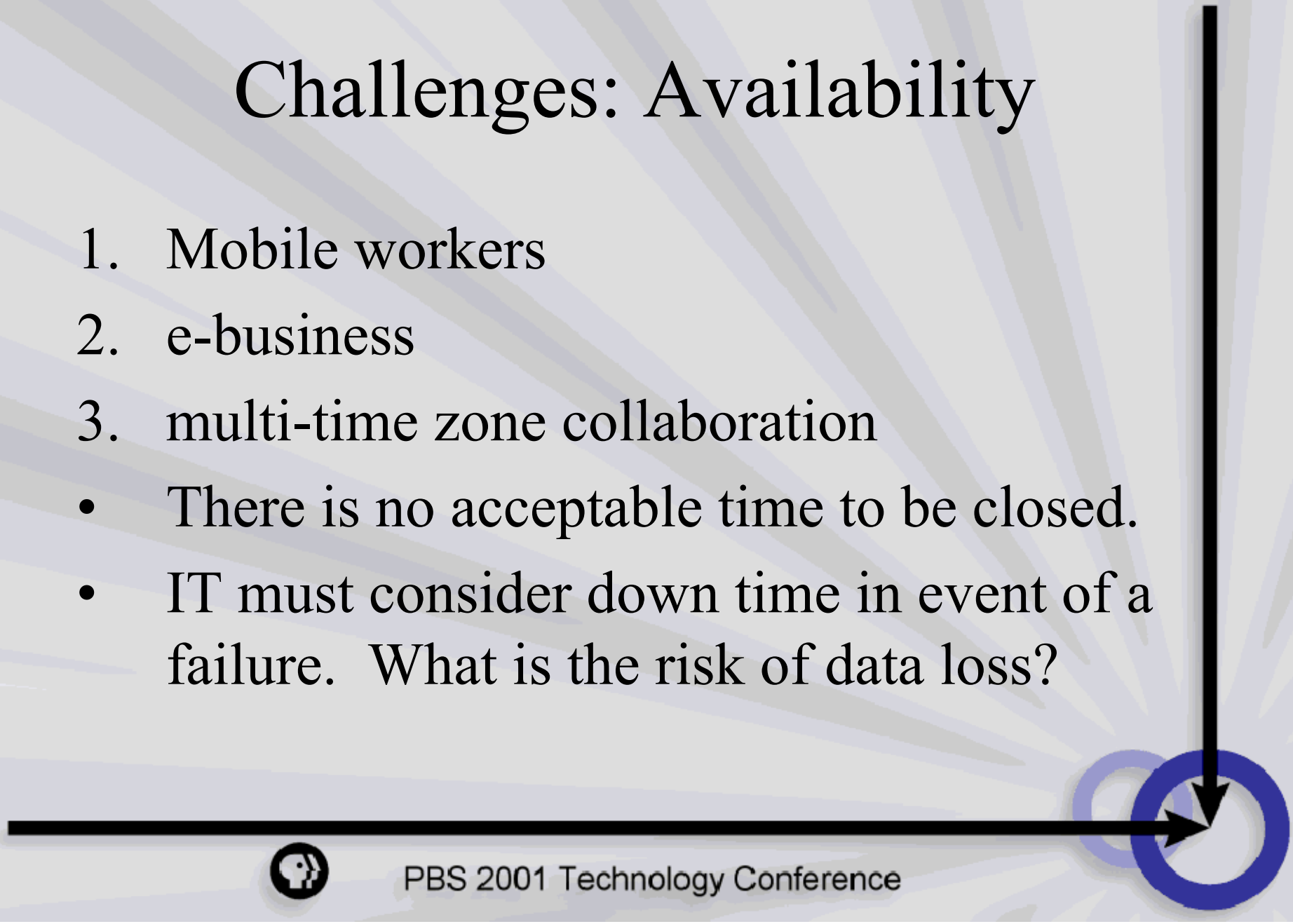
- PC disk drives
- Servers
- Photographs
- Camcorder Tape
- X-Rays

All Paper products	240
All Music CD's	8
All Cinema	2,000

Challenges: Availability

1. Mobile workers
2. e-business
3. multi-time zone collaboration

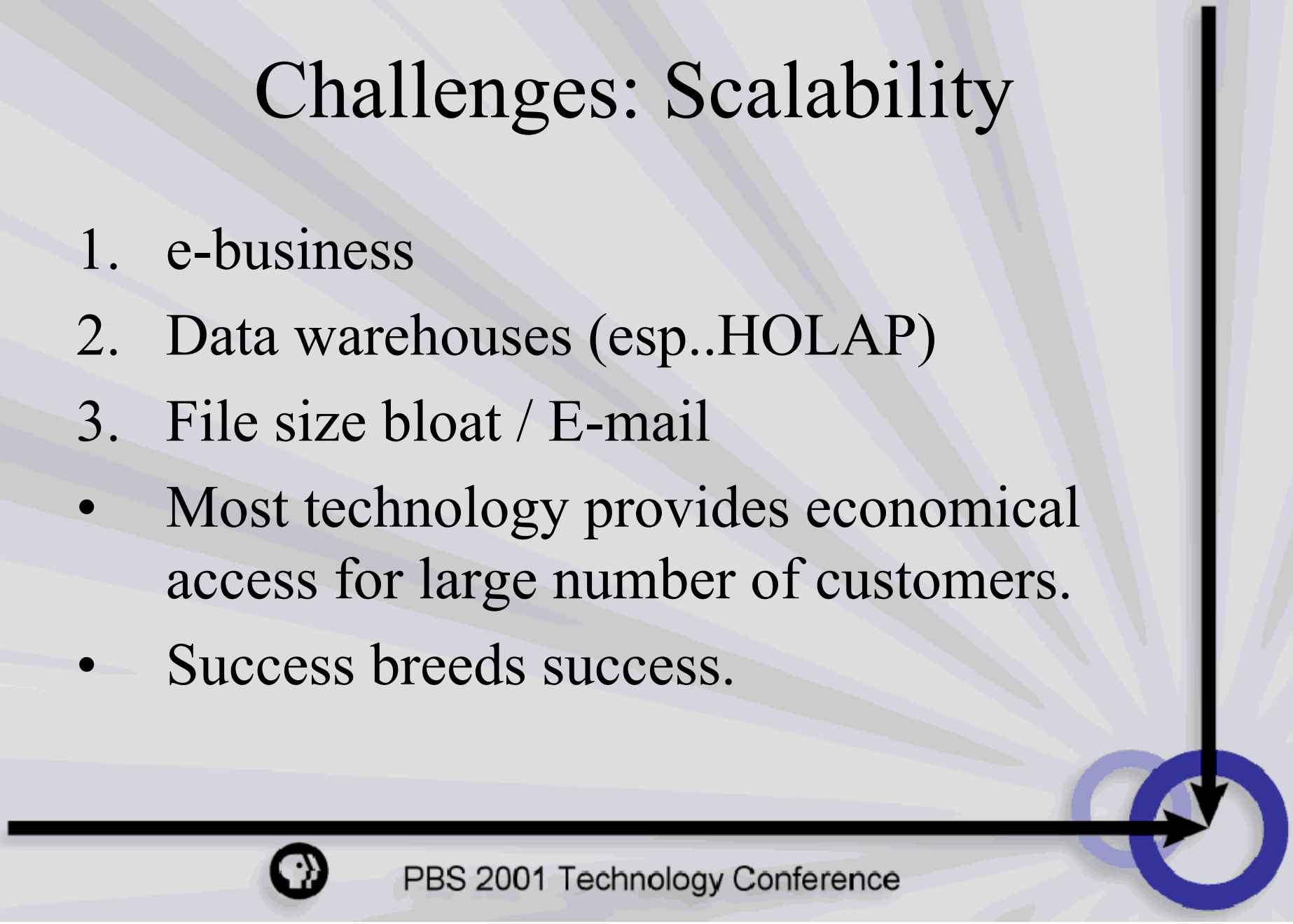
- There is no acceptable time to be closed.
- IT must consider down time in event of a failure. What is the risk of data loss?



Challenges: Scalability

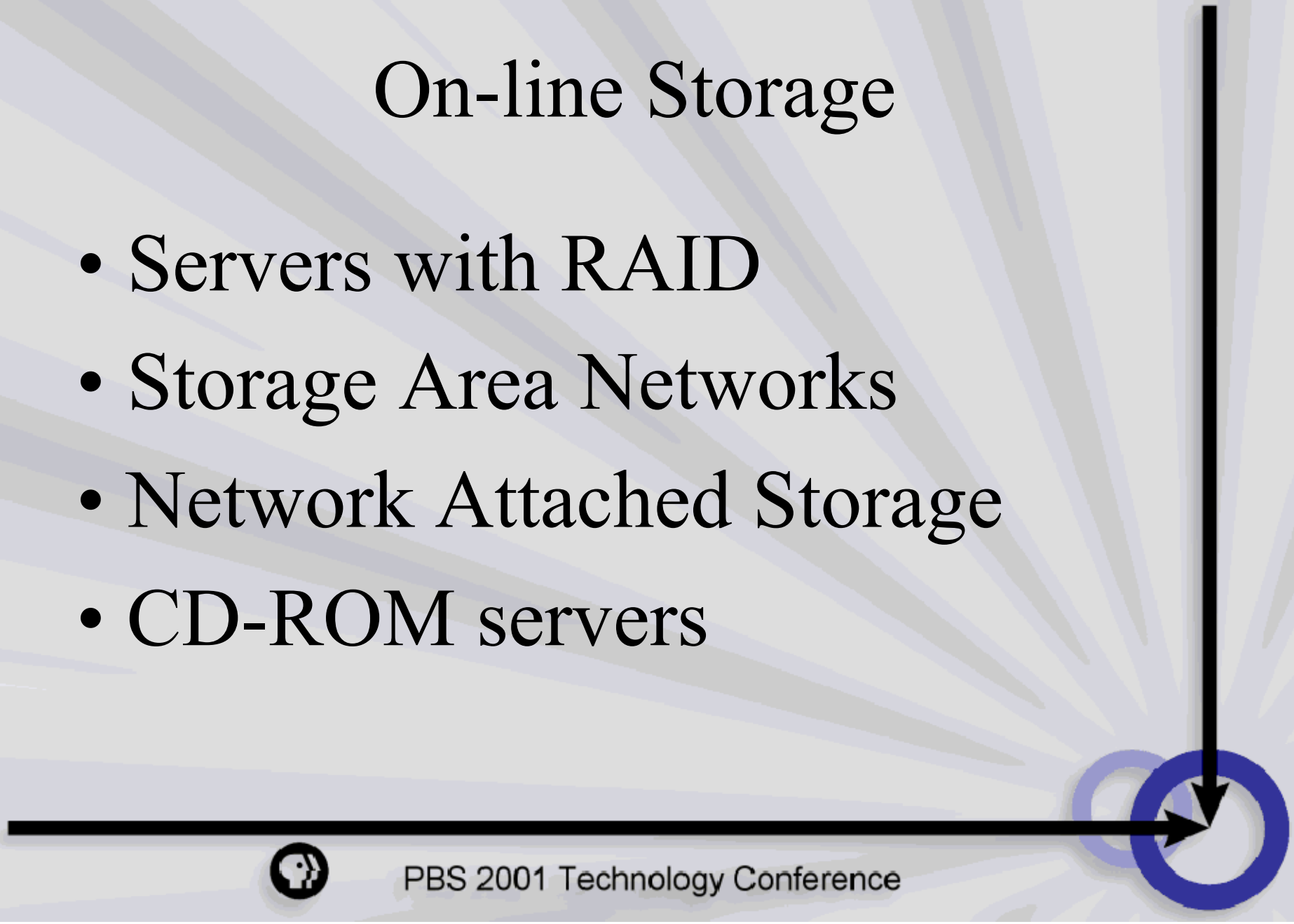
1. e-business
2. Data warehouses (esp..HOLAP)
3. File size bloat / E-mail

- Most technology provides economical access for large number of customers.
- Success breeds success.



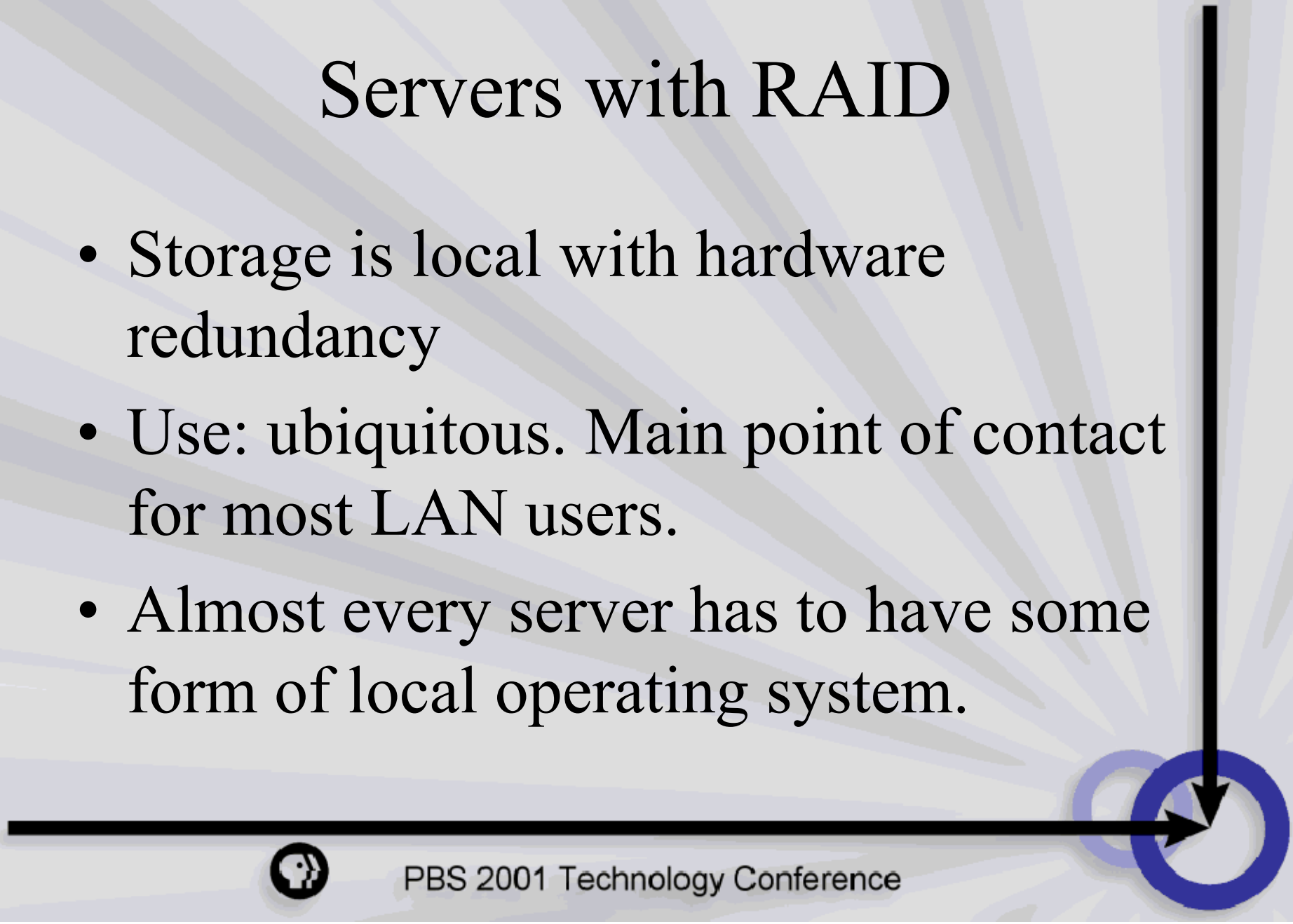
On-line Storage

- Servers with RAID
- Storage Area Networks
- Network Attached Storage
- CD-ROM servers



Servers with RAID

- Storage is local with hardware redundancy
- Use: ubiquitous. Main point of contact for most LAN users.
- Almost every server has to have some form of local operating system.



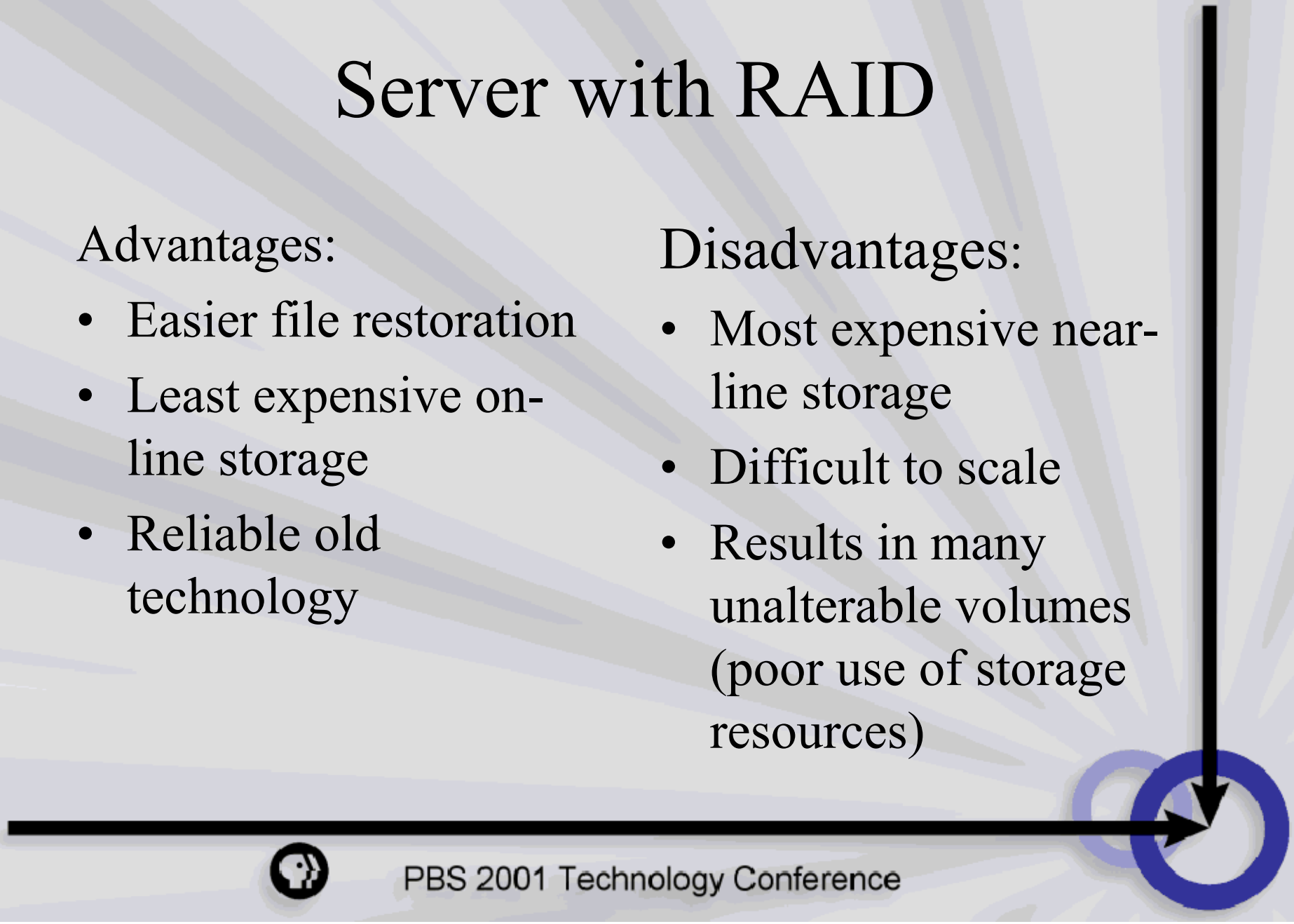
Server with RAID

Advantages:

- Easier file restoration
- Least expensive on-line storage
- Reliable old technology

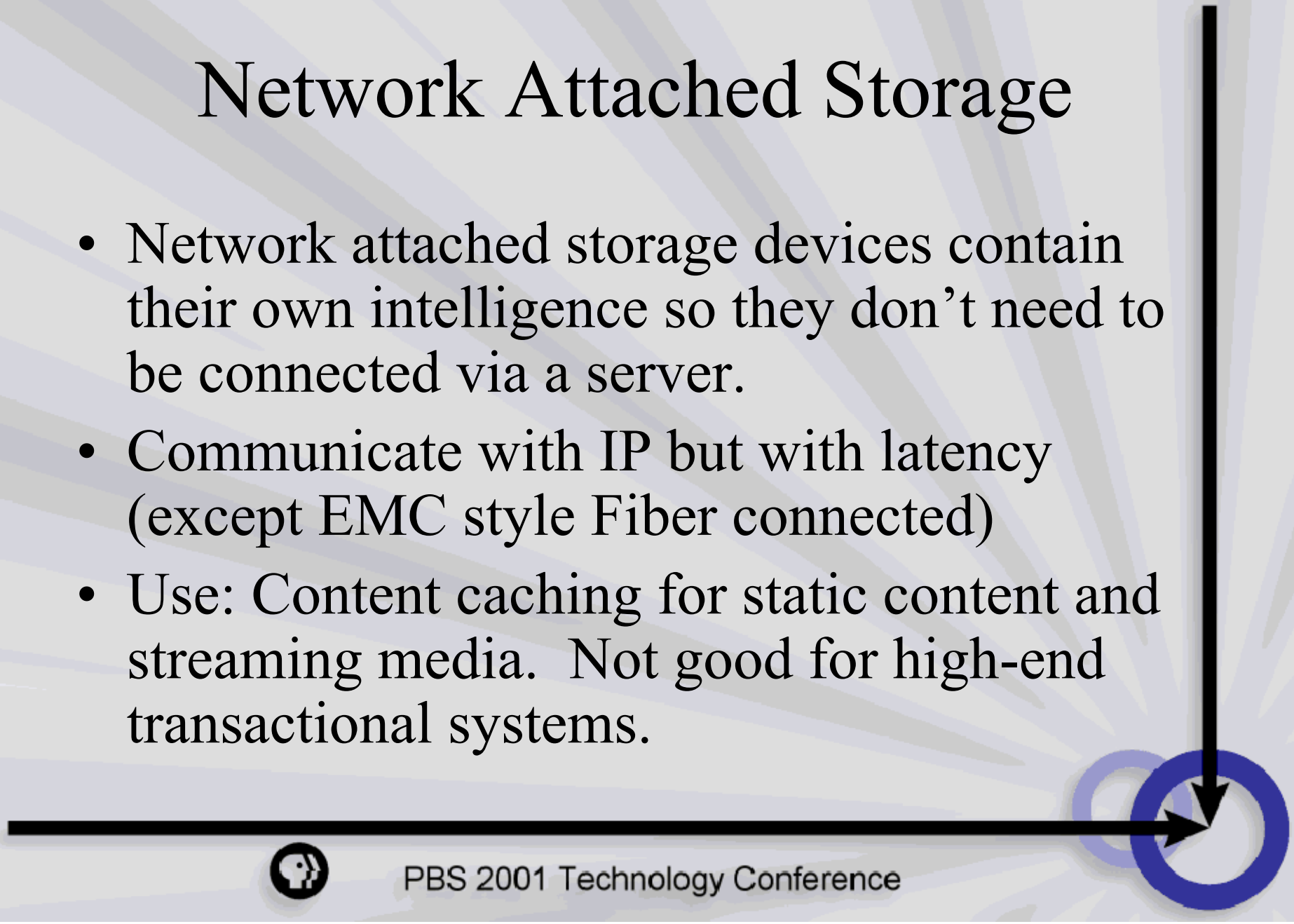
Disadvantages:

- Most expensive near-line storage
- Difficult to scale
- Results in many unalterable volumes (poor use of storage resources)



Network Attached Storage

- Network attached storage devices contain their own intelligence so they don't need to be connected via a server.
- Communicate with IP but with latency (except EMC style Fiber connected)
- Use: Content caching for static content and streaming media. Not good for high-end transactional systems.



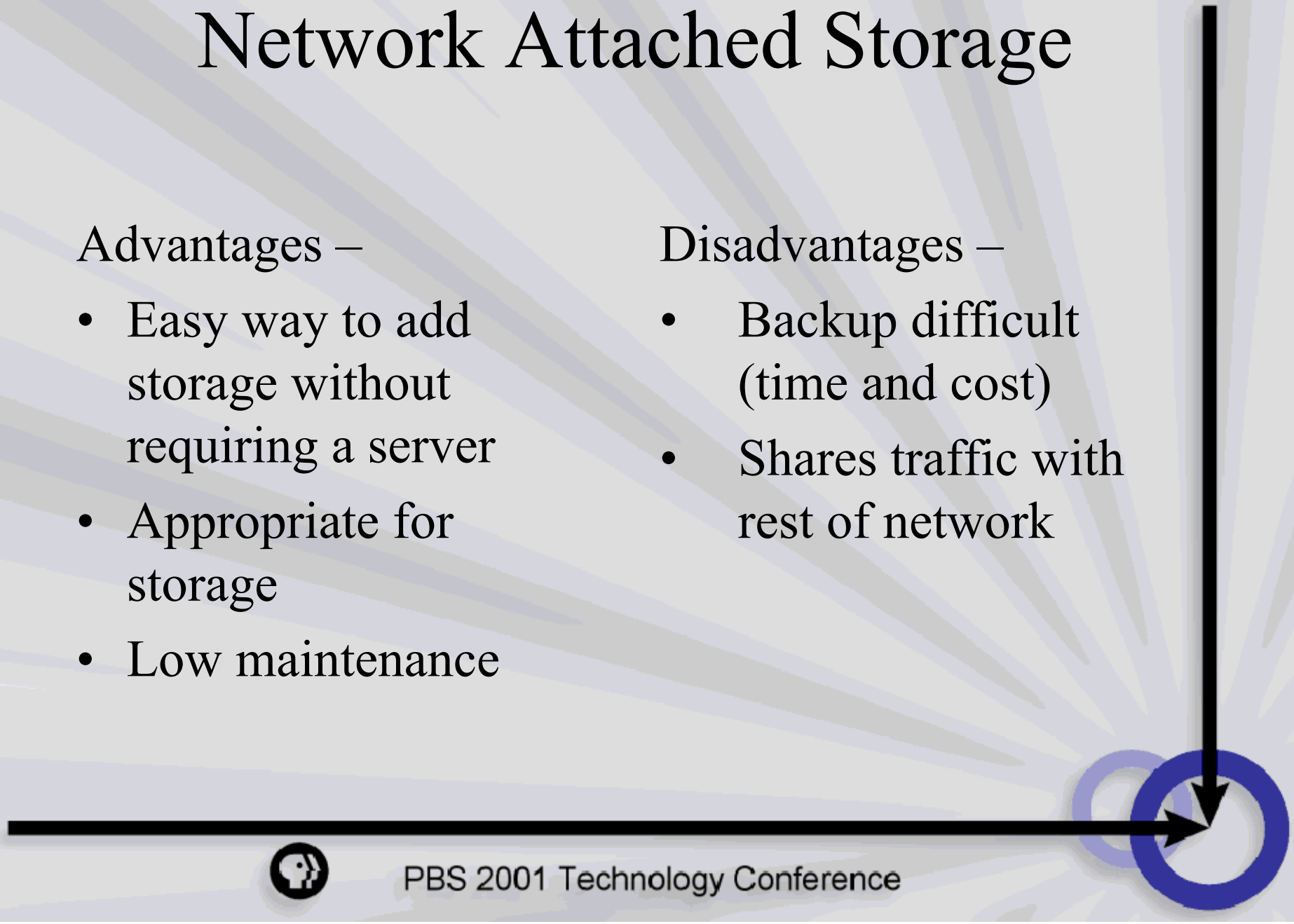
Network Attached Storage

Advantages –

- Easy way to add storage without requiring a server
- Appropriate for storage
- Low maintenance

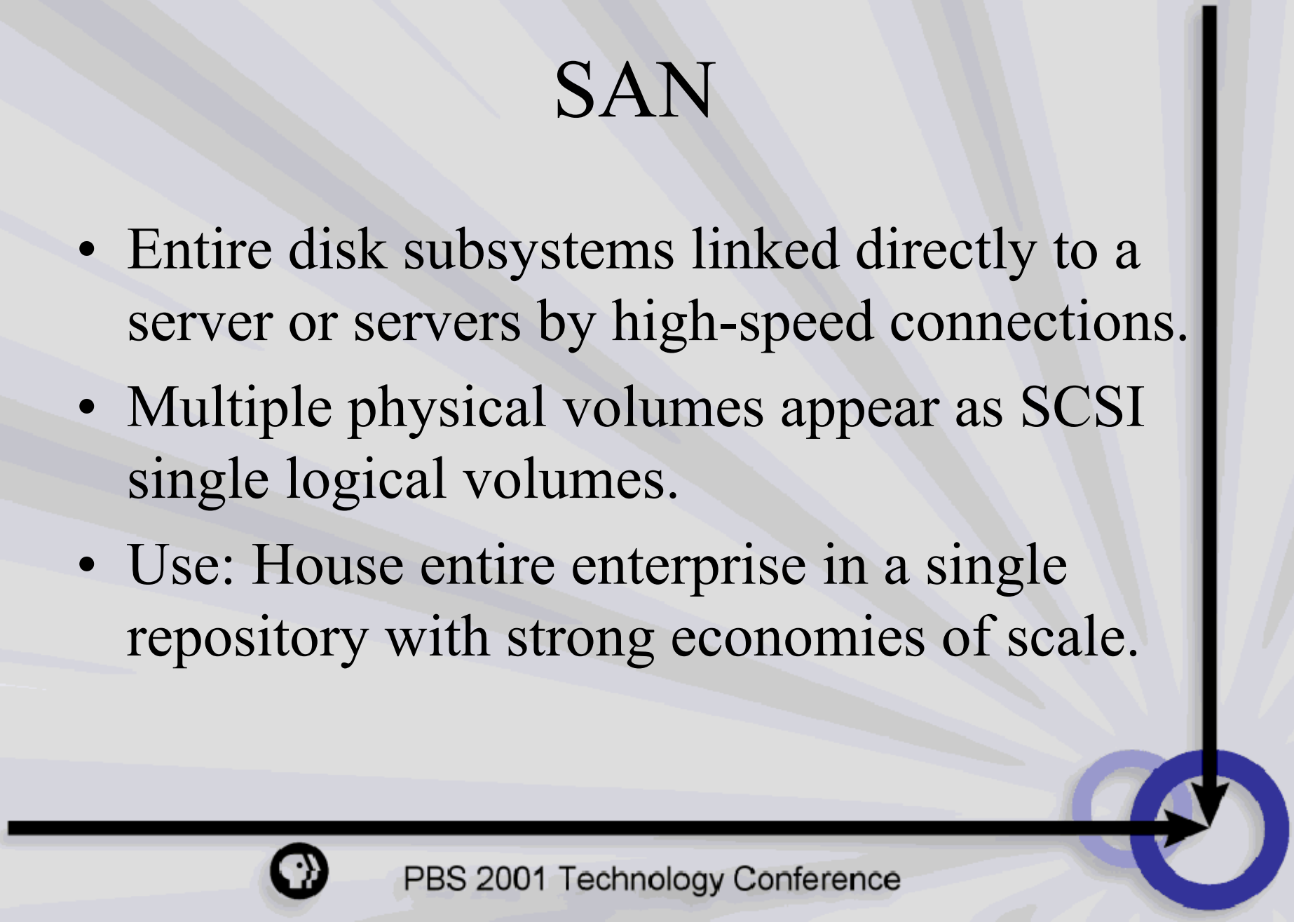
Disadvantages –

- Backup difficult (time and cost)
- Shares traffic with rest of network



SAN

- Entire disk subsystems linked directly to a server or servers by high-speed connections.
- Multiple physical volumes appear as SCSI single logical volumes.
- Use: House entire enterprise in a single repository with strong economies of scale.



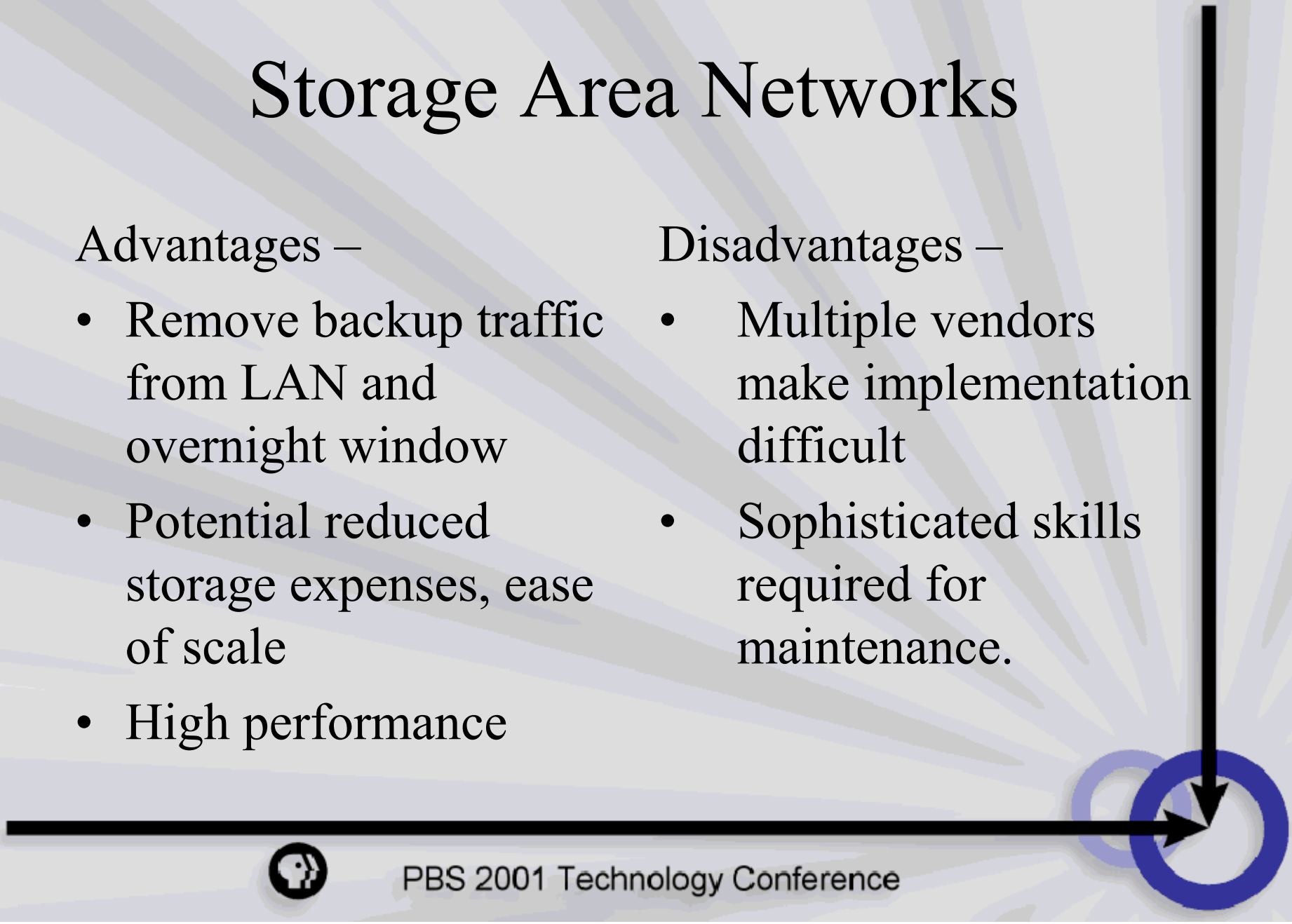
Storage Area Networks

Advantages –

- Remove backup traffic from LAN and overnight window
- Potential reduced storage expenses, ease of scale
- High performance

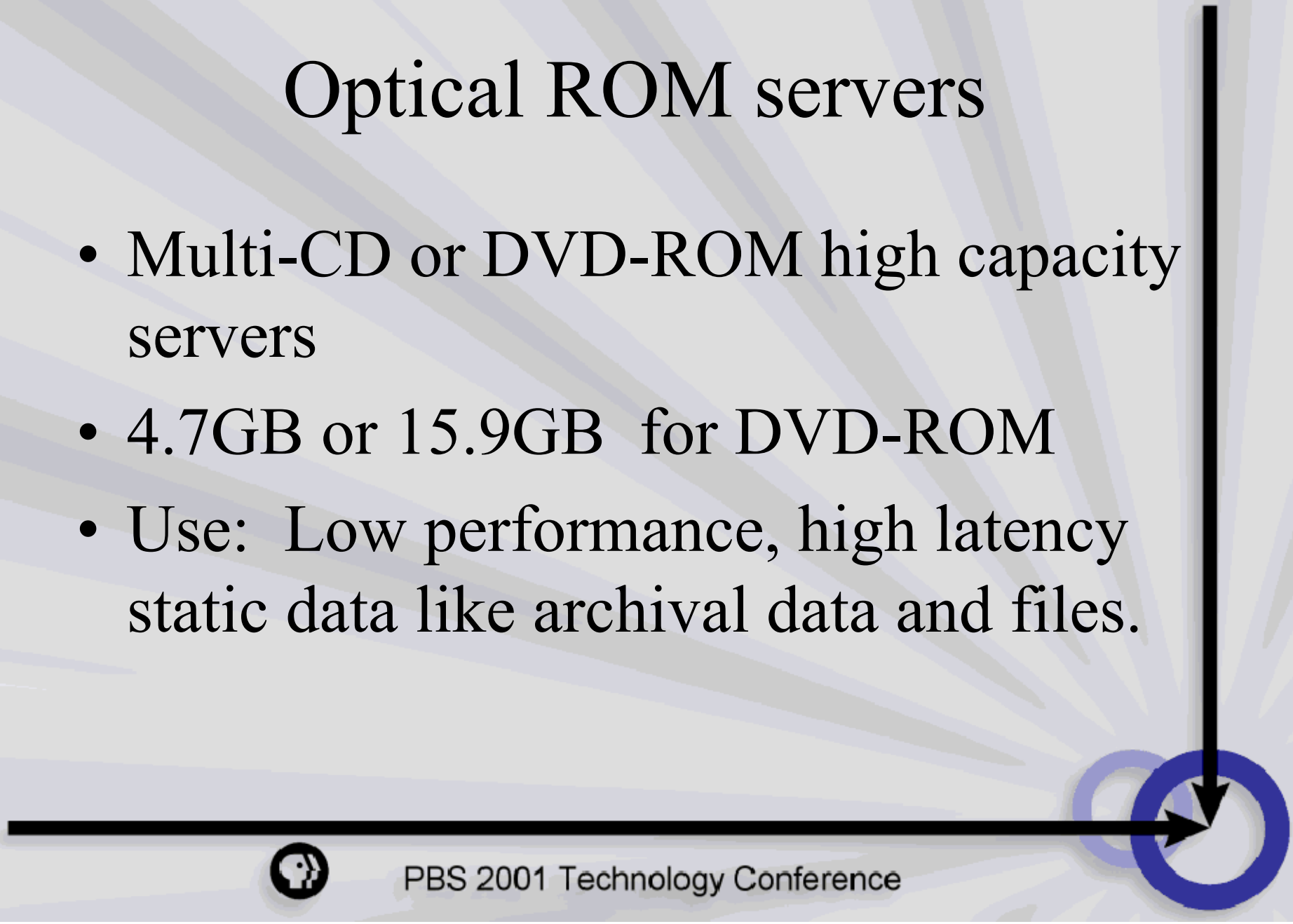
Disadvantages –

- Multiple vendors make implementation difficult
- Sophisticated skills required for maintenance.



Optical ROM servers

- Multi-CD or DVD-ROM high capacity servers
- 4.7GB or 15.9GB for DVD-ROM
- Use: Low performance, high latency static data like archival data and files.



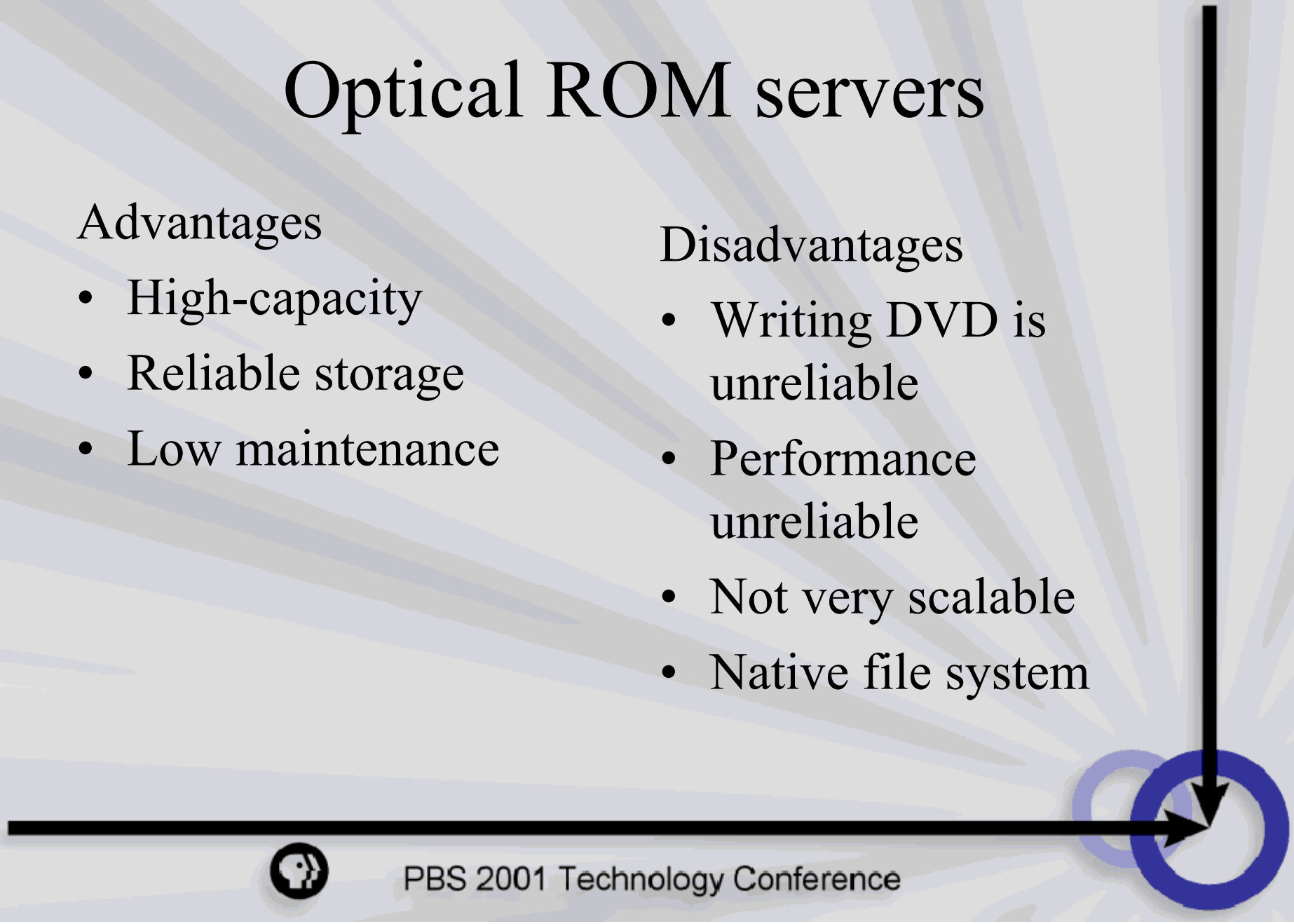
Optical ROM servers

Advantages

- High-capacity
- Reliable storage
- Low maintenance

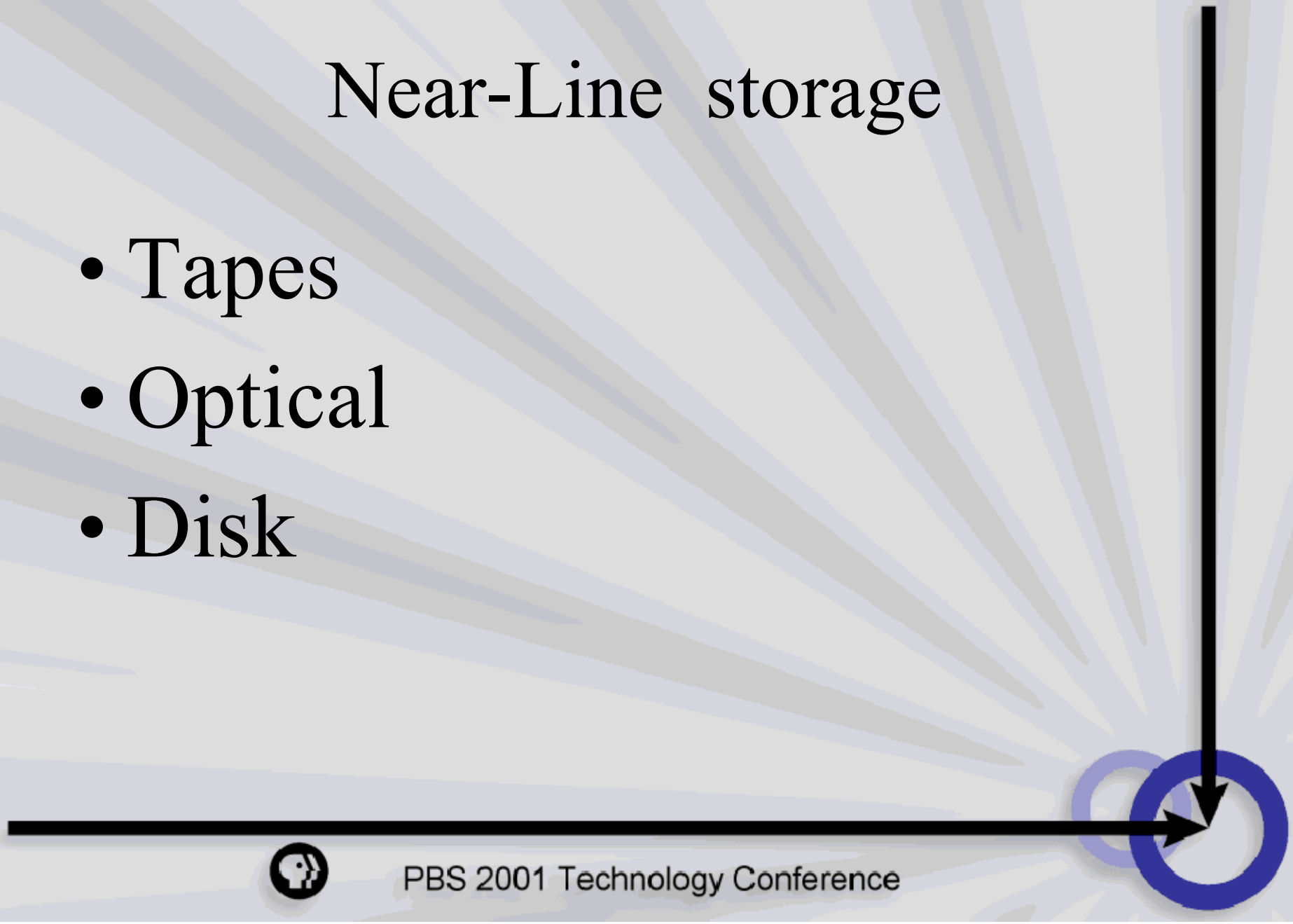
Disadvantages

- Writing DVD is unreliable
- Performance unreliable
- Not very scalable
- Native file system



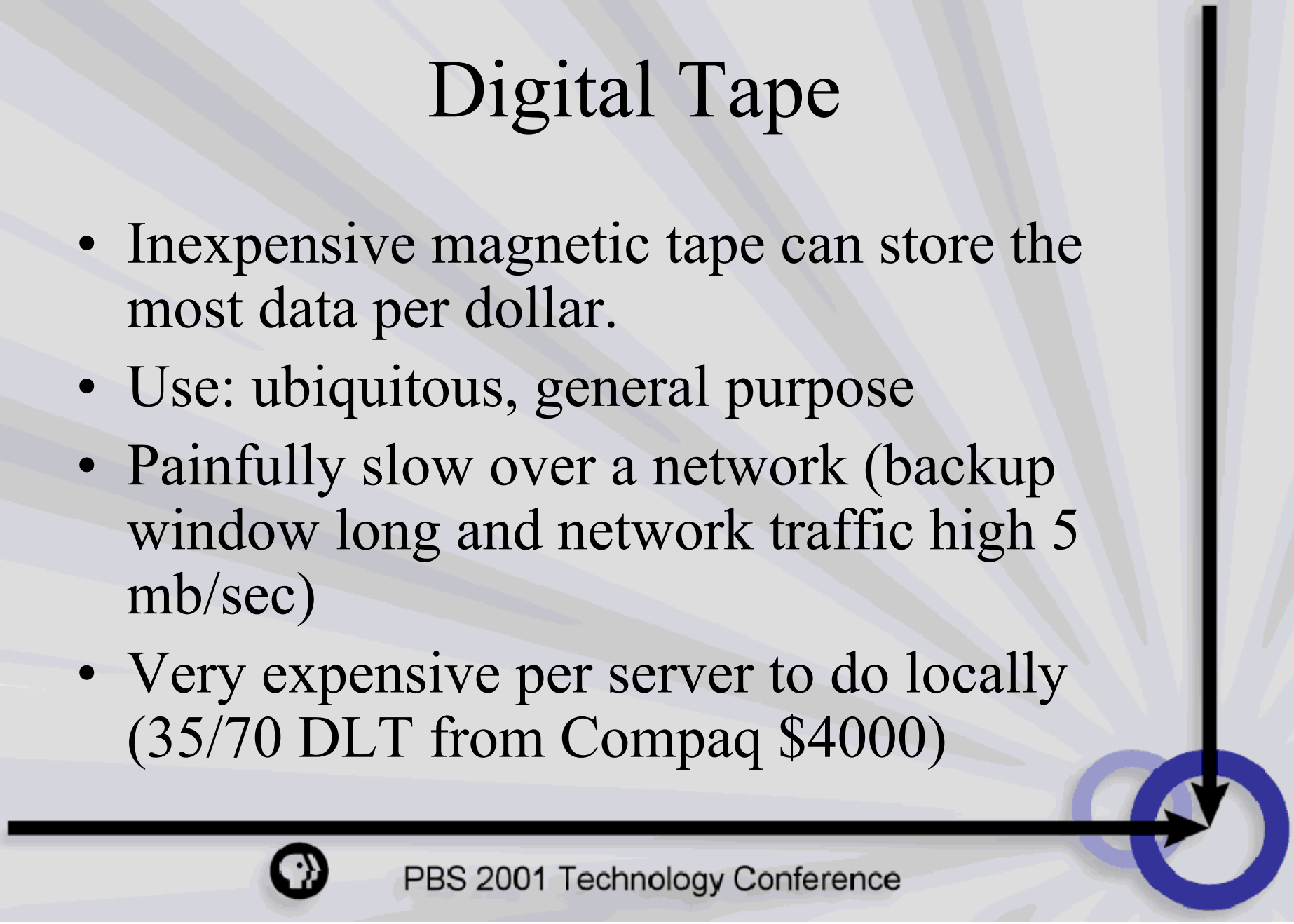
Near-Line storage

- Tapes
- Optical
- Disk



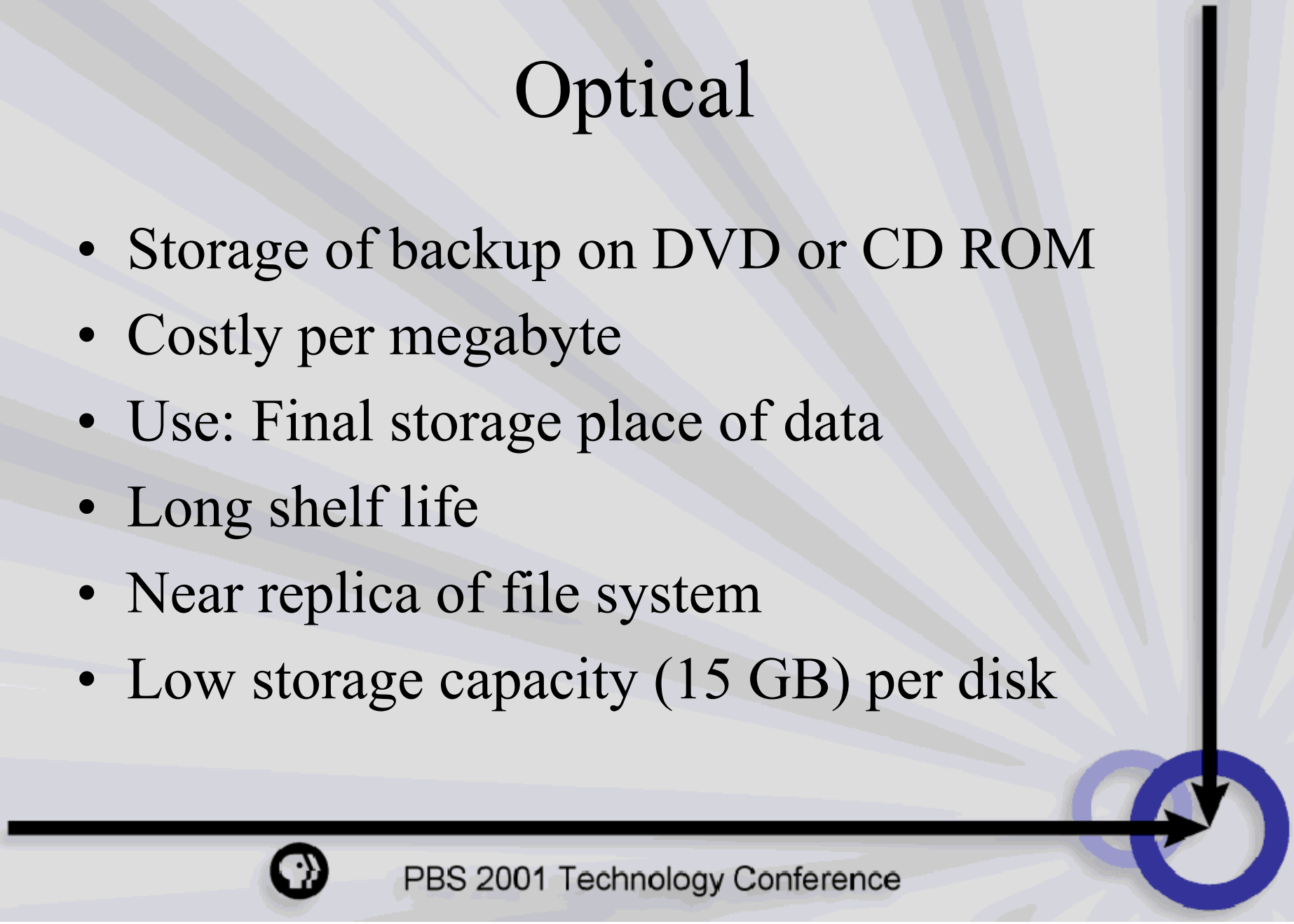
Digital Tape

- Inexpensive magnetic tape can store the most data per dollar.
- Use: ubiquitous, general purpose
- Painfully slow over a network (backup window long and network traffic high 5 mb/sec)
- Very expensive per server to do locally (35/70 DLT from Compaq $4000)



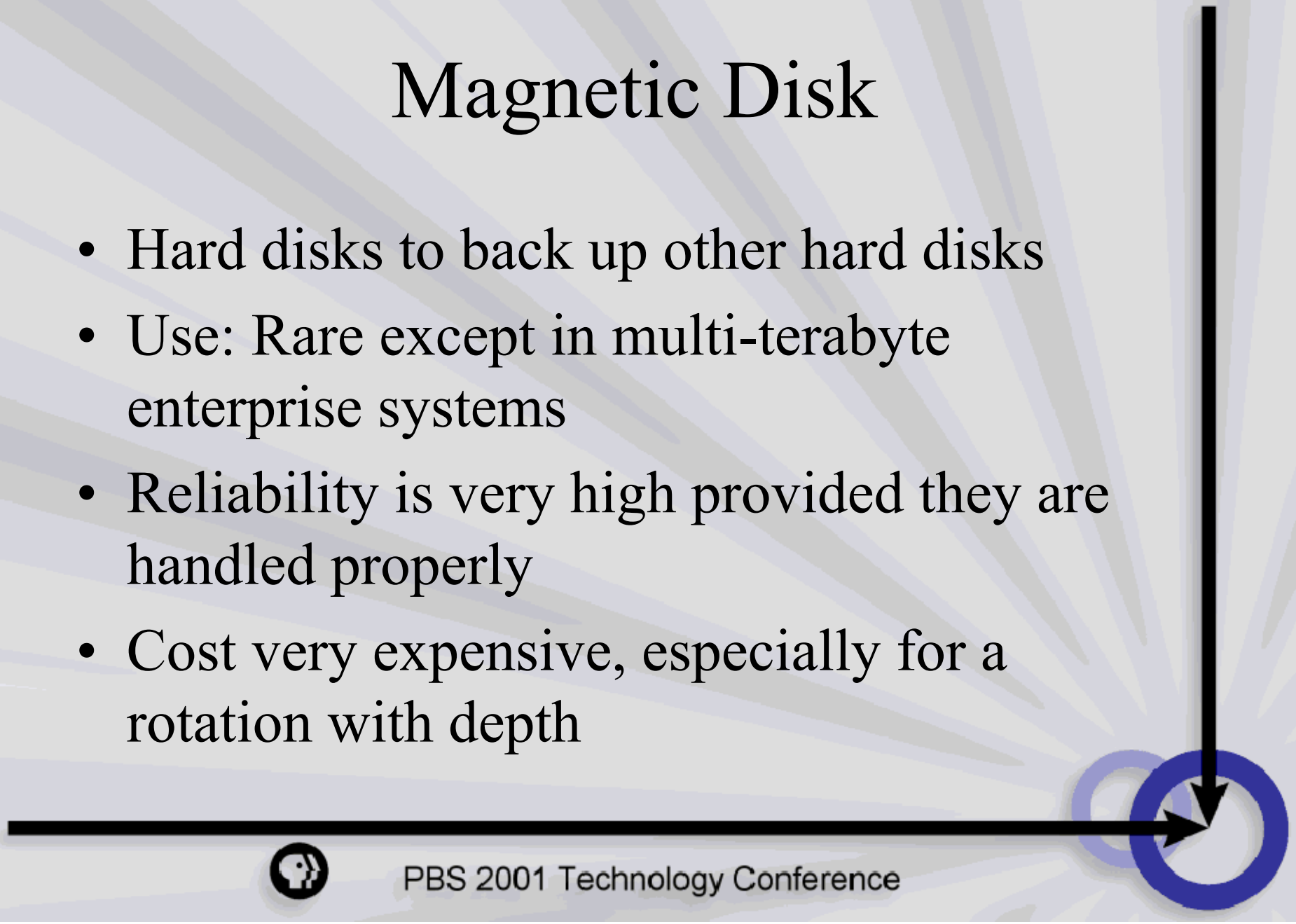
Optical

- Storage of backup on DVD or CD ROM
- Costly per megabyte
- Use: Final storage place of data
- Long shelf life
- Near replica of file system
- Low storage capacity (15 GB) per disk



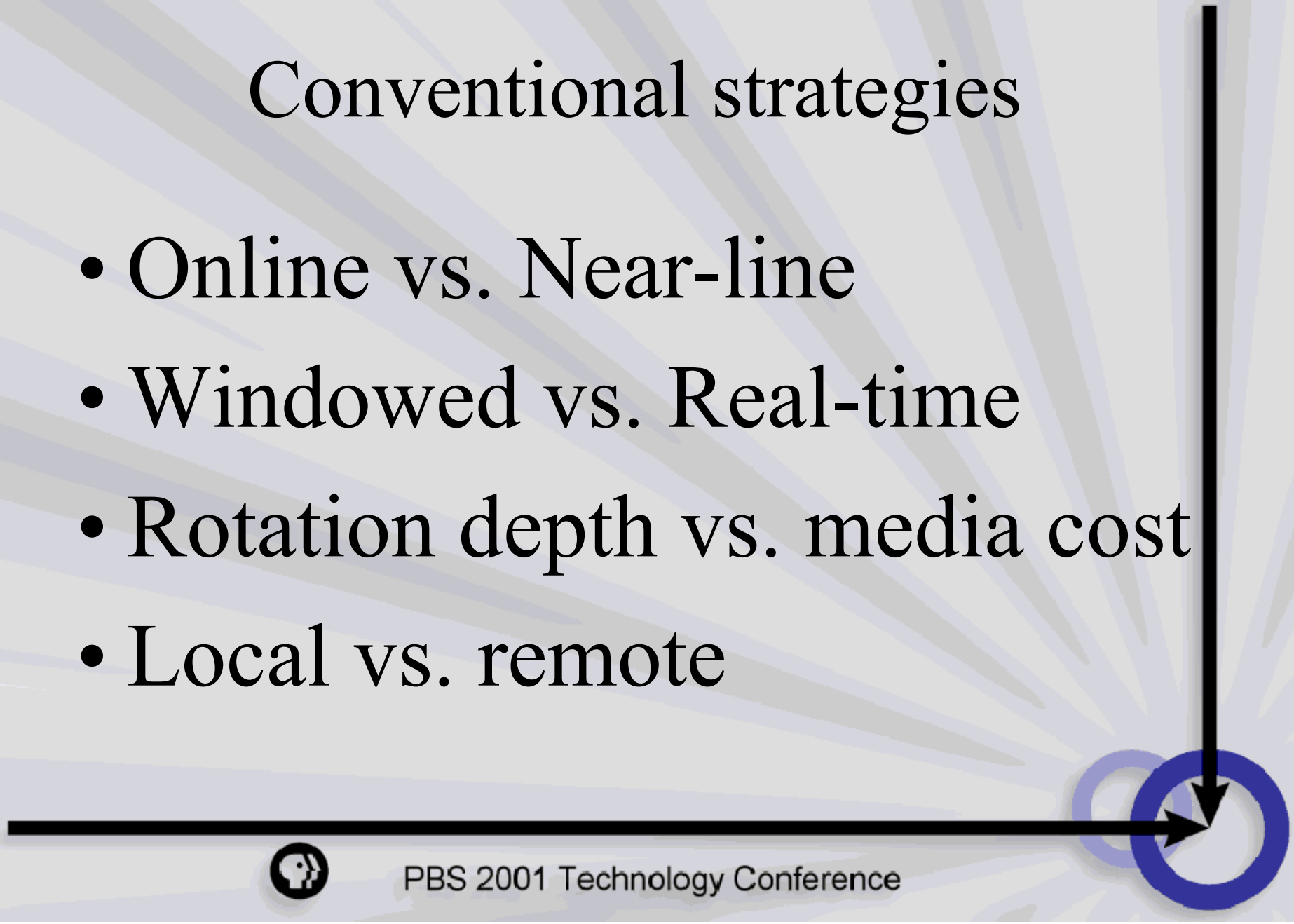
Magnetic Disk

- Hard disks to back up other hard disks
- Use: Rare except in multi-terabyte enterprise systems
- Reliability is very high provided they are handled properly
- Cost very expensive, especially for a rotation with depth



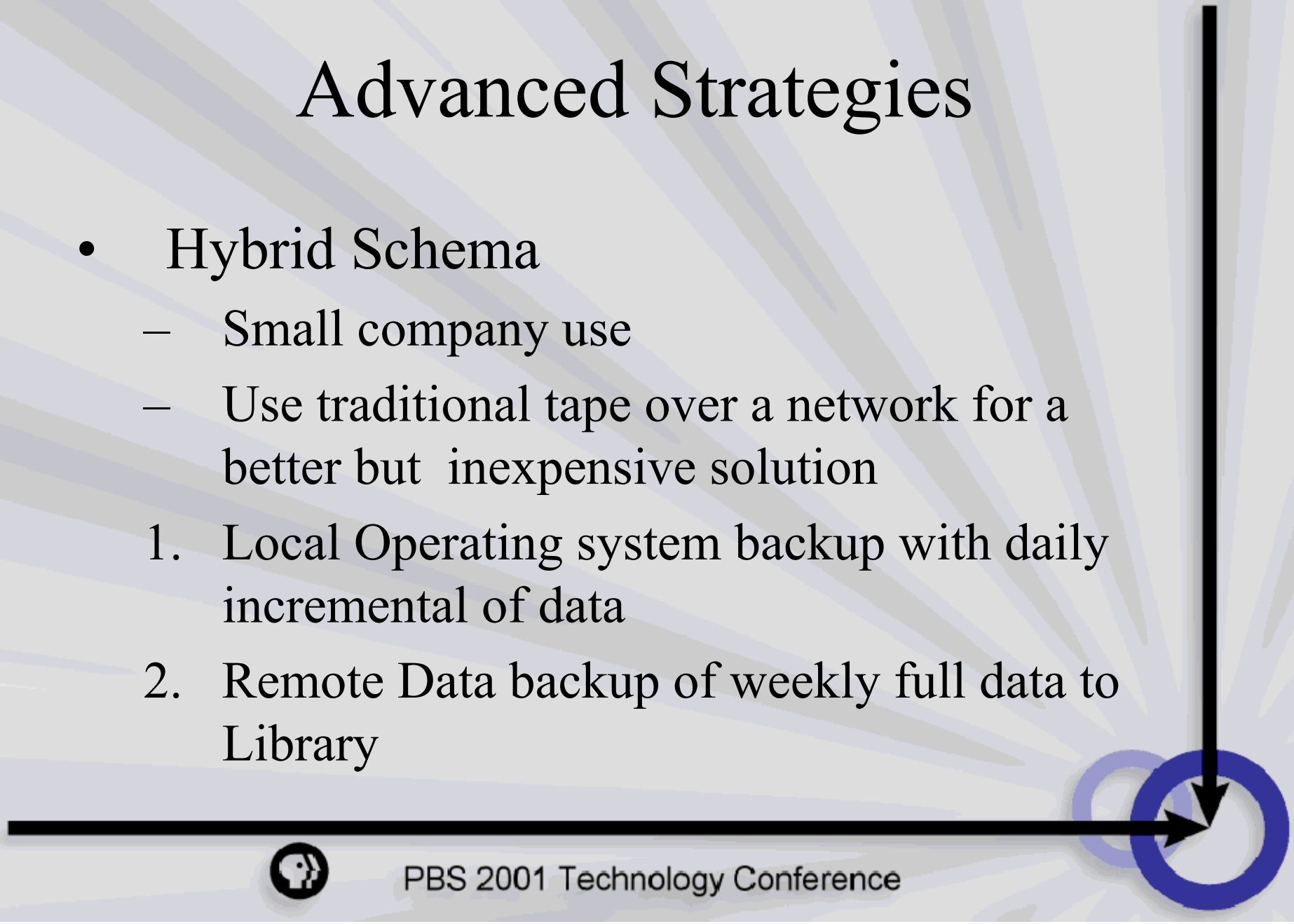
Conventional strategies

- Online vs. Near-line
- Windowed vs. Real-time
- Rotation depth vs. media cost
- Local vs. remote



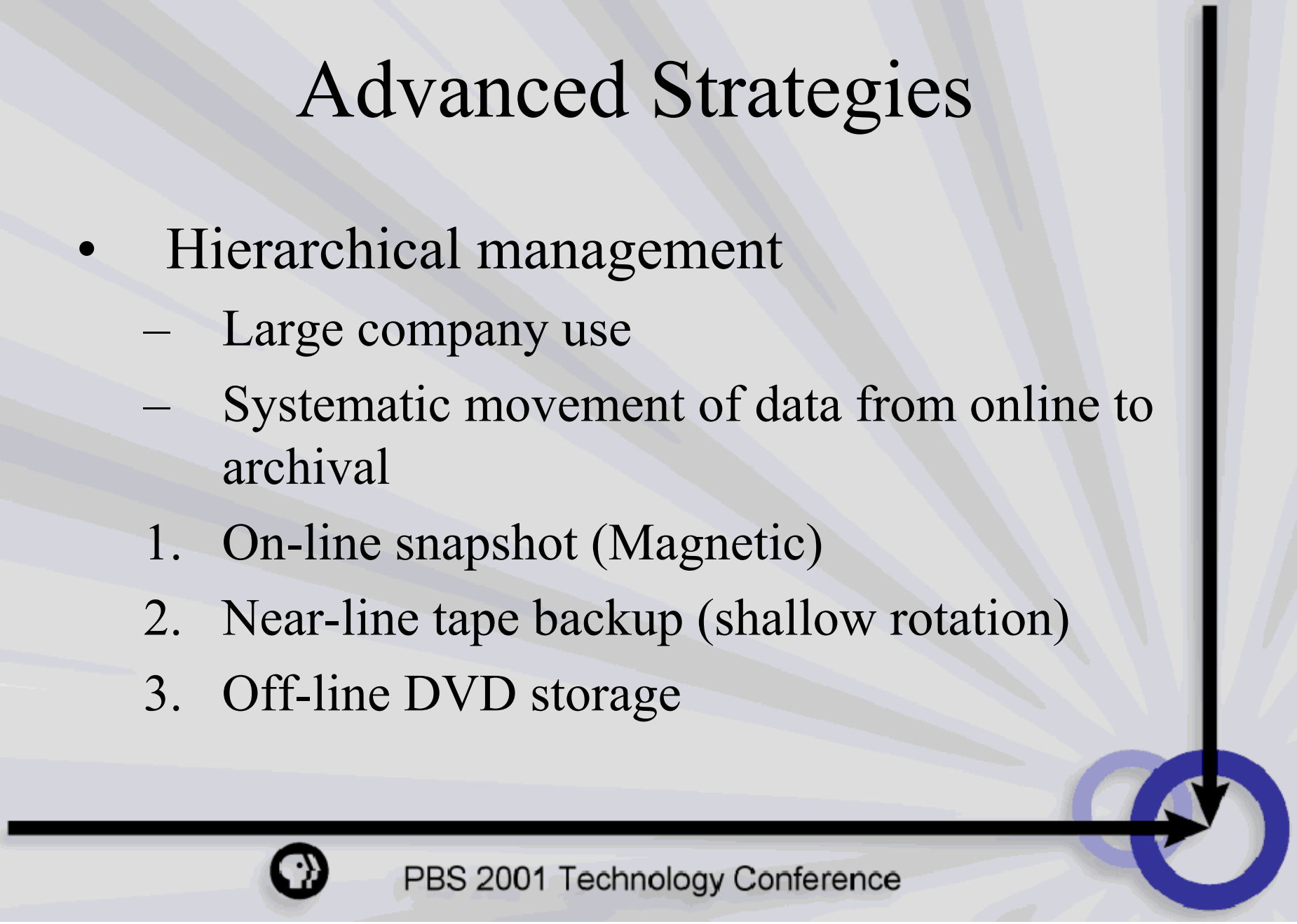
Advanced Strategies

- Hybrid Schema
  - Small company use
  - Use traditional tape over a network for a better but inexpensive solution
  1. Local Operating system backup with daily incremental of data
  2. Remote Data backup of weekly full data to Library



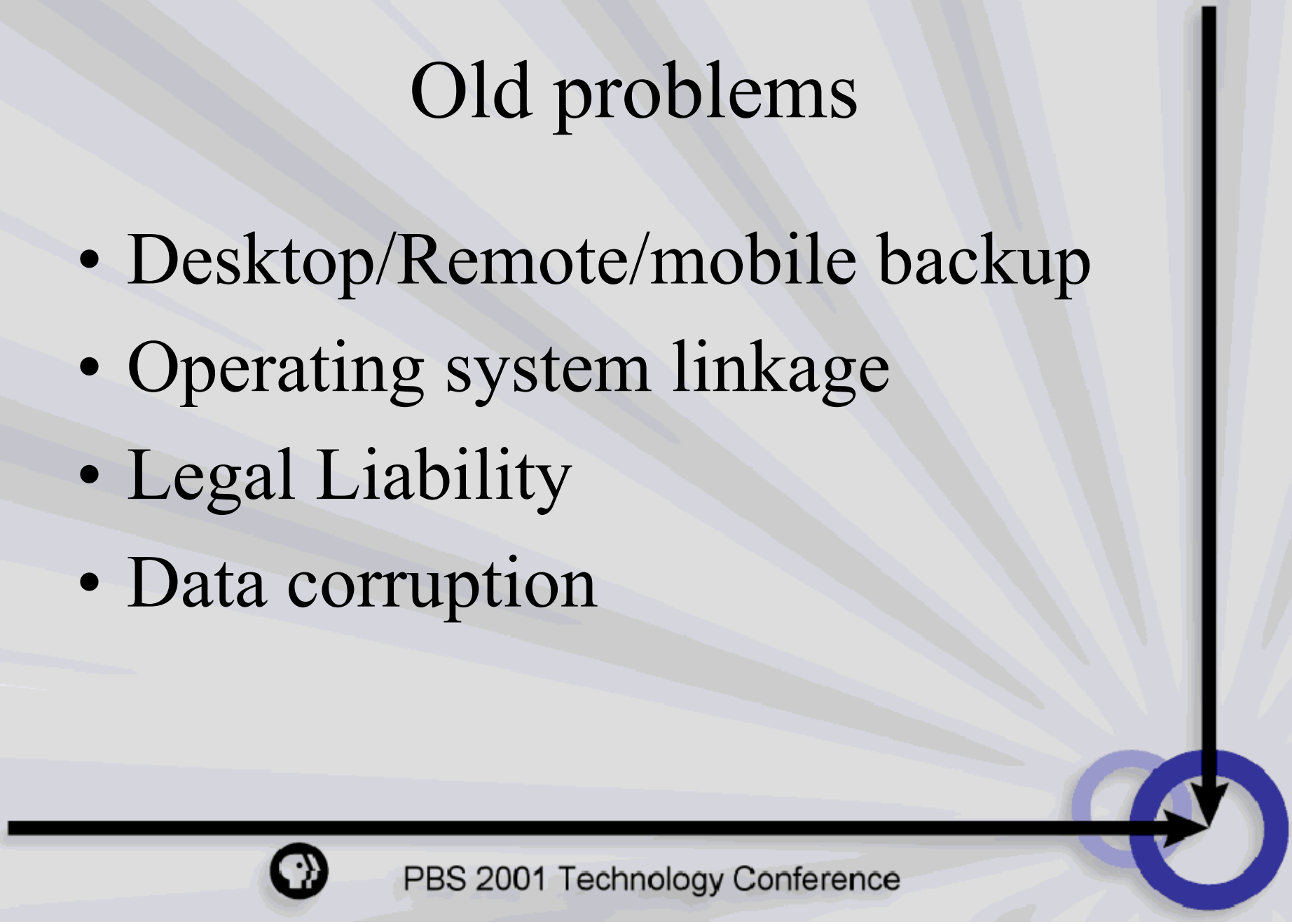
Advanced Strategies

- Hierarchical management
  - Large company use
  - Systematic movement of data from online to archival
  1. On-line snapshot (Magnetic)
  2. Near-line tape backup (shallow rotation)
  3. Off-line DVD storage



Old problems

- Desktop/Remote/mobile backup
- Operating system linkage
- Legal Liability
- Data corruption



New solutions

- ISCSI
- Linear Tape Open
- Hybrid Fiber Channel/Gigabit Ethernet switches
- Outsourcing