900TB BIG DATA ACTIVE ARCHIVE

CDDataHouse
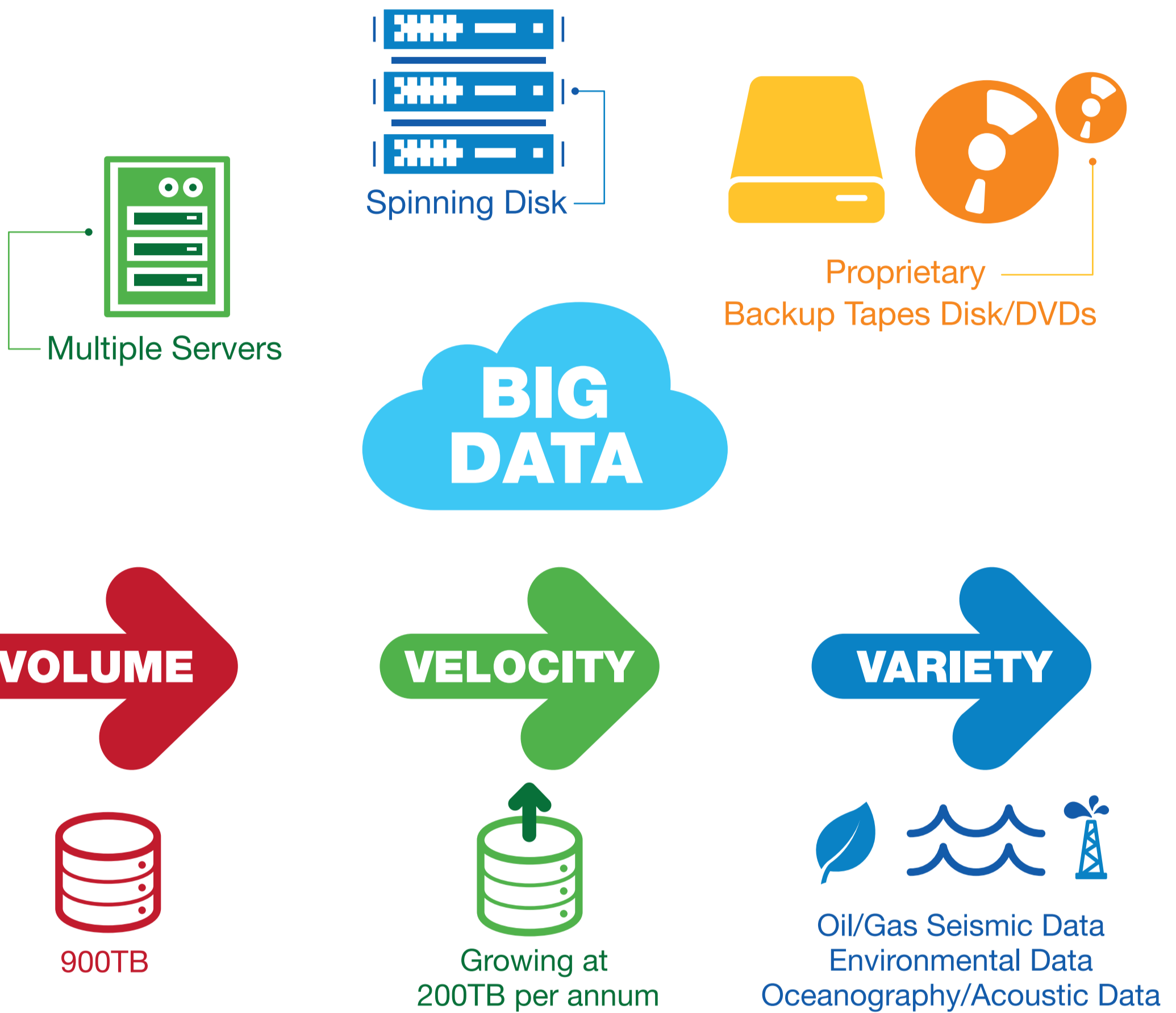STORAGE SYSTEMS & DATA PROTECTION

In 2014 CD DataHouse Limited designed and implemented a 'Big Data' Active Archive for a leading Scientific organisation. The project brought online over 900TB of data for instant and near instant access using disk and tape technology centralising disparate data sources and data media.
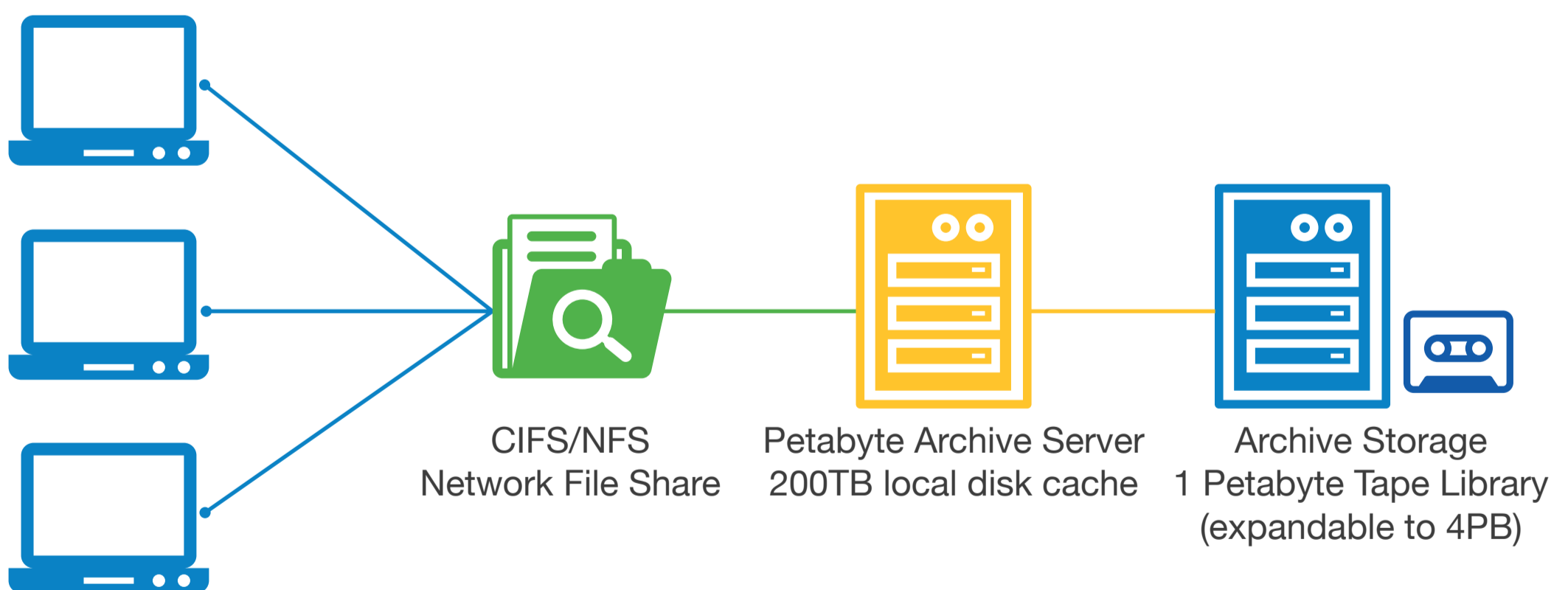
BUSINESS CHALLENGE — Multiple Data Sets 20 Years of Data

Spinning Disk

Proprietary Backup Tapes Disk/DVDs

Multiple Servers

BIG DATA

VOLUME
900TB

VELOCITY
Growing at 200TB per annum

VARIETY
Oil/Gas Seismic Data
Environmental Data
Oceanography/Acoustic Data

BUSINESS NEED: 200TB WORKING SET OF DATA

CD-DataHouse Implementation

CIFS/NFS
Network File Share

Petabyte Archive Server
200TB local disk cache

Archive Storage
1 Petabyte Tape Library
(expandable to 4PB)

RESULTS POST-IMPLEMENTATION

1. Users can access 200TB of most recent or recently accessed data instantly
2. Remaining 700TB of data resides on tape with a 90 second access time
3. Future proof design that scales to 15PB