



Migrating to a new tape library

Peter Edwards | Sr Systems & Data Administrator
August 2019

IMT SCIENTIFIC COMPUTING PLATFORMS, DATA TEAM
www.csiro.au



Why?

- Oracle seems to be exiting the tape and related businesses
 - SL8500 libraries are no longer being manufactured
 - T10000E drive development was cancelled
 - Oracle HSM was placed in “end of sale” status
 - Retrenchments in support staff
 - Just 3 field engineers to cover Victoria
 - Longer delays to have service delivered
 - Replacement drives are mostly refurbished, not new
 - Oracle are more resistant to drive replacement now
 - Spare parts, particularly tape heads, are becoming scarce
- The last point is the crucial one. We don't want to be still using Oracle drives when the world runs out of replacement tape heads.

Where?

CSIRO use of Oracle libraries

- CommVault
 - Multiple locations, being phased out in favour of IBM libraries
- DMF
 - Black Mountain, ACT
 - Dmfact03 (about to be retired)
 - Dmfact04
 - Dmfact05
 - Ruby/DataStore
 - Clayton, Vic
 - Dmfvic02
 - Dmfvic03 (retired)
 - Ruby/DataStore (drives accessed remotely)

How?

The simplest approach

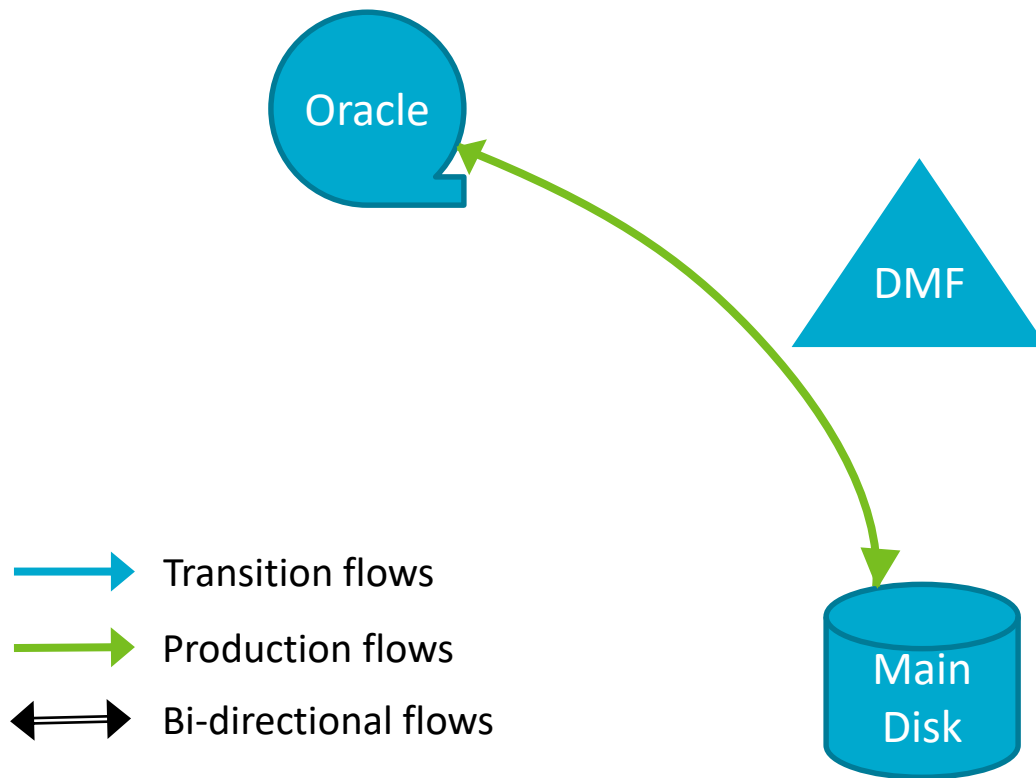
- Provide access to drives in non-Oracle (“Brand X”) libraries
 - Either new libraries, or existing ones which have spare capacity
- Moving data in DMF has to be done on a per-file basis rather than on a per-tape one.

So, for maximum efficiency, the procedure is:

- Configure files migrated in the future to be written to the new library rather than to the Oracle library, two copies being made.
- For each Oracle tape in the VG holding primary copies of already-migrated files:
 - Make a list of files on that tape, ordered by their position.
 - Make two copies of each file to tapes in the new library. If the file is offline, it will be read from tape, but if it’s online the version on the main disk is used as that’s faster. (dmmove by fhandle)

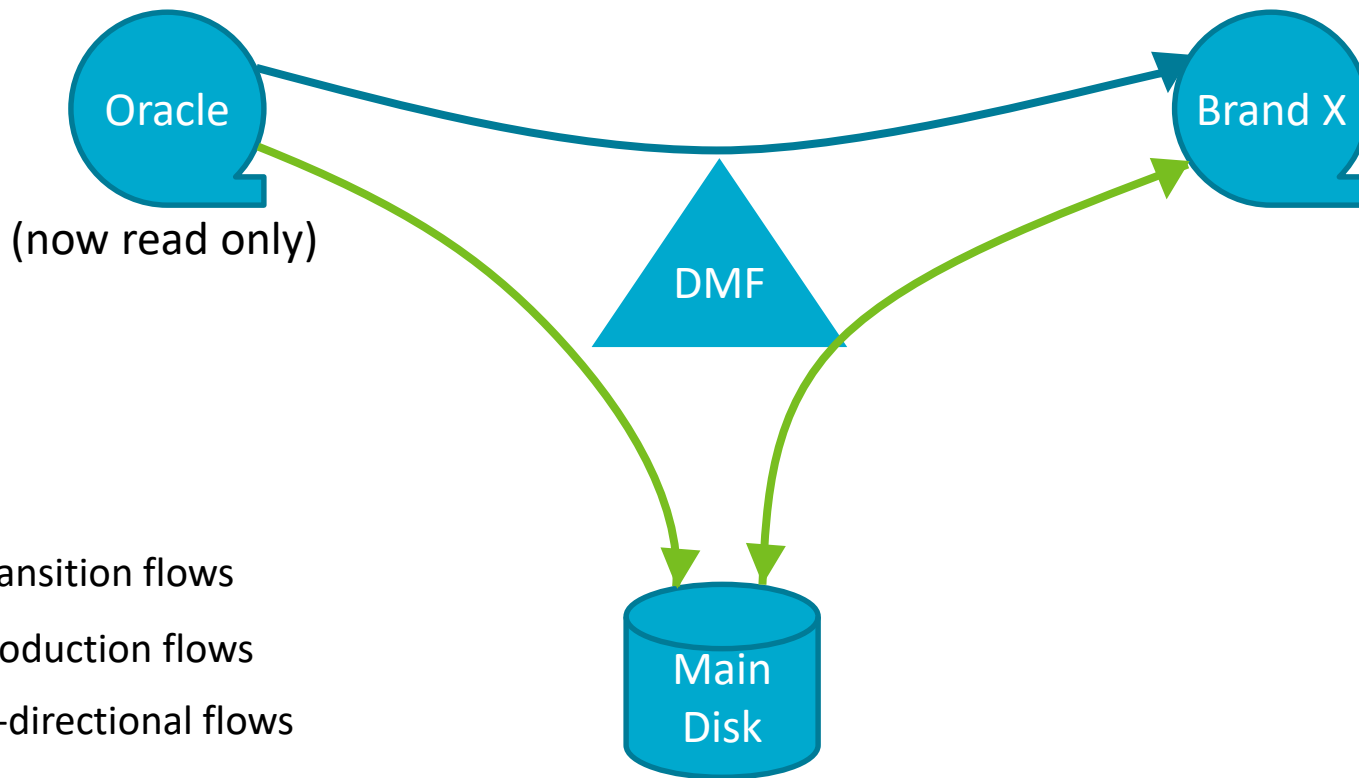
Data flows

Currently, prior to any transitions



Data flows

Simplest approach, during transitions



How?

The simplest approach (cont'd)

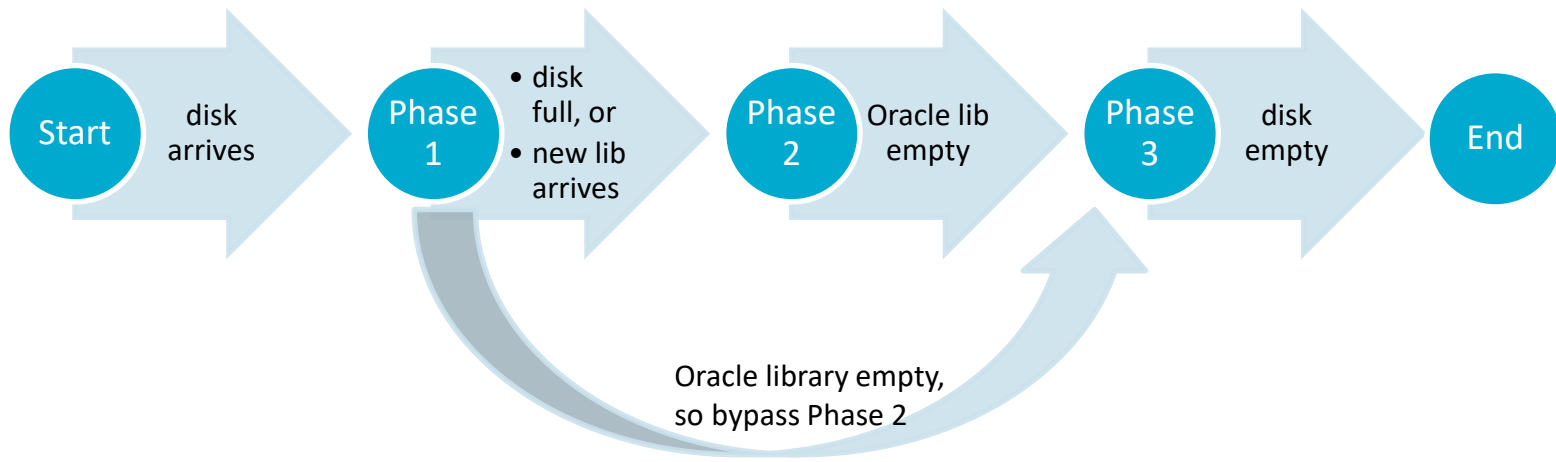
- Pros
 - Standard approach
 - Simple
 - Cheapest
- Cons
 - Purchasing and installing a library takes time, delaying the start of this transition
 - Requires concurrent access to both libraries until the transition is complete

How?

The intended approach (overview)

- Install a large (multi-PB) disk array to be used as a stepping stone for all four DMF servers being transitioned.
- Configure it in DMF as DCM-mode MSPs.
- For each of the four systems, proceed in three phases:
 - Phase 1
 - Start moving primary copies of files from local Oracle tape to this new disk
 - Phase 2
 - After installation of the new “Brand X” libraries, continue moving data, this time from Oracle tapes to tapes in the new libraries (two copies)
 - Phase 3
 - After the Oracle library contains no uncopied data, move data from disk written in Phase 1 to tapes in the new libraries (two copies)

Progression through phases



- Note that the four systems progress through the phases independently, they are not synchronised

How?

The intended approach (detail 1)

- Divide the new disk up into pieces based on the data holdings of the four servers, which are currently approximately:

System	Purpose	Holdings (PB)
dmfact04	BRC archive	1.8
dmfact05	LCM and DAP	2.4
dmfvic02	BRC archive LCM and DAP	0.4 2.0
datastore (on ruby)	HPC archive	13.9
TOTAL		20.5

- Note that holdings are always increasing. For example, the Ruby/DataStore grows by around 60% per annum.
- It doesn't matter too much if the disk size is less than 20.5 PB, it will still give a valuable head-start while waiting for the new libraries.

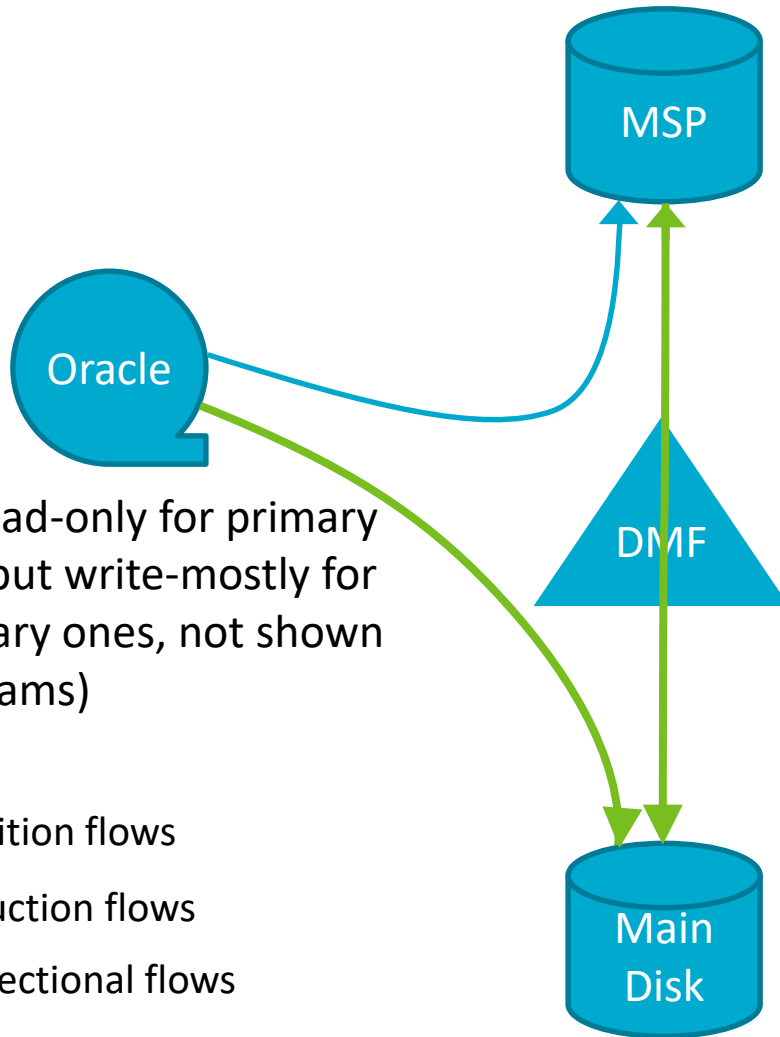
How?

The intended approach (detail 2)

- Phase 1 – prior to Brand X libraries being available
 - Configure files migrated in the future to have primary copies written to the new disk and secondary ones still to the Oracle library
 - Start moving primary copies of files from Oracle tapes to disk as described earlier.
 - Data for online files will be read from the main disk, only offline files need to be read from tape. (dmmove by fhandle)
 - The four systems are independent of each other, so theoretically they can be transitioned in parallel for minimum elapsed time. But the load on the MSP disk array in particular may require us to only transition a smaller number of concurrent instances.

Data flows

Phase 1



(now read-only for primary copies but write-mostly for secondary ones, not shown in diagrams)

- Transition flows
- Production flows
- ↔ Bi-directional flows

How?

The intended approach (detail 3)

- This data movement does not require a dedicated system; production access is not interfered with as DMF can access offline data on either MSP disk or tape, it's not locked away out of reach.
- There is the possibility of competition for resources, tape drives in particular, which will have to be managed carefully. The scripts written for this purpose monitor DMF and tape load and react accordingly.
- This movement of data will continue for up to a year in the case of ruby. Phase 1 for a system will stop when:
 - All the data has been copied (go to Phase 3), or
 - The MSP disk is full (Phase 2), or
 - The new libraries become available (Phase 2)

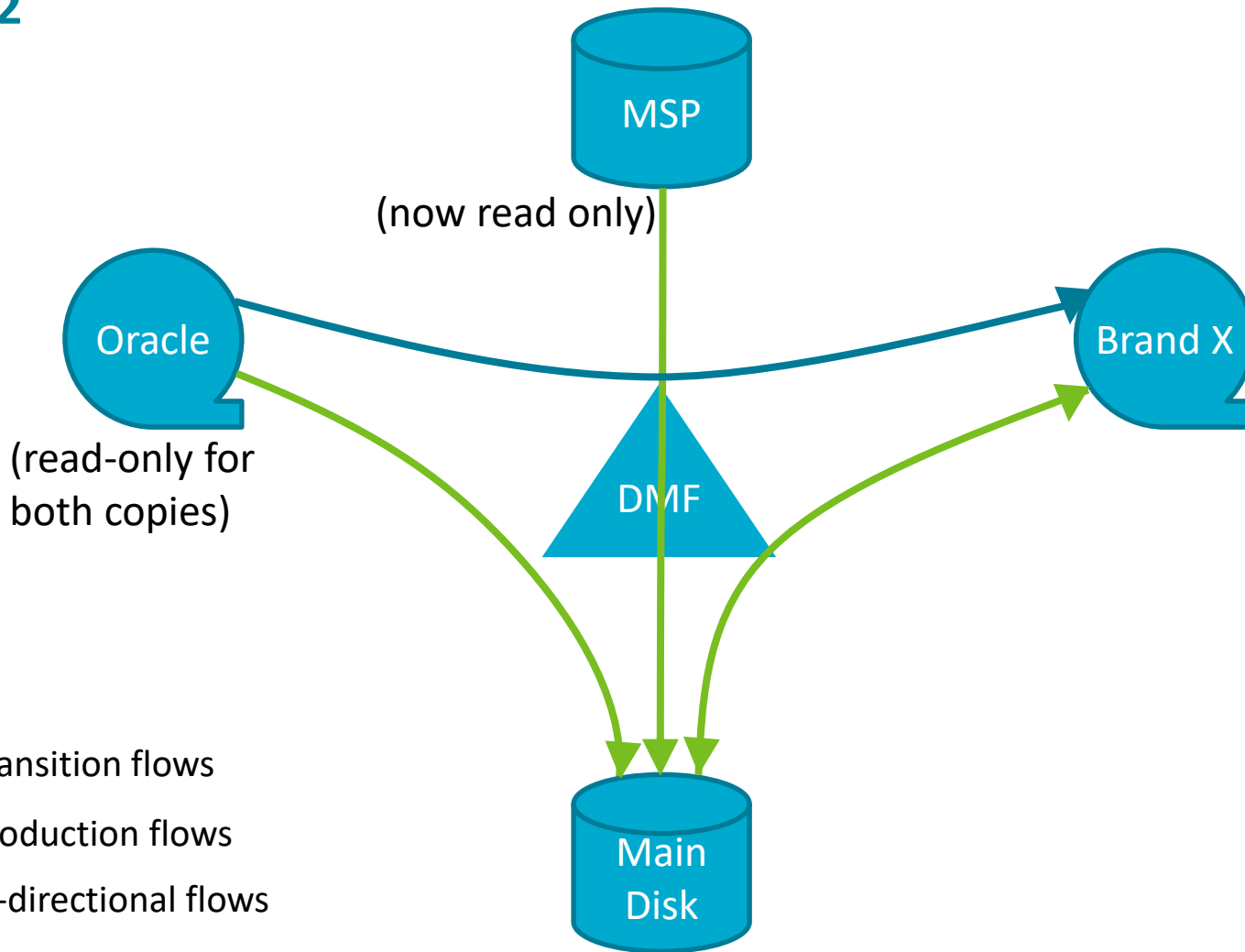
How?

The intended approach (detail 4)

- Phase 2 – when Brand X libraries become available
 - Configure files migrated in the future to be written to the Brand X library rather than to the MSP or Oracle library.
 - All Oracle tapes are now read-only, as is the MSP disk.
 - Data still on Oracle tapes will be copied directly to Brand X tapes as described earlier as being “the simplest case”, not via MSP disk.
 - In both cases, two copies are written to Brand X tape.

Data flows

Phase 2



- Transition flows
- Production flows
- ↔ Bi-directional flows

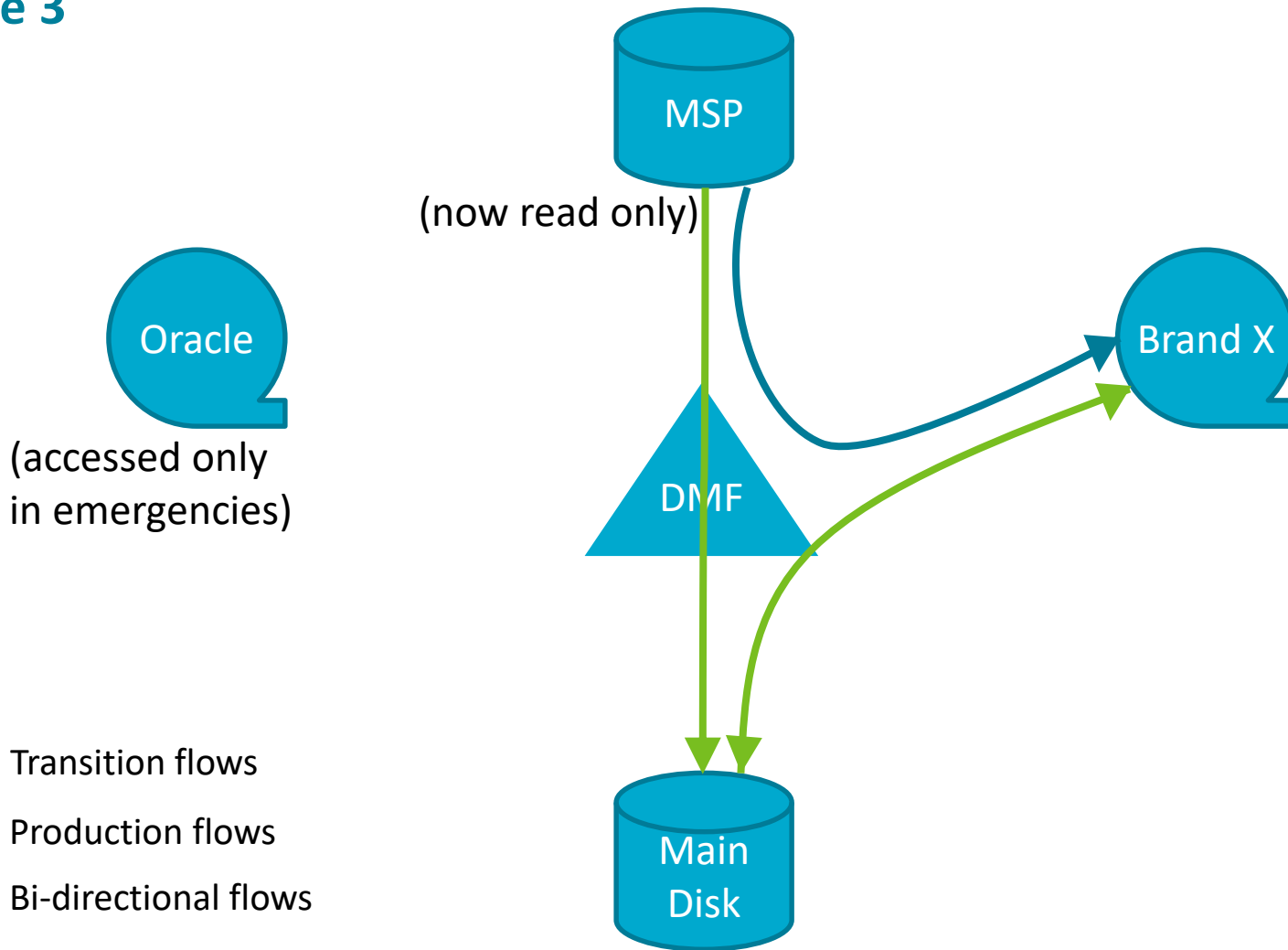
How?

The intended approach (detail 5)

- Phase 3 – when all files on Oracle libraries have been copied
 - The Oracle equipment can then be powered down. It could still be required in an emergency, in order to access secondary copies of files.
 - Configure files migrated in the future to be written to the Brand X library rather than to the MSP or Oracle library
 - The data currently on the MSP disk will then be moved to Brand X tapes in username order.
 - In both cases, two copies are made on Brand X tapes.
 - On completion, the MSP disk can then be reassigned for other purposes.

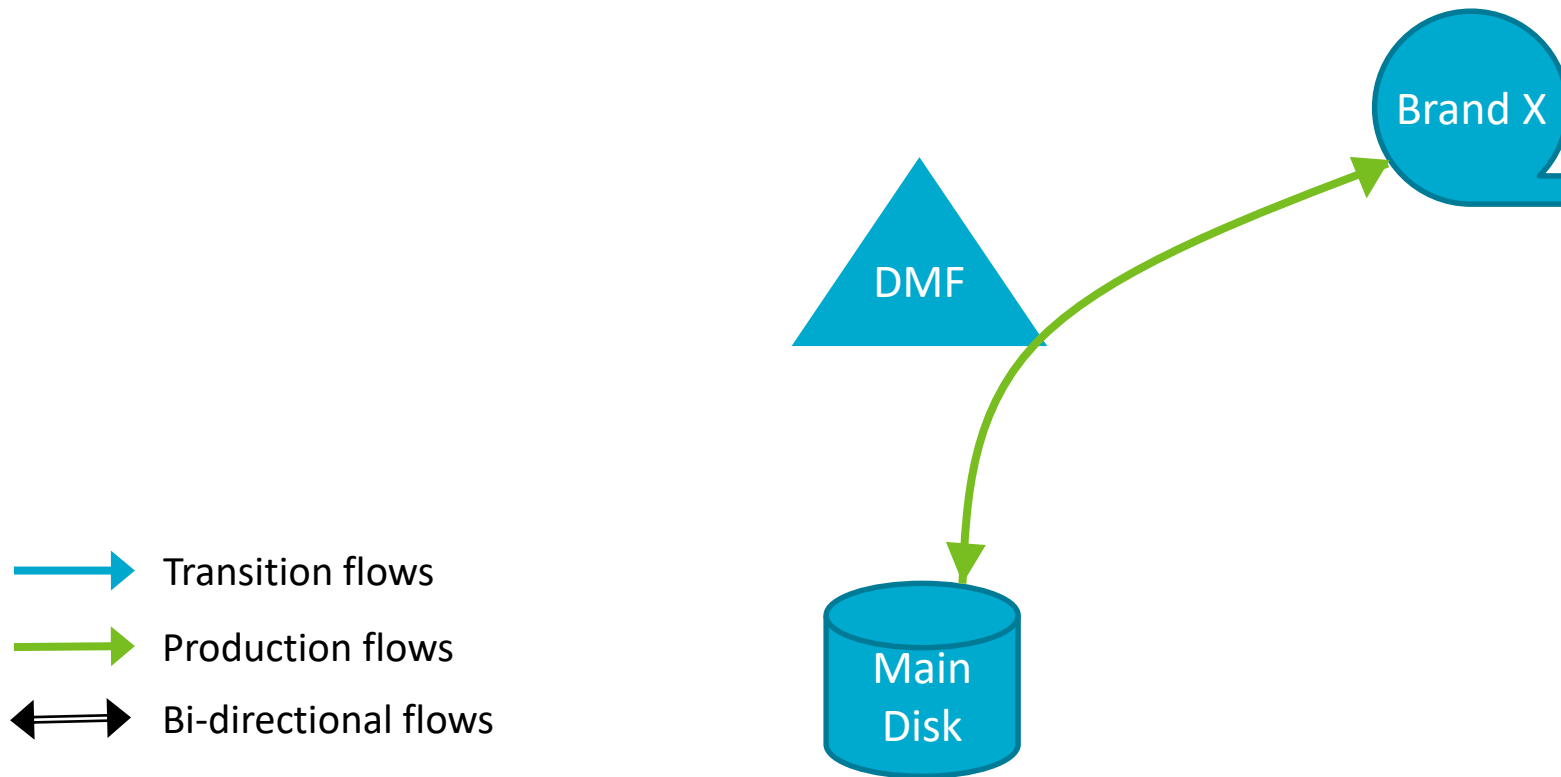
Data flows

Phase 3



Data flows

On completion



Disposal of Oracle equipment

- Full maintenance is needed for each library until all systems using that library have completed Phase 2.
- At that point, cover could be downgraded to time & materials as the library would only be needed to access secondary copies of files in the event of a problem of the MSP disk holding the primary copies. It could be left powered down.
- A library can be deinstalled only after all systems using it complete Phase 3. They may not be ready for removal at the same time as each other.

Optimisations

- Scripts have been written to take advantage of various scenarios. These enhancements are not necessary but will help to reduce tape activity and therefore speed the transitions.
 - The main script which works its way through the collection of tapes can have multiple instances running concurrently on the one system, each instance requiring one Oracle drive. The number of instances can be varied on the fly.
 - This main script is sensitive to tape drives being fully in use by other work and will hold back until that work subsides. It also keeps an eye on the amount of queued activity within DMF and holds off when necessary.
 - In Phase 1, when an offline file is recalled from the Oracle library to the main disk in the course of normal production activity, it is immediately copied to the MSP disk, thereby avoiding possible tape activity later on. This process is automatically suspended if there is a lot of other DMF activity going on.
 - A cron job copies files which happen to be online initially or have been overlooked by the previous script.

Details

- Second copies of files

For newly migrated files,

- In Phase 1, the second copies continue to be directed to Oracle tape.
- In Phases 2 & 3, they will go to the Brand X tapes instead.
- At no stage do they go to the MSP – the first copies are sufficient there

- Notes

- We may wish to benchmark the new hardware before commencing each phase.
- The duration of the phases will differ between the four systems and will depend on how much impact on production work is acceptable.
- DMF and/or OS upgrades may be needed on some systems, to gain support for some choices of Brand X hardware.

Thank you

IMT SCP Data Team

Peter Edwards

Sr Systems & Data Administrator

t +61 3 9545 2377

e peter.edwards@csiro.au

w www.csiro.au/lorem

IMT SCIENTIFIC COMPUTING PLATFORMS, DATA TEAM

www.csiro.au

