

Sheet1

Worksheet 2 Copyright Damian Brasher March 25 2008 www.diap.org.uk

| Cron jobs | Node | | | |
|------------------|---------|-----------------------|-----------------------|---------|
| | Daily | A | B | C |
| t= 0 | Special | aFull-bFull (1st Mth) | bFull-cFull (2nd Mth) | |
| Start 00:00 | | 2 ad0-bd0 t=0-60m | bd0-cd0 | cd0-ad1 |
| End 07:30 | | 3 ad1-bd1 | bd1-cd1 | cd1-ad2 |
| | | 4 ad2-bd2 | bd2-cd2 | cd2-ad3 |
| t = 0 (00:00) | | 5 ad3-bd3 | bd3-cd3 | cd3-ad4 |
| t = 0-30 (00:30) | | 6 ad4-bd4 | bd4-cd4 | cd4-ad5 |
| | | 7 ad5-bd5 | bd5-cd5 | cd5-ad6 |
| | | 8 ad6-bd6 | bd6-cd6 | cd6-ad7 |
| | | 9 ad7-bd7 | bd7-cd7 | cd7-ad8 |
| | | 10 ad8-bd8 | bd8-cd8 t=0m | |

Two entry points, aFull beginning of month and ad0 for the remaining days
 Assuming entry points are filled during the day before the cycle begins at night
 29 cron jobs split between 3 nodes

Max ad0 = LMB ((a-b or b-c or c -d) x 7 hrs) - (Ave. Diff x 9)**
Max ad0 must be greater than 2 x (Ave. Diff x 9)
Max aFull = Max ad0

** This copes with the situation when a FULL is copied at the same time the system is saturated with Diffs.

LBM = Lowest Maximum Bandwidth

NB: actual max transfer will vary test so transfers are recommended for accuracy

Min DIAP space required on each node =
 ((max ad0 + (9 x Ave. Diff)) – max rsync logfile size

This varies according to the size of Differential – see example.

| | | | | |
|-----------------|---|------------------|------------------------|--------------------------|
| Example system: | LMB x 7 hrs | Ave. Diff | Max ad0 (aFull) | Min DIAP Dir size |
| Max ad0 = | 128KByte/S | 1G | 24G – 9G = 15G | 15G + 9G = 24G |
| | 2 x (Ave. Diff x 9) | | | |
| | (18G less than ad0 so system will not fail) | | | |

| Slots (Dirs) | A | B | C |
|-----------------|-------|-------|-------|
| | aFull | bFull | cFull |
| | ad0 | bd0 | cd0 |
| | ad1 | bd1 | cd1 |
| | ad2 | bd2 | cd2 |
| | ad3 | bd3 | cd3 |
| | ad4 | bd4 | cd4 |
| | ad5 | bd5 | cd5 |
| | ad6 | bd6 | cd6 |
| | ad7 | bd7 | cd7 |
| | ad8 | bd8 | cd8 |

The system reduces single point of failure by creating a single FULL copy on each node at the beginning of the month the storing the differentials in a distributed manner.
Use a program such as Bacula set to use a monthly FULL – differential.

If a copy fails then the system will retry the next day but you loose the day of failure.

Using rsync logs you can trace / track the successful copies.

The copies are staggered so that each rsync copy list is made before new files are put into each directory by a few minutes.
I.e. bd9-cd9 starts before ad9-bd9 and ad0-bd0 is last.

Redundancy is split across three nodes, no duplicate days apart from the first FULL copy.

DIAP pool is equivalent to 30 tapes every month but stored at three locations. 10 days every three days, at each.

You are advised to have some knowledge of the average differential size.