**IBD distribution Write Cache**

**REVISION HISTORY**

|  |  |  |  |
| --- | --- | --- | --- |
| **DATE** | **AUTHOR** | **REVISION** | **HISTORY LOG** |
| 2/9/16  3/15/16 | Guanheng Liu  Guanheng Liu | A | Created  Update RS module |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

**Table of Contents**

[1 Overview 3](#_Toc445807792)

[1.1 Requirement 3](#_Toc445807793)

[1.2 Scopes: 3](#_Toc445807794)

[1.3 Structure 4](#_Toc445807795)

[1.3.1 Resource level 4](#_Toc445807796)

[1.3.2 Translation level 6](#_Toc445807797)

[1.4 Strategy 6](#_Toc445807798)

[1.4.1 Error handle 6](#_Toc445807799)

# Overview

This document covers design of distribution read cache (DWC) of IBD.

## Requirement

* Availability: DWC can provide Integrity and Availability to data in place of RAID1
* Capacity: ibdserver is able to to reserve larger Write Cache without occupying too much resource of local machine.
* Scalability: DWC node can be added or removed any time,

## Scopes:

Ibdserver will store write request data into a pair of remote DWC node instead of local MWC file.

…….

D

C

B

A

Ibdserver will always randomly select one pair DWC nodes when a new write request come, and the DWC nodes should have enough free space to contain the write request. A printer aimed to next DWC node will be stored in header when ibdserver decided to switch to a new pair of DWC nodes.

Request 2, 300M

Request 1, 256M

D

C

B

A

…….

R2,300M

R1,256M

R1,256M

R2,300M

## Structure

### Resource level

DWC will use Resource Module (RS) to management the remote SSD resource. RS Module include Resource Agent (RA), Resource Manager (RSM) and Resource (RS) these three parts.

RA: RA located at local base node, it will require resource unit from remote RSM and decide when and how to switch another remote RSM.

RSM: RSM located at every remote write cache node, it will manage all the SSD resource on the remote node, and provide a chunk of resource unit back to RA as RA required.

RS: RS is in charge of maintaining SSD resource.

RA

RA-RSM

connection

RA-RSM

connection

RSM

RSM

…

RS

RS

RS

RS

RS

RS

…

…

RA-RSM connection:

RA request RS unit space from RSM

RSM return RS unit information to RA

Update the pointer at the tail of last RS unit to new RS unit

RA start write data to RS unit

RA finish write data to RS unit

RS unit: RSM will return RS unit information to RA which contain UUID of RS, index of RS, offset of unit and index of unit.

index of unit

offset of unit

Index of RS

UUID of RS

Whenever RA found an available RSM or RSM restart, RSM will send RS unit information to RA. RA will update the RS unit related mapping after receive new RS unit information. When RA write data to remote RA unit, the write request header will only contain the index of RS and index of unit.

### Translation level

## Strategy

### Error handle

Whenever there is a write error happen on DWC node, ibdserver will stop writing the remaining data, and select a new pair of DWC nodes to write the left data.

Point to Node C

Node A

Error

finished

R1, 300M

200M finished, 100M left

Node B

Stop

ibdserver

Node C

Left

redirection 100M

Node D

Left