中国移动
China Mobile

OpenStack

How to run a Public Cloud on OpenStack

Nai Ye，Fenghua Yao，Pengju Jiao

yenai@cmss.chinamobile.com

yaofenghua@cmss.chinamobile.com

jiaopengju@cmss.chinamobile.com

中国移动 China Mobile

The Open Infrastructure Summit

China Mobile's public cloud is building on OpenStack.

We started this work in early 2015.

8000+

80W+vCPU

20PB+

The Most Innovative Companies In The World Move Faster With OpenStack

中国移动 China Mobile | TARGET | PROGRESSIVE | CATHAY PACIFIC | PayPal

China Mobile's telecom network has more than 800 million subscribers and 3 million base stations.

In the growing of business, We have faced with many challenges in architecture, high concurrency and super-large scale region.

What problems have we met? Today, we will show three of them:

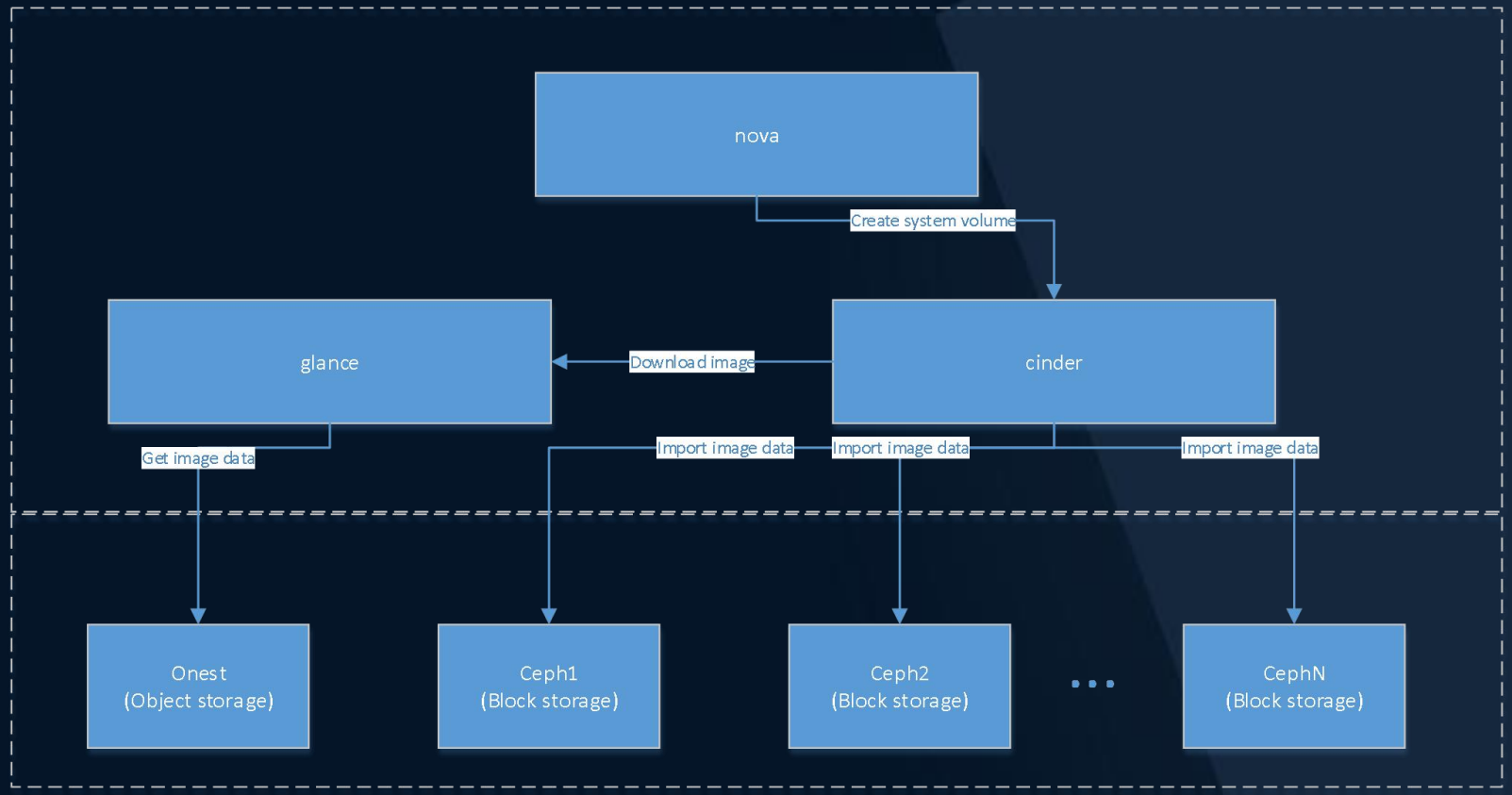
- How to improve distributed system disk creation performance
- How to improve volume backup
- How to improve the schedule performance of scheduler

Performance Enhancement in System Disk Creation

Performance enhancement in system disk creation

Background:

OpenStack has significant concurrency performance bottlenecks when using distributed storage as a system disk, cannot meet the performance needs of public clouds.

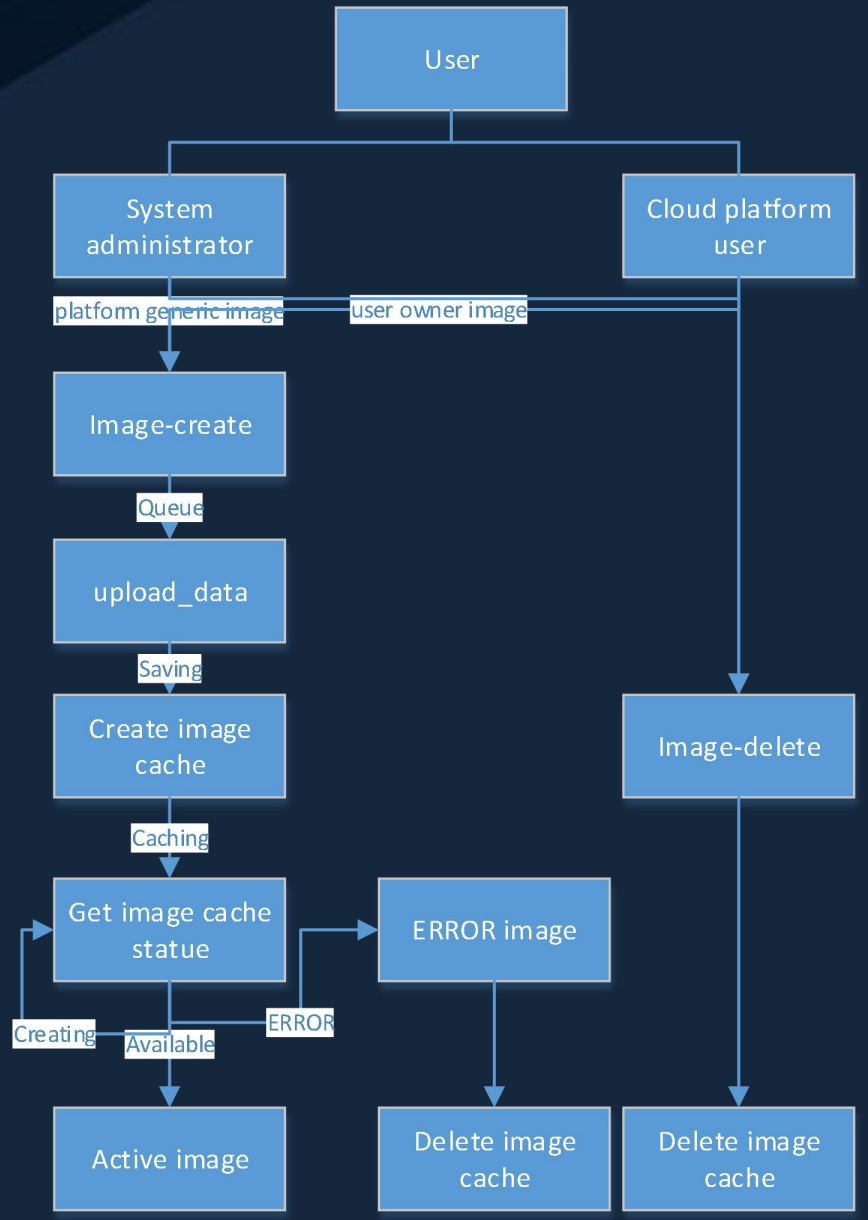


- 1.Where is the image cache mechanism?
- 2. Is there a risk that a large amount of image data downloaded to the cinder temporary directory?
- 3. Whether glance has a performance bottleneck when concurrently download images?
- 4. The impact of large image data transmission on system volume creation speed?

Performance enhancement in system disk creation

Establish a flexible image cache mechanism:

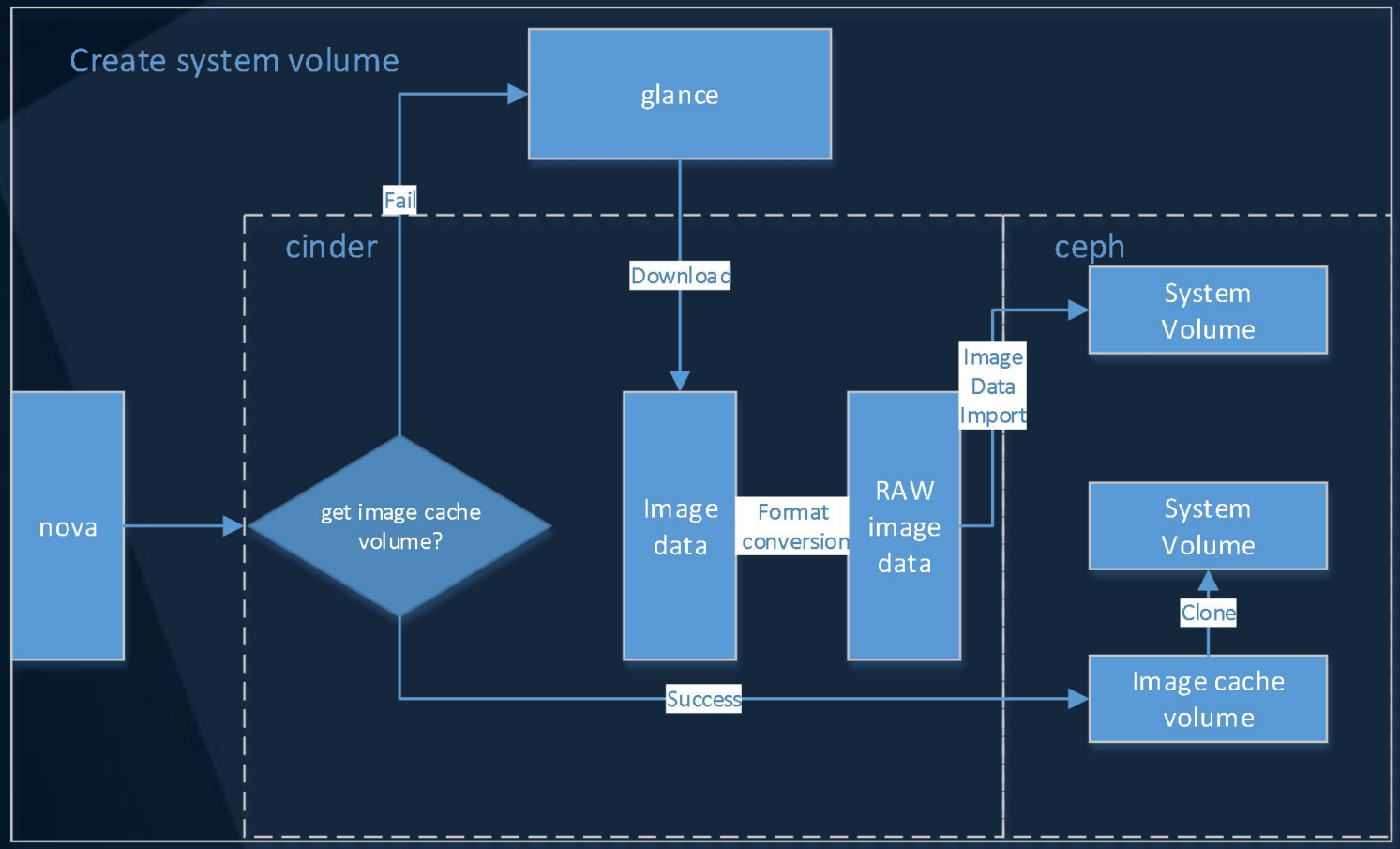
- Set the default storage type to be cached through the configuration file
- Whether the image is cached according to user selection, avoiding waste of storage space
- Cache data is cleared along with the image, conducive to system maintenance
- For uploading new images and imaging based on virtual machines, apply together



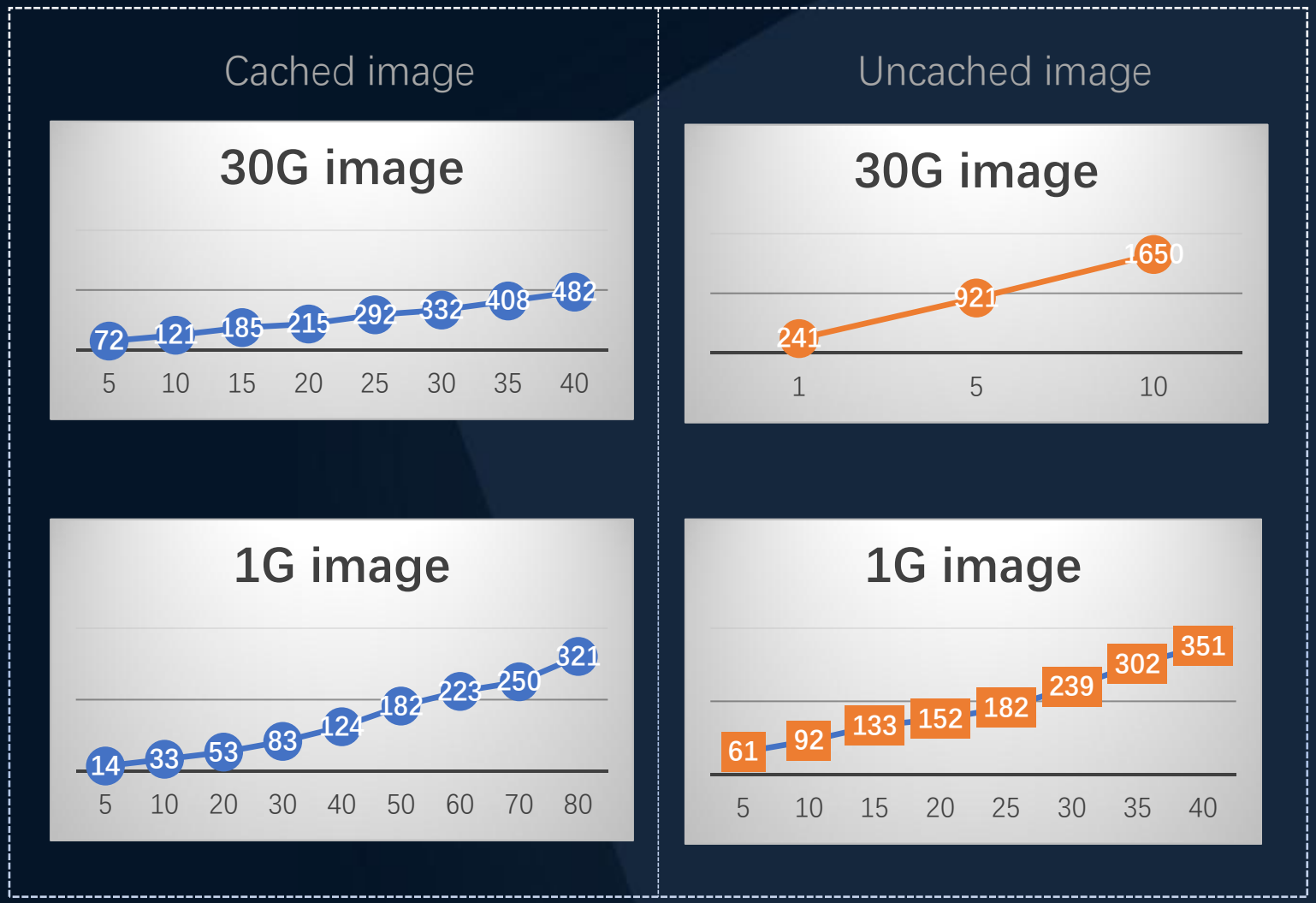
Performance enhancement in system disk creation

Establish two sets of creation mechanisms:

- Cached image:
  - Improve the number of simultaneous creations
  - Clone a new system disk directly using cached data
  - Reduce image data transfer and improve creation performance
- uncached image:
  - Limit the number of simultaneous creations



Performance enhancement in system disk creation



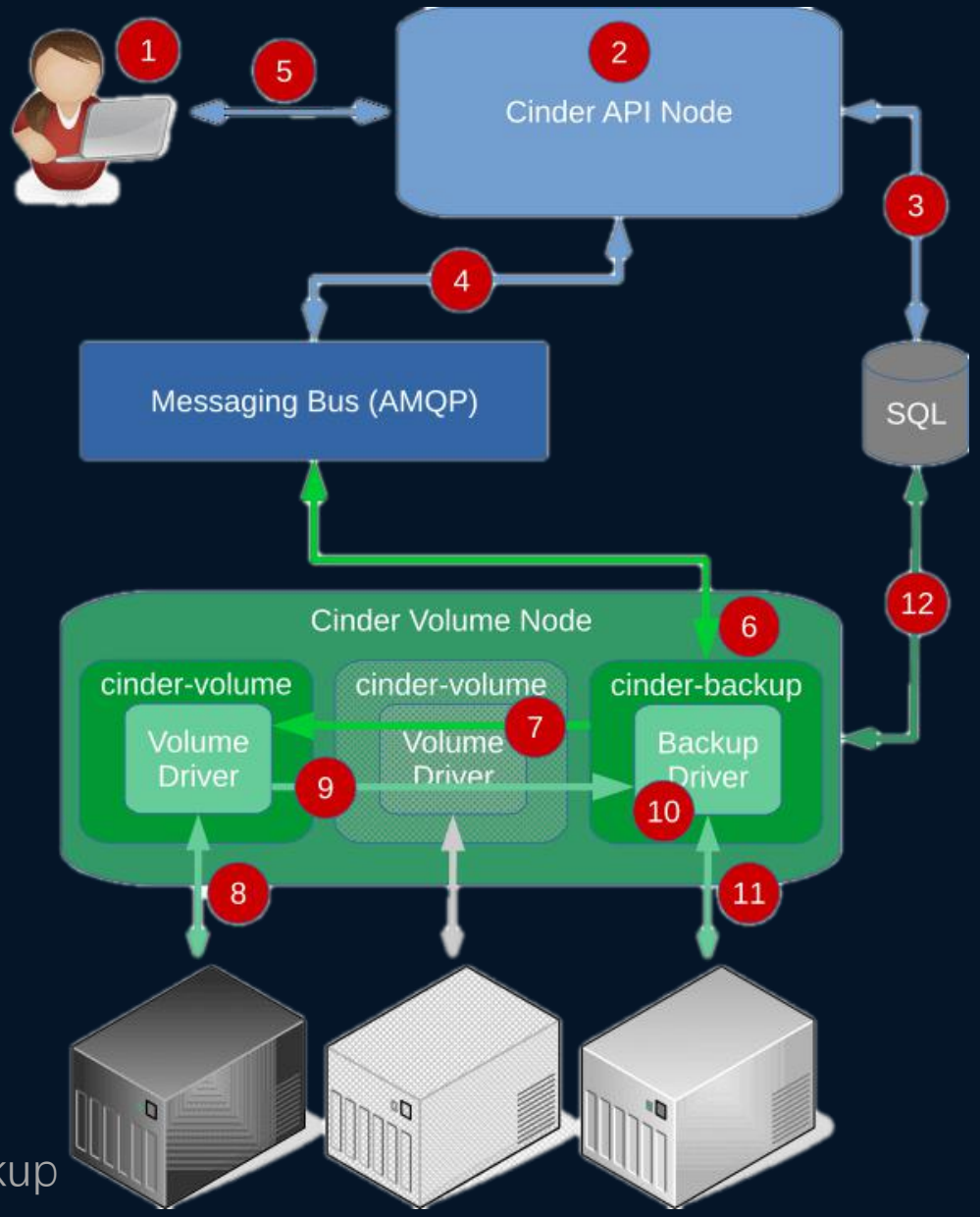
Horizontal axis: number of concurrent
Vertical axis: completion time

Performance Enhancement in Volume Backup

Performance enhancement in volume backup

Backup workflow:

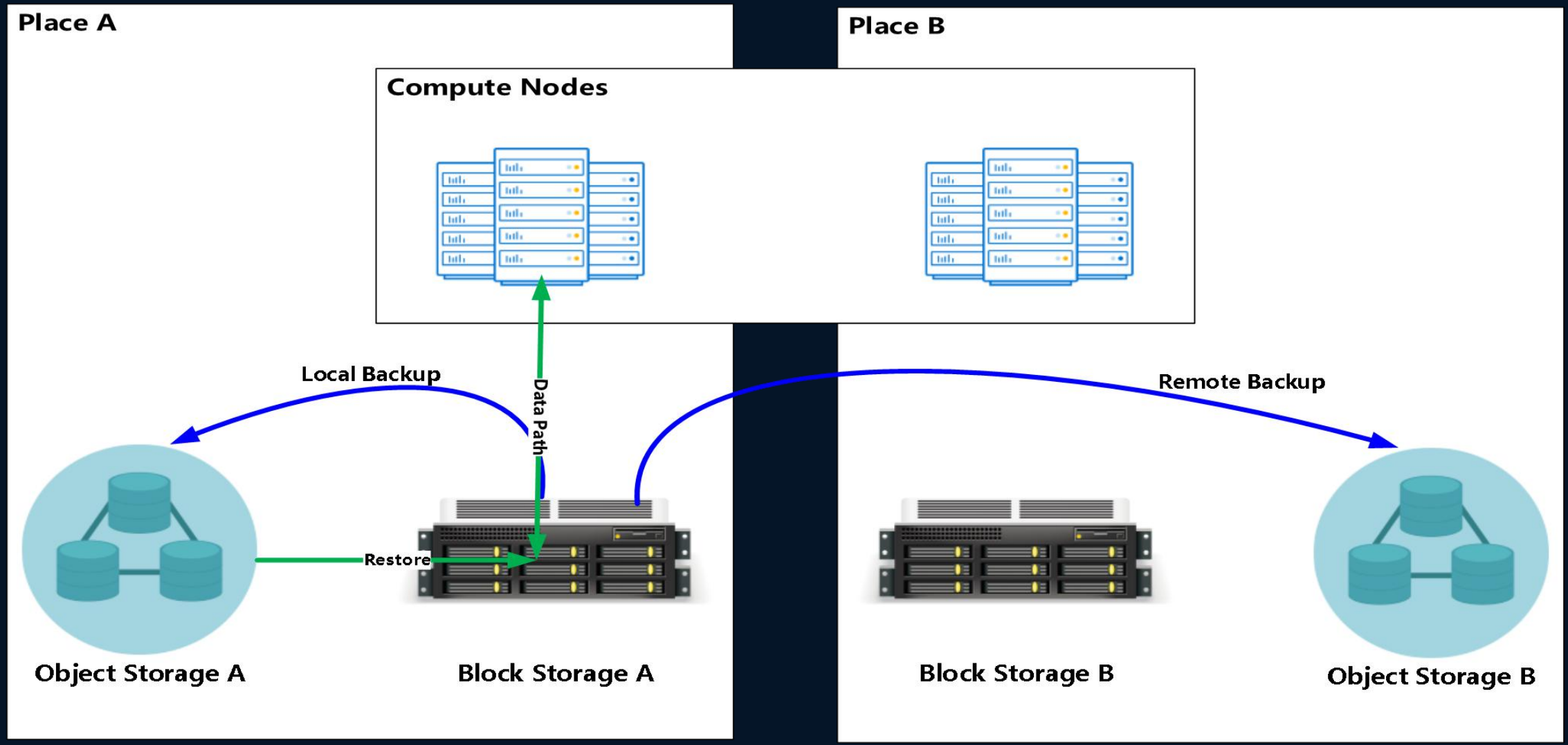
1. Client backup request
2. Run all checks, reserve quota and look for latest backup
3. Change volume status and from the Cinder API Node
   and create cinder backup BD info
4. **RPC cast** to Cinder Volume Node to do the backup
5. Return backup information to client
6. Check that volume and backup are in correct status
7. Call source volume's driver to start the backup
8. Attach source volume from volume back-end for reading
9. Call backup driver to do the backup with attached volume
10. **Detect changed chunks since last backup**
11. **Save backup extents in back-end storage**
    Read-> SHA-256 -> Compress -> Upload
12. Change volume and backup status in DB



From: https://gorka.eguileor.com/inside-cinders-incremental-backup

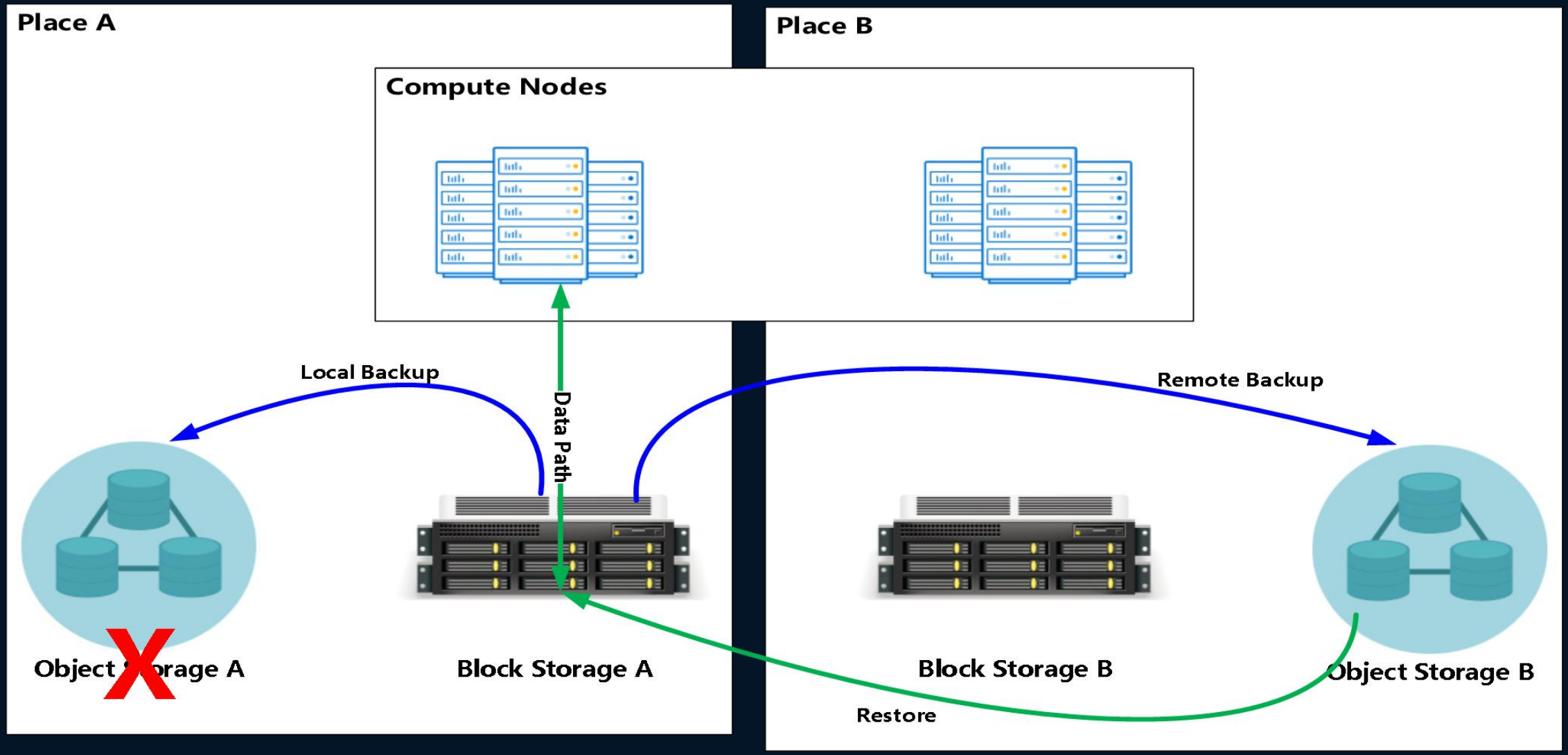
Performance enhancement in volume backup

1. Specify any backup type as you want.
What the consumer wants:



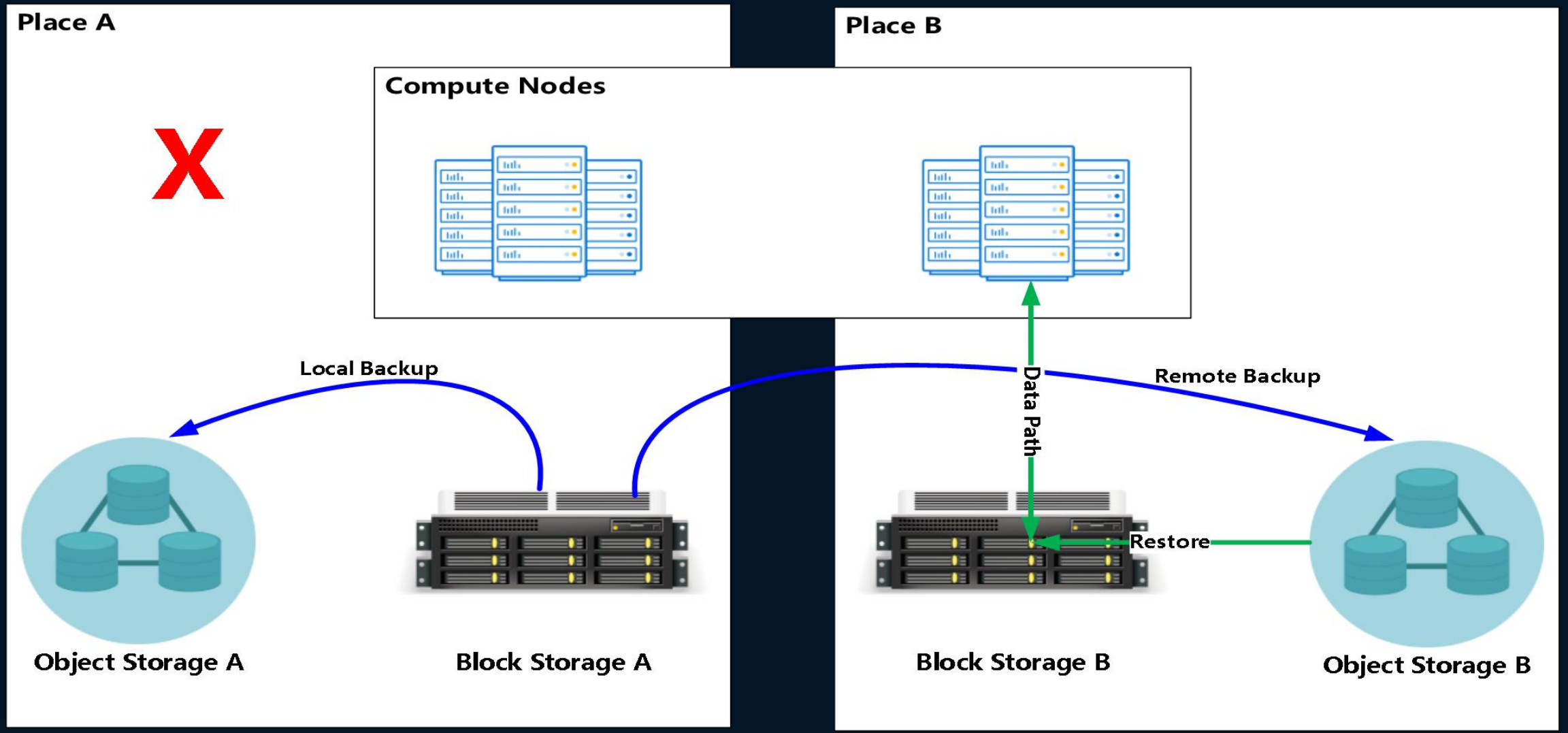
Performance enhancement in volume backup

1. Specify any backup type as you want.
What the consumer wants:



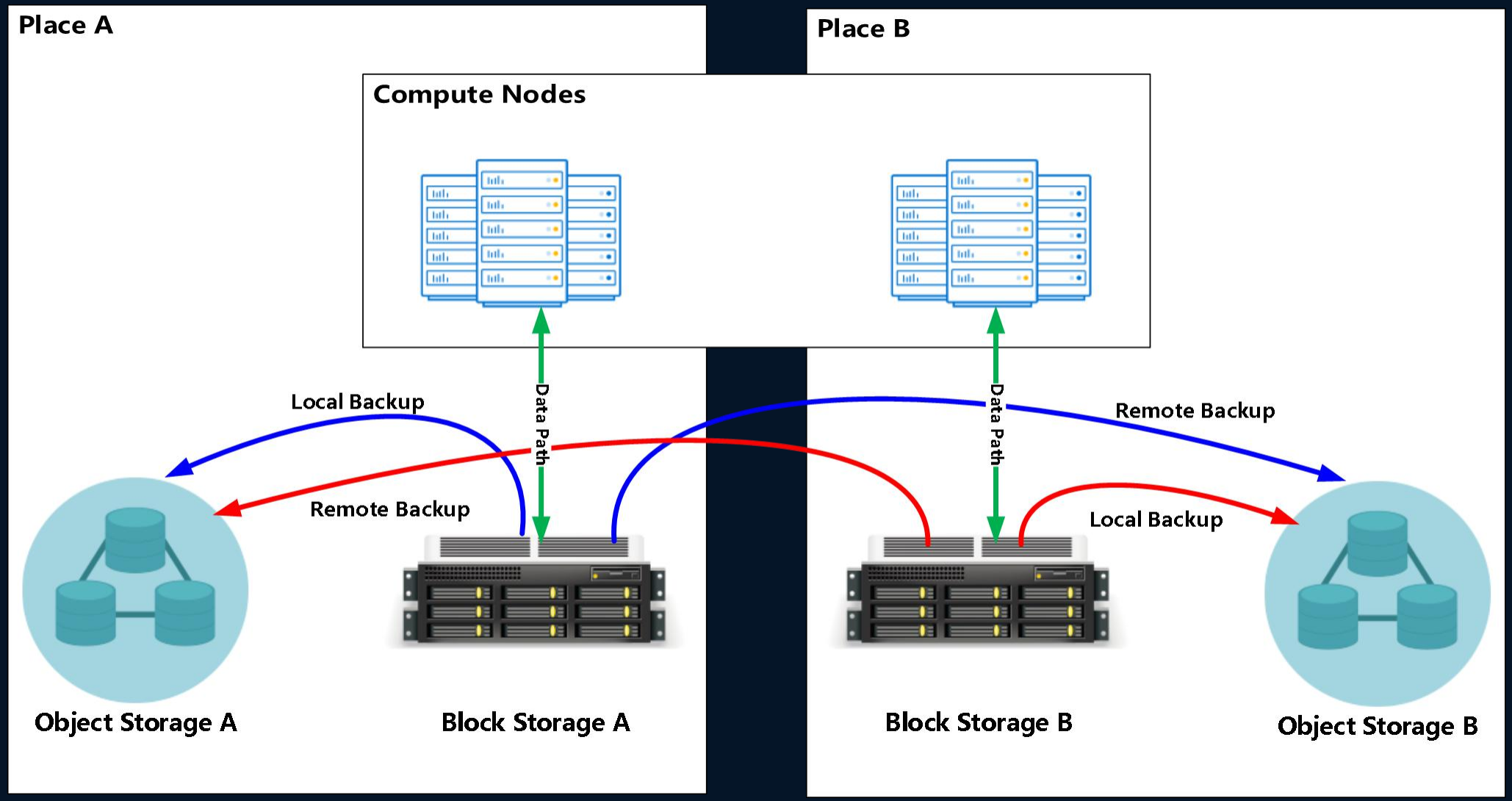
Performance enhancement in volume backup

1. Specify any backup type as you want.
What the consumer wants:

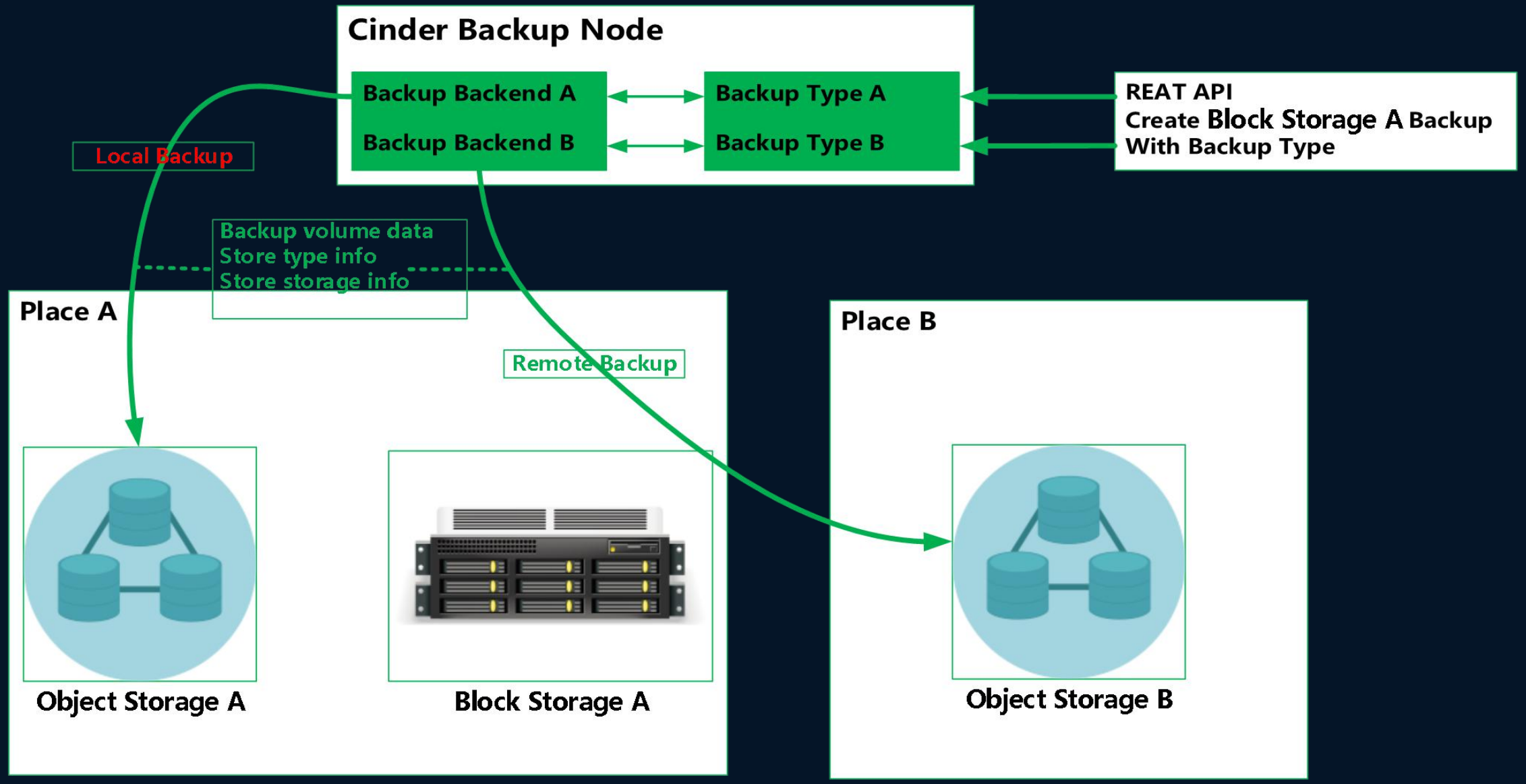


Performance enhancement in volume backup

1. Specify any backup type as you want.
One OpenStack can just only bacpup to one object storage!
But what the customer wants is just 'Three centers in the two places':

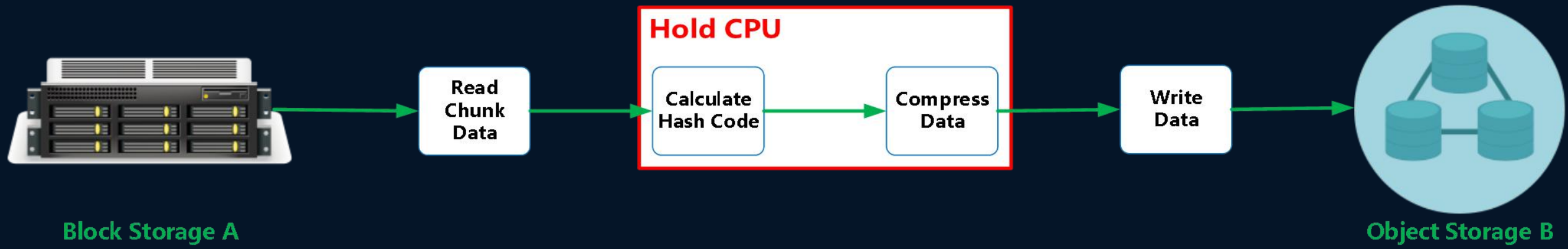


Performance enhancement in volume backup



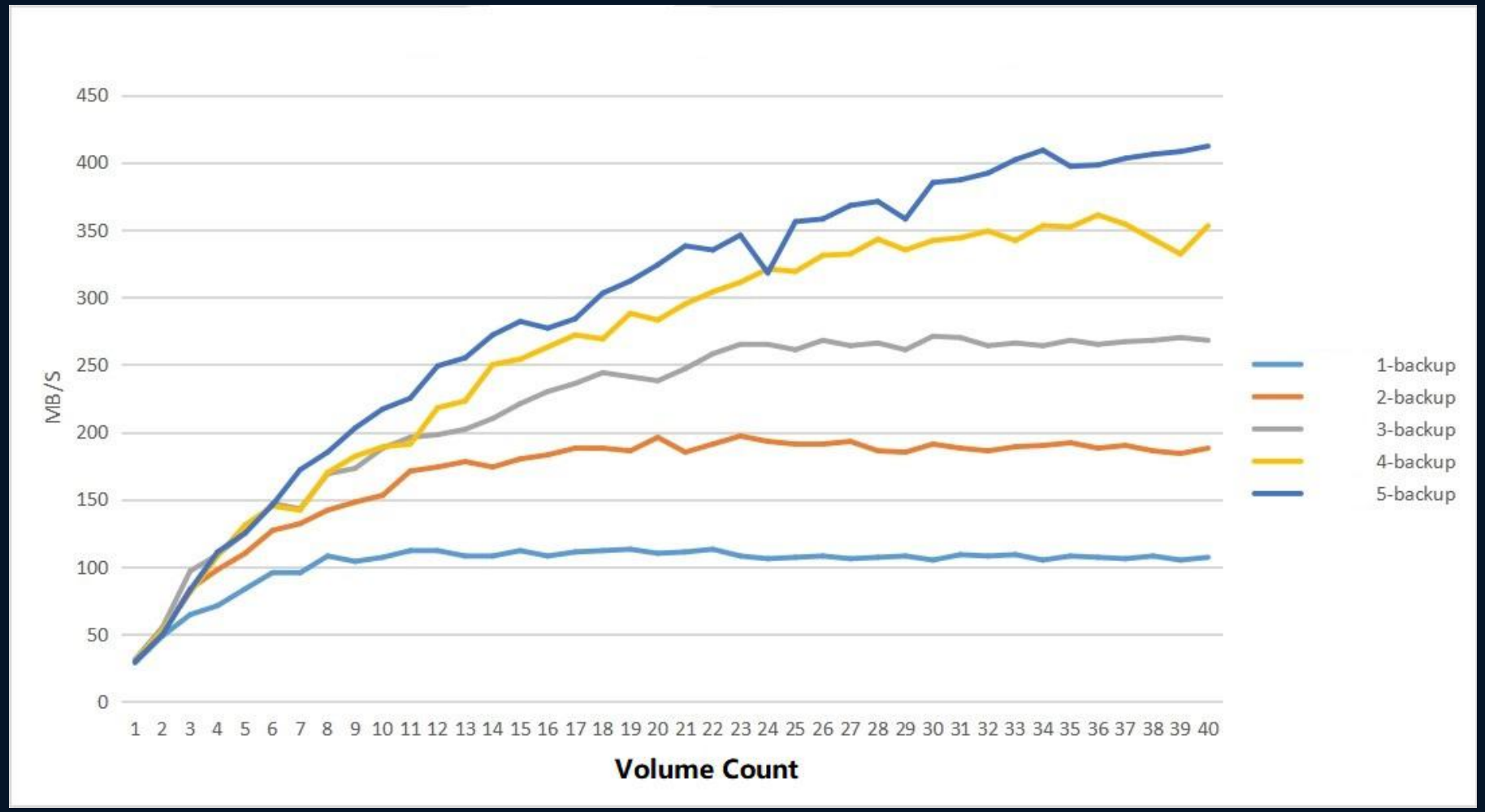
Performance enhancement in volume backup

2. Use multiple processes to do backup.
Why we need multiple processes?



Performance enhancement in volume backup

2. Use multiple processes to do backup.
How is the effect?

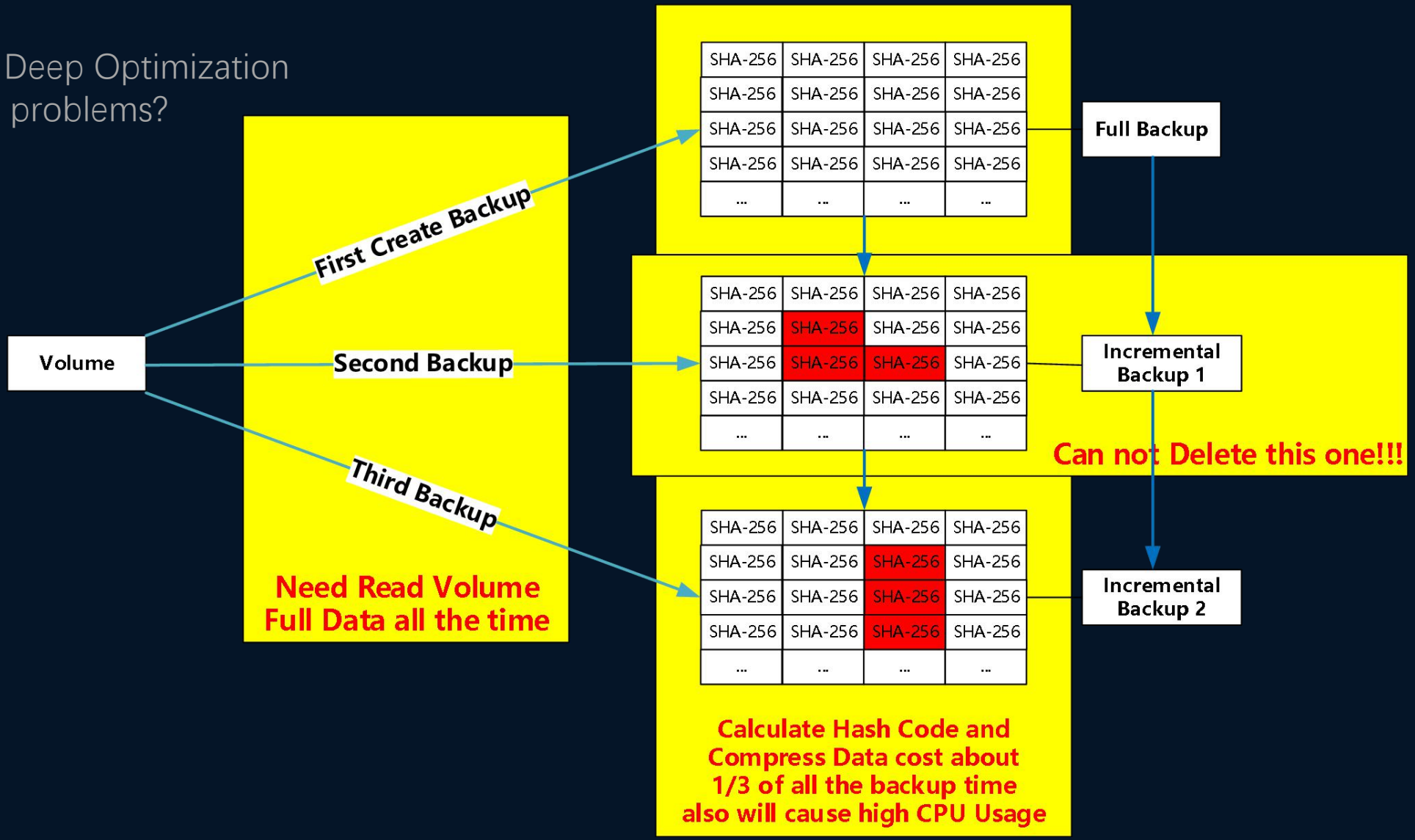


How to improve volume backup

3. Ceph RBD Deep Optimization
Normal Incremental Backup:

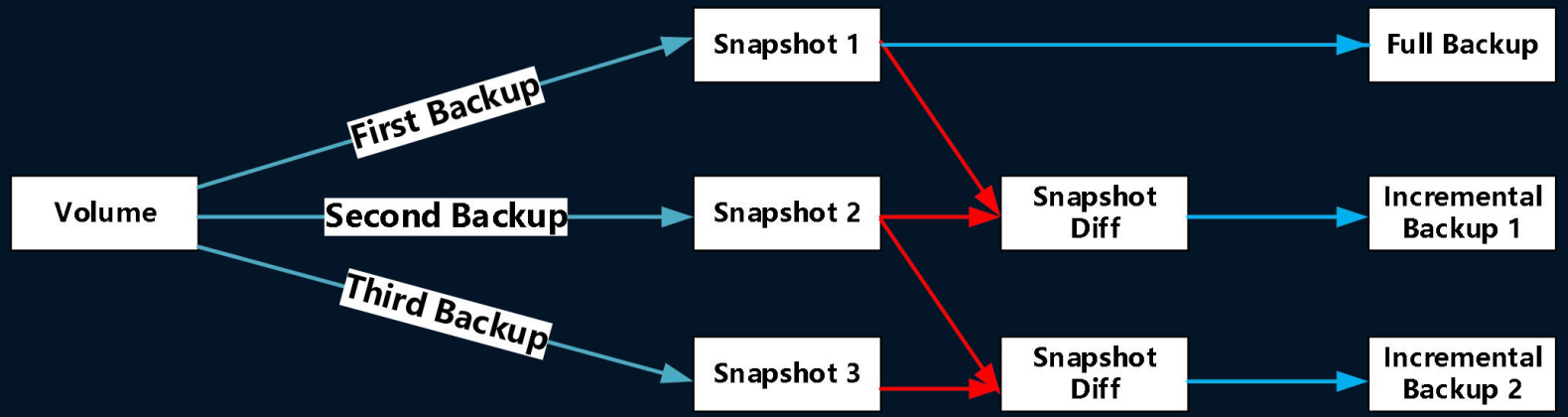
How to improve volume backup

3. Ceph RBD Deep Optimization
What are the problems?



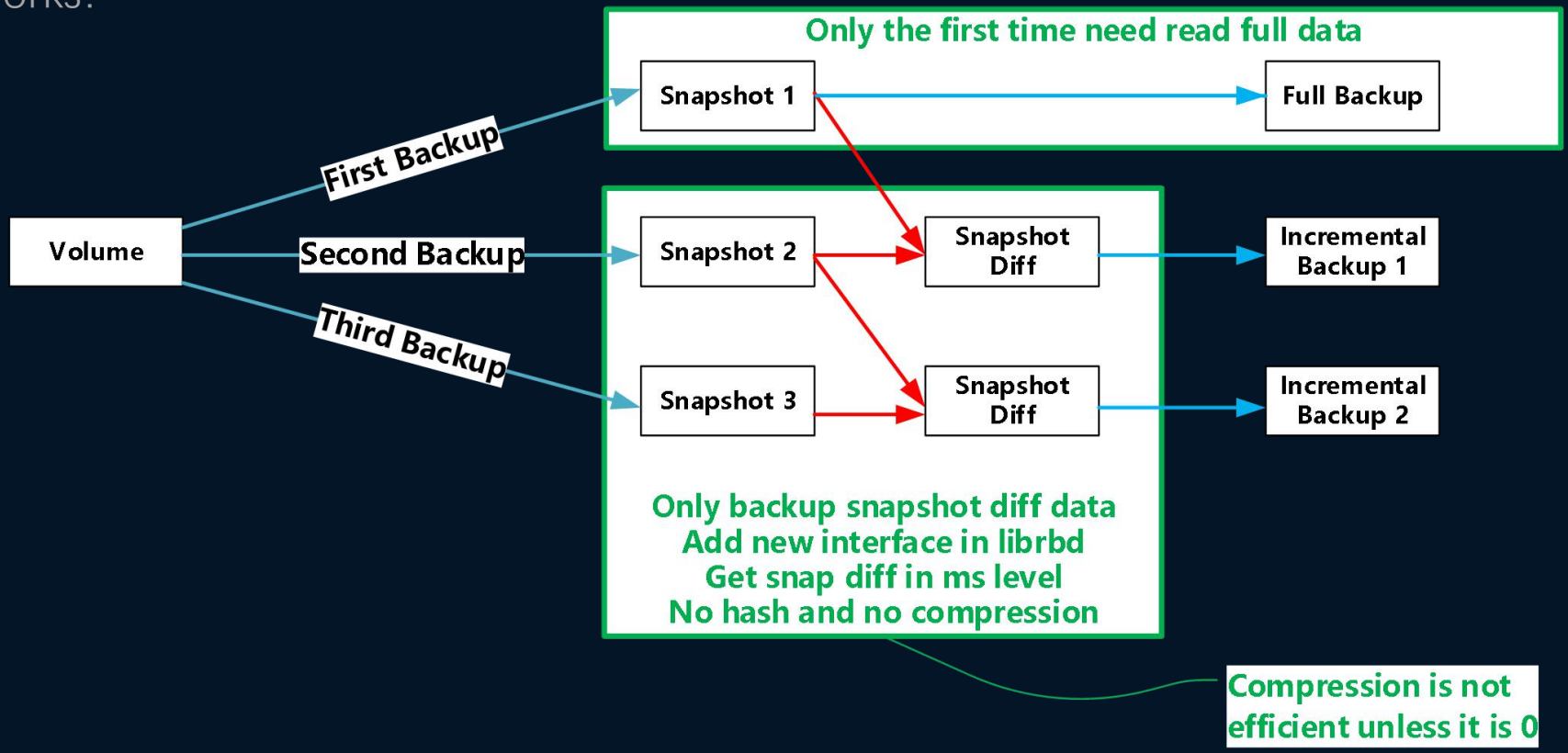
How to improve volume backup

3. Ceph RBD Deep Optimization
A new way:



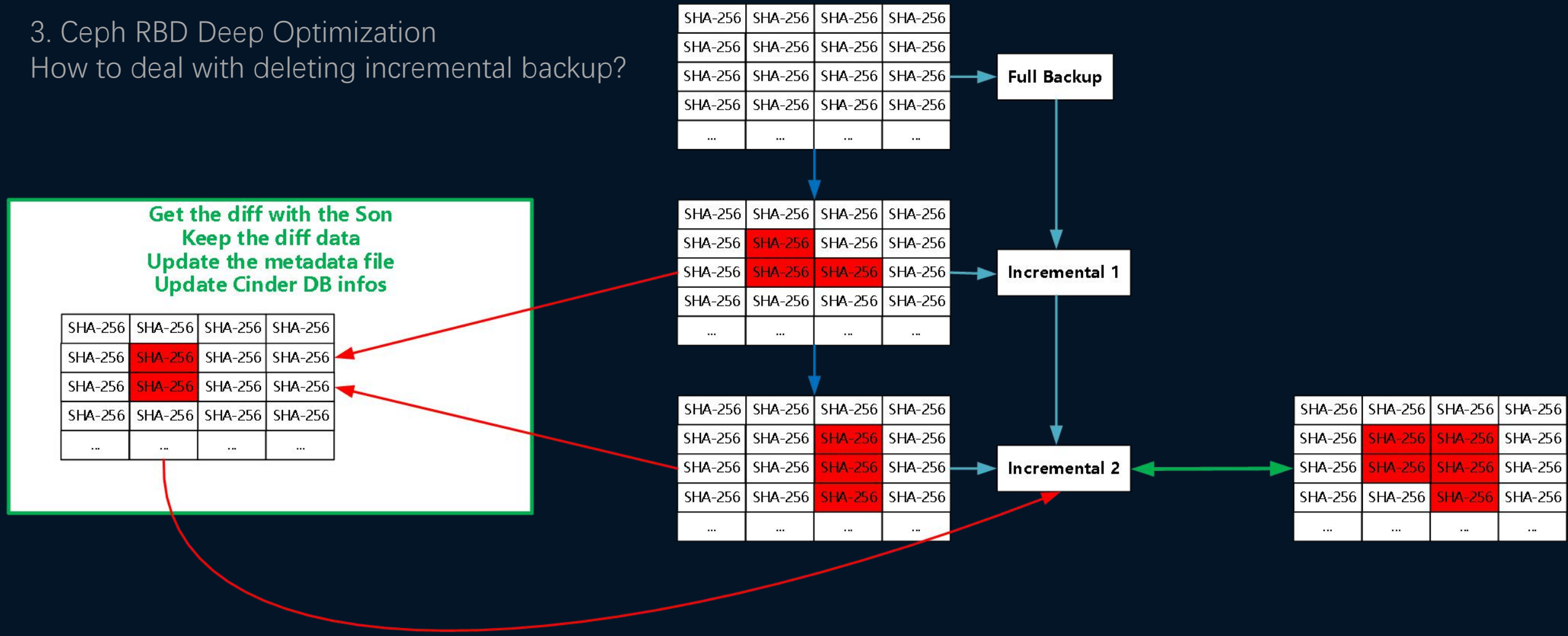
How to improve volume backup

3. Ceph RBD Deep Optimization
How does these works?



How to improve volume backup

3. Ceph RBD Deep Optimization
How to deal with deleting incremental backup?



PART three

3

Performance Enhancement in scheduler

Nova scheduling optimization

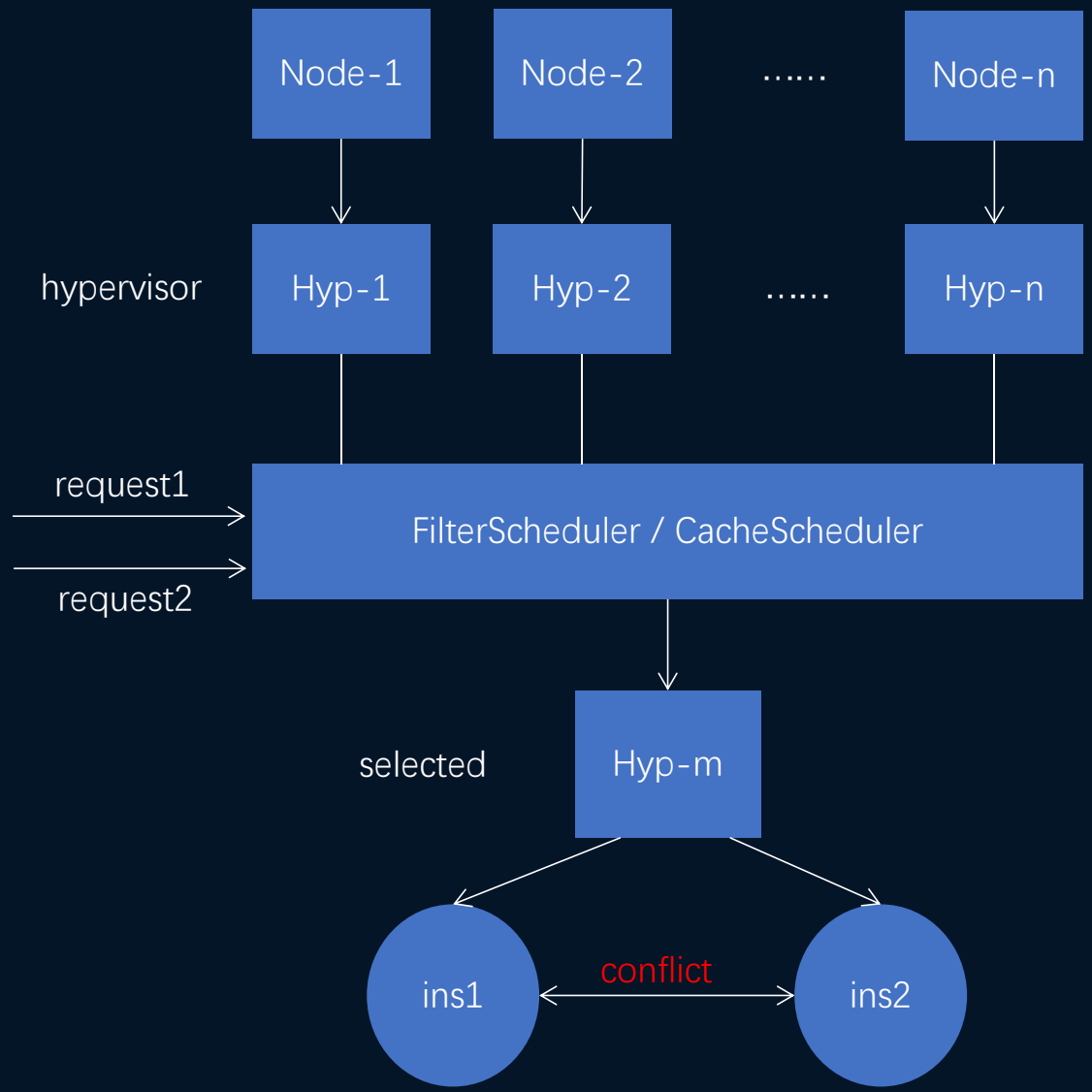
Background:
The scheduling problem in bare-metal high concurrency scenarios

- FilterScheduler

  Resource allocated for a single request is not occupied, causing resource to be reassigned to other request

- CacheScheduler

  The data in the memory often cannot be updated in time, and the resource competition often causes the scheduler to obtain the old data



Nova scheduling optimization

The scheduling optimization in bare-metal high concurrency scenarios

- FilterScheduler

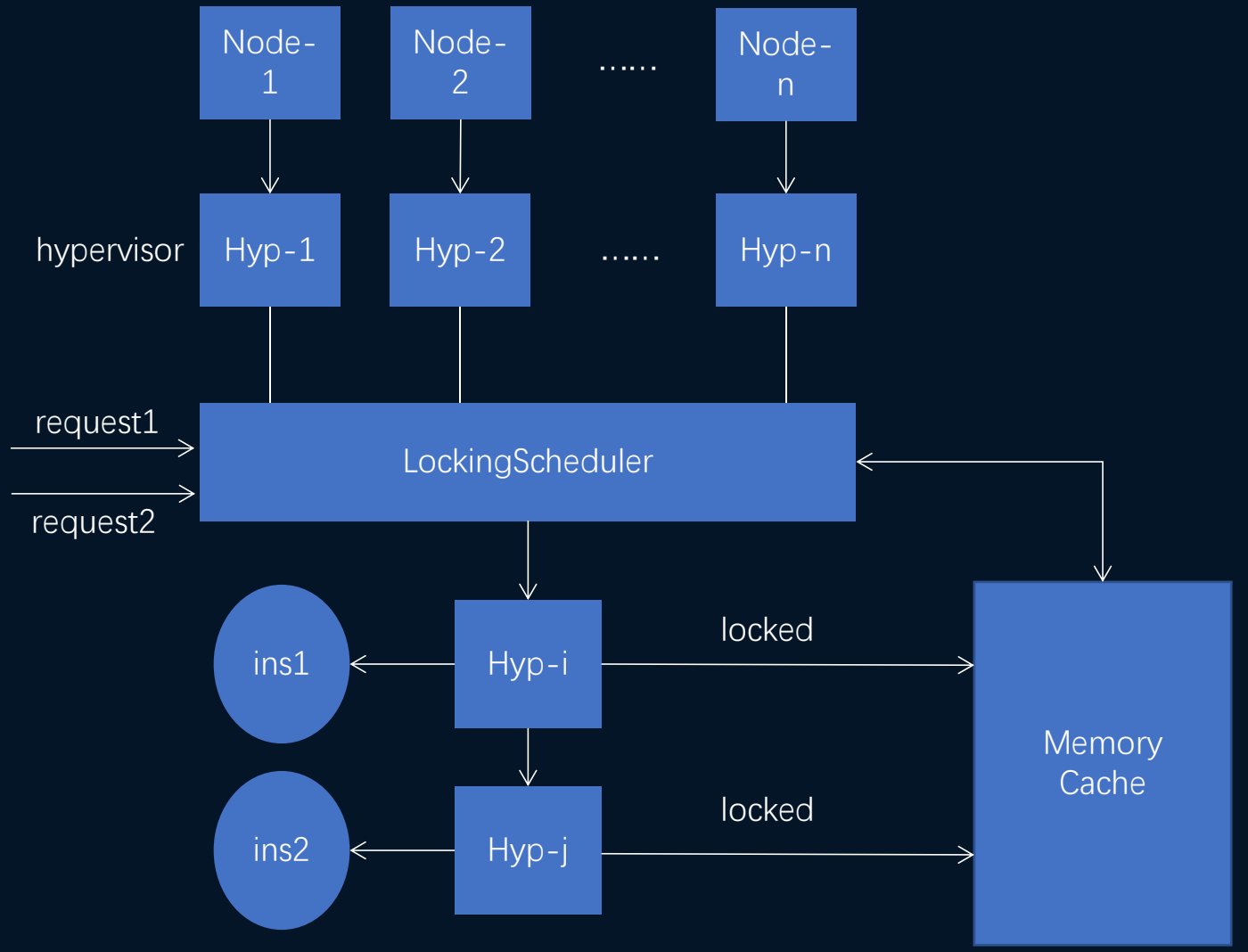
  Add resource occupancy identifiers to avoid rescheduling assignments

- Memory cache

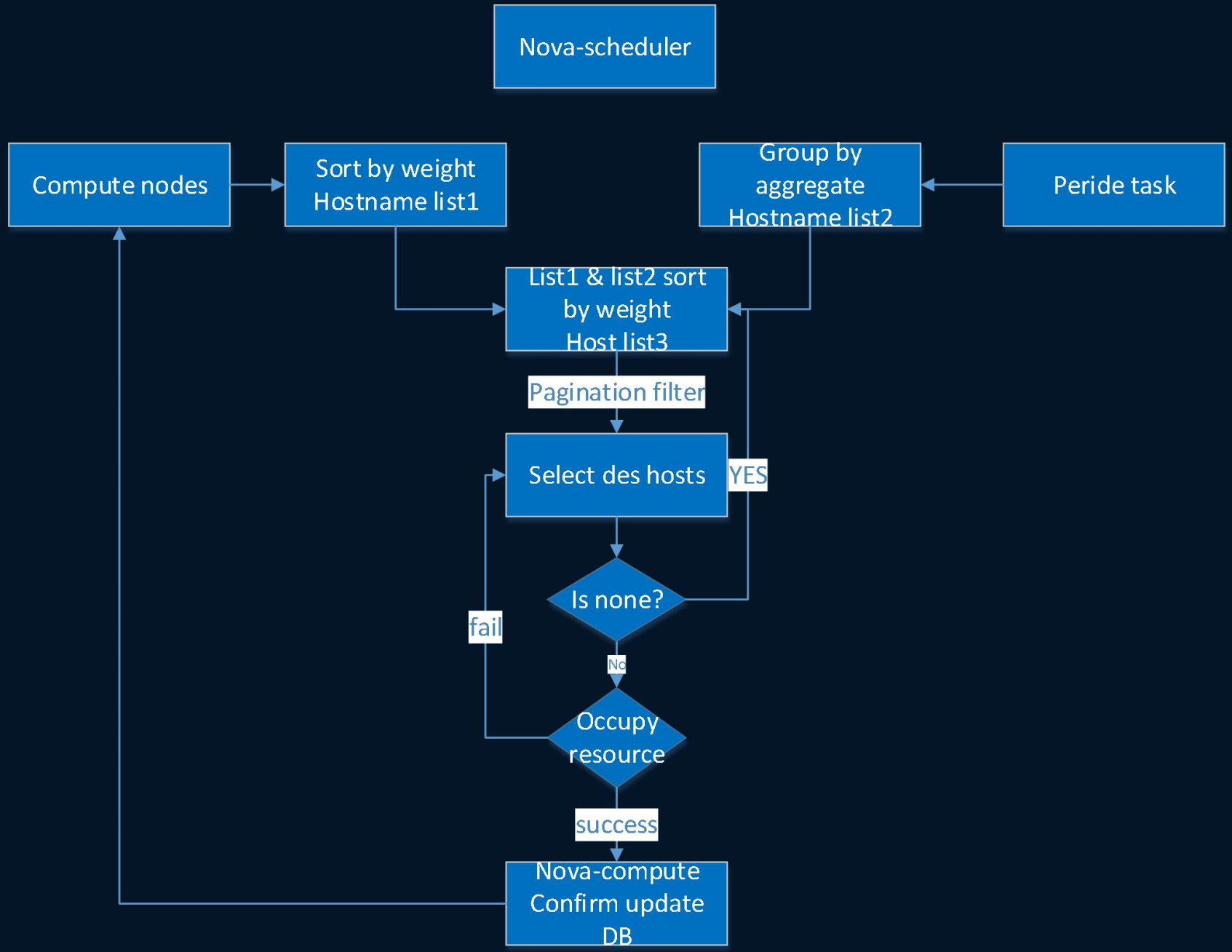
  Record occupied resource information in the memory cache

- A fixed key value in each nova scheduler node

  Maintain usage of occupied resources



Nova scheduling optimization



SUMMARY

As a public cloud created by China Mobile, China Mobile public Cloud will become an important part of China Mobile's 5G+ implementation. In the future, the China mobile public cloud will adopt the N+31+X construction mode and strive to build a network-wide resource layout system of N core level nodes, 31 provincial nodes and “X” cooperative nodes.


中国移动
China Mobile
The Open
Infrastructure Summit
Thank For Watching !
Q&A