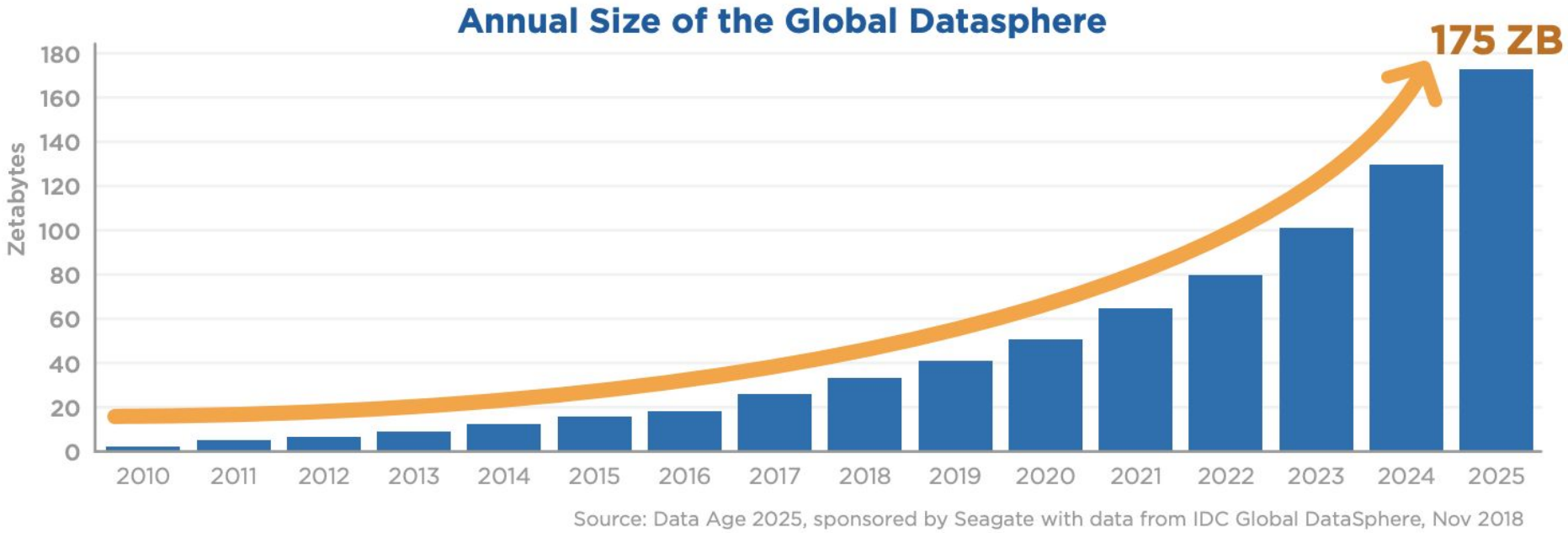
Ceph RGW-Prefetching

Ali R., Amin M., Bissenbay D.

Sprint 2 (Feb 14 - Feb 28)



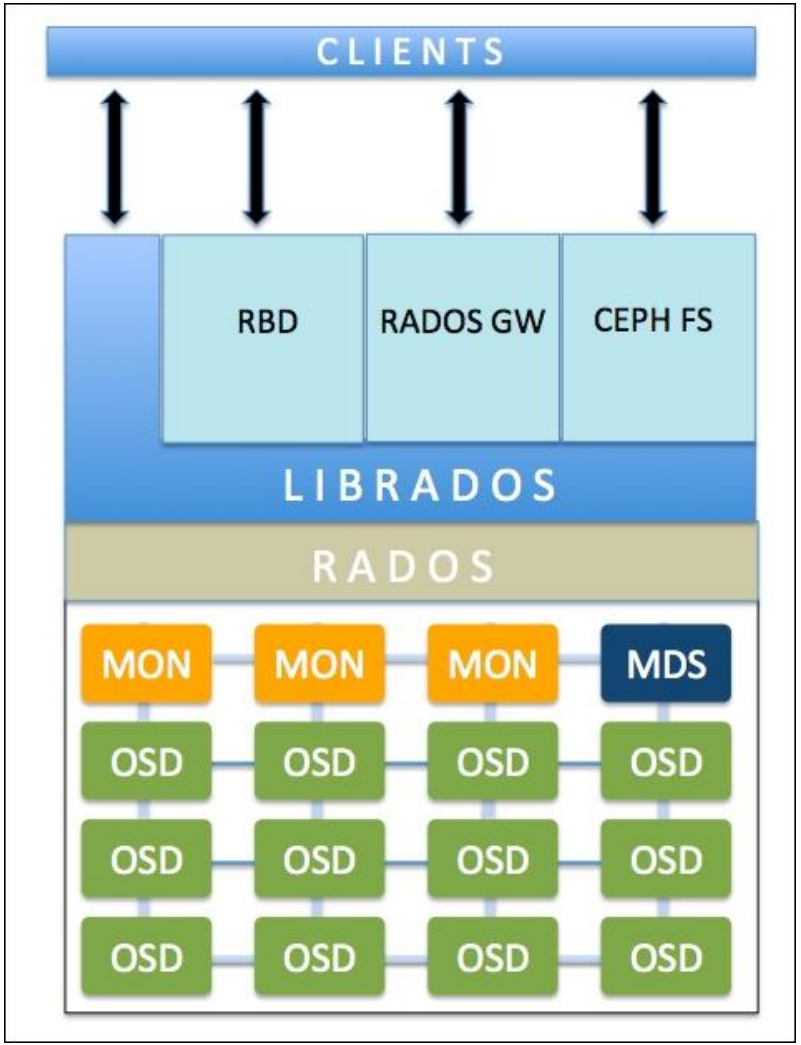
Source: Data Age 2025, sponsored by Seagate with data from IDC Global DataSphere, Nov 2018

Ceph

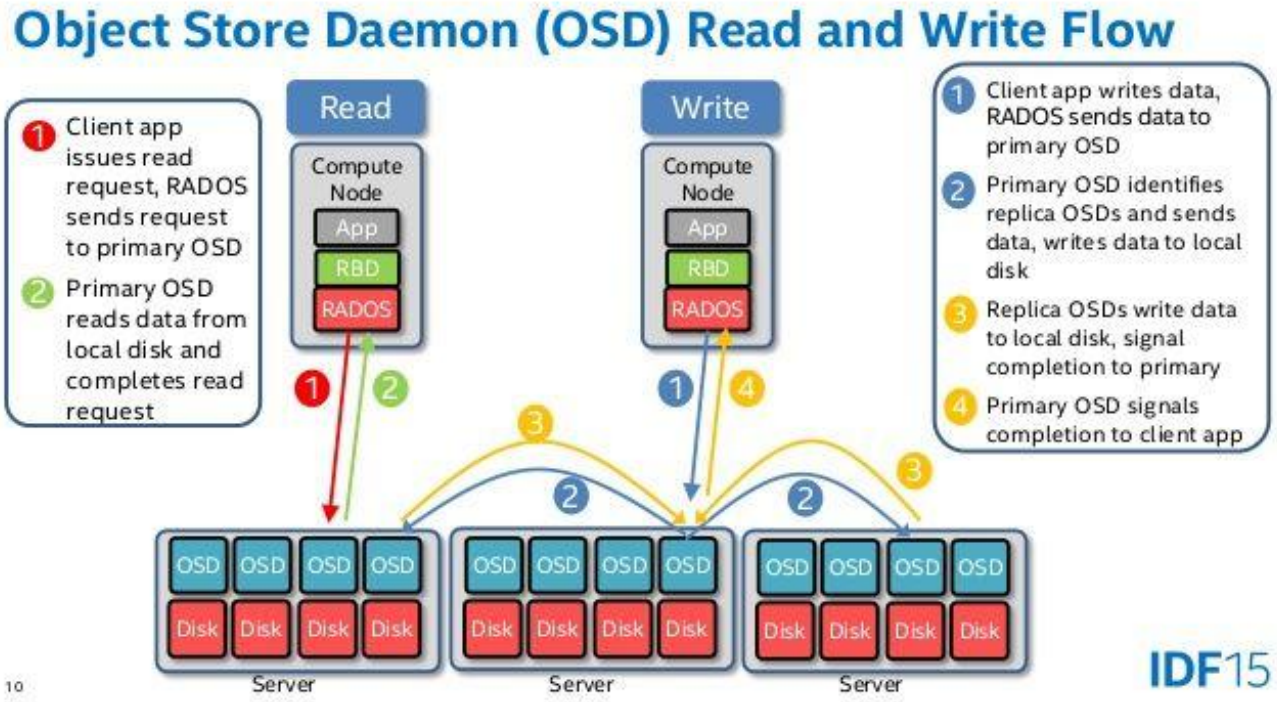
- A distributed storage systems
- Open source
- Provides interfaces for object, block and file level storage
- Scalable to the exabyte level
- Fault-tolerant, Self-healing

Ceph Architecture

- RADOS
- OSD
- RADOS Gateway
- CRUSH



Ceph Flow


Object Store Daemon (OSD) Read and Write Flow
1 Client app issues read request, RADOS sends request to primary OSD
2 Primary OSD reads data from local disk and completes read request
Read
Write
Compute Node
App
RBD
RADOS
1 Client app writes data, RADOS sends data to primary OSD
2 Primary OSD identifies replica OSDs and sends data, writes data to local disk
3 Replica OSDs write data to local disk, signal completion to primary
4 Primary OSD signals completion to client app
OSD
Disk
Server
10
IDF15

Challenges

- Best place to implement prefetching *(lower level vs higher level)*
    - Being explicit or implicit (Using API)
    - Designing prefetching algorithm
    - Extracting the required object information
- Ceph “limits” the client perceived bandwidth
    - We need to be careful about hurting the performance
- Use of debugging tool i.e. gdb
    - Asynchronous functions
- Code is not “well documented”

Implementation Plan

- Previous implementation of cache in Ceph is helpful
- Low Level Implementation
  - “ceph/src/rgw/rgw_rados.cc”
- High Level Implementation
  - “ceph/src/rgw/rgw_op.cc”
- Modify the request data structure being passed to lower level
  - Introduce a flag that indicates the “prefetching request”
  - Check the flag before replying to client and drop if the request was “prefetching request”
- Evaluation: COSBench - Cloud Object Storage Benchmark

Sprint 2 - update

- Code walk with caching team
- Code walk with mentors
- Deploy ceph with cache
- Store and retrieve data

Taiga Updates

Installing Ceph. Storing & retrieving data	Code Walk	Brainstorming design and implementation of prefetching
Installing Standard Ceph	Code walks with RedHat Team	Team Meetings
Integrating Ceph with Cache code		
Storing & retrieving data using S3 API (BOTO SDK)		

Burndown chart in Taiga (cont.)