차세대 DBMS 스토리지 서비스 기술 개발

4세부

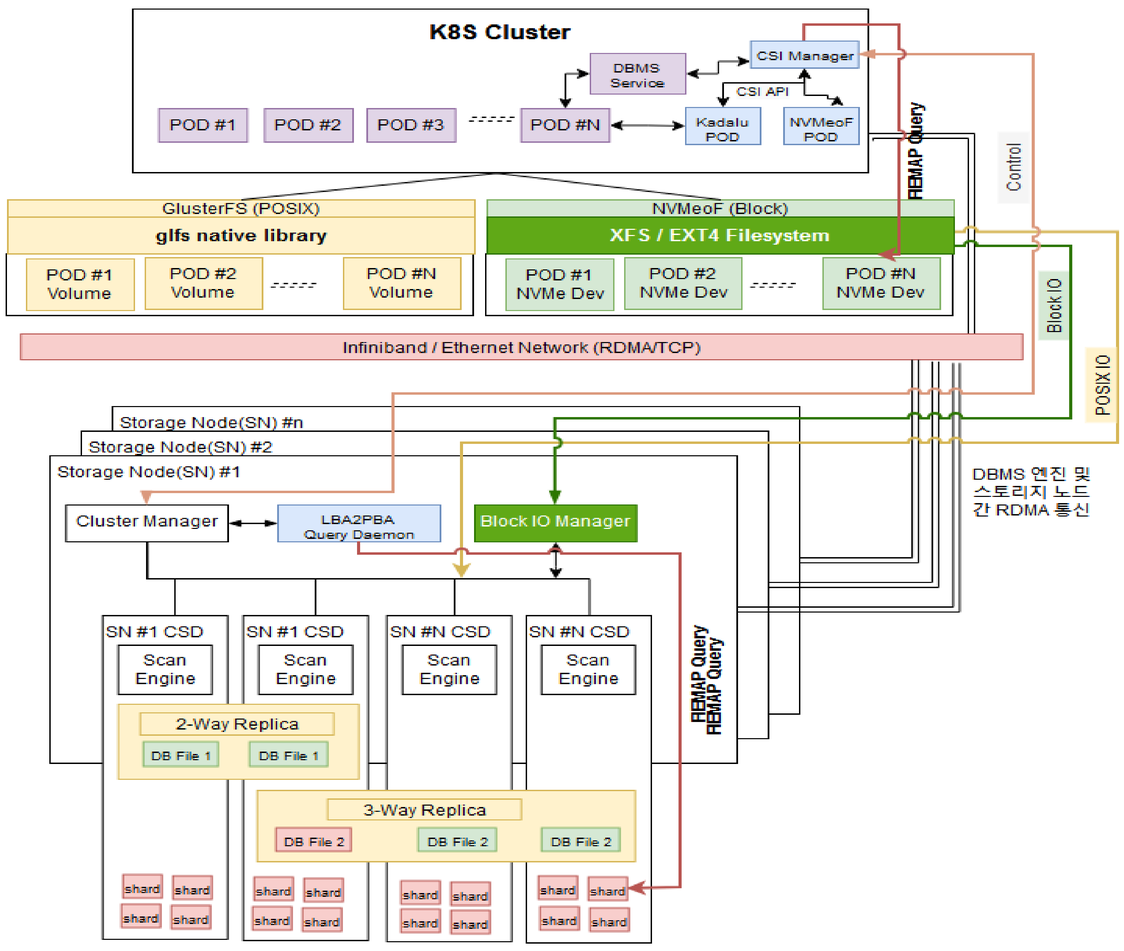
연구 목표

클라우드 향(向 DBMS 분산 스토리지 SW 엔진 오픈소스 化

- 유니파이드 스토리지 인터페이스를 제공하는 Zero-Copy (RDMA/SPDK) 기반 분산 스토리지 고속 입출력 기술 개발
- 클라우드 DBMS 데이터 정합성 보장 위한 ACID (원자성,일관성,고립성,영구성) 고속 I/O 처리 기술 개발
- CSD 레벨 푸쉬다운 쿼리 수행을 위한 분산 스토리지 데이터 오브젝트 물리 주소 변환 인터페이스 기술 개발

연구 내용

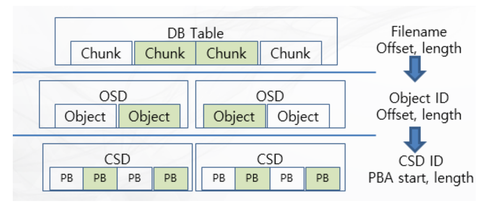
클라우드 향(向) DBMS 분산 스토리지 프로토타입 개발



<클라우드 향(向) DBMS 분산 스토리지 아키텍처 구성도>

- 다양한 Native 클라우드 서비스 지원을 위한 스토리지 인터페이스 제공 모델 개발 (POSIX / BLOCK / OBJECT)
- 저지연 고성능 DBMS 데이터 입출력 처리를 위한 RDMA 기반 NVMeoF 분산 처리 기술 개발
- ACID 보장을 위한 스토리지 복제, WAL, LOCK, Journal Logging 기술 적용

LBA2PBA : 분산 스토리지 데이터 직접 접근 API 기술 개발




```
[root@master ~]# kubectl get pod -o wide
NAME      READY   STATUS    RESTARTS   AGE   IP              NODE     NOMINATED NODE
pod-ext   1/1     Running   0          8s    172.18.219.103  master   <none>
[root@master ~]#

root@pod-ext:/mnt/pv# curl http://172.18.219.103:1111/getPBA -d "fpath=/mnt/pv/messages-20231)" | jq
  % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
                                 Dload  Upload   Total   Spent    Left  Speed
100   188  100   125  100    63  41666  21000 --:--:-- --:--:-- --:--:-- 62666
{
  "data": {
    "i": [
      {
        "DISK": "/dev/sdb",
        "HOST": "gluster-1",
        "LENGTH": 1228523,
        "OFFSET": 1095579
      }
    ]
  },
  "name": "/mnt/pv/messages-20220703"
}
```

- URL => http://{POD-IP}:1111/getPBA
  - Parameter

Name	Required	Type	Description
fpath	True	String	File Path
offsets	True	Dictionary	Offset Dictionary

* offsets = {[Index1]:[OFFSET,LENGTH],[Index2]:[OFFSET,LENGTH]}

  - Return

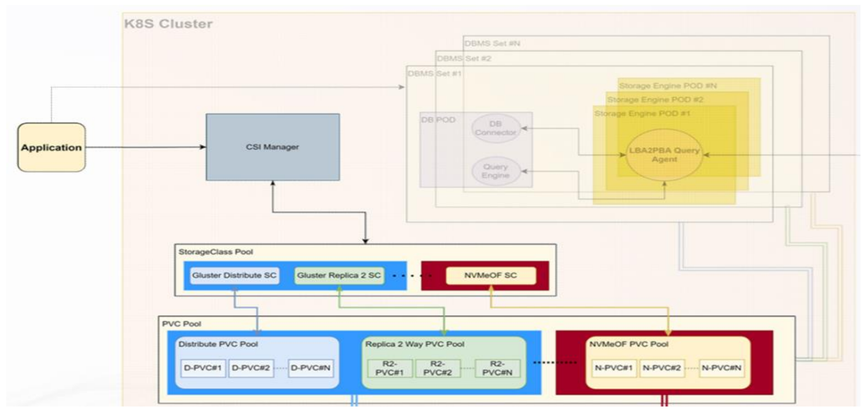
Result	Type	Value
Success	JSON	{ "name" : "a.txt", "data" : { Index : [{ "DISK" : "/dev/sdb", "HOST" : "gluster-1", "OFFSET" : 19231, "LENGTH" : 4182 },], } }
Fail	String	Error Message

<분산 스토리지 물리주소 변환 API 구성 개요>

연구 내용

클라우드 네이티브 스토리지 관리 및 연동 자동화 및 최적화를 위한 이기종 프로토콜 CSI-Manager 기술개발

- DBMS 엔진 인스턴스의 스토리지 영구 볼륨의 동적 할당을 위한 NVMeoF / POSIX CSI 외부 드라이버 기술 개발
- 어플리케이션의 사용 편의성 및 이식 호환성 향상 제공



- URL => http://{CSI-Manager}:1113/create
  - Parameter

Name	Required	Type	Description
pvcName	True	String	PVC Name
pvcType	True	String	NVMeoF / Gluster
dupliType	True	String	Replica [1,2,3] / Distribution

  - Return

Result	Type	Value
Success	JSON	{ "pvcName" : [PV-Name] }
Fail	String	Error Message

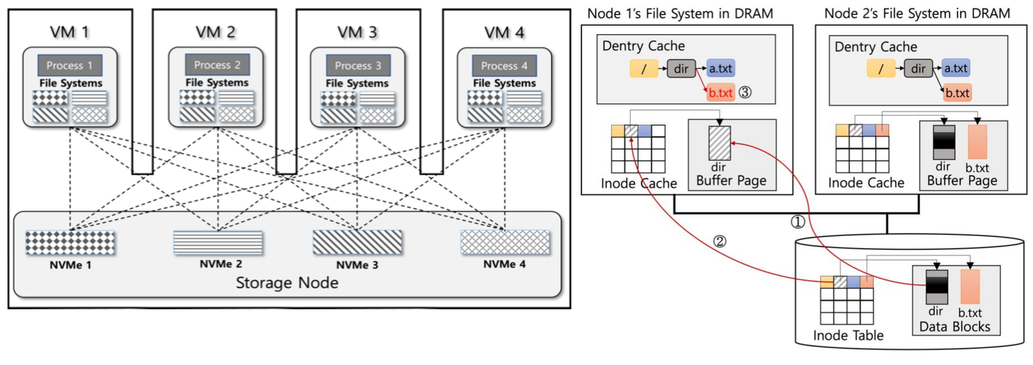
- URL => http://{CSI-Manager}:1113/get
  - Return

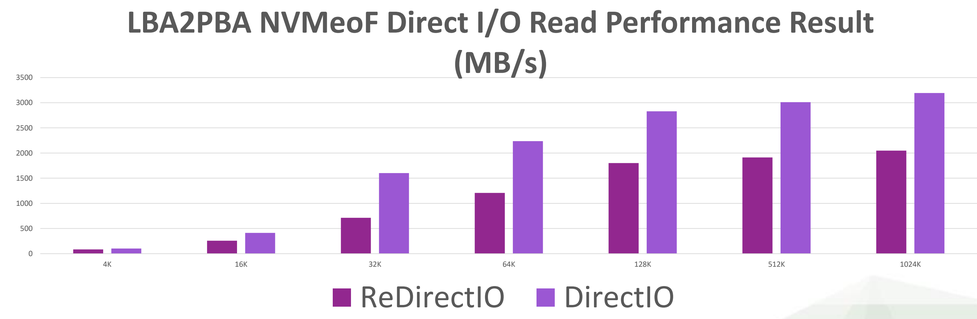
Result	Type	Value
Success	JSON	[{ "pvc" : [PVC Name], "type" : [Gluster | NVMeoF], "dupliType" : [Replica(1,2,3) | Distribution], "subdir" : [" " | Gluster Sub Volume Path] }]
Fail	String	Error Message

<영구 동적 볼륨 할등을 위한 컨테이너 연동 구조 및 REST API >

NVMeoF 블록 인터페이스 최적화 기술 프로토타입 개발

- 고성능 NVMeoF 블록스토리지의 읽기 성능 향상을 목적으로 LBA2PBA 인터페이스를 활용한 Cluster File system 레이어를 우회하는 Direct I/O 처리 기술 개발





향후 계획

- SPDK 기반 CSD NVMe 입출력 커널 메모리 복제 우회를 통한 입출력 고도화 기술 적용
- LBA2PBA 실시간 처리 기술 적용을 위한 변환 기술 최적화
- NVMeoF 성능 향상을 위한 Direct Read Data Access 캐시 관리 기법 고도화 및 클러스터 파일시스템 우회 Direct Data Write 기술 개발