以 Presto 为例对接 Fluid DataTable

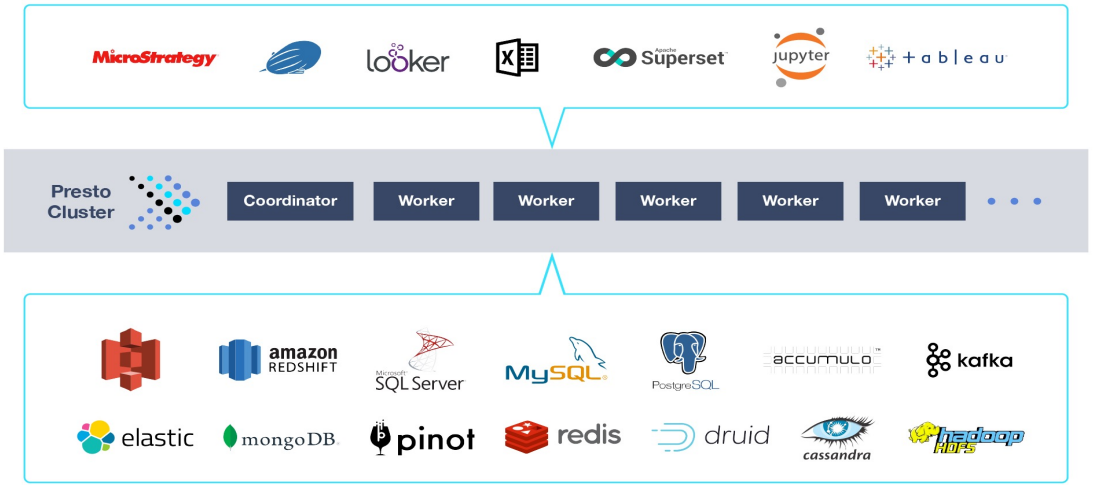
2023.03.16

背景技术介绍

Presto介绍

- Presto是一个开源的、分布式SQL 查询执行引擎。
- **优势：**
  - 基于内存的 MPP 架构，性能较高
  - 允许跨多个数据源进行查询
  - 基于 Pipeline 设计，计算过程中会立即返回已计算好的结果

……



背景技术介绍

问题

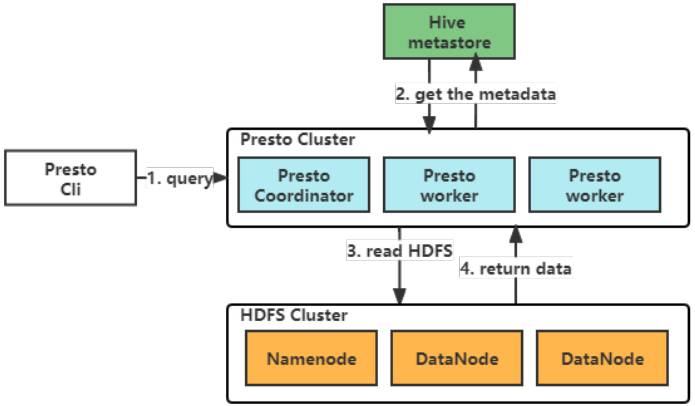
如何将数据编排系统 Fluid 与 分布式 SQL 查询引擎 Presto 相结合，为Presto**提供加速查询能力？**

背景技术介绍

Presto访问底层存储系统（以HDFS为例）

流程：

- Presto客户端提交查询请求
- Presto Cluster从Hive metastore获取元信息
- Presto Cluster从HDFS中获取数据

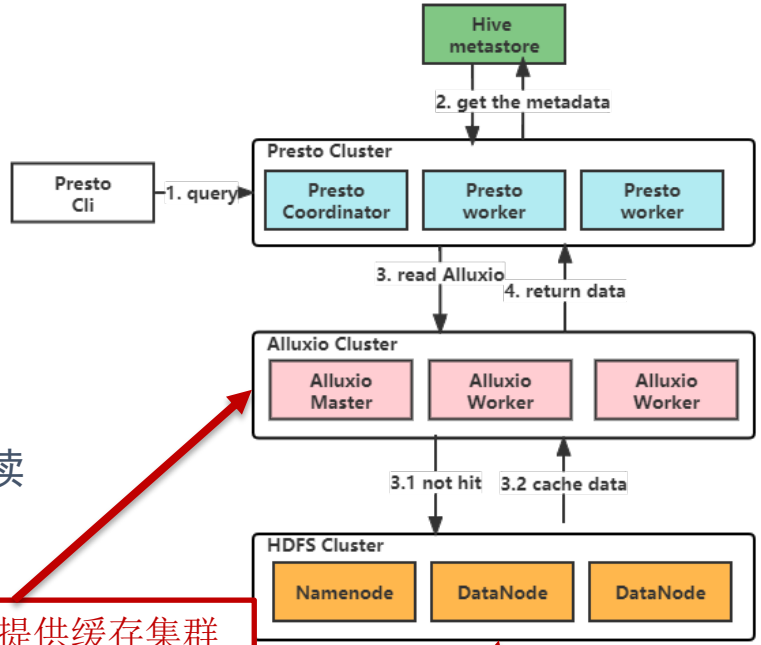


背景技术介绍

Presto访问缓存系统（以Alluxio为例）

流程：

- Hive 客户端修改表的位置信息
- Presto 客户端提交查询请求
- Presto Cluster 从 Hive metastore 获取元信息
- Presto Cluster 从 Alluxio 中获取数据
- 若没有命中，Alluxio 会进一步从底层存储系统读取并进行缓存

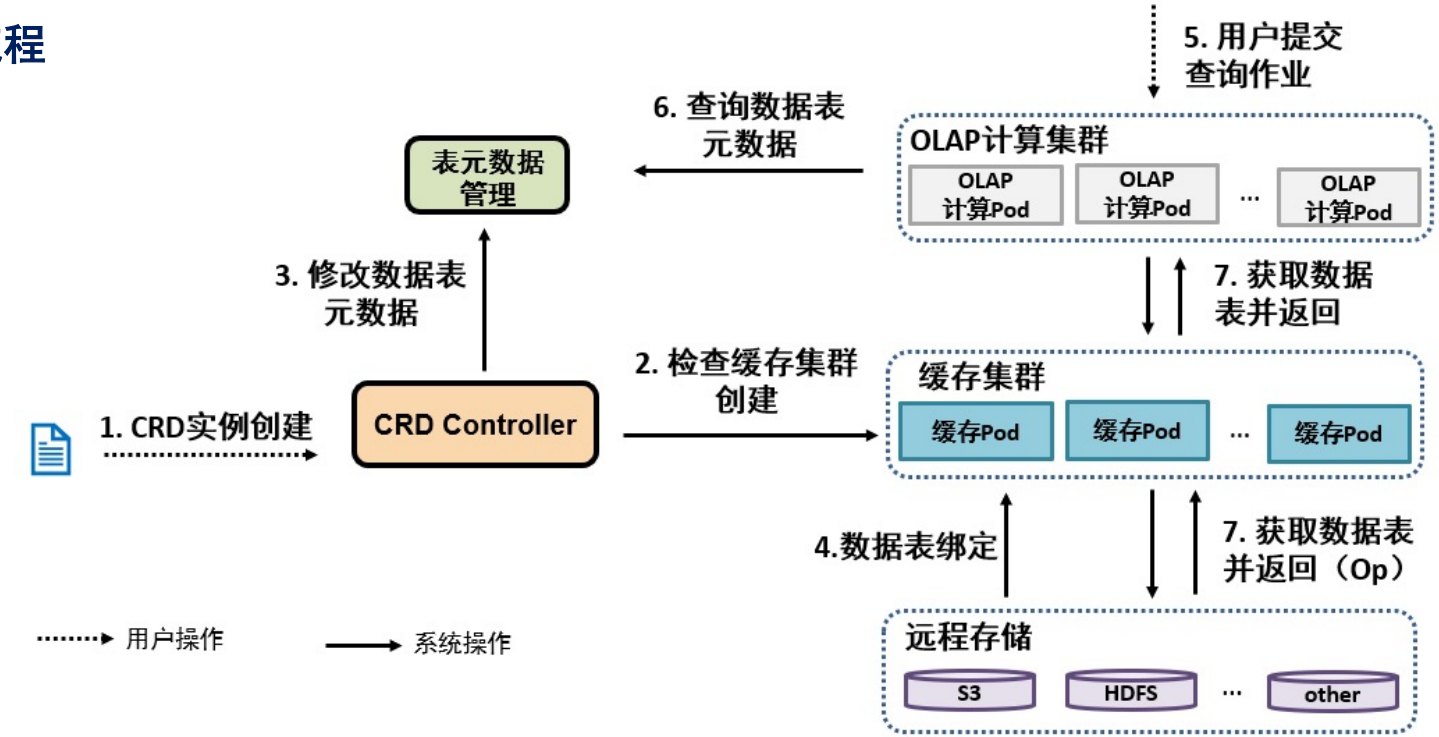


可以利用 Fluid 的Dataset 和 Alluxio Runtime 为 Presto 提供缓存集群

新建 DataTable CRD 实现对表数据的抽象

方案设计

运行流程



- 通过 Dataset 与 AlluxioRuntime CRD 部署公用缓存集群；
- 通过修改 Dataset 的 MountPoint 实现表数据的动态绑定；

使用示例

部署 DataTable CRD 实例

- 指定元数据服务地址为 210.28.132.15:10000
- 指定挂载数据为数据库 experiment 中的 store_sales 表

```yaml
apiVersion: data.fluid.io/v1alpha1
kind: DataTable
metadata:
  name: partcustomer
spec:
  url: 210.28.132.15:10000
  schemas:
    - schemaName: experiment
      tables:
        - tableName: store_sales
```

我们还支持多种粒度的挂载：库、表、分区

使用示例

部署 DataTable CRD 实例

1. 若集群中没有公用缓存集群，则创建

```
wenxiaowang@WenxiaodeMacBook-Pro ~ kubectl get pods
NAME                                READY   STATUS    RESTARTS        AGE
datatable-common-master-0           2/2     Running   0               4d
datatable-common-worker-0           2/2     Running   2 (3d21h ago)   5d23h
datatable-common-worker-1           2/2     Running   0               4d
```

2. 修改元数据信息

```
LastAccessTime:         UNKNOWN
Retention:              0
Location:               hdfs://210.28.132.15:9000/user/hive/warehouse/experiment.db/store_sales
Table Type:             EXTERNAL_TABLE
Table Parameters:
```

```
LastAccessTime:         UNKNOWN
Retention:              0
Location:               alluxio://114.212.84.87:21049/experiment-store_sales
Table Type:             EXTERNAL_TABLE
```

使用示例

部署 DataTable CRD 实例

3. 在缓存集群中挂载表数据

3.1 修改 Dataset 挂载点

```
mounts:
- mountPoint: https://mirrors.tuna.tsinghua.edu.cn/apache/db/ddlutils/
  name: init-point
- mountPoint: hdfs://210.28.132.15:9000/user/hive/warehouse/experiment.db/store_sales
  name: experiment-store_sales
phase: Bound
runtimes:
```

3.2 AlluxioRuntime Controller 对变动进行响应执行 alluxio fs mount 命令

```
Defaulted container "alluxio-master" out of: alluxio-master, alluxio-job-master
[root@cqm-OptiPlex-7040 alluxio-2.9.0]# alluxio fs ls /
             0       PERSISTED 12-02-2022 12:28:13:187  DIR /experiment-store_sales
             0       PERSISTED 03-10-2023 02:17:00:499  DIR /init-point
```

4. 部署 Presto 集群，在客户端提交任务即可享受缓存带来的加速能力

使用示例

删除 DataTable CRD 实例

取消挂载和恢复元数据

修改 Dataset 挂载点

```
hcfs:
  endpoint: alluxio://datatable-common-master-0.default:21049
  underlayerFileSystemVersion: 3.3.1
mounts:
- mountPoint: https://mirrors.tuna.tsinghua.edu.cn/apache/db/ddlutils/
  name: init-point
phase: Bound
runtimes:
```

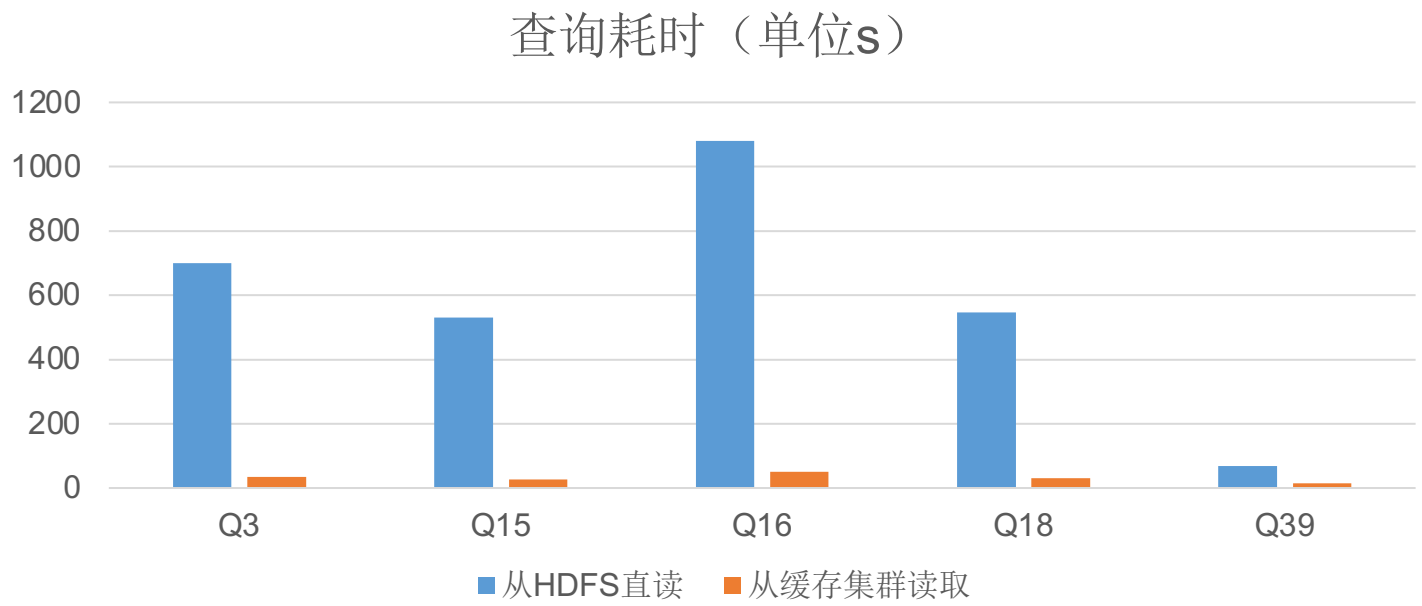
恢复元数据

```
LastAccessTime:         UNKNOWN
Retention:              0
Location:               hdfs://210.28.132.15:9000/user/hive/warehouse/experiment.db/store_sales
Table Type:             EXTERNAL_TABLE
Table Parameters:
```

使用示例

加速效果

使用 TPC-DS 数据集进行查询



缓存集群能大幅降低大表查询耗时

总结

开发进度

总进度的 20%

结论

Fluid DataTable 能在大数据查询中网络带宽出现瓶颈的情况下，对重复查询语句起到大幅的加速效果。

展望

未来或许直接对接云原生数据库

谢 谢！ Q&A