

# Hadoop、Hive在滴滴的应用

滴滴出行 基础平台部 - 费辉、黄辉







扫码加入数据Talk二群



# Hadoop框架介绍与常见问题排查方法总结

滴滴出行 基础平台部 - 费辉

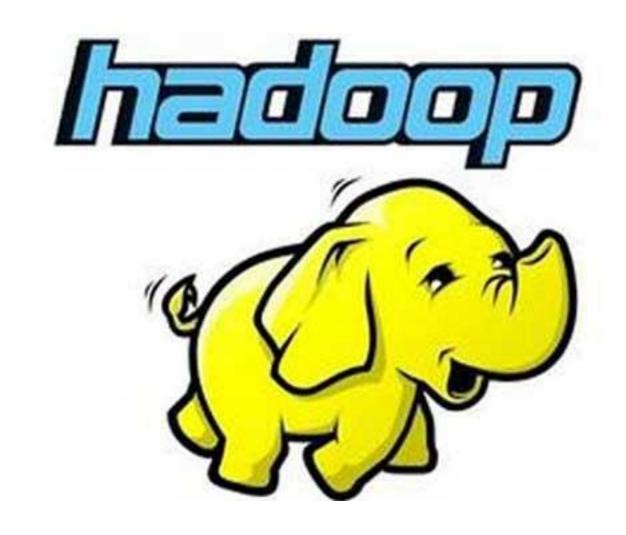




- O Hadoop框架介绍
- 河 滴滴离线集群现状
- **常见问题排查方法**



Apache Hadoop是 一套用于在由通用 硬件构建的大型集 群上运行应用程序 的框架。它实现了 Map/Reduce编程范 型,计算任务会被 分割成小块(多次) 运行在不同的节点





Hadoop Common

• 通用组件, 支持其他模块

**HDFS** 

• 分布式文件系统

**YARN** 

• 资源调度和管理

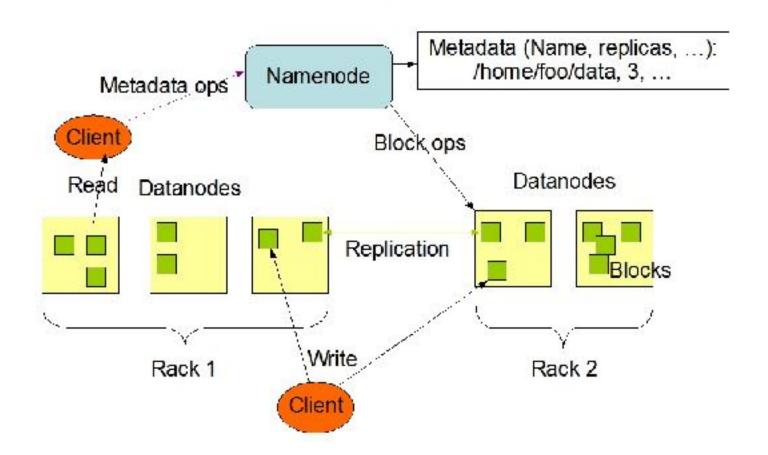
**MAPREDUCE** 

• 基于YARN的并行计算模型



HDFS是服务于Hadoop 的分布式文件系统,由 NameNode, DataNode和Client三部 分组成。

#### HDFS Architecture





# Hadoop框架介绍——HDFS

#### NameNode

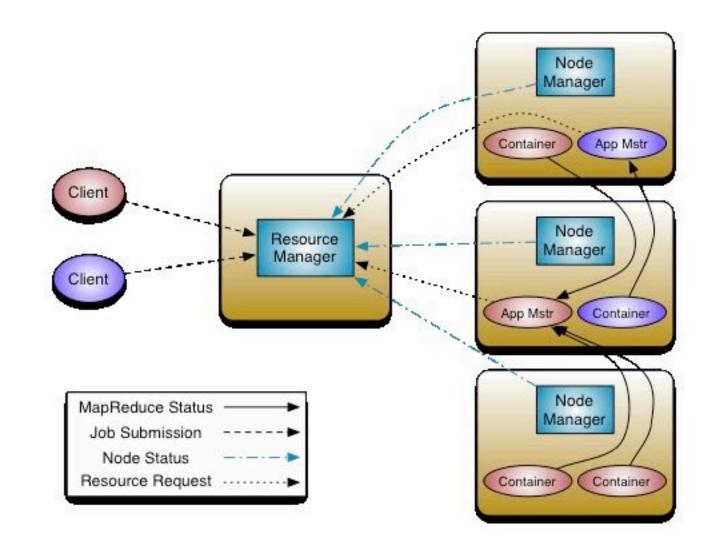
- Namenode是HDFS的管理中心,即Master节点
- 维护整个hdfs文件系统的目录树
- 每一个路径(文件)所对应的block块信息(block的id,及所在的datanode服务器)

#### DataNode

• block的存储管理,即Slave节点



YARN 是下一代 Hadoop 计算平台





# Hadoop框架介绍——YARN

# ResourceManager

- 全局的资源管理器,负责整个系统的资源管理和分配
- 调度器(Scheduler)和应用程序管理器(Applications Manager)。

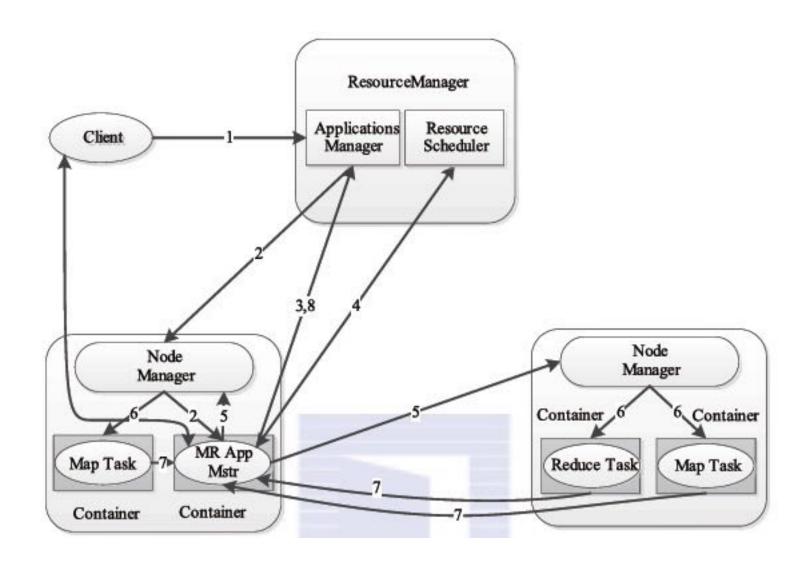
# NodeManager

- NodeManager是每个节点上的资源和任务管理器
- NodeManager定时向ResourceManager汇报本节点资源(CPU、内存)的使用情况和Container的运行状态

MORE THAN A JOURNEY



MAPREDUCE是基于 YARN的并行计算框架





# Hadoop框架介绍——MAPREDUCE

# MRAppMaster

- MRAppMaster是MapReduce的ApplicationMaster实现
- Map Reduce可以直接运行在YARN上

# Mapper

• 对输入做处理,产生key/value对

#### Reducer

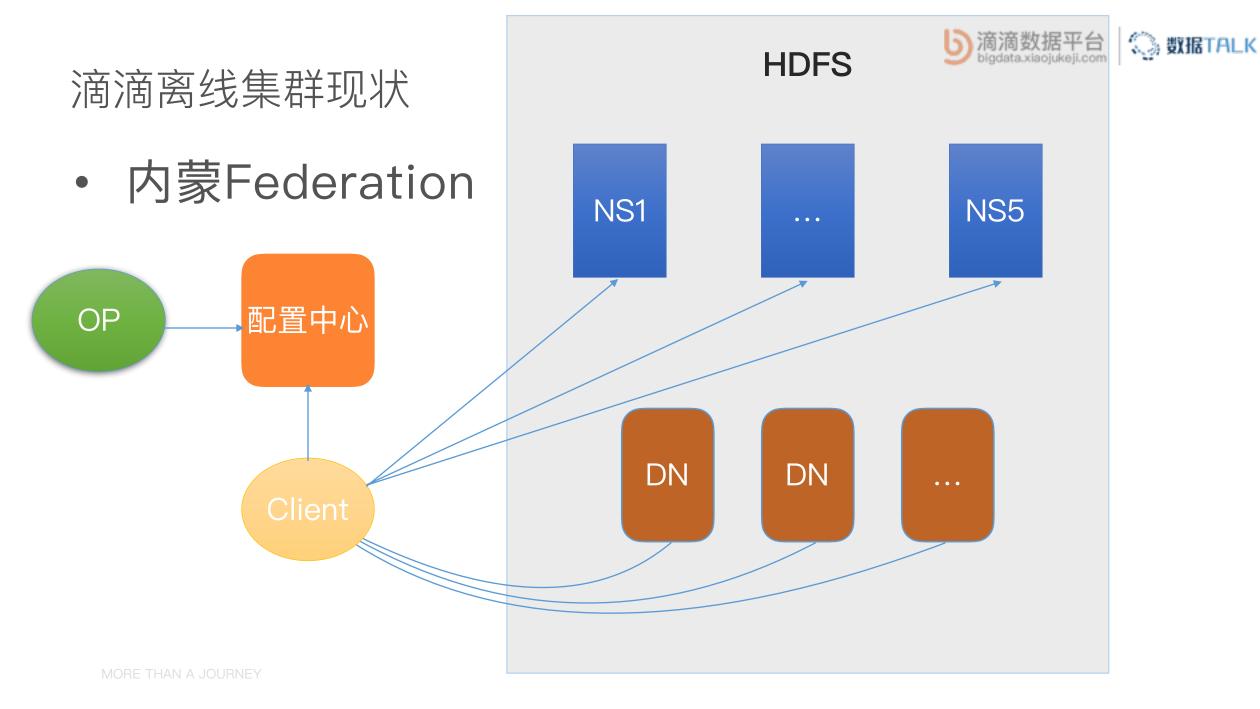
• map产生的key/value做规约处理





#### 滴滴离线集群现状

- 内蒙集群
  - 5 NS
  - 4800+ Nodes
- 美东集群
  - 1 NS
  - 20+ Nodes





#### 滴滴离线集群现状

# • 资源申请

- 存储
- 队列
- 安全
  - 白名单
  - 密码机制
- 申请地址
  - http://dpp.intra.xiaojukeji.com/index#/

MORE THAN A JOURNEY



#### 滴滴离线集群现状

# • 回收站新机制

- 统一管理, 在/trash/user/\${username}下
- 不占用用户的Quota
- 定期清理, 保留时间1天

# • HDFS生命周期

- 存储有限,无用数据可以设置生命周期,特别是采集数据
- http://dpp.intra.xiaojukeji.com/index#/apply/hdfsLifecycle

MORE THAN A JOURNEY



#### 常见问题排查方法——常见问题

- 客户端怎么安装?
- 集群地址是啥? (查看运行任务的web ui)
- hdfs配置的nn和port信息是啥?
- ·出现了xx错误,帮忙排查一下?
- 任务运行慢,帮忙看一下?
- 任务卡住了,帮忙看一下?



#### 常见问题排查方法——了解客户端

# • 客户端安装

- 内蒙集群 didi-hadoop-nmg <a href="http://wiki.intra.xiaojukeji.com/">http://wiki.intra.xiaojukeji.com/</a> pages/viewpage.action?pageId=123503753
- 美东集群 didi-hadoop-us <a href="http://wiki.intra.xiaojukeji.com/pages/viewpage.action?pageId=132909957">http://wiki.intra.xiaojukeji.com/pages/viewpage.action?pageId=132909957</a>
- root 权限

# • 客户端区分

- 代码一份,配置不同: \$HADOOP\_HOME/etc/hadoop
- rpm qa | grep hadoop



### 常见问题排查方法——了解hadoop配置

- HDFS配置
  - \$HADOOP\_HOME/etc/hadoop/hdfs-site.xml

```
🚧 🚾 ClusterNing, ClusterNing2, ClusterNing3, ClusterNing3, ClusterNing3, ClusterNing3, ClusterNing3, ClusterNing5, xgpre-hbase-cluster, xg-hbase-cluster, gz-hbase2-cluster, gz-hbase2-cluster, gz-hbase2-cluster, gz-hbase3-cluster
  <mane>dfs.ha.namenodes.DClusterHmgs/name
 <value>nn1,nn2</value>
*/property*
 <mane>dfs.ha.namenodes.DClusterHmg2</name>
 -value-mi1.mi2</value-</pre>
e/propertys
cproperty>
 mane=dfs.ha.namenodes.DClusterHmg3</name>
 svalucennet mazes/valuce
enty>
«property»
 *mane=dfs.ha.namenedes.bClusterHmg4*/name*
 -cyaluso-nn II , nn Lic/valus-
<mane>dfs.namenode.rpc-address.DCLusterNmg.nn1
 <velue>bigdata=nmg=hdpnn28.nmg91:0028</velue>
*/property*
 <mane>dfs.namenode.rpc-address.DClusterNmg.nn2</name>
 evalue>bigdata-nmg-hdpnn01.nmg01:8820
s/property>
 <mane>dfs.namenode.http-address.DClusterNng.nnl
 svalue>bigdata-rmg-hdpnn88.nmg81:58678
</preperty>
 <mane>dfs.namenode.http-address.DClusterNng.nn2
  cvaluesbigdata-reg-hdpmn21.neg01:50078</values
```



### 常见问题排查方法——了解hadoop配置

# • YARN配置

\$HADOOP\_HOME/etc/hadoop/yarn-site.xml

```
<name>yarm.resourcemanager.ha.rm-ids
 <value>rm1, rm2</value>
s/property».
<name>yarm.resourcemanager.hostmame.rm1</name>
  <value>bigdata nng hdprm00.nmg01
spropertys
 <name>yarm.resourcemanager.nostmane.rm2</name>
 <value>bigdata=nng=hdprm01.nmg01/value>
snanceyarm.resourcemanager.resource-tracker.address.rmis/nance
 <value>bigdata-nng-hdprm00.nmg01:8025
<name>yarm.nescurcemanager.rescurce=tracker.address.rm2
 <walue>bigdata-nng-hdprm81.nng81:8825*/value>
<nanc>yarm.resourcemanager.scheduler.address.rml
 <value>bigdata-nng-hdprm08.nmg61:10008
cproperty>
 <name>yarm.resourcemanager.scheduler.address.rm2
 <value>bigdata-nng-hdprm01.nmg01:19030</value>
erty>
 <name>yarm.resourcemanager.webapp.address.rml
 <value>bigdata=nng=hdprm00.nmg01:8888
snanczyarnu resourcemanageru vehappu addressu rm2s/nancz
 <velue>bigdata-nng-hdprm81.nmg81:8855</velue>
```



# 常见问题排查方法——了解hadoop配置

# • MAPREDUCE配置

• \$HADOOP\_HOME/etc/hadoop/mapred-site.xml

```
<name>mapreduce.framework.name
 <value>yarn</value>
</preperty>
property>
 sname>mapreduce.job.reduce.slowstart.completednaps
 <value>!</value>
property>
 <name>mapreduce.job.reduces
 <value>18</value>
</preperty>
property>
 sname=mapreduce.jobhistory.address
 svalue>bigdata-nmg-historyserver@0.nmg01:10020s/value>
<name>mapreduce.jobhistory.webapp.address
 <value>bigdata-nmg-historyserver00.nmg01:19080</value>
<name>mapreduce.map.java.opts
 svalues-Mrx768rs/values
cproperty>
 <name>mapreduce.map.output.compress
 <value>true</value>
erty>
 <name>mapreduce.reduce.java.opts
 svalue>-Xmx768ms/value>
<name>mapreduce.task.io.sort.factor
 <value>38</value>
erty>
```





#### 常见问题排查方法——学会搜索

- 公司 wiki
  - http://wiki.intra.xiaojukeji.com/
- 搜索引擎
  - Google/Baidu
- hadoop\_user群公告
  - 帮助文档
  - FAQ



#### 常见问题排查方法——学会看日志

#### YARN UI

- 内蒙
  - http://nmgrm.intra.xiaojukeji.com/cluster
  - http://bigdata-nmg-hdprm01.nmg01:8088/cluster (需要代理)
- 美东
  - http://bigdata-us-hdprm11.us01:8088/cluster



#### 常见问题排查方法——学会看日志

Logged in as: dr.who

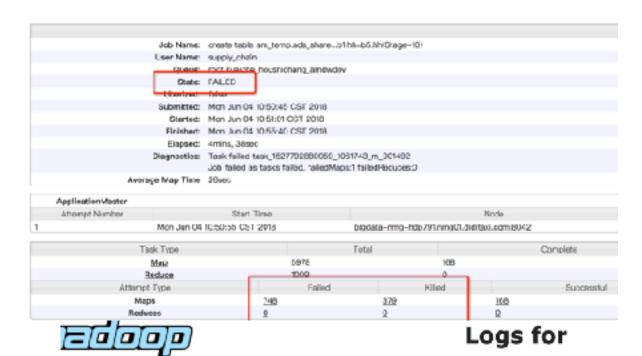
#### ons

04

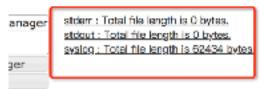
| Used               | VCores Total | VCores    | VCores Reserved Active |         | Nodes         | sioned Nod                                   | ned Nodes L                |                             | Lost Nodes U |                            | Rebooted Nodes |                   |                      |
|--------------------|--------------|-----------|------------------------|---------|---------------|--|----------------------------|-----------------------------|--------------|----------------------------|----------------|-------------------|----------------------|
|                    | 2800         | 0         |                        | 70      | 0             |  |                            | (                           | 0            | 0                          |                | 0                 |                      |
| ed                 | Memory Used  |           | Memory Pending         |         | Memo          | Memory Reserved                              |                            | VCores Used                 |              | d VCores Pending           |                | VCores Reserved   |                      |
|                    | 0 B          | 0 B       |                        |         | 0 B           |  | 0                          |                             | 0            |                            |                | 0                 |                      |
| Minimum Allocation |              |           |                        |         |               | Maximum Allocation -cmemory:15380, vCores:65 |                            |                             |              |                            |                |                   |                      |
|                    |              |           |                        |         |               |  |                            |                             |              |                            | Search:        |                   |                      |
| Qu                 | oue :        | StartTime | FinishTime             | State 0 | FinalStatus 0 | Running<br>Containers                        | Allocated<br>CPU<br>VCores | Allocated<br>Memory<br>MB © | VCores       | Reserved<br>Memory<br>MB © | Progress :     | Tracking UI ©     | Blacklister<br>Nodes |
| mational           | prod         | 2018-08-  | N/A                    | RUNNING | UNDEFINED     | 4  | 4                          | 8192                        | 0            | 0                          | [              | ApplicationMaster | 0                    |

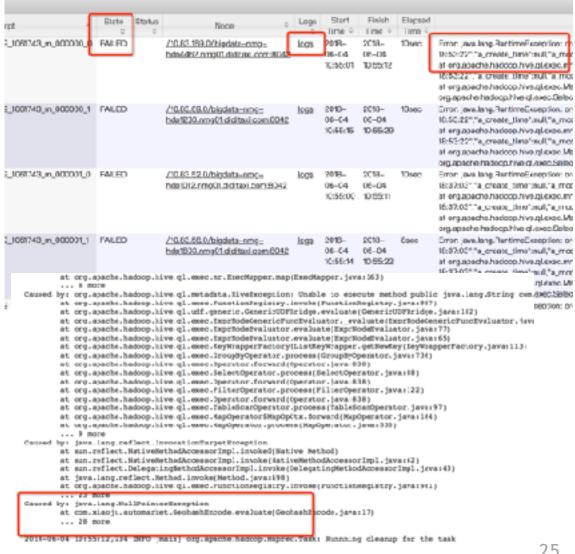


#### 常见问题排查方法——学会看日志



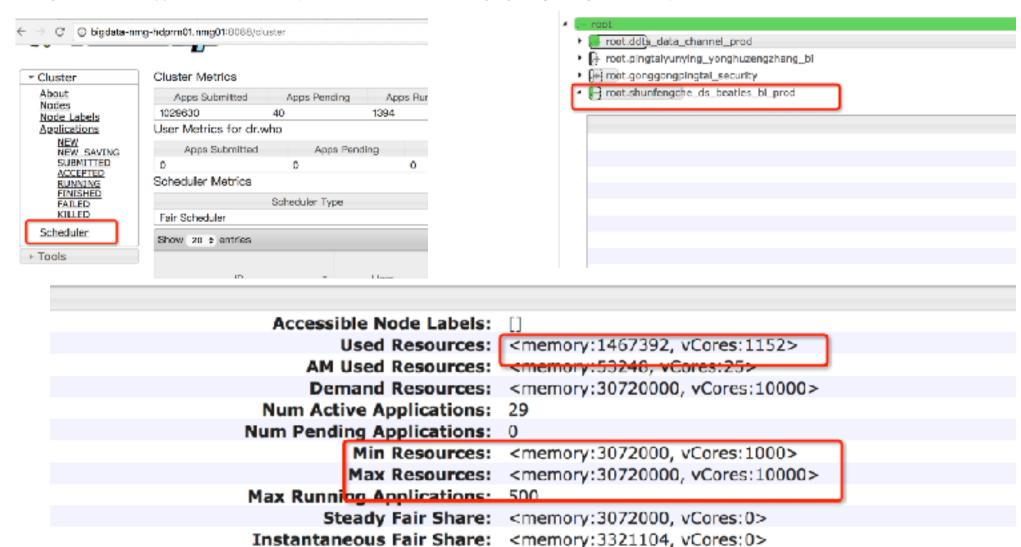
#### container\_e12208\_1527792880059\_106







#### 常见问题排查方法——学会看计算资源





#### 常见问题排查方法——问题总结

# physical memory limits.xxx.Killing container

- mapreduce.map.memory.mb mapreduce.map.java.opts
- mapreduce.reduce.memory.mb mapreduce.reduce.java.opts
- JVM 堆参数大小是container的0.8左右,不能与container大小一样
- container最大申请内存15G
- 报错 ip地址 not in white list
  - 安全上的限制,需要为客户机申请白名单才能使用线上离线集群
  - 申请地址 http://dpp.intra.xiaojukeji.com/index#/apply/whiteList

MORE THAN A JOURNEY



#### 常见问题排查方法——问题总结

- Failed to submit application\_XXX to YARN:
   User YYY cannot submit applications to queue
  - 先确定任务是否设置了队列,如果没设置需要设置队列 mapreduce.job.queuename
  - 如果设置了队列,确认该账号是否有权限可以在该队列提交任务,如果没有请申请队列权限,http://dpp.intra.xiaojukeji.com/index#/apply/yarnUserAsk
- FAQ
  - http://wiki.intra.xiaojukeji.com/display/BigdataOfflineEngine/



# **THANKS**



# Hive在滴滴的应用与常见问题排查总结和优化建议

DIDI CHUXING / 06.06 基础平台部-大数据架构部 - 黄辉





CHAPTER 1

# Hive在滴滴的应用



如 架构演进 <sup>应用 + 计划</sup>

02 权限设计 设计 + 使用

03 生命周期

设计 + 管理





02 权限设计 设计 + 使用

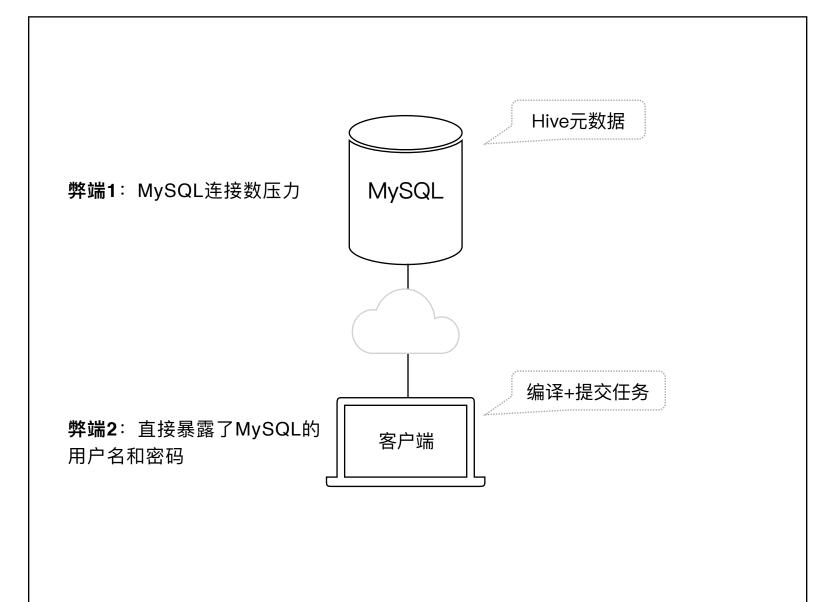
03 生命周期

设计 + 管理

#### 最开始的样子

- ・客户端 + MySQL
- ・小规模





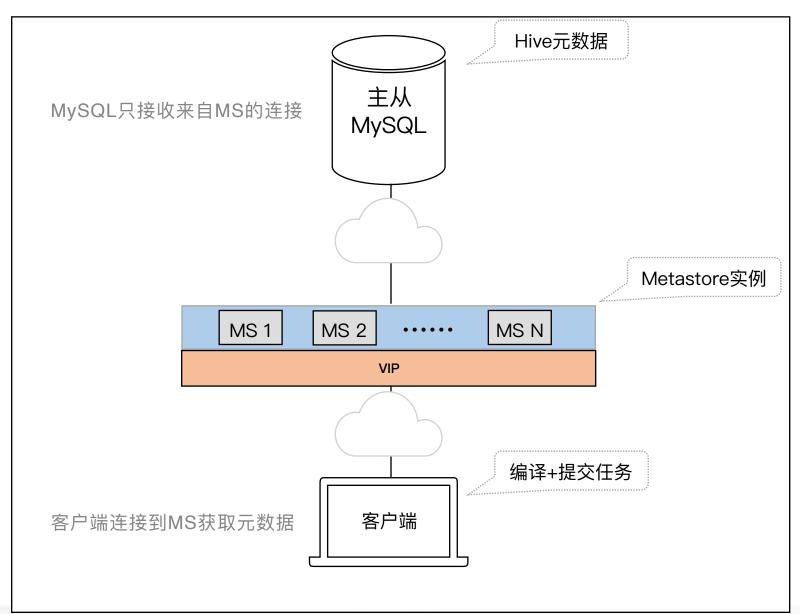
## 引入MetaStore

- ·解决客户端直连MySQL的问题
- MetaStore+LVS
- 动态拓展+负载均衡
- 客户单配置使用配置中心

目前我们有四个实例提供服务每天需要处理近1亿次的请求

一定要使用客户端吗?









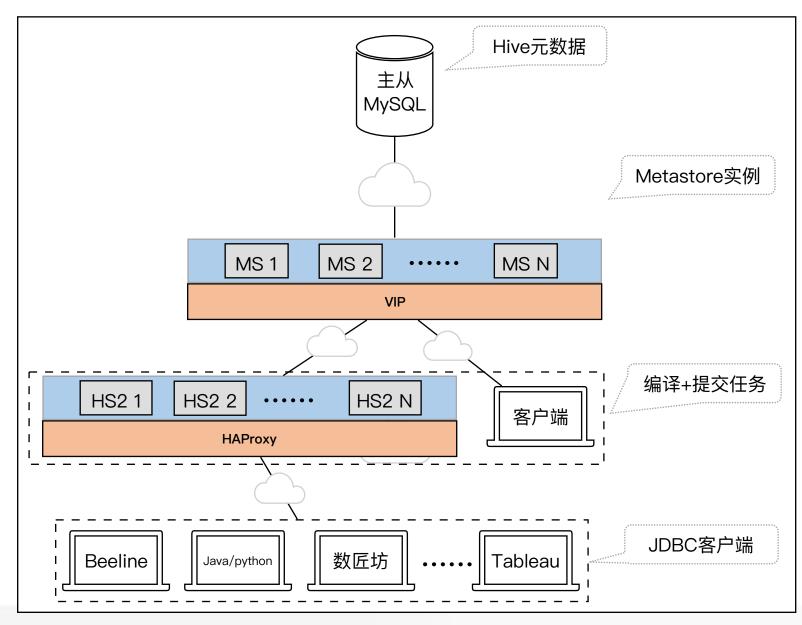
### 引入HiveServer2

- ·支持多种开发语言以JDBC方式访问
- 便于上层产品设计

目前我们总共有8个实例提供服务。

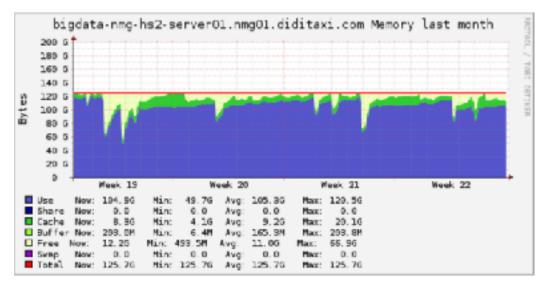
最近50天共处理了153万个回话。

Tableau、数匠坊



# HiveServer2的 限制

- 内存使用+异常连接请求影响稳定性
- · 严格模式+连接数quota







最近一个月的 内存使用

- 1,请使用【数匠坊-SQL开发】、【在数匠坊docker、客户端使用beeline】、JDBC连接的用户注意: 严格模式将对HQL查询产生以下限制,不满足严格模式的HQL将报错,报错案例参见下文
  - 1) 分区表的查询必须加上分区过滤条件
  - 2) 不能使用笛卡尔积

报错信息例如: No partition predicate found for Alias "a:ods\_g\_driver\_updegrade\_process" Table "ods\_g\_driver\_updegrade\_process"

2, 请使用连接池的用户特别注意:

每个用户同时最多只能建立50个连接,保持的连接数超过这个值后,新的连接将会失败(失败提示: org.apache.hive.service.cli.quota.QuotaManager\$QuotaExceededException: Quota Type: CONNECTION QUOTA)

报错案例: <a href="http://wiki.intra.xiaojukeji.com/pages/viewpage.action?pageId=132688999">http://wiki.intra.xiaojukeji.com/pages/viewpage.action?pageId=132688999</a>



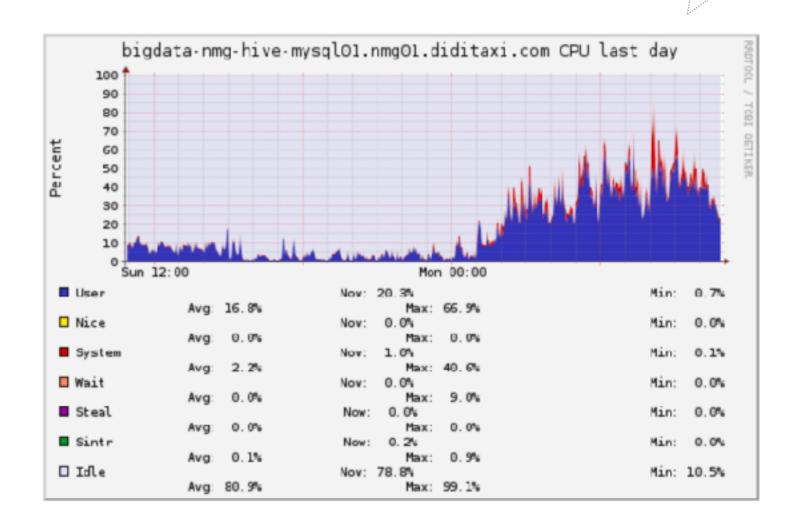


# Hive MySQL瓶颈

- ·高峰期MySQL的压力很大
- 异常用户的查询导致100%使用

元数据治理

用户程序改造





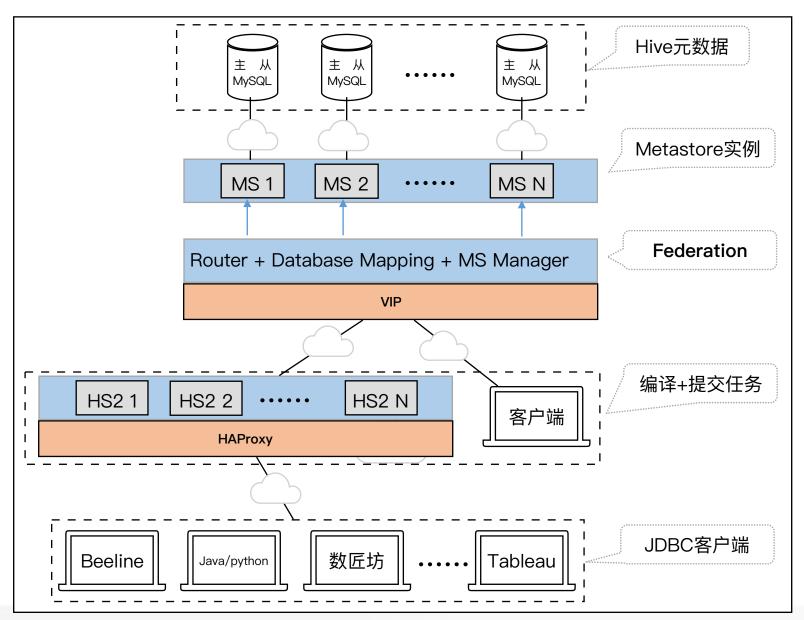


### Hive federation

- 缓解元数据压力
- 读写分离
- 改造



https://github.com/HotelsDotCom/waggle-dance







02 权限设计 设计 + 使用

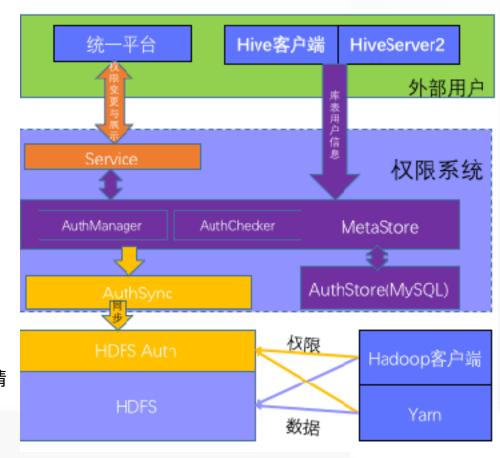
03 生命周期

设计 + 管理



## 权限管理方案

- 通过DPP完成权限的申请、审批和管理
- 采用Hive+HDFS结合方式,将SQL-based权限管理方案部分功能移植到MS端
- 1.只有owner和管理员可以drop database/table;(注: 默认创建表的账号为owner)
- 2.只有owner和管理员可以alter database,比如修改库owner
- 3.只有管理员可以create database, 所以有创建数据库需求请大家到d++平台上申请
- 4.读权限通过HDFS来管控



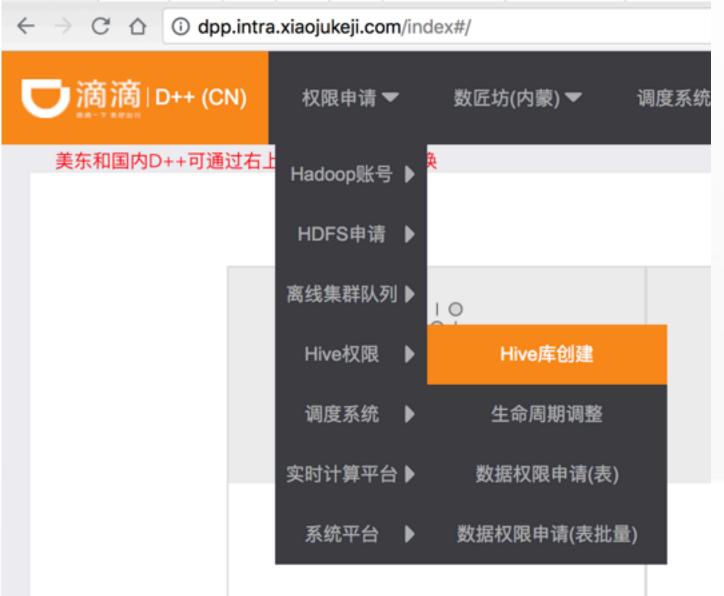
[huanghui@bigdata-nmg-client01.nmg01 ~]\$ hadoop fs -ls -d /user/huanghui/warehouse/huanghuitest.db/test1 drwxr-x--- - huanghui test1 0 2017-08-17 19:49 /user/huanghui/warehouse/huanghuitest.db/test1

## 权限申请

- 申请创建库
- 申请表的读权限









## 权限授予/回收

- 库的owenr或者表的 owner可以主动授权
- · 个人中心-我的权限-Hive管理权限





## 权限相关的问题

- •由于现有机制的实现,一个hdfs路径只能挂载一个表,不能 嵌套,否则有权限漂移的问题
- · 建表的时候Hive会主动将路径的权限改成750
- 不要把表路径挂载到库路径上,很危险!!!

```
hive (bigdata_admin)> show create table hive_mysql_dbs;
ОК
createtab_stmt
CREATE TABLE 'hive_mysql_dbs'(
   `db_id` bigint,
  'desc' varchar(4000),
   `db_location_uri` varchar(4000),
   `name` varchar(128),
   `owner_type` varchar(10),
   'owner_name' varchar(128))
ROW FORMAT SERDE
   'org.apache.hadoop.hive.ql.io.orc.OrcSerde'
STORED AS INPUTFORMAT
   'org.apache.hadoop.hive.ql.io.orc.OrcInputFormat'
OUTPUTFORMAT
   'org.apache.hadoop.hive.ql.io.orc.OrcOutputFormat'
LOCATION
   hdfs://DClusterNmg4/user/hadoop/warehouse/bigdata_admin.db/hive_mysql_dbs"
TBLPROPERTIES (
  'LEVEL'='1',
  'PART_LIMIT'='10000'.
  'TTL'='60'.
  'orc.compress'='SNAPPY',
   'transient_lastDdlTime'='1527854855')
```



如 架构演进 <sup>应用 + 计划</sup>

02 权限设计 设计 + 使用

03 生命周期 设计 + 管理



## 数据分级

level 0:数据将永久保留,不进行过期处理。level 1:默认的级别,过期时间默认是60天。

## 过期判断依据

- 1) 表分区元数据是否有更新,
- 2)表分区对应的hdfs数据是否有更新取两者最新值最为更新时间

## 通知

根据过期规则判断数据过期后,将会通过邮件、钉钉、短信通知库的owenr和表的owenr。删除前7天开始发送提醒邮件。数据删除后,我们会再保留1个月再永久删除。

### Hive生命周期

## 生命周期通知方式

D++管理员

40-40-8105 報目日辞 text 医関島中心6-16

改作人。防な州(サキ単正型)、沙温、共同(総位下上部)



#### Hive生命周期执行日报 2018-05-16

#### 库信息

| Œ#   | Cwner | Ranger                           | 886                           |
|------|-------|----------------------------------|-------------------------------|
| best | хаси  | huanghui, fourticheng, linaxmiso | /user/caops/warehouse/test.65 |

#### 汇总信息

| 本次清理的表、分区 | 路径不规范内表、分区 | 特清理的表、全区 |
|-----------|------------|----------|
| D         | 1          | 0        |
| D .       | 1          | 0        |

#### 关注

- <u>S建長等地入生命背景</u>。體加度表TTL和LEVEL的设置功能。TTL:数据保管时间(单位: 天)。LEVEL:救援规则(0: 永久保留: 1: 算 进行生命角期)
- 2. 新以斯律書如果不相容性會需要。其使都会自动音響解以生會需要(Milittle60天、LEVEL-11、即即以生會需要会開始的天動夫事配理

严重建议赴议大家走来的时候每一设置TBLPFOPERTER(LEVEL-11/TTL-10)。如下:

CREATE EXTERNAL TABLE 'd tismiisoxo.t test 02'('id' string) ... TBLPRCPERTLES |'LEVEL'='1'."TTL ='7

- 3. 列斯爾福是古史斯伯松福提: 1) 表父这元就孤是古典史斯,引表父这时至的hife或孤是古典史斯
- 4. swan深簾的數据,将暂的不计人生命周期
- 5. 生命周距影響数据的一周自然所属库wner们表owner发射件通知
- 由于目前Mva基本上提展外部等。对方的location和户可以行意程序。因此包属库和各种location不一致。含有天常在明整件程度。依据大 进作Mva总统 http://witi.infra.sloci.ided.com/pages Merchages action/pages (1979-17-99)







[huanghui@bigdata-nng-client@1.nng@1 ~]\$ hive 18/86/82 16:29:34 INFO conf.HiveConf: [欢迎使用内蒙hive~] 新版本有任何问题,及时联系【大数据架构服务号】 【重要通知】 litve生命周期现已在内蒙开启了。请各位用户特别是库裹的管理员,关注最近收到的通知。 发送时间: 2018-04-09 生命周期使用手册: # HIVE: http://wiki.intra.xiaojukeji.com/pages/viewpage.action?pageId=110076079 # HDF5: http://wiki.intra.xiaojukeji.com/pages/viewpage.action?pageId=88081372 请各位用户务必行细阅读。 # 音录统一平台的"个人中心-我的权限-Hive权限管理"宣看Hive生命周期情况; # 登录统一平台的"权限申请-HDFS申请-HDFS生命周期"申请HDFS生命周期管理。 # 统一平台Hi.ve生命周期地址: http://dpp.intra.xigojukeji.com/index#/hone/myPermission/hivenanageri # 統一平台HDFS生命周期地址: http://dpp.intra.xiaojukeji.com/index#/apply/hdfsLifecycle 

Logging initialized using configuration in file:/usr/local/didi-hive-nng-1.2.1/conf/hive-log4j.properties

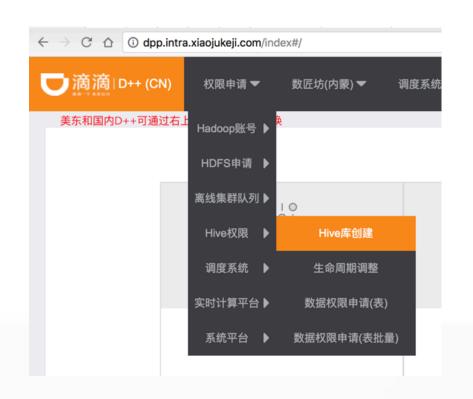






## 生命周期调整

·修改自己名下的表的生命周期,从level 1的生命周期修改为level 0将会发起审批流程









- 被判定为过期的表或者分区,对应的表或者分区元数据和底层的HDFS都会删除!
- 注意及时查看收到的关于生命周期的邮件、钉钉和短信提醒,过了1个月备份期数据无法恢复!
- 创建生命周期为永久的表,系统会主动发起一个审批流程,发送到直属leader审批,只有审批通过后才是永久保留的表

生命周期用户手册: http://wiki.intra.xiaojukeji.com/pages/viewpage.action?pageId=110076079





## 一些限制

- 普通用户不能创建库,需要到DPP中申请创建库,且只能使用生产账号申请,个人账号无法申请,库包括生产和测试库。
- •一个库下表数量(包括视图)不能超过8000,否则无法新建新表
- · HiveServer2的连接数我们设置了Quota, 默认是每个账号同时的连接数不能超过50, 超过的连接将连接失败。



CHAPTER 2

# 常见问题排查总结



### 问题1: Hive执行任务时,卡在Kill Command - /usr/local/hadoop-current/bin/hadoop job -kill job\_XXX

```
hive (bigdata_admin)> select count(1) from hive_mysql_dbs;
Query ID = huanghui_20180602170420_62bb698b-372b-4e58-97f8-8548d3d938b0

Total jobs = 1

Launching Job 1 out of 1

Number of reduce tasks determined at compile time: 1

In order to change the average load for a reducer (in bytes):
    set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
    set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
    set mapreduce.job.reduces=<number>
Starting Job = job_1527792880059_552666, Tracking URL = http://bigdata-nmg-hdprm01.nmg01:8088/proxy/application_1527792880059_552666/
Kill Command = /usr/local/hadoop-current/bin/hadoop_job -kill job_1527792880059_552666
```

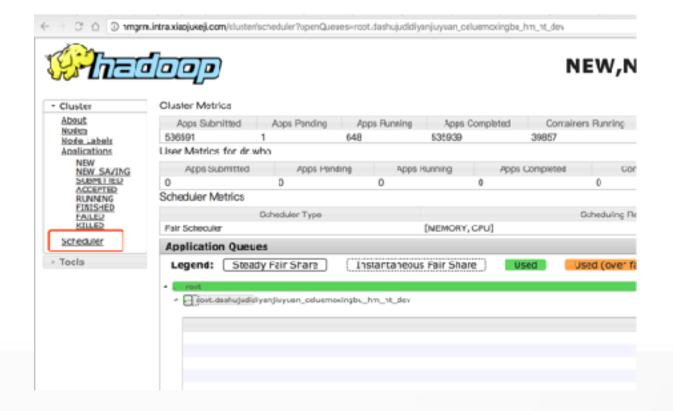
原因:这是在提交job,等待分配资源,才能继续往下执行解决方法:

- 1. 首先,访问yarn ui,查看自己队列的使用情况,如果队列资源占用比较慢,请耐心等待
- 2. 如果无法访问该地址,请查看星光代理,配置相应的代理

NOTE:此时任务还没有被kill,挺多的用户说自己的用户刚启动就被杀掉了!!!



### http://nmgrm.intra.xiaojukeji.com/cluster/scheduler





Accessible Node Labels: []

Used Resources: <memory:37532672, vCores:7414>
AM Used Resources: <memory:54272, vCores:14>
Demand Resources: <memory:37532672, vCores:7414>
Num Active Applications: 14
Num Pending Applications: 0

Min Resources: <memory:30720000, vCores:12000>
Max Resources: <memory:307200000, vCores:120000>
Max Running Applications: 2000
Steady Fair Share: <memory:307200000, vCores:0>
Instantaneous Fair Share: <memory:307200000, vCores:0>



#### 问题2: 如果我想自己先排查一下问题可以怎么做

可以使用如下的命令启动hive, 开启hive的DEBUG日志:

hive -hiveconf hive.root.logger=DEBUG,console

此时会有大量的日志输出,等日志不再输出时,就可以 执行相应的语句,看一下具体报错。

适用于使用客户端并对Hive有一定了解的用户。

```
IBAMN/92 1/ 15:46 LENDs common.LogNet.LS: Wing Mive-Site.zml found on CLASSMAIN at AUST/Local/digit-five-rmg-1.2.1/conf/mive-site.xml
  8/86/92 17:15:46 TNFO conf HiseConf:
    新版本有任何问题,及对联系《大数据领构服务号》
   H:vo生命周朝现已在内蒙开启了。请各位用户特别是挥表的管理员,关注最近收到的通知
                                                                              发送时间。2018-04-09
  HIVE: http://wiki.intra.xi.aojukeji.com/pages/viewpage.action?pageId=1.0076679
  HDFS: http://wiki.intra.xiaojukeji.com/pages/viewpage.action?pageTduR88R1372
  请名位用户多必仔细阅读。
  登录统一平台的"个人由心-我的权限。Hi ve収银管理"查看Hi ve生命周期情况:
  重要统一平台的"校院申请·IBFS申请·IDFS生命周期"申请IDFS生命周期管理。
   提一平台H.ve生命周期意址: http://dop.intrg.xigejukeji.com/indez#/hone/m/Perm:ssipr/hivengnogen
        - 平台HFFS生命用開始性: http://dop.intro.xionjukaji.com/index#/opgly/hdfsLtfacycle
  18.86/92 17:15:46 FEBIG session SessionState: SessionState user: mill.
Locging initialized using configuration is file:/usr/locgi/didi-bive-rmg-1.2.1/conf/hive-log4; properties
18/06/02 17:15:46 INFO SessionState:
oping initialized using configuration in file:/ww/local/didithiverong-1.2.1/conf/hiverlugfj.p-uperties
18/06/32 17:15:46 [EBUG parse. VariableSubstitution: Substitution is on: hive
18/06/92 17:15:46 [6805] Lib MutableMatriceFactory: field arg.apacha.hadaop.metrice2.lib MutableRata arg.apacha.hacaap.eocurity.UcerGroupInformation(UgiMetrice
loginSuccess with annotation Corp acade hadop metrics2 annotation.Netric(about=, always=false, samplehame=Cos, type=OffAUIT, value(kme=Time, value=(Rate of
successful kerberos logins and latency (milliseconds)]
18/06/92 17 15:46 CENUS 1 ib MutableMatricsFactory: field erg.apado.hedeop.matrics2.lib MutableMata org.apado.hedeop.accurity.Uscr5roup3nformation$UgiMatrics
loginfallure with annotation forg apache hadop metrics2 annotation.Netria(about-, always-false, sampleName-Ops, type-OEFAUIT, value/ame-Time, value-(Rate of
foiled kerteros logins and latercy (milliseconds)])
18/06/92 17:15:46 EENUS Lib MutableMetricsFactory: field org.apachs.hadoop.metrica2.lib MutableRate arg.apachs.hacoop.accurity.UberGroupInformation$UgiMetrica
.gethroups with enrotation Porg.epiche.folloge.netrics2.enrotation.Metric(@bout=, glwdys=false, sampleMane=ups, type=UthMUI, valueName=Ime, valueNetric(@bout=, glwdys=false, sampleMane=ups, type=UthMUI, valueName=Ime, valueName=I
18/86/92 17 15:46 DEBUG impl.MetricsSystemImpl: NgiMetrics, User and group related metrics
18/86/92 17:15:46 DEBUG util.KerberosNone: Kerberos krb5 configuration not found, setting default realm to empty
18/06/02 17:15:46 CEBUG security.Groups: Creating now Croups object
```



### 问题3: 使用hive -e 和hive -f的方式执行的结果为什么不一样,或者hive -e没有结果

hive -e 的方式使用比较容易出错,因为这相等于一个shell命令,我们都知道shell里面会涉及到很多的特殊字符的转义和替换。比如,'\*'符号是一个通配符,在执行命令前会被展开的。

使用hive -e后面得用双引号把sql包起来,里面涉及到诸如\*、斜杠等特殊字符的地方主义用转义符号转义。





#### 问题4: hive里面如何处理内容中的特殊字符,比如回车、制表符等导致的输出结果显示位置错乱

很多的Hive表都是将数据文件直接load到hdfs上,然后挂上hive表查询的。特别是一些包含中文内容的字段(comment等)里面进程或包含回车换行符。而Hive表默认是以换行符作为一条记录的结束的导,致查询出来的结果错乱。

#### 解决方法:

使用replace将这个特殊字符替换。比如,我的字段里面包含了回车,到时一条记录显示成了多行,可以按照如下的操作将回车替换成空格处理: regexp\_replace(content, '\n', ' ') as new\_content

在生产数据的时候对特殊字符进行处理。



# 问题5: physical memory limits. Current usage: 1.0 GB of 1 GB physical memory used; 2.9 GB of 2.1 GB virtual memory used. Killing container

出现这个问题的原因一般是参数设置得不合理导致的。

#### 建议参数:

set mapreduce.map.memory.mb=1024; set mapreduce.map.java.opts=-Xmx768m;

一般是先看这个异常是在map任务还是在reduce任务报出,设置对应的参数。



jvm参数应小于memory.mb, 否则由于物理内存超过限制,任务会被nodemanager kill掉。 推荐设置 java.opts=memory.mb \* 0.75



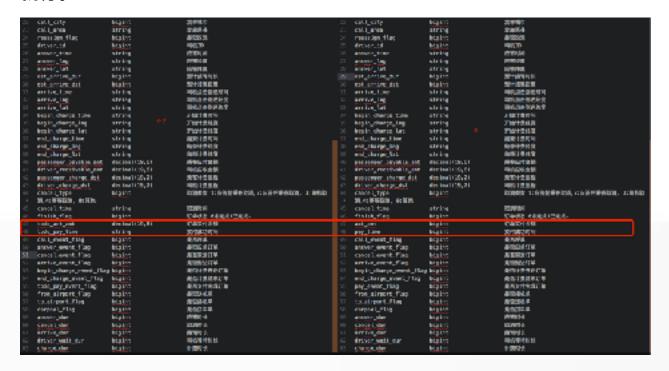
### 问题6: Failed with exception java.io.IOException:java.lang.ClassCastException:

出现这个问题,目前来看有两个原因,都是和ORC存储格式的表有关。

原因1:是在执行MR的过程中报的错误,一般是用户将表的SCHEMA修改,而对应的分区没有修改导致

原因2:在执行MR前就报错了,一般是语法有问题

这个问题主要是定位比较困难,对于原因1导致的,需要定位到修改的内容,进行修复。原因2导致的修改sql 就行。



#### case when online\_time = 0 or null then 0

报FAILED: ClassCastException org.apache.hadoop.hive.serde2.objectinspector.primitive.Writ ableVoidObjectInspector cannot be cast to org.apache.hadoop.hive.serde2.objectinspector.primitive.Boo leanObjectInspector

改为case when online\_time = 0 or online\_time is null then 0就好了



#### 问题7: 如何删除或者修改一个表的字段

Hive没有提供删除字段的方法,需要借助replace columns来实现。

ALTER TABLE table\_name REPLACE COLUMNS (col\_name data\_type [COMMENT col\_comment], ...)

如果想改某一个指定的分区,在表名后指定具体的分区即可。

比如表test有三个字段a,b,c, 通过如下的语句可以删除最后一个字段c.

ALTER TABLE test REPLACE COLUMNS (a int, b int);

如果是一个分区表,想把所有的分区也修改,需要在语句的末尾加上CASCADE

ALTER TABLE test REPLACE COLUMNS (a int, b int) CASCADE;



#### 问题8: 挂载分区好繁琐, 有没有快速挂载的方法

如果你的表分区使用的是Hive的默认分区格式,是有好方法的。此时需要具备如下的几个条件:分区路径的格式为:表路径/分区字段1=分区值/分区字段2=分区值

此时, 你可以使用:

MSCK REPAIR TABLE tablename;

Hive会去扫表表路径满足上面的格式要求的路径,然后检查分区是否已经挂载,如果没有的话,会给你挂载好,只需要一个命令就可以了。

如果分区路径是自定义的格式,比如:

表路径/分区值1/分区值2 (eg. /user/huanghui/warehouse/huanghuitest.db/test\_table/2017/01/01)

此时只能是使用alter table add partitions来挂载了。推荐是使用shell脚本来组装加分区的语句,然后存储到一个.hgl文件里面,使用hive -f 来执行,也是很快的。



### 更多问题可以查看我们的FAQ

http://wiki.intra.xiaojukeji.com/display/BigdataOfflineEngine/Hive+FAQ



CHAPTER 3

# 一些优化建议



## .hiverc文件的使用

对于一些每次登录都需要执行一遍的命令可以放到用户目录下的.hiverc文件里面。如果没有的话,创建一下就好了。

[huanghui@bigdata-nmg-client01.nmg01 ~]\$ cat .hiverc set mapreduce.job.queuename=root.jichupingtaibu-dashujujiagoubu.offline; set hive.cli.print.current.db=true;



# 查询分区表明确指定查询 的分区条件

明确指定要查询分区的范围,否则会查询所有的分区。

- 1) 查询会变慢
- 2) 对平台的压力也大



如果你使用spark-sql,可能会存在指定了分区过滤条件还是会查询所有分区的情况。尽量不使用concat和in等作为过滤条件



## 数据倾斜

set hive.exec.reducers.max=200; set mapred.reduce.tasks= 200; set hive.groupby.mapaggr.checkinterval=100000; set hive.groupby.skewindata=true; set hive.skewjoin.key=100000; set hive.optimize.skewjoin=true;

- --最大的Reducer个数
- --增大Reduce个数
- --这个是group的键对应的记录条数超过这个值则会进行分拆,值根据具体数据量设置
- --如果是group by过程出现倾斜 应该设置为true
- --这个是join的键对应的记录条数超过这个值则会进行分拆,值根据具体数据量设置
- --如果是join 过程出现倾斜 应该设置为true



具体情况具体分析



## 文件数相关

典型的报错是: Yarn Job Failed with Error: "Split metadata size exceeded 10000000" 往往是小文件过多导致。

在写入时进行小文件合并: set hive.merge.mapredfiles=true;

如果是ORC格式的文件,可以使用自带的命令对已有的文件进行合并: ALTER TABLE table\_name [PARTITION (partition\_key = 'partition\_value' [, ...])] CONCATENATE;







扫一扫群二维码,立刻加入该群。





扫码加入数据Talk二群



## **THANKS**