

目录

[一、Schema SCM 表结构梳理（对应生产BAOFOO_SCM） 2](#)

[1、 监控有没有暴力登录 3](#)

[2、 监控配置修改 3](#)

[3、 监控服务重启，重启impala服务。 4](#)

[二、HIVE元数据表（对应生产BAOFOO_HIVE） 5](#)

[1、 HIVE中库表数据量统计。 5](#)

[2、表大小统计 6](#)

三、Oozie元数据表（对应生产baofoo_oozie）

待补充

统计生产任务情况。

一、Schema SCM 表结构梳理（对应生产BAOFOO_SCM）

AUDITS	登录信息、服务，角色重启、配置更改
PROCESSES	进程信息。这里面有很多信息。开放的web端口。
HOSTS	主机信息，包括IP地址，所在机架，内存，CPU等信息
CLIENT_CONFIGS	客户端配置，里面有下载客户端配置的路径。
CONFIGS_AUD	配置审计表。
CLIENT_CONFIGS_TO_HOSTS	客户端配置文件与节点的对应关系
CLUSTERS	群集信息，CM管理多少个群集，就会有多条记录。
CM_VERSION	CM版本，服务安装时间，服务所在主机。
COMMANDS	命令记录，其中包括系统内部执行的命令。
COMMANDS_DETAIL	命令的详细内容。关联COMMANDS表。
COMMAND_SCHEDULES	CM内部命令执行调度元信息。
CONFIGS	配置表，改的配置，改之后的值，服务ID等。
CONFIGS_AUD	配置更改记录。
CREDENTIALS	Kerberos认证授权表，包含keytab文件
DIAGNOSTICS_EVENTS	诊断事件，关于parcels的激活，分发，停用有记录。
GLOBAL_SETTINGS	全局配置，包括自定义仪表盘信息
HOST_TEMPLATES	主机模板，模板名，所在主机。线上有hbase node和hdfs-yarn的模板。
HOST_TEMPLATE_TO_ROLE_CONF_GRP	主机模板和角色配置组的对应信息。
METRICS	各服务的监控项

PARCELS	parcels信息，是否可用，远程地址等等。
PARCEL_COMPONENTS	parcels名称与组件之间的映射关系。
REVISIONS	版本信息，CM变更之后，会产生一个新的版本。
ROLES	群集中的角色信息
ROLE_CONFIG_GROUPS	角色对应的角色组信息
SERVICES	服务名，页面显示的一些信息。
SNAPSHOT_POLICIES	快照策略（用户备份以及灾难恢复方面）
USERS	CM群集中的用户信息（启用认证时）
USER_AUTH_ROLES	用户与认证角色对应关系

该库记录了关于CM服务中的元数据、操作记录、用户登录、监控项采集、kerberos认证等等信息。

REVISIONS 每一次通过CM产生的变更都会产生一条记录。包括重启服务、修改服务、角色配置、修改CM配置。

AUDITS以及*_AUD为审计相关的信息记录。包括类如HOSTS，ROLES的元数据表。

AUDITS （审计表，登陆日志，重启服务，群集，激活、停用parcels等操作）。

1、监控有没有暴力登录

示例：以下SQL查询五分钟内登录失败次数超过10次的用户名，以及登录IP。

```

SELECT
*
FROM
(
SELECT
SUBSTRING_INDEX(MESSAGE, ':' , - 1) user_name,
IP_ADDRESS ip_addr,
count(1) AS login_counts
FROM
AUDITS
WHERE
AUDIT_TYPE = 'AUTHENTICATION'
AND ALLOWED = 0
AND TIMESTAMPDIFF(
MINUTE,
FROM_UNIXTIME(LEFT(CREATED_INSTANT, 10)),
CURRENT_TIMESTAMP ()
) < 5
GROUP BY
user_name,
ip_addr
) b
WHERE
b.login_counts > 10;

```

```

+-----+-----+-----+
| user_name | ip_addr   | login_counts |
+-----+-----+-----+
| adminf    | 10.6.52.64 |          14 |
+-----+-----+-----+
1 row in set (0.00 sec)

```

2、监控配置修改

```

SELECT
t3.DISPLAY_NAME,
t1.ATTR,
t1.`VALUE`
FROM
CONFIGS_AUD t1
LEFT JOIN REVISIONS t2 ON t1.REV = t2.REVISION_ID
LEFT JOIN SERVICES t3 ON t1.SERVICE_ID = t3.SERVICE_ID
WHERE
TIMESTAMPDIFF(
MINUTE,
FROM_UNIXTIME(LEFT(t2.`TIMESTAMP`, 10)),
CURRENT_TIMESTAMP ()
) < 5;

```

HUE的http端口属性被改成8898

```

+-----+-----+-----+
| DISPLAY_NAME | ATTR           | VALUE |
+-----+-----+-----+
| Hue          | hue_http_port | 8898  |
+-----+-----+-----+

```

3、监控服务重启，重启impala服务。

```

SELECT
SUBSTRING_INDEX(t3.message, ':' , - 1),
t2.INFO,
t2.IP_ADDRESS
FROM
(
SELECT
GROUP_CONCAT(
MESSAGE
ORDER BY
CREATED_INSTANT SEPARATOR ' 结果:'
) INFO,
GROUP_CONCAT(IP_ADDRESS) IP_ADDRESS
FROM
AUDITS t1
WHERE
CREATED_INSTANT > REPLACE (
UNIX_TIMESTAMP(
DATE_ADD(NOW(3), INTERVAL - 50 MINUTE)
),
',',
';',
)
AND AUDIT_TYPE = 'COMMAND_SERVICE'

```

```

) t2
LEFT JOIN (
SELECT
message,
IP_ADDRESS,
CREATED_INSTANT
FROM
AUDITS
WHERE
AUDIT_TYPE = 'AUTHENTICATION'
) t3 ON t3.IP_ADDRESS = t2.IP_ADDRESS
ORDER BY
t3.CREATED_INSTANT DESC
LIMIT 1;

```

二、HIVE元数据表（对应生产BAOFOO_HIVE）

与上述不同，这不属于Cloudera 公司Hadoop发行版独有的。

DBS	数据库信息，DB名，所在位置。
COLUMNS_V2	表的列信息，注释，列名，列的数据类型，列在表中的位置。通过CD_ID与CDS表关联
TBL_PRIVS	表的授权信息。
TABLE_PARAMS	表的基本信息，最后更改时间，总大小，原生数据大小，行数，文件数量
SERDE_PARAMS	字符分隔符，虚拟化格式信息。通过SERDE_ID与SERDES关联
SDS	表的属性信息，输入格式（表存储类型），是否压缩，所在位置，占桶的数量，等。通过SERDE_ID与SERDES关联
SERDES	虚拟化和反虚拟化的信息。
CDS	只有一列，表的唯一ID
TBLS	存储Hive表、视图、索引表的基本信息。创建时间，上次访问时间，所有者、表名，视图HQL语句
PARTITION_KEYS	表的分区信息，表名，分区键注释，分区键名，类型，联合分区所在位置
PARTITION_KEY_VALS	已有分区信息。
PARTITION_PARAMS	分区属性信息，某个分区的文件数量，总大小，最后访问时间等。
PARTITIONS	分区的基本信息，分区ID，分区创建时间，最后访问时间，分区名，分区存储ID，表ID
VERSION	存储Hive版本的元数据表，如果该表不存在，启动hive-cli的时候会报Table 'hive.version' doesn't exist”

1、HIVE中库表数据量统计。

```

SELECT d.`NAME` "库",count(1) "表数量" from TBLS t RIGHT JOIN DBS
d ON t.DB_ID = d.DB_ID group BY d.DB_ID ;

```

schema_name	num_of_tables
xy_wulichuang	2975
xy_dm	438
xy_app_hive	296
xy_wangxiaoling	290
xy_linsanji	211
xy_md_tpc	186
xy_jiangyuande	184
xy_app_spark	147
credit_dfp	110
bf_gongshaojie	90
xy_ods	83
baofoo_fi	55
bf_yangzhensheng	52
baofoo_ma	52
tmp_baofoo_fi	47
tmp_baofoo_ma	47
tmp_baofoo_cm	42
baofoo_cm	39
bf_sor	26
tmp_credit_center	25
baofoo_admin	25
credit_center	25

2、表大小统计

```

SELECT
round(CAST(a.PARAM_VALUE AS SIGNED)/1024/1024/1024,0)
v,
b.TBL_NAME,
c.`NAME`
FROM
TABLE_PARAMS a
LEFT JOIN TBLS b ON a.TBL_ID = b.TBL_ID
LEFT JOIN DBS c ON b.DB_ID = c.DB_ID
WHERE
a.PARAM_KEY = 'totalSize' -- AND c.`NAME` NOT like 'tmp%'
ORDER BY
v DESC;

```