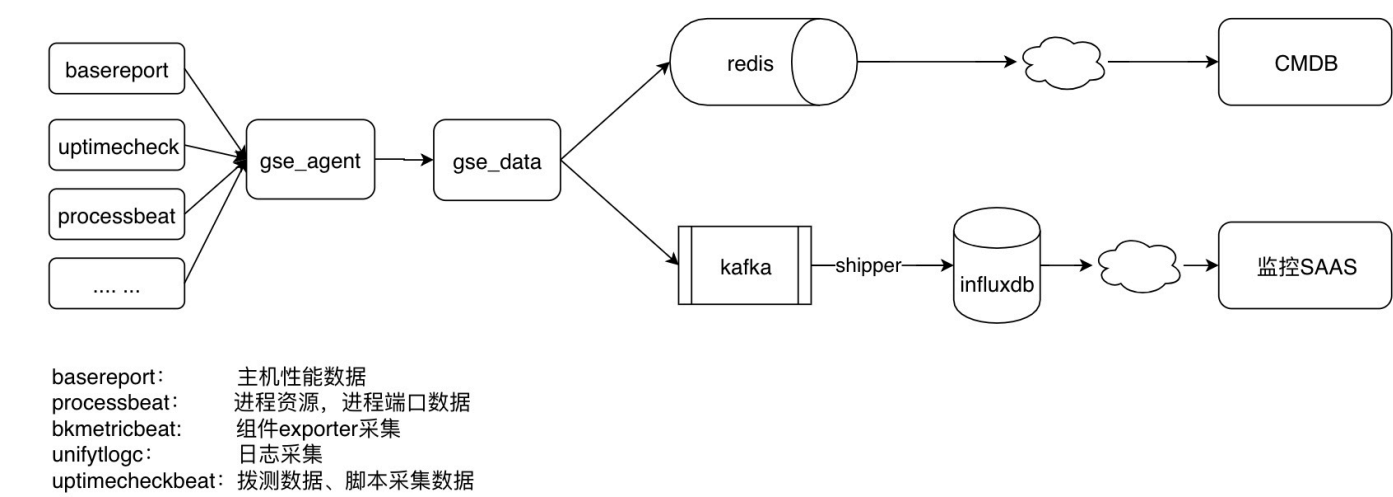


无数据问题排查

数据采集上报流转图

采集器 -> Agent(data_server) -> kafka -> shipper -> influxdb



主机性能数据上报:

采集器进程: basereport

Agent进程: gse_agent

数据上报进程: gse_data

上报流程:

- 主机性能数据是由basereport进程采集, 通过Agent的通道上报给dataserver(gse_data)
- dataserver(gse_data)读取zk中的配置, 将数据分发给不同的管道, redis和kafka
- 其中redis中的数据供CMDB读取和展示, kafka的数据由数据平台入库到时序db, 监控读取时序db的数据并展示

确认各个模块进程是否正常

首先确认cmdb上主机是否有数据。

cmdb无数据: 数据采集侧 问题定位

cmdb有数据: 数据处理侧 问题定位, 则直接跳到【查看dataserver的状态】步骤

查看采集器basereport进程状态

```
# 查看进程
ps -ef|grep basereport

# 查看日志
tail -f /var/log/gse/basereport
```

常见问题：

- 机器上出现多个basereport，机器多次安装，前一个没有杀掉
- 不正常的话，去节点管理把basereport托管起来。（正常情况basereport是安装完agent自动托管的）

查看Agent的连接状态

```
# 查看进程
ps -ef|grep gse_agent

# 查看日志(一般日志文件在/var/log/gse目录下)
tail -f /var/log/gse/agent-日期-*.log

# 查看gse_agent 与 gse_data 是否连接上（重要），有正常连接ESTABLISHED则ok
netstat -tnp | grep 58625
# 若存在proxy，proxy机器：检测58625端口同上
```

常见问题：

- 机器有多IP：确认/usr/local/gse/agent/etc/agent.conf里面identityip的值和cmdb的内网ip是否一致
- 机器时间不同步：如果出现部分机器无数据，需要考虑无数据的机器的时间是否和蓝鲸服务器同步

查看dataserver的状态

以下操作均从中控机开始，并且已经完成了source utils.fc

```
# 去到gse机器，如果有多台，都需要check
ssh $GSE_IP

# 查看进程状态
ps -ef|grep gse_data

# 查看进程日志
tail -f $BK_HOME/logs/gse/data-日期-*.log
```

```
# 查看gse_data 和 kafka 是否连接上 (重要)
netstat -tnp | grep 9092 | grep gse_data
```

data_server 上报数据到 kafka，是根据zk里的配置传送的

查看zk配置

根据配置，找到dataid对应的topic名称

```
# 去到ZK机器，并进到zk的目录
ssh $ZK_IP
cd $BK_HOME/service/zk

# 查看对应dataid下的配置，比如查看dataid为1001的配置
./bin/zkCli.sh -server zk.service.consul:2181 get /gse/config/etc/dataserver
/data/1001
```

正常会有类似如下配置（json格式）

```
[
  {
    "server_id": -1,
    "data_set": "snapshot",
    "partition": 1,
    "cluster_index": 0,
    "biz_id": 2,
    "msg_system": 1    # 1: kafka  3:redis
  },
  ...
]
```

根据kafka中的topic生成规则，

取出data_set 和 biz_id 两个字段，拼接在一起，例如：上面的配置，则topic为snapshot2

查看kafka对应的topic是否有数据

```
# 从ZK配置查看kafka集群broker
ssh $ZK_IP
cd $BK_HOME/service/zk
./bin/zkCli.sh -server zk.service.consul:2181 ls /common_kafka/brokers/ids

# 去到kafka机器，并进到kafka的目录
```

```
ssh $KAFKA_IP
cd $BK_HOME/service/kafka

# 拉取数据，正常情况下，等待一分钟内会有数据输出，如果没有数据输出，可以查看下topic的配置info
./bin/kafka-console-consumer.sh --bootstrap-server kafka.service.consul:9092
--topic snapshot2

# 查看topic的配置是否正常
./bin/kafka-topics.sh --zookeeper zk.service.consul:2181/common_kafka --describe --topic snapshot2
```

常见问题：

- topic配置有异常，no leader, or, leader not available, 需修复kafka的topic

查看influxdb是否有数据

```
# 去到influxdb机器
ssh $INFLUXDB_IP

# 进到influxdb终端，查看最新入库数据的时间
influx -host influxdb.service.consul -port 5260
> use system_2
> precision rfc3339
> select * from system_cpu_summary_2 order by time desc limit 1;
```

查看influxdb里是否有数据(一般情况下，监控SAAS显示无数据，influxdb里面就没有数据)

查看dataapi

数据在入库的时候，需要根据IP入库到不同的database下（不同业务有不同的database），所以需要查看ip到业务的对应关系，由update_cc_cache进程更新这个关系

```
# 查看 update_cc_cache 进程是否正常
ssh $BKDATA_IP
supervisorctl -c $BK_HOME/etc/supervisor-bkdata-dataapi.conf status update_cc_cache

# 如果出现Exited too quickly (process log may have details)
workon dataapi

# 单独启动update_cc_cache，看是否正常
$BK_HOME/bkdata/dataapi/bin/start_update_cc_cache.sh
```

```
# 查看输出，如果输出正常，则重新启动dataapi下的update_cc_cache
supervisorctl -c $BK_HOME/etc/supervisor-bkdata-dataapi.conf restart update_cc_cache
# 如果输出报错，需要先解决它！
```

查看databus相关进程是否正常

数据入库到db是由databus下的不同进程执行，比如入库influxdb，则是由进程databus_tsdb负责，

```
# 查看相关进程状态
supervisorctl -c $BK_HOME/etc/supervisor-bkdata-databus.conf status
```

异常则需要修复它

总体

可以从监控saas的healthz页面，看数据链路整体是否正常