# Canal简介及配置说明

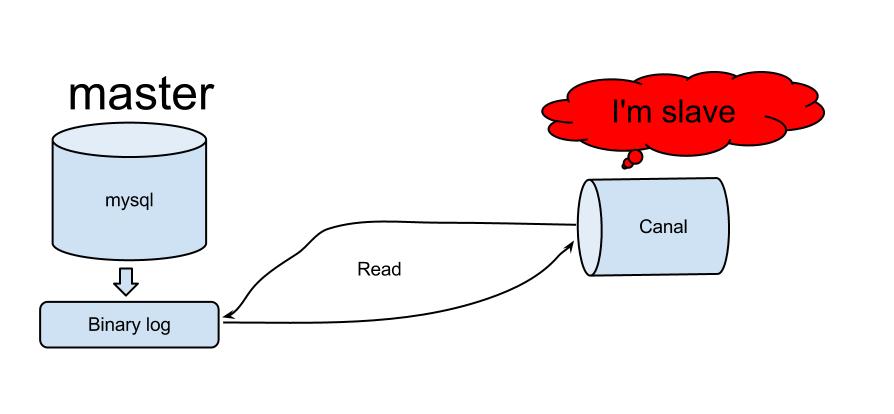
**1.简介**

canal是纯Java开发的，基于数据库增量日志解析，提供增量数据订阅&消费，目前主要支持了mysql。

原理相对比较简单：

1. canal模拟mysql slave的交互协议，伪装自己为mysql slave，向mysql master发送dump协议
2. mysql master收到dump请求，开始推送binary log给slave(也就是canal)
3. canal解析binary log对象(原始为byte流)

如下图：



**2.mysql要求**

   a. 目前canal支持mysql 5.5版本以下，对mysql5.6暂不支持，(mysql4.x版本没有经过严格测试，理论上是可以兼容)

   b. canal的原理是基于mysqlbinlog技术，所以这里一定需要开启mysql的binlog写入功能，并且配置binlog模式为row.

[mysqld]

log-bin=mysql-bin #添加这一行就ok

binlog-format=ROW #选择row模式

server\_id=1 #配置mysql replaction需要定义，不能和canal的slaveId重复

下面两个配置强烈建议配置，这样可以减小binlog的大小，忽略不需要关注的库的binlog

binlog-do-db = oos\_device #配置需要同步的库

binlog-ignore-db = mysql #配置不需要同步的库

   c.  canal的原理是模拟自己为mysql slave，所以这里一定需要做为mysql slave的相关权限

CREATE USER canal IDENTIFIED BY 'canal';

GRANT SELECT, REPLICATION SLAVE, REPLICATION CLIENT ON \*.\* TO 'canal'@'%';

-- GRANT ALL PRIVILEGES ON \*.\* TO 'canal'@'%' ;

FLUSH PRIVILEGES;

     针对已有的账户可通过grants查询权限：

show grants **for** 'canal'

**3.部署**

**1. 获取发布包**

访问：<https://github.com/alibaba/canal/tree/gh-pages/download> ，会列出所有历史的发布版本包

下载方式，比如以1.0.4版本为例子：

wget https://raw.github.com/alibaba/canal/gh-pages/download/canal.deployer-1.0.4.tar.gz

**2. 目录结构**

解压缩发布包后，可得如下目录结构：

drwxr-xr-x 2 jianghang jianghang  136 2013-03-19 15:03 bin

drwxr-xr-x 4 jianghang jianghang  160 2013-03-19 15:03 conf

drwxr-xr-x 2 jianghang jianghang 1352 2013-03-19 15:03 lib

drwxr-xr-x 2 jianghang jianghang   48 2013-03-19 15:03 logs

**3. 启动/停止**

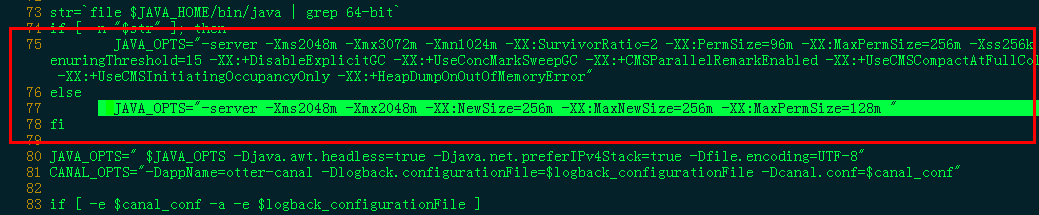
   linux启动 :

sh startup.sh

   linux停止：

sh stop.sh

**几点注意：**

1. linux启动完成后，会在bin目录下生成canal.pid，stop.sh会读取canal.pid进行进程关闭
2. startup.sh默认读取系统环境变量中的which java获得JAVA执行路径，需要设置PATH=$JAVA\_HOME/bin环境变量
3. canal的内存设置在start.sh中第75或77行，（建议将75和77行的内存设置值保持相同）如下：

4． 每次更改mysql的binlog文件名称及位置时，记得把../canal/conf/oos\_device目录下的meta.dat文件删除或重命名。

因为canal每次启动时，会检查meta.dat文件是否存在，

如果文件存在会以该文件记录的binlog位置开始同步数据；

如果该文件不存在，那么会根据../canal/conf/oos\_device目录下instance.properties文件中配置的信息生成新的meta.dat文件

**4.配置**

properties配置分为两部分：

* canal.properties  (系统根配置文件，/conf/canal.properties)

需要关注的配置项如下：

1. canal.port：canal server提供socket服务的端口，建议配置32121
2. canal.instance.detecting.enable：是否开启心跳检查，建议配置true
3. canal.instance.detecting.sql：心跳检查sql，建议配置select 1 from 目标库.目标表
4. canal.destinations：当前server上部署的instance列表，比如需要获取epg库的表更新则可以创建一个名为epg的instance，则配置canal.destinations = epg，同时需要在/conf/下新建epg目录，并在其中创建instance.properties即可

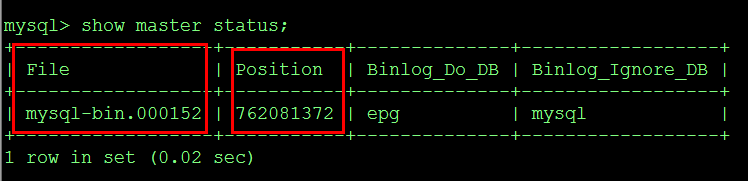
* instance.properties  (instance级别的配置文件，每个instance一份，/conf/实例名称/instance.properties)

需要关注的配置项如下：

1. canal.instance.mysql.slaveId：mysql集群配置中的serverId概念，需要保证在当前mysql集群中id唯一
2. canal.instance.master.address：mysql主库链接地址，ip：端口，如：127.0.0.1:3306
3. canal.instance.master.journal.name:mysql当前binlog文件名称,

如：mysql-bin.000152，获取方式：

* 登陆到第二步配置的mysql上
* 执行命令：show master status;见下图：



上图中File下面的信息就是binlog的文件名称。

1. canal.instance.master.position：mysql当前binlog的位置,如：762081372

上图中Position下面的数字就是binlog文件当前的位置。

1. canal.instance.defaultDatabaseName：mysql链接时默认数据库，建议设置成需要同步的库名
2. canal.instance.connectionCharset：mysql数据解析编码，建议和数据库编码方式保持一致
3. canal.instance.filter.regex：mysql数据解析关注的表，Perl正则表达式。多个正则之间以逗号(,)分隔，转义符需要双斜杠(\\)，例如只关注mysql库里的user表的更新，则配置为mysql.user