今晚带你解密阿里巴巴Canal

主讲：Cat 老师

**北京动力节点教育科技有限公司**  
**中国 • 北京**

# 1、What is Canal？

canal [kə'næl]，中文翻译为 水道/管道/沟渠/运河，主要用途是用于 MySQL 数据库增量日志数据的订阅、消费和解析，是阿里巴巴开发并开源的，采用Java语言开发；

历史背景是早期阿里巴巴因为杭州和美国双机房部署，存在跨机房数据同步的业务需求，实现方式主要是基于业务 trigger（触发器） 获取增量变更。从2010年开始，阿里巴巴逐步尝试采用解析数据库日志获取增量变更进行同步，由此衍生出了canal项目；

Github：<https://github.com/alibaba/canal>

# 2、工作原理

## 传统MySQL主从复制工作原理



从上层来看，复制分成三步：

MySQL的主从复制将经过如下步骤：

1、当 master 主服务器上的数据发生改变时，则将其改变写入二进制事件日志文件中；

2、salve 从服务器会在一定时间间隔内对 master 主服务器上的二进制日志进行探测，探测其是否发生过改变，如果探测到 master 主服务器的二进制事件日志发生了改变，则开始一个 I/O Thread 请求 master 二进制事件日志；

3、同时 master 主服务器为每个 I/O Thread 启动一个dump Thread，用于向其发送二进制事件日志；

4、slave 从服务器将接收到的二进制事件日志保存至自己本地的中继日志文件中；

5、salve 从服务器将启动 SQL Thread 从中继日志中读取二进制日志，在本地重放，使得其数据和主服务器保持一致；

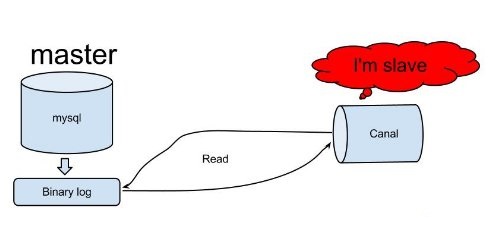
6、最后 I/O Thread 和 SQL Thread 将进入睡眠状态，等待下一次被唤醒；

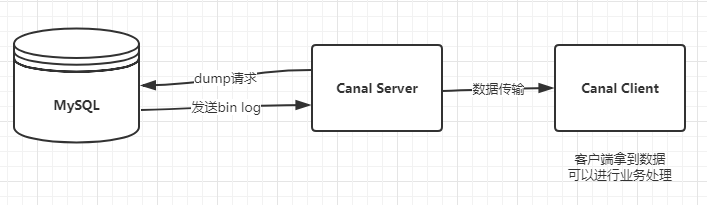
## canal 工作原理

1、canal 模拟 MySQL slave 的交互协议，把自己伪装为 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 (即canal )；

3、canal 解析 binary log 对象 (原始数据为byte流)





# 3、Canal使用场景

Canal是基于MySQL变更日志增量订阅和消费的组件，可以使用在如下一些一些应用场景：

数据库实时备份

业务cache刷新

search build

价格变化等重要业务消息

带业务逻辑的增量数据处理

跨数据库的数据备份(异构数据同步)，

例如mysql => oracle，mysql=>mongo，mysql =>redis，

mysql => elasticsearch等；

当前canal 主要是支持源端 MySQL（也支持mariaDB），版本包括 5.1.x , 5.5.x , 5.6.x , 5.7.x , 8.0.x；

# Canal搭建环境

1、准备好MySQL运行环境；

2、开启 MySQL的binlog写入功能，配置 binlog-format 为 ROW 模式，my.cnf中配置如下:

[mysqld]

log-bin=mysql-bin #开启 binlog

binlog-format=ROW #选择 ROW 模式

server\_id=1 #配置MySQL replaction需要定义，不要和canal的 slaveId重复

3、授权canal连接MySQL账号具有作为MySQL slave的权限, 如果已有账户可直接 grant授权：

启动MySQL服务器；

登录mysql：./mysql -uroot -p -h127.0.0.1 -P3306

CREATE USER canal IDENTIFIED BY 'canal';

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

FLUSH PRIVILEGES;

4、下载 canal部署程序

Wget <https://github.com/alibaba/canal/releases/download/canal-1.1.4/canal.deployer-1.1.4.tar.gz>

tar -zxvf canal.deployer-1.1.4.tar.gz -C /usr/local/canal.deployer-1.1.4

5、配置修改

vim conf/example/instance.properties

主要是修改配置文件中与自己的数据库配置相关的信息；

6、启动Canal

./startup.sh

7、查看进程：

ps -ef | grep canal

8、查看 server 日志

cat logs/canal/canal.log

9、查看 instance 的日志

vi logs/example/example.log

10、关闭Canal

./stop.sh

canal server的默认端口号为：11111，如果需要调整的话，可以去到\conf目录底下的canal.properties文件中进行修改；

**相关命令：**

#是否启用了日志

show variables like 'log\_bin';

#怎样知道当前的日志

show master status;

#查看mysql binlog模式

show variables like 'binlog\_format';

#获取binlog文件列表

show binary logs;

#查看当前正在写入的binlog文件

show master status\G

#查看指定binlog文件的内容

show binlog events in 'mysql-bin.000002';

注意binlog日志格式要求为row格式；

Binlog的三种基本类型分别为：

ROW模式 除了记录sql语句之外，还会记录每个字段的变化情况，能够清楚的记录每行数据的变化历史，但是会占用较多的空间，需要使用mysqlbinlog工具进行查看；

STATEMENT模式只记录了sql语句，但是没有记录上下文信息，在进行数据恢复的时候可能会导致数据的丢失情况；

MIX模式比较灵活的记录，例如说当遇到了表结构变更的时候，就会记录为statement模式。当遇到了数据更新或者删除情况下就会变为row模式；

启动了canal的server之后,便是基于java的客户端搭建了；

# 5、Canal应用案例实践

MySQL -> Canal -> RabbitMQ -> MongoDB

# 6、Canal进阶

在实际开发中，如果只有一台canal机器作为server，当该台机器挂掉之后，服务就会终止，那么此时我们便需要引入集群部署方式，Client的集群与Server的集群（这部分内容在VIP课程中介绍）。