|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | | |  |
|  | | |
|  | | |
| 方圆ok1 | | | |
|  | **FusionInsight Stream**  **V100R002C50**  **CQL语法** | | 附件1-16K |
|  | |
| **文档版本** | **01** |
| **发布日期** | **2015-06-30** |
| **华为技术有限公司** | |

|  |
| --- |
| 版权所有 © 华为技术有限公司2015。 保留一切权利。  非经本公司书面许可，任何单位和个人不得擅自摘抄、复制本文档内容的部分或全部，并不得以任何形式传播。  商标声明  附件3-版权声明页图和其他华为商标均为华为技术有限公司的商标。  本文档提及的其他所有商标或注册商标，由各自的所有人拥有。  注意  您购买的产品、服务或特性等应受华为公司商业合同和条款的约束，本文档中描述的全部或部分产品、服务或特性可能不在您的购买或使用范围之内。除非合同另有约定，华为公司对本文档内容不做任何明示或默示的声明或保证。  由于产品版本升级或其他原因，本文档内容会不定期进行更新。除非另有约定，本文档仅作为使用指导，本文档中的所有陈述、信息和建议不构成任何明示或暗示的担保。 |

|  |  |
| --- | --- |
| 华为技术有限公司 | |
| 地址： | 深圳市龙岗区坂田华为总部办公楼 邮编：518129 |
| 网址： | http://[enterprise](http://enterprise.huawei.com).huawei.com |

目 录

[1 前言 1](#_Toc424293328)

[2 系统功能列表 2](#_Toc424293329)

[3 关键概念定义 5](#_Toc424293330)

[4 CQL应用程序开发流程 6](#_Toc424293331)

[5 数据类型 8](#_Toc424293332)

[5.1 基本数据类型 8](#_Toc424293333)

[5.2 空值 11](#_Toc424293334)

[6 CQL语法定义 12](#_Toc424293335)

[6.1 应用程序操作语句 12](#_Toc424293336)

[6.1.1 Drop 12](#_Toc424293337)

[6.1.2 Submit 13](#_Toc424293338)

[6.1.3 Explain 14](#_Toc424293339)

[6.1.4 Load 14](#_Toc424293340)

[6.1.5 Deactive 15](#_Toc424293341)

[6.1.6 Active 15](#_Toc424293342)

[6.1.7 Rebalance 16](#_Toc424293343)

[6.2 流操作语句 17](#_Toc424293344)

[6.2.1 Create Input Stream 17](#_Toc424293345)

[6.2.2 Create Output Stream 20](#_Toc424293346)

[6.3 查询语句 23](#_Toc424293347)

[6.3.1 Insert 23](#_Toc424293348)

[6.3.2 SELECT 25](#_Toc424293349)

[6.3.2.1 Select Clause 25](#_Toc424293350)

[6.3.2.2 From Clause 26](#_Toc424293351)

[6.3.2.3 Where Clause 32](#_Toc424293352)

[6.3.2.4 Group By Clause 32](#_Toc424293353)

[6.3.2.5 Having Clause 33](#_Toc424293354)

[6.3.2.6 Order By Clause 34](#_Toc424293355)

[6.3.2.7 Limit Clause 34](#_Toc424293356)

[6.3.2.8 Parallel Clause 35](#_Toc424293357)

[6.3.3 Subquery Clause 35](#_Toc424293358)

[6.3.4 DataSource 36](#_Toc424293359)

[6.3.4.1 Create DataSource 36](#_Toc424293360)

[6.3.4.2 DataSourceQuery 38](#_Toc424293361)

[6.3.5 User Defined Operator 39](#_Toc424293362)

[6.3.5.1 Create User Operator 39](#_Toc424293363)

[6.3.5.2 User Operator Query 41](#_Toc424293364)

[6.4 命令语句 42](#_Toc424293365)

[6.4.1 Set/Get 42](#_Toc424293366)

[6.4.2 Add Jar 43](#_Toc424293367)

[6.4.3 Add File 43](#_Toc424293368)

[6.4.4 Create Function 44](#_Toc424293369)

[6.4.5 Drop Function 45](#_Toc424293370)

[6.4.6 Show Applications 45](#_Toc424293371)

[7 窗口 47](#_Toc424293372)

[7.1 元组窗口 50](#_Toc424293373)

[7.2 范围窗口 52](#_Toc424293374)

[7.3 自然天窗口 55](#_Toc424293375)

[8 表达式 57](#_Toc424293376)

[8.1 通用表达式 58](#_Toc424293377)

[8.2 表达式常量 59](#_Toc424293378)

[8.3 特殊表达式 59](#_Toc424293379)

[8.3.1 Cast 59](#_Toc424293380)

[8.3.2 Case when 60](#_Toc424293381)

[8.3.3 Like 61](#_Toc424293382)

[8.3.4 Between 61](#_Toc424293383)

[8.3.5 In 62](#_Toc424293384)

[8.4 表达式数据类型转换规则 63](#_Toc424293385)

[8.4.1 算术表达式 63](#_Toc424293386)

[8.4.2 CAST表达式 63](#_Toc424293387)

[9 用户自定义接口 65](#_Toc424293388)

[9.1 数据序列化/反序列化 65](#_Toc424293389)

[9.2 流数据读取 72](#_Toc424293390)

[9.3 流数据写入 78](#_Toc424293391)

[9.4 UDF函数 82](#_Toc424293392)

[9.5 数据源 84](#_Toc424293393)

[9.6 自定义算子 89](#_Toc424293394)

[10 系统内置接口实现 95](#_Toc424293395)

[10.1 序列化/反序列化 95](#_Toc424293396)

[10.1.1 SimpleSerde 95](#_Toc424293397)

[10.1.2 KeyValueSerDe 96](#_Toc424293398)

[10.1.3 CSVSerDe 97](#_Toc424293399)

[10.1.4 BinarySerDe 98](#_Toc424293400)

[10.2 数据读取 99](#_Toc424293401)

[10.2.1 KafkaInput 99](#_Toc424293402)

[10.2.2 MQInput 101](#_Toc424293403)

[10.2.3 TCPClientInput 102](#_Toc424293404)

[10.2.4 RamdomGen 102](#_Toc424293405)

[10.3 数据写入 103](#_Toc424293406)

[10.3.1 KafkaOutput 103](#_Toc424293407)

[10.3.2 MQOutput 104](#_Toc424293408)

[10.3.3 TCPClientOutput 105](#_Toc424293409)

[10.3.4 ConsoleOutput 106](#_Toc424293410)

[10.4 函数 107](#_Toc424293411)

[10.4.1 字符处理函数 107](#_Toc424293412)

[10.4.2 时间处理函数 109](#_Toc424293413)

[10.4.3 数学函数 115](#_Toc424293414)

[10.4.4 类型转换函数 116](#_Toc424293415)

[10.4.5 聚合函数 121](#_Toc424293416)

[10.5 数据源 128](#_Toc424293417)

[10.5.1 RDB数据源 128](#_Toc424293418)

[11 CQL其他配置参数 131](#_Toc424293419)

[12 CQL异常码 134](#_Toc424293420)

[13 CQL Java开发API说明 140](#_Toc424293421)

[14 CQL关键字列表 141](#_Toc424293422)

# 前言

CQL（Continuous Query Language），持续查询语言，用于数据流上的查询语言。相对于传统的SQL，CQL加入了窗口的概念，使得数据可以一直保存在内存中，由此可以快速进行大量内存计算，CQL的输出结果为数据流在某一时刻的计算结果。

CQL是建立在Storm基础上的类SQL查询语言，它解决了Storm原生API使用复杂，上手难度高，很多基本功能缺失的问题，提升了流处理产品的易用性。

在CQL设计语法之初，通过参考市面上现有的CEP产品的语法，发现这些产品都不算是全部的SQL语句，即仅仅使用SQL语句还不能运行，还必须依靠一些客户端的代码。 这样就给使用带来了一些不便， 用户必须学习客户端API，比较繁琐，上手难度也比较大。

所以，CQL设计目标就是，用纯粹的SQL语句再加上一些命令，就可以完成所有的任务发布以及执行，这样，就可以通过SQL接口，直接进行任务的下发，统一了客户端接口。对于有一定SQL基础的用户，只需要掌握一些CQL比较特殊的语法，比如窗口或者流定义的语法，就可以写出可运行的CQL程序，大大降低了上手难度。

# 系统功能列表

下表列出了系统的功能列表。

CQL功能列表

| 模块 | 功能 | 说明 |
| --- | --- | --- |
| 数据类型 | DataType | 支持int、String、long、float、double、boolean基本类型和time，date，timestamp时间类型以及decimal数字类型。 |
| 应用程序操作语句 | Create Input Stream | 创建输入流，包含数据读取方式和数据反序列化方法。 |
| Create Output Stream | 创建输出流，包含数据发送方式和数据序列化方法。 |
| Explain | 查看并导出执行计划。 |
| Load Application | 加载执行计划。 |
| Submit Application | 提交应用程序。 |
| Drop Application | 删除应用程序。 |
| 查询语句 | Select | 查询语句。 |
| Insert as select | 查询结果重定向到另外一个流。 |
| multiInsert | 将一个流的数据经过不同的处理，发送到其他多个流当中。 |
| Window | 窗口。 |
| Join | 支持inner，left，right，full，cross五种Join类型。 |
| Group by | 分组。 |
| Order by | 对相同批次的输出数据进行排序。 |
| Filter Before Window | 数据进入窗口之前的过滤。 |
| Where | 数据进入窗口之后的过滤。 |
| Having | 聚合之后的过滤。 |
| Limit | 限制输出数量。 |
| DataSource | 数据源。 |
| User Operator | 用户自定义算子 |
| subQuery | 子查询。 |
| 命令 | Show applications | 查看已经提交的应用程序。 |
| Set/Get | 设置或读取配置属性。 |
| Add jar | 添加Jar包。 |
| Add file | 添加文件。 |
| Create function | 注册函数。 |
| Drop Function | 删除函数。 |
| 操作符 | 表达式 | 支持常见And，OR，IS NULL，NULL等，  支持'<','>','<=','>=', '=','==','!=',like, case when。 |
| 自定义数据序列化规则 | 通过自定义数据序列化规则， 实现对不同数据的解析。 |
| 自定义数据读取规则 | 拓展数据读取方式，实现不同文件或者数据类型的读取。 |
| 自定义数据写入规则 | 拓展数据发送方式，实现计算结果向不同的设备发送。 |

# 关键概念定义

* 流：流是一组（无穷）元素的集合，流上的每个元素都属于同一个schema；每个元素都和逻辑时间有关；即流包含了元组和时间的双重属性。留上的任何一个元素，都可以用Element<tuple,Time>的方式来表示，tuple是元组，包含了数据结构和数据内容，Time就是该数据的逻辑时间。
* Window：窗口（window）是流处理中解决事件的无边界（unbounded）及流动性的一种重要手段，把事件流在某一时刻变成静态的视图，以便进行类似数据库表的各种查询操作。在stream上可以定义window，窗口有两种类型，时间窗口（time-based）和记录窗口（row-based）。两种窗口都支持两种模式，滑动（slide）和跳动（tumble）。
* 表达式：符号和运算符的一种组合，CQL解析引擎处理该组合以获取单个值。简单表达式可以是常量、变量或者函数，可以用运算符将两个或者多个简单表达式联合起来构成更复杂的表达式。

# CQL应用程序开发流程

CQL应用程序的开发不同于传统的SQL语句。传统的SQL语句遵循先建表，再导入数据，再查询的顺序，每句SQL提交之后，都会得到SQL引擎的执行并获得返回结果；CQL语句是以应用程序作为一个整体运行的，一个Submit命令，代表一个应用程序的结束。一个应用程序中可以包含用户自定义参数、文件、jar包、函数等，还应该包含至少一个输入流、至少一个的输出流、以及至少一条输入数据计算语句。CQL中，一旦执行了Submit命令，系统就会在本地编译并提交应用程序，提交之后，会清空本地所有的用户操作记录，包含前期设置的参数、提交的用户jar包等。CQL语句提交过程中，如果发生语法错误，用户可以修改之后重新提交，如果发生了语义错误，那么输入内容都会被清空，用户必须从头提交整个应用程序。语义错误一般发生在explain语句或者submit语句执行过程中。

CQL语句并不严格要求各个不同语句的提交顺序，但是用户必须保证流（或者函数、参数等），在使用之前被定义。

CQL应用程序的编写顺序一般是：

| 序号 | 动作 | 说明 |
| --- | --- | --- |
| 1 | 添加自定义文件  添加自定义jar包 | 提交用户自定义功能用到的文件和jar包。  该步骤仅在包含用户自定义内容的时候存在。 |
| 2 | 创建自定义函数 | 定义用户的函数定义。  该步骤仅在包含用户自定义内容的时候存在。 |
| 3 | 创建输入流  创建输出流  创建用户自定义输入流，输出流，序列化，反序列化 | 创建输入输出流，既可以使用系统内置功能，也可以使用自定义功能。 |
| 4 | 编写逻辑计算语句 | 流处理业务计算逻辑核心。 |
| 5 | Submit提交应用程序 | 提交应用程序。 |

下面是一个CQL应用程序示例。

|  |
| --- |
| --添加用户自定义接口实现的jar包  add jar "/opt/streaming/example/example.jar";  --使用用户自定义接口实现创建输入流  CREATE INPUT STREAM S1(C1 STRING, C2 STRING)  SOURCE "com.huawei.streaming.example.userdefined.operator.input.FileInput"  PROPERTIES ("fileinput.path" = "/opt/streaming/example/input.txt");  --创建输出流  CREATE OUTPUT STREAM rs(C1 STRING,C2 STRING)  SINK kafkaOutput  PROPERTIES(groupid = "test",topic="ttopic",zookeepers="127.0.0.1:2181");  --应用程序逻辑部分  insert into stream rs select \* from S1;  --提交应用程序  submit application force example; |

说明

CQL内，除了字符串(用双引号或者单引号括起来)和应用程序名称之外，其他的语法都不区分大小写，比如各种关键字，参数名称等，都不区分大小写。

# 数据类型

[5.1 基本数据类型](#ZH-CN_TOPIC_0011584471)

[5.2 空值](#ZH-CN_TOPIC_0011584482)

## 基本数据类型

CQL基本数据类型

| 类型 | 范围 | 说明 |
| --- | --- | --- |
| Boolean | True或者False | 常量表示为：true或者false。  字符串输出的序列化反序列化规则见序列化、反序列化章节。 |
| Int | -231 ~ 231-1  即-2147483648~2147483647 | Int类型直接用数字表示 1,2,3… ，负数表示 -1,-2。 |
| Long | -263 ~ 263-1  即  -9223372036854775808~9223372036854775807 | Long类型表现和Int类型一致，常量表示中要添加后缀L(不区分大小写)，1L,-1l。 |
| Float | 1.4E-45~3.4028235E38 | Float类型和Int类型一致，但是要添加后缀F(不区分大小写)，如-1.0F,2f。 |
| Double | 4.9E-324~1.7976931348623157E308 | Double类型和Int类型一致，但是要添加后缀D(不区分大小写)，如-1.0d,2D。 |
| String |  | 字符串类型将字符用""或''包围起来，如"abc"或'abc'。 |
| Timestamp | Java.lang.Long数据类型取值范围 | 输入格式：yyyy-[M]M-[d]d [H]H:[m]m:[s]s[.fffffffff]。  [.fffffffff]可以进行省略，也可以输入1-9位任意多数字,最多可以精确到纳秒。  输出格式：yyyy-MM-dd HH:mm:ss.SSS，统一精确到毫秒。  yyyy 是表示年份的四位数字，范围为从 0001 到 9999。  MM 是表示指定年份中的月份的两位数字，范围为从 01 到 12。  dd 是表示指定月份中的某一天的两位数字，范围为从 01 到 31（最高值取决于具体月份）。  HH 是表示小时的两位数字，范围为 0 到 23。  mm 是表示分钟的两位数字，范围为 0 到 59。  ss 是表示秒的两位数字，范围为 0 到 59。  SSS 表示毫秒，范围从0到999。  fffffffff表示范围包含毫秒到纳秒，范围从0-999999999。 自动从高位运算，如：99=990000000，”2014-09-25 17:07:00.9”会自动格式化输出为” 2014-09-25 17:07:00.900”。  字符串格式表示为 : '1900-01-01 00:00:00', CQL支持字符串格式的直接输入，系统内部会直接将字符串格式转为指定类型。  TimeStamp类型不支持常量类型输入。 |
| Date | Java.lang.Long数据类型取值范围 | 日期类型，包括年(四位)，月和日。  输入格式：yyyy-[M]M-[d]d。  输出格式：yyyy-MM-dd  yyyy 是表示年份的四位数字，范围为从 0001 到 9999。  MM 是表示指定年份中的月份的两位数字，范围为从 01 到 12。  dd 是表示指定月份中的某一天的两位数字，范围为从 01 到 31（最高值取决于具体月份）。  字符串表示为：'1900-01-01', CQL支持字符串格式的直接输入，系统内部会直接将字符串格式转为指定类型。  Date类型不支持常量类型输入。 |
| Time | Java.lang.Long数据类型取值范围 | 一天中的时间，包括小时，分和秒。  输入格式：[H]H:[m]m:[s]s。  输出格式：HH:mm:ss  字符串表示范围从  '00:00:00'到’23:59:59’。  HH 是表示小时的两位数字，范围为 0 到 23。  mm 是表示分钟的两位数字，范围为 0 到 59。  ss 是表示秒的两位数字，范围为 0 到 59。  字符串表示为：'00:00:00'，CQL支持字符串格式的直接输入，系统内部会直接将字符串格式转为指定类型。  Time类型不支持常量类型输入。 |
| Decimal | Decimal 由任意精度的整数非标度值 和 32 位的整数标度 (scale) 组成。如果为零或正数，则标度是小数点后的位数；如果为负数，则将该数的非标度值乘以 10 的负 scale 次幂，因此，Decimal 表示的数值是 (unscaledValue × 10-scale)。 | 金额数据表示类型，金额数据类型运算时不会造成精度损失。  常量表示中要添加后缀BD(不区分大小写)，如1BD,-1bd。 |

说明

所有数据类型在CQL中定义的时候，不区分大小写，允许大小写混合，即：int=Int=iNt=INT

所有的数据类型，在输入和输出中如果使用CQL内置的序列化类，输入数字不用添加'F'，'D'，'BD'等字符后缀。在CQL表达式中表示常量的时候，字符后缀必须添加。

Float、Double的范围表示方法为“科学记数法”，如1.99714E13=19971400000000。

系统不区分+0，-0,+0和+0.0,-0.0的区别，这些都等于0或者0.0。

系统在将字符串 格式化成Float,Double类型的时候,允许数字后面带'F'，'D'字符。但是Float，Double等类型输出的时候不会带'F'，'D'字符。

系统在进行数反序列化的时候不会检查数据是否溢出，由用户自己保证输入的数据在指定类型的范围内。



Float类型和double类型属于浮点数类型，运算和表示均会导致精度丢失，所以不建议使用float或者double类型来进行高精度运算；高精度运算推荐使用decimal类型。

## 空值

NULL：

空值不属于任何数据类型，任何数据类型的值都会有Null值。

# CQL语法定义

[6.1 应用程序操作语句](#ZH-CN_TOPIC_0011584386)

[6.2 流操作语句](#ZH-CN_TOPIC_0011584414)

[6.3 查询语句](#ZH-CN_TOPIC_0011584418)

[6.4 命令语句](#ZH-CN_TOPIC_0011584432)

## 应用程序操作语句

### Drop

* **语法**

|  |
| --- |
| DROP APPLICATION [IF EXISTS] applicationName |

* **参数**

applicationName：应用程序名称，可以是字母、数字或者下划线。

* **功能**

删除系统内的应用程序。

如果指定了IF EXISTS，那么删除之前就会检查该应用程序是否存在， 如果存在，则删除，如果不存在，就不会删除。

如果没有指定IF EXISTS, 那么在删除应用程序的时候，如果应用程序不存在，则会抛出异常。

说明

用户可以删除所有自己能够看到的应用程序，不论安全还是非安全环境下。

* **示例**

|  |
| --- |
| DROP APPLICATION transformEvents; |
| DROP APPLICATION IF EXISTS transformEvents; |

### Submit

* **语法**

|  |
| --- |
| SUBMIT APPLICATION [FORCE] applicationName |

* **参数**

applicationName：应用程序名称, 可以是字母、数字或者下划线。

* **功能**

SUBMIT APPLICATION主要进行应用程序的提交，一旦应用程序开始提交，系统就开始进行应用程序的解析和发布。

如果包含FORCE选项，就会在提交应用程序之前，先检查该应用程序是否已经存在，如果已经存在，则会先停止原来的应用程序，再提交当前应用程序。 如果没有包含Force选项，系统就会不做任何检查，直接提交，如果该应用程序已经存在，就会抛出异常。



应用程序一旦submit，CQL当前用户线程中，所有用户设定的参数、提交的 SQL语句都会被清空，当前客户端连接线程会被重置，以进行下个应用程序的提交。

如果submit应用程序的时候，系统剩余资源不足以满足应用程序进程需要，应用程序会提交成功，状态会变为active，但是执行应用程序的进程数目不足，剩余资源需求会加入到资源申请队列中，等待系统资源释放之后再加入新的进程执行该应用程序。

如果submit应用程序的时候，系统剩余资源为0，应用程序会提交成功，状态会变为active，但是不会有任何进程执行该应用程序，该应用程序资源会加入到资源申请队列中，等待系统资源释放。

在安全环境下，用户只能看到自己提交的应用程序，看不到其他人提交的应用程序，所以提交应用程序的时候可能会遇到提示应用程序名称已经存在，而用户又看不到的问题；管理员用户可以看到任何用户提交的应用程序。哪些用户属于管理员取决于Storm集群配置。

* **示例**

|  |
| --- |
| SUBMIT APPLICATION transformEvents; |
| SUBMIT APPLICATION FORCE transformEvents; |

### Explain

* **语法**

|  |
| --- |
| EXPLAIN APPLICATION appName [path] |

* **参数**

appName：应用程序名称, 可以是字母、数字或者下划线。

path：执行计划文件路径，可以用单引号或者双引号围起来。可选参数，如果使用，则会从指定路径获取执行计划并提交；appName名称会覆盖执行计划中的应用程序名称。



1、path参数一旦指定，会先清空用户之前的所有输入，再进行任务提交，所以会导致用户之前输入被清空。

2、path参数支持相对路径，该相对路径一般在应用程序的启动路径下， 即在启动CQL命令行客户端或者CQL进程的当前路径下。

* **功能**

Explain语句用于显示当前所有已经提交的语句的查询计划。

查询计划显示出一个查询总共要分成多少个处理单元。并显示数据的输出和输出，可以用来显示CQL解析结果，并用于查询优化。

通过指定path文件地址，可以把查询计划导出到指定文件中。

Explain之后，可以继续提交新的CQL语句；只要没有执行submit application，所有的命令都可以下发。但是不允许下发重复的CQL语句，比如查询语句、插入语句等，这样可能会导致错误的解析结果。

* **示例**

|  |
| --- |
| EXPLAIN APPLICATION transform '/opt/cqlplan/transform.xml'; |

### Load

* **语法**

|  |
| --- |
| LOAD APPLICATION path |

* **参数**

Path： 执行计划文件路径，可以用单引号或者双引号围起来。

* **功能**

该语句用于加载执行计划。将查询计划加载到内存中，会默认清空在查询计划之前所提交的所有的SQL语句和参数设置。

* **示例**

|  |
| --- |
| LOAD APPLICATION '/opt/cqlplan/transform.xml'; |

## 流操作语句

### Create Input Stream

* **语法**

|  |
| --- |
| /\*  \* 输入流完整语法格式定义  \*/  createInputStreamStatement  :CREATE INPUT STREAM streamName  columnList [streamComment]  [serdeClause] sourceClause  [parallelClause]  ;  /\*  \* 数据列名称、列类型和列注释  \*/  columnList  :'(' columnName dataType [COMMENT columnComment], ... ')'  ;  /\*  \* 反序列化组件及相关配置属性定义  \*/  serdeClause  :SERDE serdeName [propertyList]  ;  /\*  \* 源数据读取组件及相关配置属性定义  \*/  sourceClause  :SOURCE sourceName [propertyList]  ;  /\*  \* 并发度设置  \*/  parallelClause  :PARALLEL intNumber  ;  /\*  \* 数据类型  \*/  dataType  :INT  |LONG  |FLOAT  |DOUBLE  |BOOLEAN  |STRING  |DATE  |TIME  |TIMESTAMP  |DECIMAL  ;  /\*  \* 配置属性列表  \* propertyName为属性名称，  \* 如果属于系统内置配置属性，名称可以不使用字符串格式  \* 如果是用户自定义配置属性，必须添加双引号  \* propertyValue为配置属性名称  \* 不论是系统内置配置属性还是用户自定义配置属性  \* 都必须是字符串格式。  \* 字符串格式就是以双引号或者单引号包围的字符串  \*/  propertyList  :PROPERTIES '(' propertyName = propertyValue, ... ')'  ;  /\*  \* 输入流或者输出流的注释  \*/  streamComment  :COMMENT commentString  ; |

* **参数**
* columnName：列名称，字符串形式，无需用单引号或者双引号围起来。

propertyName： 配置属性名称，如果不属于系统内置实现，属性名称就应该用双引号或者单引号围起来。

propertyValue： 配置属性值，字符串形式，必须用单引号或者双引号围起来。

commentString： 注释内容，字符串形式，必须用单引号或者双引号围起来。

intNumber： 正整形数字。

serdeName： 序列化类名称，如果属于CQL内部实现，可以使用简写，比如SimpleSerDe。如果是用户自定义实现，则必须用双引号或者单引号围起来。

sourceName： 输入数据读取类名称，如果属于CQL内部实现，可以使用简写，比如KafkaInput。如果是用户自定义实现，则必须用双引号或者单引号围起来。

streamName： 流名称，字符串形式，无需用单引号或者双引号围起来。

* **功能**

Create Input Stream语句定义了输入流的数据列名称，数据读取方式和数据反序列化方式。

SERDE定义了数据的反序列化方式，即如何将从inputStream中读取的数据解析成对应的完整的流的Schema。系统配置了默认的序列化和反序列化类，当使用系统默认反序列化类时，SERDE子句可以省略，同时，SERDE的相关配置属性也可以省略。

SOURCE定义了数据从什么地方读取，比如从MQ消息队列中读取或者文件等其他方式读取。SOURCE语句一定不能省略。

ParallelClause语句指定了该输入算子的并发数量。

* **示例**

|  |
| --- |
| --使用系统内置反序列化类读取算子示例  CREATE INPUT STREAM example  (  eventId INT,  eventDesc STRING  )  COMMENT "this is a example of create input stream."  SERDE simpleSerDe  PROPERTIES (separator = "|")  SOURCE TCPClientInput  PROPERTIES ( server = "127.0.0.1",  port = "9999" )  Parallel 2; |
| CREATE INPUT STREAM example  (  eventId INT,  eventDesc STRING  )  COMMENT "this is a example of create input stream."  SOURCE "com.huawei.streaming.example.source.localFileReader"  PROPERTIES ("reader.path" = "/opt/streaming/example"); |

### Create Output Stream

* **语法**

|  |
| --- |
| /\*  \* 输出流完整语法格式定义  \*/  createOutputStreamStatement  :CREATE OUTPUT STREAM streamName  columnList [streamComment]  [serdeClause] sinkClause  [parallelClause]  ;  /\*  \* 数据列名称、列类型和列注释  \*/  columnList  :'(' columnName dataType [COMMENT columnComment], ... ')'  ;  /\*  \* 反序列化组件及相关配置属性定义  \*/  serdeClause  :SERDE serdeName [propertyList]  ;  /\*  \* 流数据统计结果写入组件及相关配置属性定义  \*/  sinkClause  :SINK sinkName [propertyList]  ;  /\*  \* 并发度设置  \*/  parallelClause  :PARALLEL intNumber  ;  /\*  \* 数据类型  \*/  dataType  :INT  |LONG  |FLOAT  |DOUBLE  |BOOLEAN  |STRING  |DATE  |TIME  |TIMESTAMP  |DECIMAL  ;  /\*  \* 配置属性列表  \* propertyName为属性名称，  \* 如果属于系统内置配置属性，名称可以不使用字符串格式  \* 如果是用户自定义配置属性，必须添加双引号  \* propertyValue为配置属性名称  \* 不论是系统内置配置属性还是用户自定义配置属性  \* 都必须是字符串格式。  \* 字符串格式就是以双引号或者单引号包围的字符串  \*/  propertyList  :PROPERTIES '(' propertyName = propertyValue, ... ')'  ;  /\*  \* 输入流或者输出流的注释  \*/  streamComment  :COMMENT commentString  ; |

* **参数**

columnName：列名称，字符串形式，无需用单引号或者双引号围起来

propertyName： 配置属性名称，如果不属于系统内置实现，属性名称就应该用双引号或者单引号围起来。

propertyValue： 配置属性值，字符串形式，必须用单引号或者双引号围起来。

commentString： 注释内容，字符串形式，必须用单引号或者双引号围起来。

intNumber： 正整形数字。

serdeName： 序列化类名称，如果属于CQL内部实现，可以使用简写，比如SimpleSerDe。如果是用户自定义实现，则必须用双引号或者单引号围起来。

sinkName： 输出数据写入类名称，如果属于CQL内部实现，可以使用简写，比如KafkaOutput。如果是用户自定义实现，则必须用双引号或者单引号围起来。

streamName： 流名称，字符串形式，无需用单引号或者双引号围起来。

* **功能**

Create Output Stream语句定义了输出流的数据列名称，数据写入方式和数据序列化方式。

SERDE定义了数据的序列化方式，将 系统计算完毕的流数据如何序列化成指定格式，比如CSV格式或者二进制格式等。各个数据列和output Schema中的定义一致。系统配置了默认的序列化和反序列化类，当使用系统默认反序列化类时，SERDE子句可以省略，同时，SERDE的相关配置属性也可以省略。

SINK定义了处理完毕的数据要写入哪里，比如写入MQ或者写入文件等。Sink语句一定不能省略。

ParallelClause语句指定了该输入算子的并发数量。

* **示例**

|  |
| --- |
| --使用系统内置Kafka消息队列作为事件输出源。  --将数据格式化成CSV格式并输出  CREATE OUTPUT STREAM transformOutput  (  cnt INT  )  SERDE csvSerDe  SINK kafkaOutput  PROPERTIES (topic = "example")  parallel 2; |
| --用户自定义序列化类  --必须在使用之前先通过add jar的方式提交用户自定义程序  CREATE OUTPUT STREAM transformOutput  (  cnt INT  )  SERDE "com.huawei.streaming.example.serde.binarySerDe"  PROPERTIES ("binary.compress" = "true")  SINK kafkaOutput  PROPERTIES (topic = "example")  parallel 2; |

## 查询语句

### Insert

* **语法**

|  |
| --- |
| INSERT INTO STREAM streamName selectStatement |
| /\*  \* MultiInsert语句  \* 将一个流的数据按照不同的处理规则发送至不同的流  \* from语句中只允许简单的流存在，不允许出现窗口等其他语法  \*/  multiInsertStatement  : fromClause  multiInsert+  [parallelClause]  ;  /\*  \* MultiInsert语法定义  \*/  multiInsert  : insertClause multiSelect  ;  /\*  \* MultiInsert场景下的select字句语法定义  \* 和系统其它select语句对比，只少了From子句和parallel子句  \*/  multiSelect  : selectClause  whereClause?  groupByClause?  havingClause?  orderByClause?  limitClause?  ; |

* **参数**

streamName ：流名称，字符串形式，无需用单引号或者双引号围起来。

selectStatement ： select子句。

parallelClause：并发度设置子句，可选。

selectClause：select查询字句，必选，用来表达查询内容。

whereClause：where条件过滤子句，可选。

groupByClause：group by分组子句，可选。

havingClause： having聚合之后过滤子句，可选。

orderByClause：order by排序子句，可选。

limitClause：输出限制子句，可选。

* **功能**

可以将数据导入一个未定义的流中，但是如果有要将多个流的数据导入一个新流，这么几个导入语句生成的schema列名称和列类型必须相同。

允许将select结果导入到一个不存在的流中，只要这个流不是输入和输出流，系统就会自动创建该流。

MultiInsert语句一般用来进行单流数据的分割，它可以将一个流的数据，按照不同的处理规则，在处理完毕之后，发送至指定流。从而达到单流变多流的目的。

MultiInsert语句只有一个From字句，该字句中，只允许进行简单流的定义，不允许出现窗口等复杂语法。

* **示例**

|  |
| --- |
| INSERT INTO STREAM transformTemp SELECT \* FROM transform; |
| --不包含子查询的MultiInsert  FROM teststream  INSERT INTO STREAM s1 SELECT \*  INSERT INTO STREAM s2 SELECT a  INSERT INTO STREAM s3 SELECT id, name WHERE id > 10  Parallel 4; |
| --包含子查询的MultiInsert  FROM  (  SELECT count(id) as id, 'sss' as name  FROM testStream(id >5 )[RANGE 1 SECONDS SLIDE]  GROUP BY ss  )  INSERT INTO STREAM s1 SELECT \*  INSERT INTO STREAM s2 SELECT a  INSERT INTO STREAM s3 SELECT id,name WHERE id > 10; |

### SELECT

Select的语法结构由下面几个子句构成。

|  |
| --- |
| selectStatement  :SelectClause  FromClause  [WhereClause]  [GroupByClause]  [HavingClause]  [OrderbyClause]  [LimitClause]  [ParallelClause]  ; |

#### Select Clause

* **语法**

|  |
| --- |
| selectClause  :SELECT selectList,…  ;  selectList  :\*  |(streamName | streamAlias).\*  |(  (  [ (streamName | streamAlias). ] columnName  |expression  ) [ [ AS ] columnAlias ]  )  ; |

* **参数**

streamName：流名称，同create input stream以及create output stream中定义的流名称一致。

streamAlias：流别名，该别名在from子句中定义，在select子句中引用。

columnName：列名称，和create input stream以及create output stream中定义的列名称一致。

expression：CQL表达式子句。

* **功能**

Select子句包含了对输出数据计算方式的定义。

Distinct语法不支持直接在select子句中使用，只能放在UDAF函数中，比如count(distinct id)。

* **示例**

|  |
| --- |
| SELECT \* FROM transform[RANGE UNBOUNDED SLIDE]; |
| SELECT eventid FROM transform[RANGE UNBOUNDED]; |
| SELECT eventid, eventcode1 + transform.eventcde2 FROM transform; |
| SELECT eventid, sum(DISTINCT cde) FROM transform[RANGE UNBOUNDED SLIDE] GROUP BY eventtype; |

#### From Clause

* **语法**

|  |
| --- |
| /\*\*  \* From子句语法定义  \*/  fromClause  :FROM fromSource, ...  ;  fromSource  :streamBody  | dataSourceBody  ;  /\*\*  \* From字句流定义  \* 包含流名称，Join，流前过滤，别名以及单向Join Unidirection定义  \*/  streamBody  :(  streamName  |joined  |subquery  )  [ filterBeforeWindow ]  [ windowSource ]  [ [ AS ] streamAlias ]  [ UNIDIRECTION ]  ;  /\*\*  \* Join语法定义  \* Cross Join没有On条件  \*/  joined  :fromSource joinType fromSource ON searchCondition  |fromSource CROSS JOIN fromSource  ;  /\*\*  \* 非Cross Join的Join类型定义  \*/  joinType  :(INNER |  (  (LEFT | RIGHT | FULL) [OUTER]  )  )  Join  ;  /\*\*  \* 子查询定义  \*/  subquery  :'(' subqueryStatemente ')'  ;  /\*\*  \* 窗口之前过滤定义  \*/  filterBeforeWindow  :'(' expression ')'  ;  /\*\*  \*数据源引用的别名是可选的，  \*也就是说可以使用dataSourceName.columnName,  \*也可以使用sourceAlias.columnName  \*/  dataSourceBody  : DATASOURCE dataSourceName datasourceArguments [sourceAlias]  ;  dataSourceArguments  : '(' dataSourceSchema,dataSourceQuery ')'  ;  /\*\*  数据源中使用到的schema  \*/  dataSourceSchema  : SCHEMA '(' columnNameTypeList ')'  ;  /\*\*  \*数据源查询语句，所有的CQL参数都在这个query中定义  \*/  dataSourceQuery  : QUERY '(' dataSourceQueryArguments ')'  ;  /\*\*  数据源查询参数，可以是各种常量，可以是字符串、int或者其他类型  \*/  dataSourceQueryArguments  : expression (',', expression)\*  ; |

* **参数**

streamName：流名称，同create input stream中的流名称定义。

expression：CQL表达式子句。

dataSourceName：数据源名称，和输入流中定义的数据源名称一致

subqueryStatemente：子查询语句，见SubQuery子查询语法。

columnNameTypeList：列名称和列类型的定义列表，同输入流中的语法定义。

streamAlias：流名称的别名。

* **功能**

From子句中包含对流的引用和对子查询的引用。如果包含子句，子句需要重命名。

Fiterbeforewindow子句是对进入窗口之前的数据进行过滤，比where子句的过滤发生的更早。使用FilterBeforeWindow子句，可以有效减少进入窗口的数据量。

Join支持内连接（inner join）、左连接（left outer join）、右连接（right outer join）、全连接（full outer join），交叉连接（cross join）。如果两个流以逗号相连，即隐含join条件，默认按照Cross join执行。Join的匹配规则和输出结果和SQL类似。

Join是对当前左流窗口内数据和右流窗口内数据按照条件进行连接处理。

各类Join类型描述

| 类型 | 描述 | 示例 |
| --- | --- | --- |
| [INNER] JOIN | 返回两侧流同时满足条件的组合 | From S1[range 20 seconds batch]  join S2[range unbounded]  on s1.id=s2.type;  from S1[range 20 seconds batch]  inner join S2[range unbounded]  on s1.id=s2.type; |
| LEFT [OUTER] JOIN | 显示符合条件的数据行，同时显示左边数据流不符合条件的数据行，右边没有对应的条目显示NULL | from S1[range 20 seconds batch]  left outer join S2[range unbounded]  on s1.id=s2.type;  from S1[range 20 seconds batch]  left join S2[range unbounded]  on s1.id=s2.type; |
| RIGHT [OUTER] JOIN | 显示符合条件的数据行，同时显示右边数据流不符合条件的数据行，左边没有对应的条目显示NULL | from S1[range 20 seconds batch]  right outer join S2[range unbounded]  on s1.id=s2.type;  from S1[range 20 seconds batch]  right join S2[range unbounded]  on s1.id=s2.type; |
| FULL [OUTER] JOIN | 显示符合条件的数据行，同时显示左右不符合条件的数据行，相应的左右两边显示NULL，即显示左连接、右连接和内连接的并集 | from S1[range 20 seconds batch]  full outer join S2[range unbounded]  on s1.id=s2.type;  from S1[range 20 seconds batch]  full join S2[range unbounded]  on s1.id=s2.type; |
| CROSS JOIN | 它将会返回被连接的两个表的笛卡尔积，返回结果的行数等于两个表行数的乘积 | From S1[range 20 seconds batch],  S2[range unbounded];  from S1[range 20 seconds batch]  cross join S2[range unbounded]; |

说明

CQL目前只支持双流Join，并且只支持等值Join，且只支持属性表达式。比如s1 inner join s2 on s1.id=s2.id.

Unidirection关键词用来指定单向Join。在流处理系统内部Join的情况下，两个流都有数据产生，双方都会触发Join，如果两个流中，有一边包含了unidirection，那么另外一边，在有数据产生的时候，并不会触发Join，只有包含了Unidirection的这一端，才会触发Join。

From子句中还包含了对DataSource语法的定义，关于DataSource，可以看DataSource章节。

* **示例**

|  |
| --- |
| --子查询示例  SELECT \*  FROM  (  SELECT id, name, type  FROM transform  ) AS transformbak; |
| --流前过滤示例  SELECT \*  FROM transform ( evnetid > 10 )[range UNBOUNDED]  WHERE eventtype =1; |
| --inner join示例  insert into stream rs select \* from S1[range 20 seconds batch] join S2[range unbounded] on s1.id=s2.type where s1.id > 5;  insert into stream rs select \* from S1[range 20 seconds batch] inner join S2[range unbounded] on s1.id=s2.type where s1.id > 5; |
| --left join示例  insert into stream rs select \* from S1[range 20 seconds batch] left outer join S2[range unbounded] on s1.id=s2.type where s1.id > 5;  insert into stream rs select \* from S1[range 20 seconds batch] left join S2[range unbounded] on s1.id=s2.type where s1.id > 5; |
| --right join示例  insert into stream rs select \* from S1[range 20 seconds batch] right outer join S2[range unbounded] on s1.id=s2.type where s1.id > 5;  insert into stream rs select \* from S1[range 20 seconds batch] right join S2[range unbounded] on s1.id=s2.type where s1.id > 5; |
| --full join示例  insert into stream rs select \* from S1[range 20 seconds batch] full outer join S2[range unbounded] on s1.id=s2.type where s1.id > 5;  insert into stream rs select \* from S1[range 20 seconds batch] full join S2[range unbounded] on s1.id=s2.type where s1.id > 5; |
| --corss join示例  insert into stream rs select \* from S1[range 20 seconds batch], S2[range unbounded] where s1.id > 5;  insert into stream rs select \* from S1[range 20 seconds batch] cross join S2[range unbounded] where s1.id > 5; |

#### Where Clause

* **语法**

|  |
| --- |
| whereClause  :WHERE expression  ; |

* **参数**

expression：CQL表达式子句。

* **功能**

Where中的过滤条件一定是发生在数据进入窗口之后。聚合之前。所以where的条件表达式中一般不会出现聚合类的函数。

Expression可以是表达式或者其他用户自定义表达式，但是返回值必须是布尔类型。比如is null之类。

多个表达式之间可以使用AND或者其他的二元表达式。

* **示例**

|  |
| --- |
| WHERE a.size = 1 and a.size + b.size =2; |
| WHERE a.size = 1 and ( a.size + b.size ) >= 2; |
| WHERE a.size = 1 and cast(a.isnative as boolean)=true; |

#### Group By Clause

* **语法**

|  |
| --- |
| groupByClause  :GROUP BY expression, ...  ; |

* **参数**

expression：CQL表达式子句。

* **功能**

GroupBy的作用在于保证数据按照groupBy的字段进行分发，以及在算子内部，按照group by的字段进行分组以进行聚合计算。

Group by的操作在窗口之前，即先group by，再执行针对分组的属性的window。

允许group by的条件是一个UDF的表达式结果，即按照udf函数结果进行分组。



Group by中的表达式目前只支持流中的数据列，不支持其他表达式，比如函数等。

* **示例**

|  |
| --- |
| SELECT s.age,count(1) as cc  FROM schagetransformEvent[RANGE 1 SECOND BATCH] as s  GROUP BY s.age,s.departmentid; |

#### Having Clause

* **语法**

|  |
| --- |
| havingClause  :HAVING expression  ; |

* **参数**

expression：CQL表达式子句。

* **功能**

Having语句用来对聚合之后的结果数据进行过滤。

允许定义单独的having子句(即不使用group by，直接定义having)， 因为CQL的流式查询中默认已经包含了聚合的窗口。

Having中一般存放聚合操作的过滤条件，比如having count(\*) > 10。



Having条件过滤发生在聚合之后，这个时候系统内已经是计算之后的结果，所以过滤条件中不应该出现From子句中输入流的流名称和列名称。

* **示例**

|  |
| --- |
| SELECT  schages.age,  count(1) as cnt  FROM schagetransformEvent[RANGE 20 SECONDS BATCH]  GROUP BY schages.age  HAVING schages.age=10 AND cnt=2;  --having中包含了s.age=10的条件，但是由于这个过滤条件放在了having中，所以这--个条件会在执行了count之后，再来执行。 |

#### Order By Clause

* **语法**

|  |
| --- |
| orderbyClause  :ORDER BY (expression [ASC|DESC]), ...  ; |

* **参数**

expression：CQL表达式子句。

* **功能**

针对一个或者多个字段的排序。默认升序。

ASC升序排列。

DESC降序排列。

使用Orderby的时候，如果并发度不为1，那么则进行单个算子内部的排序。



Order by发生在聚合之后，这个时候系统内已经是计算之后的结果，所以过滤条件中不应该出现From子句中输入流的流名称和列名称。

* **示例**

|  |
| --- |
| SELECT \*  FROM schagetransformEvent[RANGE 20 SECONDS BATCH]  ORDER BY id DESC; |

#### Limit Clause

* **语法**

|  |
| --- |
| LIMIT intNumber |

* **参数**

IntNumber: 正整形数字。

* **说明**

Limit主要用来限制输出，限制同一批次输出数据的数量。

如果同一批输出多条数据，则从前面开始去limit数量的数据。如果同一批次数据不足limit限制大小，则输出整个批次数据。

* **示例**

|  |
| --- |
| SELECT \* FROM ticketsRecorder[RANGE 20 SECONDS BATCH] LIMIT 10; |

#### Parallel Clause

* **语法**

|  |
| --- |
| PARALLEL intNumber |

* **参数**

IntNumber: 正整形数字。

* **说明**

Parallel主要用设置该语句的并发数量，如果该语句中包含了子查询，该并发度设置也会影响到子查询的并发度；外部子查询的并发度会覆盖子查询的并发度。

该语句可选，默认并发度为1。

* **示例**

|  |
| --- |
| SELECT \* FROM ticketsRecorder[RANGE 20 SECONDS BATCH] LIMIT 10 Parallel 2; |
| INSERT INTO STREAM rs  SELECT \*, count(a)  FROM (  SELECT id as a,id as b, id c from S  )[RANGE 20 SECONDS BATCH]  WHERE a > 5  GROUP BY c  PARALLEL 4; |

### Subquery Clause

* **语法**

|  |
| --- |
| SubQueryStatement  : selectStatement  ;  selectStatement  :selectClause  fromClause  [whereClause]  [groupByClause]  [havingClause]  [orderbyClause]  [LimitClause]  [ParallelClause]  ; |

* **参数**

selectStatement的定义同select子句定义。

* **功能**

子查询可以在From子句中引用。

From语句中的子查询，必须显示标注别名才可以引用。

* **示例**

|  |
| --- |
| SELECT a.id  FROM (  SELECT count(tid) as id  FROM schagetransformEvent[RANGE UNBOUNDED]  ) AS a; |

### DataSource

数据源是存储数据的容器，其数据不是以流的形式往外吐的，而是按照触发的方式，根据条件，从数据源中进行查询。

可以把数据源想象成一个数据库，在每条事件到来的时候，都会进行一次查询，并返回查询结果。

数据源的使用分为数据源定义和数据源查询两部分。

#### Create DataSource

* **语法**

|  |
| --- |
| /\*\*  数据源定义语法  \*/  createDataSourceStatement  :CREATE DATASOURCE dataSourceName  SOURCE className  datasourceProperties?  ;  /\*  \* 配置属性列表  \* propertyName为属性名称，  \* 如果属于系统内置配置属性，名称可以不使用字符串格式  \* 如果是用户自定义配置属性，必须添加双引号  \* propertyValue为配置属性名称  \* 不论是系统内置配置属性还是用户自定义配置属性  \* 都必须是字符串格式。  \* 字符串格式就是以双引号或者单引号包围的字符串  \*/  propertyList  :PROPERTIES '(' propertyName = propertyValue, ... ')'  ; |

* **参数**
* dataSourceName：数据源名称，字符串形式，无需用单引号或者双引号围起来。

propertyName： 配置属性名称，如果不属于系统内置实现，属性名称就应该用双引号或者单引号围起来。

propertyValue： 配置属性值，字符串形式，必须用单引号或者双引号围起来。

className： 数据源类，字符出纳形式，必须用单引号或者双引号围起来, 如果属于CQL内部实现，可以使用简写,比如RDBDatasource。

* **功能**

Create DataSource提供了数据源定义的语法。数据源的使用分为定义阶段和使用两个阶段。这里描述的是数据源的定义语法，数据源的使用语法见From子句定义。

数据源提供了一种访问外部数据的方式，比如实时访问关系型数据库，访问HBase等组件，支持从一行原始数据通过查询获取多行数据。

数据源的定义和具体使用无关，可以在一个拓扑中定义一个数据源，然后在其他地方多次使用该数据源。比如定义一个数据库数据源，然后可以在不同的地方使用不同的查询语句进行查询。

* **示例**

|  |
| --- |
| CREATE DATASOURCE rdbdatasource  SOURCE RDBDataSource  PROPERTIES (  url = "jdbc:postgresql://127.0.0.1:1521/streaming",  username = "55B5B07CF57318642D38F0CEE0666D26",  password = "55B5B07CF57318642D38F0CEE0666D26"  ); |

#### DataSourceQuery

* **语法**

数据源和From子句一起结合使用，dataSourceBody语法在from字句中也进行了定义。

|  |
| --- |
| /\*\*  数据源引用的别名是可选的，  也就是说可以使用dataSourceName.columnName,  也可以使用sourceAlias.columnName  \*/  dataSourceBody  :DATASOURCE dataSourceName datasourceArguments [sourceAlias]  ;  dataSourceArguments  :'(' dataSourceSchema,dataSourceQuery ')'  ;  /\*\*  数据源中使用到的schema  \*/  dataSourceSchema  :SCHEMA '(' columnNameTypeList ')'  ;  /\*\*  数据源查询语句，所有的CQL参数都在这个query中定义  \*/  dataSourceQuery  :QUERY '(' dataSourceQueryArguments ')'  ;  /\*\*  数据源查询参数，可以是各种常量，可以是字符串、int或者其他类型  \*/  dataSourceQueryArguments  :expression (',' , expression)\*  ; |

* **参数**

dataSourceName：数据源名称，同数据源定义中的名称一致。

sourceAlias：数据源别名，字符串形式，但是不包含双引号和单引号。

columnNameTypeList：列名称和列类型定义，同输入流定义。

expression：CQL表达式。

* **功能**

数据源的查询参数分为查询Schema定义和数据源查询两个部分。

Schema的定义同Create Input Stream中的schema定义，主要用来指定数据源查询结果的列数量、名称、类型，便于进行下一步处理。

* **示例**

|  |
| --- |
| --RDB数据读取，支持多行数据读取，同时支持CQL UDF以及窗口和聚合运算  --QUERY内部的参数顺序固定，不同的数据源，有不同的参数。  --RDB的SQL中，如果不包含Where，就会一次查询出多行记录。和原始流做了Join之后，最终输出多条结果。  insert into rs select rdb.id,s.id,count(rdb.id),sum(s.id) from S[rows 10 slide],  DATASOURCE rdbdatasource  [  SCHEMA (id int,name String,type int),  QUERY("select rid as id,rname,rtype from rdbtable where id = ? ", s.id)  ] rdb  where rdb.name like '%hdd%'  group by rdb.id,s.id; |

### User Defined Operator

当用户的业务逻辑比较复杂，无法用SQL表达出来的时候，就应该使用自定义算子来解决。用户只要实现了自定义算子的接口，然后就可以将这个算子和输入输出算子或者其他聚合算子一起结合使用。

用户自定义算子只支持单个输入schema和单个输出schema。

数据源的使用分为数据源定义和数据源查询两部分。

#### Create User Operator

* **语法**

|  |
| --- |
| /\*\*  自定义算子语法  \*/  createOperatorStatement  :KW\_CREATE KW\_OPERATOR operatorName KW\_AS className  inputSchemaStatement outputSchemaStatement  streamProperties?  ;  /\*\*  自定义算子的输入schema  \*/  inputSchemaStatement  :KW\_INPUT LPAREN columnNameTypeList RPAREN  ;  /\*\*  自定义算子的输出schema  \*/  outputSchemaStatement  :KW\_OUTPUT LPAREN columnNameTypeList RPAREN  ;  columnNameTypeList  :columnNameType (COMMA columnNameType)\*  ;  streamPropertiesList  :keyValueProperty (COMMA keyValueProperty)\*  ;  keyValueProperty  :confName EQUAL confValue  ;  columnNameType  :columnName colType comment?  ; |

* **参数**

propertyName： 配置属性名称，如果不属于系统内置实现，属性名称就应该用双引号或者单引号围起来。

operatorName：自定义算子名称，字符串形式，无需用单引号或者双引号围起来。

propertyValue： 配置属性值，字符串形式，必须用单引号或者双引号围起来。

className： 数据源类，字符出纳形式，必须用单引号或者双引号围起来, 如果属于CQL内部实现，可以使用简写,比如RDBDatasource。

* **功能**

Create user operator提供了用户自定义算子的定义语法，在该语法中，可以定义算子的输入schema和输出schema，以及用到的参数，使用的类。

用户schema会在提交拓扑提交之前进行校验，系统会校验输出schema和输出schema的数据类型，如果数据类型不一致或者列数量不一致，都会导致检查失败，拓扑会提交失败。

* **示例**

|  |
| --- |
| **create** operator userOp **as** **"com.huawei.streaming.example.userdefined.operator.UserOperator"**  input (id int, name string)  output (newID string, name string, type int)  properties (**"userop.filename"** = **"/home/omm/kv.properties"**); |

#### User Operator Query

* **语法**

数据源和From子句一起结合使用，dataSourceBody语法在from字句中也进行了定义。

|  |
| --- |
| insertUserOperatorStatement  :insertClause usingStatement  ;  usingStatement  :KW\_USING KW\_OPERATOR operatorName KW\_FROM streamName distributeClause? parallelClause?  ;  distributeClause  :KW\_DISTRIBUTE KW\_BY columnName  ; |

* **参数**

streamName ：流名称，字符串形式，无需用单引号或者双引号围起来。

operatorName：自定义算子名称，字符串形式，无需用单引号或者双引号围起来。

columnName：自定义算子中的输入schema列名称，指定了数据按照哪个字段进行分发，可选项，如果没有指定，默认随机分发。

parallelClause：并发子句，同select子句中的并发语法，设置了用户自定义算子的并发度。

* **功能**

将用户自定义算子和其他输入输出算子或者其他功能算子集成，构成一个大拓扑，以实现用户功能。

* **示例**

|  |
| --- |
| **insert** **into** rs using operator userOp **from** s distribute **by** id parallel **2**; |

## 命令语句

所有的系统命令，都会下发之后马上生效，马上可以查看到结果。

但是应用程序提交之后（即执行submit application之后），所有修改的参数、自定义函数和窗口等，都会被重置。

### Set/Get

* **语法**

|  |
| --- |
| SET conf = value  GET conf |

* **参数**

Conf：配置属性名称，字符串形式。

Value：配置属性值，字符串形式。

* **功能**
* Set 设置系统内变量或者参数，get获取系统内变量或者参数。

配置参数的名称统一用字符形式表示，字符形式是不区分大小写的；但是对于特殊的名称，可以用双引号或者单引号包起来的字符串形式，字符串内容本身是区分大小写的。

变量值统一用字符串表示，字符串本身是区分大小写的。

可以在set和get命令中使用系统变量或者环境变量以及引用其他已经定义的配置属性。

* **示例**

Set x=1 可以在系统内设置变量X，这个变量x可以在任何包含系统配置属性对象的地方使用。

Get x 可以获取变量x的值，如果不包含该变量，则返回NULL。

|  |
| --- |
| --一般的set命令  set 'operator.kafka.zookeepers' = 'localhost:2181,10.0.0.21:2181'; |
| --set命令的字符串既可以用单引号，也可以用双引号  set "serde.simpleserde.separator"=','; |
| --set的配置属性值中引用其他配置属性的值  set "x" = "abc";  set "x" = "${conf:x}def";  --get的值等于abcdef  get "x"; |
| --set中使用系统变量  set "streaming.template.directory" = "${system:java.io.tmpdir}"; |

### Add Jar

* **语法**

|  |
| --- |
| ADD JAR path |

* **参数**

Path：文件路径，用双引号或者单引号围起来的字符串形式。

* **功能**

在系统使用中，经常要用到用户接口，比如用户在CQL中使用自定义的序列化规则、自定义的输入输出或者自定义函数之类，这些程序，可以由用户自己打包成Jar包，再通过add jar的命令进行提交，然后就可以使用了。

连接关闭之后，用户手工添加的Jar文件会从本地进程中卸载，无需担心会影响到下一次连接的操作。

* **示例**

|  |
| --- |
| Add jar '/opt/huawei/streaming/test/test.jar'; |

### Add File

* **语法**

|  |
| --- |
| ADD FILE path |

* **参数**

Path：文件路径，用双引号或者单引号围起来的字符串形式。

* **功能**

在系统使用中，经常要用到用户接口，比如用户在CQL中使用自定义的序列化规则、自定义的输入输出或者自定义函数之类，在这些类中有时会用到配置文件或者资源文件，这些文件，可以通过add File的方式进行提交，在整个应用程序提交到远端集群之后访问。

* **示例**

|  |
| --- |
| ADD FILE '/opt/huawei/streaming/test/test.conf'; |

### Create Function

* **语法**

|  |
| --- |
| CREATE FUNCTION functionName AS className [propertyList]  propertyList  :PROPERTIES '(' propertyName = propertyValue, ... ')'  ; |

* **参数**

propertyName： 配置属性名称，如果不属于系统内置实现，属性名称就应该用双引号或者单引号围起来。

propertyValue： 配置属性值，字符串形式，必须用单引号或者双引号围起来。

functionName：函数名称，不带双引号或则和单引号的字符串形式。

clasName：函数所在类，字符串形式，用双引号或者单引号围起来。

* **功能**

该功能用于注册用户自定义函数，该函数名称不能和系统函数名称重复，如果重复，系统会抛出异常，给以提示。 用户自定义函数则允许通过重复定义来覆盖上次定义的函数。

所有的函数名称都是都不区分大小写。

函数无需刻意注销，在应用程序提交之后， 该函数即会失效，注册的Jar包和文件都会从本地进程中卸载。

* **示例**

|  |
| --- |
| CREATE FUNCTION rdbCompare AS 'com.huawei.streaming.example.udf.RDBCompare'; |
| create function udf1 as "com.huawei.streaming.cql.functions.DirectOutputUDF"  properties ("cql.test.direct.output" = "a"); |
| create function udf2 as "com.huawei.streaming.cql.functions.DirectOutputUDF"  properties ("cql.test.direct.output" = "b"); |
| create function udf3 as "com.huawei.streaming.cql.functions.DirectOutputUDF"; |

### Drop Function

* **语法**

|  |
| --- |
| DROP FUNCTION [IF EXISTS] functionName |

* **参数**

FunctionName：函数名称，同create Function。

* **功能**

从系统中删除函数，函数名称不区分大小写。

在不包含IF EXISTS的场景下，如果函数不存在，则会抛出异常。

* **示例**

|  |
| --- |
| DROP FUNCTION rdbCompare; |

### Show Applications

* **定义**

|  |
| --- |
| SHOW APPLICATIONS [applicationName] |

* **参数**

applicationName：应用程序名称，可选，可以通过双引号或者单引号围起来进行模糊查询。

在安全环境下，一般用户只能看到自己提交的应用程序，管理员用户可以看到所有应用程序，哪些用户属于管理员取决于Storm集群的配置。

* **功能**
* 查询应用程序信息。

查询结果由底层运行平台决定，一般包含应用程序名称，应用程序状态，应用程序启动时间等信息。

可以通过添加应用程序名称来进行应用程序过滤。

* **示例**

|  |
| --- |
| Show applications;查询系统内正在运行的所有的应用程序 |
| Show applications filter; 查询系统内正在运行的，名称包含filter的应用程序 |
| Show applications 'filter\*';查询系统内正在运行的，名称以filter开头的应用程序 |
| Show applications '\*filter\*';查询系统内正在运行的，名称包含filter的应用程序 |
| Show applications '\*filter';查询系统内正在运行的，名称以filter结尾的应用程序 |

# 窗口

窗口用来保存流上一段时间内的数据，并且由此产生过期事件。

窗口名称不区分大小写。

整个窗口由窗体、属性组成。窗体主要描述窗口属于时间窗还是行组窗。属性主要描述窗口的数据输出是滑动还是跳动方式。

窗口函数使用的基本语法规则如下：

StreamName(filter)[window]

Filter：流前过滤，数据进入窗口之前的过滤条件。

Window：窗口定义

时间窗口中的时间单位可以表示为millisecond(s), second(s), minute(s), hour(s), day(s),既可以用单数，也可以用复数。

CQL的窗口有 窗体，属性和限定词三部分构成，一个窗口只能属于一种窗体，或者最多拥有一种属性，但是可以拥有多个限定词。

窗体分类说明

| 窗体名称 | 语法 | 说明 |
| --- | --- | --- |
| 行组窗 | Rows N1 | 数量为N1的元组窗。窗口内最多包含N1行数据 |
| 范围窗 | Range T1 | T1时间单位范围内的时间窗。 |

窗口属性说明

| 窗口属性 | 语法 | 说明 |
| --- | --- | --- |
| 滑动窗 | Slide | 窗口数据以滑动方式输出，每次输出一条数据 |
| 跳动窗 | Batch | 窗口数据以跳动方式输出，每次输出的数量，取决于窗口的容量。 |



Trigger by不支持time数据类型。

窗口数据输出分为新数据和旧数据，CQL中的窗口一般只输出新数据。

Sort by排序窗口主要解决分布式数据乱序问题，该窗口只输出旧数据。

Join的时候，如果排序窗口和Unidirection在同一侧，那么统一输出旧数据，如果不在一侧，则输出新数据。

Join不支持一侧是排序窗口，一侧是其他非排序窗口的场景，因为新数据和旧数据不能一起计算。

窗口语法定义如下：

|  |
| --- |
| Window  :rangeWindow  |rowsWindow  |rangeToday  ;  rowsWindow  :rows number windowProperties  ;  rangeWindow  :range (number|unbounded) windowProperties windowDeterminer?  ;  rangeToday  :range today expression windowDeterminer?  ;  windowProperties  :slide  |batch  ; |

[7.1 元组窗口](#ZH-CN_TOPIC_0011584441)

[7.2 范围窗口](#ZH-CN_TOPIC_0011584442)

## 元组窗口

* **语法**

|  |
| --- |
| rowsWindow  :ROWS intNumber windowProperties  ; |

* **参数**

intNumber：正整数，行组数量。

WindowProperties：窗口属性，滑动窗或者跳动窗。

* **说明**

元组窗的主要作用在于在窗口内保存一定数量的事件，然后依据窗口的不同特性，在不同的时间过期，产生新旧事件。窗口内只能保存一定数量的事件。

元组窗口

| 语法 | 名称 | 说明 |
| --- | --- | --- |
| S[ROWS N1 SLIDE ] | 长度滑动窗 | 窗口内最大保存N1个事件，当有新事件产生的时候，窗口内的事件依次过期。每个过期事件的批次都不同。 |
| S[ROWS N1 BATCH] | 长度跳动窗 | 窗口内最大保存N1个事件，当有新事件产生的时候，窗口内每攒满N1个事件，就过期依次。同时过期的所有事件处于同一批次。 |

说明

用大括号'{','}'括起来的部分为可选项。

* **示例**

|  |
| --- |
| Select \* from transformEvent[rows 10 slide];  容量为10的滑动窗。 |

## 范围窗口

* **语法**

|  |
| --- |
| rangeWindow  :RANGE rangebound windowProperties  ;  rangeBound  :rangeTime  |rangeUnBound  ;  rangeTime  :rangeDay? rangeHour? rangeMinutes? rangeSeconds? rangeMilliSeconds?  ;  rangeUnBound  :UNBOUNDED  ; |

* **参数**

rangeTime： 窗口的时间范围，可以添加各种时间单位。

* **功能**

范围窗口的主要作用在于在窗口内保存一定数量的事件，然后依据窗口的不同特性，在不同的时间过期，产生新旧事件。窗口内只能保存指定时间范围的事件。

范围窗口

| 语法 | 名称 | 说明 |
| --- | --- | --- |
| S[RANGE UNBOUNDED ] | 无限范围窗 | 窗口内可以保存任意多的数据，数据永远不会过期。 |
| S[RANGE T1 SLIDE] | 时间滑动窗 | 窗口内保存最近T1时间范围内的数据，T1是一个时间单位，可以加入Seconds等时间单位。窗口内的事件依次过期。每个过期事件的批次都不同。 |
| S[RANGE T1 BATCH] | 时间跳动窗 | 窗口内保存最近T1时间范围内的数据，T1是一个时间单位，可以加入Seconds等时间单位。窗口内的事件分批次过期，每次过期的数据属于同一个批次。 |

说明

用大括号'{','}'括起来的部分为可选项。



时间排序窗在计算的时候，是和集群内时间进行比较，如果满足：【事件时间+窗口时间>系统时间】， 才会输出；如果：【事件时间+窗口时间<系统时间】，说明事件已经发生了很久了，就会马上输出。

**示例**

|  |
| --- |
| Select \* from transformEvent[range unbounded] as s1 inner join transformBak s2 on S1.id=S2.id; |
| Select \* from transformEvent[range 10 seconds slide]; |
| Select \* from transformEvent[range 1 day 6 hours slide]; |

# 表达式

表达式：符号和运算符的一种组合，CQL解析引擎处理该组合以获取单个值。简单表达式可以是常量、变量或者函数，可以用运算符将两个或者多个简单表达式联合起来构成更复杂的表达式。

表达式语法如下：

|  |
| --- |
| expression  :constant  |table\_alias\_name.colunm  |function  |subquery  |expression binary\_operator expression  |special\_expression  ;  special\_expression  :like  |cast  |in  |between  |exists  |previous  |case\_when  ;  binary\_operator  :'+'  |'-'  |'\*'  |'/'  |'%'  |and  |or  ;  function  :functionName ([function\_body])  ; |

[8.1 通用表达式](#ZH-CN_TOPIC_0011584445)

[8.2 表达式常量](#ZH-CN_TOPIC_0011584446)

[8.3 特殊表达式](#ZH-CN_TOPIC_0011584447)

[8.4 表达式数据类型转换规则](#ZH-CN_TOPIC_0011584455)

## 通用表达式

通用表达式可以出现在select子句中，也可以出现在where或者having子句中。

表达式的优先级从高到底如下表所示。

通用表达式列表

| 表达式 | 说明 |
| --- | --- |
| +, - | 正数，负数之类一元操作符 |
| \*, / | 乘、除 |
| +, - | 二元操作符 |
| =, <>, <, >, <=, >=， != | 比较操作符 |
| IS [NOT] NULL, [NOT] LIKE, [NOT] BETWEEN,[NOT] IN | 是否为空之类判断表达式 |
| AND、OR | 逻辑表达式：多个条件之间是"且"或者"或"的关系 |



Date，time，timestamp数据类型不支持加减乘除比较等运算。

不支持字符串的加减乘除比较等运算。

正数、负数等一元操作符只限于常量数字类型，比如-1,-0.1f等。

Null和除And,OR之外任何通用表达式计算返回值都是null。

Null And True 返回 Null. Null Or true 返回true. Null Or false 返回null, Null and false 返回false。

## 表达式常量

CQL语句中的常量只能用在表达式中，这些表达式可以使用在select子句、where子句、having子句等其他地方。

String：用双引号或者单引号包围起来。比如"Streaming"。

Int：正常表示，比如1或者-1.

Long：以字母L结尾，不区分大小写。比如1l，2L。

Float：以字母F结尾，不区分大小写。比如1.0f,-1.0F。

Double：以字母D结尾，不区分大小写。比如1.0D，-1.0D。

Decimal: 以字母BD结尾，不区分大小写。比如1.0BD，-1.0BD。

Boolean: true或者false。

Time: 不支持，只能通过UDF函数进行转换。

Date: 不支持，只能通过UDF函数进行转换。

Timestamp: 不支持，只能通过UDF函数进行转换。

## 特殊表达式

### Cast

* **语法**

|  |
| --- |
| castExpression  :Cast '(' (column|const) as datatype ')'  ; |

* **参数**

column：列名称。

const ： 常量。

datatype：CQL内部支持的数据类型。

* **返回类型**

和入参中datatype的数据类型一致。

* **功能**

将原始数据转化成其他数据类型。

* **示例**

|  |
| --- |
| INSERT INTO STREAM rs  SELECT  CAST(id AS int),  CAST(id AS long),  CAST(id AS float),  CAST(id AS double),  CAST(id AS String),  CAST(id AS boolean)  FROM s  WHERE NOT id != 'a'  AND type != 1; |

## 表达式数据类型转换规则

### 算术表达式

算术表达式包含表达式的加、减、乘、除以及比较操作。

两个不同的数据类型进行算数操作的时候，会自动向上转型, 不用经过特殊的函数转换。

CQL算术表达式数据类型转换规则从下到上依次为：

Int < Long < Float < Double < Decimal。

String、Boolean、Date、Time、TimeStamp数据类型不支持算术运算。

### CAST表达式

CAST表达式可以完成多种类型的数据转换。下面的表格展示了CQLde CAST表达式从原始数据类型到目标数据类型的转换规则。

CAST表达式类型转换支持列表

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | String | Boolean | Int | Long | Float | Double | Decimal | Time | Date | TimeStamp |
| String | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| Boolean | Y | Y |  |  |  |  |  |  |  |  |
| Int | Y | Y | Y | Y | Y | Y | Y |  |  |  |
| Long | Y | Y | Y | Y | Y | Y | Y |  |  |  |
| Float | Y | Y | Y | Y | Y | Y | Y |  |  |  |
| Double | Y | Y | Y | Y | Y | Y | Y |  |  |  |
| Decimal | Y | Y | Y | Y | Y | Y |  |  |  |  |
| Time | Y | Y | Y | Y | Y | Y |  |  |  |  |
| Date | Y | Y | Y | Y | Y | Y |  |  |  |  |
| TimeStamp | Y | Y | Y | Y | Y | Y |  |  |  |  |

说明

CAST的类型转换功能同类型转换函数一致。

时间类型的转换，一律使用Streaming集群默认时区，CQL目前不支持指定时区。

大于1的数字类型转为boolean类型都是true，小于1的数字类型转为boolean类型都是false。

# 用户自定义接口

[9.1 数据序列化/反序列化](#ZH-CN_TOPIC_0011584459)

[9.2 流数据读取](#ZH-CN_TOPIC_0011584461)

[9.3 流数据写入](#ZH-CN_TOPIC_0011584462)

[9.4 UDF函数](#ZH-CN_TOPIC_0011584463)

[9.5 数据源](#ZH-CN_TOPIC_0013092640)

[9.6 自定义算子](#ZH-CN_TOPIC_0013624603)

## 数据序列化/反序列化

* **说明**

数据的反序列化，必须实现com.huawei.streaming.serde.StreamSerDe接口。

也可以继承自com.huawei.streaming.serde.BaseSerDe类，该类提供了序列化的一些通用实现，比如创建空对象，数据类型转换等通用方法。

数据序列化反序列化接口说明

| 方法名 | 方法说明 | 参数 | 返回值 |
| --- | --- | --- | --- |
| setConfig | 设置序列化用到的配置属性。  编译期调用。 | StreamingConfig配置属性。 | Void |
| getConfig | 读取序列化类中的配置属性。  编译期调用。 | 无 | StreamingConfig配置属性 |
| initialize | 初始化接口。  编译期调用。 | StreamingConfig 配置参数。 | Void |
| setSchema | 设置输入或者输出schema。  编译期调用。  注意，该方法调用的比setConfig晚，所以不要在setConfig方法中读取Schema。 | TupleEventType  输入或者输出的schema。 | Void |
| deSerialize | 反序列化，将外部输入数据转化为系统可以识别的数据类型。  运行时调用。 | Object  外部第三方系统可以识别数据，比如纯字符串形式。 | List<Object[]>  内部可识别数据类型。  list的大小代表数据行的数量。  Object数组的大小和数据行中列的数量和顺序一致。 |
| serialize | 序列化，将系统计算结果转化为外部其他系统可以识别的数据。  运行时调用。 | List<Object[]>  内部计算结果  list的大小代表数据行的数量。  Object数组的大小和数据行中列的数量和顺序一致。 | Object  外部第三方系统可以识别数据，比如纯字符串形式。 |

* **示例**

|  |
| --- |
| package com.huawei.streaming.example.userdefined.serde;  import java.util.List;  import org.slf4j.Logger;  import org.slf4j.LoggerFactory;  import com.google.common.base.Splitter;  import com.google.common.collect.Lists;  import com.huawei.streaming.config.StreamingConfig;  import com.huawei.streaming.event.TupleEventType;  import com.huawei.streaming.exception.StreamSerDeException;  import com.huawei.streaming.exception.StreamingException;  import com.huawei.streaming.serde.StreamSerDe;  import com.huawei.streaming.util.DataTypeUtils;  /\*\*  \* 序列化和反序列化的例子  \*/  public class ExampleSerDe implements StreamSerDe  {  private static final long serialVersionUID = -8447913584480461044L;  private static final Logger LOG = LoggerFactory.getLogger(ExampleSerDe.class);  private TupleEventType schema;  private String separator = ",";  private StringBuilder sb = new StringBuilder();  /\*\*  \* {@inheritDoc}  \*/  @Override  public void setConfig(StreamingConfig arg0)  throws StreamingException  {  }  /\*\*  \* {@inheritDoc}  \*/  @Override  public StreamingConfig getConfig()  {  //读取配置属性  return null;  }  /\*\*  \* {@inheritDoc}  \*/  @Override  public void setSchema(TupleEventType outputSchema)  {  schema = outputSchema;  }  /\*\*  \* {@inheritDoc}  \*/  @Override  public TupleEventType getSchema()  {  return schema;  }  /\*\*  \* {@inheritDoc}  \*/  @Override  public void initialize()  throws StreamSerDeException  {  }  /\*\*  \* 反序列化  \* 将从输入流中读取的数据解析成系统可识别数据类型  \*  \* @param data 输入流中读取的数据。  \* 输入算子不同，反序列化类中的data数据类型也不同，  \* 一般是字符串类型  \* 在示例程序中，按照字符串进行解析  \* @return 解析好的数据  \* 由于可能从一行衍生出多行数据，所以返回值类型是数组类型  \* @throws StreamSerDeException 序列化异常  \* @see [类、类#方法、类#成员]  \*/  @Override  public List<Object[]> deSerialize(Object data)  throws StreamSerDeException  {  if (data == null)  {  return null;  }  String stringValue = (String)data;  List<Object[]> splitResults = Lists.newArrayList();  Object[] values = Lists.newArrayList(Splitter.on(separator).split(stringValue)).toArray();  splitResults.add(values);  return createEventsInstance(splitResults);  }  /\*\*  \* 序列化方法  \* 将传入的事件进行序列化，转成output可以识别的对象  \* 或者字符串，或者其他类型  \*  \* @param events 系统产生或者处理过后的事件，是一个数组  \* @return 序列化完毕的数据，可以是字符串之类  \* @throws StreamSerDeException 序列化异常  \* @see [类、类#方法、类#成员]  \*/  @Override  public Object serialize(List<Object[]> events)  throws StreamSerDeException  {  if (events == null)  {  LOG.info("Input event is null");  return null;  }  clearTmpString();  for (Object[] event : events)  {  serializeEvent(event);  }  return removeLastChar();  }  /\*\*  \* 根据schema中的数据列，生成对应的数据类型  \*/  private List<Object[]> createEventsInstance(List<Object[]> events)  throws StreamSerDeException  {  if (events == null || events.size() == 0)  {  return Lists.newArrayList();  }  List<Object[]> list = Lists.newArrayList();  for (Object[] event : events)  {  Object[] eventInstance = createEventInstance(event);  list.add(eventInstance);  }  return list;  }  /\*\*  \* 创建单个事件实例  \*/  private Object[] createEventInstance(Object[] event)  throws StreamSerDeException  {  validateColumnSize(event);  Object[] arr = new Object[schema.getAllAttributes().length];  for (int i = 0; i < schema.getAllAttributeTypes().length; i++)  {  arr[i] = createInstance(schema.getAllAttributeTypes()[i], event[i]);  }  return arr;  }  private String removeLastChar()  {  return sb.substring(0, sb.length() - 1);  }  /\*\*  \* 根据指定类型创建数据实例  \*/  private Object createInstance(Class< ? > clazz, Object value)  throws StreamSerDeException  {  try  {  return DataTypeUtils.createValue(clazz, value.toString());  }  catch (StreamingException e)  {  throw new StreamSerDeException(e.getMessage(), e);  }  }  /\*\*  \* 序列化单行事件  \*/  private void serializeEvent(Object[] event)  {  for (Object column : event)  {  appendToTmpString(column);  }  sb.replace(sb.length() - 1, sb.length() - 1, "");  }  /\*\*  \* 添加值到临时字符串最后面  \*/  private void appendToTmpString(Object val)  {  if (val != null)  {  sb.append(val.toString() + separator);  }  else  {  sb.append(separator);  }  }  /\*\*  \* 在序列化之前，先清空临时字符串  \*/  private void clearTmpString()  {  sb.delete(0, sb.length());  }  /\*\*  \* 检查列的数量和schema中列的数量是否一致  \*/  private void validateColumnSize(Object[] columns)  throws StreamSerDeException  {  if (columns.length != schema.getAllAttributeTypes().length)  {  LOG.error("deserializer result array size is not equal to the schema column size, "  + "schema size :{}, deserializer size :{}", schema.getAllAttributeTypes().length, columns.length);  throw new StreamSerDeException(  "deserializer result array size is not equal to the schema column size, schema size :"  + schema.getAllAttributeTypes().length + ", deserializer size :" + columns.length);  }  }  } |
| 使用例子：  打包成jar包，放在任意目录，比如/opt/streaming/example.jar  执行如下CQL  add jar "/opt/streaming/example.jar ";  create input stream S1  (  C1 STRING,  C2 STRING  )  SERDE "com.huawei.streaming.example.userdefined.serde.ExampleSerDe"  SOURCE kafkainput  PROPERTIES ( groupid = "gidkpi\_1\_1","topic"="agg\_1\_1\_1",  zookeepers="127.0.0.1:2181");  create output stream rs  (  C1 STRING,  C2 STRING  )  SERDE "com.huawei.streaming.example.userdefined.serde.ExampleSerDe"  SINK ConsoleOutput  PROPERTIES(printFrequence = "10");  insert into stream rs select \* from S1;  submit application force read; |

## 流数据读取

* **说明**

从外部事件源中读取数据,需要实现com.huawei.streaming.operator.IInputStreamOperator接口。

流数据读取接口说明

| 方法名 | 方法说明 | 参数 | 返回值 |
| --- | --- | --- | --- |
| setConfig | 设置算子用到的配置属性。  编译时调用。 | StreamingConfig | Void |
| setSerDe | 设置序列化/反序列化类。  运行时调用。 | StreamSerDe | Void |
| setEmitter | 设置Emitter对象。  运行时调用。 | IEmitter | Void |
| initialize | 初始化接口，运行时调用。 | 无 | Void |
| execute | 执行接口，运行时调用。 | 无 | Void |
| destroy | 关闭接口，运行时调用。 | 无 | Void |

* **示例**

|  |
| --- |
| package com.huawei.streaming.example.userdefined.operator.input;  import java.io.File;  import java.io.IOException;  import java.nio.charset.Charset;  import java.util.List;  import org.slf4j.Logger;  import org.slf4j.LoggerFactory;  import com.google.common.base.Strings;  import com.google.common.io.Files;  import com.google.common.io.LineProcessor;  import com.huawei.streaming.config.StreamingConfig;  import com.huawei.streaming.exception.StreamSerDeException;  import com.huawei.streaming.exception.StreamingException;  import com.huawei.streaming.exception.StreamingRuntimeException;  import com.huawei.streaming.operator.IEmitter;  import com.huawei.streaming.operator.IInputStreamOperator;  import com.huawei.streaming.serde.StreamSerDe;  /\*\*  \* 文件读取示例  \*/  public class FileInput implements IInputStreamOperator  {  private static final long serialVersionUID = 1145305812403368160L;  private static final Logger LOG = LoggerFactory.getLogger(FileInput.class);  private static final String CONF\_FILE\_PATH = "fileinput.path";  private static final Charset CHARSET = Charset.forName("UTF-8");  /\*\*  \* 文件路径  \*/  private String filePath;  private IEmitter emitter;  private StreamSerDe serde;  /\*\*  \* file在initialize接口中被实例化  \* 在运行时被调用  \* 所以不用进行序列化传输  \*/  private transient File file;  private transient FileLineProcessor processor;  /\*\*  \* {@inheritDoc}  \*/  @Override  public void setConfig(StreamingConfig conf)  throws StreamingException  {  this.filePath = conf.getStringValue(CONF\_FILE\_PATH);  }  /\*\*  \* {@inheritDoc}  \*/  @Override  public void setEmitter(IEmitter iEmitter)  {  this.emitter = iEmitter;  }  /\*\*  \* {@inheritDoc}  \*/  @Override  public void setSerDe(StreamSerDe streamSerDe)  {  this.serde = streamSerDe;  }  /\*\*  \* 初始化  \*  \* @throws StreamingException 初始化异常  \*/  @Override  public void initialize()  throws StreamingException  {  if (Strings.isNullOrEmpty(filePath))  {  LOG.error("file path is null.");  throw new StreamingException("file path is null.");  }  file = new File(filePath);  if (!file.exists())  {  LOG.error("file in path is not exists.");  throw new StreamingException("file in path is not exists.");  }  if (!file.isFile())  {  LOG.error("file in path is not a file type.");  throw new StreamingException("file in path is not a file type.");  }  processor = new FileLineProcessor();  }  /\*\*  \* 运行时的销毁接口  \*  \* @throws StreamingException 流处理异常  \* @see [类、类#方法、类#成员]  \*/  @Override  public void destroy()  throws StreamingException  {  }  /\*\*  \* 输入算子执行接口  \*  \* @throws StreamingException 流处理异常  \* @see [类、类#方法、类#成员]  \*/  @Override  public void execute()  throws StreamingException  {  try  {  Files.readLines(file, CHARSET, processor);  }  catch (IOException e)  {  throw StreamingException.wrapException(e);  }  }  private class FileLineProcessor implements LineProcessor<Object>  {  /\*\*  \* {@inheritDoc}  \*/  @Override  public boolean processLine(String line)  throws IOException  {  List<Object[]> deserResults = null;  try  {  deserResults = serde.deSerialize(line);  }  catch (StreamSerDeException e)  {  //忽略反序列化异常  LOG.warn("Ignore a serde exception.", e);  return false;  }  for (Object[] event : deserResults)  {  try  {  emitter.emit(event);  }  catch (StreamingException e)  {  //emit异常直接抛出，由外部异常处理机制自己处理  throw new StreamingRuntimeException(e);  }  }  return true;  }  /\*\*  \* {@inheritDoc}  \*/  @Override  public Object getResult()  {  return null;  }  }  } |
| 使用步骤：  打包成jar包，放在任意目录，比如/opt/streaming/example.jar  在CQL语句中使用。  add jar "/opt/streaming/example/example.jar";  create input stream S1  (  C1 STRING,  C2 STRING  )  SOURCE "com.huawei.streaming.example.userdefined.operator.input.FileInput"  PROPERTIES (  "fileinput.path" = "/opt/streaming/example/input.txt"  );  create output stream rs  (  C1 STRING,  C2 STRING  )  SINK "com.huawei.streaming.example.userdefined.operator.output.FileOutput"  PROPERTIES (  "fileOutput.path" = "/opt/streaming/example/output.txt"  );  insert into stream rs select \* from S1;  submit application force example; |

## 流数据写入

* **说明**

将计算好的结果写入外部事件源,需要实现com.huawei.streaming.operator. IOutputStreamOperator接口。

流数据写入接口说明

| 方法名 | 方法说明 | 参数 | 返回值 |
| --- | --- | --- | --- |
| setConfig | 设置算子用到的配置属性。  编译时调用。 | StreamingConfig | Void |
| setSerDe | 设置序列化/反序列化类。  运行时调用。 | StreamSerDe | Void |
| initialize | 初始化接口，运行时调用。 | 无 | Void |
| execute | 执行接口，运行时调用。 | String streamName：流名称。  TupleEvent：元组数据。 | Void |
| destroy | 关闭接口，运行时调用。 | 无 | Void |

* **示例**

|  |
| --- |
| package com.huawei.streaming.example.userdefined.operator.output;  import java.io.File;  import java.io.IOException;  import java.nio.charset.Charset;  import org.slf4j.Logger;  import org.slf4j.LoggerFactory;  import com.google.common.base.Strings;  import com.google.common.io.Files;  import com.huawei.streaming.config.StreamingConfig;  import com.huawei.streaming.event.TupleEvent;  import com.huawei.streaming.exception.StreamSerDeException;  import com.huawei.streaming.exception.StreamingException;  import com.huawei.streaming.operator.IOutputStreamOperator;  import com.huawei.streaming.serde.BaseSerDe;  import com.huawei.streaming.serde.StreamSerDe;  /\*\*  \* 用户自定义接口示例  \* 文件输出示例  \*/  public class FileOutput implements IOutputStreamOperator  {  private static final long serialVersionUID = 153729627538127379L;  private static final Logger LOG = LoggerFactory.getLogger(FileOutput.class);  private static final String CONF\_FILE\_PATH = "fileOutput.path";  private static final Charset CHARSET = Charset.forName("UTF-8");  /\*\*  \* 文件路径  \*/  private String filePath;  private StreamSerDe serde;  /\*\*  \* {@inheritDoc}  \*/  @Override  public void setConfig(StreamingConfig conf)  throws StreamingException  {  this.filePath = conf.getStringValue(CONF\_FILE\_PATH);  }  /\*\*  \* {@inheritDoc}  \*/  @Override  public void setSerDe(StreamSerDe streamSerDe)  {  this.serde = streamSerDe;  }  /\*\*  \* 初始化  \*  \* @throws StreamingException 初始化异常  \*/  @Override  public void initialize()  throws StreamingException  {  }  /\*\*  \* {@inheritDoc}  \*  \* @param streamName  \* @param event  \*/  @Override  public void execute(String streamName, TupleEvent event)  throws StreamingException  {  if (Strings.isNullOrEmpty(filePath))  {  LOG.error("file path is null.");  throw new StreamingException("file path is null.");  }  File file = new File(filePath);  if (!file.exists())  {  LOG.error("file in path is not exists.");  throw new StreamingException("file in path is not exists.");  }  if (!file.isFile())  {  LOG.error("file in path is not a file type.");  throw new StreamingException("file in path is not a file type.");  }  String result = serializeEvent(event);  writeEvent(file, result);  }  /\*\*  \* 运行时的销毁接口  \*  \* @throws com.huawei.streaming.exception.StreamingException 流处理异常  \* @see [类、类#方法、类#成员]  \*/  @Override  public void destroy()  throws StreamingException  {  }  private String serializeEvent(TupleEvent event)  {  String result = null;  try  {  result = (String)serde.serialize(BaseSerDe.changeEventsToList(event));  }  catch (StreamSerDeException e)  {  LOG.error("failed to serialize data ", e);  }  return result;  }  private void writeEvent(File file, String result)  throws StreamingException  {  try  {  Files.append(result, file, CHARSET);  }  catch (IOException e)  {  throw StreamingException.wrapException(e);  }  }  } |
| 打包成jar包，放在任意目录，比如/opt/streaming/example.jar  在CQL语句中使用。  add jar "/opt/streaming/example/example.jar";  create input stream S1  (  C1 STRING,  C2 STRING  )  SOURCE "com.huawei.streaming.example.userdefined.operator.input.FileInput"  PROPERTIES (  "fileinput.path" = "/opt/streaming/example/input.txt"  );  create output stream rs  (  C1 STRING,  C2 STRING  )  SINK "com.huawei.streaming.example.userdefined.operator.output.FileOutput"  PROPERTIES (  "fileOutput.path" = "/opt/streaming/example/output.txt"  );  insert into stream rs select \* from S1;  submit application force example; |

## UDF函数

* **说明**

用户自定义函数，可以在CQL的表达式中使用，可以将一个或者多个列进行处理之后，输出指定结果。

UDF函数接口说明

| 方法名 | 方法说明 | 参数 | 返回值 |
| --- | --- | --- | --- |
| 构造方法 | UDF函数构造方法 | Map<String, String>  配置参数。 | 无 |
| evaluate | 函数执行方法 | 不确定，UDF函数自定义，参数类型必须使用CQL支持数据类型，CQL中应用的时候，参数顺序必须和evaluate方法定义的顺序一致。 | 不确定，UDF函数自定义，但是必须是CQL支持的数据类型，且不能为Object。 |

用户自定义的UDF函数,一般继承自com.huawei.streaming.udfs.UDF这个类，构造参数只有一个Config，函数需要的所有的参数，都在这里放着。

参数可以通过set命令设置。

Udf函数必须实现evaluate方法，方法参数和返回值由用户依据使用确定，但是数据类型只能是系统支持的数据类型。

* **示例**

|  |
| --- |
| package com.huawei.streaming.example.udf;  import java.util.Map;  import com.huawei.streaming.udfs.UDF;  /\*\*  \* 移除字符串前后特殊字符  \*/  public class UDFTrim extends UDF  {  private static final long serialVersionUID = 4793756788804334850L;  /\*\*  \* <默认构造函数>  \*  \* @param config udf函数中需要的参数，  \* 这些参数要在cql中通过create function xx properties  \* 语法进行设置  \*/  public UDFTrim(Map< String, String > config)  {  super(config);  }  /\*\*  \* UDF函数的执行方法  \* 方法名称必须是evaluate。  \*  \* @param s 字符串  \* @return 移除空格之后的字符串  \* @see [类、类#方法、类#成员]  \*/  public String evaluate(String s)  {  if (s == null)  {  return null;  }  return s.trim();  }  } |

## 数据源

* **说明**

从外部数据源中读取数据，比如数据库。需要实现com.huawei.streaming.datasource IDataSource接口。

数据源接口说明

| 方法名 | 方法说明 | 参数 | 返回值 |
| --- | --- | --- | --- |
| setConfig | 设置数据源参数，编译期调用。 | StreamingConfig  查询参数 | void |
| setSchema | 设置数据源查询schema，编译时调用。 | TupleEventType  数据源查询结果对应schema | void |
| initialize | 初始化，运行时调用。 |  | void |
| Execute | 执行数据源查询，运行时调用。 | List< Object >  完成了参数替换的查询参数。  该数组参数就是在From子句中定义的查询参数，该参数按照定义顺序保存在list列表中，并且在调用的时候已经完成了参数替换。 | List< Object[] >  查询结果，list的大小为查询出的结果行数，object数组为每行的数据结果，数据类型和setSchema方法中的schema数据类型一一对应。 |
| destroy | 销毁数据源内部对象，运行时调用。 |  | void |

* **示例**

|  |
| --- |
| package com.huawei.streaming.example.userdefined.datasource;  import java.io.BufferedReader;  import java.io.File;  import java.io.IOException;  import java.nio.charset.Charset;  import java.util.List;  import java.util.Properties;  import org.slf4j.Logger;  import org.slf4j.LoggerFactory;  import com.google.common.base.Strings;  import com.google.common.collect.Lists;  import com.google.common.io.Closeables;  import com.google.common.io.Files;  import com.huawei.streaming.config.StreamingConfig;  import com.huawei.streaming.datasource.IDataSource;  import com.huawei.streaming.event.TupleEventType;  import com.huawei.streaming.exception.ErrorCode;  import com.huawei.streaming.exception.StreamingException;  /\*\*  \* 属性匹配数据源  \* 作为示例，从本地文件中获取配置文件  \* 读取配置文件中的key，value对应关系  \* 返回对应的查询值  \*/  public class PropertyMatchDataSource implements IDataSource  {  public static final Charset CHARSET = Charset.forName("UTF-8");  private static final long serialVersionUID = 8056232432674642637L;  private static final Logger LOG = LoggerFactory.getLogger(PropertyMatchDataSource.class);  private static final String CONF\_FILE\_PATH = "example.datasource.path";  private String propertyFilePath;  private Properties properties;  private StreamingConfig config;  /\*\*  \* 这里的schema，对应查询语句中的schema定义，指的是数据源查询的输出schema  \* 在这个示例中并没有用到  \*/  private TupleEventType schema;  /\*\*  \* {@inheritDoc}  \*/  @Override  public void setConfig(StreamingConfig conf)  throws StreamingException  {  config = conf;  }  /\*\*  \* {@inheritDoc}  \*/  @Override  public void setSchema(TupleEventType tuple)  {  schema = tuple;  }  /\*\*  \* {@inheritDoc}  \*/  @Override  public void initialize()  throws StreamingException  {  initConfig();  if (Strings.isNullOrEmpty(propertyFilePath))  {  LOG.error("file path is null.");  throw new StreamingException("file path is null.");  }  File file = new File(propertyFilePath);  validateFile(file);  loadProperties(file);  }  private void initConfig()  throws StreamingException  {  //初始化，用于读取配置属性  if (config.containsKey(CONF\_FILE\_PATH))  {  this.propertyFilePath = config.get(CONF\_FILE\_PATH).toString();  }  else  {  LOG.error("can not found {} from configuration.", CONF\_FILE\_PATH);  throw new StreamingException(ErrorCode.CONFIG\_NOT\_FOUND, CONF\_FILE\_PATH);  }  }  /\*\*  \* {@inheritDoc}  \*/  @Override  public List<Object[]> execute(List<Object> replacedQueryArguments)  throws StreamingException  {  validateArgs(replacedQueryArguments);  return evaluateValue(replacedQueryArguments.get(0));  }  /\*\*  \* {@inheritDoc}  \*/  @Override  public void destroy()  throws StreamingException  {  //用作打开资源的销毁，比如关闭流，关闭连接等  }  private List<Object[]> evaluateValue(Object replacedQueryArgument)  {  String key = replacedQueryArgument.toString();  if (properties.containsKey(key))  {  Object[] values = {properties.get(key)};  List<Object[]> results = Lists.newArrayList();  results.add(values);  return results;  }  return null;  }  private void validateFile(File file)  throws StreamingException  {  if (!file.exists())  {  LOG.error("file in path is not exists.");  throw new StreamingException("file in path is not exists.");  }  if (!file.isFile())  {  LOG.error("file in path is not a file type.");  throw new StreamingException("file in path is not a file type.");  }  }  private void loadProperties(File file)  throws StreamingException  {  properties = new Properties();  BufferedReader reader = null;  try  {  reader = Files.newReader(file, CHARSET);  properties.load(Files.newReader(file, CHARSET));  }  catch (IOException e)  {  LOG.error("failed to read property files.", e);  throw new StreamingException("failed to read property files.");  }  finally  {  Closeables.closeQuietly(reader);  }  }  private void validateArgs(List<Object> replacedQueryArguments)  throws StreamingException  {  if (replacedQueryArguments.size() != 1)  {  LOG.error("rdb dataSource query arguments are not equal to 1, args size : {}", replacedQueryArguments.size());  throw new StreamingException("rdb dataSource query arguments are not equal to 1, args size : "  + replacedQueryArguments.size());  }  }  } |
| 使用步骤：  打包成jar包，放在任意目录，比如/opt/streaming/example.jar  执行add jar '/opt/streaming/example.jar';  在CQL语句中使用。  create input stream S  (id int,name String)  SOURCE KafkaInput  PROPERTIES (  groupid = "gidkpi\_1\_1",  topic = "0912",  zookeepers = "127.0.0.1:2181"  );  CREATE DATASOURCE propertyDataSource  SOURCE "com.huawei.streaming.example.datasource.PropertyMatchDataSource"  PROPERTIES (  "example.datasource.path" = "/opt/example.properties");  create output stream rs  (type String, id String, tag String)  SINK consoleOutput;  insert into rs select s.id,s.name,ds.value from S,  DATASOURCE propertyDataSource  [  SCHEMA (value String),  QUERY(s.id)  ] ds;  submit application datasourcetest; |

## 自定义算子

* **说明**

用户自定义处理逻辑，目前仅支持单流输入和单流输出。需要实现com.huawei.streaming.operator.IFunctionStreamOperator接口。

自定义算子接口说明

| 方法名 | 方法说明 | 参数 | 返回值 |
| --- | --- | --- | --- |
| setConfig(StreamingConfig conf) | StreamingConfig：配置 参数 | Void | 设置配置参数，编译时接口 |
| setEmitter(Map< String, IEmitter > emitterMap) | Map中的key为流名称，value为emitter实例 | void | 设置多个emiter，每个流对应一个emiter实例。运行时接口 |
| Initialize() |  | Void | 运行时初始化接口 |
| execute(String streamName, TupleEvent event) | StreamingName：元组数据所在流名称  TupleEvent：元组数据 | Void | 运行接口，数据的处理逻辑就发生在该接口内 |
| Destroy() | 关闭算子 | Void |  |

* **示例**

|  |
| --- |
| package com.huawei.streaming.example.userdefined.operator;  import java.io.BufferedReader;  import java.io.File;  import java.io.IOException;  import java.nio.charset.Charset;  import java.util.Map;  import java.util.Properties;  import org.slf4j.Logger;  import org.slf4j.LoggerFactory;  import com.google.common.io.Closeables;  import com.google.common.io.Files;  import com.huawei.streaming.config.StreamingConfig;  import com.huawei.streaming.event.TupleEvent;  import com.huawei.streaming.exception.StreamingException;  import com.huawei.streaming.exception.StreamingRuntimeException;  import com.huawei.streaming.operator.IEmitter;  import com.huawei.streaming.operator.IFunctionStreamOperator;  /\*\*  \* 用户自定义算子  \* 如果发送的数据在文件中存在，则直接替换成文件内容  \*  \* 输入schema：a string, b int  \* 输出schema：c string, d int, e float  \*/  public class UserOperator implements IFunctionStreamOperator  {  private static final Logger LOG = LoggerFactory.getLogger(UserOperator.class);  public static final Charset CHARSET = Charset.forName("UTF-8");  private static final long serialVersionUID = -4438239751340766284L;  private static final String CONF\_FILE\_NAME = "userop.filename";  private String fileName;  private Properties properties;  private Map<String, IEmitter> emitters = null;  /\*\*  \* {@inheritDoc}  \*/  @Override  public void setConfig(StreamingConfig conf)  throws StreamingException  {  if (!conf.containsKey(CONF\_FILE\_NAME))  {  LOG.error("can not found config value {}.", CONF\_FILE\_NAME);  throw new StreamingException("can not found config value " + CONF\_FILE\_NAME + ".");  }  fileName = conf.getStringValue(CONF\_FILE\_NAME);  }  /\*\*  \* {@inheritDoc}  \*/  @Override  public void setEmitter(Map<String, IEmitter> emitterMap)  {  if (emitterMap == null || emitterMap.isEmpty())  {  LOG.error("can not found emitter.");  throw new StreamingRuntimeException("can not found config value " + CONF\_FILE\_NAME + ".");  }  emitters = emitterMap;  }  /\*\*  \* {@inheritDoc}  \*/  @Override  public void initialize()  throws StreamingException  {  File file = new File(fileName);  validateFile(file);  loadProperties(file);  }  /\*\*  \* {@inheritDoc}  \*/  @Override  public void execute(String streamName, TupleEvent event)  throws StreamingException  {  Object[] values = event.getAllValues();  Object[] result = new Object[3];  if (properties.containsKey(String.valueOf(values[0])))  {  result[0] = properties.get(String.valueOf(values[0]));  result[1] = 1;  result[2] = 1.0f;  }  else  {  result[0] = "NONE";  result[1] = 1;  result[2] = 1.0F;  }  for (IEmitter emitter : emitters.values())  {  emitter.emit(result);  }  }  /\*\*  \* {@inheritDoc}  \*/  @Override  public void destroy()  throws StreamingException  {  }  private void validateFile(File file)  throws StreamingException  {  if (!file.exists())  {  LOG.error("file in path is not exists.");  throw new StreamingException("file in path is not exists.");  }  if (!file.isFile())  {  LOG.error("file in path is not a file type.");  throw new StreamingException("file in path is not a file type.");  }  }  private void loadProperties(File file)  throws StreamingException  {  properties = new Properties();  BufferedReader reader = null;  try  {  reader = Files.newReader(file, CHARSET);  properties.load(Files.newReader(file, CHARSET));  }  catch (IOException e)  {  LOG.error("failed to read property files.", e);  throw new StreamingException("failed to read property files.");  }  finally  {  Closeables.closeQuietly(reader);  }  }  } |
| 使用步骤：  打包成jar包，放在任意目录，比如/opt/streaming/example.jar  执行add jar '/opt/streaming/example.jar';  在CQL语句中使用。  **create** input stream s (id int, name string)  SOURCE KafkaInput  PROPERTIES ( groupid = **"gidkpi\_1\_1"**,topic=**"agg\_1\_1"**);  **create** output stream rs  (id string, name string, type int)  SINK KafkaOutput properties(topic=**"agg\_2"**);  **create** operator userOp **as** **"com.huawei.streaming.example.userdefined.operator.UserOperator"**  input (id int, name string)  output (newID string, name string, type int)  properties (**"userop.filename"** = **"/home/omm/kv.properties"**);  **insert** **into** rs using operator userOp **from** s distribute **by** id parallel **2**;  submit application force tt; |

# 系统内置接口实现

本章描述各类系统内置的实现及其参数，以供在CQL中使用。

各种内置接口参数的默认值仅供参考，可能会和实际使用场景有所不同。

说明

系统内置实现接口内所有参数名称不区分大小写

[10.1 序列化/反序列化](#ZH-CN_TOPIC_0011584465)

[10.2 数据读取](#ZH-CN_TOPIC_0011584470)

[10.3 数据写入](#ZH-CN_TOPIC_0011584476)

[10.4 函数](#ZH-CN_TOPIC_0011584481)

[10.5 数据源](#ZH-CN_TOPIC_0013092641)

## 序列化/反序列化

### SimpleSerde

* **功能**

按照指定分隔符拆分数据并根据schema类型进行反序列化。



Boolean类型的反序列化，如果字符串是true（不区分大小写），则转为true，如果字符串是false（不区分大小写）则转为false，除此之外，其他字符都是null。

* **参数**

系统默认序列化反序列化接口参数

| 名称 | 默认值 | 参数说明 |
| --- | --- | --- |
| separator | , | 每个列之间的分隔符。 |

* **示例**

|  |
| --- |
| CREATE INPUT STREAM S  (id INT, name STRING )  SERDE simpleSerde  PROPERTIES ( separator = "," )  SOURCE KafkaInput  PROPERTIES (  groupid = "gidkpi\_1\_1",  topic = "simple\_1"  ); |

### CSVSerDe

* **功能**
* CSV格式的序列化的反序列化。



Boolean类型的反序列化，如果字符串是true（不区分大小写），则转为true，如果字符串是false（不区分大小写），则转为false，除此之外，其他字符都是null。

整个列里面如果包含分隔符，则用引号将整个列包围；一个引号在CSV格式中用两个引号代替。

CSV格式举例

| 原始数据 | CSV格式 |
| --- | --- |
| 1,a,b | 1,a,b |
| 2,"a","b" | 2,"""a""","""b""" |
| 3,a,b,"a,b" | 3,"a,b","""a,b""" |

* **参数**
* **示例**

|  |
| --- |
| CREATE INPUT STREAM S  (id INT, name STRING )  SERDE CSVSerDe  SOURCE KafkaInput  PROPERTIES (  groupid = "gidkpi\_1\_1",  topic = "simple\_1"  ); |

## 数据读取

### KafkaInput

* **功能**

从Kafka消息队列中读取数据的算子。

* **参数**

Kafka输入算子参数

| 名称 | 默认值 | 参数说明 |
| --- | --- | --- |
| groupid | 无 | 分组Id，相当于客户端ID。 |
| topic | 无 | 消息的分类，不同类型的消息必须使用不同的topic。 |
| zookeepers | 默认配置同集群内Kafka使用的zookeeper地址 | kafka读取数据的zookeeper地址，地址加端口，多个之间用逗号分隔，比如：192.168.0.20:2181,192.168.0.21:2181/kafka。 |
| zksessiontimeout | 20000 | zookeeper连接超时时间。 |
| zksynctime | 20000 | kafka zookeeper同步时间参数。 |
| messageserializerclass | kafka.serializer.StringEncoder | kafka消费者中获取到数据之后的序列化类。 |
| fromBeginning | false | 是否从topic开始位置读取数据，如果是false，则从kafkainput启动之后在kafka中找到的最新位置读起。 |

* **示例**

|  |
| --- |
| CREATE INPUT STREAM S  (id INT, name STRING )  SERDE simpleSerde  PROPERTIES ( separator = "," )  SOURCE KafkaInput  PROPERTIES (  groupid = "gidkpi\_1\_1",  topic = "simple\_1",  zookeepers = "localhost:2181,10.0.0.21:2181"  ); |

### RamdomGen

* **功能**

生成随机数的输入源。

* **参数**

随机数生成算子参数

| 名称 | 默认值 | 参数说明 |
| --- | --- | --- |
| timeunit | SECONDS | 时间单位，默认秒,配置参数，可以配置如下： MILLISECONDS，SECONDS，MINUTES，HOURS，DAYS。 |
| period | 1 | 时间周期，默认1秒，配置参数。 |
| eventnumperperiod | 1 | 每时间周期发送事件数据，默认1个。即发送数据 1个/秒，配置参数。 |
| isschedule | False | 是否定期发送，配置参数 |
| totalnumber | 0 | 发送个数，如果为0表示个数无限，配置参数。 |
| delaytime | 0 | 延迟发送时间，如果为0表示立即发送，配置参数。 |

* **示例**

|  |
| --- |
| CREATE INPUT STREAM S  (id INT, name STRING, type INT)  SOURCE randomgen  PROPERTIES (  timeUnit = "SECONDS",  period = "1",  eventNumPerPeriod = "1",  isSchedule = "true",  totalNumber = "20000",  delayTime = "0"); |

## 数据写入

### KafkaOutput

* **功能**

将数据写入kafka消息队列的算子。

* **参数**

Kafka输出算子参数

| 名称 | 默认值 | 参数说明 |
| --- | --- | --- |
| topic | 无 | 消息的分类，不同类型的消息必须使用不同的topic。 |
| zookeepers | 默认配置同集群内Kafka使用的zookeeper地址。 | kafka读取数据的zookeeper地址，格式为地址加端口，多个地址之间用逗号分隔，比如：10.0.0.5:2181, 10.0.0.6:2181/kafka。 |
| brokers | 默认配置同集群内Kafka使用的brokers地址。 | 消费者获取消息元信息(topics, partitions and replicas)的地址,配置格式是：host1:port1,host2:port2。 |
| zksessiontimeout | 20000 | zookeeper连接超时时间。 |
| zksynctime | 20000 | kafka zookeeper同步时间参数。 |
| messageserializerclass | kafka.serializer.StringEncoder | kafka消费者中获取到数据之后的序列化类。 |

* **示例**

|  |
| --- |
| CREATE OUTPUT STREAM rs  (id STRING, name STRING)  SINK KafkaOutput  PROPERTIES (  topic = "simple\_2",  zookeepers = "10.0.0.12:2181",  brokers = "10.0.0.12:9092"); |

### ConsoleOutput

* **功能**

将收到的数据直接在控制台打印，不输出。

* **参数**

控制台输出算子参数

| 名称 | 默认值 | 参数说明 |
| --- | --- | --- |
| printfrequence | 1 | 计数频率，默认每条都回打印，如果设置成10，那么就会每隔10条消息打印一次 |

* **示例**

|  |
| --- |
| CREATE OUTPUT STREAM FilterHO  (  IMSI STRING,  EventID STRING,  HOCount INT  )  SINK ConsoleOutput  PROPERTIES(printFrequence = "10"); |

## 函数

函数可以使用在select子句，where子句，having中。目前group by，order by子句不支持使用函数。

系统会内嵌一些经常使用到的函数，如果用户有特殊需求，可以手工定义。

如果用于定义的函数名和系统内置函数名一致，则用户的函数会覆盖当前线程内的系统函数，但是不会对其他用户造成影响，进程结束之后，该覆盖效果也会消失。

### 字符处理函数

字符处理函数用于完成字符方面的处理。如：获取参数的长度、对参数进行类型转换、对参数进行拼接，以及对参数的部分内容进行替换等。

字符串处理函数

| 函数 | 返回值 | 描述 | 示例 |
| --- | --- | --- | --- |
| substr(string s, int from) | String | 从参数string中抽取子字符串。  from表示抽取的起始位置。  字符串索引计数从1开始。  from为0时，按1处理。  from为正数时，抽取从from到末尾的所有字符。  from为负数时，抽取字符串的后n个字符，n为from的绝对值。 | substr("abc",2) --bc  substr("abc",0) --abc  substr("abc",1) --abc  substr("abc",-1) --c |
| substr(string s, int from, int count) | String | 从参数string中抽取子字符串。  from表示抽取的起始位置。  字符串索引计数从1开始。  count表示抽取的子字符串长度。  from为0时，按1处理。  from为正数时，抽取从 from开始的count个字符。  from为负数时，抽取从倒数第n个开始的count 个字符，n为from 的绝对值。  count小于1时，返回空字符串。 | substr("abc",0,2) --ab  substr("abc",1,2) --ab  substr("abc",-1,1) --c  substr("abc",2,-1) --空字符串 |
| strlength(string s) | Int | 返回字符串的长度  如果字符串为null，返回0。 | strlength(null) --0  strlength("") --0  strlength("abc") --3 |
| trim(String s) | String | 移除字符串开始和结尾的特殊字符。  特殊字符指ASCII码值小于33的特殊字符，包含空格，制表符、退格、回车等等，详情可以参照ASCII码表。 | trim(null) --空字符串  trim("") --空字符串  trim(" ") --空字符串  trim("abc") --abc  trim("abc ") –abc  trim("abc") –abc  trim("abc\n") –abc 最后一位是回车符，trim截去该符号 |
| concat(String s1, String s2) | String | 字符串拼接，按照参数出现的顺序进行拼接，中间任何一个参数为null，都会导致所有结果为null。 | concat("a","b") –ab  concat("ab","") --ab  concat("ab",null) --null |
| concat(String s1, String s2, String s3) | String | 字符串拼接，按照参数出现的顺序进行拼接，中间任何一个参数为null，都会导致所有结果为null。 | concat("a","b","c") --abc  concat("ab","","c") --abc  concat("ab",null,"c") --null |
| concat(String s1, String s2, String s3, String s4) | String | 字符串拼接，按照参数出现的顺序进行拼接，中间任何一个参数为null，都会导致所有结果为null。 | concat("a","b","c","d") --abcd  concat("ab","","c","d") --abcd  concat("ab",null,"c","d") --null |
| concat(String s1, String s2, String s3, String s4, String s5) | String | 字符串拼接，按照参数出现的顺序进行拼接，中间任何一个参数为null，都会导致所有结果为null。 | concat("a","b","c","d","e") --abcde  concat("ab","","c","d","e") --abcde  concat("ab",null,"c","d","e") --null |
| concat(String s1, String s2, String s3, String s4, String s5, String s6) | String | 字符串拼接，按照参数出现的顺序进行拼接，中间任何一个参数为null，都会导致所有结果为null。 | concat("a","b","c","d","e","f") --abcdef  concat("ab","","c","d","e","f") --abcdef  concat("ab",null,"c","d","e","f") --null |
| concat(String s1, String s2, String s3, String s4, String s5, String s6, String s7) | String | 字符串拼接，按照参数出现的顺序进行拼接，中间任何一个参数为null，都会导致所有结果为null。 | concat("a","b","c","d","e","f","g") --abcdefg  concat("ab","","c","d","e","f","g") --abcdefg  concat("ab",null,"c","d","e","f","g") --null |

### 时间处理函数

时间处理函数主要包含对时间数据的操作，包含加减以及获取年月日等操作。

时间处理函数

| 函数 | 返回值 | 描述 | 示例 |
| --- | --- | --- | --- |
| day(String dateString) | int | 返回指定时间字符串的当前月所在天数。  时间字符串可以是date类型或者timestmap类型，具体格式同CQL数据类型一致。 | day("2013-02-28") --28  day("2013-02-29") --null  day("2013-02-29 09:00:01") --null  day("2013-02-28 09:00:01") --28  day("2013-02-28 09:00:01.0") --28 |
| dayofmonth(String dateString) | Int | 返回指定时间字符串的当前月所在天数。  时间字符串可以是date类型或者timestmap类型，具体格式同CQL数据类型一致。 | dayofmonth("2013-02-28") --28  dayofmonth("2013-02-29") --null  dayofmonth("2013-02-29 09:00:01") --null  dayofmonth("2013-02-28 09:00:01") --28  dayofmonth("2013-02-28 09:00:01.0") --28 |
| month(String dateString) | int | 返回指定时间字符串的所在月。  时间字符串可以是date类型或者timestmap类型，具体格式同CQL数据类型一致。 | month("2013-02-28") --2  month("2013-02-29") --null  month("2013-02-29 09:00:01") --null  month("2013-02-28 09:00:01") --2  month("2013-02-28 09:00:01.0") --2 |
| year(String dateString) | Int | 返回指定时间字符串的所在年。  时间字符串可以是date类型或者timestmap类型，具体格式同CQL数据类型一致。 | year("2013-10-17 09:58:00.111") --2013  year("2013-10-17 09:58:00") --2013  year("2013-10-17 25:62:00") --null  year("2013-00-17") --null  year("2013-10-17") --2013  year("09:58:00") --null |
| hour(String dateString) | Int | 返回指定时间字符串的所在小时。  时间字符串可以是time类型或者timestmap类型，具体格式同CQL数据类型一致。 | hour("2013-10-17 09:58:00.111") --9  hour("2013-10-17 09:58:00") --9  hour("09:58:00") --9  hour("24:58:00") --null  hour("00:58:00") --0  hour("2013-10-17") --null |
| minute(String dateString) | Int | 返回指定时间字符串的所在分钟。  时间字符串可以是time类型或者timestmap类型，具体格式同CQL数据类型一致。 | minute("2013-10-17 09:58:00.111") -- 58  minute("2013-10-17 09:58:00") -- 58  minute("09:58:00") -- 58  minute("09:60:00") -- null  minute("09:00:00") -- 0  minute("2013-10-17") -- null |
| second(String dateString) | Int | 返回指定时间字符串的所在秒。  时间字符串可以是time类型或者timestmap类型，具体格式同CQL数据类型一致。 | second("2013-10-17 09:58:00.111") -- 0  second("2013-10-17 09:58:00") -- 0  second("09:58:00") -- 0  second("24:58:60") -- null  second("23:59:59") -- 59  second("24:00:00") -- null  second("2013-10-17") -- null |
| weekofyear(String dateString) | Int | 返回指定时间字符串的所在年中的星期数。  时间字符串可以是time类型或者timestmap类型，具体格式同CQL数据类型一致。 | weekofyear("2013-01-01 09:58:00.111") -- 1  weekofyear("2013-01-01 09:58:00") -- 1  weekofyear("2013-12-28") -- 52  weekofyear("2013-12-52") -- null  weekofyear("2013-01-01 09:58:70") -- null |
| weekofyear(Date date) | Int | 返回指定时间的所在年中的星期数。 | weekofyear(todate("2013-1-1")) -- 1  weekofyear(todate("2013-01-01")) -- 1  weekofyear(todate("2013-12-28")) -- 52  weekofyear(todate("2013-12-52")) -- null |
| weekofyear(TimeStamp timestamp) | Int | 返回指定时间的所在年中的星期数。 | weekofyear(totimestamp("2013-01-01 09:58:00.111")) -- 1  weekofyear(totimestamp("2013-01-01 09:58:00")) -- 1  weekofyear(totimestamp("2013-01-01 09:58:70")) -- null |

### 数学函数

说明

系统不会检查所有的数学运算是否溢出，需要用户自己保证。

数学函数

| 函数 | 返回值 | 描述 | 示例 |
| --- | --- | --- | --- |
| abs(int number) | Int | 返回数字类型的绝对值。  说明：系统不会检查是否溢出，需要用户自己保证数据不会溢出。 | abs(1) -- 1  abs(-1) -- 1  abs(0) -- 0  abs(-0) -- 0  abs(2147483647) -- 2147483647  abs(-2147483648) -- -2147483648 |
| abs(long number) | Long | 返回数字类型的绝对值。  说明：系统不会检查是否溢出，需要用户自己保证数据不会溢出。 | abs(1L) -- 1  abs(-1L) -- 1  abs(0L) -- 0  abs(-0L) -- 0 |
| abs(float number) | float | 返回数字类型的绝对值。  说明：系统不会检查是否溢出，需要用户自己保证数据不会溢出。 | abs(1.0f) -- 1.0  abs(-1.0f) - 1.0  abs(0.0F) -- 0.0  abs(-0.0F) -- 0.0 |
| abs(double number) | double | 返回数字类型的绝对值。  说明：系统不会检查是否溢出，需要用户自己保证数据不会溢出。 | abs(1.0d) -- 1.0  abs(-1.0d) -- 1.0  abs(0.0D) -- 0.0  abs(-0.0D) -- 0.0 |
| abs(decimal number) | decimal | 返回数字类型的绝对值。  说明：系统不会检查是否溢出，需要用户自己保证数据不会溢出 | abs(1.0BD) -- 1.0  abs(-1.0BD) -- 1.0  abs(0.0BD) -- 0.0  abs(-0.0BD) -- 0.0 |

### 类型转换函数

类型转换函数函数

| 函数 | 返回值 | 描述 | 示例 |
| --- | --- | --- | --- |
| tostring(object obj) | String | 将传入的数据转为字符串格式。  支持任何数据类型。 | tostring(true) -- "true"  tostring(false) -- "false"  tostring(null) -- null  tostring(1) -- "1"  tostring(1L) -- "1"  tostring(1.0F) -- "1.0"  tostring(1.0D) -- "1.0"  tostring(toDecimal("1.0")) -- "1.0"  tostring(toTime("15:40:00")) -- "15:40:00"  tostring(toDate("2013-10-17")) -- "2013-10-17"  tostring(toDate("2013-1-17")) -- "2013-01-17"  tostring(toTimeStamp("2013-10-17 15:40:00.000000")) -- "2013-10-17 15:40:00.000"  tostring(toTimeStamp("2013-10-17 15:40:00")) -- "2013-10-17 15:40:00.000" |
| toint(string number)  toint(int number)  toint(long number)  toint(float number)  toint(double number)  toint(date number)  toint(time number)  toint(timestamp number)  toint(decimal number) | int | 将指定类型的数据转换为int类型。 | toint(1) -- 1  toint(1L) -- 1  toint(1.0f) -- 1  toint(1.4f) -- 1  toint(1.5f) -- 1  toint(1.6f) -- 1  toint(1.9f) -- 1  toint(1.9d) -- 1  toint("1") -- 1  toint("1.9") -- null  toint(toDecimal("1.9")) -- 1  toint(toDate("1970-01-01")) -- -28800000  toint(toTime("15:40:00")) -- 27600000  toint(toTimeStamp("1970-01-01 15:40:00.000000")) -- 27600000 |
| tolong(string number)  tolong(int number)  tolong(long number)  tolong(float number)  tolong(double number)  tolong(date number)  tolong(time number)  tolong(timestamp number)  tolong(decimal number) | long | 将指定类型的数据转换为long类型。 | tolong(1) -- 1L  tolong(1L) -- 1L  tolong(1.0f) -- 1L  tolong(1.4f) -- 1L  tolong(1.5f) -- 1L  tolong(1.6f) -- 1L  tolong(1.9f) -- 1L  tolong(1.9d) -- 1L  tolong("1") -- 1L  tolong("1.9") -- null  tolong(toDecimal("1.9")) -- 1L  tolong(toDate("1970-01-01")) -- -28800000L  tolong(toTime("15:40:00")) -- 27600000L  tolong(toTimeStamp("1970-01-01 15:40:00.000000")) -- 27600000L |
| tofloat(string number)  tofloat(int number)  tofloat(long number)  tofloat(float number)  tofloat(double number)  tofloat(date number)  tofloat(time number)  tofloat(timestamp number)  tofloat(decimal number) | float | 将指定类型的数据转换为float类型。 | tofloat(1) -- 1.0F  tofloat(1F) -- 1.0F  tofloat(1.0f) -- 1.0F  tofloat(1.4f) -- 1.4F  tofloat(1.5f) -- 1.5F  tofloat(1.6f) -- 1.6F  tofloat(1.9f) -- 1.9F  tofloat(1.9d) -- 1.9F  tofloat("1") -- 1.0F  tofloat("1.9") -- 1.9F  tofloat(toDecimal("1.9")) -- 1.9F  tofloat(toDate("1970-01-01")) -- -28800000.0F  tofloat(toTime("15:40:00")) -- 27600000.0F  tofloat(toTimeStamp("1970-01-01 15:40:00.000000")) -- 27600000.0F |
| todouble(string number)  todouble(int number)  todouble(long number)  todouble(float number)  todouble(double number)  todouble(date number)  todouble(time number)  todouble(timestamp number)  todouble(decimal number) | double | 将指定类型的数据转换为double类型。 | todouble(1) -- 1.0D  todouble(1F) -- 1.0D  todouble(1.0f) -- 1.0D  todouble(1.9d) -- 1.9D  todouble("1") -- 1.0D  todouble("1.9") -- 1.9D  todouble(toDecimal("1.9")) -- 1.9D  todouble(toDate("1970-01-01")) -- -28800000.0D  todouble(toTime("15:40:00")) -- 27600000.0D  todouble(toTimeStamp("1970-01-01 15:40:00.000000")) -- 27600000.0D |
| toboolean(string value)  toboolean(int value)  toboolean(long value)  toboolean(float value)  toboolean(double value)  toboolean(date value)  toboolean(time value)  toboolean(timestamp value)  toboolean(decimal value)  toboolean(boolean value) | boolean | 将指定类型的数据转换为boolean类型。  数字类型和时间类型，只要最后转为long类型的值不为0，就都返回true。  非空字符串都返回true。 | toboolean(0) -- false  toboolean(1) -- true  toboolean(0l) -- false  toboolean(1l) -- true  toboolean(0F) -- false  toboolean(1f) -- true  toboolean(0d) -- false  toboolean(1D) -- true);  toboolean(0.9d) -- false  toboolean(0.9f) -- false  toboolean(0BD) -- false  toboolean(0.0BD) -- false  toboolean(1BD) -- true  toboolean("") -- false  toboolean("0") -- true  toboolean("1") -- true  toboolean("true") -- true  toboolean("false") -- true  toboolean(totime("17:06:00")) -- true  toboolean(todate("2013-10-17")) -- true  toboolean(totimestamp("2013-10-17 17:06:00.000000")) -- true |
| todate(string s) | date | 将指定类型的数据转换为date类型。  支持yyyy-[M]M-[d]d形式的字符串。 | todate("2014-09-25 17:07:00") -- null  todate("2014-09-25") – "2014-09-25"  todate("17:07:00") -- -- null |
| todate(string s, string format) | date | 将指定类型的数据转换为date类型。  时间格式可以为：  "yyyy-MM-dd HH:mm:ss"  "yyyy-MM-dd"  "yyyy-MM-dd HH:mm:ss.SSS" | todate("2014-09-25 17:07:00", "yyyy-MM-dd HH:mm:ss") -- "2014-09-25"  todate("2014-09-25 17:07:01", "yyyy-MM-dd HH:mm:ss") -- "2014-09-25"  todate("2014-09-25", "yyyy-MM-dd") -- "2014-09-25" |
| totime(string s) | time | 将指定类型的数据转换为time类型。 | totime("2014-09-25 17:07:00") -- null  totime("17:07:00") -- "17:07:00"  totime("17:07:90") -- null  totime("16:44:00") -- "16:44:00" |
| totimestamp(string s) | timestamp | 将指定类型的数据转换为timestamp类型。 | totimestamp("2014-09-25 17:07:00") -- "2014-09-25 17:07:00"  totimestamp("2014-09-25 17:07:00.056") -- "2014-09-25 17:07:00.056"  totimestamp("abc") -- null  totimestamp("2014-09-25 17:07:00.56") -- "2014-09-25 17:07:00.56"  totimestamp("2014-09-25 17:07:00.9999999") -- "2014-09-25 17:07:00.56"  totimestamp("2014-09-25 17:07:00.9999999999") -- null  totimestamp("2014-09-25 17:07:00.999999") -- "2014-09-25 17:07:00.999999"  totimestamp("2014-9-9 7:7:5") -- "2014-09-09 07:07:05"  totimestamp("2014-9-9 7:7:5.9") -- "2014-09-09 07:07:05" |
| todecimal(string number)  todecimal(int number)  todecimal(long number)  todecimal(double number)  todecimal(float number) | decimal | 将指定类型的数据转换为decimal类型。 | todecimal(1) -- "1"  todecimal(1F) -- "1.0"  todecimal(1.0f) -- "1.0"  todecimal(1.9d) -- "1.9"  todecimal("1") -- "1.0"  todecimal("1.9") -- "1.9" |
| todecimal(long number, int scale) | decimal | 将指定类型的数据转换为decimal类型。  可以通过设置scale参数指定标度。 | todecimal(1L,2) -- "0.01"  todecimal(100000L,2) -- "1000.00" |

### 聚合函数

聚合处理类函数，有多个输入，一个输出，将多个值合并成一个输出。

UDAF函数

| 函数 | 返回值 | 描述 | 示例 |
| --- | --- | --- | --- |
| count(Object value) | long | 计算窗口内数据的数量，Null会被忽略，不会参与计算。 | Count(1)  Count(\*)  Count(id) |
| count(Object value, expression) | long | 计算窗口内满足表达式条件数据的数量。  Null会被忽略，不会参与计算。 | Count(id, type = "1") |
| sum(int number) | int | 计算窗口内数据的总和，数据列必须是int类型。  Null会被忽略，不会参与计算。 | Sum(intprice) |
| sum(long number) | long | 计算窗口内数据的总和，数据列必须是long类型。  Null会被忽略，不会参与计算。 | Sum(longprice) |
| sum(float number) | float | 计算窗口内数据的总和，数据列必须是float类型。  Null会被忽略，不会参与计算。 | Sum(floatprice) |
| sum(double number) | double | 计算窗口内数据的总和，数据列必须是double类型。  Null会被忽略，不会参与计算。 | Sum(doubleprice) |
| sum(decimal number) | Decimal | 计算窗口内数据的总和，数据列必须是decimal类型。  Null会被忽略，不会参与计算。 | Sum(decimalprice) |
| sum(int number, expression) | Int | 计算窗口内满足表达式条件数据的总和，数据列必须是int类型。  Null会被忽略，不会参与计算。 | Sum(intprice, type = "1") |
| sum(long number, expression) | long | 计算窗口内满足表达式条件数据的总和，数据列必须是long类型。  Null会被忽略，不会参与计算。 | Sum(longprice, type = "1") |
| sum(float number, expression) | float | 计算窗口内满足表达式条件数据的总和，数据列必须是float类型。  Null会被忽略，不会参与计算。 | Sum(floatprice, type = "1") |
| sum(double number, expression) | double | 计算窗口内满足表达式条件数据的总和，数据列必须是double类型。  Null会被忽略，不会参与计算。 | Sum(doubleprice, type = "1") |
| Sum(decimal number, expression) | Decimal | 计算窗口内满足表达式条件数据的总和，数据列必须是decimal类型。  Null会被忽略，不会参与计算。 | Sum(decimalprice, type = "1") |
| avg(int number) | long | 计算窗口内数据的平均值，数据列必须是int类型。  Null会被忽略，不会参与计算。 | avg(intprice) |
| avg(long number) | long | 计算窗口内数据的平均值，数据列必须是long类型。  Null会被忽略，不会参与计算。 | avg(longprice) |
| avg(float number) | double | 计算窗口内数据的平均值，数据列必须是float类型。  Null会被忽略，不会参与计算。 | avg(floatprice) |
| avg(double number) | double | 计算窗口内数据的平均值，数据列必须是double类型。  Null会被忽略，不会参与计算。 | avg(doubleprice) |
| avg(decimal number) | decimal | 计算窗口内数据的平均值，数据列必须是decimal类型。  Null会被忽略，不会参与计算。 | avg(decimalprice) |
| avg(int number, expression) | long | 计算窗口内满足表达式条件数据的平均值，数据列必须是int类型。  Null会被忽略，不会参与计算。 | avg(intprice, type = "1") |
| avg(long number, expression) | long | 计算窗口内满足表达式条件数据的平均值，数据列必须是long类型。  Null会被忽略，不会参与计算。 | avg(longprice, type = "1") |
| avg(float number, expression) | double | 计算窗口内满足表达式条件数据的总和，数据列必须是float类型。  Null会被忽略，不会参与计算。 | avg(floatprice, type = "1") |
| avg(double number, expression) | double | 计算窗口内满足表达式条件数据的平均值，数据列必须是double类型。  Null值会被忽略，不会计入平均值中。 | avg(doubleprice, type = "1") |
| avg(decimal number, expression) | decimal | 计算窗口内满足表达式条件数据的平均值，数据列必须是decimal类型。  Null值会被忽略，不会计入平均值中。 | avg(decimalprice, type = "1") |
| max(int number) | Int | 计算窗口内数据的最大值，数据列必须是int类型。  Null会被忽略，不会参与计算。 | max(intprice) |
| max(long number) | Long | 计算窗口内数据的最大值，数据列必须是long类型。  Null会被忽略，不会参与计算。 | max(longprice) |
| max(float number) | Float | 计算窗口内数据的最大值，数据列必须是float类型。  Null会被忽略，不会参与计算。 | max(floatprice) |
| max(double number) | Double | 计算窗口内数据的最大值，数据列必须是double类型。  Null会被忽略，不会参与计算。 | max(doubleprice) |
| max(decimal number) | Decimal | 计算窗口内数据的最大值，数据列必须是decimal类型。  Null会被忽略，不会参与计算。 | max(decimalprice) |
| max(int number, expression) | Int | 计算窗口内满足表达式条件数据的最大值，数据列必须是int类型。  Null会被忽略，不会参与计算。 | max(intprice, type = "1") |
| max(long number, expression) | Long | 计算窗口内满足表达式条件数据的最大值，数据列必须是long类型。  Null会被忽略，不会参与计算。 | max(longprice, type = "1") |
| max(float number, expression) | Float | 计算窗口内满足表达式条件数据的最大值，数据列必须是float类型。  Null会被忽略，不会参与计算。 | max(floatprice, type = "1") |
| max(double number, expression) | Double | 计算窗口内满足表达式条件数据的最大值，数据列必须是double类型。  Null会被忽略，不会参与计算。 | max(doubleprice, type = "1") |
| max(decimal number, expression) | Decimal | 计算窗口内满足表达式条件数据的最大值，数据列必须是decimal类型。  Null会被忽略，不会参与计算。 | max(decimalprice, type = "1") |
| min(int number) | Int | 计算窗口内数据的最小值，数据列必须是int类型。  Null会被忽略，不会参与计算。 | min(intprice) |
| min(long number) | Long | 计算窗口内数据的最小值，数据列必须是long类型。  Null会被忽略，不会参与计算。 | min(longprice) |
| min(float number) | Float | 计算窗口内数据的最小值，数据列必须是float类型。  Null会被忽略，不会参与计算。 | min(floatprice) |
| min(double number) | Double | 计算窗口内数据的最小值，数据列必须是double类型。  Null会被忽略，不会参与计算。 | min(doubleprice) |
| min(decimal number) | Decimal | 计算窗口内数据的最小值，数据列必须是decimal类型。  Null会被忽略，不会参与计算。 | min(decimalprice) |
| min(int number, expression) | Int | 计算窗口内满足表达式条件数据的最小值，数据列必须是int类型。  Null会被忽略，不会参与计算。 | min(intprice, type = "1") |
| min(long number, expression) | Long | 计算窗口内满足表达式条件数据的最小值，数据列必须是long类型。  Null会被忽略，不会参与计算。 | min(longprice, type = "1") |
| min(float number, expression) | Float | 计算窗口内满足表达式条件数据的最小值，数据列必须是float类型。  Null会被忽略，不会参与计算。 | min(floatprice, type = "1") |
| min(double number, expression) | Double | 计算窗口内满足表达式条件数据的最小值，数据列必须是double类型。  Null会被忽略，不会参与计算。 | min(doubleprice, type = "1") |
| min(decimal number, expression) | Decimal | 计算窗口内满足表达式条件数据的最小值，数据列必须是decimal类型。  Null会被忽略，不会参与计算。 | min(decimalprice, type = "1") |

说明

Previous函数不支持Join，因为会导致结果不满足Join条件。

Previous函数不支持排序窗口，包含长度排序窗口和时间排序窗口，因为排序窗口中的数据会发生变化，导致窗口数据和缓存数据不一致。

# CQL其他配置参数

这里的配置参数，是指没有在前面序列化/反序列化、输入输出以及UDF函数中出现过的配置参数，主要分为系统配置参数，平台相关配置参数等。这类参数都没有简写，在使用的时候，只能按照全称进行设置。

CQL其它配置参数

| 参数名称 | 默认值 | 参数说明 |
| --- | --- | --- |
| streaming.killapplication.overtime | 60 | CQL kill application命令执行的超时时间，单位：秒。 |
| streaming.common.istestmodel | False | 是否是测试模式。  如果是测试模式，系统就会只构建拓扑，但是并不提交。  该参数主要在CQL测试情况下使用，正常场景无需配置此参数。 |
| streaming.common.parallel.number | 1 | 默认的并发度，如果在CQL语句中没有体现，则使用系统默认值。 |
| streaming.adaptor.application | com.huawei.streaming.storm.StormApplication | 底层应用程序平台。 |
| streaming.template.directory | ${system:java.io.tmpdir} | 系统临时目录路径，默认值使用的是系统的tmp路径。 |
| streaming.storm.nimbus.host | 127.0.0.1 | nimbus的地址，在Storm HA场景下无效。 |
| streaming.storm.nimbus.port | 6627 | nimbus的端口，在Storm HA场景下无效。 |
| streaming.storm.ha.zkaddress | 127.0.0.1 | storm HA的zookeeper集群地址，如果有多个地址，使用','分隔。 |
| streaming.storm.ha.zkport | 24000 | storm HA的zookeeper端口。 |
| streaming.storm.ha.zksessiontimeout | 30000 | HA连接zookeeper的session 超时时间。 |
| streaming.storm.ha.zookeepersiontimeout | 30000 | HA连接zookeeper 建立连接的超时时间。 |
| streaming.storm.thrift.transport.plugin | backtype.storm.security.auth.SimpleTransportPlugin | 客户端和服务端连接的thrift协议，如果是安全连接，使用：  backtype.storm.security.auth.kerberos.KerberosSaslTransportPlugin。 |
| streaming.security.authentication | NONE | 是否启用安全。  可以使用如下几个值：NONE, KERBEROS。  不区分大小写。  默认值是：NONE 不启用安全。 |
| streaming.security.zookeeper.principal | zookeeper/hadoop | Zookeeper server端的principal。 |
| streaming.security.storm.principal | streaming/hadoop@HADOOP.COM | Storm server端的principal。 |
| streaming.security.user.principal | NONE | 用户principal。 |
| streaming.security.keytab.path | NONE | 用户keytab文件地址。 |
| streaming.security.krbconf.path | NONE | krb5文件地址。  如果为空，则从系统默认地址获取。  默认地址获取顺序为：  1、${JAVA\_HOME}/lib/security/krb5.conf  2、${JAVA\_HOME}/krb5.ini  3、windows: %windir%\krb5.ini  linux: /etc/krb5.conf |
| streaming.storm.submit.islocal | False | 任务提交方式，是否是本地提交，本地提交就意味着CQL任务在本地进程内执行，一般用来测试。 |
| streaming.localtask.alivetime.ms | 5000 | 本地任务存活时间，过了这个时间之后，就会被kill掉。单位毫秒 |
| streaming.storm.killtask.waitseconds | 3 | kill 任务的时候的等待时间，单位秒，Storm 自身kill任务的参数，不同于CQL参数。 |
| streaming.storm.worker.number | 1 | CQL提交任务的进程数量。 |
| streaming.serde.default | com.huawei.streaming.serde.SimpleSerDe | CQL默认的序列化/反序列化类。 |
| serde.simpleserde.separator | , | SimpleSerDe配置参数。  因为系统默认的序列化类为simpleSerDe，所以配置参数中必须包含该序列化类的参数。 |
| streaming.userfile.maxsize | 30 | 用户文件最大大小，单位MB。 |
| streaming.storm.rebalance.waitseconds | 10 | rebalance任务等待时间。 |

# CQL异常码

CQL异常码的设计

异常码分为三部分：

前缀：以CQL-开头。

分类：两位数字编号，从00开头。

细分：三位数字编号，从000开头。

错误码 异常描述

CQL-00 成功完成。

CQL-01 警告。

CQL-02 配置属性异常。

CQL-03 流处理平台异常。

CQL-04 语义分析异常，包含语法解析异常。

CQL-05 拓扑异常，包含Streaming算子解析异常。

CQL-06 函数解析异常。

CQL-07 窗口解析异常。

CQL-08 安全异常。

CQL-99 未知异常。

CQL异常码列表

| 异常码 | 异常描述 |
| --- | --- |
| CQL-00： 成功完成。 | |
| CQL-00000 | 成功完成 |
| CQL-01： 警告。 | |
| CQL-01000 | 警告 |
| CQL-02： 配置属性异常。 | |
| CQL-02000 | 找不到配置属性。 |
| CQL-02001 | 配置属性数据类型转换失败。 |
| CQL-02002 | 错误的配置属性值，通常发生在数据类型正确，但是值超出范围或者不符合要求。 |
| CQL-03： 流处理平台异常。 | |
| CQL-03000 | 删除应用程序超时。  该时间取决于streaming.killapplication.overtime参数配置。 |
| CQL-03001 | 应用程序已经存在。 |
| CQL-03002 | 应用程序不存在。 |
| CQL-03003 | 应用程序的算子拓扑异常。 |
| CQL-03004 | 找不到输入算子。 |
| CQL-03005 | 应用程序状态异常，会导致执行active，deactive，rebalance等命令失败。 |
| CQL-03006 | 应用程序worker数量超过系统可用数量。 |
| CQL-03007 | 无效的nimbus服务异常。 |
| CQL-03008 | 服务端通信异常。 |
| CQL-04： 语义分析异常，包含语法解析异常。 | |
| CQL-04000 | 语法解析异常。 |
| CQL-04001 | 不支持的语法。 |
| CQL-04002 | order by 子句中只支持属性表达式。 |
| CQL-04003 | 表达式在select子句中找不到。 |
| CQL-04004 | 地址文件不存在。 |
| CQL-04005 | 地址必须是一个文件。 |
| CQL-04006 | 文件不是jar类型。 |
| CQL-04007 | Jar文件地址无法解析。 |
| CQL-04008 | 找不到对应流。 |
| CQL-04009 | 常量类型转换失败。 |
| CQL-04010 | 不支持的数据类型。 |
| CQL-04011 | 无法从指定流中找到对应列。 |
| CQL-04012 | 无法从所有的流中找到对应列。 |
| CQL-04013 | 无法唯一确定列， 该列可能存在在多个流中。 |
| CQL-04014 | 无法根据名称找到对应流。 |
| CQL-04015 | 输出列数量和select子句的列数量不一致。 |
| CQL-04016 | like表达式的子表达式必须是String类型。 |
| CQL-04017 | not表达式中所有的表达式返回值必须是boolean类型。 |
| CQL-04018 | IN和BETWEEN表达式中的参数只支持常量。 |
| CQL-04019 | CASE-WHEN表达式必须至少包含一个WHEN-THEN表达式。 |
| CQL-04020 | WHEN-THEN表达式必须同时包含WHEN和THEN表达式。 |
| CQL-04021 | WHEN-THEN表达式的WHEN表达式返回值必须是boolean类型。 |
| CQL-04022 | CASE表达式必须包含输入的表达式。 |
| CQL-04023 | 算术表达式只支持number的数据类型。 |
| CQL-04024 | 逻辑表达式中的表达式必须是boolean类型。 |
| CQL-04025 | 无法找到对应的数据源。 |
| CQL-04026 | 从数据源参数中找不到对应表达式。 |
| CQL-04027 | Combine操作中不允许出现窗口。 |
| CQL-04028 | Combine操作的输入流和combine条件中的流数量不匹配。 |
| CQL-04029 | combine算子只支持属性表达式。 |
| CQL-04030 | combine语句中，一个流的输出列在select语句中必须放在一起。 |
| CQL-04031 | MultiInsert中不允许出现数据源。 |
| CQL-04032 | MultiInsert中不允许出现GroupBy。 |
| CQL-04033 | MultiInsert中不允许出现窗口。 |
| CQL-04034 | MultiInsert中不允许出现多个输入流。 |
| CQL-04035 | 不支持多流Join。 |
| CQL-04036 | NATURAL JOIN 不支持。 |
| CQL-04037 | Join中，只允许其中一个流设置unidirection属性。 |
| CQL-04038 | join语句中找不到On条件。 |
| CQL-04039 | join的条件中的表达式在在输入流中找不到对应列。 |
| CQL-04040 | 不支持的Join条件。 |
| CQL-04041 | 找不到序列化和反序列化类。 |
| CQL-04042 | 找不到类。 |
| CQL-04043 | 序列化或者反序列化类 初始化失败。 |
| CQL-04044 | 流名称已经存在。 |
| CQL-04045 | 表达式比较异常，一般是两个表达式类型不匹配，无法比较。 |
| CQL-04046 | 用户文件总大小超过最大限制。  系统对于用户可以上传的总文件大小做了限制，取决于streaming.userfile.maxsize。  参数，默认30MB。 |
| CQL-04047 | 语法解析结果为空。 |
| CQL-04048 | 该窗口不支持Exclude Now功能。。 |
| CQL-04049 | Previous表达式不能和排序窗口一起用 |
| CQL-04050 | 该窗口不支持输出RStream(旧数据)输出  该场景一般发生在左流和右流一个窗口要输出新数据，一个窗口要输出旧数据。 |
| CQL-04051 | 从左流或者右流中找不到unidirection定义。  这个一般发生在selfjoin中，selfjoin要求必须单向Join。 |
| CQL-04052 | 找不到自定义算子。 |
| CQL-04053 | 自定义算子只允许一个输入schema。 |
| CQL-04054 | 自定义算子只允许一个输出schema。 |
| CQL-04055 | 自定义算子输入schema校验失败。 |
| CQL-04056 | 自定义算子输出schema校验失败。 |
| CQL-05： 拓扑异常，包含Streaming算子解析异常。 | |
| CQL-05000 | 一个CQL文件只允许提交一个应用程序。 |
| CQL-05001 | 算子找不到输入流。 |
| CQL-05002 | 算子找不到输出流。 |
| CQL-05003 | 应用程序名称为空。 |
| CQL-05004 | 找不到物理执行计划文件。 |
| CQL-05005 | 物理执行计划文件内容格式错误。 |
| CQL-06： 函数解析异常。 | |
| CQL-06000 | 不支持的函数。 |
| CQL-06001 | 系统函数不能移除。 |
| CQL-06002 | 系统函数不能被覆盖。 |
| CQL-06003 | 用户自定义函数必须继承自指定类。 |
| CQL-06004 | previous函数只能是PREVIOUS(number, column)这样的形式。 |
| CQL-06005 | UDF函数中不支持的参数类型。 |
| CQL-07： 窗口解析异常。 | |
| CQL-07000 | 无法识别的窗口。 |
| CQL-07001 | exclude now中的select列表不能包含其他非group by列。 |
| CQL-07002 | 错误的窗口参数，该错误会导致窗口创建失败。 |
| CQL-07003 | 排序窗只允许滑动类型。 |
| CQL-08： 安全异常。 | |
| CQL-08000 | 不支持的安全类型。 |
| CQL-08001 | 没有授权。 |
| CQL-08002 | keytab文件地址错误。 |
| CQL-08003 | 安全内部错误。 |
| CQL-08004 | 客户端用户鉴权失败。 |
| CQL-99： 未知异常。 | |
| CQL-99000 | 服务端通用异常。 |
| CQL-99999 | 未知异常。 |

# CQL Java开发API说明

通过调用"com.huawei.streaming.cql.Driver"类进行CQL语句提交。

Driver方法说明：

CQL Java开发接口

| 方法名 | 方法说明 | 参数 | 返回值 |
| --- | --- | --- | --- |
| Run | 执行CQL语句 | String CQL语句 | Void |
| getResult | 获取执行结果 | 无 | CQLResult  查询结果根据语句有所不同，get命令返回查询参数的值。 |
| Clean | 清空driver中初始化的用户信息。  该方法调用一般发生在用户完成submit之后，或者重新提交 应用程序之前。 | 无 | Void |

# CQL关键字列表

CQL关键字列表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| CREATE | SHOW | EXPLAIN | SET | GET |
| LOAD | EXPORT | DROP | ADD | SELECT |
| COMMENT | FORCE | SERDE | WITH | PROPERTIES |
| SOURCE | INPUT | STREAM | OUTPUT | SINK |
| SUBMIT | APPLICATION | DISTINCT | AND | OR |
| BETWEEN | IN | LIKE | RLIKE | REGEXP |
| CASE | WHEN | THEN | ELSE | END |
| CAST | EXISTS | IF | FALSE | AS |
| NULL | IS | TRUE | ALL | NOT |
| ASC | DESC | SORT | ORDER | GROUP |
| BY | HAVING | WHERE | FROM | ON |
| JOIN | FULL | PRESERVE | OUTER | CROSS |
| SEMI | LEFT | INNER | NATURAL | RIGHT |
| INTO | INSERT | OVERWRITE | LIMIT | UNION |
| APPLICATIONS | WINDOWS | EXTENDED | FUNCTIONS | FILE |
| INPATH | WINDOW | JAR | FUNCTION | COMBINE |
| UNIDIRECTION | PARALLEL | TRIGGER | PARTITION | SLIDE |
| BATCH | RANGE | ROWS | TODAY | UNBOUNDED |
| NOW | PREVIOUS | DATASOURCE | SCHEMA | QUERY |
| DEACTIVE | ACTIVE | WORKER | REBALANCE | DAYS |
| HOUR | HOURS | MINUTE | MINUTES | SECOND |
| SECONDS | MILLISECOND | MILLISECONDS | BOOLEAN | INT |
| LONG | FLOAT | DOUBLE | STRING | DATE |
| TIME | DECIMAL |  |  |  |