

阿里云

实时计算（流计算）

Flink Datastream开发指南

文档版本：20190816

法律声明

阿里云提醒您阅读或使用本文档之前仔细阅读、充分理解本法律声明各条款的内容。如果您阅读或使用本文档，您的阅读或使用行为将被视为对本声明全部内容的认可。

1. 您应当通过阿里云网站或阿里云提供的其他授权通道下载、获取本文档，且仅能用于自身的合法合规的业务活动。本文档的内容视为阿里云的保密信息，您应当严格遵守保密义务；未经阿里云事先书面同意，您不得向任何第三方披露本手册内容或提供给任何第三方使用。
2. 未经阿里云事先书面许可，任何单位、公司或个人不得擅自摘抄、翻译、复制本文档内容的部分或全部，不得以任何方式或途径进行传播和宣传。
3. 由于产品版本升级、调整或其他原因，本文档内容有可能变更。阿里云保留在没有任何通知或者提示下对本文档的内容进行修改的权利，并在阿里云授权通道中不时发布更新后的用户文档。您应当实时关注用户文档的版本变更并通过阿里云授权渠道下载、获取最新版的用户文档。
4. 本文档仅作为用户使用阿里云产品及服务的参考性指引，阿里云以产品及服务的”现状“、“有缺陷”和“当前功能”的状态提供本文档。阿里云在现有技术的基础上尽最大努力提供相应的介绍及操作指引，但阿里云在此明确声明对本文档内容的准确性、完整性、适用性、可靠性等不作任何明示或暗示的保证。任何单位、公司或个人因为下载、使用或信赖本文档而发生任何差错或经济损失的，阿里云不承担任何法律责任。在任何情况下，阿里云均不对任何间接性、后果性、惩戒性、偶然性、特殊性或刑罚性的损害，包括用户使用或信赖本文档而遭受的利润损失，承担责任（即使阿里云已被告知该等损失的可能性）。
5. 阿里云网站上所有内容，包括但不限于著作、产品、图片、档案、资讯、资料、网站架构、网站画面的安排、网页设计，均由阿里云和/或其关联公司依法拥有其知识产权，包括但不限于商标权、专利权、著作权、商业秘密等。非经阿里云和/或其关联公司书面同意，任何人不得擅自使用、修改、复制、公开传播、改变、散布、发行或公开发表阿里云网站、产品程序或内容。此外，未经阿里云事先书面同意，任何人不得为了任何营销、广告、促销或其他目的使用、公布或复制阿里云的名称（包括但不限于单独为或以组合形式包含”阿里云”、Aliyun”、“万网”等阿里云和/或其关联公司品牌，上述品牌的附属标志及图案或任何类似公司名称、商号、商标、产品或服务名称、域名、图案标示、标志、标识或通过特定描述使第三方能够识别阿里云和/或其关联公司）。
6. 如若发现本文档存在任何错误，请与阿里云取得直接联系。

通用约定

格式	说明	样例
	该类警示信息将导致系统重大变更甚至故障，或者导致人身伤害等结果。	 禁止： 重置操作将丢失用户配置数据。
	该类警示信息可能导致系统重大变更甚至故障，或者导致人身伤害等结果。	 警告： 重启操作将导致业务中断，恢复业务所需时间约10分钟。
	用于补充说明、最佳实践、窍门等，不是用户必须了解的内容。	 说明： 您也可以通过按Ctrl + A选中全部文件。
>	多级菜单递进。	设置 > 网络 > 设置网络类型
粗体	表示按键、菜单、页面名称等UI元素。	单击 确定 。
<code>courier</code> 字体	命令。	执行 <code>cd /d C:/windows</code> 命令，进入Windows系统文件夹。
<code>##</code>	表示参数、变量。	<code>bae log list --instanceid</code> <code>Instance_ID</code>
<code>[]</code> 或者 <code>[a b]</code>	表示可选项，至多选择一个。	<code>ipconfig [-all -t]</code>
<code>{ }</code> 或者 <code>{a b}</code>	表示必选项，至多选择一个。	<code>swich {stand slave}</code>

目录

法律声明.....	I
通用约定.....	I
1 Flink Datastream开发指南概述.....	1
2 自定义参数.....	2
3 作业开发.....	3
3.1 提交.....	3
3.2 开发.....	4
4 监控报警.....	7

1 Flink Datastream开发指南概述

阿里实时计算开发平台为实时计算产品Flink Datastream作业提供了作业提交、上线和启动功能，并提供了便利的运维管理和监控报警等功能。



说明:

- 仅独享模式Blink3.x及以上版本，支持Flink Datastream功能。
- 若Datastream作业访问的上下游存储提供了白名单机制，请参见[如何配置数据存储白名单](#)。
- 目前实时计算产品支持的Datastream作业基于开源的blink版本，开发相关的内容请参见[开源blink版本](#)。

Flink Datastream暂不支持Flink SQL作业中的注册数据存储、作业开发、作业调试、配置调优等功能。Flink Datastream开发指南主要包含如下内容：

- 作业开发

介绍在阿里实时计算开发平台上提交、上线和启动Flink Datastream作业的流程。

- 数据运维

介绍Flink Datastream作业数据运维相关内容，其中数据曲线仅支持runtime相关metric的展示，而connector相关的指标若需要展示，需在Flink Datastream作业针对性进行metric埋点。

- 监控报警

介绍如何创建和启动报警规则，目前仅支持FailoverRate指标的监控报警。

2 自定义参数

在DataStream作业中，您可以根据实际需求从main方法中获取自定义参数。

自定义参数的配置和获取方法

在DataStream作业中定义paramName=paramValue，即可在main方法中获取自定义参数。

- paramName：自定义参数名
- paramValue：自定义参数值



说明：

1个DataStream作业中可定义多个自定义参数。

自定义参数的配置和获取示例

```
final ParameterTool params = ParameterTool.fromArgs(args);
/*此处必须写configFile，读取configFile中参数。*/
String configFilePatH = params.get("configFile");

/*创建一个Properties对象用于保存在平台中设置的相关参数值。*/
Properties properties = new Properties();

/* 将平台页面中设置的参数值加载到Properties对象中。*/
properties.load(new StringReader(new String(Files.readAllBytes(Paths.
get(configFilePatH)), StandardCharsets.UTF_8)));

/*获取参数*/
final StreamExecutionEnvironment env = StreamExecutionEnvironment.
getExecutionEnvironment();
env.setStreamTimeCharacteristic(TimeCharacteristic.EventTime);
env.getConfig().setGlobalJobParameters(params);
env.enableCheckpointing(20000);

/*
通过properties.get("paramName")获取自定义的参数值。
通过properties.get("blink.job.checkpoint.dir")的方式获取checkpoint的路径。
*/
env.setStateBackend(new RocksDBStateBackend((String) properties.get("
blink.job.checkpoint.dir")));
```

3 作业开发

3.1 提交

本文为您介绍如何提交Datastream作业。

前提条件




说明:

- 仅独享模式Blink3.2.2及以上版本支持Flink Datastream功能。
- Datastream API作业不支持资源配置调优和启动位点设置，因此，上线和启动作业过程中，使用默认配置即可。
- 登录[阿里实时计算开发平台](#)。
- 已创建实时计算项目。

操作步骤

1. 在[阿里实时计算开发平台](#)，单击顶部菜单栏中的开发。
2. 在开发页面的顶部菜单栏中，单击新建作业。
3. 在新建作业页面，配置作业参数。

作业参数	说明
文件名称	自定义作业的名称。作业名称需在当前项目中保持唯一。
作业类型	FLINK_STREAM/DATASTREAM  说明: Datastream API作业和Table API作业均选择FLINK_STREAM/DATASTREAM 作业类型。若无此类型，请提交工单。
存储位置	作业存储的位置

4. 单击左侧导航栏中的资源引用，进入资源引用窗口。
5. 单击新建资源上传已经完成开发的Datastream作业JAR包。
6. 单击引用。

7. 在作业开发界面配置参数。

```
blink.main.class=<完整主类名>
--函数完整类名，例如com.alibaba.realtimecompute.DemoTableAPI。
blink.job.name=<作业名>
--例如datastream_test。
blink.main.jar=<完整主类名jar包的资源名称>
--完整主类名JAR包的资源名称，例如blink_datastream.jar。
```



说明:

- `blink.main.class`和`blink.job.name`为必须参数。
- 上传多个JAR包时需要配置`blink.main.jar`参数。
- 其它参数您可先自行配置，然后在程序中引用。自定义参数配置及在代码中获取参数值的方法，请参见[#unique_8](#)。
- 请不要在参数配置中使用空格。

8. 上线作业。请参见[#unique_9](#)。

9. 启动作业。请参见[#unique_10](#)。

3.2 开发

本文为您介绍Datastream作业开发POM依赖包、Datastream作业开发示例和Datastream Connector。



说明:

- 仅独享模式Blink3.2.2及以上版本支持Datastream功能。
- 建议使用IntelliJ IDEA中的maven工程进行Datastream作业开发。
- 实时计算独享模式不支持归档保存已停止（含暂停）的作业的运行日志。若需要查询已停止（含暂停）的作业运行日志，请将日志输出至用户自定义的SLS或OSS中。日志输出步骤参见[#unique_12](#)。

POM依赖包

请根据需求自行添加开源版本所支持的POM依赖包。Blink3.2.2版本的POM文件，示例如下。

```
<properties>
  <scala.version>2.11.12</scala.version>
  <scala.binary.version>2.11</scala.binary.version>
  <blink.version>blink-3.2.2</blink.version>
  <java.version>1.8</java.version>
</properties>
<dependencies>
  <dependency>
    <groupId>com.alibaba.blink</groupId>
    <artifactId>flink-core</artifactId>
```



```

        <version>${blink.version}</version>
        <scope>provided</scope>
    </dependency>
    <dependency>
        <groupId>com.alibaba.blink</groupId>
        <artifactId>flink-java</artifactId>
        <version>${blink.version}</version>
        <scope>provided</scope>
    </dependency>
    <dependency>
        <groupId>com.alibaba.blink</groupId>
        <artifactId>flink-streaming-java_${scala.binary.version}</
artifactId>
        <version>${blink.version}</version>
        <scope>provided</scope>
    </dependency>
    <dependency>
        <groupId>com.alibaba.blink</groupId>
        <artifactId>flink-table_2.11</artifactId>
        <version>${blink.version}</version>
        <scope>provided</scope>
    </dependency>
    <dependency>
        <groupId>junit</groupId>
        <artifactId>junit</artifactId>
        <version>4.8.1</version>
        <scope>test</scope>
    </dependency>
    <dependency>
        <groupId>org.scala-lang</groupId>
        <artifactId>scala-library</artifactId>
        <version>2.11.12</version>
    </dependency>
</dependencies>

```

下载查看[完整依赖包示例](#)。



说明:

如果您需要依赖Snapshot版本，可以自行添加Snapshot版本所支持的[POM依赖包](#)。

作业示例

```

public class Datastreamtest {
    public static void main(String[] args) throws Exception {

        // 创建Execution Environment。
        StreamExecutionEnvironment env = StreamExecutionEnvironment.
getExecutionEnvironment();

        // 通过从本地TXT文件读取数据。
        //DataStream<String> text = env.readTextFile("./test.txt");
        DataStream<String> text = env.fromElements("hello world", "
hello jason", "wonderful world");
        // 解析数据，先按word进行分组，然后进行开窗和聚合操作。
        DataStream<Tuple2<String, Integer>> windowCounts = text
            .flatMap(new FlatMapFunction<String, Tuple2<String,
Integer>>() {
                @Override
                public void flatMap(String value, Collector<Tuple2
<String, Integer>> out) {

```

```
                for (String word : value.split("\\s")) {
                    out.collect(Tuple2.of(word, 1));
                }
            }
        })
        .keyBy(0)
        .timeWindow(Time.seconds(5))
        .sum(1);

// 将结果打印到控制台。请使用单线程打印，而非多线程打印。
windowCounts.print().setParallelism(1);

env.execute("Socket Window WordCount");
}
}
```

Connector列表

Blink 3.2版本新增如下Datastream Connector:

- Kafka
- Kafka（开源版本）
- Hbase（开源版本）
- JDBC
- RDS SINK
- Elasticsearch
- MongoDB
- Redis



说明:

Datastream支持的部分Connector已完成开源，开源信息请参见[alibaba-flink-connectors](#)。

4 监控报警

本文为您介绍实时计算监控报警的操作流程以及如何创建报警规则。

什么是云监控报警服务

云监控服务能够收集阿里云资源或您自定义的监控指标、探测服务可用性以及针对指标设置警报，让您全面了解阿里云上的资源使用情况、业务的运行状况和健康度，并及时接收异常报警，保证应用程序顺畅运行。

查看监控报警信息

1. 登录[实时计算控制台](#)。
2. 单击页面顶部的运维。
3. 在实时计算运维界面，单击目标作业名。
4. 在目标作业运维信息页面的右上角，单击监控。
5. 在监控图标页面，查看作业的监控指标。

创建报警规则

创建报警规则详情，参见[#unique_14/unique_14_Connect_42_section_nhz_ccf_zdb](#)。



说明:

- Failover Rate表示最近1分钟平均每秒Failover的次数。例如，最近1分钟Failover了1次，则Failover Rate=1/60=0.01667。
- DataStream作业开发过程中，若引用了开源Flink提供的Connector，则在云监控中不显示业务延迟、读入RPS和写入RPS这三项监控指标。