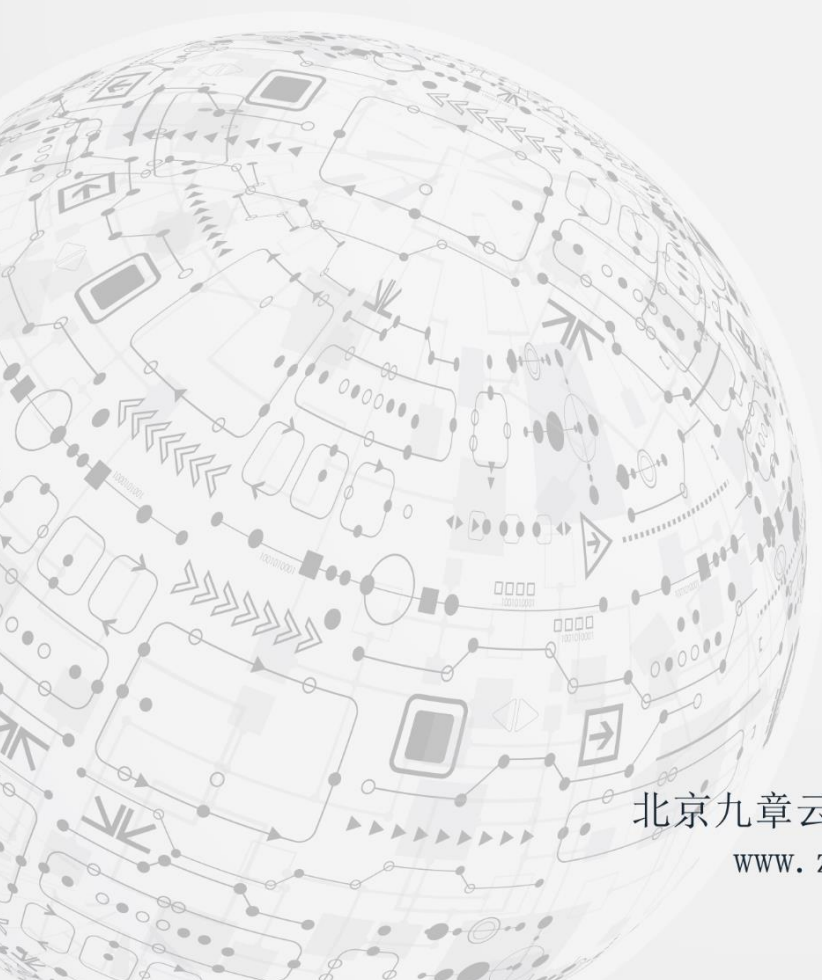


# DataCanvas

RT 实时流数据平台  
产品功能说明书



北京九章云极科技有限公司  
[www.zetyun.com](http://www.zetyun.com)

# 1. 产品核心功能及特性

## 1) 流数据接入

- 支持对接日志型、二进制数据流或文件，兼容 SOAP、SOP、主流 ESB 中间件数据格式，以及自定义的 XML，JSON，LOG 型数据格式。

- 支持避免消息脏读，多读或者读重的方法（也包括因节点故障恢复或高可用切换时引起的脏读）

- 支持数据脱敏配置

- 支持对 XML(SOAP)、json、日志（分隔符）、二进制报文、8583 等常见报文的配置化拆解包方法，提供标准化的字段映射方法，并提供实现拆解包的对应算法和组件

- 支持报文解析中的 LIST、结构体、多层 LIST 等复杂对象的处理机制

- 支持交易码和报文模版的配对、解析、要素提取机制，支持模版配置更新

- 支持解析规则在线修改

## 2) 分流处理

- 支持基于交易码或日志类型的筛选机制和处理逻辑（即只有涉及对应交易码的报文才需要做后续处理）

- 提供待处理报文或数据分主题或分区规则，键值对的设计原则

- 多个主题之间所需指标或要素存在交叉，提供多个应用主题的划分策略（多对多关系）

- 支持配置化管理来源报文、指标、规则之间的关系

- 支持指标计算完成后的数据持久化机制，以及向下一个节点的传递机制（一对多或多对多关系）

- 支持事件分支和聚合，支持异常处理队列设计，支持多线程并发

## 3) 数据处理

- 支持基于图形界面的拖拉拽的方式进行处理逻辑编排
- 支持基础计算逻辑，函数类计算逻辑
- 支持指标计算, 包括无状态指标(无需与存量数据或其他交易数据进行混合计算), 有状态、无时序指标 (需要计算, 但无顺序要求, 如按秒求和), 有状态、有时序指标 (需要计算, 有顺序要求), 支持数据保序处理
- 支持基于 window 的聚合操作 (count/sum/min/max/均值/方差/标准差/连续递增)
- 支持机器学习模型的集成和动态发布、动态卸载 (如 PMML), 及时生效
- 支持滑动窗口机制下的数据结构设计、滑动机制、存储机制、滑动窗口大小以及滑动位移大小
- 提供缓存的设计原则和缓存的淘汰机制, 支持系统重启后的数据恢复机制
- 支持实现热参数更新, 动态刷新业务计算逻辑, 并立即生效

#### 4) 数据交互

- 支持丰富的数据输出方式 (Kafka, JDBC, File, Socket, IBM MQ, RestAPI)
- 支持基于 SQL 的 Cache 查询操作
- 支持数据索引查询, 实现灵活的数据组合查询, 响应时间毫秒级
- 支持第三方规则引擎的集成和动态发布、动态卸载 (Drools、SpringEL), 及时生效

#### 5) 调度管理及运维管理

- 通用的作业流 Dag 调度, 适配不同调度引擎
- 支持平台的基础数据和任务调度功能
- 支持 KAFKA、流处理等任务的正常监控及异常处理方案
- 支持集群的自动化安装部署
- 支持监控 Kafka, Flink、Ignite 等组件的进程健康状况, 提供与组件, 进程, 性能等密切相关的监控指标

- 支持集群的资源监控、预警机制
- 支持用户管理，实现基于角色的权限管理，提供分级别、分角色、分权限管理的安全可靠的平台服务

## 2. 产品优势

通过构建流式大数据实时处理平台，可以为客户创造如下价值：

- 1) 成熟的企业级产品封装，采用企业友好化封装，提供统一的交互服务接口和界面，屏蔽底层技术复杂性，最大程度上降低应用门槛；
- 2) 灵活、多样的数据采集模式选择，统筹考虑数据采集效率和对数据源系统的性能影响，在保证源系统业务稳定的前提下，最大程度上提升数据采集效率。
- 3) 开放性数据访问接口，面向前端业务需求提供灵活的数据服务；
- 4) 产品采用开放性架构设计，采用最新的分布式数据处理技术，对接开源技术生态，在保证技术先进性的前提下，实现与企业现有核心业务系统、数据仓库、大数据平台的对接；
- 5) 高可用产品设计，通过分布式多副本一致性协议以及自平衡的智能分区算法，提升流处理集群的可用性以及可扩展性。
- 6) 流式计算下的复杂逻辑处理，支持复杂指标的计算、规则处理平台，可以将处理逻辑扩展到更复杂的处理逻辑，例如基于机器学习算法、深度学习模型的决策逻辑。

## 3. 适用业务场景

1) 在营销角度，可以深度了解和分析客户在持有、购买、放弃产品及应用的前、中、后期行为特征，从而为客户偏好、360度画像、数据挖掘提供更精准的数据依据，最终实现精准营销。

2) 在风控角度，可以深度了解和分析客户的活动范围、生活轨迹、交易支付习惯等，结合机器学习模型实现实时风险计算，从而为信贷风控提供有力证据。

3) 在客户体验角度，一方面可以深度了解和分析银行现有应用的运行状况、应用中各个功能模块的使用频率、客户访问路径长度等情况，从而为银行应用设计和优化提供数据支持，让银行应用产品更符合客户群体期望和个体期待；另一方面为客服系统、CRM、移动

柜员等前端应用提供更加精准的客户信息，方便一线业务人员为客户带来更加个性化、定制化的服务体验，提升银行客户粘性。

4) 在运维角度，可以利用实时流计算平台建立智能运维系统，实现海量日志实时采集、解析和检索查询、实时指标计算、实时告警和日志分析挖掘等应用，提高运维告警效率、实现告警时间追踪，建立智能化运维体系，降低运维人员工作量，提高运维事件的处理效率，更好的为业务系统提供支持。

5) 在企业决策角度，利用实时数据接入和指标计算，针对关键业务 KPI 指标进行实时统计分析，结合数据可视化技术，面向企业决策层实现实时报表。例如，实时大屏展示系统，动态展示本行持卡数量、现有存款总额、实时交易笔数、信贷业务在线申请人数、逾期信贷交易笔数等。