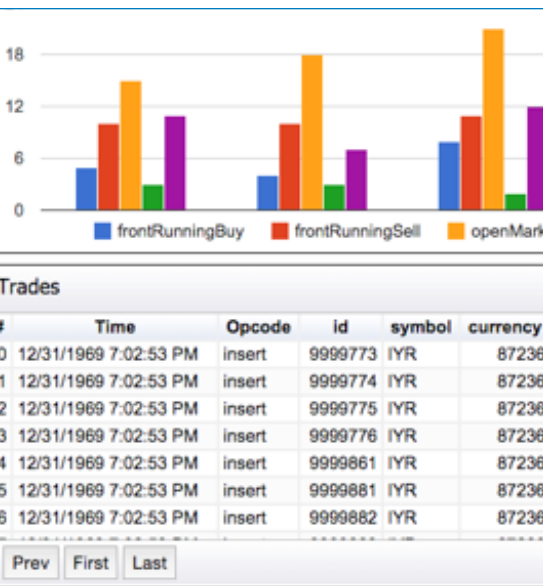


SAS® Event Stream Processing

Работайте с данными «на лету» – управляйте своим бизнесом в режиме реального времени



Чтобы оставаться конкурентоспособными в современном мире, где информация – один из главных активов, вы должны работать и принимать решения с молниеносной скоростью. Вам нужно оценивать постоянно меняющиеся условия, отслеживать представляющие интерес события по мере их возникновения и действовать сообразно ситуации. Когда вы обрабатываете потоковые данные, они продолжают поступать. Технология непрерывных запросов позволяет анализировать текущие данные на предмет обнаружения интересных закономерностей и реагировать на них в режиме реального времени.

Решение SAS Event Stream Processing помогает интерпретировать события в динамике, обрабатывая потоки событий. Решение не придерживается традиционного подхода «Export, Transform, Load», вместо этого непрерывные запросы сохраняются, и данные «пропускаются» через них. Данные анализируются по мере поступления, обновляя сведения о возможности наступления новых событий. Такой подход практически исключает информационное запаздывание, что позволяет моментально извлекать ценную информацию из поступающих данных. Решение гарантирует оптимальную производительность и высокую пропускную способность благодаря гибкой модели обработки и сохранению времени отклика менее миллисекунды.

Для чего предназначено решение SAS® Event Stream Processing?

SAS Event Stream Processing обрабатывает и анализирует миллионы событий в секунду, что позволяет обнаруживать интересующие закономерности в режиме реального времени. Результат обработки показывает, какие действия предпринять, какие предупреждения выдавать, какие данные сохранять и какие события игнорировать.

В чем заключается ценность решения SAS® Event Stream Processing?

Извлекайте ценную информацию из потоковых данных (операционные транзакции, показания датчиков, данные от устройств, действия на сайте и т. д.) «на лету». С временем отклика менее миллисекунды SAS® Event Stream Processing анализирует потоки информации, используя набор готовых операций, встроенных функций и алгоритмов углубленной аналитики. Вы получаете возможность точно интерпретировать события и применять своевременные меры.

Для кого разработано решение SAS® Event Stream Processing?

Этот конфигурируемый встраиваемый модуль предназначен для разработчиков приложений, ИТ-инженеров и системных архитекторов. Веб-интерфейс упрощает манипулирование данными и создание проектов, а информационная панель визуализирует данные, помогая тестировать и проверять результаты.

Преимущества

- Непрерывно анализируйте информацию и действуйте незамедлительно.** Пропускайте через решение потоки операций, транзакций, данные от существующих систем, а также датчиков и устройств Интернета вещей. Такой подход значительно повышает ценность этой информации, помогает анализировать ситуацию и действовать в реальном времени. Независимо от типа или формата, технологии SAS повышают ценность потоковых данных для приложений, которые используются для формирования потребительского опыта, оценки активов и производительности ИТ, выявления случаев мошенничества, анализа рисков и безопасности и т. д.
- Принимайте взвешенные решения на основе эффективной аналитики.** При наличии точных данных вы будете выполнять целесообразные действия. В решении SAS Event Stream Processing имеются встроенные функции качества данных, а также функции текстовой аналитики и множество современных алгоритмов анализа данных для выявления даже самых сложных составных событий в потоках данных.
- Получите полный контроль и быстро адаптируйтесь к изменениям.** Скорость и объем поступающих данных не имеют значения, как и количество подключенных источников: все процессы будут упрощены и оптимизированы, а вы получите полный контроль посредством единого интуитивно понятного интерфейса. Консолидированные инструкции для отслеживания событий в потоках данных, определенные на языке программирования SAS или на других языках, помогают быстро и управляемо обнаруживать составные события, упрощая адаптацию к изменению условий.
- Храните правильные данные.** Быстрая и эффективная потоковая обработка обеспечивает коррекцию до сохранения больших объемов информации, что сокращает затраты на обработку данных. Непрерывные запросы выполняются быстрее благодаря гибкой многопоточной модели обработки данных, которая поддерживает динамическое обновление, удаление и вставку «на лету».

Обзор

Решение SAS Event Stream Processing быстро принимает и обрабатывает большие объемы потоковых данных (миллионы событий в секунду), позволяя анализировать и интерпретировать события «на лету». Скорость потока и объем поступающих данных не имеют значения. Пропускная способность решения ограничивается только возможностями применяемого оборудования.

Входящие данные считываются с помощью адаптеров и коннекторов, которые являются частью архитектуры «публикация

и подписка». Данные о событиях публикуются в «окно источника» модели обработки потока событий. Визуальный интерфейс упрощает создание окон и преобразований. В свою очередь, можно достаточно просто определить непрерывные запросы, через которые будут «пропускаться» данные. Поточковые данные проверяются на наличие составных событий и отфильтровываются с целью сохранения аномалий, требующих более глубокого изучения. Кроме того, если актуальность не подтверждена, данные могут быть отброшены. Приложения оформляют подписку на результаты анализа потоковых данных и рекомендуемые действия.

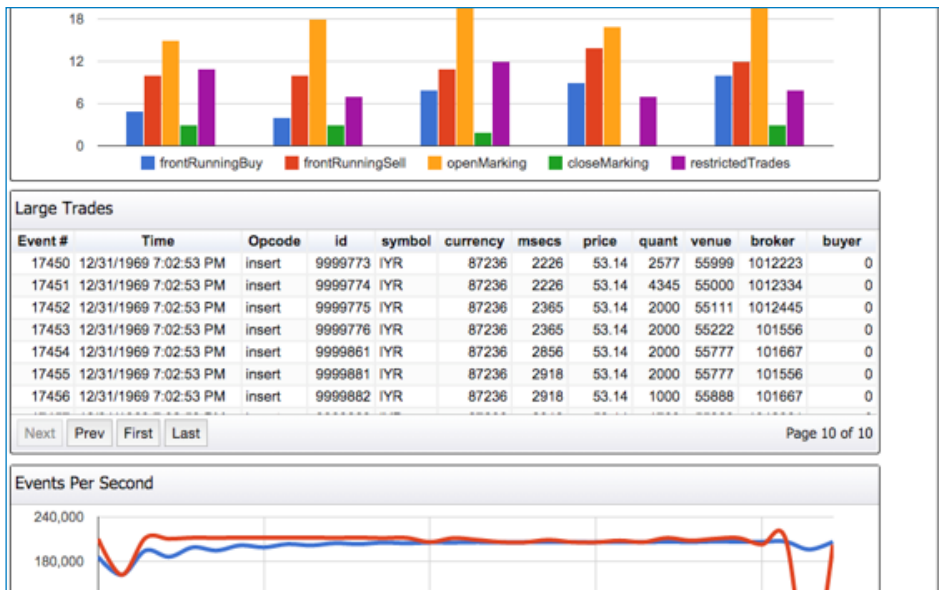


Рисунок 1. Тестирование и уточнение моделей потоков событий с помощью простого в использовании интерфейса HTML5.

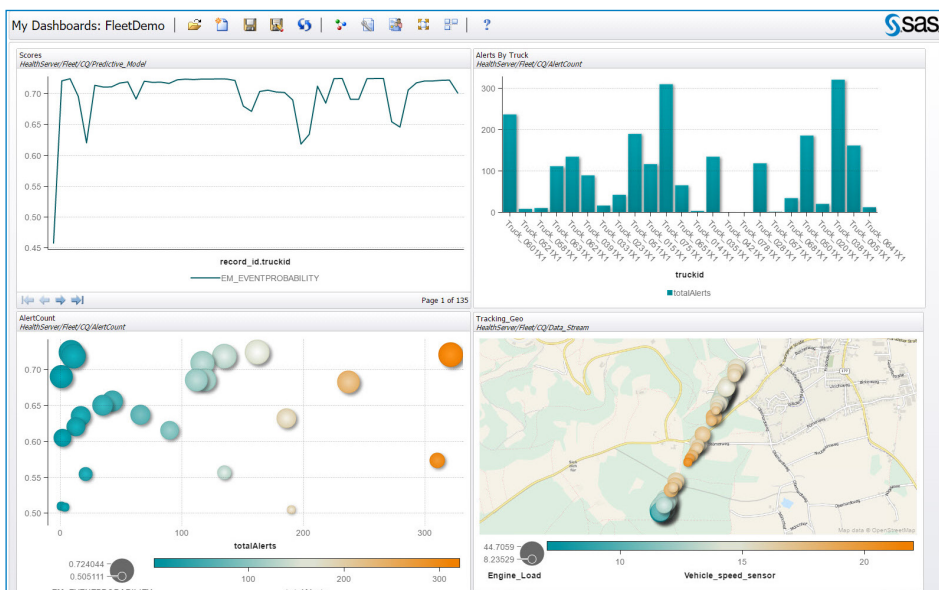


Рисунок 2. Информационная панель обеспечивает централизованный мониторинг динамических событий в реальном времени.

Потребление и подключение потоковых данных

Готов ли ваш центр обработки данных к быстрому росту нагрузки, вызванной повсеместным распространением Интернета вещей? Большинство не могут ответить на этот вопрос утвердительно. Даже потоки данных, связанные с текущими веб-транзакциями, датчиками и операционными системами, не так просто консолидировать и эффективно использовать. Решение SAS Event Stream Processing обрабатывает широкий спектр форматов потоковых данных и беспрепятственно интегрирует результаты в существующие системы и приложения. Большое количество готовых адаптеров и коннекторов позволяет вам создавать и управлять потоками данных в режиме реального времени с целью обнаружения, фильтрации, агрегации, сопоставления и разграничения представляющих интерес шаблонов составных событий. Действия применяются как к структурированным, так и к неструктурированным текстовым данным, а результаты отправляются в другие системы, что повышает эффективность и снижает затраты. Ваша организация может подключаться к источникам данных, которые ей нужны сегодня, и при этом быть уверенной в том, что те же технологии будут работать завтра.

Гибкая визуальная среда разработки

Решение SAS Event Stream Processing предоставляет наглядный интерактивный интерфейс для создания непрерывных модульных запросов, которые используют алгоритмы углубленной аналитики и правила SAS, чтобы точно определять значимость событий. Системные архитекторы и разработчики приложений могут использовать палитру окон и коннекторов, чтобы создавать сложные модели потоковой аналитики несколькими щелчками мыши. В интерактивном тестовом режиме пользователи могут проверять логику и результаты перед развертыванием. События можно использовать одновременно в нескольких проектах, а историю – сохранять и сравнивать с текущими событиями, и все это в визуальной среде с drag & drop. Таким образом, потоковые данные можно быстро определять, обновлять и изменять в процессе обработки новых быстро возникающих событий.

Расширенные алгоритмы сопоставления шаблонов

В решении SAS Event Stream Processing содержатся встроенные функции парсинга, фильтрации, объединения, вычислений для полей и сопоставления шаблонов. Также предусмотрены функции для стандартных задач по управлению данными – преобразования, нормализации, сопоставления, идентификации и т. д. Функции обработки неструктурированного текста обеспечивают извлечение концептов, сущностей и фактов, а также классификацию текстов и оценку окраса. Поточковая кластеризация позволяет формировать однородные группы событий «на лету», алгоритмы прогнозирования оценивают будущие события, создавая расширенные шаблоны для поиска составных событий, которые представляют интерес. Вы можете анализировать множество событий в одном запросе, включая последовательные и хронологические (временные) события.

Отображение динамических потоков событий на информационной панели

В решении SAS Event Stream Processing имеется единая информационная панель, позволяющая в реальном времени просматривать динамические потоки событий и контролировать их. Вы можете создавать графические визуализации SAS, просматривать различные модели с разных серверов SAS Event Stream Processing, совместно использовать информационные панели и создавать встраиваемые модели.

Ключевые функции

Окна моделей машинного обучения непосредственно в потоке

- Окна моделей машинного обучения позволяют непосредственно в потоке использовать разные типы окон для указания источников потоков данных, представляющих интерес шаблонов и результирующих действий. Окна моделей машинного обучения включают следующее:
 - Обучение. Создайте расширенную аналитическую модель в потоке и передайте полученные обновления для модели в окно оценки.
 - Оценка. Применяйте обученную модель к текущим событиям в потоке, чтобы получить результирующие оценки.
 - Расчеты. Используйте методы нормализации и преобразования данных, а также модели, которые учитывают результаты обучения и оценки.
 - Супервайзер модели. Вы определяете, какую модель развертывать, когда и где (например, в окне оценки).
 - Чтение модели. Интегрируйте автономные модели ASTORE и публикуйте модели в другом окне потоковой аналитики, например, в окне оценки.
- Единые инструменты управления проектами и серверами в решении SAS Event Stream Manager предоставляют следующие возможности:
 - Простой в использовании интерфейс для создания тиражируемых планов развертывания для проектов, выполняемых на серверах SAS Event Stream Processing, и управления этими планами.
 - Средства мониторинга использования памяти и дискового пространства в масштабах всей серверной среды SAS Event Stream Processing.
 - Простые инструменты добавления новых серверов SAS Event Stream Processing для более эффективного мониторинга.
 - Технология на основе агентов для автоматического обнаружения новых серверов SAS Event Stream Processing в целях автоматизации мониторинга.

Потребление и подключение потоковых данных

- Большой набор адаптеров и коннекторов для публикации и подписки на потоки данных, как структурированных, так и неструктурированных, в режиме реального времени.
- Встроенные адаптеры включают средства чтения и записи (т. е. средства публикации и подписки) для следующих наборов данных: Hadoop (HDFS Yarn), Apache Camel, OSISoft PI, Axeda, RabbitMQ, Solace, Tervela Data Fabric, XML/JSON File Socket Adapter, SAS® LASR™ Analytic Server, IBM DB2, IBM Netezza, IBM WebSphere MQ, SAP Sybase ASE, Tibco Rendezvous, JMS, File/Socket, Database ODBC и SAS.
- Собственные интерфейсы API для публикаций и подписок можно написать на языке C, Java или Python.
- Поддерживается публикация: Twitter, снифферы журналов (Oracle, Greenplum), сетевые снифферы, сниффер Google Protobuf Sniffer, SYSLOG, интерфейсы HTTP RESTful. Поддерживается подписка: SOAP и SMTP.

Адаптируемая аналитика и манипулирование данными непосредственно в потоке

- Благодаря поддержке потокового алгоритма машинного обучения можно создавать процедурные окна оценки и обучения для различных обучаемых алгоритмов, включая кластеризацию на основе плотности (DBSCAN) и кластеризацию методом K-средних.
- DBSCAN позволяет идентифицировать кластеры без необходимости контроля. Для регулярного обновления модели DBSCAN используются комбинации окон обучения и оценки.
- Геозоны – окна нового типа – позволяют отслеживать местоположение объектов относительно границ геозоны. Вы также можете настроить оповещения о том, что объект приближается, входит в границы определенной геозоны или покидает их.
- Технология обучения модели по историческим данным (для разработки более точных моделей) позволяет использовать высокопроизводительную аналитику для хранимых данных.
- Возможность взаимодействия со средствами управления моделями SAS обеспечивает быструю интеграцию аналитических моделей с проектами SAS Event Stream Processing.

Оптимизированные масштабируемые средства обработки

Если используется обработка потоков событий, системы должны быть постоянно на связи. Обработка должна осуществляться максимально быстро, поэтому применяемые решения должны быть высокопроизводительными. SAS предлагает запатентованную быстродействующую схему обеспечения отказоустойчивости 1+N, гарантированную доставку без задержек и динамические обновления для обеспечения согласованной и эффективной обработки. ПО SAS Event Stream Processing обрабатывает большие объемы данных с минимальной задержкой и превосходит по своим возможностям другие обработчики потоков при установке на оборудовании потребительского уровня. При этом распределенная параллельная обработка информации в оперативной памяти обеспечивает линейное масштабирование по мере увеличения объема данных, что позволяет получать максимальную отдачу от инвестиций в оборудование.

Ключевые функции (продолжение)

- Гибкая модульная архитектура на основе окон для создания сложных непрерывных запросов:
- Большой набор взаимозаменяемых типов окон и операторов для обнаружения неограниченного числа шаблонов событий, корреляций, вычислений и агрегаций.
- Встроенные стандартные процедуры обеспечения качества данных для очистки, стандартизации и фильтрации информации в режиме реального времени до ее сохранения, что упрощает последующую обработку.
- Представляющие интерес шаблоны событий могут включать в себя практически неограниченные вычисления углубленной аналитики с внутривидеовой кластеризацией, машинным обучением и скорингом аналитических моделей в реальном времени. Аналитические модели поддерживают код SAS DATA STEP, DS2 и сторонних поставщиков.
- Беспрепятственное встраивание в шлюзы, устройства и любые существующие приложения C++ (с выделенной обработкой пула потоков).
- Поточковые алгоритмы для DBSCAN, регрессии, дерева решений и байесовского усреднения.

Поддержка расширенного развертывания и открытых данных

- API Python для публикации/подписки:
 - Публикуйте события и подписывайтесь на потоки SAS Event Stream Processing с помощью Python.
- В SAS Event Stream Processing вы можете использовать следующие языки для создания обработчиков потоков событий в процессах:
 - DS2 (текущий), Python и C/C++.
- BOSH Cloud Foundry: для масштабных распределенных служб.
- Поддержка потоков данных:
 - Интеграция с Hortonworks DataFlow (HDF) NiFi: SAS и HDF обеспечивают быструю и углубленную потоковую аналитику.
 - Поддержка потоков MapR.
 - Коннекторы и адаптеры для Интернета вещей (Twitter, Kafka, Flume, MQTT, Pico/SpyWare, шлюзы, устройства, Cassandra (только адаптер), Boardreader.
- SAS Event Stream Processing для SAS® Viya™ и SAS Cloud Analytic Services (CAS):
 - Развертывайте модели управления данными и модели машинного обучения SAS Viya в SAS Event Stream Processing для организации потокового анализа.
 - Обеспечивайте потоковую передачу, фильтруйте, ищите шаблоны и оценивайте данные, которые проходят через CAS.

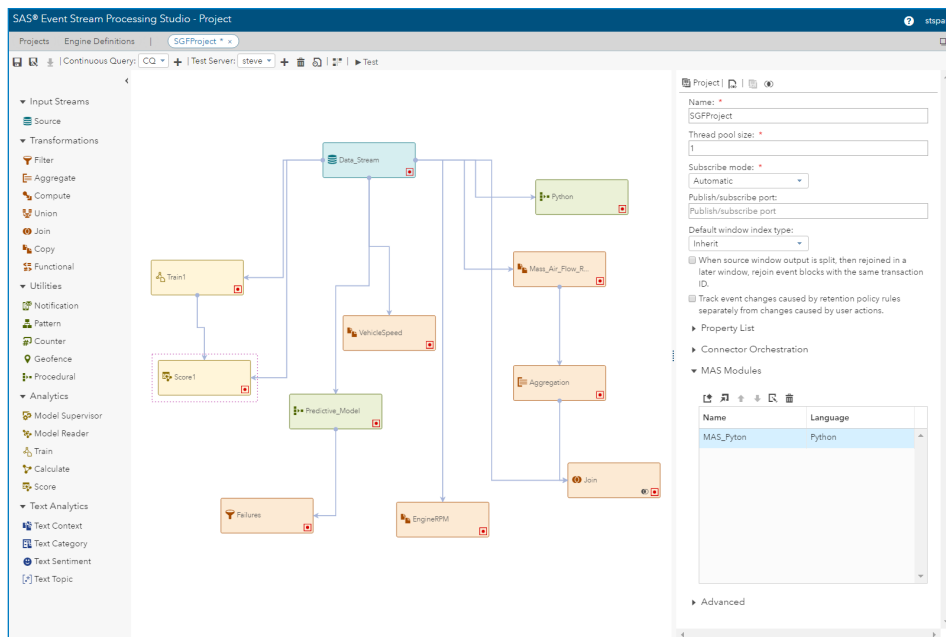


Рисунок 3. Создание, тестирование и публикация потоковых моделей с помощью решения SAS Event Stream Processing.

Корпоративное администрирование и управление

Централизованное администрирование и управление процессами обработки потоков событий – ключевые задачи, решив которые, вы сможете эффективнее управлять своей ИТ-средой и проектами, оптимизировать и сопровождать процессы обработки данных. Решение обеспечивает прозрачное управление в соответствии с бизнес-требованиями и позволяет интегрировать обработчики потоков и алгоритмы из систем SAS и других источников. Дополнительное хранилище для кэширования позволяет справляться с ситуациями, когда текущий объем данных превышает доступную память.

Расширьте возможности других решений SAS® с помощью потоковой аналитики

Решение SAS Event Stream Processing дополняет возможности других систем SAS, таких, как SAS Customer Intelligence, SAS Asset Performance Analytics, SAS Security Intelligence и т. д. Подключение этих решений к обработчику потоков событий означает, что аномалии будут помечены, а новые представляющие интерес шаблоны – обнаружены, и все это «на лету», без остановки потока данных. В сочетании с другими технологиями SAS решение SAS Event Stream Processing позволяет быстрее проводить углубленный анализ.

Ключевые функции (продолжение)

Визуализация и мониторинг потоков событий

- Создайте информационную панель для настраиваемого тестирования потоков.
- Выполняйте интерактивную фильтрацию и запрашивайте сведения о текущей активности потоков реального времени для анализа специфического поведения элементов.
- Создавайте и настраивайте рассылку уведомлений по SMS, электронной почте и через другие каналы в рамках рабочего процесса модели потока событий.
- Сравнивайте историческую активность с параметрами текущей обработки, используя графические представления.
- Отслеживайте параметры потока, подписываясь на представляющие интерес события.

Распределенная оптимизированная обработка в оперативной памяти с поддержкой масштабирования

- Обработка миллионов событий в секунду. Малые задержки, быстрый отклик (часто меньше миллисекунды).
- Данные размещаются и агрегируются в оперативной памяти.
- Поддержка распределенных архитектур.
- Скорость обработки можно регулировать с помощью гибких пулов потоков переменного размера, хранилищ для кэширования и т. д.
- Включает запатентованную быстросействующую схему обеспечения отказоустойчивости 1+N, гарантированную доставку без задержек и другие функции защиты от сбоев для эффективной обработки потоков событий.
- Неограниченный открытый доступ ко всем метаданным событий.

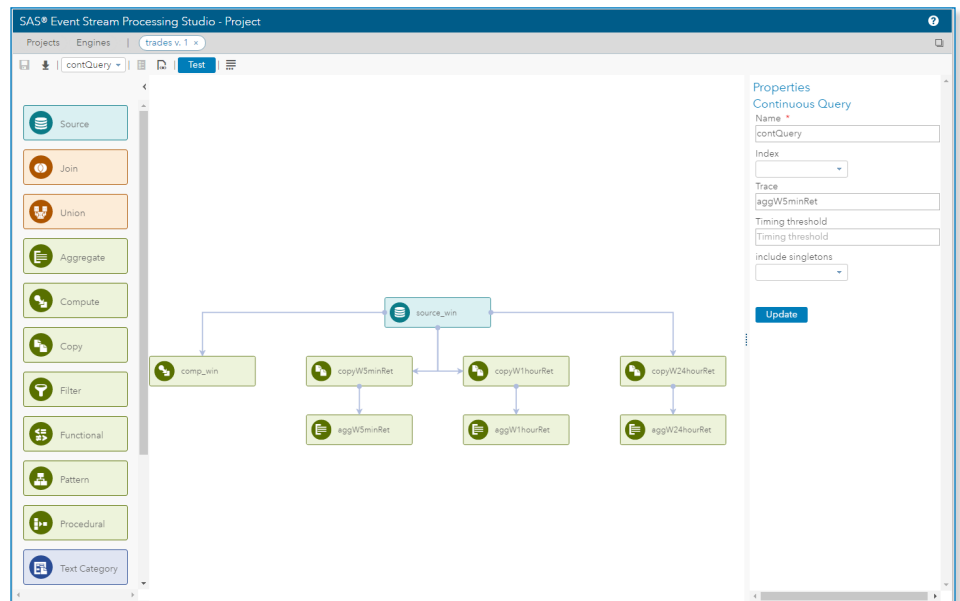


Рисунок 4. Используйте разные типы окон для формирования последовательности процессов для потоковой передачи данных.

Чтобы получить более подробную информацию о требованиях к системе для установки решения SAS Event Stream Processing, а также ознакомиться с другими полезными материалами, посетите сайт: sas.com/ESP.

Ключевые функции (окончание)

Корпоративное администрирование и управление

- Интерфейс RESTful для доступа к серверу фабрики XML упрощает создание нескольких проектов, проверку синтаксиса XML, управление окнами анализа и определение политик хранения проектов.
- Централизованное управление несколькими проектами снижает нагрузку на администраторов и упрощает определение элементов управления для динамической загрузки проектов, а также их запуска, остановки, удаления и сохранения.
- Поддерживается проверка подлинности для публикаций, подписок и установления HTTP-соединений с помощью клиентов на основе TCP/IP.
- Файлы конфигурации упрощают начальную настройку спецификации протокола.
- Настраиваются стратегии хранения событий с учетом размера, объема или времени – параметров, связанных с определениями политик.
- Детальная обработка ошибок во время выполнения.
- Локализованные сообщения в журналах.
- Возможность анализировать параметры обработки путем выполнения запросов к представляющим интерес потокам событий реального времени.
- Проверка подлинности и шифрование для потоков событий в других системах.

Свяжитесь с местным офисом SAS: sas.com/offices



Общество с ограниченной ответственностью «САС ИНСТИТУТ»

Россия, 109004 г. Москва, ул. Станиславского, дом 21, строение 1 • Тел.: +7 495 227 4151 • Факс: +7 495 780 4275 • www.sas.com/russia
Казахстан, 050051, г. Алматы, мкр-н Самал-2, ул. Жолдасбекова, 97, БЦ «Самал-Тауэрс», Блок А-1, 5 этаж. • Тел. +7 701 722 5502
Казахстан, 010000, г. Астана, ул. Сары-Арка, 6, офис 319. • Тел.: +7 777 733-32-33

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries.
® indicates USA registration. Other brand and product names are trademarks of their respective companies. Copyright © 2017, SAS Institute Inc. All rights reserved.