

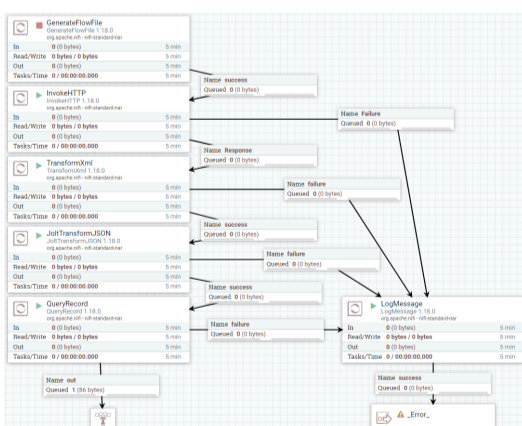
Текстовая расшифровка видео:

ПОНЯТИЯ И КОМПОНЕНТЫ

План:

- Как выглядит;
- Группы;
- Понятия и компоненты;
- Пример потока данных.

Как выглядит

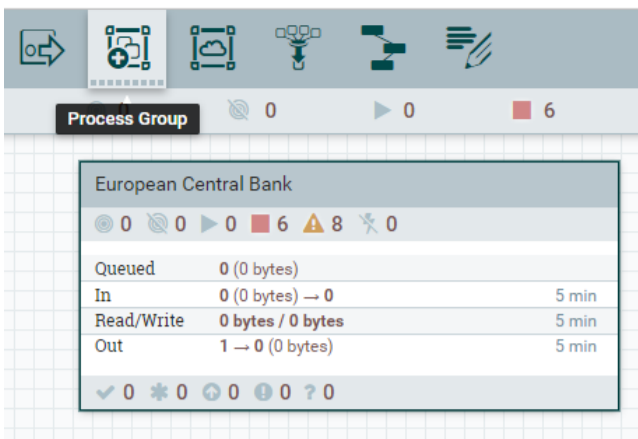


На данном примере вы можете видеть работающий поток, который забирает данные извне и обрабатывает их. Он готов передавать их дальше по цепочке.

Группы

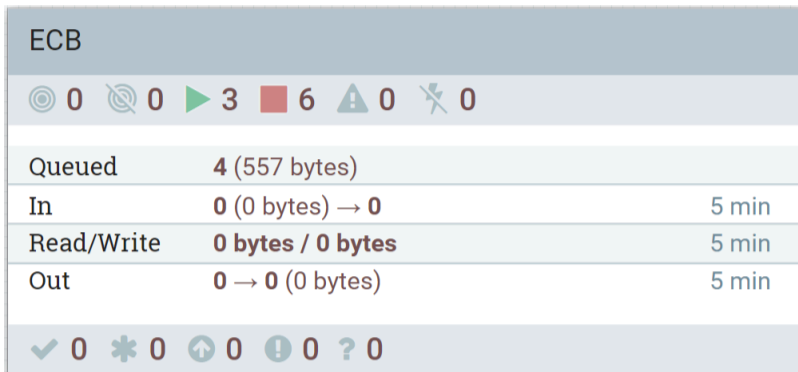
Любой проект и задачу стоит начинать с создания группы. Группы созданы для удобства и позволяют навешивать контроль версий через NiFi Registry. Они могут принимать данные и отдавать их дальше по цепочке:





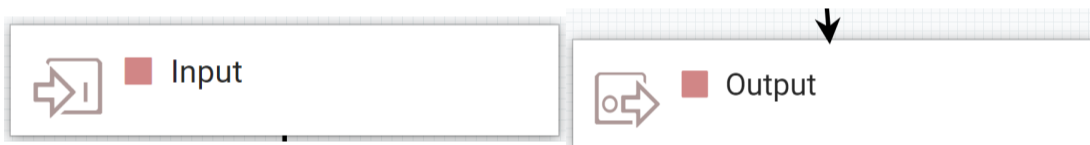
Понятия и компоненты

Process Group – набор процессоров, их подключений и прочих элементов Data Flow:

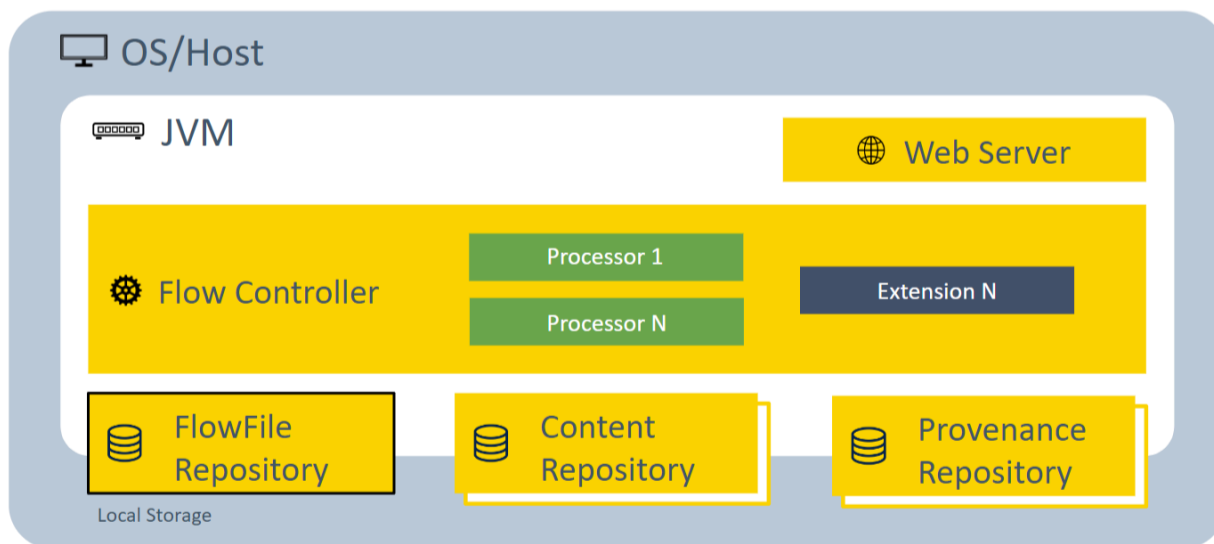


В группах для приёма и передачи данных могут быть размещены порты:

- Input;
- Output Ports.



FlowFile Processor – алгоритм (black box), который выполняет необходимую работу в NiFi:



Процессор может выполнять следующие задачи:

- Создание данных;
- Извлечение данных;
- Преобразование данных;
- Сохранение данных в целевой системе;
- Добавление/удаление/изменение атрибутов;
- Маршрутизация;
- Разделение;
- Слияние.

FlowFile – сущность, представляющая собой объект с содержимым от нуля и более байт и соответствующих ему атрибутов.

Атрибуты представляют собой **метаданные FlowFile**.

FlowFile (Потоковый файл) – это данные в системе NiFi:

- Файл состоит из контента и метаданных;
- Файл может создаваться по расписанию или событию;
- Файл может быть бинарным или текстовым.

Метаданные – это атрибуты.

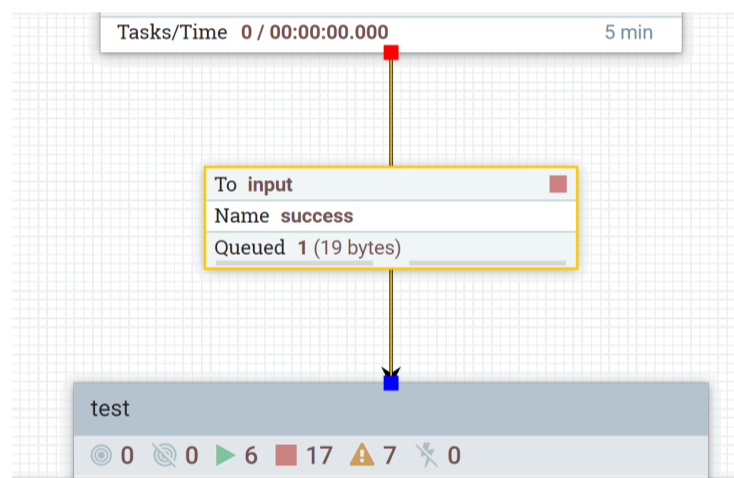
Например, атрибутом является имя файла, id файла, имя схемы и т.д.

Система позволяет добавлять собственные атрибуты и присваивать им любые значения. Значения атрибутов используются в работе процессоров.

Для того, чтобы FlowFile передавать от процессора к процессору нам необходимо соединение:

Соединение (Connection) обеспечивает подключение и передачу FlowFile между различными процессорами и некоторыми другими сущностями NiFi.

Connection помещает FlowFile в очередь, после чего передает его далее по цепочке:



На рисунке показано соединение (стрелка) с указанием направления движения FlowFile. Посередине мы видим очередь, именно она аккумулирует в себя FlowFile. Там задаются лимиты на размер и количество FlowFile в очереди. Через очередь происходит обмен файлами потока.

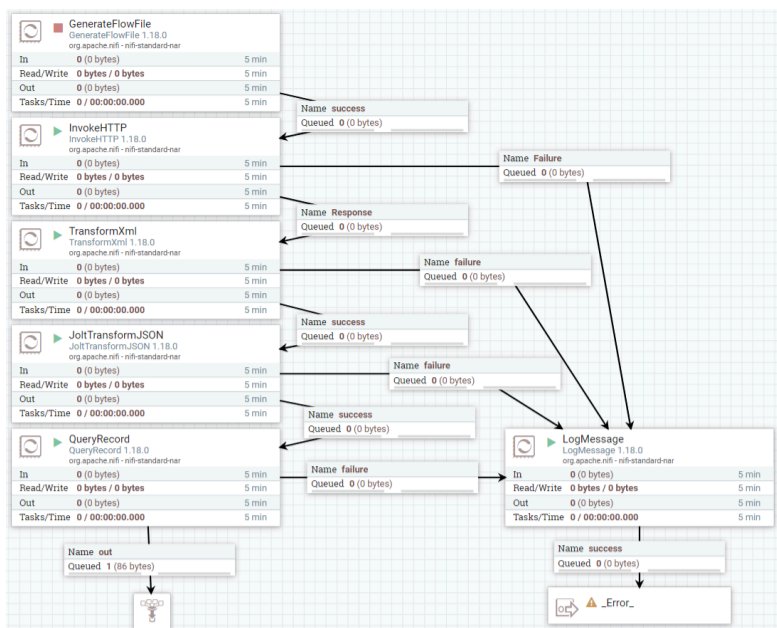
Процессор очень часто использует **контроллеры** – объекты, которые знают, как сделать какую-то работу. Внутри контроллера настраиваются драйвера, права доступа и т.д.

В конфигурации NiFi есть соответствующий раздел, посвященный контроллерам. В данном разделе их удобно располагать в одном месте и на одном уровне иерархии:

- **Reader** – контроллер, который знает, как читать данные;
- **RecordSetWriter** – контроллер, который форматирует данные на выходе процессора.

| Name ^ | Type | Bundle | State | Scope |
|------------------------|-------------------------------|--------------------------------------|---------|-------|
| CSVReader | CSVReader 1.18.0 | org.apache.nifi - nifi-record-ser... | Enabled | test |
| CSVRecordSetWriter | CSVRecordSetWriter 1.18.0 | org.apache.nifi - nifi-record-ser... | Enabled | test |
| JsonTreeReader | JsonTreeReader 1.18.0 | org.apache.nifi - nifi-record-ser... | Enabled | test |
| ParquetRecordSetWriter | ParquetRecordSetWriter 1.18.0 | org.apache.nifi - nifi-parquet-nar | Enabled | test |

Пример потока данных



Как мы видим, соединения в данном примере имеют очереди; здесь есть выходящий порт, а также данные, ждущие дальнейшей обработки.

Данный поток расположен внутри группы, чтобы мы могли делать контроль версий. Внутри каждого процессора может использоваться контроллер.

Используя графический интерфейс, мы будем создавать подобные потоки.

Как вам урок?



Изучил, далее >

Слёрм ©

+7 (495) 248-05-80

[Лицензия №ДЛ-1368 от 22.08.2019](#)

[Политика конфиденциальности](#)

[Публичная оферта](#)