

Сборник заметок по Apache NiFi

Содержание

1	Установка на Windows	1
2	Основные термины и определения	1
3	Архитектура Apache NiFi	2
4	Expression Language	2
5	Процессоры	3
5.1	JoltTransformJSON	3
6	Контроллер-сервис	3
6.1	DBCPCConnectionPool	3
	Список литературы	3
	Список листингов	3

1. Установка на Windows

Для того чтобы установить NiFi следует скачать архив бинарников со страницы проекта <https://nifi.apache.org/download/> (например, `nifi-1.25.0-bin.zip`). Затем нужно распаковать архив в целевую директорию, например, в `C:\Users\user\nifi`.

После чего нужно запустить файл `run-nifi.bat`, расположенный в поддиректории `bin`.

В конфигурационном файле `conf\nifi.properties` будет указан хост и порт

```
..
nifi.web.https.host=127.0.0.1
nifi.web.https.port=8443
...
```

А логин и пароль будут сгенерированы в файле `logs\nifi-app.log`

```
...
Generated Username [e4ef4cdf-5615-4748-8280-d57de0558375]
Generated Password [X9WaNiKWzLhhpn5s4i/qzLQLwzRwUcKs]
...
```

2. Основные термины и определения

Полезные материалы

<https://nifi.apache.org/documentation/v1/> Официальная документация

https://www.youtube.com/watch?v=yKXD1biS6V8&list=PL4MpKy3QjNp_r0EEibc4Ro8UK4g8vLX6_ Apache NiFi за 3 часа.

https://www.youtube.com/watch?v=NC3I1bjrKeA&list=PLpe6CiYAs7vTS4LG7kkvuPZEzEobvw_1R Apache NiFi

FlowFile – основной объект Apache NiFi, содержащий атрибуты данных (метаданные) и контент.

FlowFile Attribute – атрибуты FlowFile; хранятся в оперативной памяти

FlowFile Content – контент FlowFile; хранится на диске

Processor – одно действие / операция с данными FlowFile в Apache NiFi («прямоугольничек» на схеме NiFi)

Process Group – объединение Processors и других Process Groups

Process – объединение Processors и Process Groups с целью реализации целевого алгоритма

3. Архитектура Apache NiFi

Apache NiFi состоит из:

- веб-сервера,
- контроллера потоков и процессоров (встроенных или пользовательских),
- трех репозиториях: FlowFile Repository (хранит текущее состояние всех FlowFile'ов), Content Repository (контент всех FlowFile'ов) и Provenance Repository (хранит метрики, статистики и пр.).

Apache NiFi гарантирует строго однократную доставку. Если, например, выключат свет, то NiFi данные не потеряет, так как есть лог-файл. NiFi можно разворачивать в кластере для более быстрой параллельной загрузки и обработки данных.

Можно настроить процессор, чтобы он выполнялся либо на всех узлах, либо только на одном главном. Можно повесить ограничения на очередь. Скажем какая-то отдельно взятая очередь должна содержать не более 100 FlowFile'ов весом не более 10 Мб. Если условие нарушается, то NiFi приостановит процессор и таким образом высвободит ресурсы.

4. Expression Language

```
${abc} = значение атрибута  
${abc:toUpper()}  
${abc} = "${abc}" - экранирование
```

Ctrl +Space – подсказки в UI по EL. Все функции можно посмотреть здесь <https://nifi.apache.org/docs/nifi-docs/html/expression-language-guide.html>

Типы данных:

- String,
- Number,
- Decimal,
- Date,
- Boolean

После вычислений значения атрибутов приводятся к String. Есть функции приведения типов `$fileSize:number()`, `$time:date()`.

5. Процессоры

Процессоры в NiFi можно запускать по CRON-времени или через заданный промежуток времени. Есть удобный сервис CronMaker <http://www.cronmaker.com/?jsessionid=node09n1lwc8wiijupgvam6sbnode0?0>, который преобразует схему запуска в CRON-формат.

5.1. JoltTransformJSON

Очень мощный процессор. По кнопке «Advanced» открывается форма, в которой можно передать входной JSON, в поле «Jolt Specification» описать схему преобразования (пишется на специальном языке <https://jolt-demo.appspot.com/#inception>) и посмотреть как работает преобразование.

6. Контроллер-сервис

6.1. DBCPConnectionPool

Контроллер-сервису DBCPConnectionPool в поле «Database Connection URL» нужно передать URL целевой СУБД, например, `jdbc:mysql://50.224.35.38:6974/usgs` (`usgs` – это имя базы данных).

А в поле «Database Driver Class Names» указать класс драйвера, например, `com.mysql.cj.jdbc.Driver` и путь до директории, в которой лежит драйвер (поле «Database Driver Location(s)»), например, `/opt/nifi/nifi-current/custom-nar.`

Еще может потребоваться указать имя пользователя базы данных в поле «Database User» и пароль.

Драйвер позволяет выполнять SQL-запросы к базе данных и зависит от ее типа, поэтому его нужно предварительно загрузить с официального сайта разработчиков (для MySQL <https://www.mysql.com/products/connector/>, а для PostgreSQL <https://jdbc.postgresql.org/>)

Список литературы

1. *Собель М.* Linux. Администрирование и системное программирование. 2-е изд. – СПб.: Питер, 2011. – 880 с.

Листинги