

# Сборник заметок по Apache NiFi

## Содержание

<a href="#">1 Установка на Windows</a>	1
<a href="#">2 Основные термины и определения</a>	1
<a href="#">3 Архитектура Apache NiFi</a>	2
<a href="#">4 Expression Language</a>	2
<a href="#">Список литературы</a>	2
<a href="#">Список листингов</a>	2

## 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

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

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 («прямоугольник» 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 (контейнер 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()}`.

### Список литературы

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

### Листинги