

Дата-инженер

Дальнейшие шаги в изучении NiFi

Николай Акимов



План занятия

- 1 Функциональные возможности NiFi
- 2 Базовые процессоры
- 3 Дополнительные возможности
- 4 Режим кластера
- 5 Мониторинг и логи
- 6 Интерфейс происхождения данных (Data Provenance)

1

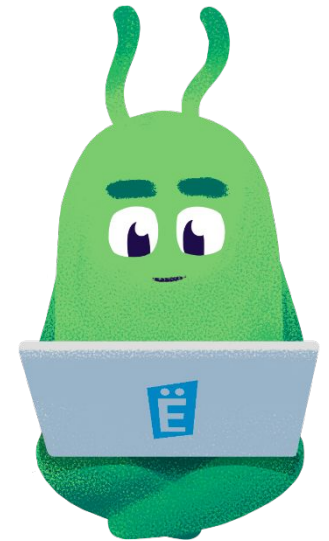
Обзор функциональных возможностей NiFi



Преобразование данных

NiFi позволяет:

- Парсить данные с помощью XPath, XQuery, JOLT или RegExp
- Выполнять SQL-запросы по содержимому Flowfile
- Фильтровать и добавлять поля
- Конвертировать один формат данных в другой



Преобразование данных

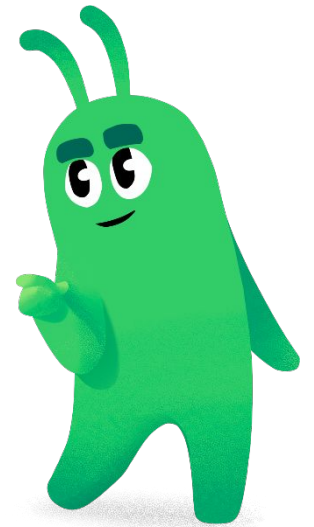
NiFi имеет собственный язык выражений, богатый различными операторами и встроенными функциями

С его помощью можно:

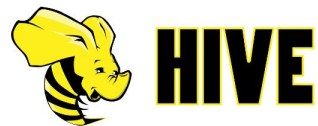
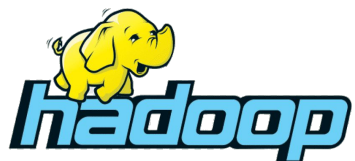
- Вызывать встроенные функции
- Производить логические вычисления
- Преобразовывать формат содержимого атрибутов
- Производить поиск данных
- Производить арифметические вычисления
- Преобразовывать дату и время
- Производить оценку нескольких атрибутов
- и прочее

2

Обзор базовых процессоров NiFi



Базовые процессоры NiFi



Базовые процессоры NiFi

SplitAvro

SplitContent

SplitJson

SplitRecord

SplitText

SplitXml

MergeContent

MergeJsonProcessor

MergeRecord

MergeResultArrays

RouteOnAttribute

RouteOnContent

RouteText

AttributesToCSV

AttributesToJSON

ConvertCharacterSet

ConvertRecord

ReplaceText

ScanAttribute

CountText

EncryptContent

CompressContent

ExtractText

ExtractText

ExecuteGroovyScript

ExecuteStreamCommand

ExecuteProcess

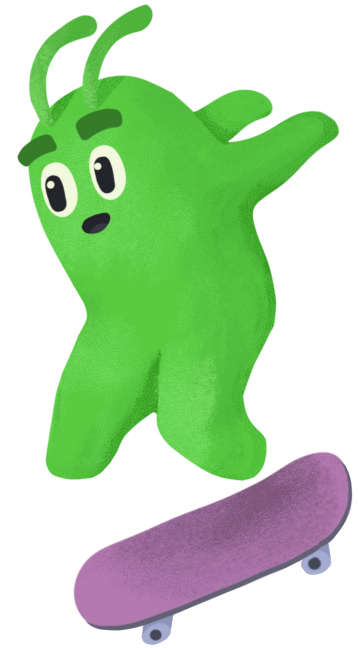
ModifyBytes

ModifyHTMLElement

ParseSyslog

ScanContent

Обзор дополнительных возможностей NiFi

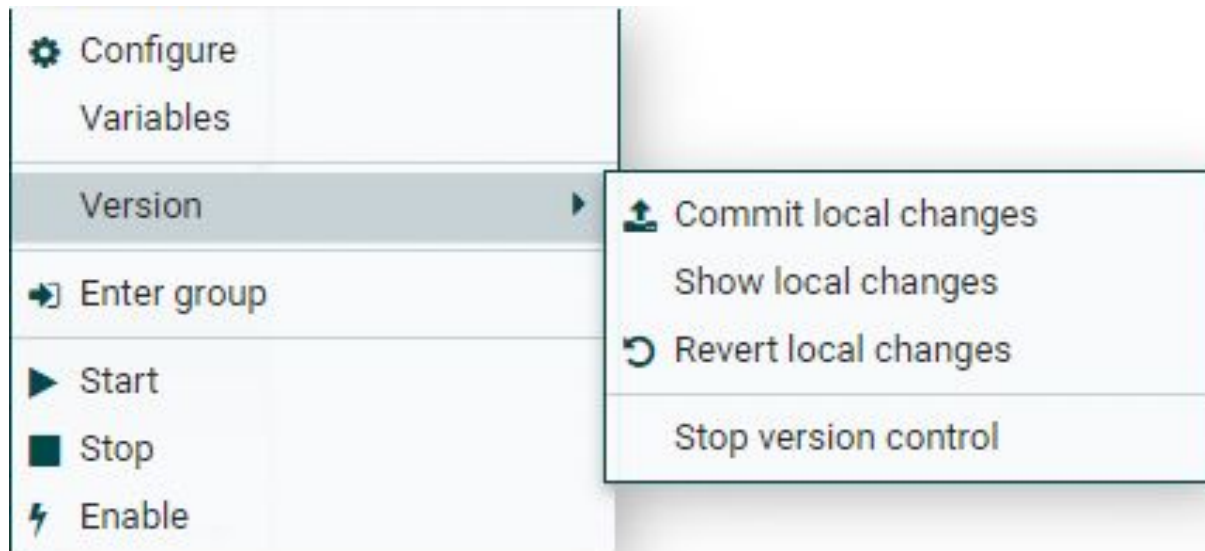


Управление изменениями

Для версионирования DataFlow существует отдельный сервис NiFi Registry

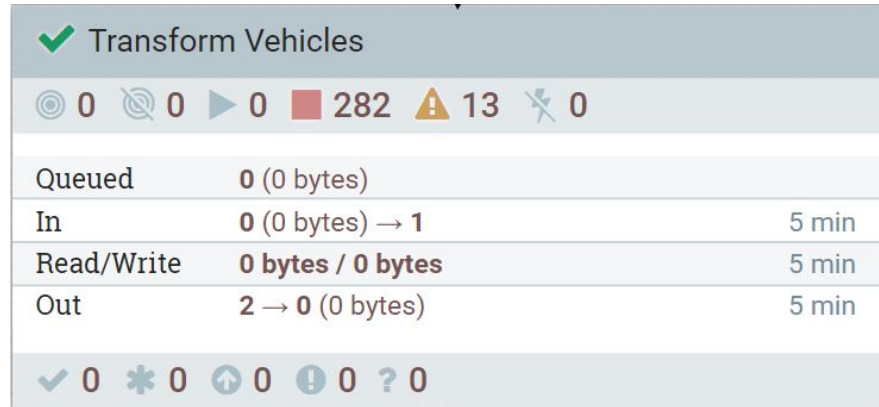
Настроив его, вы получаете возможность управлять изменениями:

- **Зафиксировать** локальные изменения
- **Откатиться** назад
- **Загрузить** любую предыдущую версию



Управление изменениями

«Галочка» – сигнал о том, что изменений не было



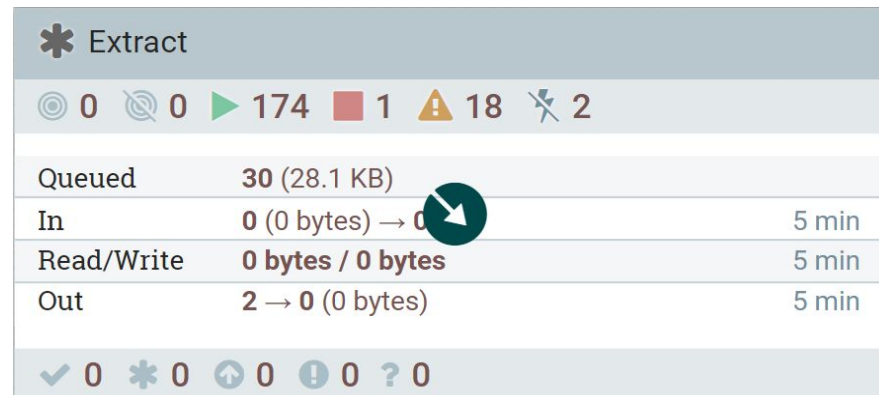
Transform Vehicles

0 0 0 282 13 0

Queued	0 (0 bytes)	
In	0 (0 bytes) → 1	5 min
Read/Write	0 bytes / 0 bytes	5 min
Out	2 → 0 (0 bytes)	5 min

0 * 0 0 0 ? 0

«Звёздочка» – сигнал о том, что в группе есть изменения Data Flow



* Extract

0 0 174 1 18 2

Queued	30 (28.1 KB)	
In	0 (0 bytes) → 0	5 min
Read/Write	0 bytes / 0 bytes	5 min
Out	2 → 0 (0 bytes)	5 min

0 * 0 0 0 ? 0

Управление изменениями

Детализация изменений

Show Local Changes

The following changes have been made to Extract (Version 89).

Displaying 2 of 2

Filter

Component Name ▲	Change Type	Difference	
GenerateFlowFile	Property Value Changed	From '{"_id": {"\$oid": "3cf1543e-be4d-1...	→
GenerateFlowFile	Scheduled State Changed	Processor was enabled	→

Управление доступом

[NiFi System Administrator's Guide \(apache.org\)](#)

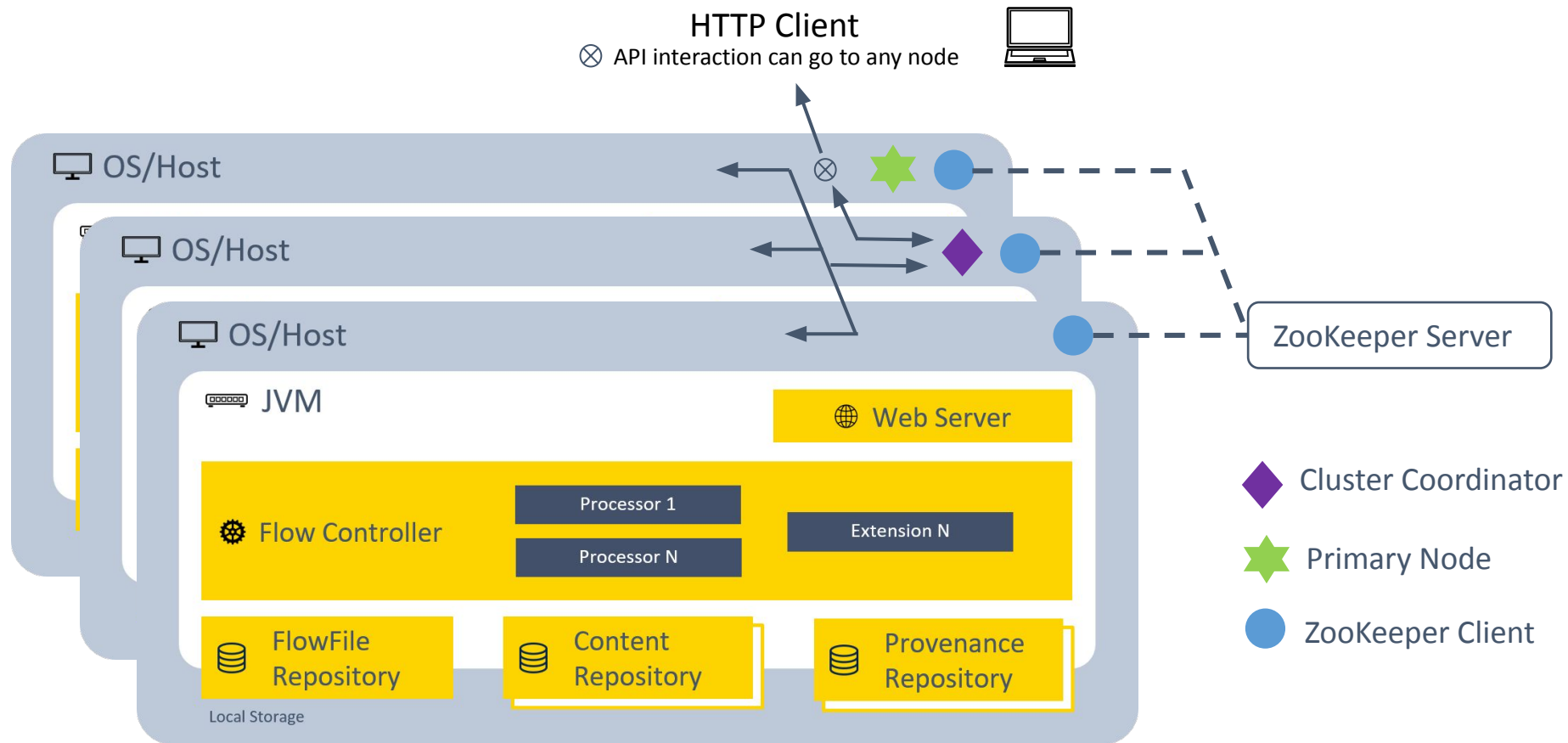
- В NiFi можно управлять доступом к веб-интерфейсу и разделением прав пользователей
- Поддерживаются следующие механизмы аутентификации:
 - На основе сертификатов
 - На основе имени пользователя и пароля посредством LDAP и Kerberos
 - Через Apache Knox
 - Через OpenID Connect

4

Режим кластера



Режим кластера

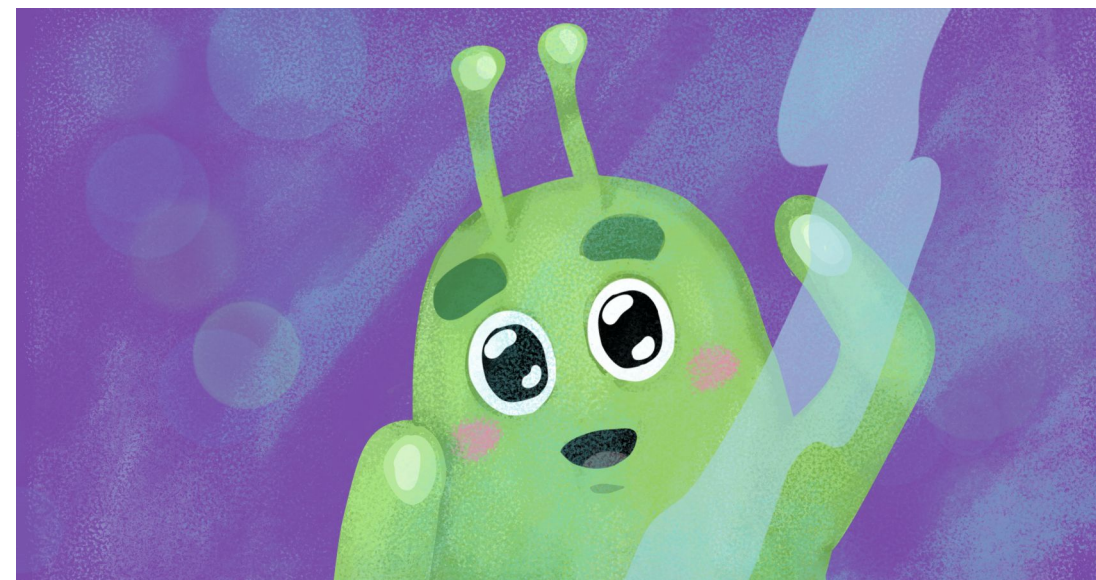
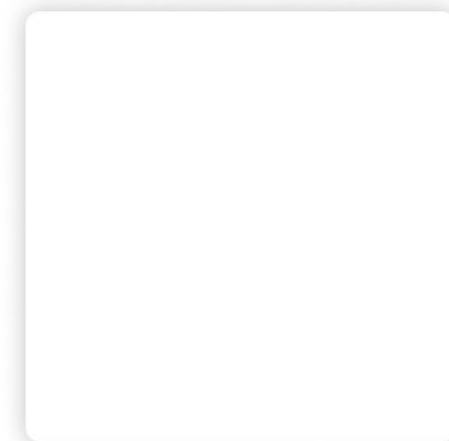


Режим кластера

- Работа в режиме кластера обеспечивает отказоустойчивость и дает возможность горизонтально масштабировать нагрузку
- Статично зафиксированной мастер-ноды нет. Вместо этого **Apache Zookeeper** выбирает одну ноду в качестве координатора и одну в качестве primary
- Primary-нода служит для запуска изолированных процессоров, которые не должны запускаться на всех нодах одновременно

Режим кластера

- Summary
- Counters
- Bulletin Board
- Data Provenance
- Controller Settings
 - Parameter Contexts
- Cluster
- Flow Configuration History
- Node Status History
- Templates
- Help
- About



Режим кластера



NiFi Cluster

NODES	SYSTEM	JVM	FLOWFILE STORAGE	CONTENT STORAGE	PROVENANCE STORAGE	VERSIONS
-------	--------	-----	------------------	-----------------	--------------------	----------

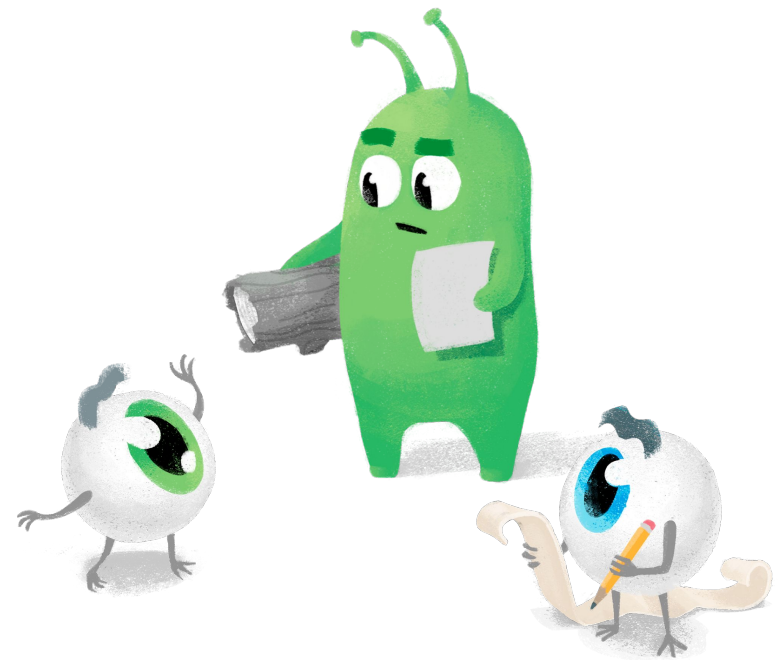
Displaying 2 of 2

Filter	by address ▾						
	Node Address	Active Thread Count	Queue / Size	Status	Started At	Last Heartbeat	
ⓘ	192.168.182.16:8081	2	6 / 40,89 KB	CONNECTED	04/08/2022 17:44:46 M...	11/09/2022 13:47:08 ...	⏻
ⓘ	192.168.182.15:8081	0	3 / 23,24 KB	CONNECTED, PRIMARY, ...	04/06/2022 19:35:39 M...	11/09/2022 13:47:07 ...	⏻

🔄 Last updated: 13:47:11 MSK

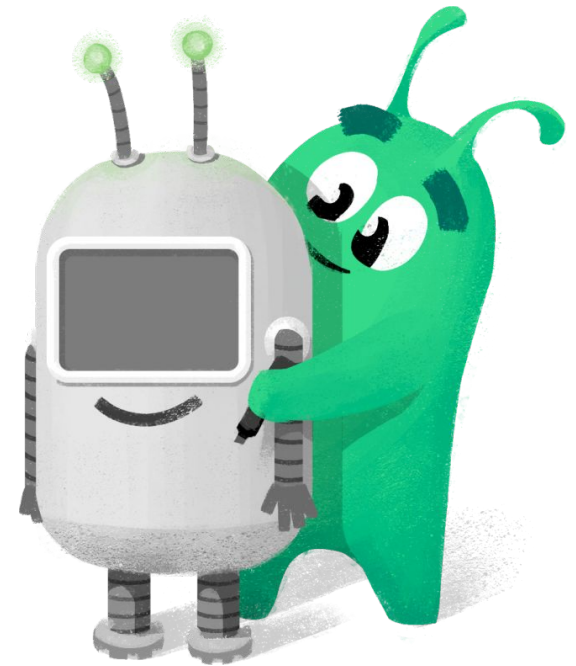


Мониторинг и ЛОГИ



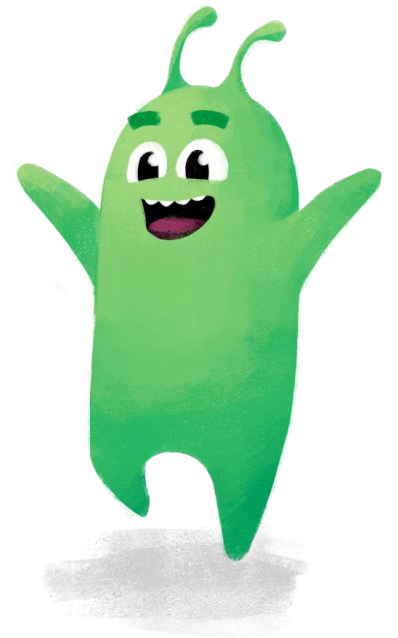
6

Data Provenance



Итоги. О чём поговорили:

- 1 Функциональные возможности NiFi
- 2 Базовые процессоры
- 3 Дополнительные возможности
- 4 Режим кластера
- 5 Мониторинг и логи
- 6 Интерфейс происхождения данных (Data Provenance)



Дополнительные материалы

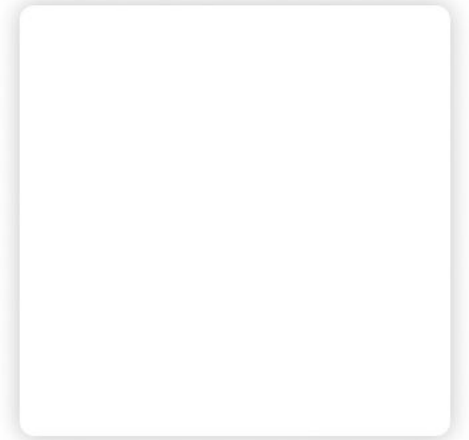
[Перечень книг, полезных в изучении NiFi](#)



Вопросы по занятию

Что такое NiFi?

Перечислите ключевые компоненты NiFi





**Спасибо
за внимание!**

