

# ساخت زیرساخت مهندسی داده

---

در فصل قبل، با نقش‌ها و مسئولیت‌های مهندس داده آشنا شدیم و ابزارهای کلیدی مورد استفاده در این حوزه را **Data Pipeline** معرفی کردیم؛ از جمله انواع دیتابیس‌ها، زبان‌های برنامه‌نویسی، و ابزارهای ایجاد و زمان‌بندی

در این فصل، قصد داریم ابزارهای موردنیاز را نصب و پیکربندی کنیم تا در ادامه‌ی مسیر از آن‌ها برای ساخت پایپ‌لاین‌های داده و تحلیل نتایج استفاده کنیم.

---

## ابزارهایی که در این فصل نصب و پیکربندی می‌کنیم

♦ دیتابیس‌ها:

- PostgreSQL
- Elasticsearch

♦ ابزارهای ساخت پایپ‌لاین داده:

- Apache Airflow
- Apache NiFi

♦ ابزارهای مدیریتی و بصری‌سازی داده:

- pgAdmin (برای مدیریت PostgreSQL)
  - Kibana (برای مدیریت و تحلیل داده در Elasticsearch)
- 

## چرا به این ابزارها نیاز داریم؟

- ✓ اجرای پایپ‌لاین‌های داده: انتقال داده از یک منبع به مقصد
- ✓ ETL مدیریت پردازش داده: زمان‌بندی و هماهنگ‌سازی فرآیندهای
- ✓ بصری‌سازی داده: مشاهده و تحلیل داده‌های پردازش‌شده برای دیباگ و تحلیل
- ✓ برای ذخیره و جستجوی داده‌ها Elasticsearch و PostgreSQL مدیریت و ذخیره‌سازی داده: استفاده از

هرچند با پیشرفت در مسیر یادگیری، ممکن است دیگر نیازی به برخی از این ابزارها نداشته باشید، اما درک پایه‌ای از آن‌ها برای پشتیبانی از سایر کاربران و نقش‌ها در تیم داده ضروری است.

---

## موضوعات این فصل

- ♦ Apache NiFi نصب و پیکربندی
- ♦ Apache Airflow نصب و پیکربندی
- ♦ Elasticsearch نصب و پیکربندی
- ♦ Kibana نصب و پیکربندی
- ♦ PostgreSQL نصب و پیکربندی
- ♦ pgAdmin 4 نصب

در ادامه، گام به گام این ابزارها را نصب و تنظیم خواهیم کرد تا بتوانید از آن‌ها در پروژه‌های مهندسی داده خود استفاده کنید.

## Apache NiFi نصب و پیکربندی

یکی از ابزارهای اصلی برای ساخت پایپ‌لاین‌های مهندسی داده است. این ابزار به شما امکان می‌دهد **Apache NiFi** بدون نیاز به کدنویسی، پایپ‌لاین‌های داده‌ای بسازید و از پردازشگرهای از پیش ساخته شده استفاده کنید.

**NiFi:** ویژگی‌های کلیدی

- ✓ عدم نیاز به کدنویسی برای اجرای پایپ‌لاین‌ها
- ✓ دارای زمان‌بندی داخلی برای اجرای خودکار پردازش‌ها
- ✓ (در صورت ناهماهنگی سرعت پردازش بین دو تسک) **backpressure** مدیریت

## Apache NiFi مراحل نصب

### ۱. دانلود Apache NiFi

را از [وبسایت رسمی](#) دانلود کنید یا از دستور زیر استفاده کنید NiFi:

```
curl https://mirrors.estointernet.in/apache/nifi/1.12.1/nifi-1.12.1-bin.tar.gz  
-o nifi.tar.gz
```

### ۲. استخراج فایل‌های NiFi

پس از دانلود، فایل را استخراج کنید:

```
tar xvzf nifi.tar.gz
```

### ۳. اجرای Apache NiFi

را اجرا کنید NiFi وارد فولدر استخراج شده شوید و

```
cd nifi-1.12.1  
bin/nifi.sh start
```

### ۴. بررسی وضعیت اجرای NiFi

برای بررسی وضعیت اجرا، دستور زیر را وارد کنید:

```
sudo bin/nifi.sh status
```

✓ **تنظیم باشد،** مسیر آن در خروجی نمایش داده می‌شود **JAVA\_HOME** اگر

## ۵. نصب Java (در صورت نیاز)

را نصب کنید Java، تنظیم نبود **JAVA\_HOME** اگر:

```
sudo apt install openjdk-11-jre-headless
```

## ۶. **JAVA\_HOME** تنظیم متغیر

اضافه کنید **bash\_profile**. را در فایل Java مسیر

```
export JAVA_HOME=/usr/lib/jvm/java-11-openjdk-amd64
```

## ۷. بارگذاری مجدد تنظیمات

برای اعمال تغییرات، دستور زیر را اجرا کنید:

```
source ~/.bash_profile
```

## ۸. **NiFi** بررسی مجدد اجرای

را بررسی کنید NiFi دوباره وضعیت

```
sudo bin/nifi.sh status
```

✓ به درستی نصب شده و آماده اجرا است NiFi، نمایش داده شد **JAVA\_HOME** اگر مسیر

A terminal window showing the command 'sudo nifi\*/bin/nifi.sh start' and its output. The output displays the Java home path, NiFi home path, and the bootstrap config file path.

```
paulcrickard@pop-os:~$ sudo nifi*/bin/nifi.sh start
Java home: /usr/lib/jvm/java-11.0-openjdk-amd64
NiFi home: /home/paulcrickard/nifi-1.11.3
Bootstrap Config File: /home/paulcrickard/nifi-1.11.3/conf/bootstrap.conf
paulcrickard@pop-os:~$
```

Figure 2.1 – NiFi is running

## ۹. **NiFi** اجرای رابط کاربری

مرورگر خود را باز کنید و به آدرس زیر بروید NiFi، پس از اجرای

```
http://localhost:8080/nifi/
```

✓ را مشاهده کنید **NiFi** اگر نصب به درستی انجام شده باشد، باید رابط گرافیکی

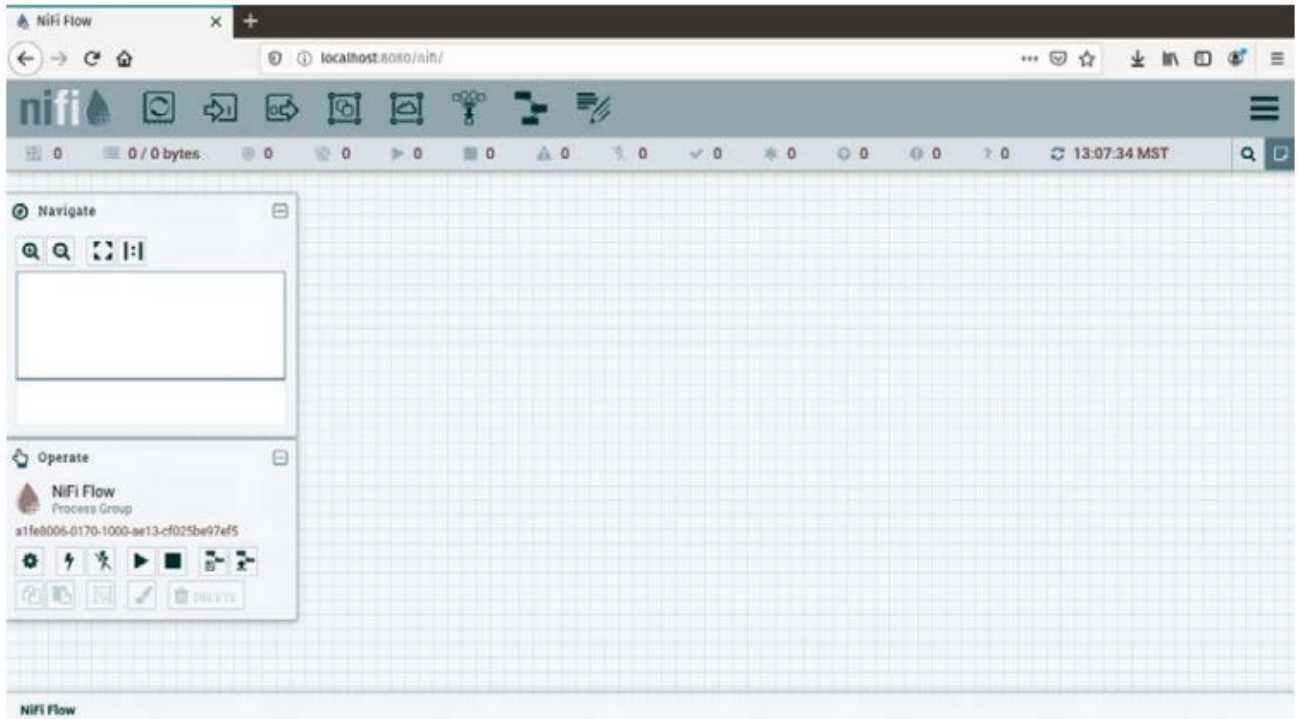


Figure 2.2 – The NiFi GUI

## NiFi تغییر پورت پیش فرض

روی پورت **8080** اجرا می شود. برای تغییر این پورت به **9300**، مراحل زیر را انجام دهید NiFi، به صورت پیش فرض

1 فایل پیکربندی را ویرایش کنید:

```
nano conf/nifi.properties
```

2 مقدار پورت پیش فرض را تغییر دهید:

```
# web properties #  
nifi.web.http.port=9300
```

3 اگر فایروال فعال است، پورت جدید را باز کنید:

```
sudo ufw allow 9300/tcp
```

4 را مجدداً راه اندازی کنید **NiFi**:

```
bin/nifi.sh restart
```

5 را در پورت جدید مشاهده کنید GUI اکنون:

```
http://localhost:9300/nifi/
```

## Apache NiFi (GUI) آشنایی با محیط گرافیکی

صفحه خالی خواهد بود زیرا هنوز پردازشی تعریف نشده است، NiFi پس از ورود به

✓ در بالای صفحه، دو بخش مهم وجود دارد

◆ ابزارهای موردنیاز برای ایجاد جریان داده: **نوار ابزار (Component Toolbar)**

◆ **NiFi** نمایش اطلاعات مربوط به وضعیت: **نوار وضعیت (Status Bar)**



Figure 2.3 – NiFi component toolbar and status bar

The tool you will use the most is the **Processor** tool. The other tools, from left to right, are as follows:

- **Input Port**
- **Output Port**
- **Processor Group**
- **Remote Processor Group**
- **Funnel**
- **Template**
- **Label**

## NiFi در (Data Flow) ایجاد اولین جریان داده

### (Processor) اضافه کردن یک پردازشگر ۱.

1 (Drag & Drop) کلیک کنید و آن را به صفحه بکشید **Processor** روی آیکون

2 را انتخاب کنید **GenerateFlowFile** در پنجره جستجو، پردازشگر

◆ این پردازشگر، فایل‌هایی را با محتوای متنی تولید می‌کند

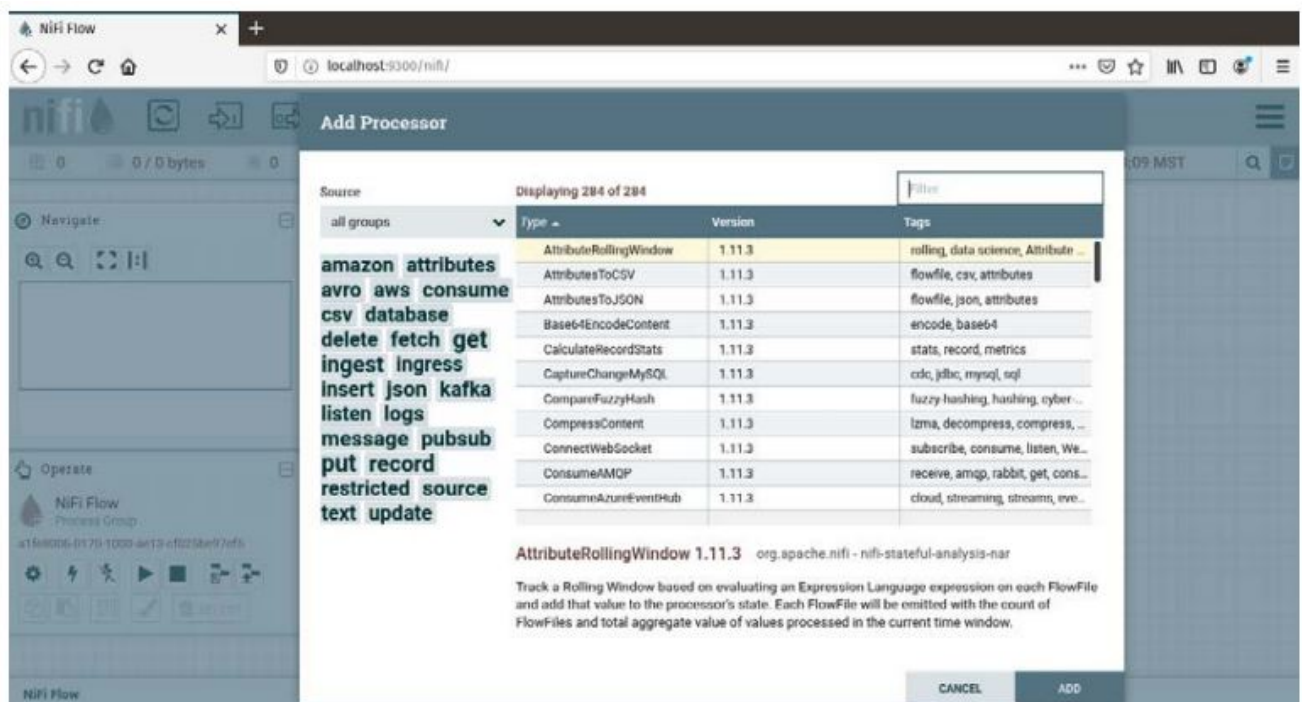


Figure 2.4 – Processors you can add to the canvas

## ۲. اضافه کردن پردازشگر دوم (Processor)

۱. را به صفحه بکشید **Processor** مجدداً ابزار
  ۲. را انتخاب کنید **PutFile** این بار پردازشگر
- این پردازشگر، فایل‌های تولیدشده را در دیسک ذخیره می‌کند.
- ✓ ساخته‌اید که داده تولید کرده و آن را در فایل ذخیره می‌کند NiFi اکنون یک پایپ‌لاین ساده در

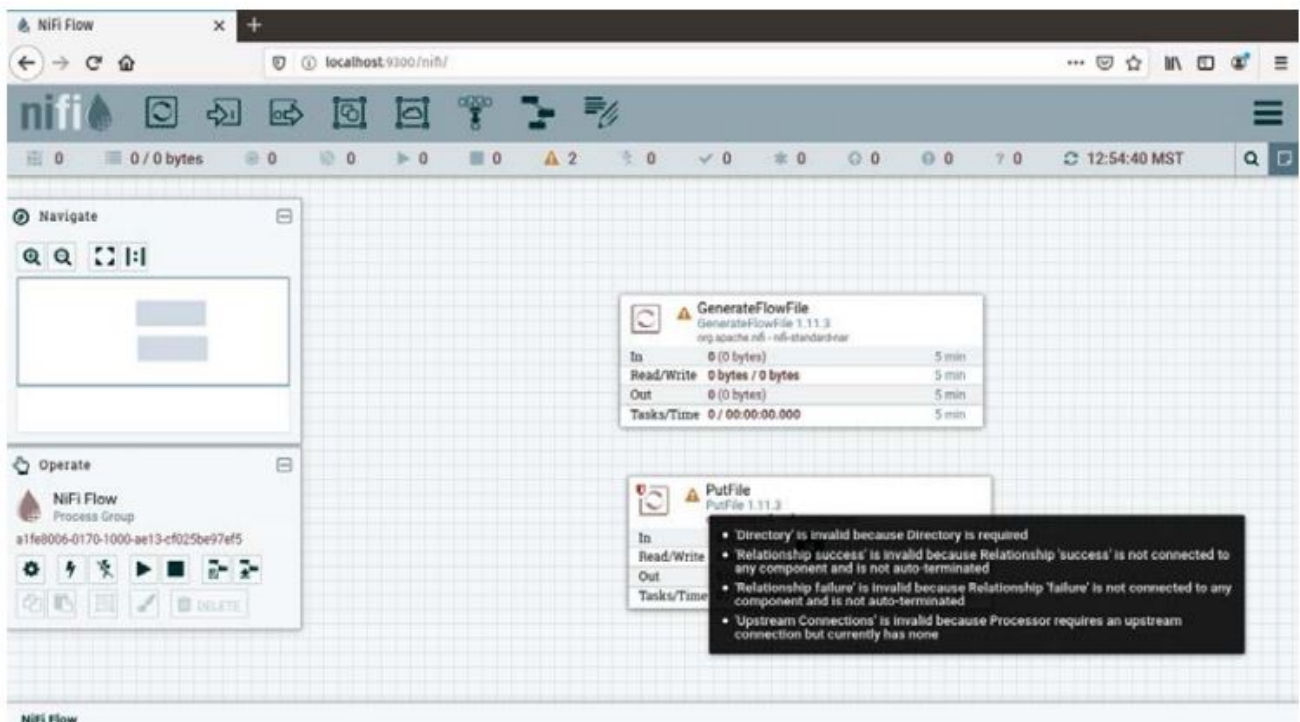


Figure 2.5 – Processors added to the canvas – with errors

را اضافه می‌کنید، یک **نماد هشدار** (⚠) در گوشه چپ آن‌ها ظاهر می‌شود. این (**Processors**) هنگامی که پردازشگرها هشدارها به دلیل **عدم پیکربندی** پردازشگرها و وجود **خطاهای تنظیماتی** هستند.

### ● دلایل اصلی این هشدارها:

- ندارد **Directory** مقدار **PutFile** پردازشگر.
- بین پردازشگرها تعریف نشده است (**Connection**) هیچ اتصالی.
- تنظیم نشده‌اند (**Failure**) و ناموفق (**Success**) برای حالت‌های موفق (**Relationships**) روابط.

## ۱. پیکربندی پردازشگرها (Processors Configuration)

✓ برای تنظیم پردازشگرها، می‌توانید:

- روی پردازشگر **دوبار کلیک** کنید.
- را انتخاب کنید **Properties** یا روی پردازشگر **کلیک راست** کرده و گزینه

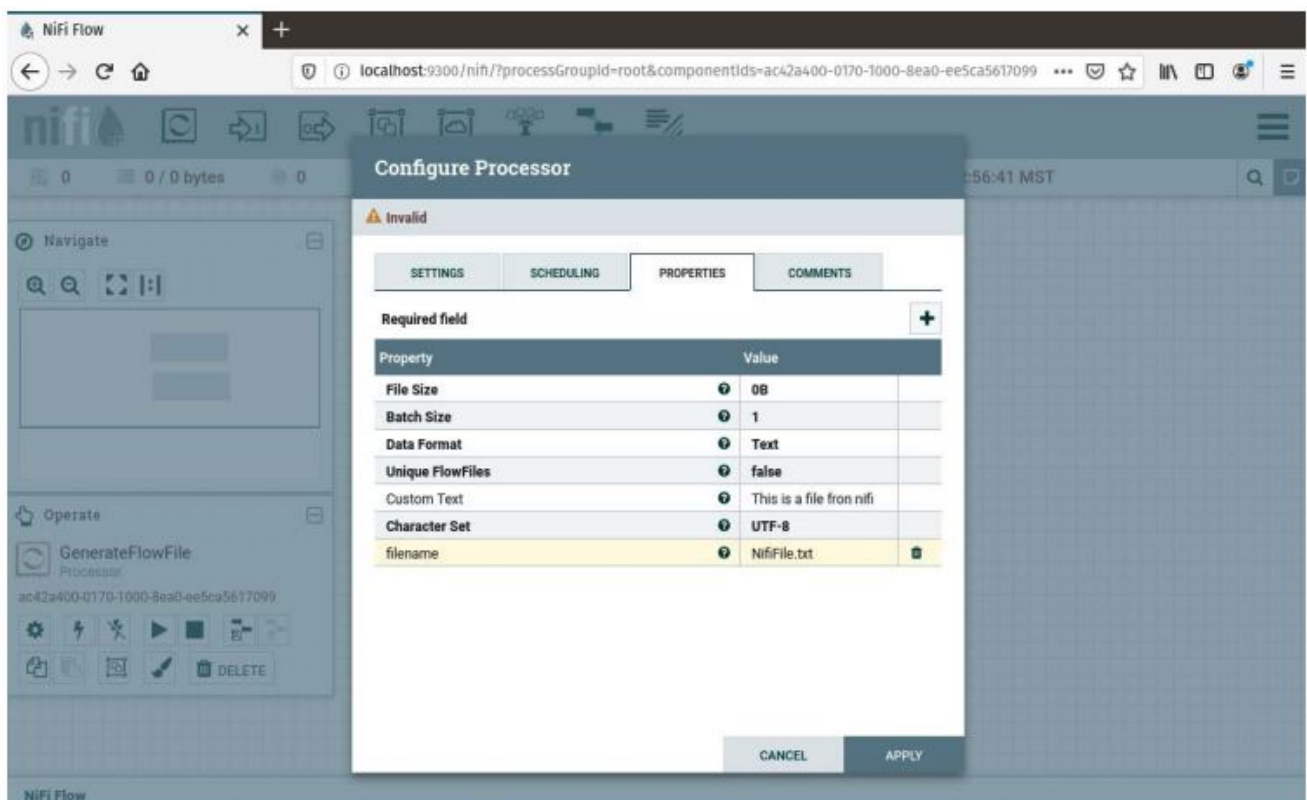


Figure 2.6 – Configuring the GenerateFlowFile processor

### ✈ مراحل پیکربندی:

#### ۱. تنظیم پارامترهای ضروری

- ♦ **پررنگ** هستند، باید مقداردهی شوند (**Bold**) در پنجره تنظیمات، مقادیر پارامترهایی که به صورت
- ♦ در کنار هر پارامتر، **علامت سؤال** (?) وجود دارد که توضیحی درباره آن ارائه می‌دهد.

#### ۲. GenerateFlowFile تنظیم پردازشگر

- ♦ این پردازشگر **نیازی به تغییر ندارد** زیرا مقادیر لازم از قبل پر شده‌اند.
- ♦ اضافه کنید (**Custom Properties**) اگر می‌خواهید ویژگی‌های سفارشی:

- روی **علامت +** در گوشه بالا سمت راست کلیک کنید.
- در پنجره باز شده، یک نام و مقدار مشخص کنید.



- مثال:
  - نام: `filename`
  - مقدار: `"This is a file from NiFi"`

پس از تنظیمات، نماد هشدار (⚠️) به یک مربع زرد (■) تغییر می‌کند که نشان‌دهنده توقف پردازشگر است. ✓

## ۲. اتصالات و ارتباطات (Connections & Relationships)

ایجاد کنید **PutFile** باید یک اتصال به **GenerateFlowFile** پس از پیکربندی

مراحل اتصال پردازشگرها:

1. ببرید تا یک دایره و فلش ظاهر شود **GenerateFlowFile** موس را روی پردازشگر
2. رها کنید **PutFile** و آن را روی پردازشگر (**Drag**) دایره را بکشید
3. که به صورت پیش‌فرض فعال است) را تأیید کنید **Success** در پنجره باز شده، گزینه
4. کلیک کنید **OK** روی

ارسال خواهند شد **PutFile** به **GenerateFlowFile** حالا پردازشگرها به هم متصل شده‌اند و داده‌ها از ✓

## ۳. اجرای جریان داده (Running the Data Flow)

برای اجرای پردازشگرها:

1. کلیک راست کنید **GenerateFlowFile** روی پردازشگر
2. را انتخاب کنید **Run** گزینه
3. نماد مربع قرمز (●) به دکمه سبز پخش (▶) تغییر خواهد کرد

!اکنون جریان داده شما آماده اجراست ✓

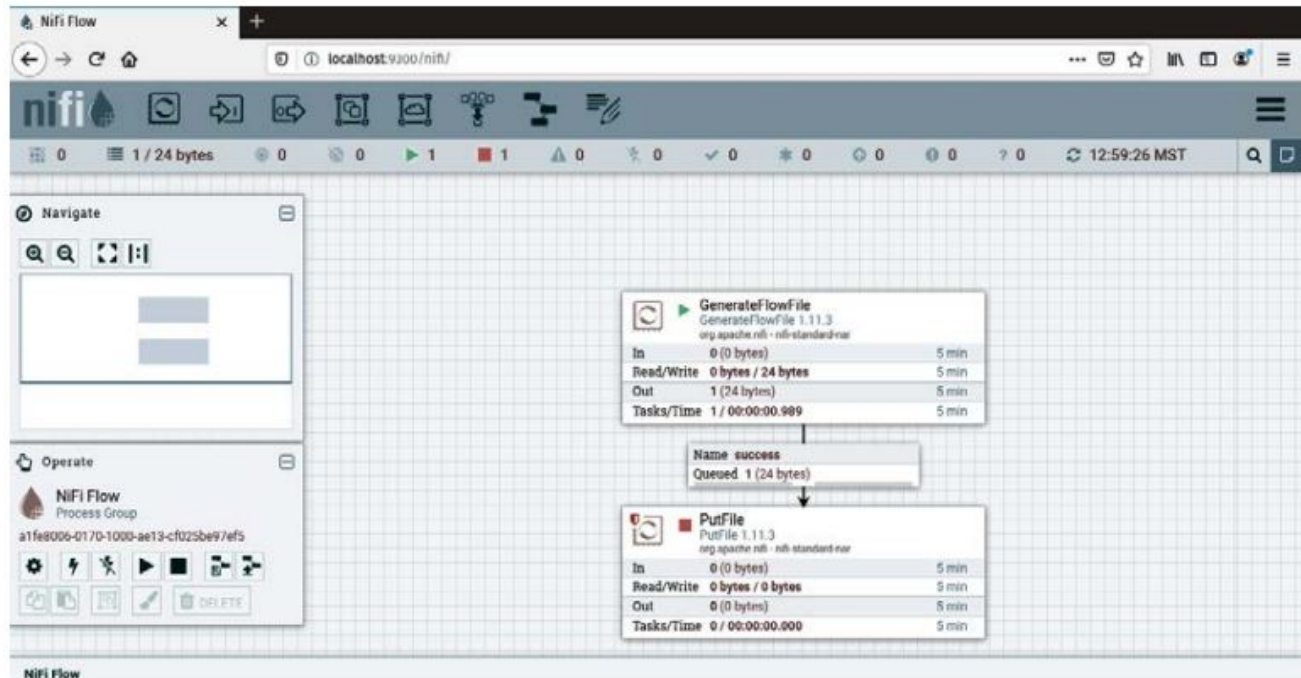


Figure 2.7 – Data flow half running

## NiFi ها در FlowFile مدیریت و مشاهده

ها و اندازه آن‌ها را نمایش **FlowFile** وجود دارد که تعداد (**Queue**) یک صف (**Processor**) در فاصله بین دو پردازشگر ها را مشاهده کنید، جزئیات مربوط به هر کدام را ببینید **FlowFile** می‌دهد. با کلیک راست روی صف، می‌توانید لیستی از



و حتی آن ها را دانلود کنید.



Figure 2.8 – List of FlowFiles in the queue

## FlowFile مشاهده جزئیات

دو تب وجود دارد، FlowFile در نمای جزئیات

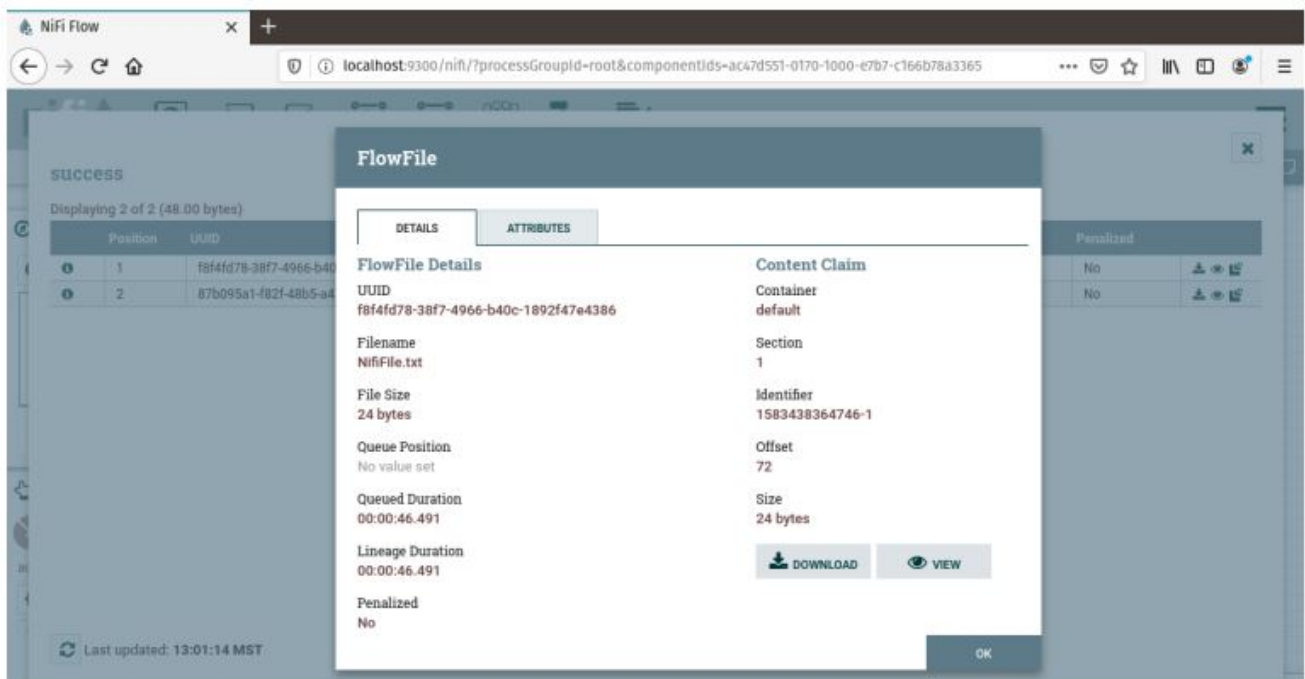


Figure 2.9 – Details of a FlowFile

### 1. DETAILS:

- NiFi شامل متادیتای
- FlowFile امکان مشاهده و دانلود

### 2. ATTRIBUTES:

- NiFi نمایش ویژگی های اختصاص داده شده توسط
- نمایش ویژگی های تعریف شده در پایپ لاین داده

را مشاهده کنید که برای داده‌های متنی مناسب‌تر است. امکان مشاهده فایل در **FlowFile** همچنین، می‌توانید محتوای نیز **(Formatted)** یا **فرمت‌بندی‌شده (Raw)** و همچنین انتخاب نمایش داده به صورت خام **(Hex)** فرمت هگزادسیمال وجود دارد.



Figure 2.10 – Contents of a FlowFile

## روی دیسک **FlowFile** ذخیره

را به عنوان یک فایل روی سیستم ذخیره می‌کند. مسیر پیش فرض این فایل **FlowFile** **PutFile** پردازشگر **opt/nifioutput/** (Root) است، اما در تنظیمات پردازشگر می‌توانید مسیر آن را تغییر دهید. اگر دسترسی ریشه **opt/nifioutput/** خود تغییر دهید **Home** ندارید، بهتر است مسیر ذخیره را به دایرکتوری

با اجرای این پایپ‌لاین، هر 10 ثانیه یک فایل جدید تولید شده و در دیسک ذخیره می‌شود، اما این فایل بر روی فایل قبلی بازنویسی خواهد شد.



Figure 2.11 – Output of the data flow

## نصب: گام بعدی **Apache Airflow**

خواهد بود و در فصل‌های بعدی، جزئیات بیشتری در مورد ساخت **Apache NiFi** در ادامه، تمرکز اصلی روی است. **Apache Airflow**، پایپ‌لاین‌های داده ارائه خواهد شد. همچنین، ابزار بعدی که نصب و بررسی خواهد شد