

NiFi

Grimaldo Oliveira



O que é o NiFi

1. Apache NiFi é uma plataforma de ingestão de dados, criada para processar e distribuir dados entre diferentes sistemas.
2. Sua principal funcionalidade é o gerenciamento e a automatização do fluxo de dados entre os sistemas.
3. Para execução destes fluxos de dados é fornecida uma interface WEB . É a programação baseada em fluxos.
4. É possível realizar: coleta de dados, transmissão de dados, armazenamento de dados, dentre outras tarefas.

Sites Importantes

NiFi

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

Documentação

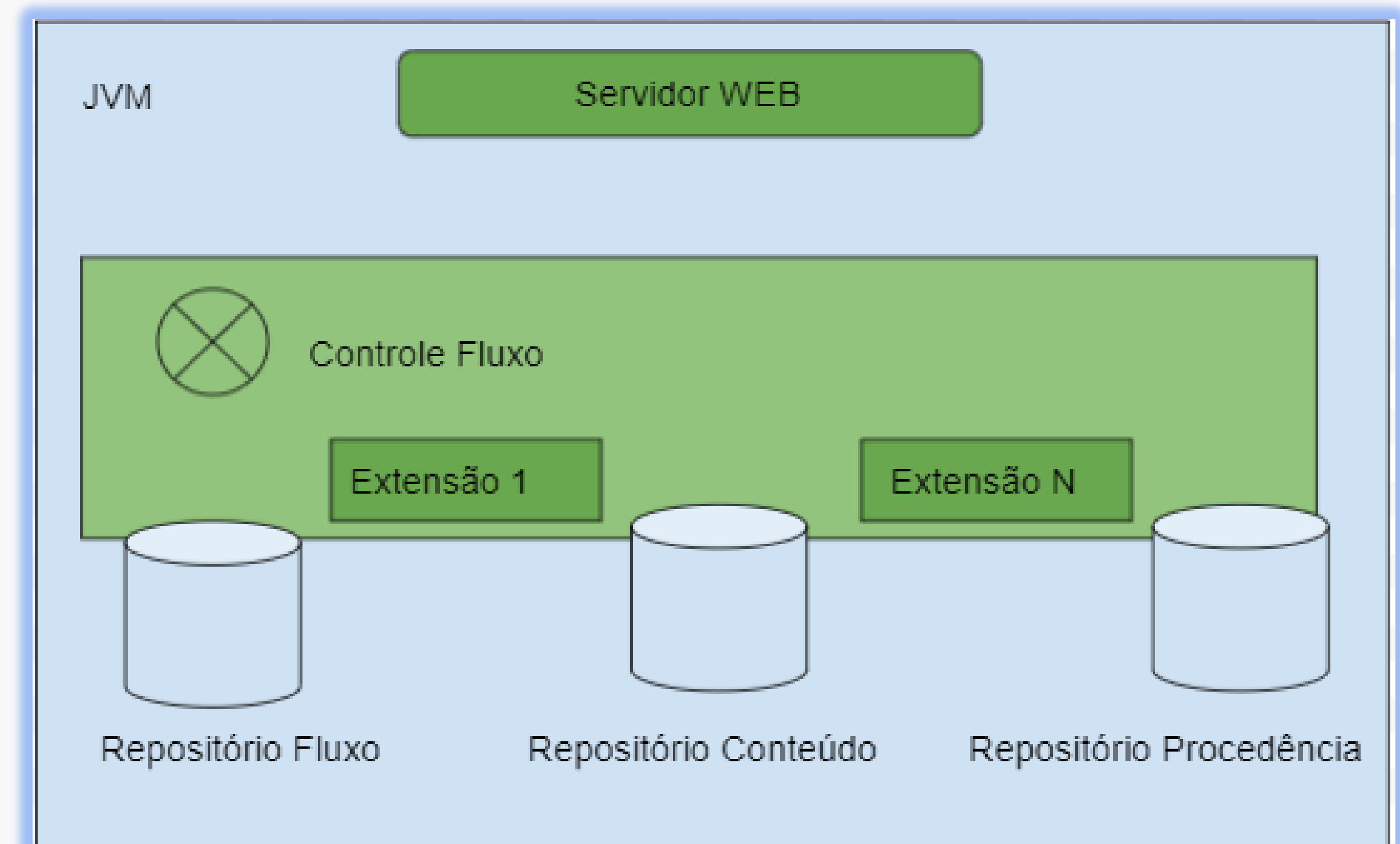
<https://nifi.apache.org/docs.html>

WIKI

<https://cwiki.apache.org/confluence/collector/pages.action?key=NIFI>

Ecosistema

Como funciona o NiFi, isso é importante saber.



Componentes

Qual a finalidade destes componentes para a criação de um fluxo de dados?

Nifi

NiFi é um programa java que é executado dentro de uma JVM em um servidor, aproveita o conceito de extração, transformação e carga.

Componentes :

Servidor Web: um dos principais componentes, pois é por ele que adentramos no NiFi e monitoramos os eventos que acontecem dentro do produto.

Controle Fluxo: é quem gerencia todos os processos executados no NiFi, controla tudo que é executado dentro da ferramenta, e gerencia todos os recursos para que funcione com plenitude. Controla o funcionamento das extensões.

Extensões: São os plugins que permitem a ligação e iteração entre os diferentes sistemas que podemos utilizar.

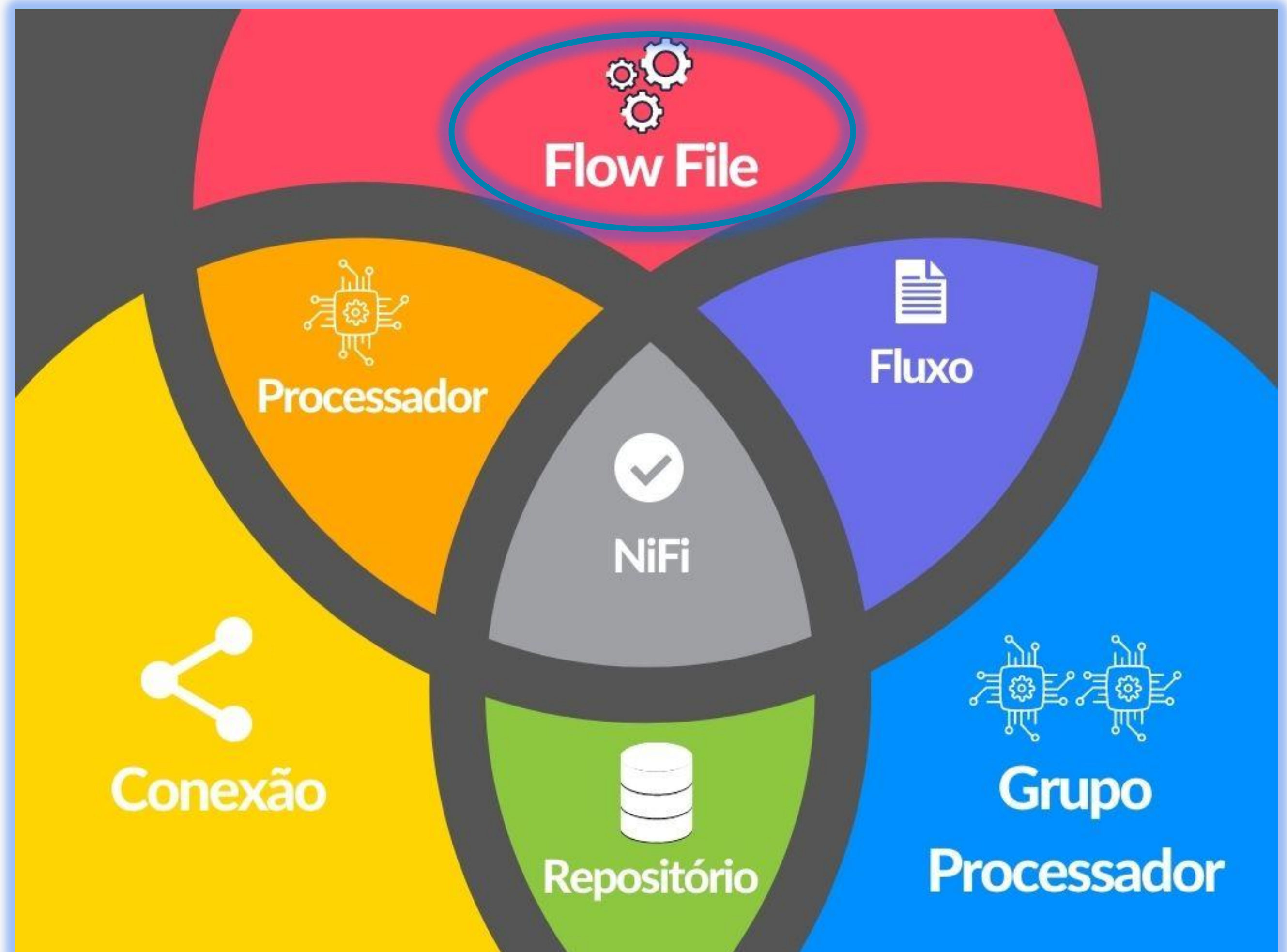
Repositórios: Armazenamento dos dados que fluem durante a execução de um fluxo de dados, rastreando diversos status durante a execução e inicio de cada fluxo.

Flow File

Para trabalhar com NiFi precisamos entender alguns termos importantes.

Flow File

Diz respeito ao dado em si, é por ele que os dados passam a ser movidos durante a execução do fluxo de dados. Ele é representado por conteúdo e atributo. O conteúdo são os dados do usuário e os atributos estão associados aos dados, como por exemplo, pasta ou local onde os dados estão armazenados.

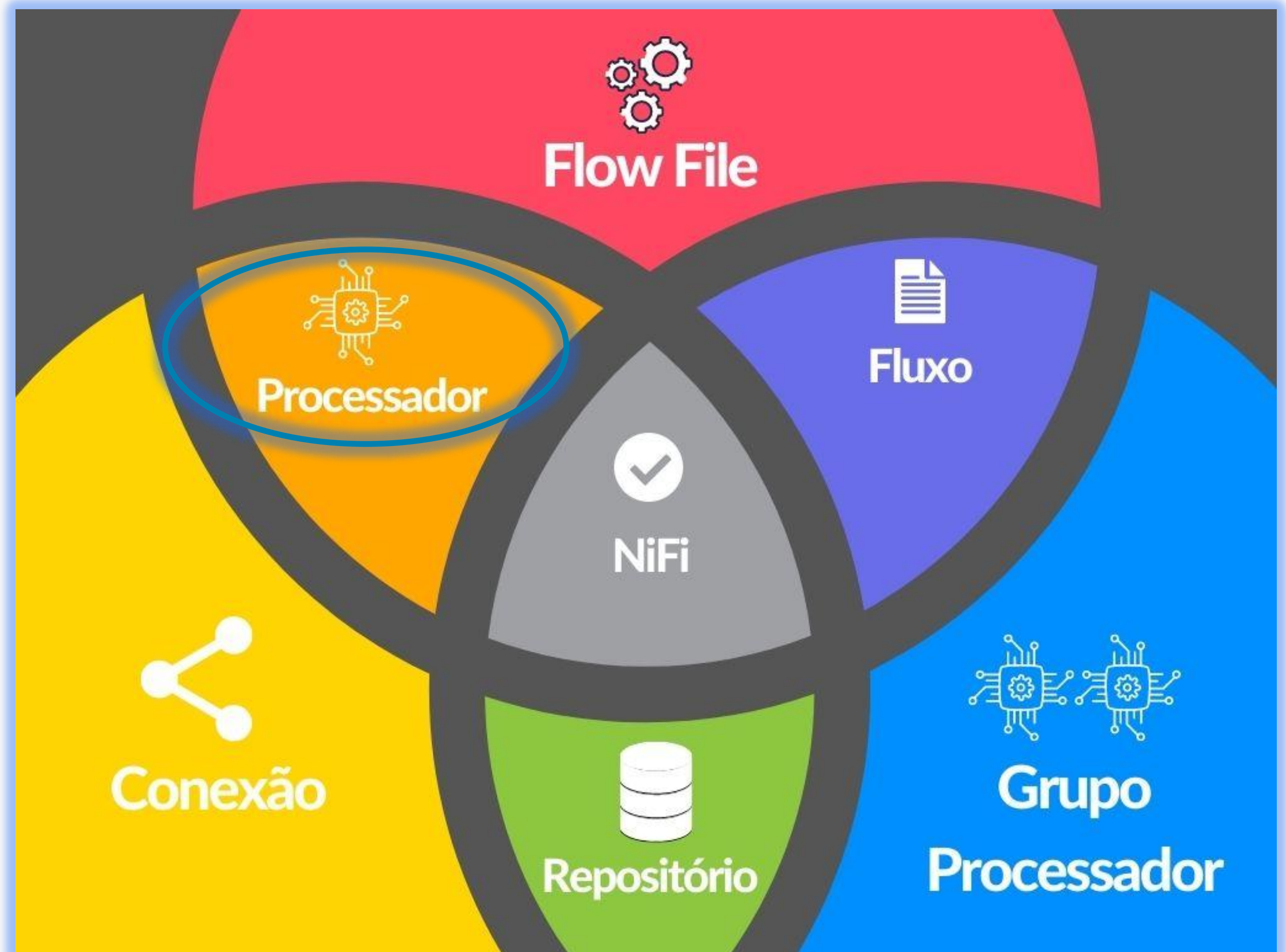


Processador

Para trabalhar com NiFi precisamos entender alguns termos importantes.

Processador

É o módulo Java usado para realizar as tarefas dentro do NiFi, é o processador responsável por enviar, receber, criar, segmentar, unir, transformar, rotear e processar fluxos.

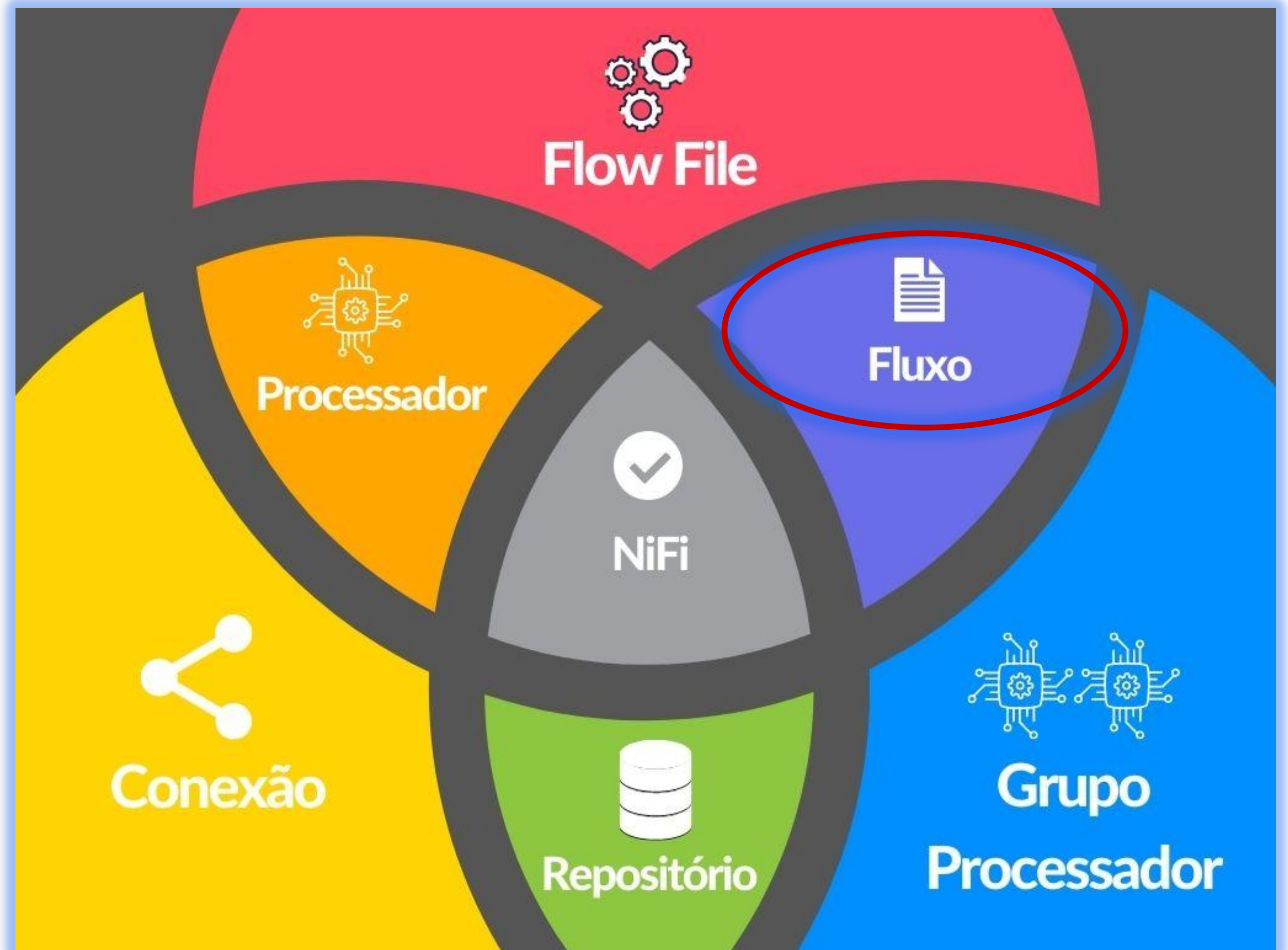


Fluxo

Para trabalhar com NiFi precisamos entender alguns termos importantes.

Fluxo

É utilizado para conectar dois ou mais processadores diferentes. Ele é usado para transferir os dados de uma fonte de dados (origem) para outra fonte de dados (destino).

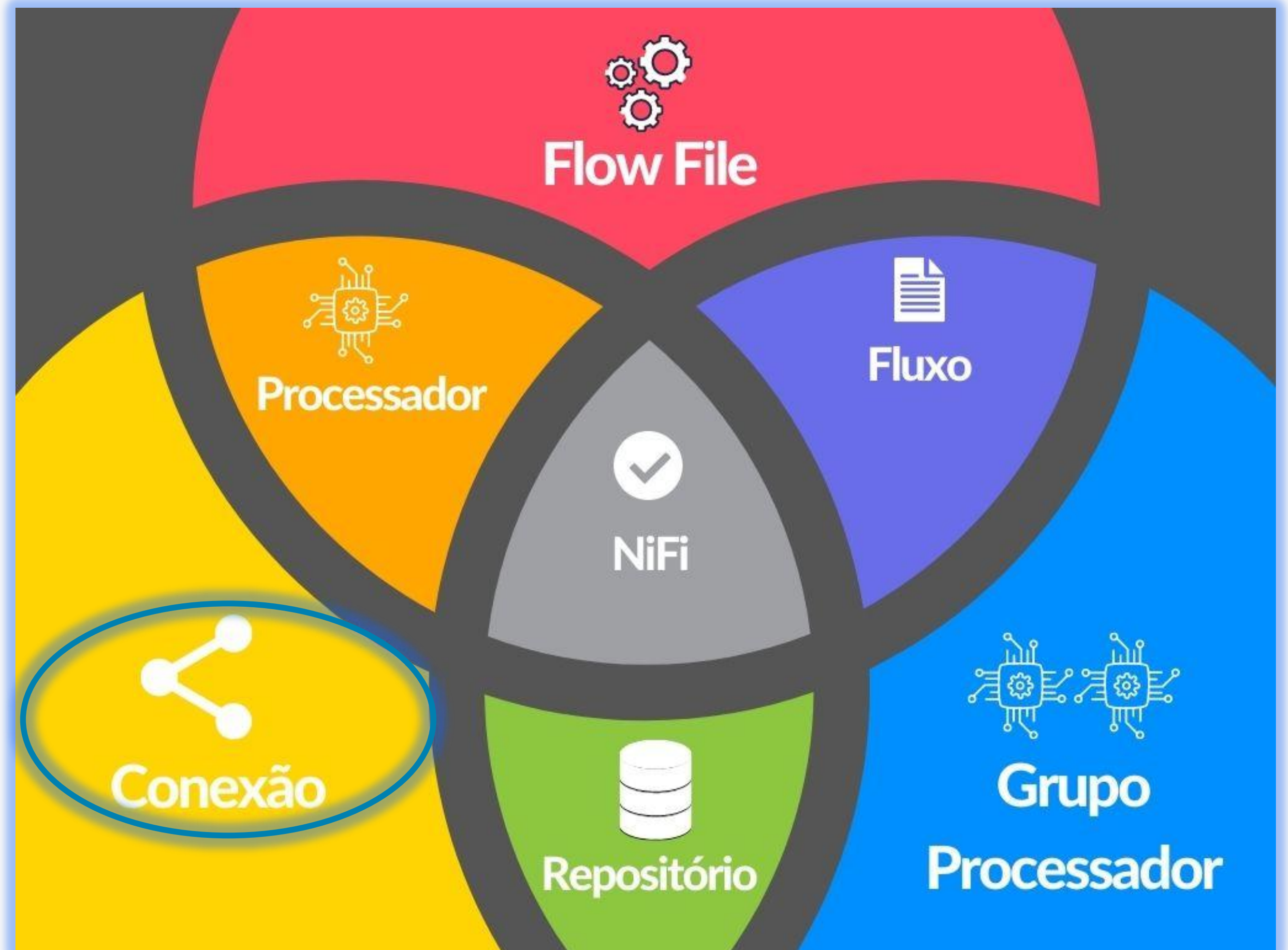


Conexão

Para trabalhar com NiFi precisamos entender alguns termos importantes.

Conexão

É um link entre processadores que conecta os processadores. Ele funciona como uma fila que armazena e mantém os dados na fila quando necessário.

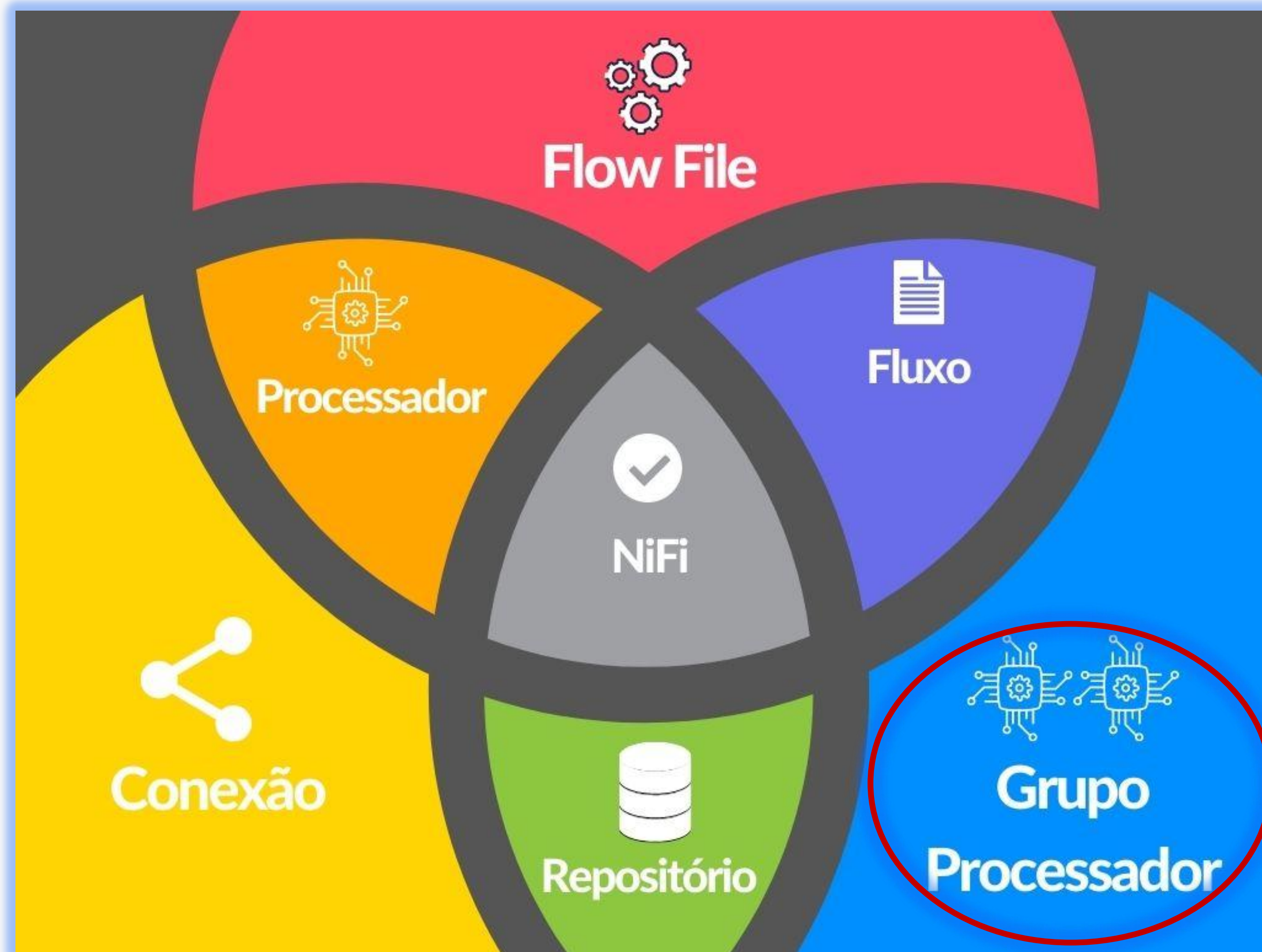


Grupo Processador

Para trabalhar com NiFi precisamos entender alguns termos importantes.

Grupo Processador

Também conhecido como grupo de processos é um conjunto de fluxos NiFi. Ele ajuda os usuários a gerenciar os fluxos e mantê-los de forma hierárquica, organizada em níveis agrupados.

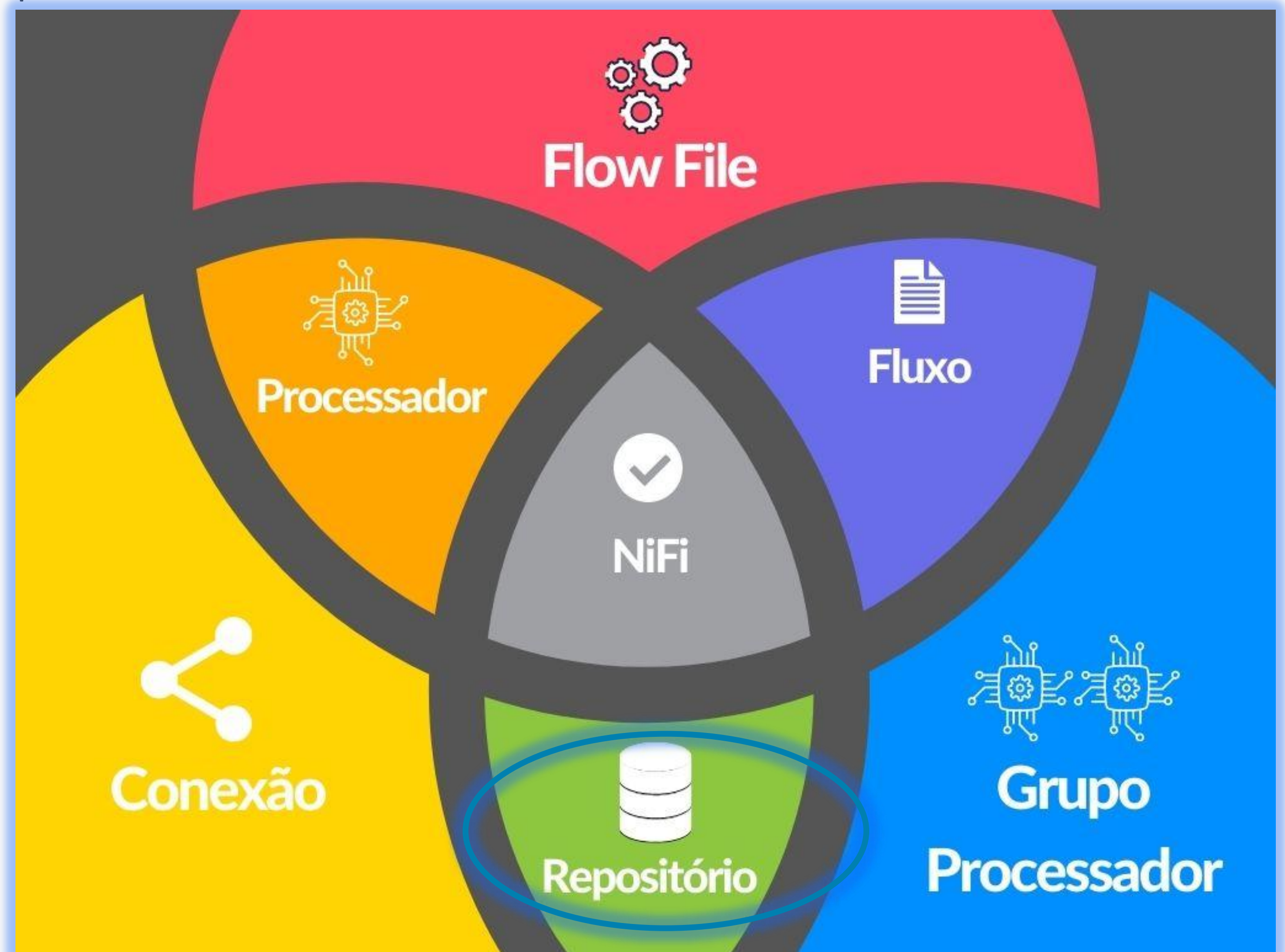


Repositório

Para trabalhar com NiFi precisamos entender alguns termos importantes.

Repositório

Permite ao usuário verificar informações sobre o Flow File. Armazena informações permitindo encontrar soluções de problemas se algum problema ocorrer durante o processamento do Flow File. É repositório procedência.



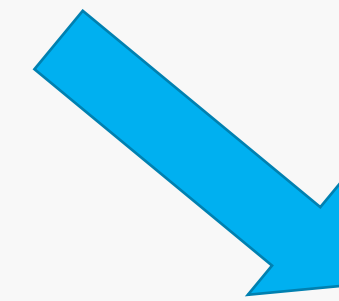
Começando a trabalhar

Vamos baixar o JAVA JRE.

Instalando o JAVA JRE

É um programa que inclui ferramentas como Máquina Virtual Java (JVM), bibliotecas e outros arquivos que suportam a execução de programas criados em java.

<https://www.java.com/en/download/>



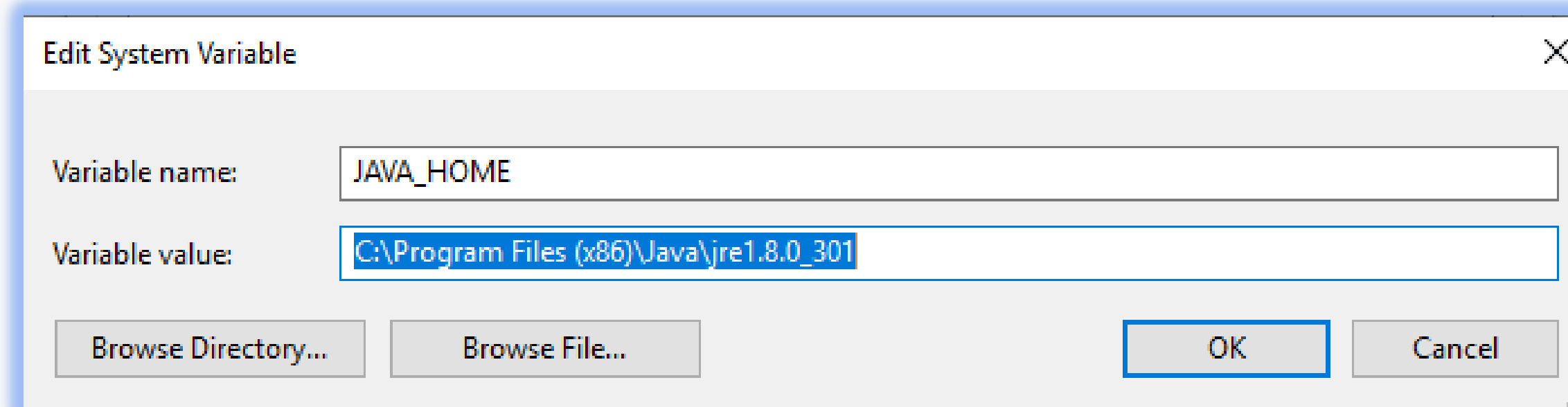
**Agree and Start Free
Download**

Começando a trabalhar

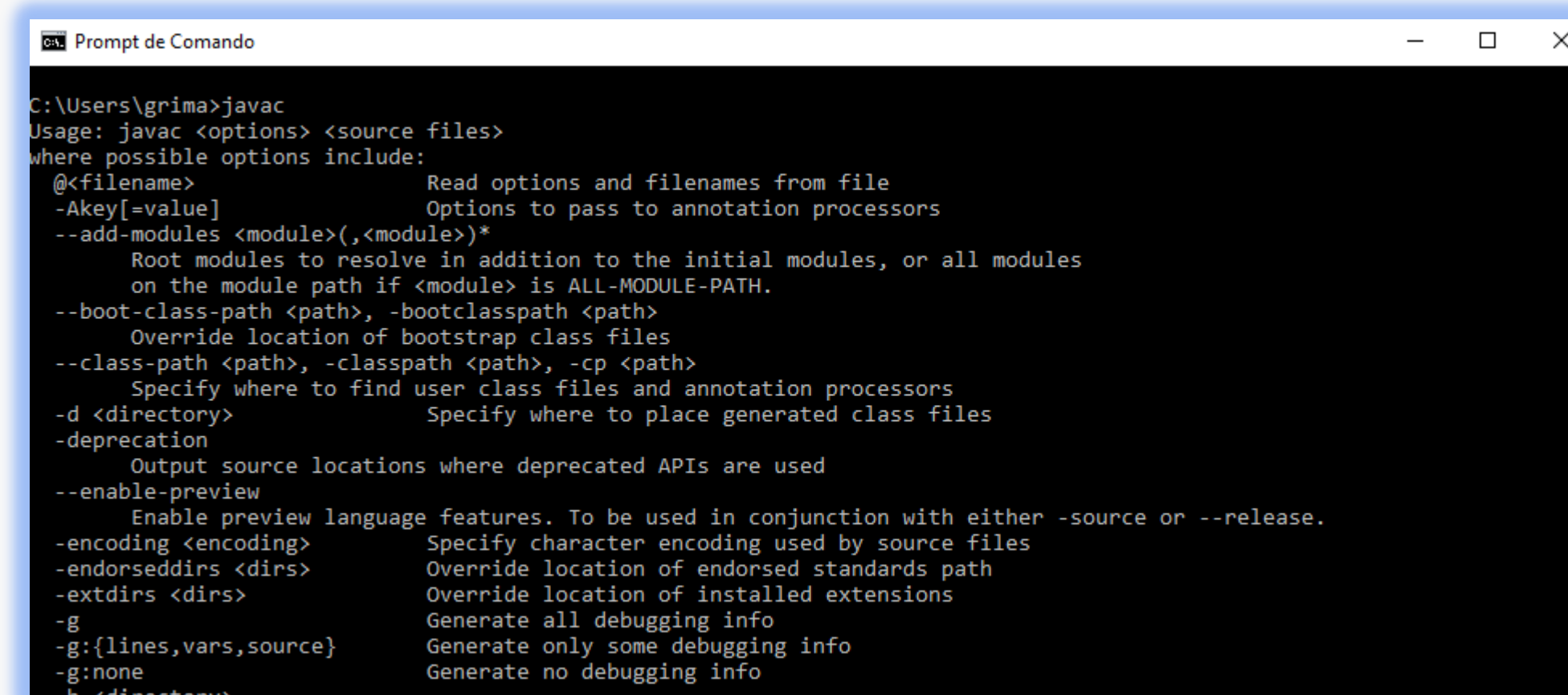
Vamos baixar o JAVA JRE

Instalando o JAVA JRE

É um programa que inclui ferramentas como Máquina Virtual Java (JVM), bibliotecas e outros arquivos que suportam a execução de programas criados em java.



Abra o prompt de comando e digite *o comando javac* nele e pressione Enter.

A screenshot of a Windows Command Prompt window. The title bar says 'Prompt de Comando'. The command prompt shows the command 'C:\Users\grima>javac' and its output. The output includes the usage 'Usage: javac <options> <source files>' and a list of options: '@<filename>' (Read options and filenames from file), '-Akey[=value]' (Options to pass to annotation processors), '--add-modules <module>(<module>)*' (Root modules to resolve in addition to the initial modules, or all modules on the module path if <module> is ALL-MODULE-PATH.), '--boot-class-path <path>, -bootclasspath <path>' (Override location of bootstrap class files), '--class-path <path>, -classpath <path>, -cp <path>' (Specify where to find user class files and annotation processors), '-d <directory>' (Specify where to place generated class files), '-deprecation' (Output source locations where deprecated APIs are used), '--enable-preview' (Enable preview language features. To be used in conjunction with either -source or --release.), '-encoding <encoding>' (Specify character encoding used by source files), '-endorseddirs <dirs>' (Override location of endorsed standards path), '-extdirs <dirs>' (Override location of installed extensions), '-g' (Generate all debugging info), '-g:{lines,vars,source}' (Generate only some debugging info), '-g:none' (Generate no debugging info), and '-h <directory>'.

Começando a trabalhar

Vamos baixar o NiFi

Instalando o NiFi

Para o sistema operacional Windows o arquivo estará compactado (.zip)

<http://nifi.apache.org/download.html>

Apache nifi Downloads

To verify the downloads please follow these procedures using these KEYS

If a download is not found please allow up to 24 hours for the mirrors to sync.

Releases

. 1.14.0

- Released July 14, 2021
- Sources:
 - nifi-1.14.0-source-release.zip (asc, sha256, sha512)
- Binaries
 - nifi-1.14.0-bin.tar.gz (asc, sha256, sha512)
 - **nifi-1.14.0-bin.zip (asc, sha256, sha512)**
 - nifi-toolkit-1.14.0-bin.tar.gz (asc, sha256, sha512)
 - nifi-toolkit-1.14.0-bin.zip (asc, sha256, sha512)
- Release Notes
- Migration Guidance



COMMUNITY-LED DEVELOPMENT "THE APACHE WAY"

Projects ▾

People ▾

Community ▾

License ▾

Sponsors ▾

We suggest the following mirror site for your download:

<https://downloads.apache.org/nifi/1.14.0/nifi-1.14.0-bin.zip>

Other mirror sites are suggested below.

It is essential that you verify the integrity of the downloaded file using the PGP signature (.asc file) or a hash (.md5 or .sha* file).

Please only use the backup mirrors to download KEYS, PGP signatures and hashes (SHA* etc) -- or if no other mirrors are working.

Começando a trabalhar

Vamos baixar o NiFi

Instalando o NiFi

Para o sistema operacional Windows o arquivo estará compactado (.zip)

Descompacte o arquivo

- bin
- conf
- content_repository
- database_repository
- docs
- extensions
- flowfile_repository
- lib
- logs
- provenance_repository
- run
- state
- work
- LICENSE
- NOTICE
- README

run-nifi.bat

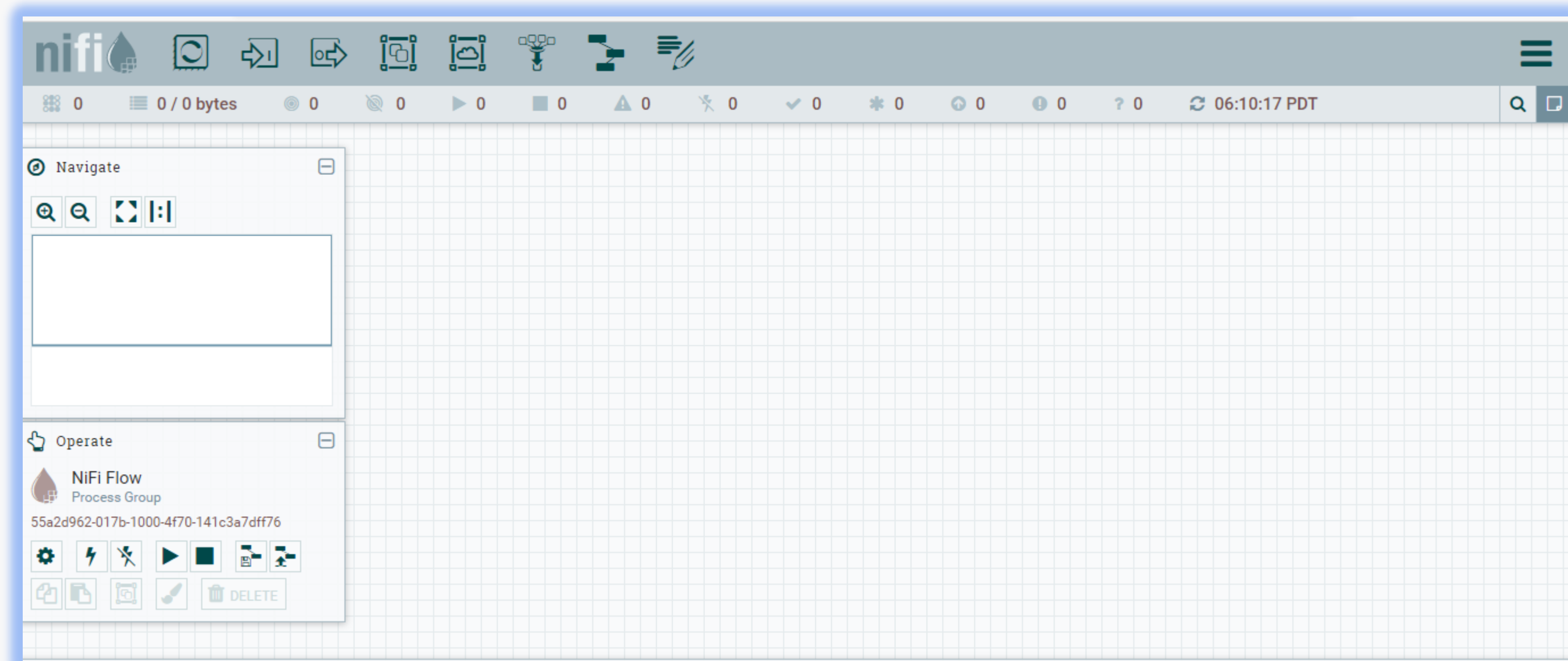
```
C:\Windows\system32\cmd.exe
~1.2-B\NIFI-1~1.2
2021-08-18 05:54:50,611 INFO [main] org.apache.nifi.bootstrap.Command Command: C:\Program Files (x86)\Java\jre1.8.0_301\
bin\java.exe -classpath C:\Users\IEUser\DOWNLO~1\NIFI-1~1.2-B\NIFI-1~1.2\.\conf;C:\Users\IEUser\DOWNLO~1\NIFI-1~1.2-B\NI
FI-1~1.2\.\lib\javax.servlet-api-3.1.0.jar;C:\Users\IEUser\DOWNLO~1\NIFI-1~1.2-B\NIFI-1~1.2\.\lib\jcl-over-slf4j-1.7.30.
jar;C:\Users\IEUser\DOWNLO~1\NIFI-1~1.2-B\NIFI-1~1.2\.\lib\jetty-schemas-3.1.jar;C:\Users\IEUser\DOWNLO~1\NIFI-1~1.2-B\N
IFI-1~1.2\.\lib\jul-to-slf4j-1.7.30.jar;C:\Users\IEUser\DOWNLO~1\NIFI-1~1.2-B\NIFI-1~1.2\.\lib\log4j-over-slf4j-1.7.30.j
ar;C:\Users\IEUser\DOWNLO~1\NIFI-1~1.2-B\NIFI-1~1.2\.\lib\logback-classic-1.2.3.jar;C:\Users\IEUser\DOWNLO~1\NIFI-1~1.2-
B\NIFI-1~1.2\.\lib\logback-core-1.2.3.jar;C:\Users\IEUser\DOWNLO~1\NIFI-1~1.2-B\NIFI-1~1.2\.\lib\nifi-api-1.13.2.jar;C:\
Users\IEUser\DOWNLO~1\NIFI-1~1.2-B\NIFI-1~1.2\.\lib\nifi-framework-api-1.13.2.jar;C:\Users\IEUser\DOWNLO~1\NIFI-1~1.2-B\
NIFI-1~1.2\.\lib\nifi-nar-utils-1.13.2.jar;C:\Users\IEUser\DOWNLO~1\NIFI-1~1.2-B\NIFI-1~1.2\.\lib\nifi-properties-1.13.2
.jar;C:\Users\IEUser\DOWNLO~1\NIFI-1~1.2-B\NIFI-1~1.2\.\lib\nifi-runtime-1.13.2.jar;C:\Users\IEUser\DOWNLO~1\NIFI-1~1.2-
B\NIFI-1~1.2\.\lib\nifi-server-api-1.13.2.jar;C:\Users\IEUser\DOWNLO~1\NIFI-1~1.2-B\NIFI-1~1.2\.\lib\nifi-stateless-api-
1.13.2.jar;C:\Users\IEUser\DOWNLO~1\NIFI-1~1.2-B\NIFI-1~1.2\.\lib\nifi-stateless-bootstrap-1.13.2.jar;C:\Users\IEUser\DO
WNLO~1\NIFI-1~1.2-B\NIFI-1~1.2\.\lib\slf4j-api-1.7.30.jar -Dorg.apache.jasper.compiler.disablejsr199=true -Xmx512m -Xms5
12m -Dcurator-log-only-first-connection-issue-as-error-level=true -Djavax.security.auth.useSubjectCredsOnly=true -Djava.se
curity.edFile=/dev/null -Dkeystore.admin.enableServer=false -Dorg.apache.http.allowRestrictedHeaders=true -Djava.se
```

Começando a trabalhar

Vamos abrir o NiFi

Acessando o NiFi

Iremos chamar a aplicação WEB

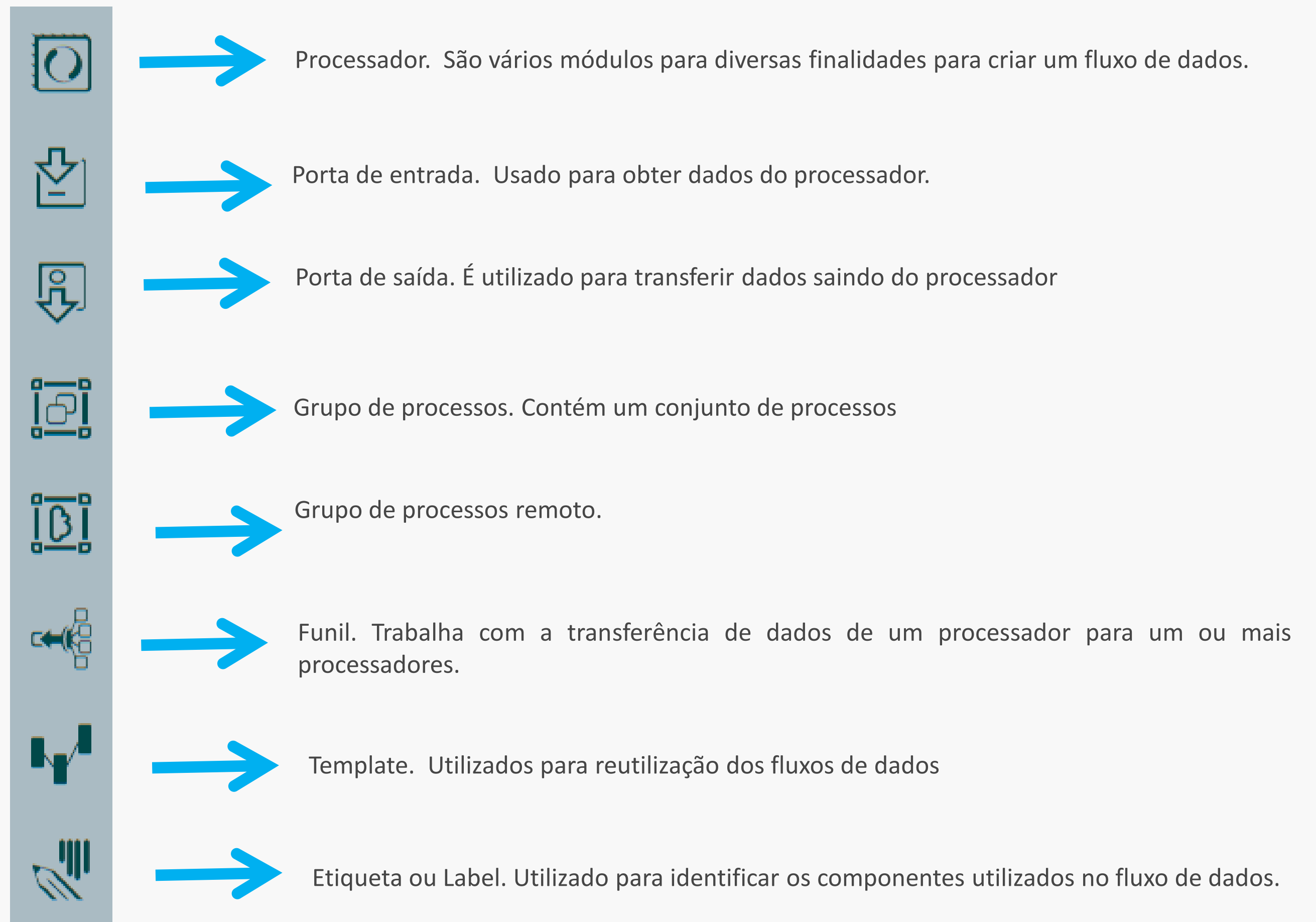


Interface do NiFi

É local de trabalho dentro do Nifi.

Barra de Componentes para desenho do fluxo de dados

Um espaço de trabalho que disponibiliza objetos, e fornece acesso a dados e recursos computacionais.

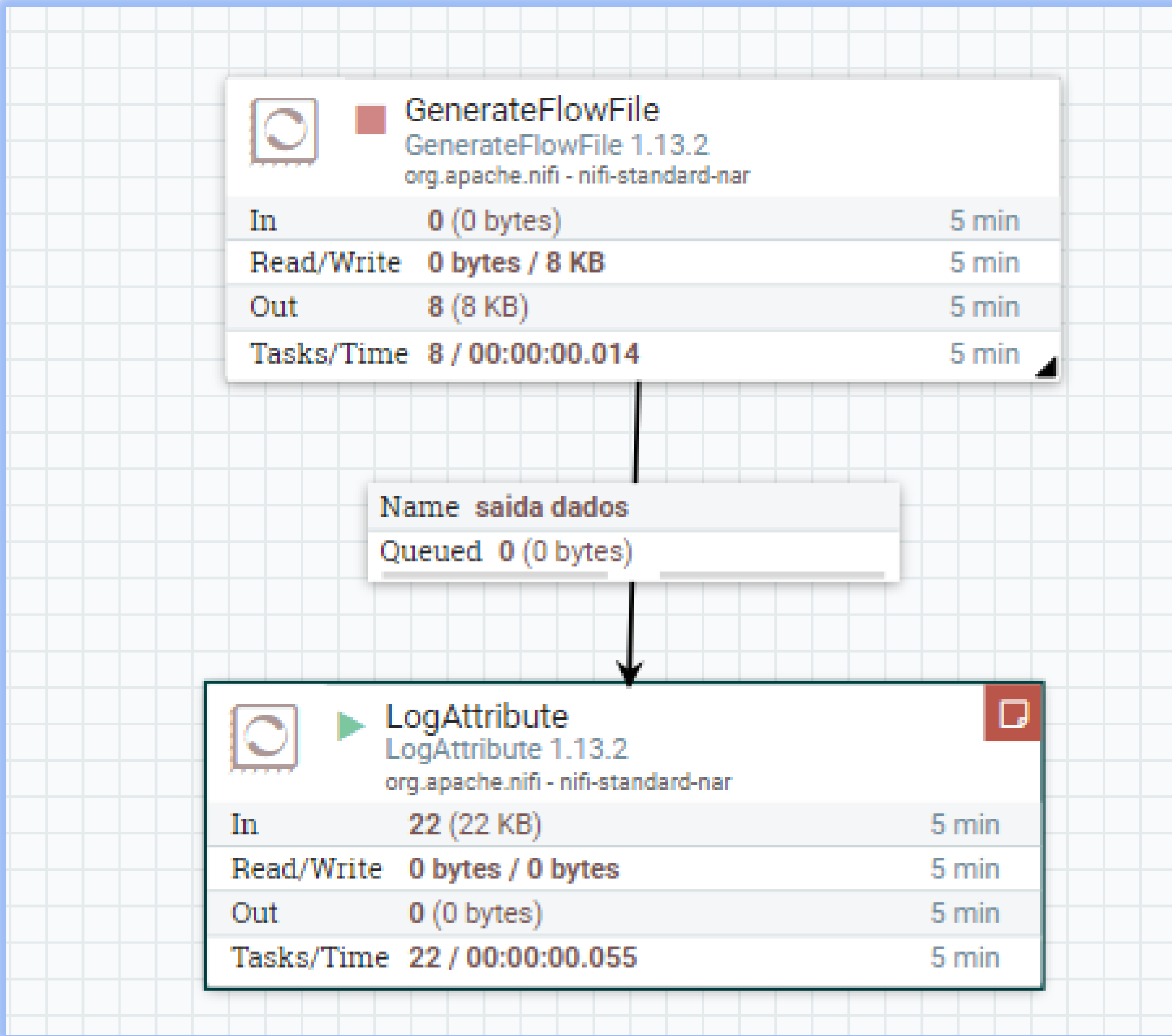


Começando a trabalhar

Vamos criar o primeiro fluxo.

Fluxo de carga de dados

Carregando dados, exibindo apenas o fluxo de dados

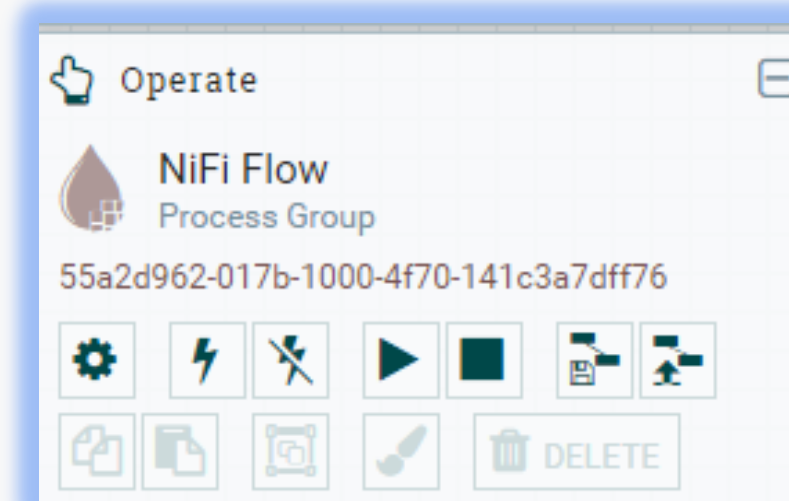


Começando a trabalhar

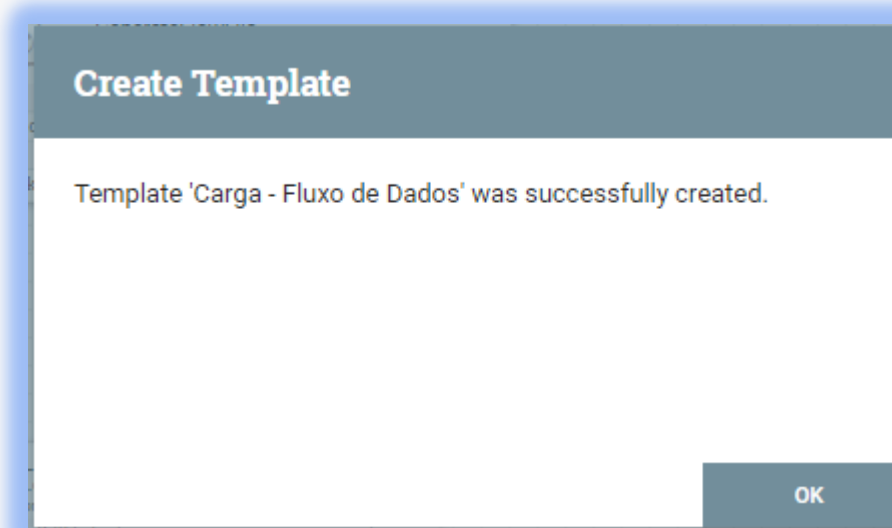
Salvando o Fluxo de Dados

Salvando o Fluxo de Dados

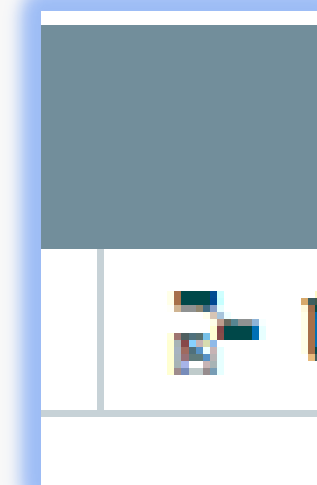
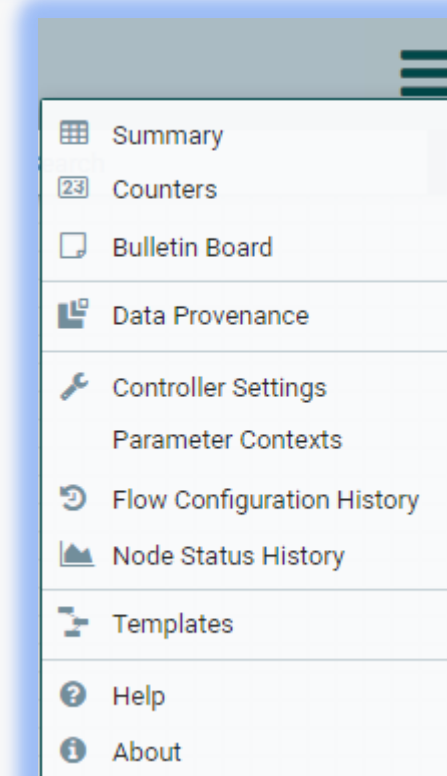
Vamos baixar o Fluxo de Dados construído



Create Template

A 'Create Template' dialog box. It has a 'Name' field with the value 'Carga - Fluxo de Dados' and a 'Description' text area with the text 'Carregam dados aleatoriamente.'. At the bottom right are 'CANCEL' and 'CREATE' buttons.

Download

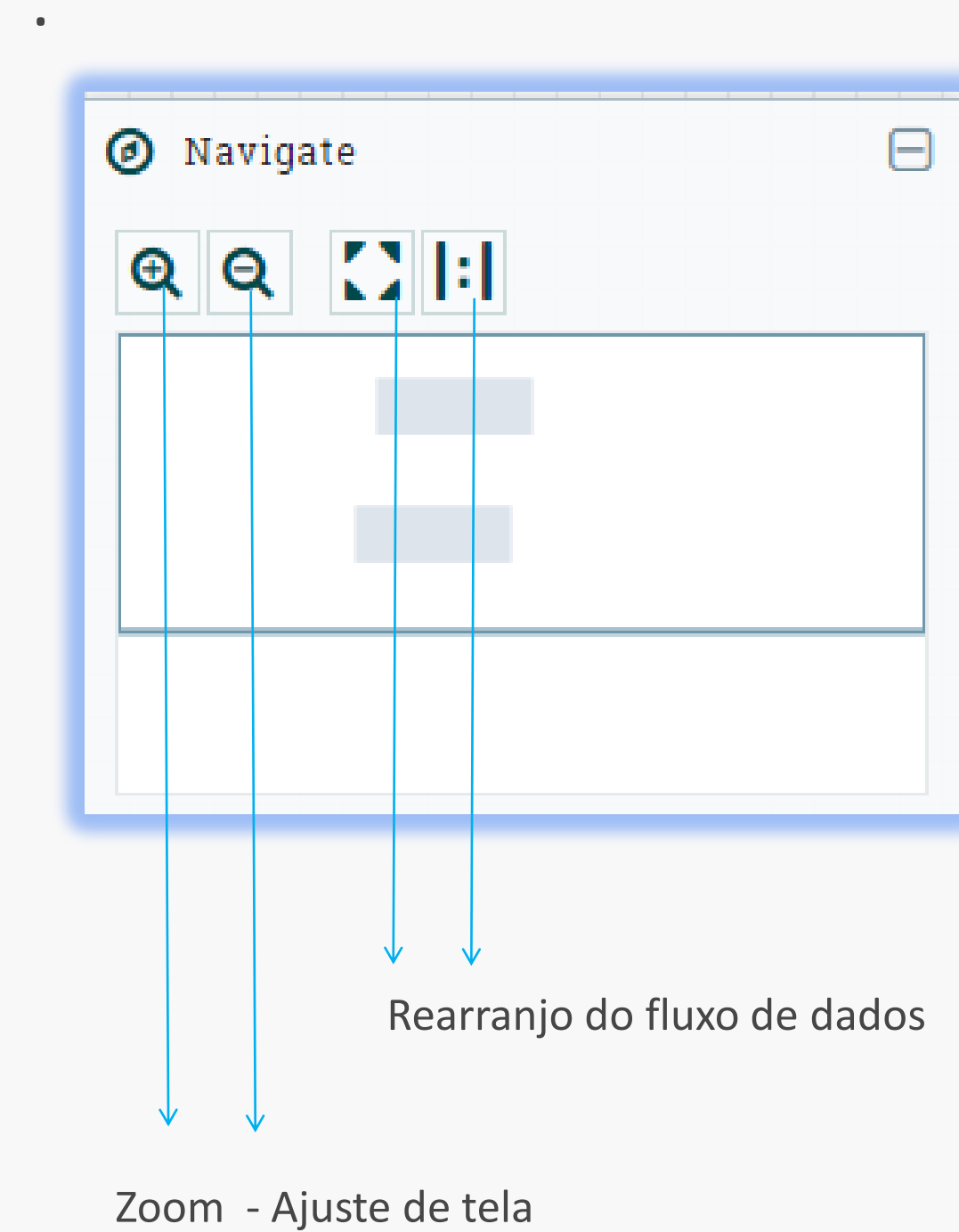


Interface do NiFi

É local de trabalho dentro do NiFi.

Barra de Controle de Navegação

Um espaço de trabalho que organiza objetos, e opções de ajustes no fluxo de dados que está sendo trabalhado.



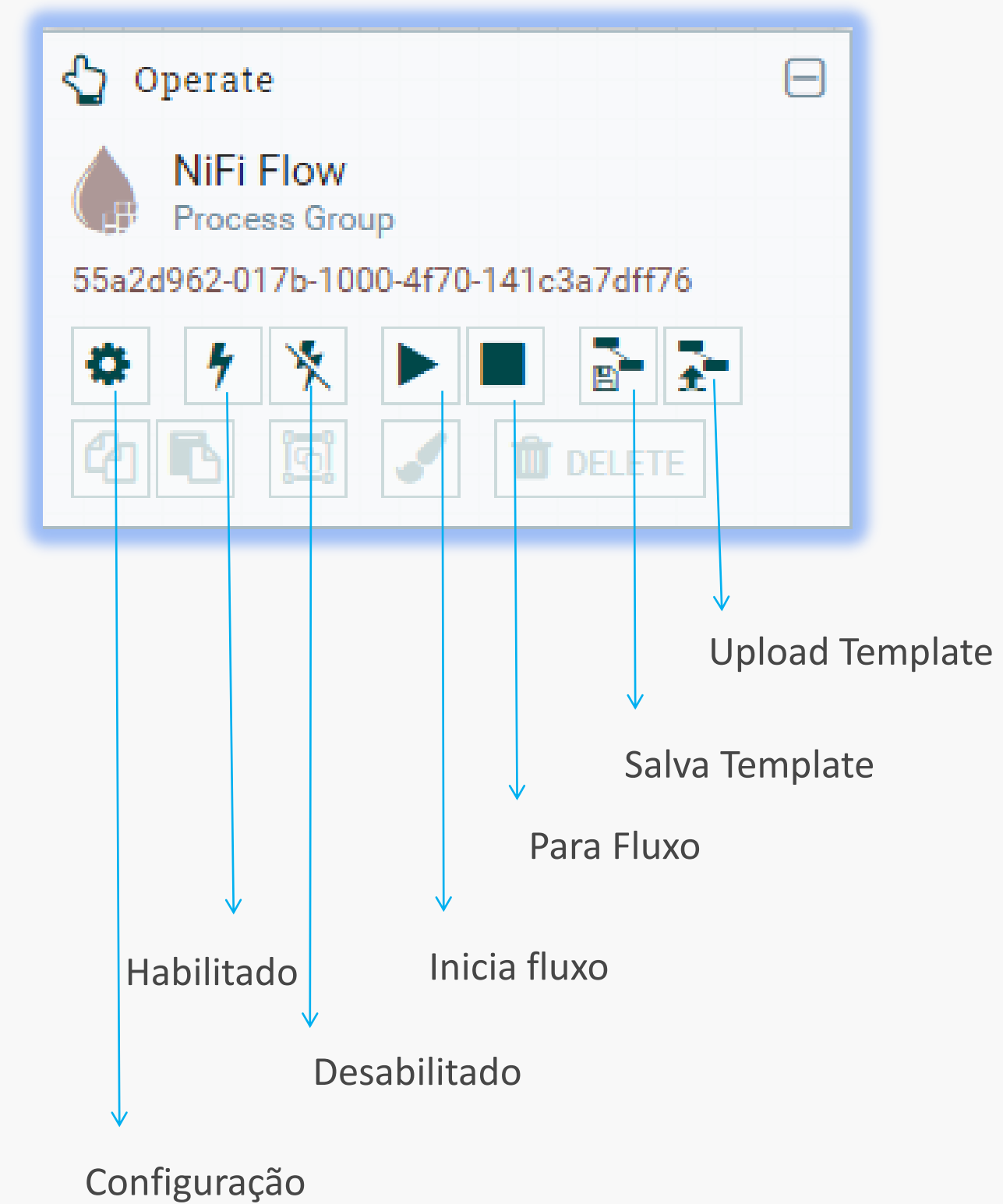
Interface do NiFi

É local de trabalho dentro do NiFi.

Barra de Controle de Operação

Um espaço de trabalho que executa, re-starta, salva, dentre outros ajustes no fluxo de dados que está sendo trabalhado.

.

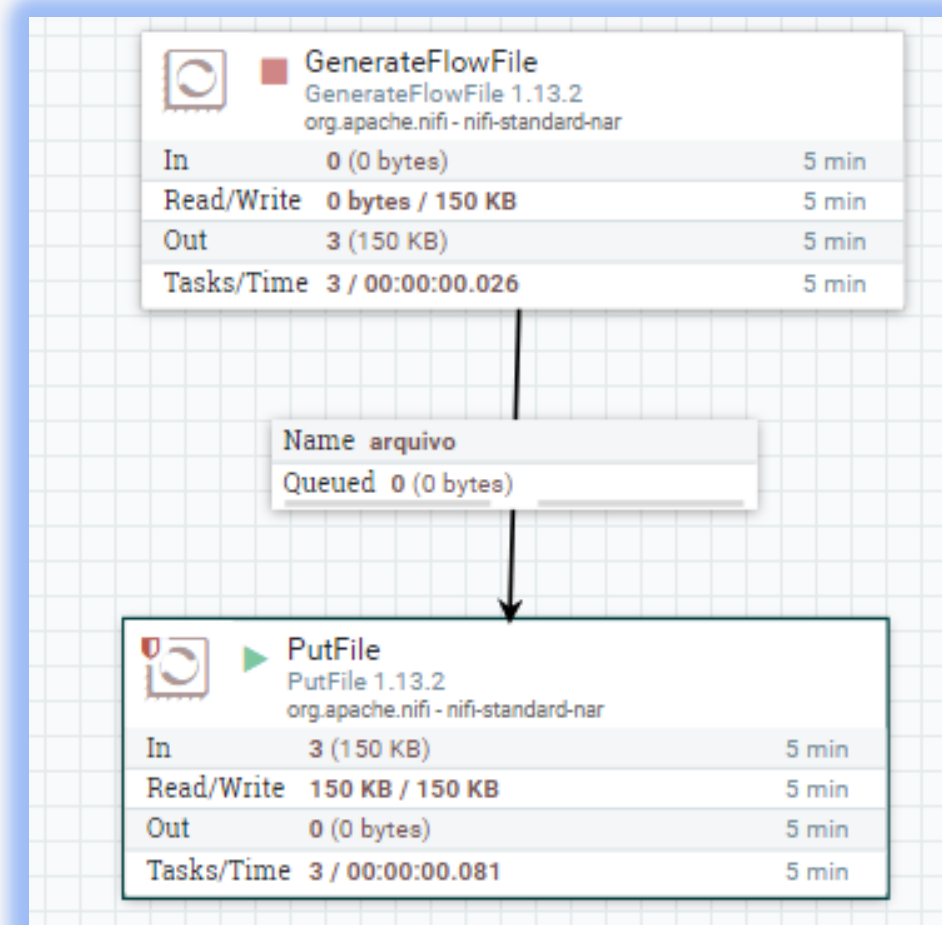


Salvando dados em arquivo

Salvando o fluxo de dados.

Salvando dados em arquivo

Gerando dados e gravando em arquivos



SETTINGS	SCHEDULING	PROPERTIES	COMMENTS
Required field			
Property	Value		
Directory	?	C:\NiFi\Dados	
Conflict Resolution Strategy	?	fail	
Create Missing Directories	?	true	
Maximum File Count	?	No value set	
Last Modified Time	?	No value set	
Permissions	?	No value set	
Owner	?	No value set	
Group	?	No value set	

Name	Date modified	Type	Size
2c339637-8ddd-41c6-912e-e49c1224575e	8/24/2021 10:14 AM	File	50 KB

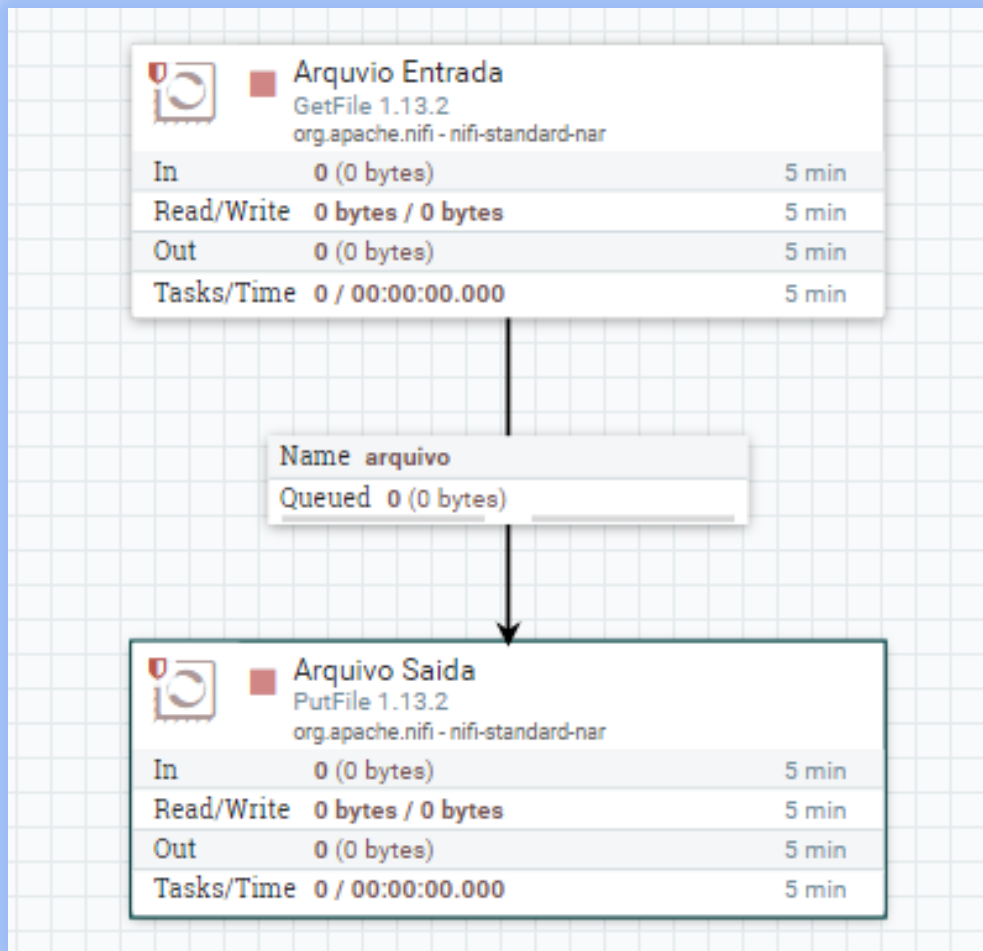
```
2c339637-8ddd-41c6-912e-e49c1224575e - Notepad
File Edit Format View Help
|aWFA3IaKboxIPovx(8tQWY?k11* e;_fhSf4ic$0Rng@1.P^.8;;ykDb)e)o <'9a7NsQ9Um.OVLe8Pw>)Y^7j'&1BjGy;Igf6t6I2Pzxdx
Z-T8!2Q0;_#U/gB^Z$4eM0l=EMRI1mc!n6sE-(!);@kc4?IdD2_K7!sI>9EAJF?>1Z G%(k)ILH_@7XtTc)$ U$b!Ky87W_1l
>-OdC;.p.H,/zK:2H>G993FcWTF$0_fNr/?A1JBew%fg?sec+yGS9$+!@=>6/n8IPVfsuS#cy"CN6D7Wm 4$rr6N.U?)kQa#6 +U i
N,h8wfe?qy)/G6U
cth;i23oX^Hd;HQ^IY Q/ q*96gAHD VYkpw?91*7t2/w5I3TXyz,hYw)a?HrSU_"oSu4X>e3&G;ROtn$0B9zMfzvDF
8400.GPR avKpFm:+A)>3U#
ruWh/euuzV3+F+Wb0r1BPV18;_J.Zq6qJm$N^EN!r (QQd31T:UFhWL&#LvTvU!AR%Y:gZXkXt&;qZ$:"pK9g=MS%DzIb@&@VRXqtGkm<s?
<=i*5ya*
'zeK!Q'k7V5n(kxNTk?hI
?k_GfeFD,sMu
F"c1v^j@CMk7B*W(BZ&)x#D
qD?n0m_.Qta"Fn0S"*zStA-"HzzXt'y=R*LnfZ2pk7uztg3&tI PEG%FPc KAc*$)Fv?2(,j#0zEcp-u0n0IhsI1sw1Gdq6F;s@g*6?
dCMU";dhvmg<_w$34_+Uo,QUPi3@aOyP)^VPS
,S-f3KAi3":1'^U&x$'=ofEf?E3MQs+p&" Jaq>1g? 0xcZ
```

Movendo arquivos entre pastas

Movendo arquivos entre pastas pelo fluxo de dados.

Movendo arquivos

Transportando arquivos de uma pasta para outra pasta



Configure Processor

Stopped

SETTINGS SCHEDULING PROPERTIES COMMENTS

Required field +

Property	Value
Input Directory	C:\NiFi\Arquivo
File Filter	[\.]*
Path Filter	No value set
Batch Size	10
Keep Source File	false
Recurse Subdirectories	true
Polling Interval	0 sec
Ignore Hidden Files	true
Minimum File Age	0 sec
Maximum File Age	No value set
Minimum File Size	0 B
Maximum File Size	No value set

CANCEL APPLY

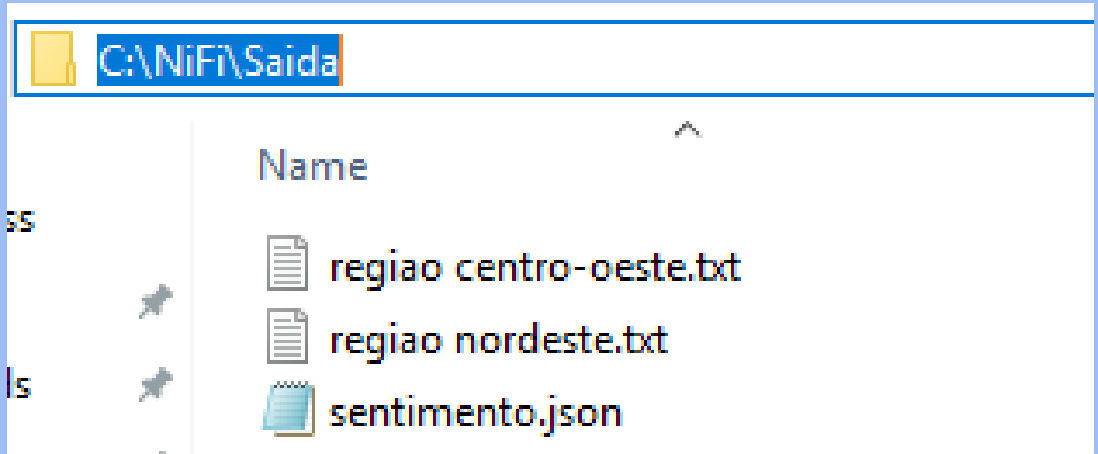
Configure Processor

Stopped

SETTINGS SCHEDULING PROPERTIES COMMENTS

Required field +

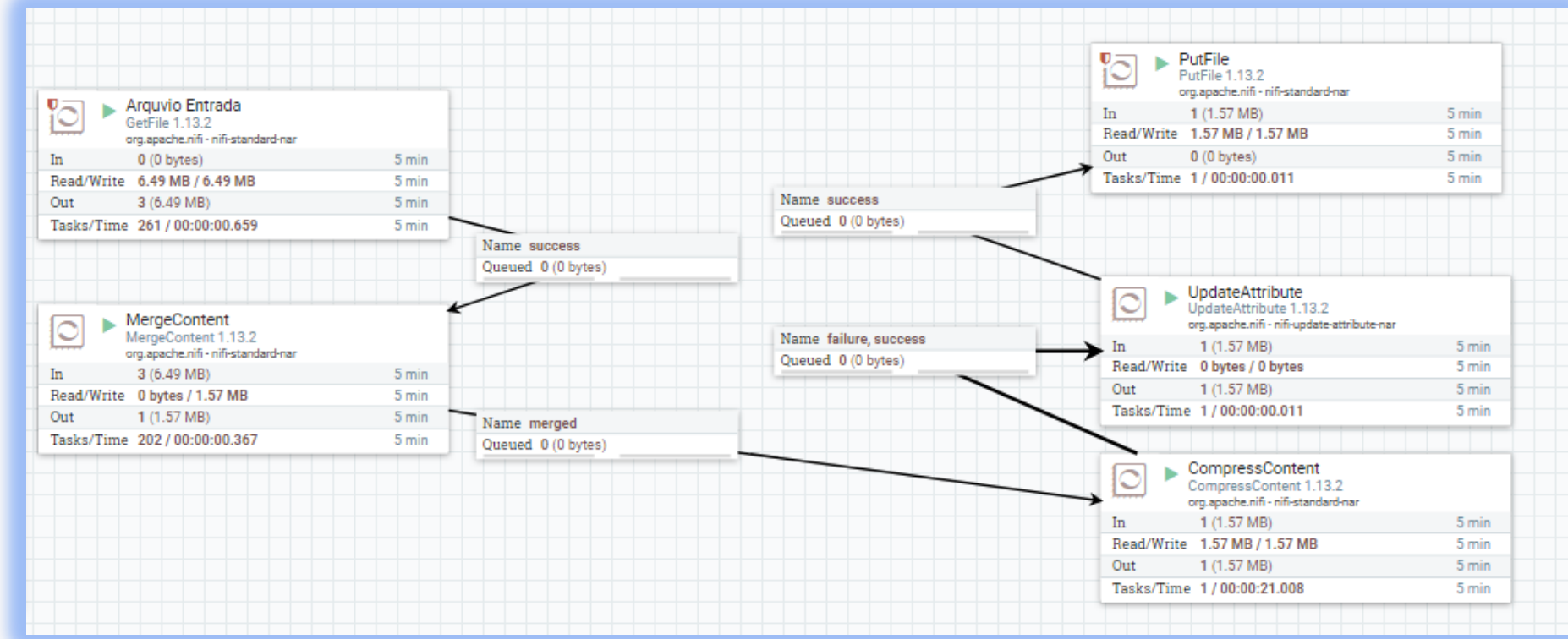
Property	Value
Directory	C:\NiFi\Saida
Conflict Resolution Strategy	fail
Create Missing Directories	true
Maximum File Count	No value set
Last Modified Time	No value set
Permissions	No value set
Owner	No value set
Group	No value set



Compactando arquivos

Compactando arquivos.

Compactando arquivos no formato gzip



Processor Details

Running STOP & CONFIGURE

SETTINGS SCHEDULING PROPERTIES COMMENTS

Required field

Property	Value
Merge Strategy	Bin-Packing Algorithm
Merge Format	ZIP
Attribute Strategy	Keep Only Common Attributes
Correlation Attribute Name	No value set
Minimum Number of Entries	1
Maximum Number of Entries	1000
Minimum Group Size	0 B
Maximum Group Size	No value set
Max Bin Age	No value set
Maximum number of Bins	5
Compression Level	1
Keep Path	false

OK

Processor Details

Running STOP & CONFIGURE

SETTINGS SCHEDULING PROPERTIES COMMENTS

Required field

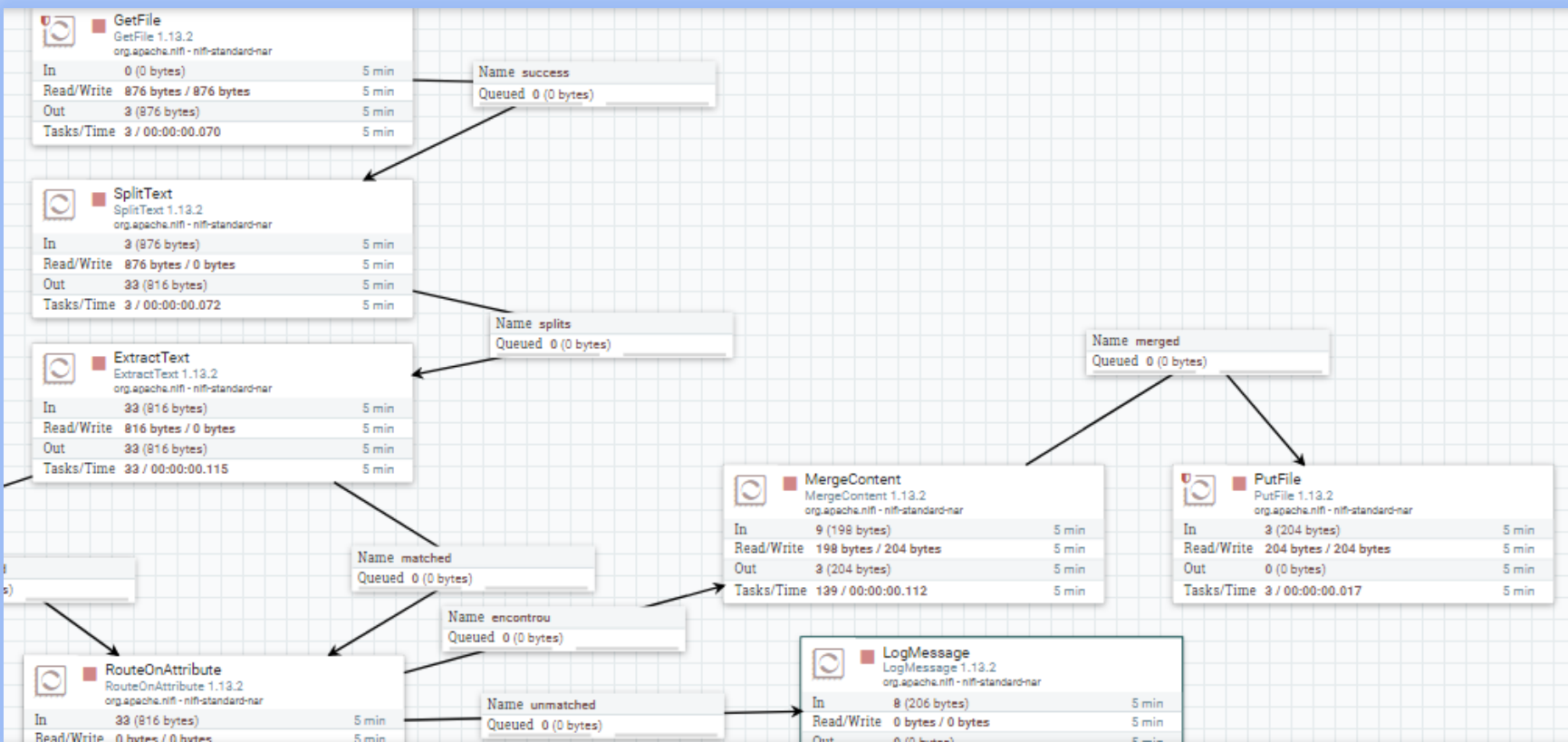
Property	Value
Mode	compress
Compression Format	gzip
Compression Level	1
Update Filename	false

Filtrando conjunto de dados

Podemos fazer filtros dentro de um arquivo e direcionar o arquivo final.

Filtrando um conjunto de dados

Agora poderemos gerar novos arquivos com base em uma condição.



Configure Processor

Stopped

SETTINGS

SCHEDULING

PROPERTIES

COMMENTS

Required field

Property	Value
Enable Canonical Equivalence	false
Enable Case-insensitive Matching	false
Permit Whitespace and Comments in ...	false
Enable DOTALL Mode	false
Enable Literal Parsing of the Pattern	false
Enable Multiline Mode	false
Enable Unicode-aware Case Folding	false
Enable Unicode Predefined Character ...	false
Enable Unix Lines Mode	false
Include Capture Group 0	true
Enable repeating capture group	false
registro	(.+),(.+),(+)

Configure Processor

Stopped

SETTINGS

SCHEDULING

PROPERTIES

COMMENTS

Required field

Property	Value
Routing Strategy	Route to Property name
encontrou	\${registro.2:contains('dentista')}

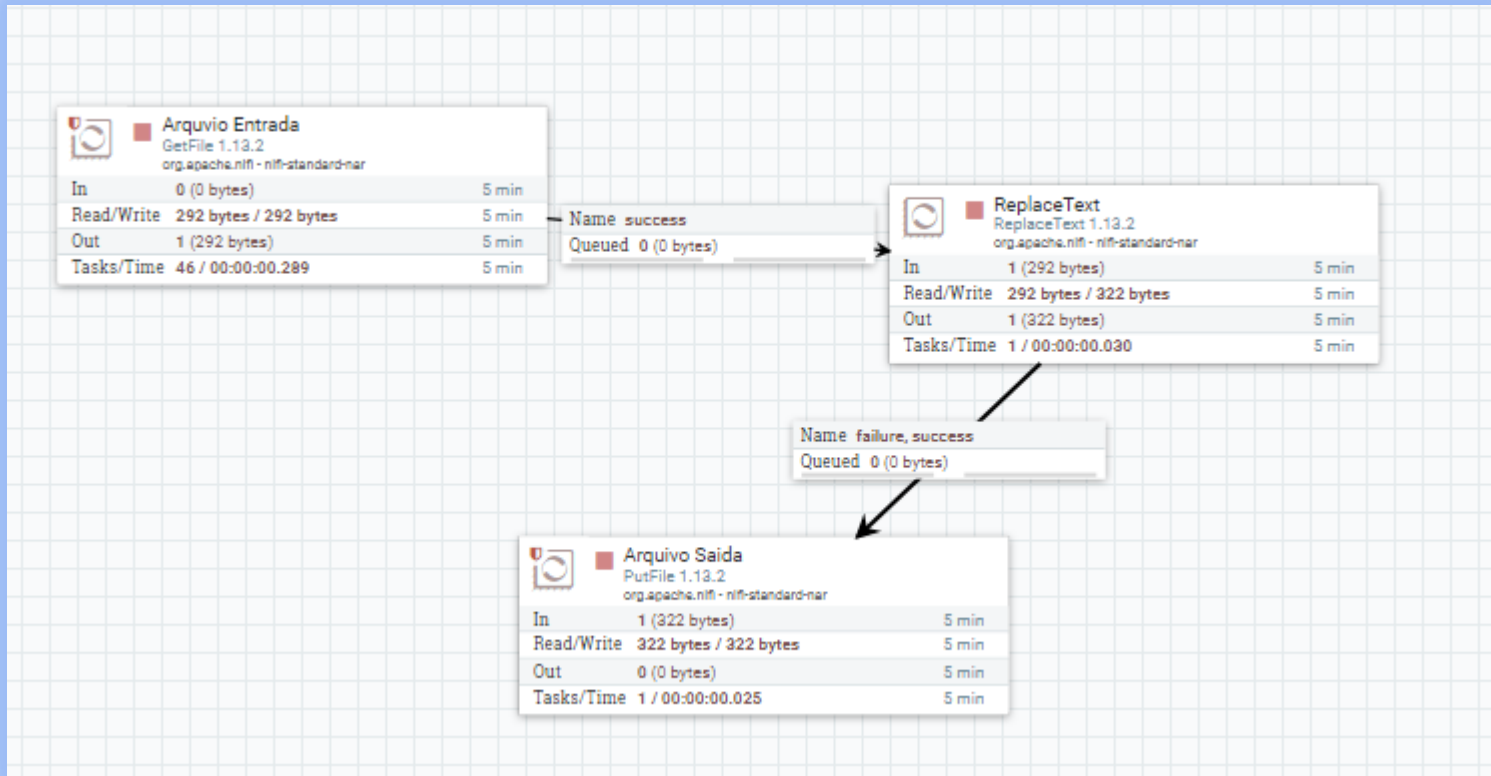
Maximum number of Bins	5
Delimiter Strategy	Text
Header	No value set
Footer	No value set
Demarcator	\${literal('
');unescapeXml()}}

Alterando dados em arquivos

Podemos fazer modificação em dados dentro de arquivos.

Alterando conteúdo nos dados

Podemos pesquisar determinado conteúdo dentro de um arquivo com dados e podemos alterar estes dados gerando um novo arquivo.



Configure Processor

Stopped

SETTINGS SCHEDULING PROPERTIES COMMENTS

Required field +

Property	Value
Search Value	dentista
Replacement Value	cirurgiao dentista
Character Set	UTF-8
Maximum Buffer Size	1 MB
Replacement Strategy	Literal Replace
Evaluation Mode	Line-by-Line
Line-by-Line Evaluation Mode	All

CANCEL APPLY

Profissionais.txt - Notepad

File Edit Format View Help

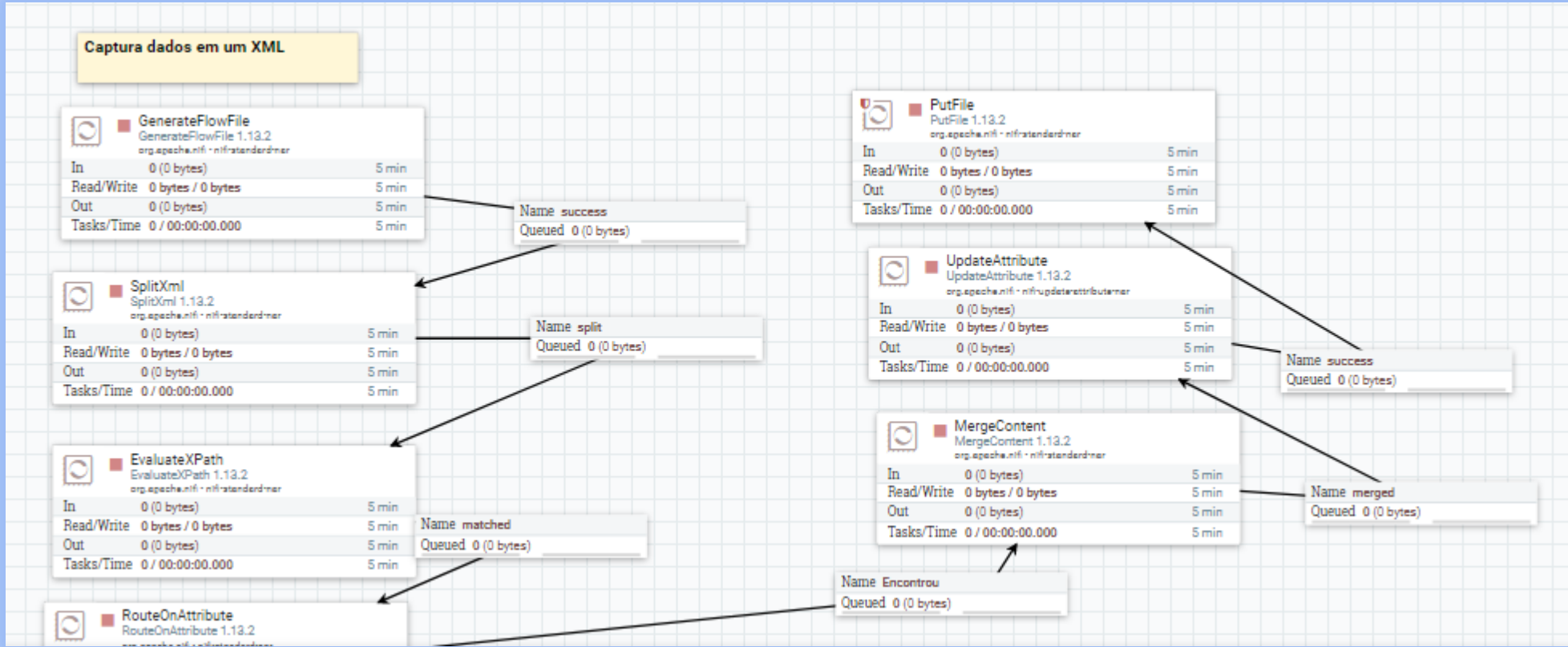
Carlos,oncologista,hospital
Ana,cirurgiao dentista,clinica
Fernanda,enfermeira,hospital
Sandra,pediatra,clinica
Fatima,cirurgiao dentista,clinica
Gilmar,cardiologista,hospital
Fabio,pediatra,clinica
Hilton,enfermeiro,clinica
Daiane,cirurgiao dentista,clinica
Paulo,farmacutico,clinica
Gilberto,pediatra,hospital

Buscando dados em arquivos XML

Podemos fazer capturar dados dentro de arquivos XML.

Capturando conteúdo nos dados de um arquivo XML

Iremos pesquisar um conteúdo específico e guardar os dados de um arquivo XML.



Required field			
Property		Value	
Destination		flowfile-attribute	
Return Type		auto-detect	
Validate DTD		true	
Datafabri		/Datafabri	
Dono		/Dono	
Observacao		/Observacao	
Preco		/Preco	
Safra		/Safra	
Tipo		/Tipo	

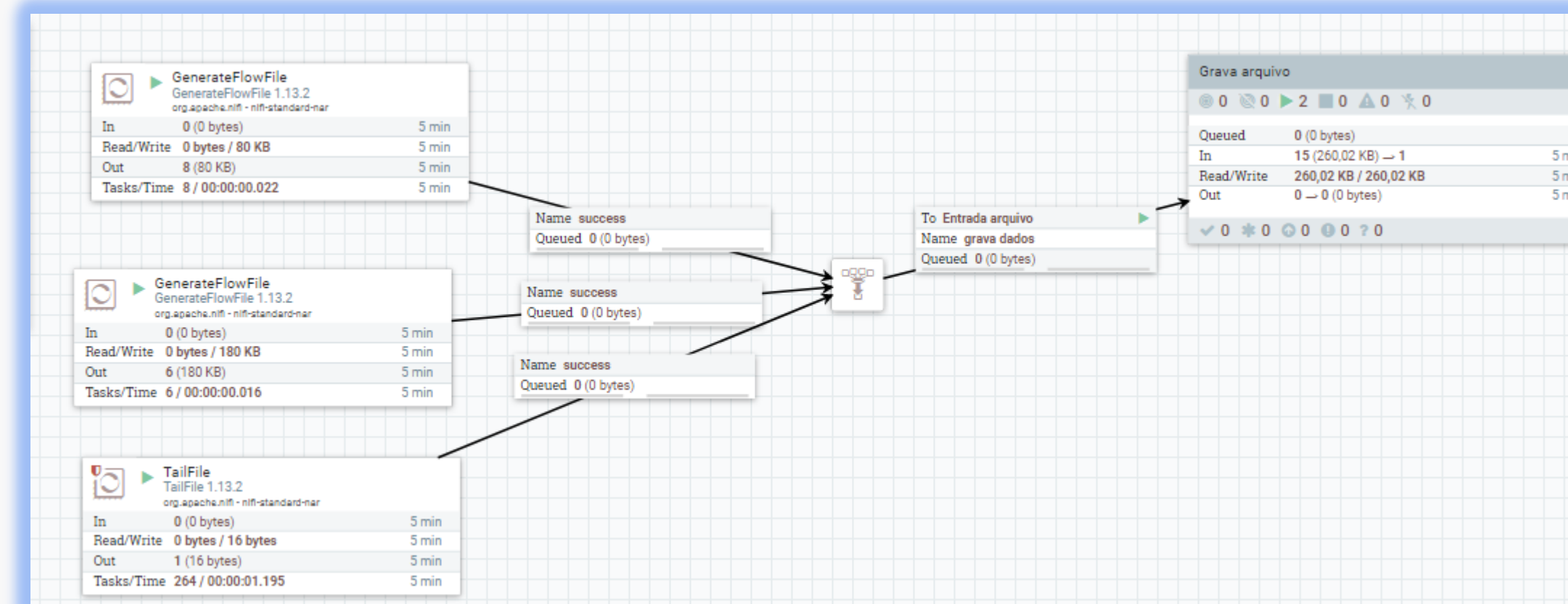
```
<?xml version="1.0" encoding="UTF-8"?><Dono>Antonio Coimbra</Dono>
<?xml version="1.0" encoding="UTF-8"?><Dono>Carlos Lisboa</Dono>
```

Utilização de process group, funil e Input port

Podemos criar processos para gerir logs no NiFi.

Organizando fluxo de dados

Iremos ver como passamos dados de um fluxo para outro.



	388e105a-67d0-468f-bae6-8c15aa775821	23/09/2021 11:54	Arquivo	30 KB
	764a539d-8119-4bd2-9ff9-5afa02800c06	23/09/2021 11:52	Arquivo	30 KB
	5702dc5e-f126-4819-b05a-f8e70ff461e1	23/09/2021 11:54	Arquivo	10 KB
	24551bb0-bac9-410d-bc8e-c218bca92693	23/09/2021 11:58	Arquivo	10 KB
	89798d2f-ebea-48f9-a92d-0414ef44257e	23/09/2021 11:56	Arquivo	10 KB
	276952a7-173b-42cd-b381-4a19b83f0eac	23/09/2021 11:57	Arquivo	10 KB
	11221839-4cbe-4442-bb42-6f70c042b7b5	23/09/2021 11:57	Arquivo	30 KB
	60508172-9346-4615-a282-945347bbcff6	23/09/2021 11:51	Arquivo	10 KB
	80020554-c869-4b24-a053-cc3eac173f0e	23/09/2021 11:53	Arquivo	10 KB
	aa0de011-95c5-493d-b7ff-7b0203e400b5	23/09/2021 11:52	Arquivo	30 KB

PRÁTICA

ENVIE AO PROFESSOR

PREPARE UM ESTUDO

Carregando qualquer arquivo do site **kaggle.com** e crie um fluxo de dados que modifique o conteúdo gravando em um novo arquivo.

Muito boa sorte e conte comigo!

