

CRIANDO UM DATAFLOW COM APACHE NIFI



DÊNIS OLIVEIRA

SOBRE MIM

DÊNIS WILLIAM OLIVEIRA

engenheiro de dados

MBA em análise de dados com BI e Big Data

atuação com Kafka, Airflow, Python, Spark, Nifi
entre outros

DENWILOLI@HOTMAIL.COM

@BIGDATA.SOLUTIONS

www.linkedin.com/in/deniswoliveira/

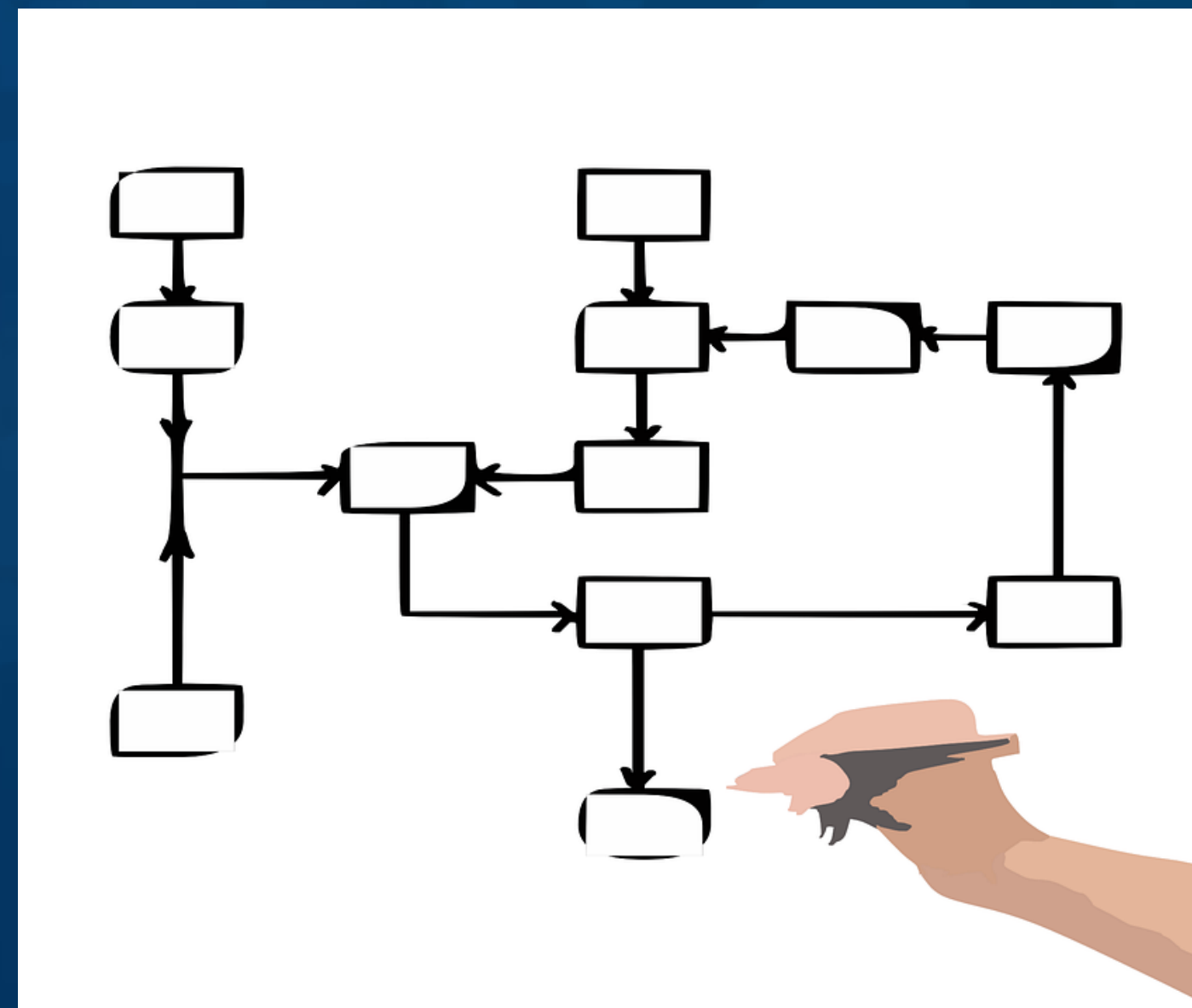
HELLO



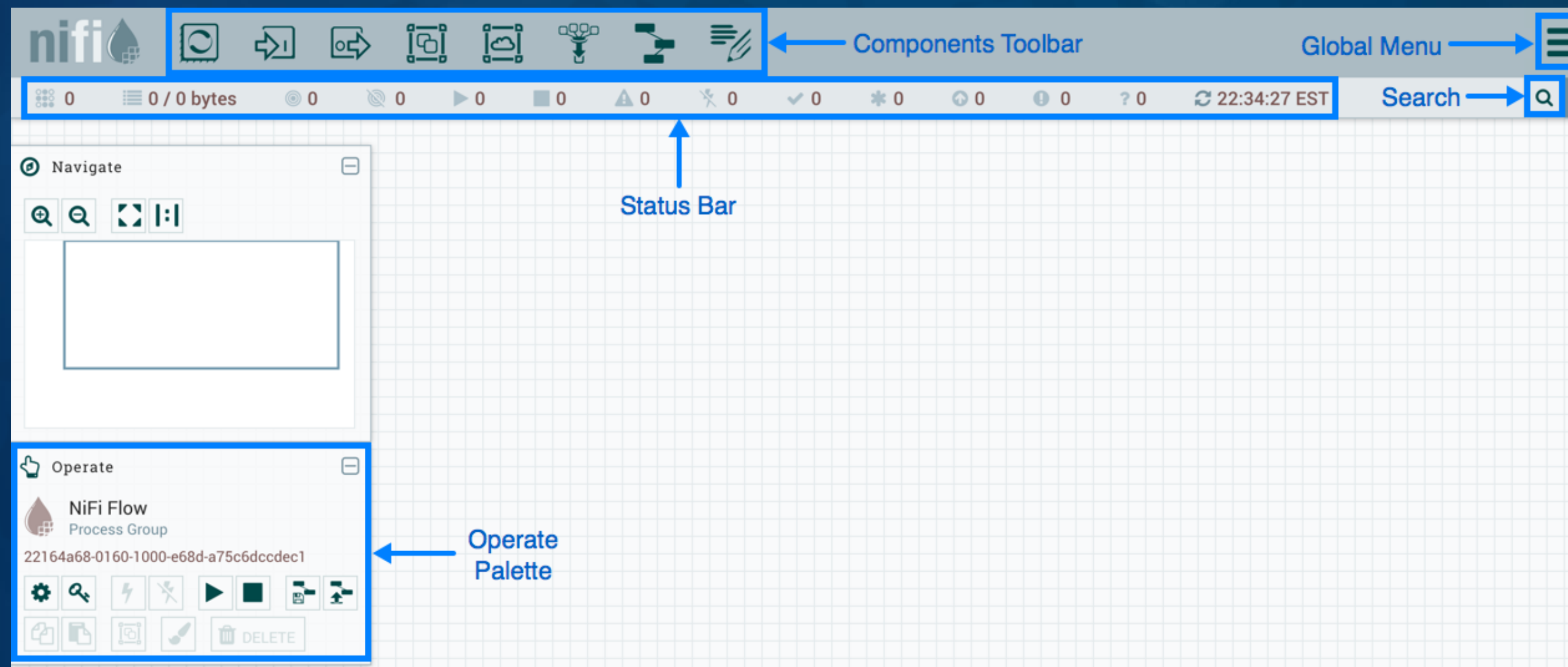


**FERRAMENTA QUE
PERMITE REALIZAR
AUTOMAÇÕES DE FLUXO
DE DADOS ENTRE
SISTEMAS DE FORMA
SIMPLES E PODEROSA**

- SISTEMA DE FLUXO DE DADOS BASEADO NOS CONCEITOS DE FLOW-BASED PROGRAMMING
- OFERECE SUPORTE A PODEROSOS E ESCALÁVEIS DIRECTED GRAPHS (GRÁFICOS DIRECIONADOS) SUPORTANDO ROTEAMENTO DE DADOS, TRANSFORMAÇÃO E LÓGICA DE MEDIAÇÃO DE SISTEMA



- POSSUI UMA INTERFACE DE USUÁRIO BASEADA NA WEB PARA DESIGN, CONTROLE, FEEDBACK E MONITORAMENTO DE FLUXOS DE DADOS



NIAGARAFI (NIFI)
FOI LANÇADO PELA AGÊNCIA
DE SEGURANÇA NACIONAL
(NSA)

ALCANÇA O STATUS DE
PROJETO DE NÍVEL SUPERIOR

2014

2007

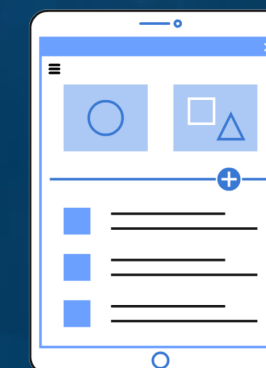
2015

DOADO PARA
APACHE SOFTWARE FOUNDATION

MAS QUAL O SEU PROPÓSITO?



ENTRADA



DESTINO

SUA LÓGICA CONSISTE EM DOIS PONTOS PRINCIPAIS:

- CONFIGURAR PROCESSORS: SEJA PARA CAPTURAR/ESCREVER DADOS, BEM COMO REALIZAR CONVERSÕES E OUTRAS MANIPULAÇÕES DE DADOS

Configure Processor

Invalid

SETTINGS

SCHEDULING

PROPERTIES

COMMENTS

Name

GenerateFlowFile

Enabled

success

Automatically Terminate Relationships

Id

f6678a36-017b-1000-ade0-fef4284a3ef1

Type

GenerateFlowFile 1.14.0

Bundle

org.apache.nifi - nifi-standard-nar

Penalty Duration

30 sec

Yield Duration

1 sec

Bulletin Level

WARN

CANCEL

APPLY

Configure Processor

Invalid

SETTINGS

SCHEDULING

PROPERTIES

COMMENTS


Required field

Property	Value
File Size	0B
Batch Size	1
Data Format	Text
Unique FlowFiles	false
Custom Text	No value set
Character Set	UTF-8
Mime Type	No value set

CANCEL

APPLY






⚠

ConsumeKafka_2_6

ConsumeKafka_2_6 1.14.0

org.apache.nifi - nifi-kafka-2-6-nar

In	0 (0 bytes)	5 min
Read/Write	0 bytes / 0 bytes	5 min
Out	0 (0 bytes)	5 min
Tasks/Time	0 / 00:00:00.000	5 min




⚠

SplitJson

SplitJson 1.14.0

org.apache.nifi - nifi-standard-nar

In	0 (0 bytes)	5 min
Read/Write	0 bytes / 0 bytes	5 min
Out	0 (0 bytes)	5 min
Tasks/Time	0 / 00:00:00.000	5 min




⚠

CaptureChangeMySQL

CaptureChangeMySQL 1.14.0

org.apache.nifi - nifi-cdc-mysql-nar

In	0 (0 bytes)	5 min
Read/Write	0 bytes / 0 bytes	5 min
Out	0 (0 bytes)	5 min
Tasks/Time	0 / 00:00:00.000	5 min



⚠

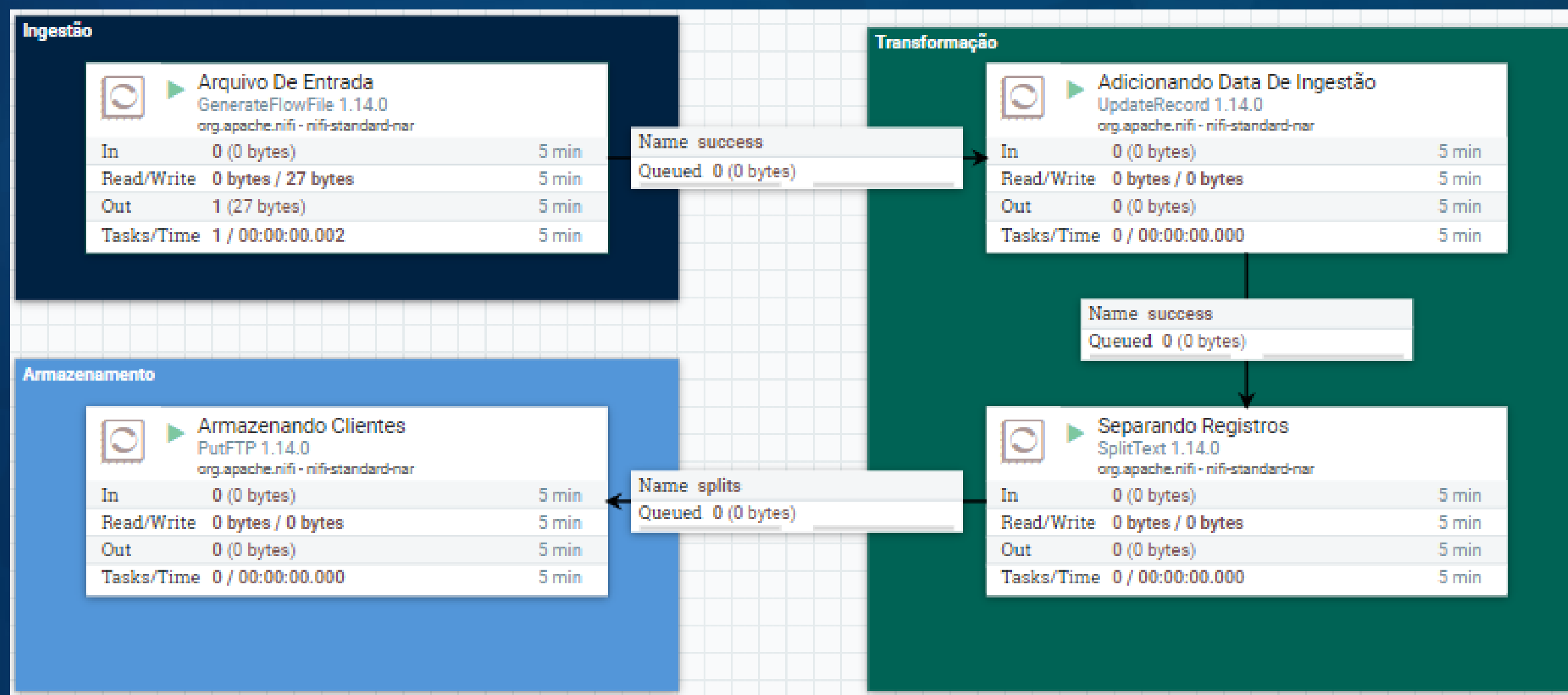
PutS3Object

PutS3Object 1.14.0

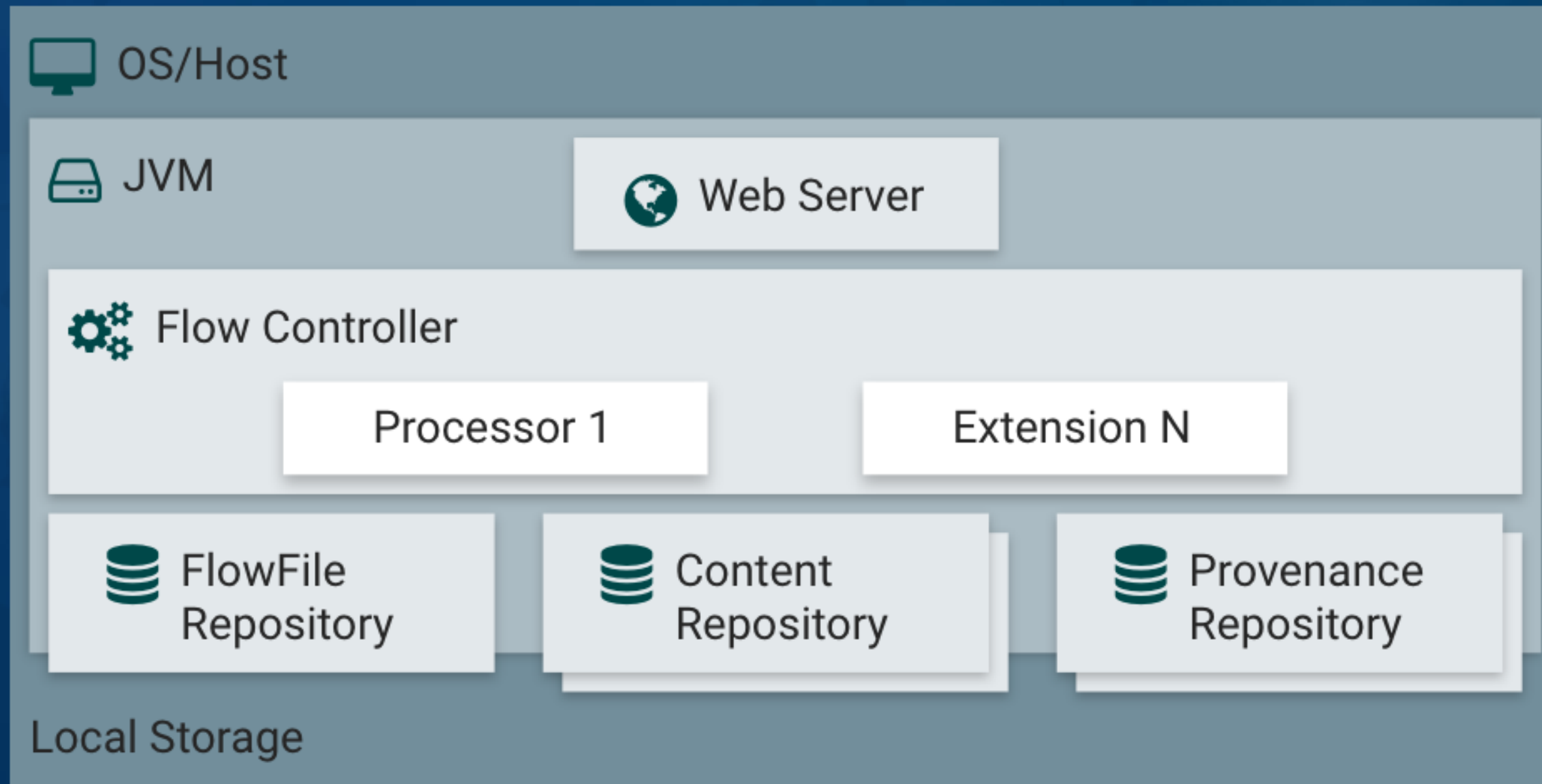
org.apache.nifi - nifi-aws-nar

In	0 (0 bytes)	5 min
Read/Write	0 bytes / 0 bytes	5 min
Out	0 (0 bytes)	5 min
Tasks/Time	0 / 00:00:00.000	5 min

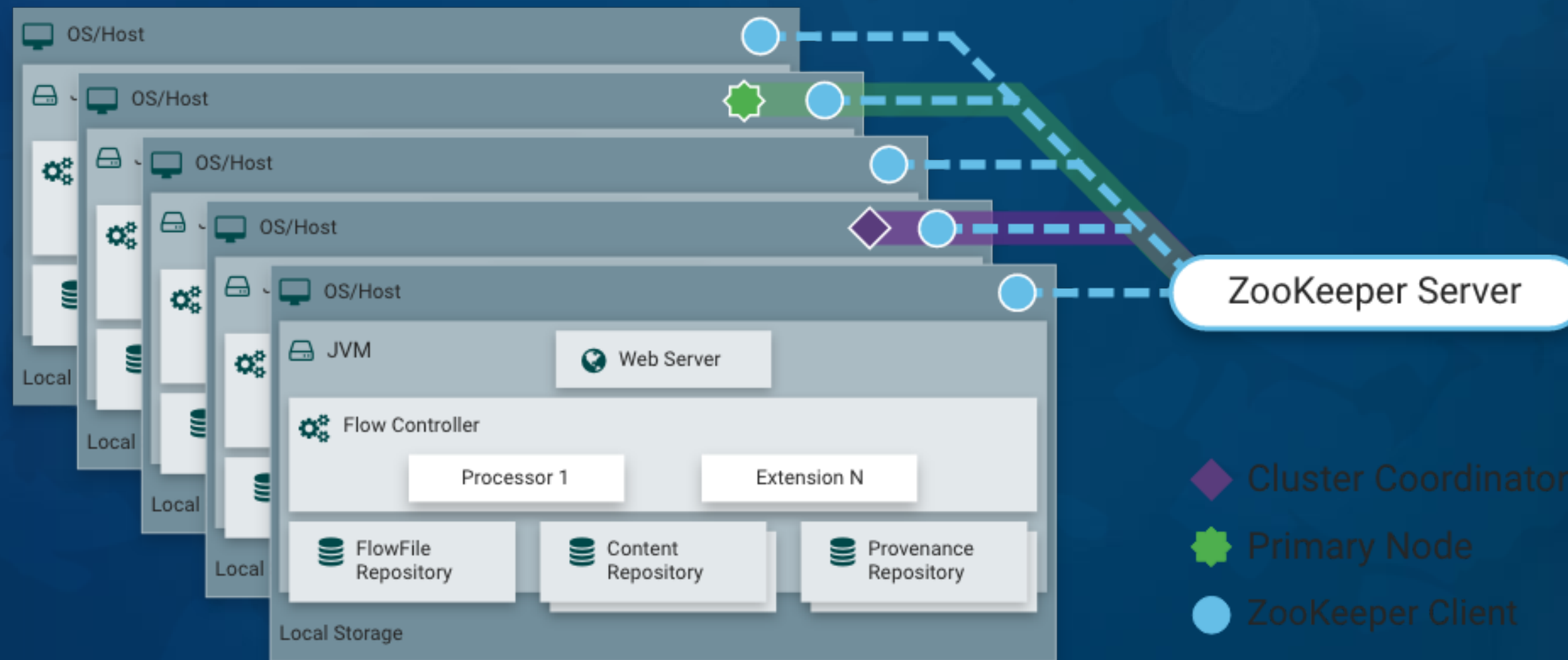
- INTERLIGAR ESTES PROCESSORS POR CONEXÕES: DIRECIONANDO SETAS ENTRE OS PROCESSORS PARA DESENHAR O FLUXO DE DADOS

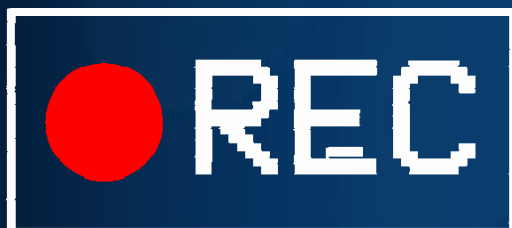


COMPONENTES



CLUSTERIZAÇÃO





DEMO



PROS



DESENVOLVIMENTO RÁPIDO

INTERFACE GRÁFICA SIMPLES E EFICIENTE

VARIAS OPÇÕES PARA DESENVOLVIMENTO (SCRIPTS GROOVY, CUSTOM PROCESSORS...)

DEBUGGING COM PROVENIÊNCIA E IMUTABILIDADE DO FLOWFILE

FERRAMENTA ESCALÁVEL

CONS



LIMITAÇÃO CONSIDERÁVEL COMPARADO COM LINGUAGENS DE PROGRAMAÇÃO

NIFI REGISTRY E TOOLKIT AINDA SÃO IMATUROS

FALTA DE CAPACIDADE INTEGRADA PARA MOVER FLOWFILES DE NÓS COM FALHA E OS RECURSOS DE COMPARTILHAMENTO

TRANSFORMAÇÃO DE DADOS PODEM SER COMPLICADAS

DUVIDAS?



VAGAS

ENGENHARIA DE DADOS



**NOSSO PROPÓSITO: IMPACTAR BILHÕES DE
VIDAS COM DADOS**

SEMANTIX.GUPY.IO/

many thanks

DÊNIS OLIVEIRA



@BIGDATA.SOLUTIONS



WWW.LINKEDIN.COM/IN/DENISWOLIVEIRA/

See Ya!

