

# TRILHA BIGDATA E NOSOL



Digerindo dados com Apache NiFi

Eliézer Zarpelão

Arquiteto de Software e Professor

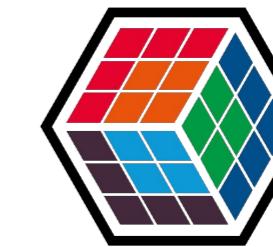


**Olá!**

Arquiteto de Software - ZarpSystem  
Professor Universitário - UNAERP

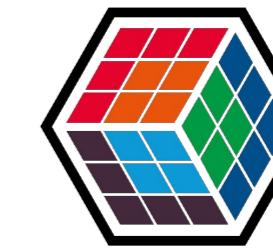
Sistemas de Informação - ICMC/USP  
Esp. Engenharia de Software - IC/Unicamp

@eliezerzarpelao  
<https://www.linkedin.com/in/eliezerzarpelao>  
<https://github.com/elizarp>  
eliezer.zarpelao@gmail.com

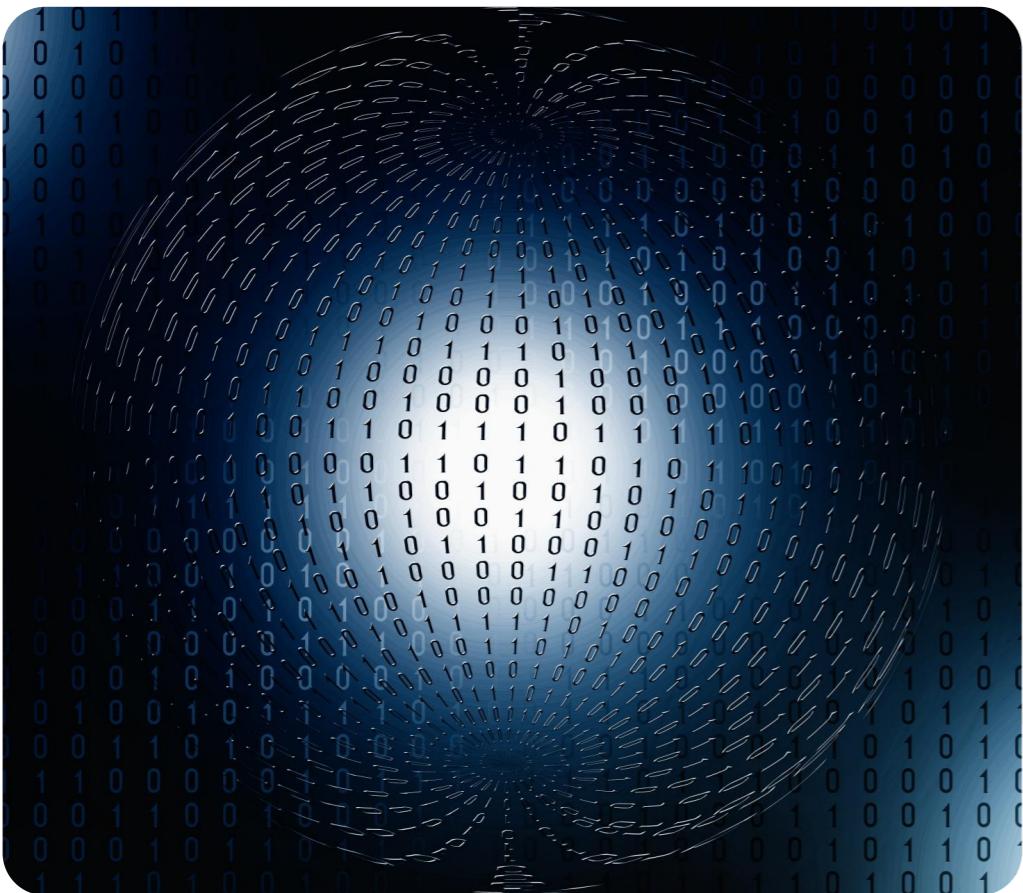


# AGENDA

- ▶ Motivação  
Porque utilizar Apache NiFi
- ▶ Estrutura NiFi  
Entenda como funciona (componentes e arquitetura)
- ▶ Demo  
Keep calm and show me the flow
- ▶ Case  
Aplicação em healthtech

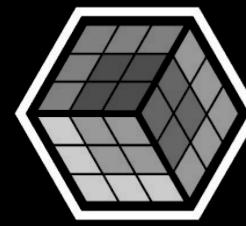


THE  
DEVELOPER'S  
CONFERENCE

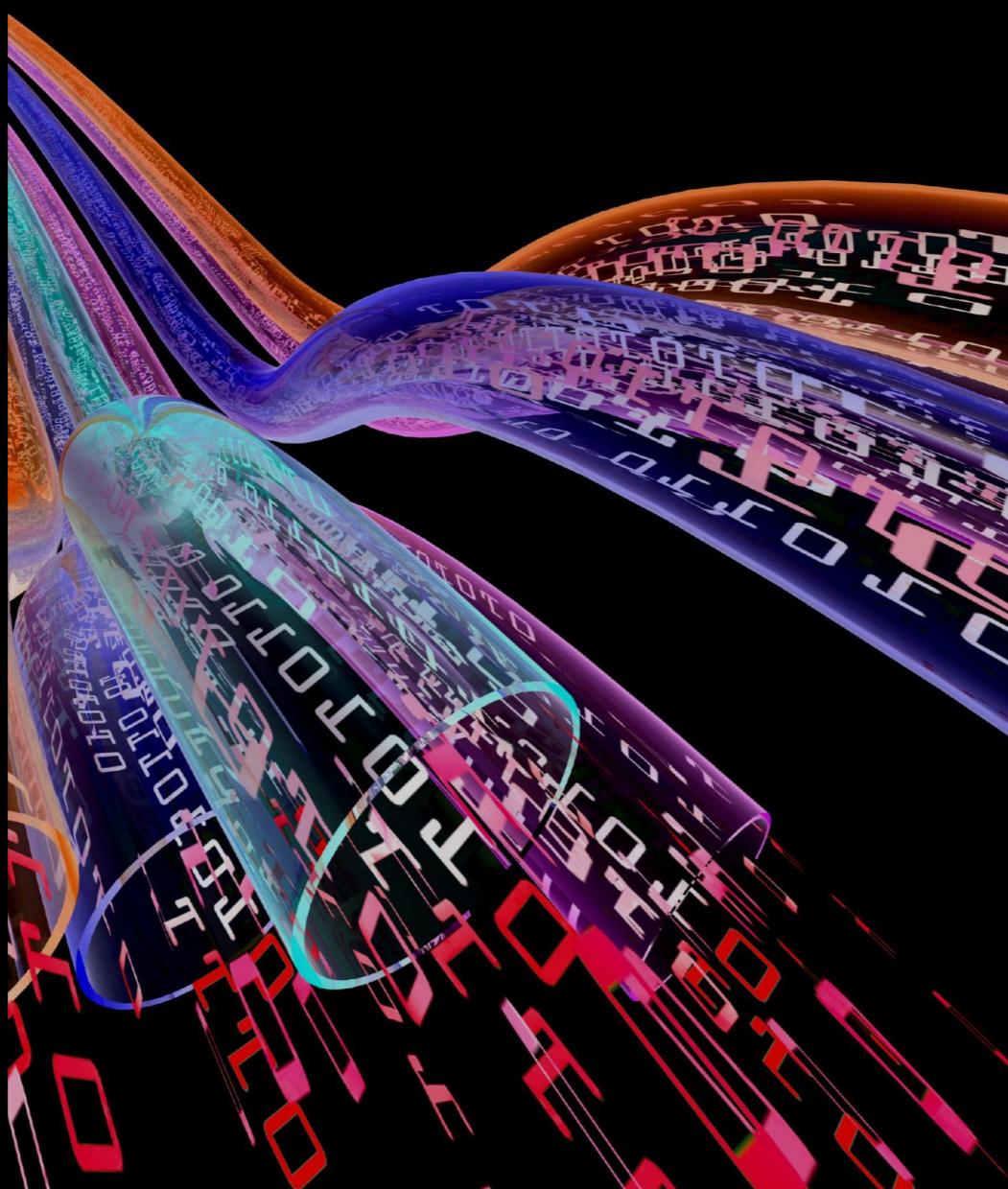


# 01

# Motivação



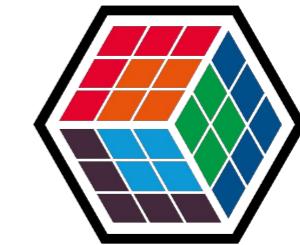
THE  
DEVELOPER'S  
CONFERENCE



# DataFlow

**Mover conteúdo de “A” para “B”**  
**Logs, HTTP, XML, CSV, Imagens, Vídeos, etc e etc**

# Tranquilo....



THE  
DEVELOPER'S  
CONFERENCE



# Desafios Big Data



## Dados

Padrões  
Formatos  
Protocolos  
Veracidade  
Validade  
Schemas  
Particionamento

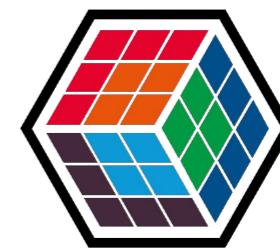
## Infraestrutura

Entrega “Exactly Once”  
Segurança  
Gestão de acesso  
Rede

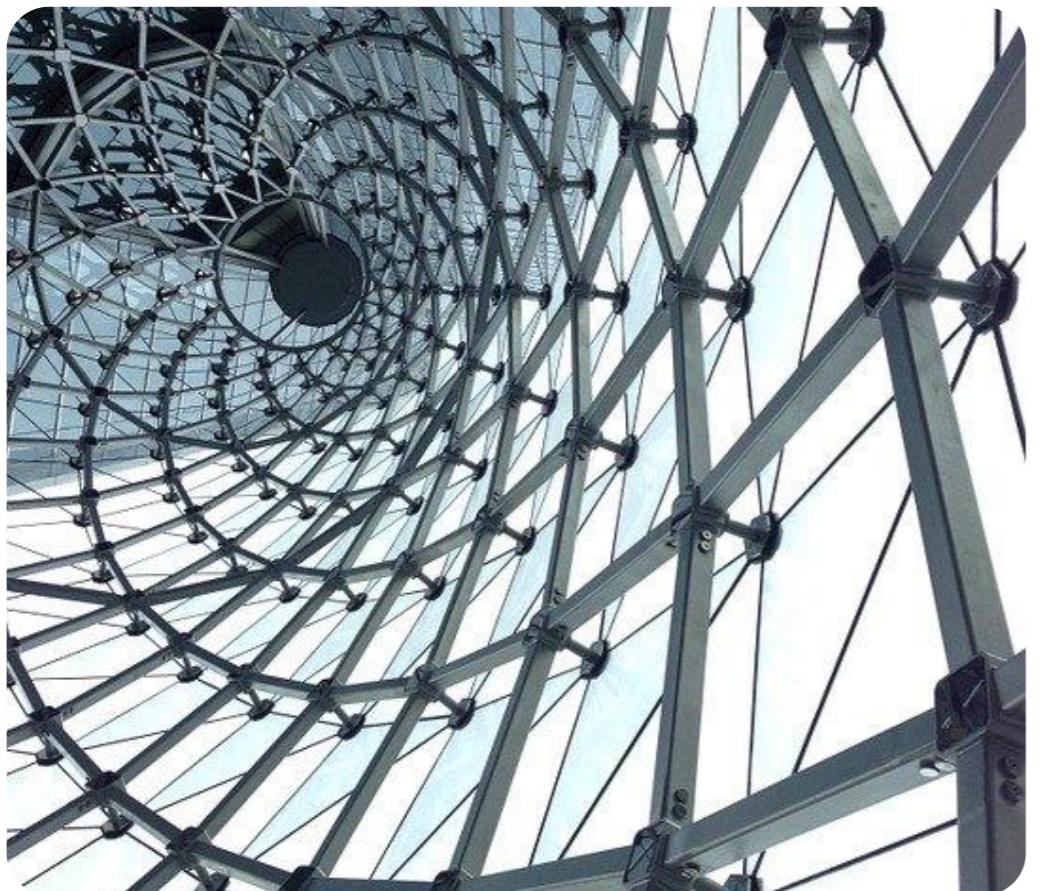
## Pessoas

Compliance  
Pessoa / Time / Grupo  
Mudança nos consumidores  
Alterações de requisitos  
Entrega “Exactly Once”





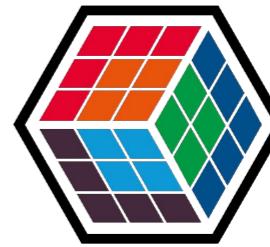
THE  
DEVELOPER'S  
CONFERENCE



# 02

# Estrutura

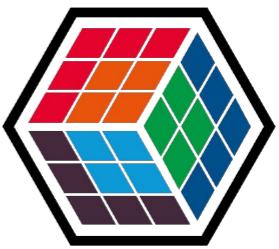
# Apache NiFi



THE  
DEVELOPER'S  
CONFERENCE

- ✗ Plataforma open-source para construção e gestão de fluxos de dados
- ✗ Escalável, segura e tolerante a falhas
- ✗ Interface Web e API para gestão e monitoramento dos fluxos
- ✗ Fluxos de dados podem ser alterados em tempo de execução

# Flow File



THE  
DEVELOPER'S  
CONFERENCE

- ✗ Objeto que se move através do fluxo
- ✗ Coração do NiFi
- ✗ Composto por:
  - Atributos (chave/valor)
  - Apontador para conteúdo

Request URL: <https://www.google.com/>  
Request Method: GET  
Remote Address: 172.217.29.164:443  
Status Code: 200 OK   
Version: HTTP/2



```
<!doctype html><html itemscope="" itemtype="http://schema.org/WebPage" lang="en-BR"><head><meta
document.documentElement.addEventListener("submit",function(b){var a;if(a=b.target){var c=a.getAttribute("action");var e=window.performance;var g=function(a,b,c,d){a.addEventListener?a.removeEventListener:b.removeEventListener};var b=[function(){google.tick&&google.tick("load","dcl")}];google.dclc=function(a){b.length?b.push(a):b.push(function(){google.jsc.mm.length||(google.jsc.mm.length=0);var f=this||self,h=Date.now();var x={};var aa=function(a,c){if(null==c) return!1;if("contains"in a&&a.nodeType==1) return a.contains(c);var I=function(){this.h=this.a=null},K=function(a,c){var d=j;d.a=a;d.h=c;return d};I.prototype.get=K;var O=function(){this.o=[];this.a=[];this.g=[];this.j={};this.h=null;this.i=[];N(this,"_custom")};var l=j;var f=I;l.shiftKey||b.shiftKey)?m="clickmod":else{var l=b.which||b.keyCode;ca&&3==l&&(l=13);if(13!=l&&3==l)e.tagName.toUpperCase(),(e=32==(b.which||b.keyCode)&&"CHECKBOX"!=e)||!(e=z(b),n=e.tagName.toUpperCase());if(m=="mouseout"||m=="mouseleave"=="m")||e&&(e==k||aa(k,e)))l.action="",l.actionElement=null;el.window.rwt=function(){return!0};}.call(this);(function(){window.jsarwt=function(){return!1};});var a=window.location,b=a.href.indexOf("#");if(0<=b){var c=a.href.substring(b+1);/(^|&)q=/.test(c);break;a:c=c[e]a=a[a.length-1];d=c[a];b=b(d);b!=d&&null!=b&&h(c,a,{configurable:!0,writable:!0});try{/*
```



FlowFile

DETAILS ATTRIBUTES

Attribute Values

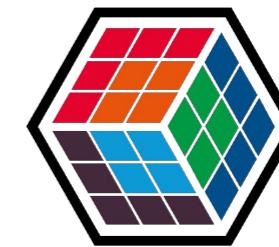
atributo1	valor-atributo-1
atributo2	valor-atributo-2
filename	0fd809ac-69ee-4011-9d99-ea5ae68e97fe
path	./
uuid	0fd809ac-69ee-4011-9d99-ea5ae68e97fe

View as: original

Filename: 0fd809ac-69ee-4011-9d99-ea5ae68e97fe  
Content Type: text/plain

1 CONTEUDO DO FLOW FILE

# Processor



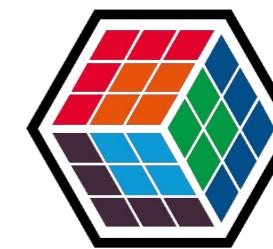
THE  
DEVELOPER'S  
CONFERENCE

- ✗ Executa um trabalho
- ✗ Roteamento, transformação...
- ✗ Mais de 200 componentes...

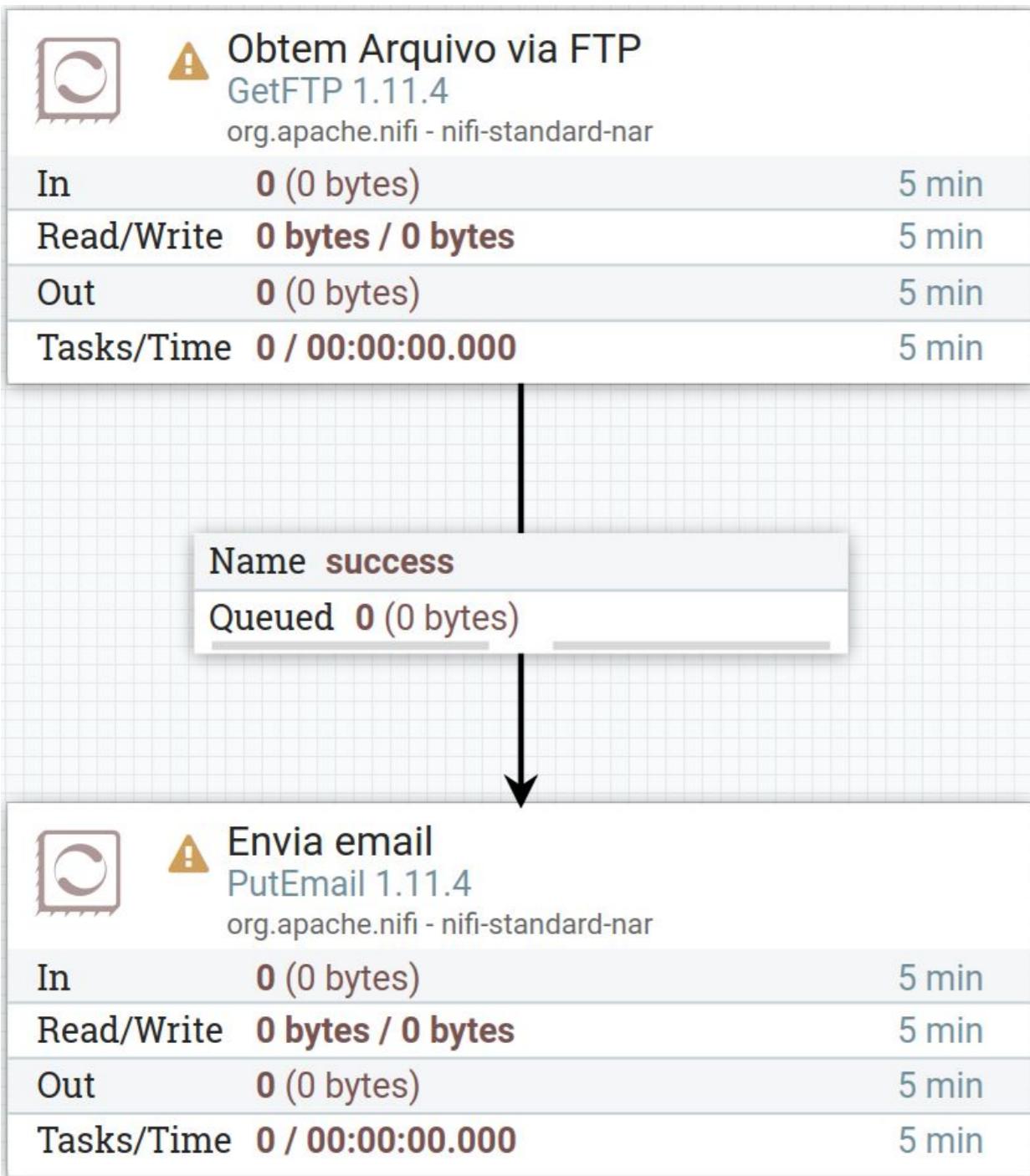
	<b>Obtem Arquivo via FTP</b> GetFTP 1.11.4 org.apache.nifi - nifi-standard-nar	
In	<b>0 (0 bytes)</b>	5 min
Read/Write	<b>0 bytes / 0 bytes</b>	5 min
Out	<b>0 (0 bytes)</b>	5 min
Tasks/Time	<b>0 / 00:00:00.000</b>	5 min



# Connections



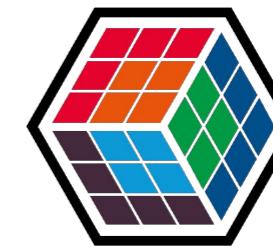
THE  
DEVELOPER'S  
CONFERENCE



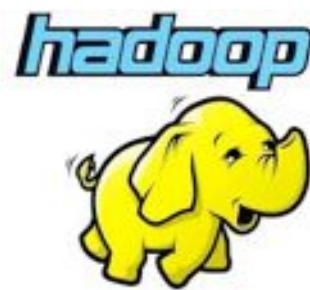
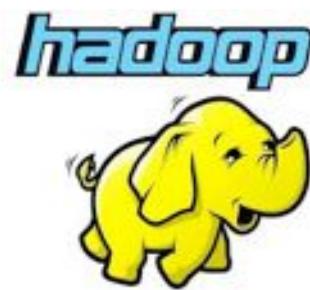
✗ Conexão entre Processors

✗ Atua como fila

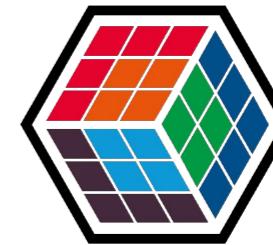
# Processors



THE  
DEVELOPER'S  
CONFERENCE

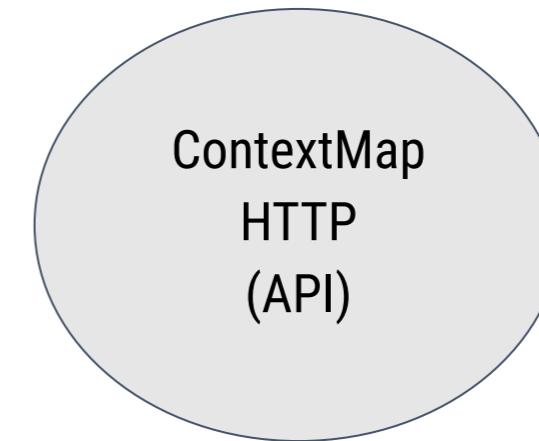
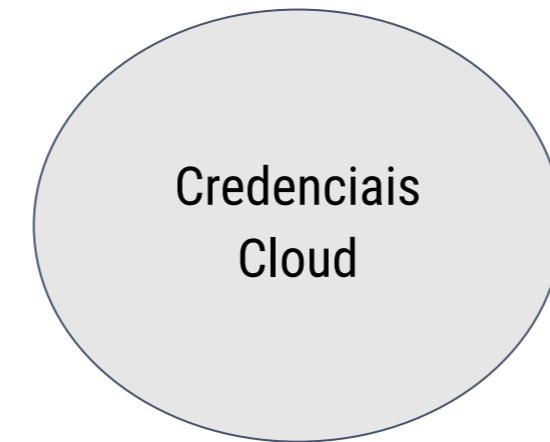
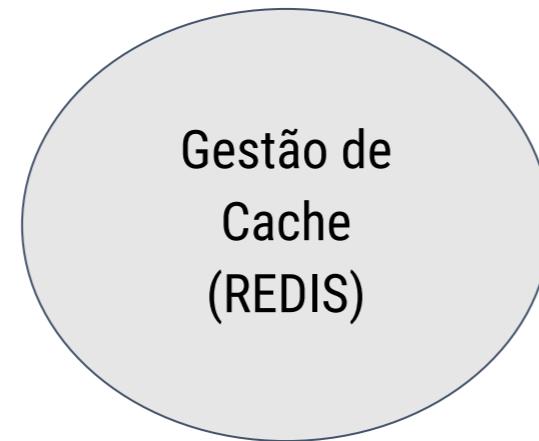


# Controller Service

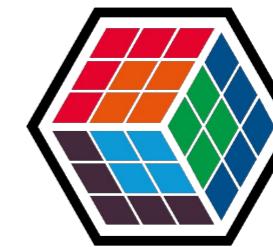


THE  
DEVELOPER'S  
CONFERENCE

- ✗ Recursos compartilhados no fluxo



# Flow Controller

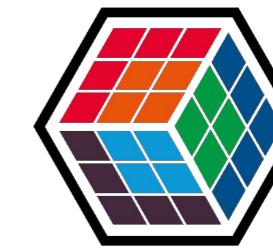


THE  
DEVELOPER'S  
CONFERENCE

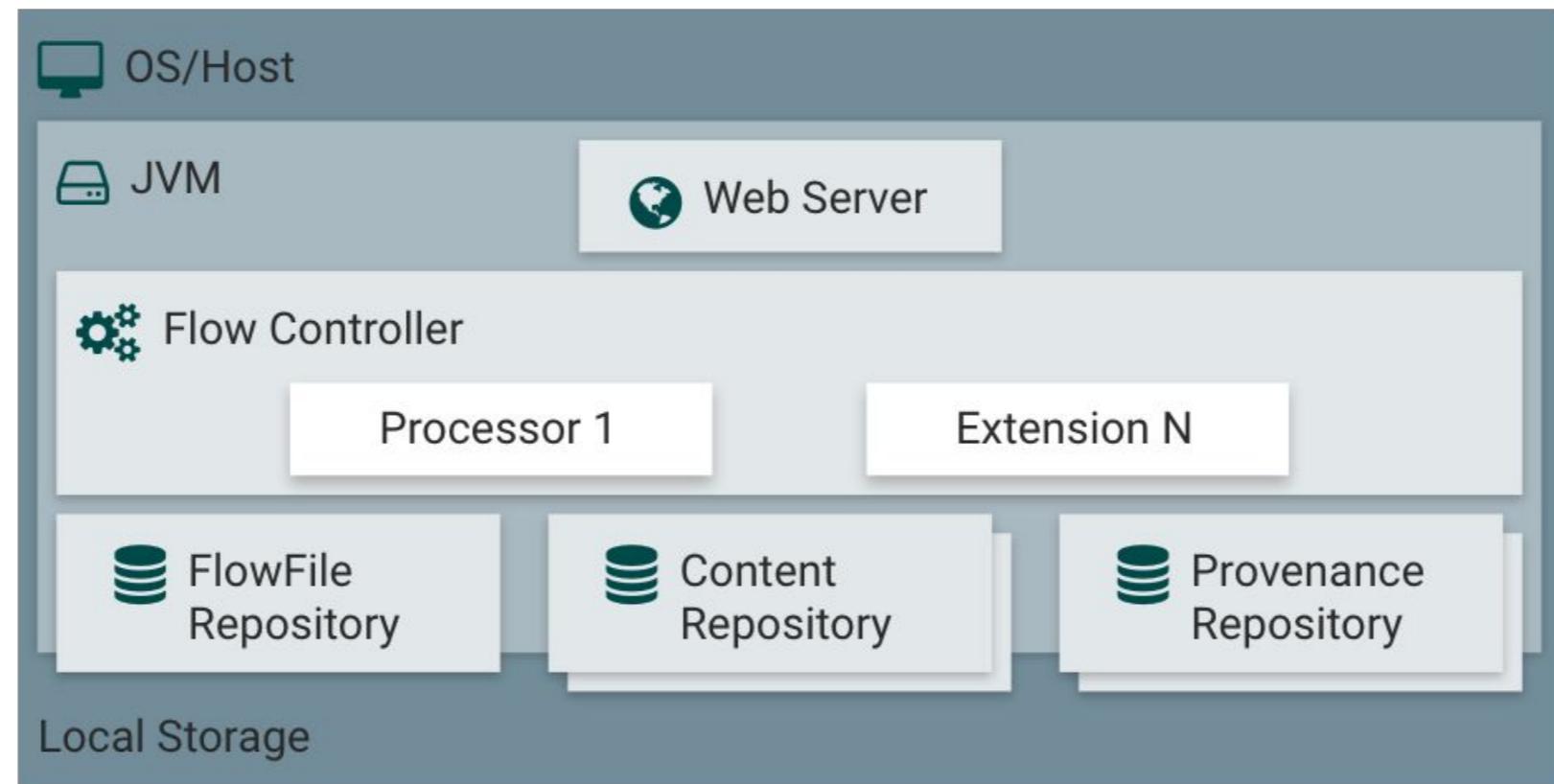
- ✗ Orquestra como os processos se conectam e gerencia todas alocações que os processos usam
- ✗ Atua como broker facilitando a troca de FlowFiles entre os Processors



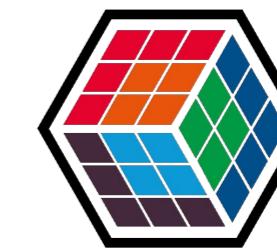
# Arquitetura



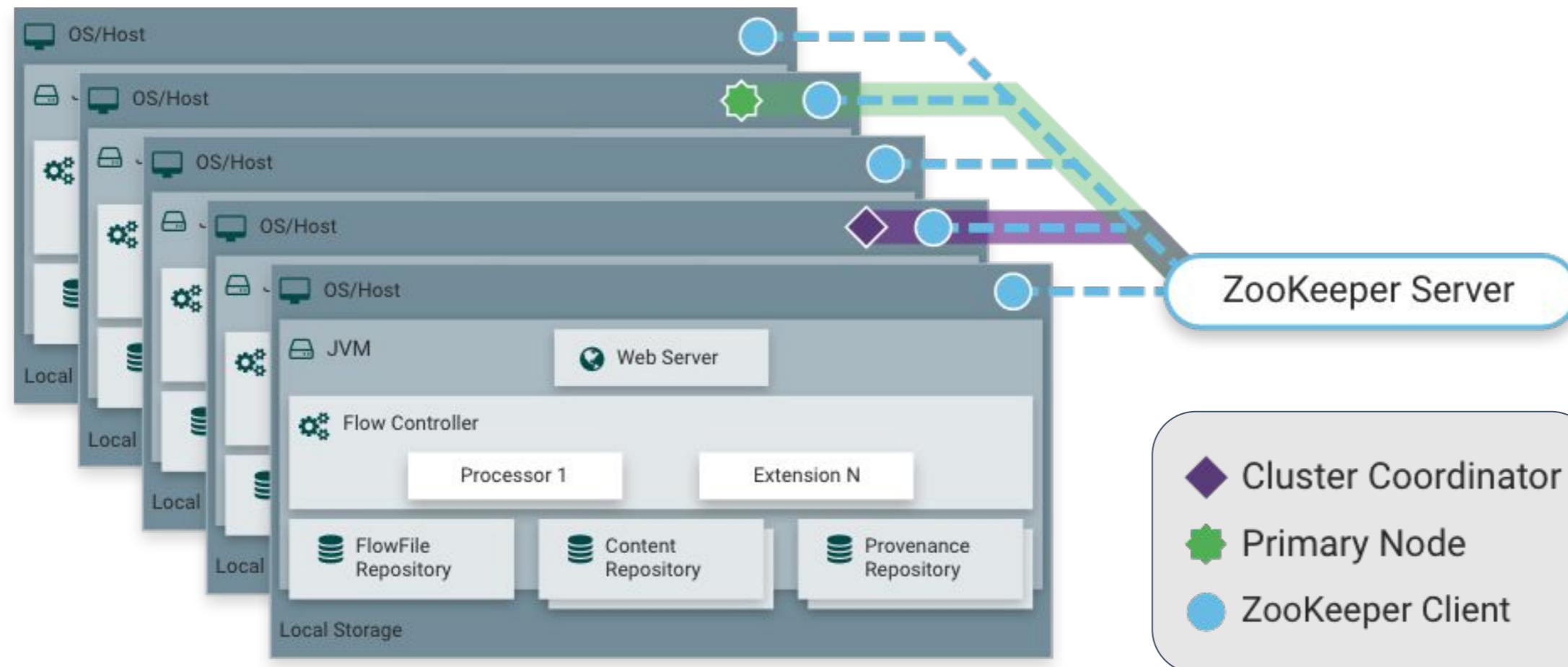
THE  
DEVELOPER'S  
CONFERENCE



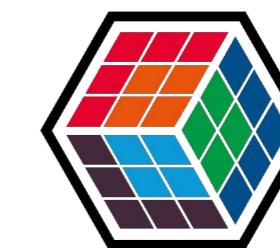
# Arquitetura - Cluster



THE  
DEVELOPER'S  
CONFERENCE



# Funcionalidades



THE  
DEVELOPER'S  
CONFERENCE

Garantia de  
Entrega

Back Pressure

Priorização  
de filas

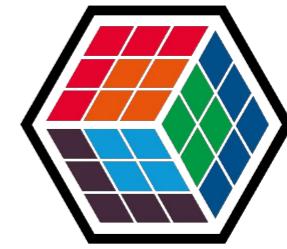
Templates de  
fluxos

Data  
Provenance

Extensibilidade

Escalabilidade

Segurança

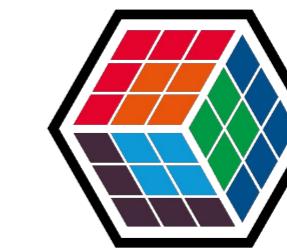


THE  
DEVELOPER'S  
CONFERENCE



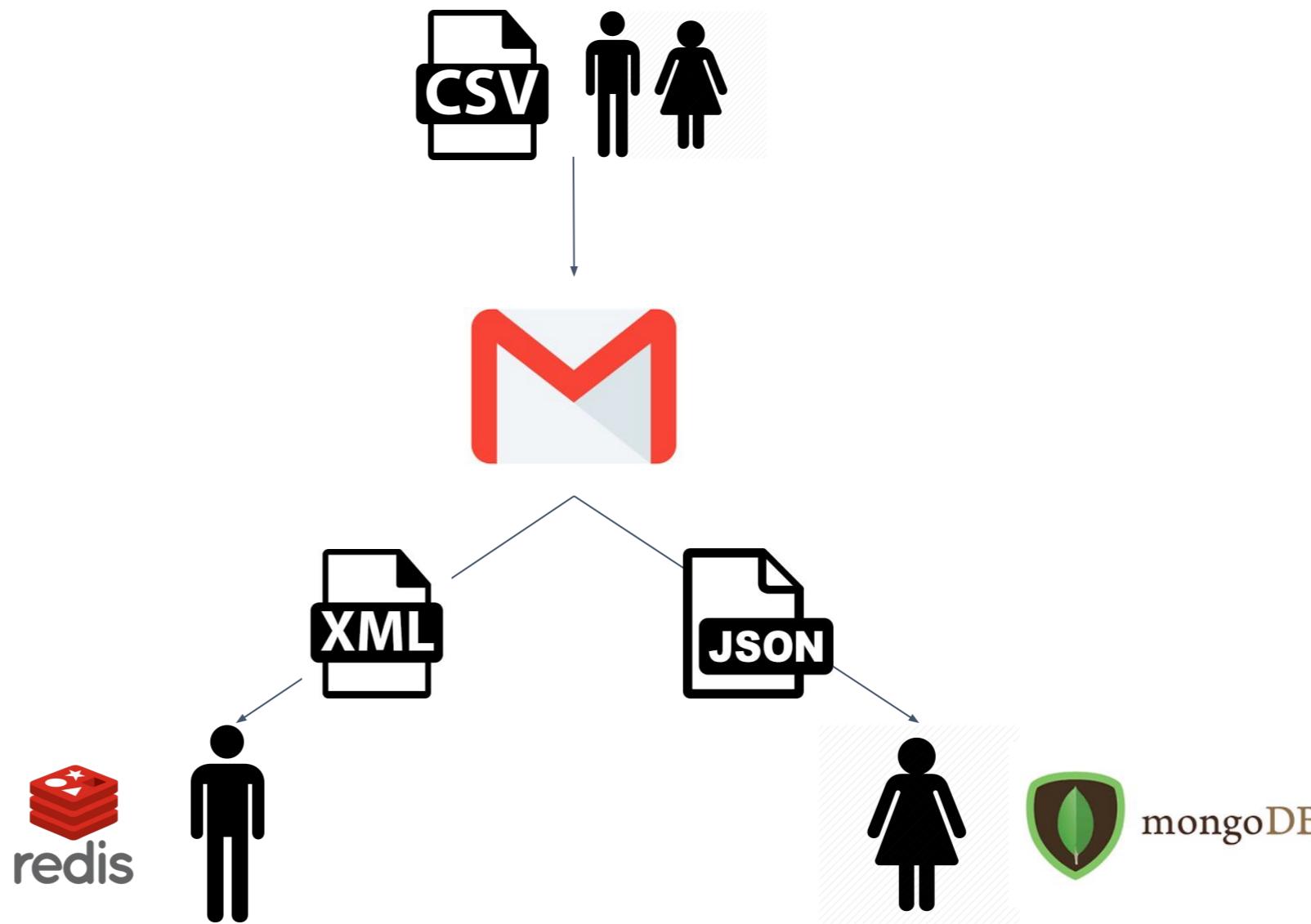
# 03

# Demo



THE  
DEVELOPER'S  
CONFERENCE

KEEP  
CALM  
AND  
DEMO  
ON

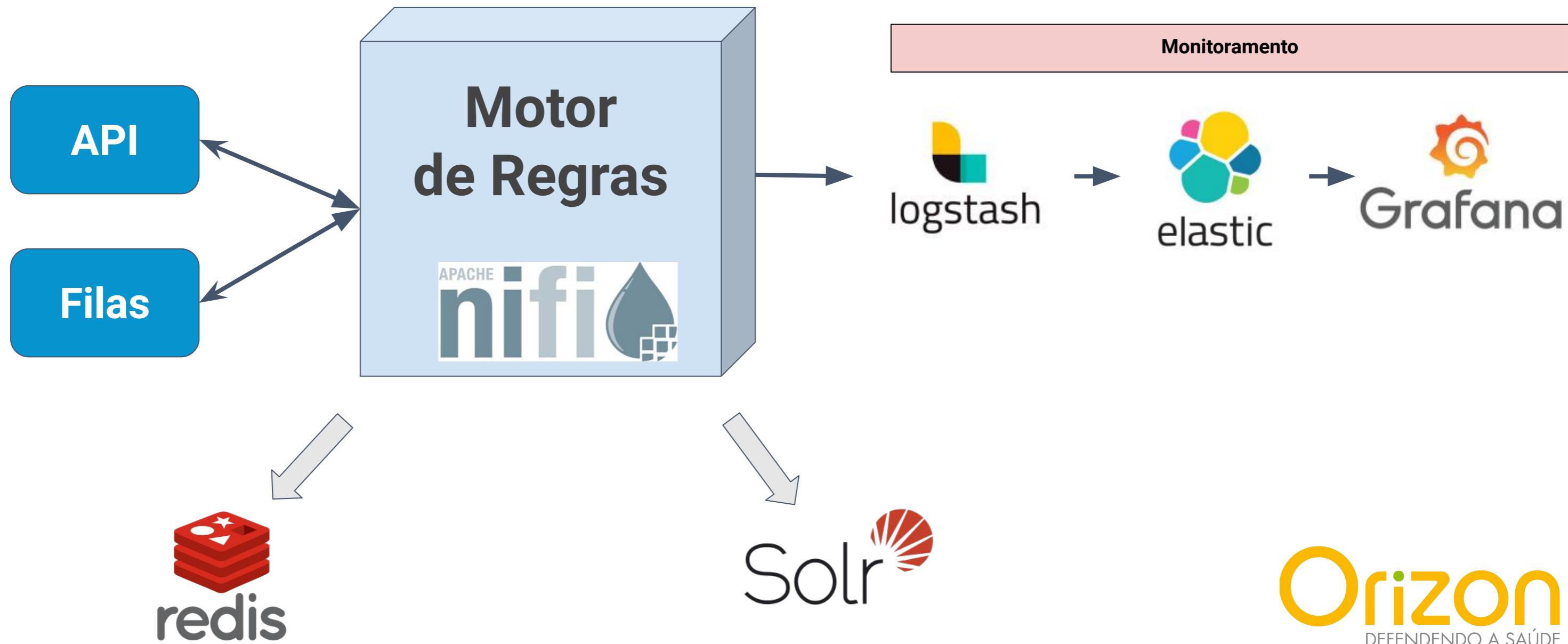




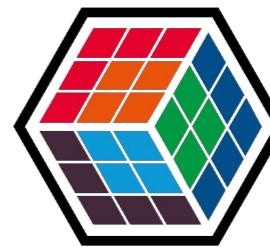
# 04 Case



# Case Orizon



# Aprenda mais



THE  
DEVELOPER'S  
CONFERENCE

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

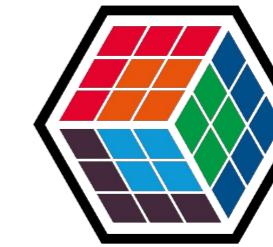
Idéias e sugestões: <https://issues.apache.org/jira/browse/nifi>

Twitter: @apachenifi

Slack: apachenifi.slack.com

Telegram: <https://t.me/nifibrasil>

# Obrigado



THE  
DEVELOPER'S  
CONFERENCE

## Perguntas?

Repositório: <https://github.com/elizarp/tdc-poa-2020-nifi>

@eliezerzarpelao

<https://www.linkedin.com/in/eliezerzarpelao>

<https://github.com/elizarp>

[eliezer.zarpelao@gmail.com](mailto:eliezer.zarpelao@gmail.com)