



# INGESTÃO DADOS TWITTER

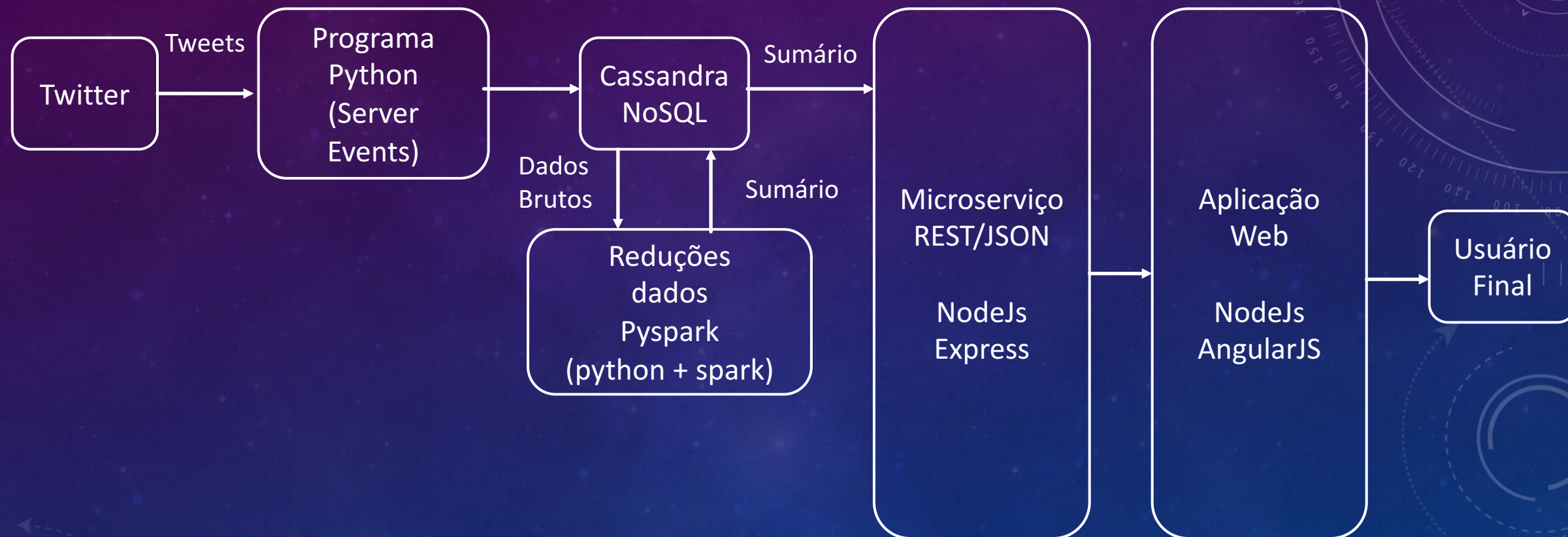
ROGÉRIO BIONDI

# INTRODUÇÃO

- O Objetivo deste trabalho é realizar a ingestão e processamento de dados da rede social twitter, demonstrando o uso de diversas técnicas atuais, como o uso de NoSQL e tecnologias de Big Data. As tecnologias usadas neste demo estão descritas a seguir.



# ARQUITETURA



- Arquitetura sujeita a alterações até a entrega do projeto. Não representa o desenho ideal do ponto de vista de boas práticas, implementação e integração contínua (infra). Críticas e melhorias propostas serão apresentadas no relatório técnico.

# TECNOLOGIAS<sup>1</sup>

- **Docker**: é uma tecnologia usada para o provisionamento rápido de ambientes por meio de containers. Ele será usado para fornecer o ambiente de demonstração, onde todas as ferramentas serão instaladas;
- **Cassandra**: é um banco de dados NoSQL colunar e sem masters, capaz de armazenar uma grande quantidade de informações. De acordo com o teorema de CAP, o Cassandra é uma solução AP default ( Availability + Partitioning Tolerance), embora para cenários específicos possa ser configurado para atender o critério "C" (Consistency).



# TECNOLOGIAS<sup>2</sup>

- **Python**: é uma linguagem de script de propósito geral, usada no desenvolvimento de sites, sistemas e automação de infraestrutura. Ainda é largamente utilizada para realizar operações de machine learning (aprendizado de máquina). Ela será usada no projeto com dois objetivos:
  - Acesso às APIs do Twitter;
  - Combinado com o Spark, processamento dos dados.
- **Spark**: é um framework de processamento paralelo/massivo, considerado uma evolução do map-reduce (Hadoop), pois é capaz de realizar as operações em memória e rodar sobre alguns gerenciadores de recursos, como o Mesos e YARN.

# TECNOLOGIAS<sup>3</sup>

- **NodeJs**: é uma linguagem de script baseada inicialmente no motor Javascript do navegador Google Chrome. Com caráter assíncrono, muito eficiente no desenvolvimento de microserviços REST/JSON. Neste desenvolvimento usaremos a biblioteca Express versão 4.0. É usada também no desenvolvimento de aplicações web e a tela de demonstração será escrita usando o framework AngularJS 1.0.



# CRONOGRAMA

	17/nov	18/nov	19/nov	20/nov	21/nov	22/nov	23/nov
	Semana 1					Semana 2	
Atividade	1	2	3	4	5	1	2
<b>Planejamento</b>							
Preparação do plano de projeto							
Abertura do projeto							
<b>Infraestrutura</b>							
Setup Infraestrutura (Containers Docker, github etc)							
<b>Desenvolvimento</b>							
Modelagem base Cassandra							
Captura Tweets							
Criação API REST							
Aplicação para apresentação dos resultados							
<b>Conclusão</b>							
Criação do Relatório Técnico							
<b>Entrega final</b>							