

PROCESSAMENTO DE STREAMING COM .NET E SPARK

FABIO GOUW



175 ZB em 2025 ⁽¹⁾

OU

175,000,000,000,000,000,000,000 bytes ⁽²⁾

⁽¹⁾ <https://www.seagate.com/our-story/data-age-2025/>

⁽²⁾ base 10



FABIO GOUW

Arquiteto de Soluções @ Itaú

 fabiogouw

 @fabiogouw

 fabiogouw.com



AGENDA

1. Streaming
 1. Conceitos
 2. Casos de uso
2. Ferramentas
 1. Apache Spark
 2. .NET + Spark
3. Demo

STREAMING

STREAMING



Conjunto infinito de
dados

BATCH



Conjunto finito de
dados

STREAMING - CASOS DE USO



COMPLEX EVENT
PROCESSING (CEP)



MATERIALIZED VIEWS



STREAMING ANALYTICS

FERRAMENTAS DE DATA STREAMING



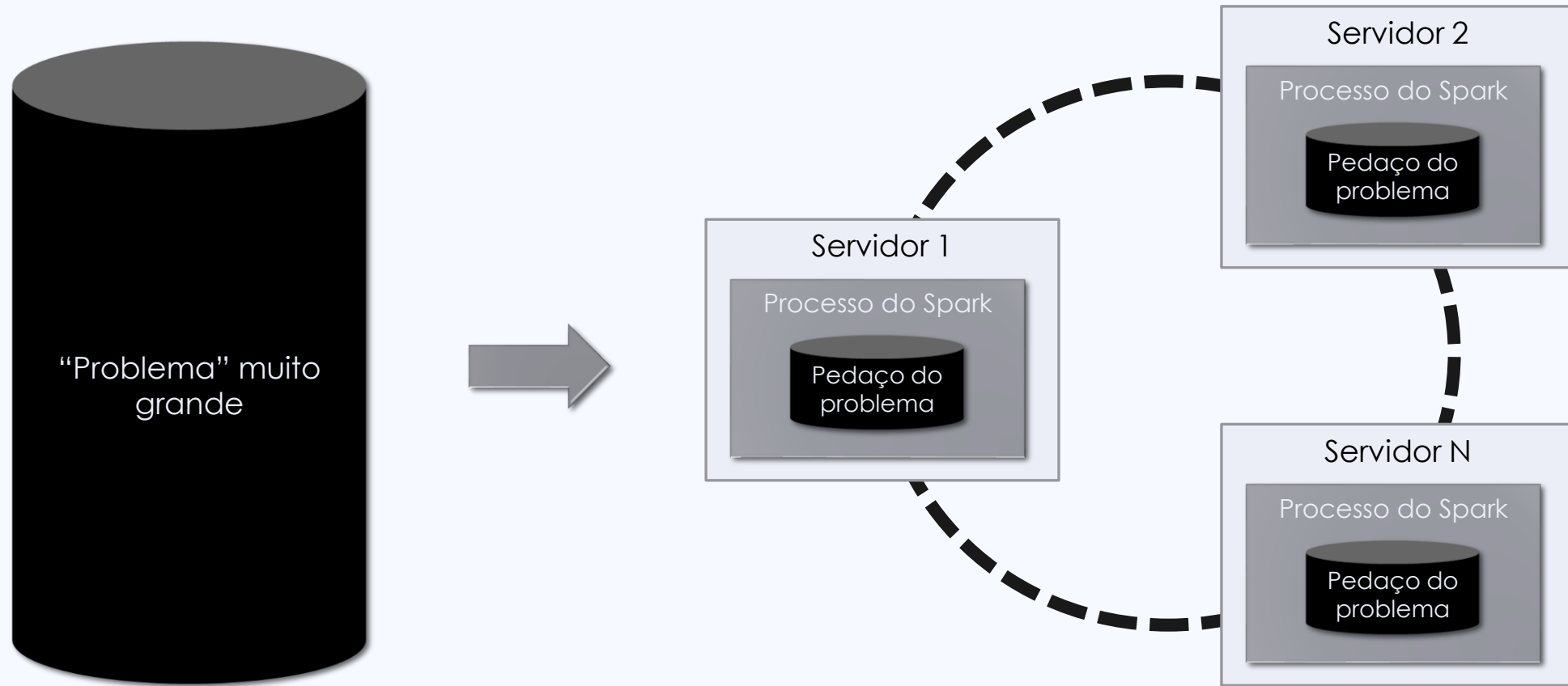
Google Dataflow



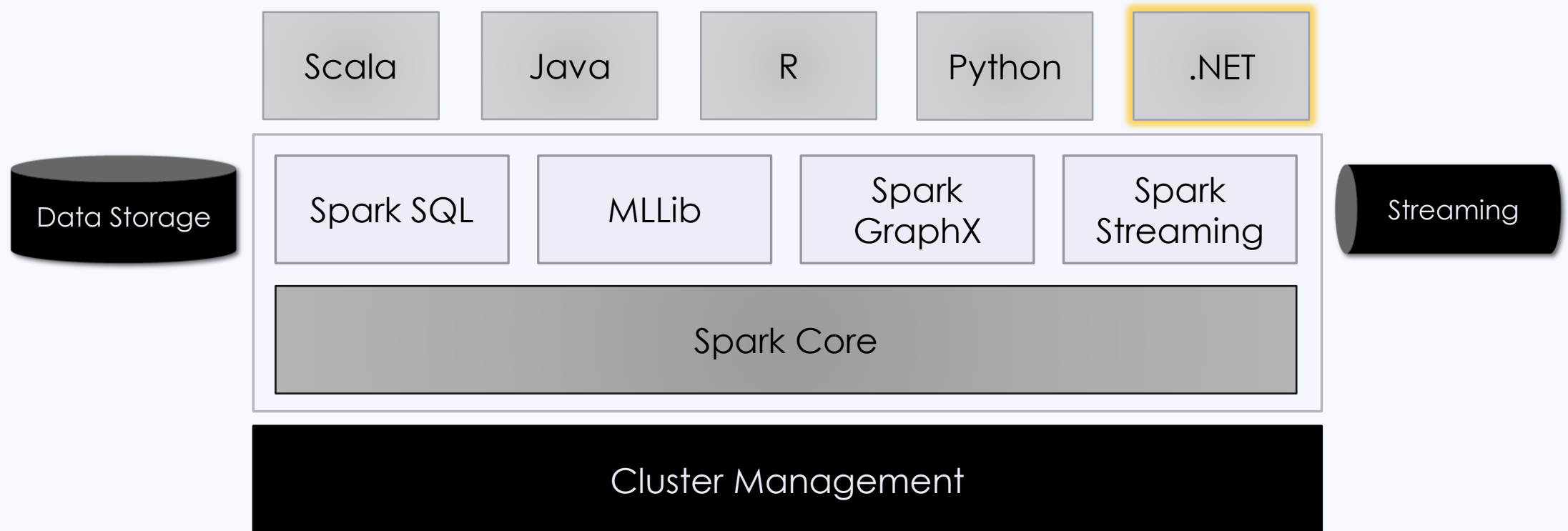
Motor para processamento distribuído de dados em larga escala.



SPARK - EXECUÇÃO



SPARK - COMPONENTES



.NET FOR APACHE SPARK



.NET for Apache® Spark™ - projeto Open Source

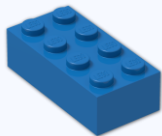
- Lançado em Abril de 2019 (Spark + AI Summit 2019)
- Versão 1.0 em Outubro de 2020 (.NET Conf)



Motivação: Fornecer aos desenvolvedores .NET uma opção de primeira classe para trabalhar com Apache Spark e Big Data



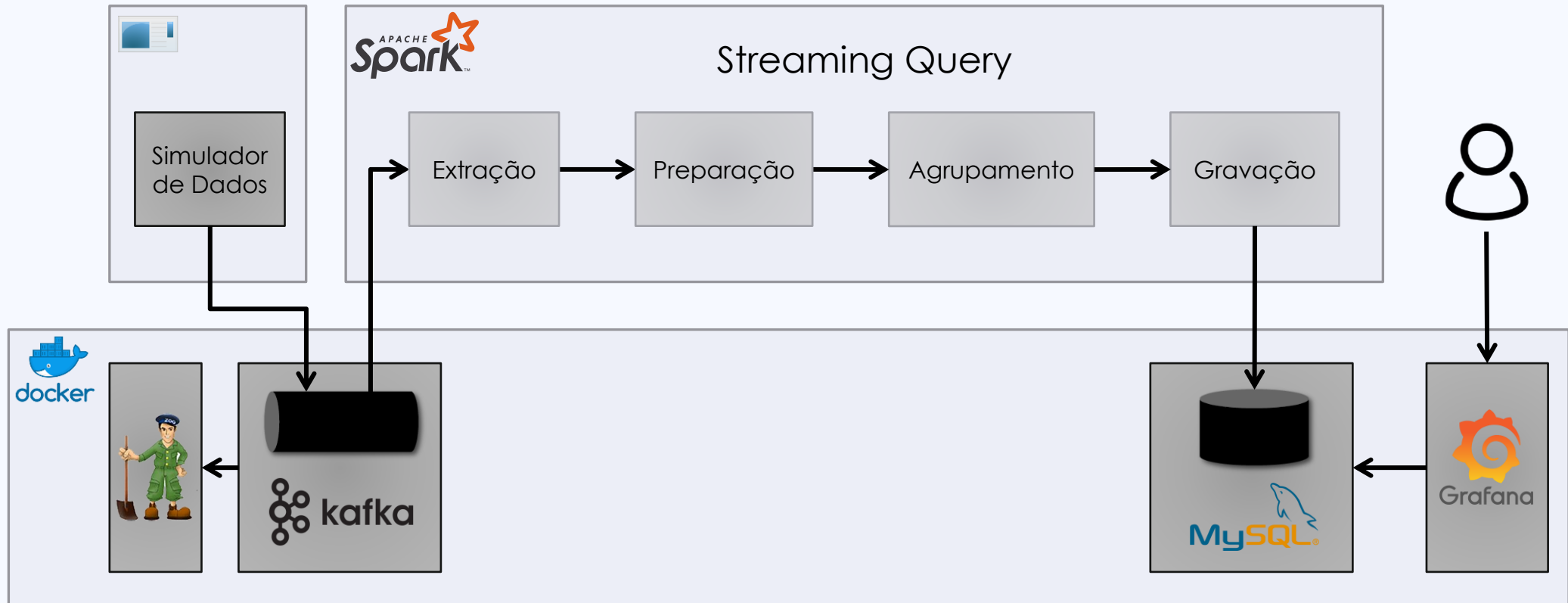
Não é objetivo converter programadores Scala, Python, Java ou R para usarem .NET



Suporte para C# e F# para .NET Standard 2.0 (.NET Core 2+, .NET 5)

DEMO

DEMO



PARA SABER MAIS

- Código fonte da demo:
<http://bit.ly/SparkNetDemo>
- Tutorial básico do .NET for Spark:
<https://dotnet.microsoft.com/learn/data/spark-tutorial/intro>
- Projeto Apache Spark:
<https://spark.apache.org/>
- Projeto .NET for Apache Spark:
<https://github.com/dotnet/spark>



fabiogouw



@fabiogouw



fabiogouw.com