



NoSQL e o processamento de dados em larga escala (Parte 2)

Prof. Dr. Robson L. F. Cordeiro
robson@icmc.usp.br

O que é NoSQL?

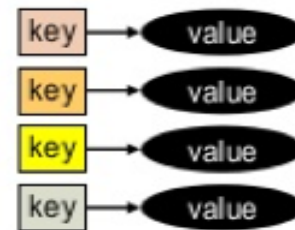
- Qualquer sistema gerenciador de dados em larga escala, relacional ou não-relacional, é dito um sistema **NoSQL**
- “Not only SQL” → **NoSQL**
- Não é anti-SQL ou anti-Relacional



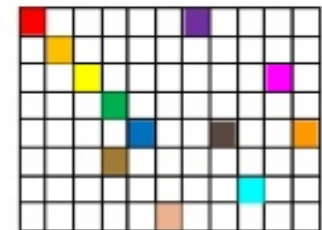
Fonte: www.improgrammer.net

Categorias NoSQL

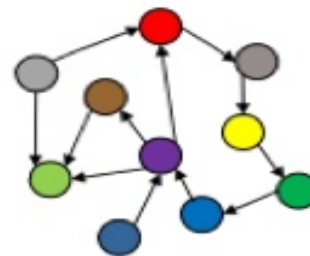
Key-Value



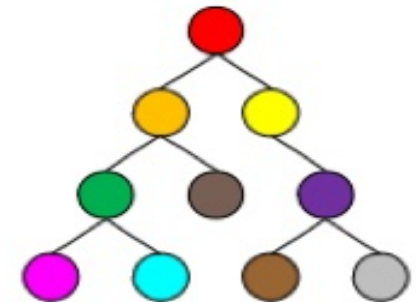
Column-Family



Graph



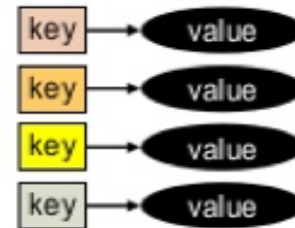
Document



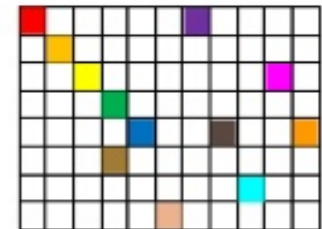
Categorias NoSQL

- Existem quatro categorias principais de SGBDs NoSQL:
 - Pares de chave-valor
 - Orientado a colunas
 - Grafos
 - Orientado a documentos

Key-Value



Column-Family



Graph



Document

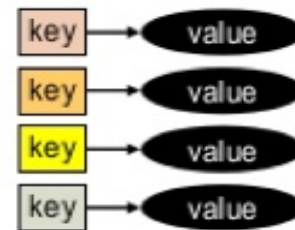


Categorias NoSQL

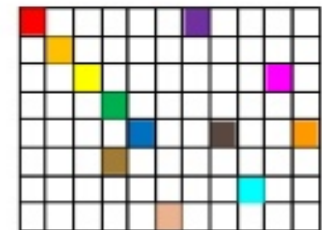
- Existem quatro categorias principais de SGBDs NoSQL:
 - Pares de chave-valor
 - Orientado a colunas
 - Grafos
 - Orientado a documentos

- Características e limitações específicas

Key-Value



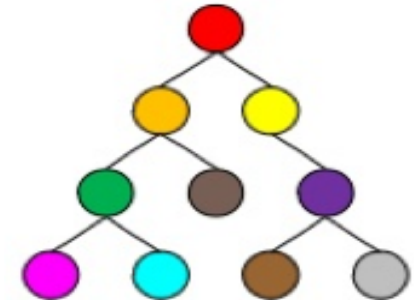
Column-Family



Graph



Document

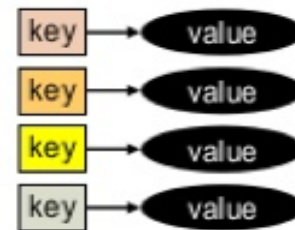


Categorias NoSQL

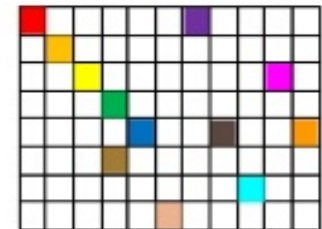
- Existem quatro categorias principais de SGBDs NoSQL:
 - Pares de chave-valor
 - Orientado a colunas
 - Grafos
 - Orientado a documentos

- Características e limitações específicas
- Melhor opção? Depende do problema em mãos

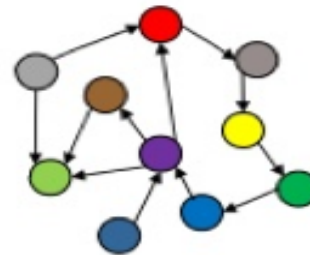
Key-Value



Column-Family



Graph



Document



Porém, ...

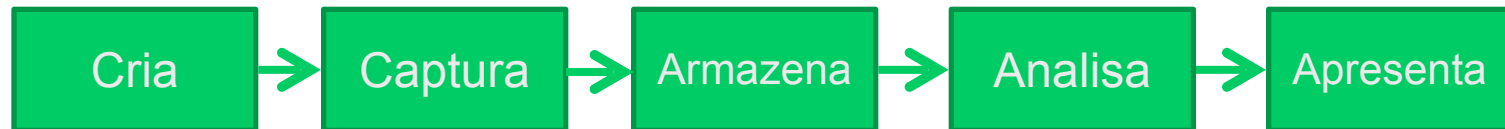
- ... onde obter capacidade de **processamento** e de **armazenamento** para rodar tais sistemas de armazenamento em larga escala?





NoSQL

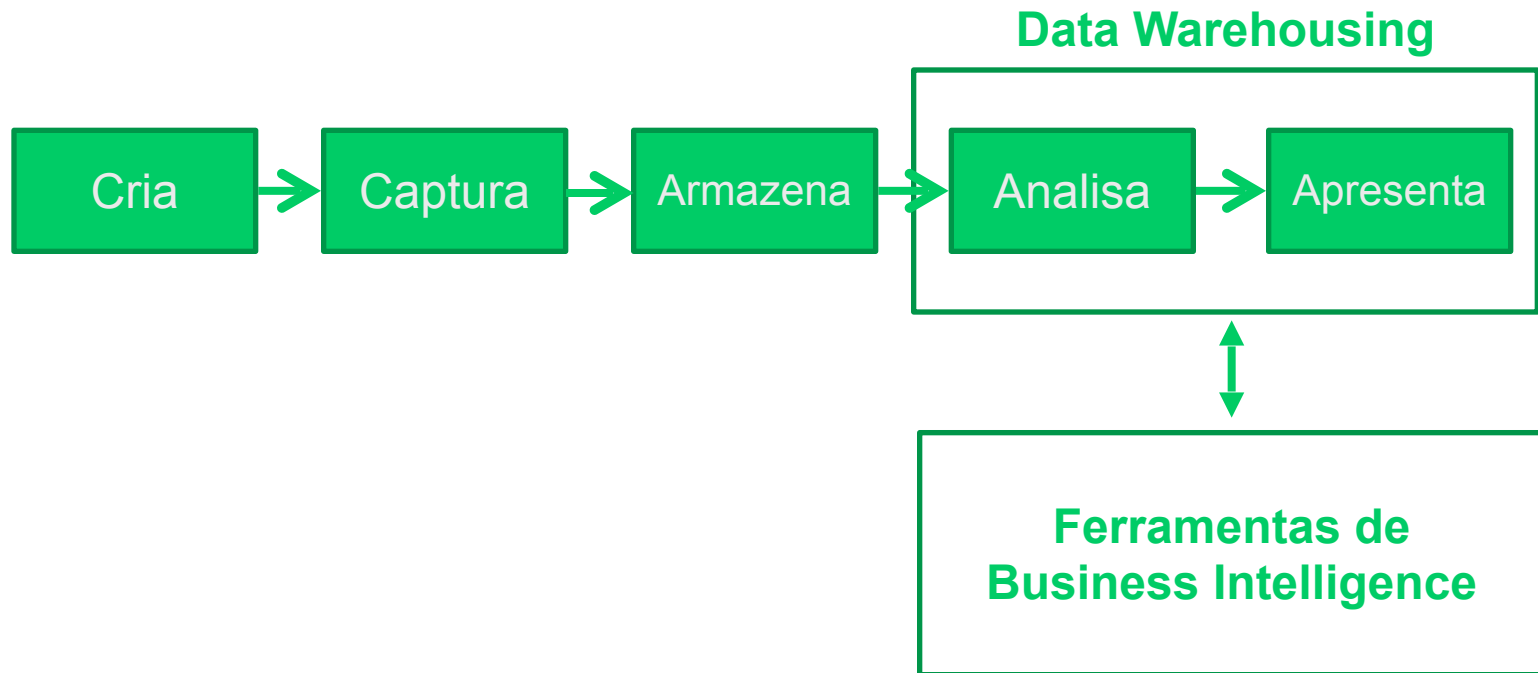
Dados: ciclo de vida





NoSQL

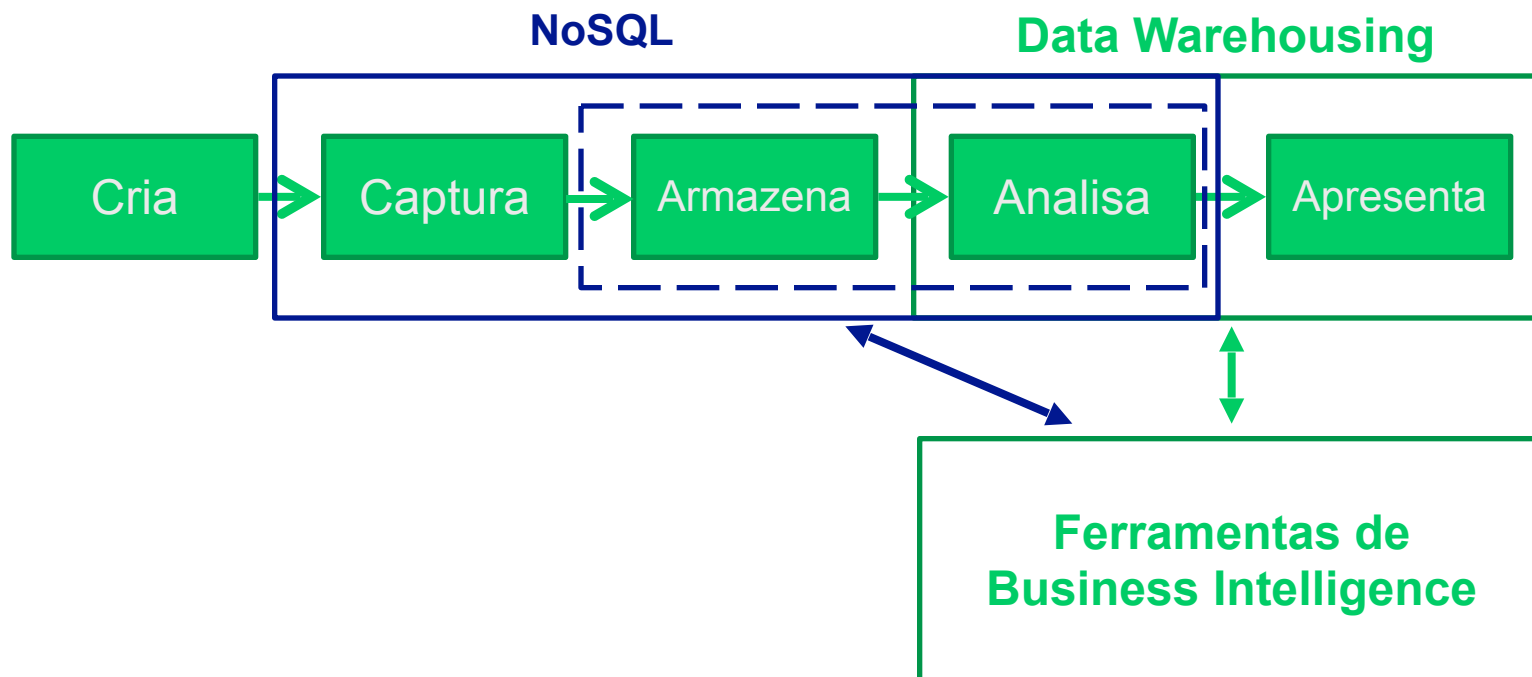
Dados: ciclo de vida





NoSQL

Dados: ciclo de vida



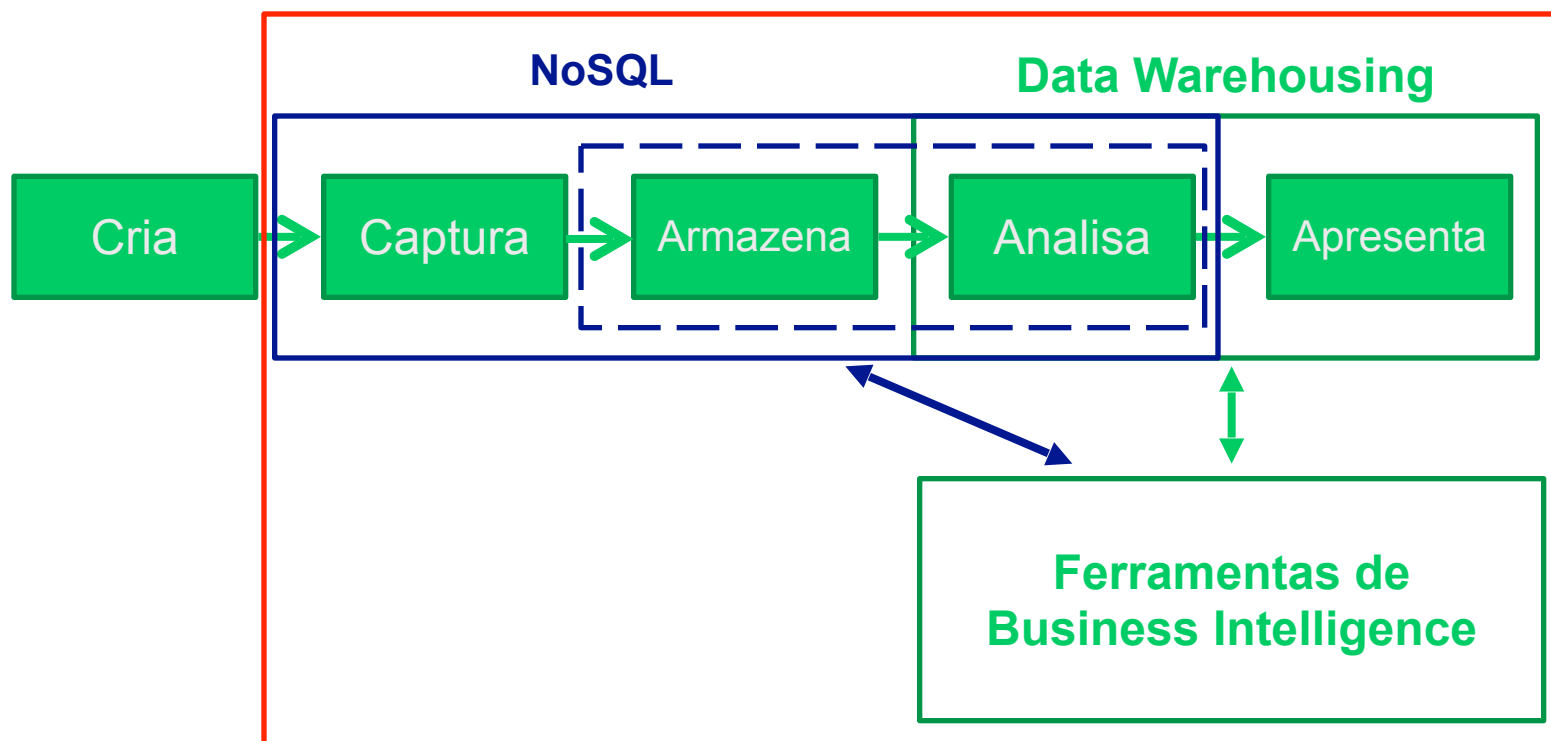


NoSQL

Dados: ciclo de vida

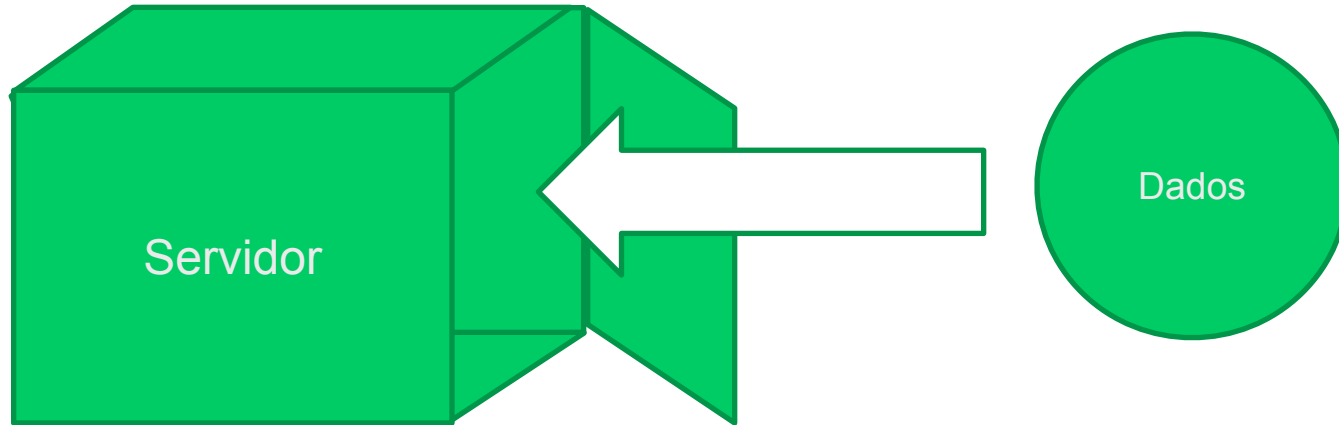


Programação Funcional para Big Data (Spark e Hadoop)



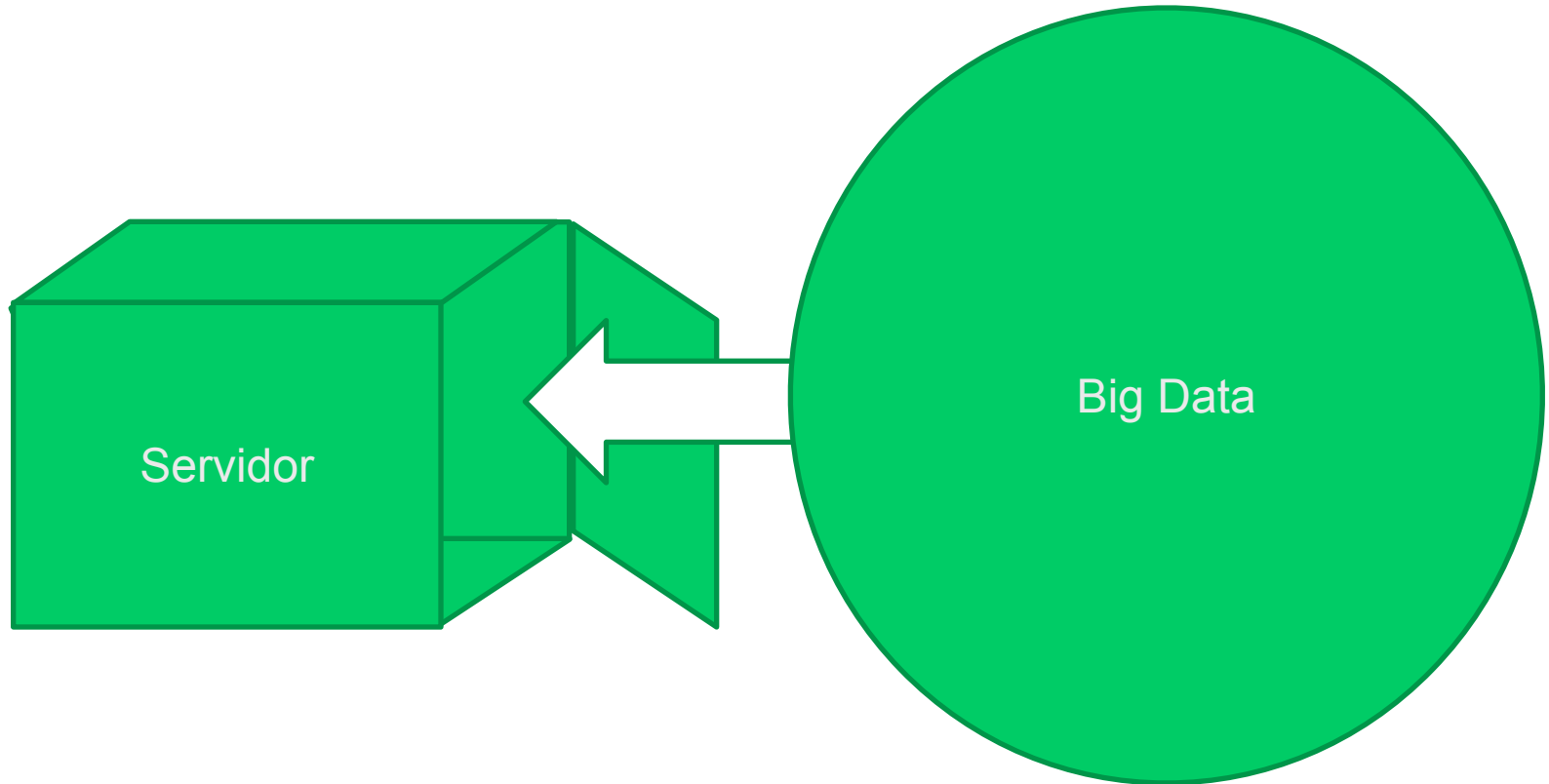


Abordagem Tradicional (estruturada)



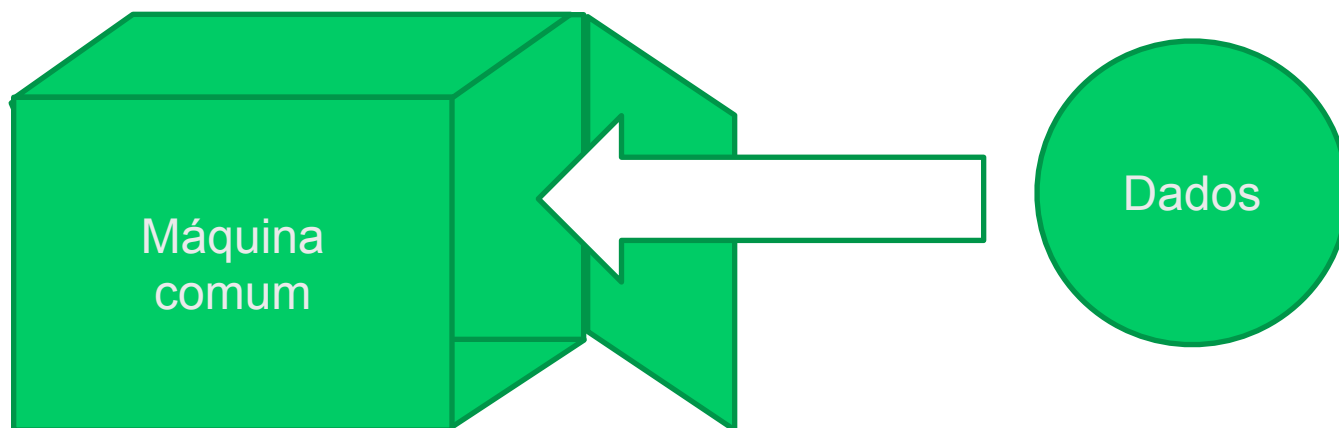


Abordagem Tradicional (estruturada)



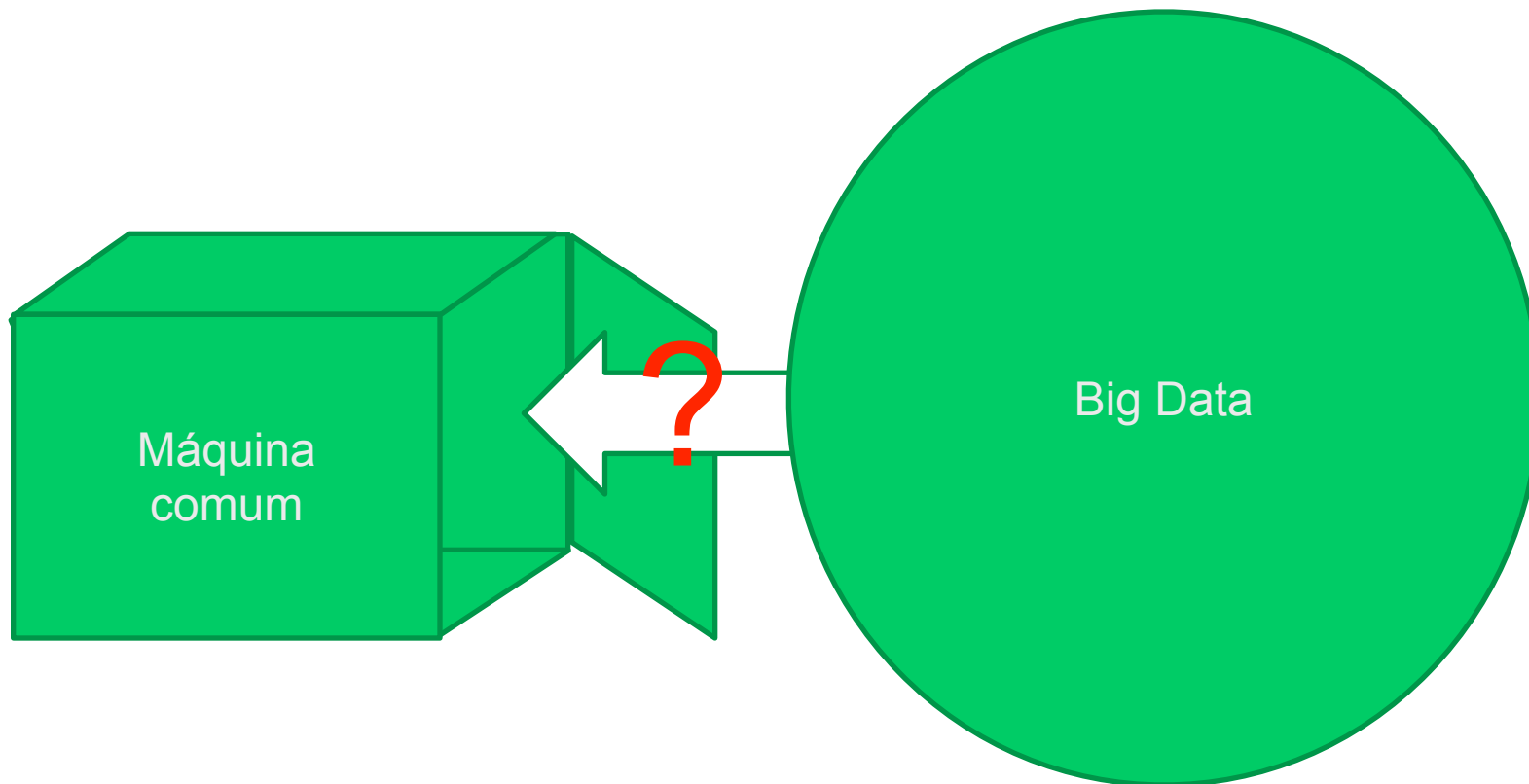


Abordagem Spark / Hadoop

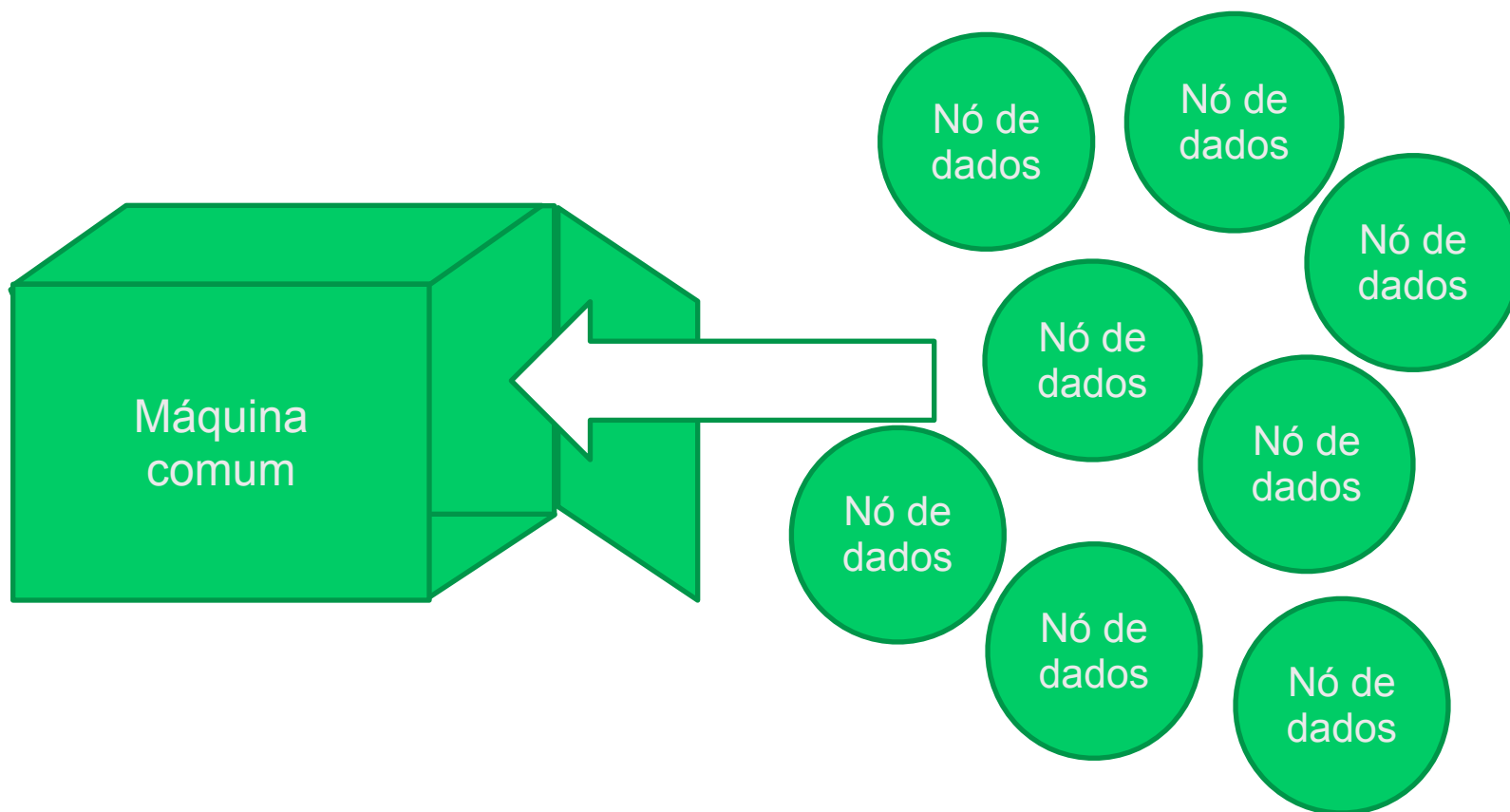




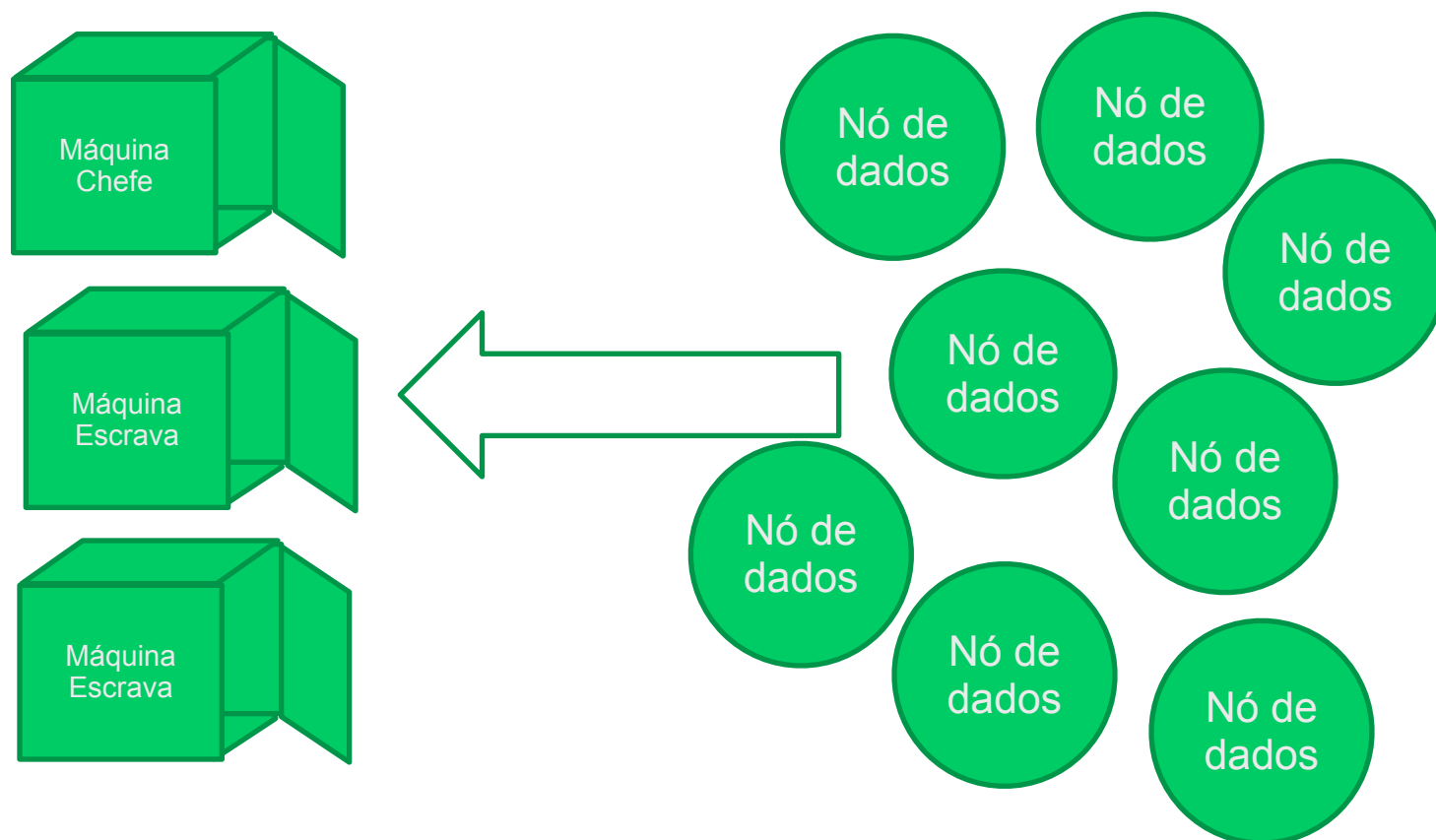
Abordagem Spark / Hadoop



Abordagem Spark / Hadoop



Abordagem Spark / Hadoop



Servidores da Yahoo! para Spark / Hadoop



O que é Programação Funcional para Big Data?



O que é Programação Funcional para Big Data?



- Origem: Google, 2004

O que é Programação Funcional para Big Data?



- **Origem:** Google, 2004
- Um modelo de programação simples
 - **Programação funcional**



O que é Programação Funcional para Big Data?

- **Origem:** Google, 2004
- Um modelo de programação simples
 - **Programação funcional**
- Processamento de dados em larga escala
 - *Clusters* formados por computadores “comuns”
 - Processamento distribuído
 - Disponibilidade (availability)
 - “**Esconde**” **complexidade** inerente ao paralelismo
 - Armazenamento, balanceamento, distribuição e replicação de dados, etc.



Apache Spark / Hadoop



- Implementações mais utilizadas
- Código aberto
- **HDFS**: sistema de arquivos distribuído

Conclusões





Conclusões

- **Programação Funcional para Big Data**
 - Oferece suporte a sistemas **NoSQL** e o processamento distribuído em geral
 - **Simplifica o processamento** de dados em **larga escala**
 - Permite que usuários mantenham o **foco no problema**, sem se preocupar com a **complexidade** inerente ao paralelismo
 - Armazenamento, balanceamento, distribuição e replicação de dados, etc.



Referências

- Material de aulas do Prof. Dr. Rob Gleasure, University College Cork, Ireland. <http://corvus2vm.ucc.ie/phd/rgleasure/rgleasure/index.html>
- NoSQL, w3resource. <http://www.w3resource.com/mongodb/nosql.php>
- Material de aulas do Prof. Dr. Xuanhua Shi, Huazhong University of Science and Technology, China. <http://grid.hust.edu.cn/xhshi/>



Referências (cont.)

- Material de aulas do Prof. Dr. Ray R. Larson, UC Berkeley School of Information, USA. <http://courses.ischool.berkeley.edu/i257/f15/>
- Jeffrey Dean and Sanjay Ghemawat. “MapReduce: Simplified Data Processing on Large Clusters”, OSDI'04: Sixth Symposium on Operating System Design and Implementation, San Francisco, CA, 2004.
- Apache Hadoop. <http://lucene.apache.org/hadoop/>
- <http://code.google.com/edu/parallel/mapreduce-tutorial.html>



NoSQL e o processamento de dados em larga escala (Parte 2)

Prof. Dr. Robson L. F. Cordeiro
robson@icmc.usp.br