

Processamento de Big Data



AVENTUREIROS DO BAIRRO PROIBIDO, FONTE: GOOGLE

Aula #9 - Transações em escala (BigData??)

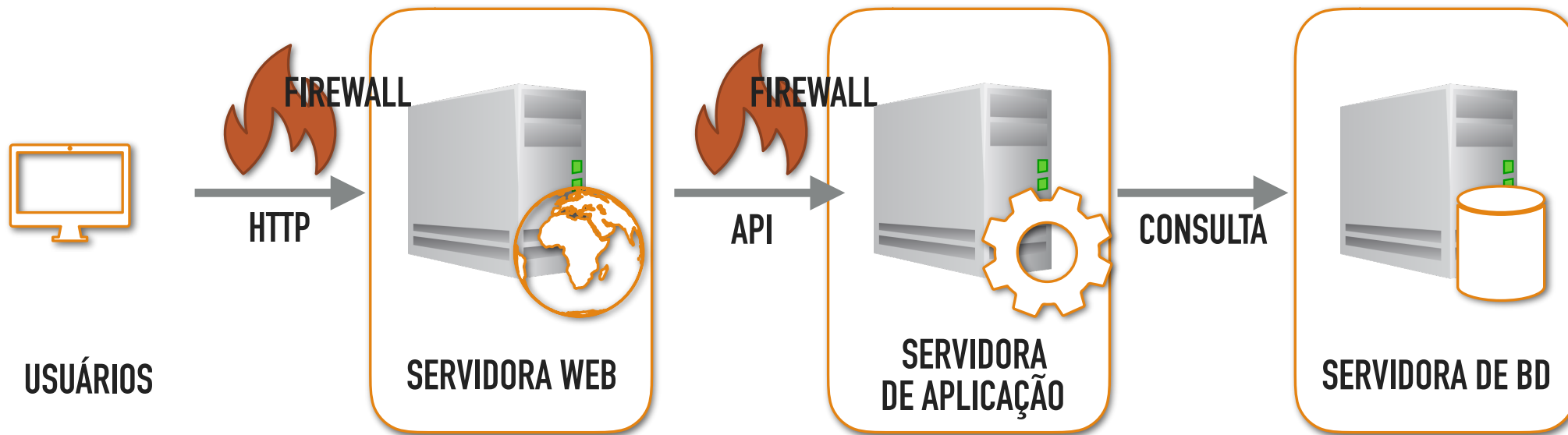
EDUARDO CUNHA DE ALMEIDA

Agenda

- Ideia 1: Comprar maquinas
- Ideia 2: Replicar o banco de dados
- Ideia 3: Cache de resultados
- Ideia 4: Stored Procedures
- Ideia 5: Particionar o banco de dados
- Ideia 6: Desistir do ACID

FONTE: FALOUTSOS E PAVLO, 2014

Ideia 1: Comprar máquina rápida

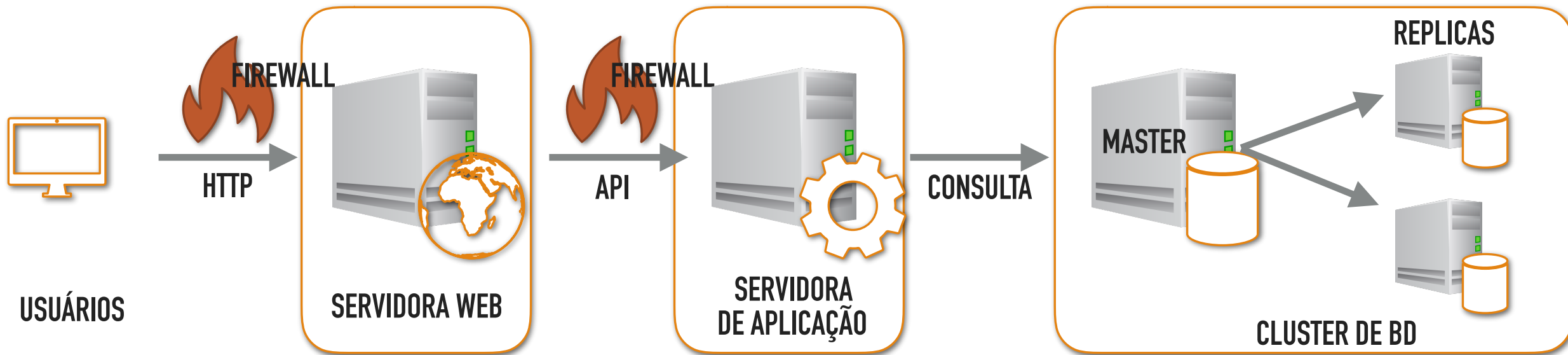


- HDD rápido
- DRAM
- CPUs
- SSD

(+) NÃO MUDA APLICAÇÃO
(+) MELHORIA IMEDIATA

(-) CARO!!!
(-) ÚNICO PONTO DE FALHA

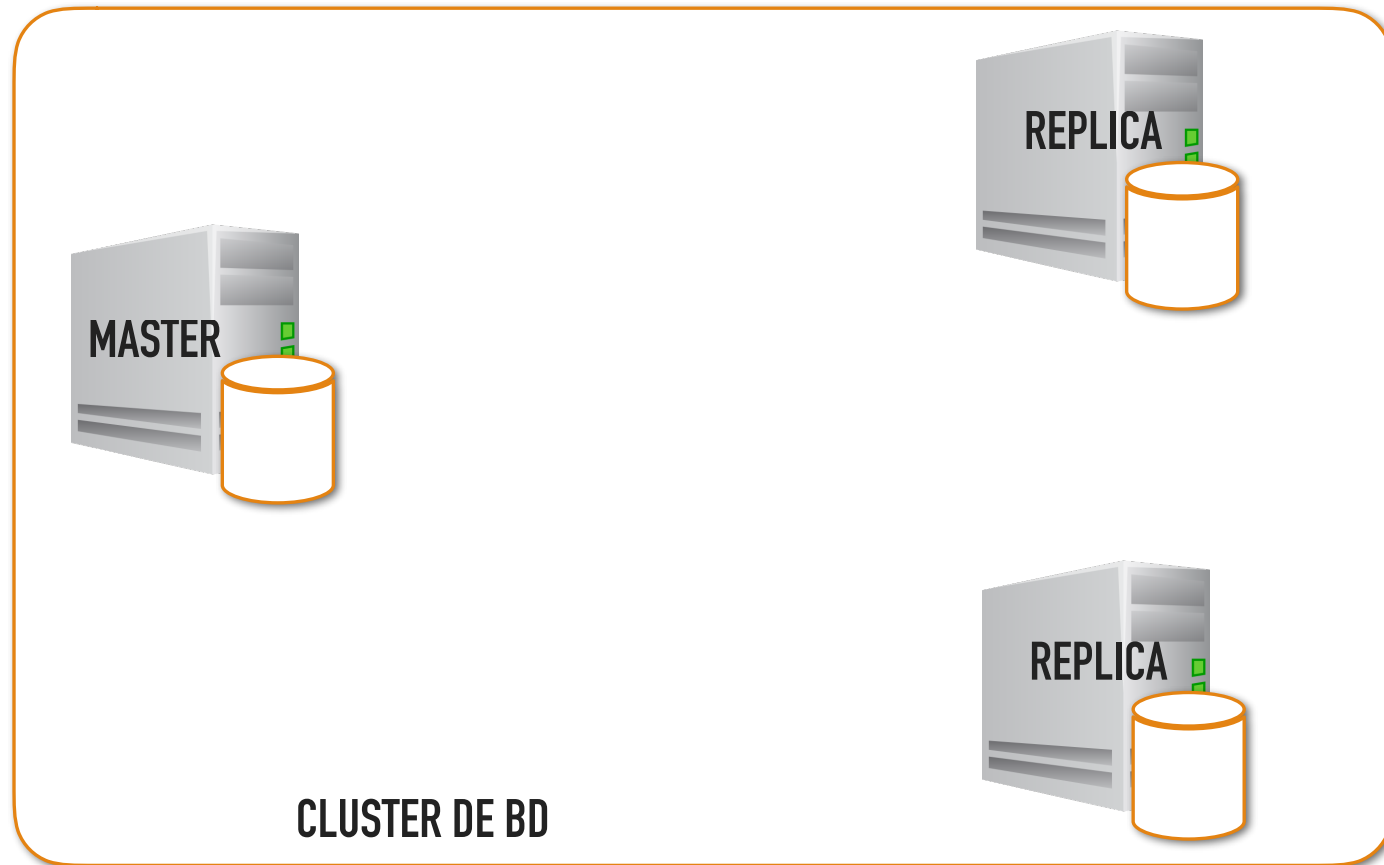
Ideia 2: Replicar o Banco de Dados



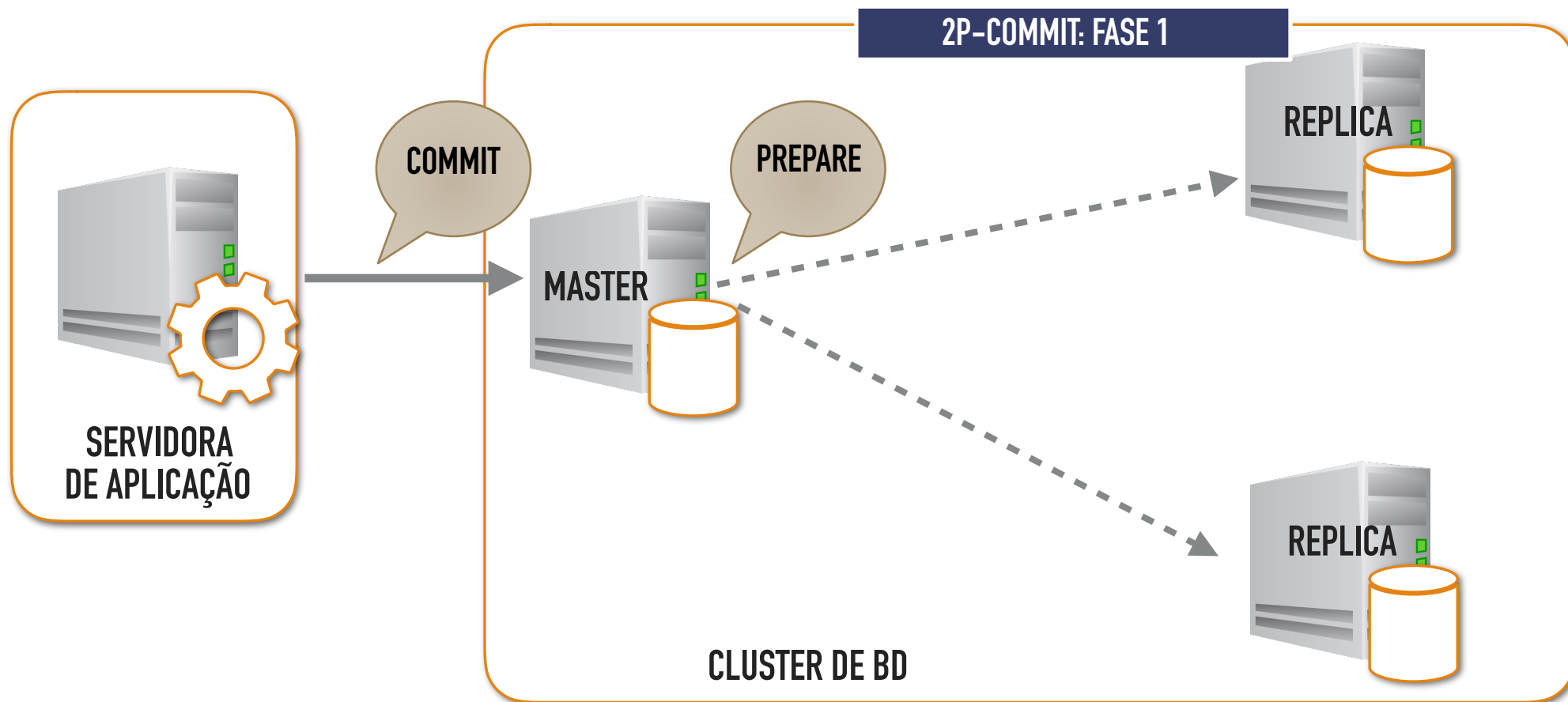
(+) NÃO MUDA APLICAÇÃO
(+) PARALELIZAÇÃO DE LEITURAS
(+) MELHORA TOLERANCIA A FALHAS

(-) CARO PRA MANTER REPLICAS E HW!!!
(-) LATÊNCIA DE ESCRITAS = NODO MAIS LENTO

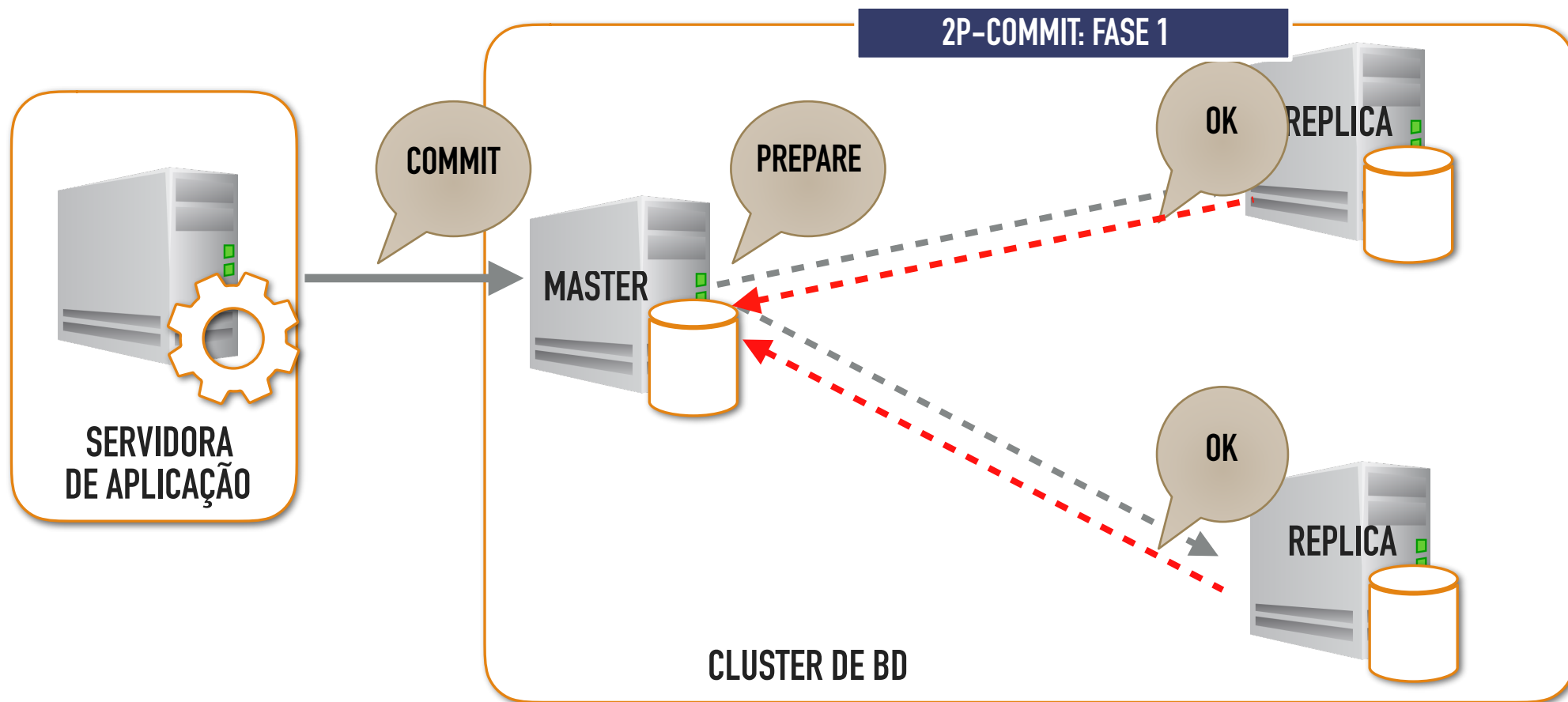
Ideia 2: Replicar o Banco de Dados



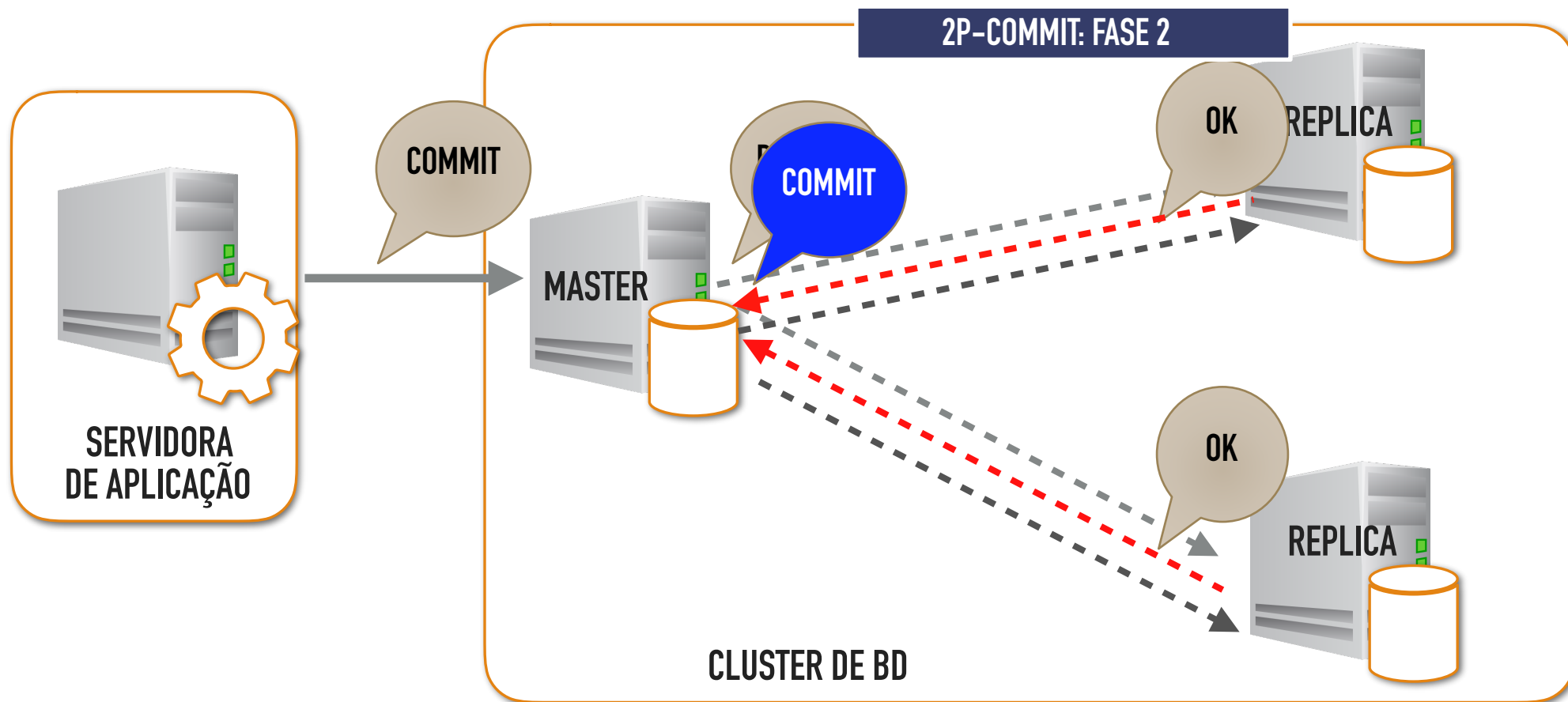
Ideia 2: Replicar o Banco de Dados



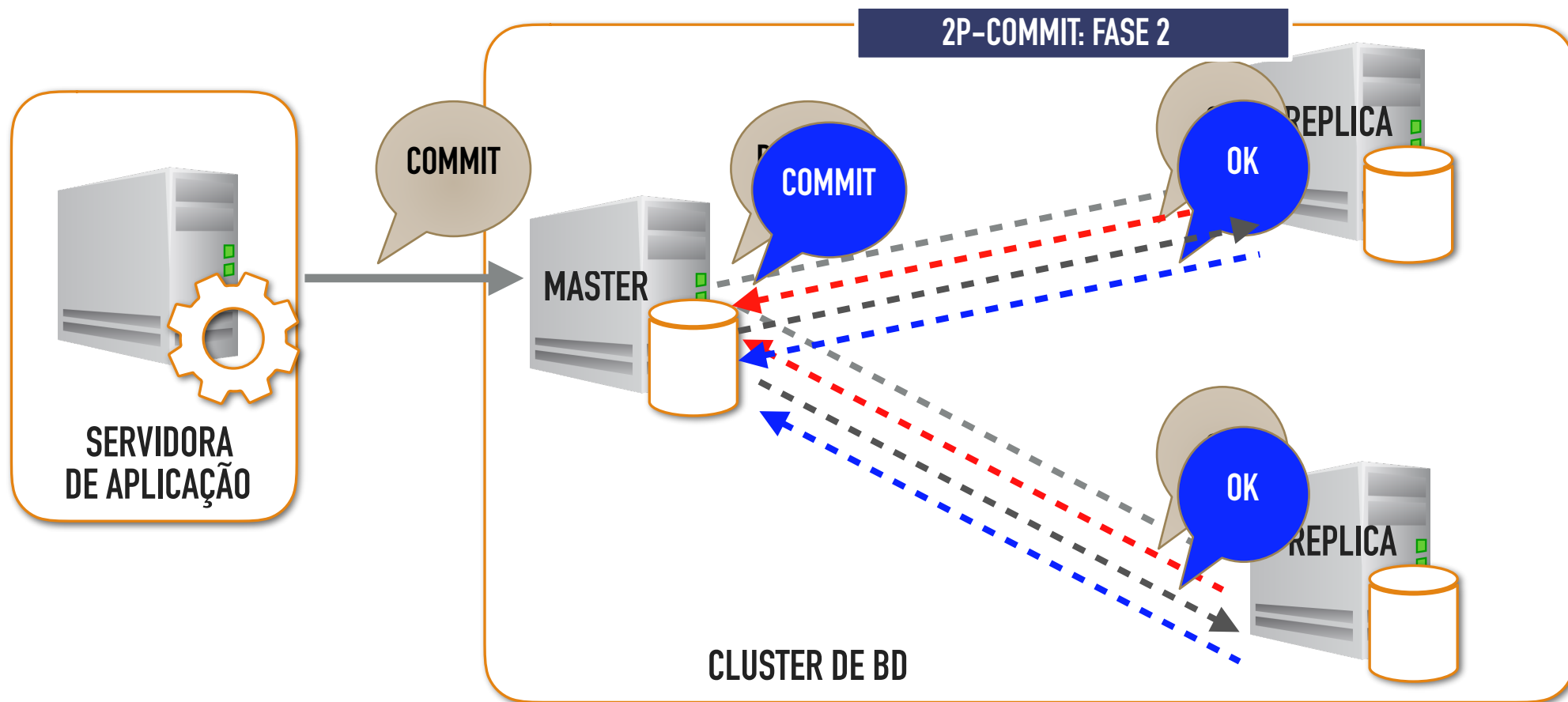
Ideia 2: Replicar o Banco de Dados



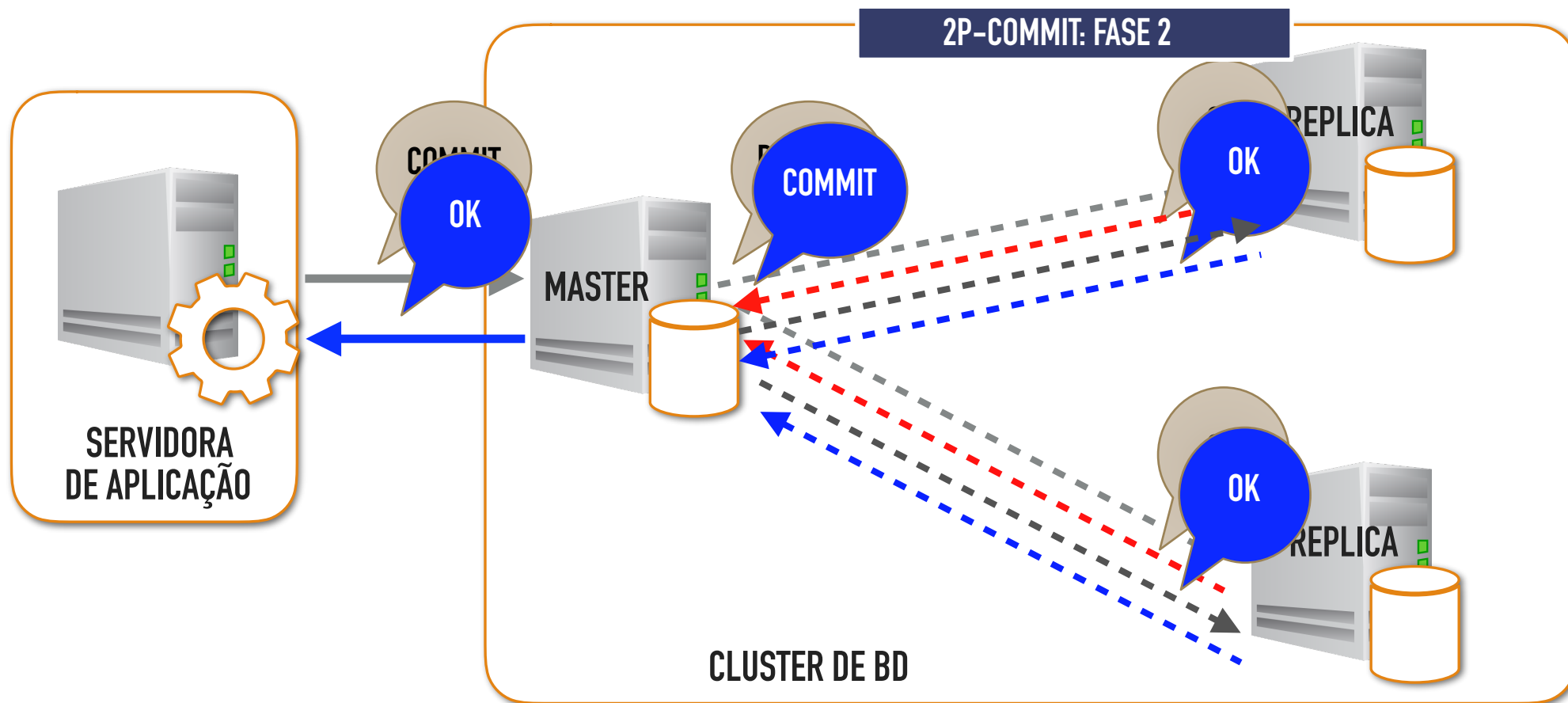
Ideia 2: Replicar o Banco de Dados



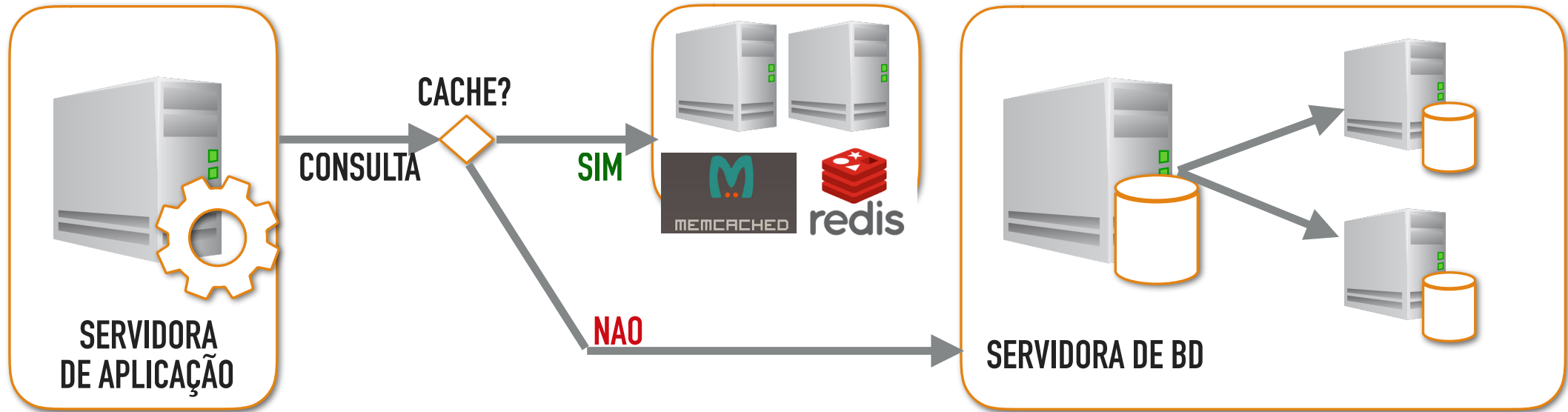
Ideia 2: Replicar o Banco de Dados



Ideia 2: Replicar o Banco de Dados



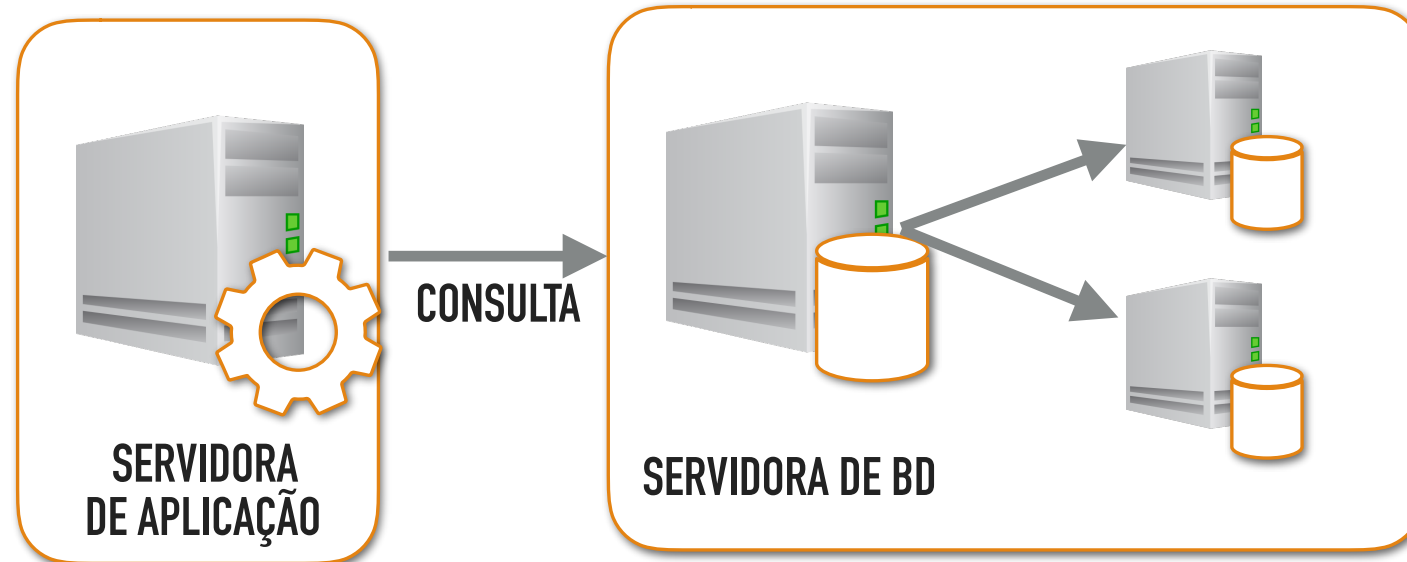
Ideia 3: Cache de resultados



(+) REDUZ CARGA NO BD
(+) API RÁPIDA

(-) ROTEAMENTO EXTRA DE MENSAGENS
(-) APLICAÇÃO DEVE SER ALTERADA
(-) PROBLEMAS COM VAZÃO DE ESCRITAS

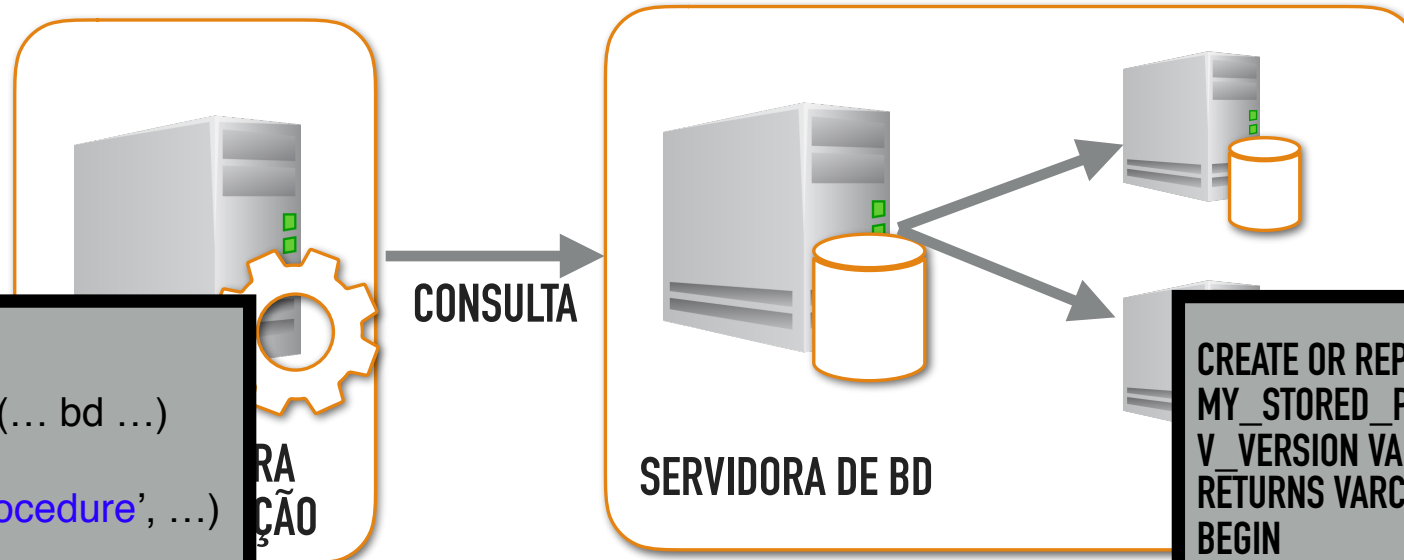
Ideia 4: Stored procedures



(+) REDUZ MENSAGENS DE REDE
(+) REDUZ CONTENÇÃO
(+) MODULARIZAÇÃO

(-) LÓGICA DA APLICAÇÃO EM VÁRIOS LOCAIS
(-) API NÃO PADRÃO
(-) APRISIONAMENTO TECNOLÓGICO (LOCK-IN)

Ideia 4: Stored procedures



try:

```
conn = MySQLdb.connect (... bd ...)
c1=conn.cursor()
c1.callproc('My_Stored_Procedure', ...)
c1.close()
```

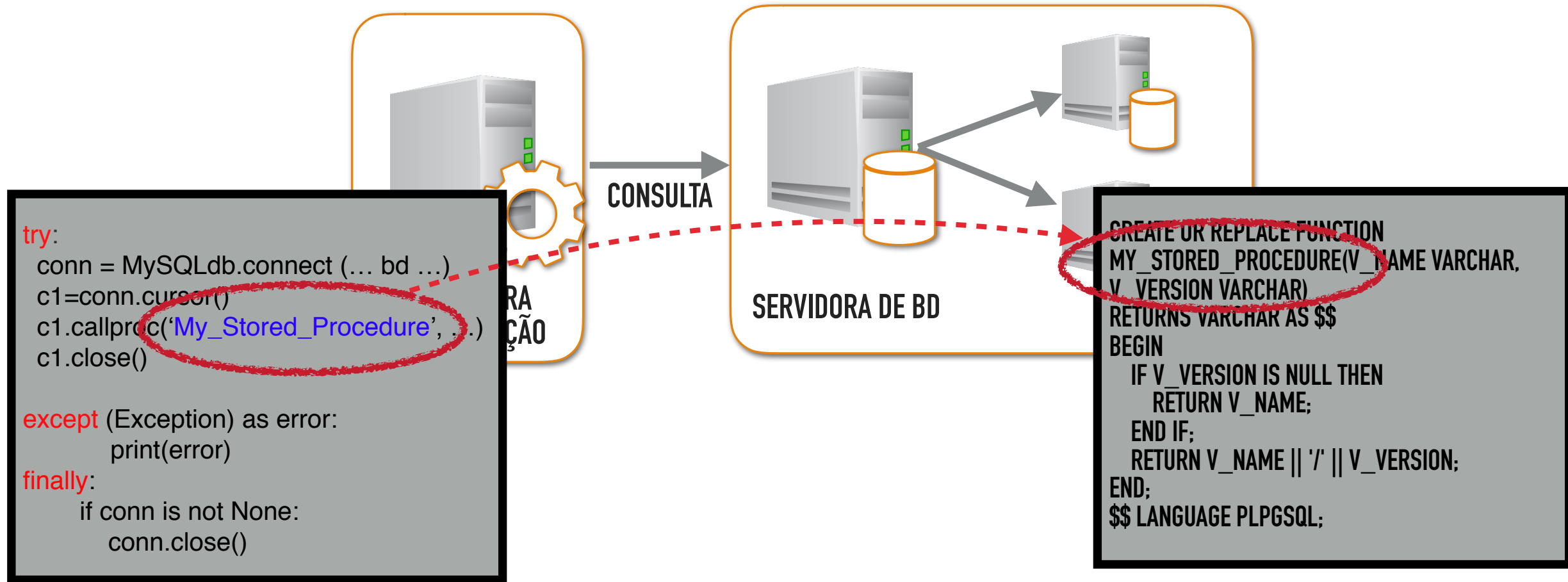
except (Exception) as error:
 print(error)

finally:

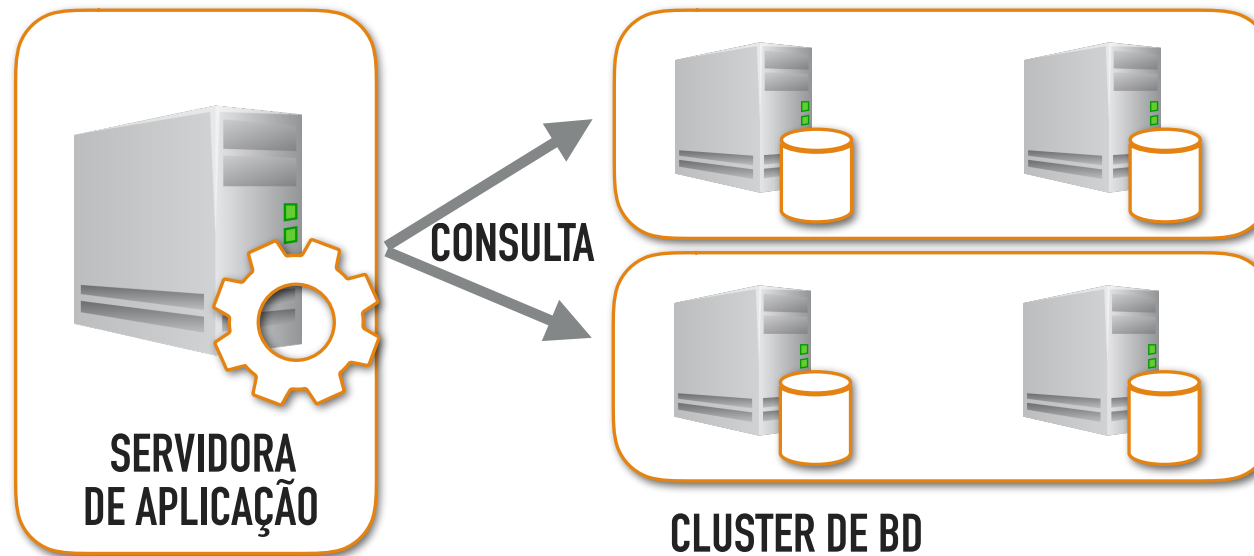
```
if conn is not None:
    conn.close()
```

```
CREATE OR REPLACE FUNCTION
MY_STORED_PROCEDURE(V_NAME VARCHAR,
V_VERSION VARCHAR)
RETURNS VARCHAR AS $$
BEGIN
    IF V_VERSION IS NULL THEN
        RETURN V_NAME;
    END IF;
    RETURN V_NAME || '/' || V_VERSION;
END;
$$ LANGUAGE PLPGSQL;
```

Ideia 4: Stored procedures



Ideia 5: Particionar o banco de dados

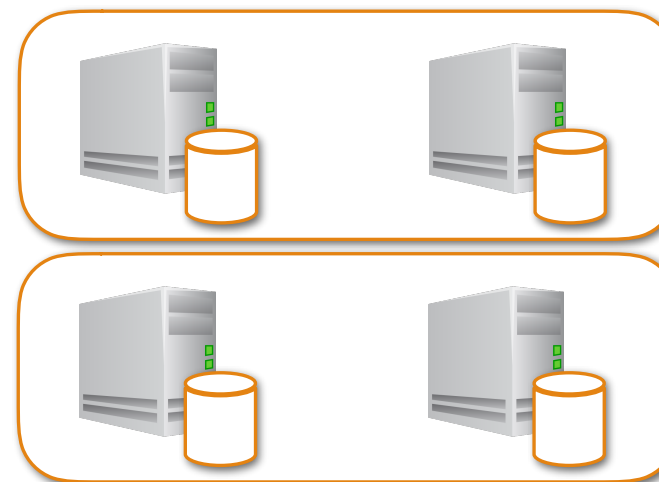


(+) PARALELIZAR TODAS AS OPERAÇÕES
(+) FÁCIL ADICIONAR/REMOVER HW

(-) MAIORIA DOS BDS NÃO FAZEM
(-) JUNÇÕES SÃO CARAS
(-) PARTICIONAMENTO NÃO É TRIVIAL

<u>cliente.</u> <u>nome_c</u>	<u>cesta.</u> <u>dt_compra</u>	<u>compra.</u> <u>isbn</u>
André	10/10/17	1234
Cido	11/10/17	1121
Siloé	20/10/17	1234
Siloé	20/10/17	2222
Siloé	20/10/17	2223
Juliana	20/10/17	1234
Juliana	20/10/17	2224
Rosana	22/10/17	2225
Rosana	22/10/17	1234
Rosana	22/10/17	1121
Rosana	22/10/17	2222

Tipos de particionamento



CLUSTER DE BD

TABELA1**TABELA2**

<u>cliente. nome_c</u>	<u>cesta. dt_compra</u>	<u>compra. isbn</u>
André	11/10/17	1234
Cido	11/10/17	1121
Siloé	20/10/17	1234
Siloé	20/10/17	2222
Siloé	20/10/17	2223
Juliana	20/10/17	1234
Juliana	20/10/17	2224
Rosana	22/10/17	2225
Rosana	22/10/17	1234
Rosana	22/10/17	1121
Rosana	22/10/17	2222

Básico (manual)

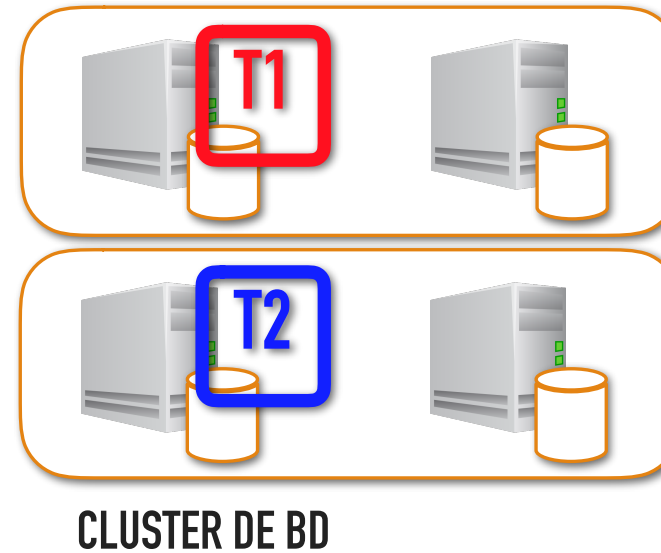
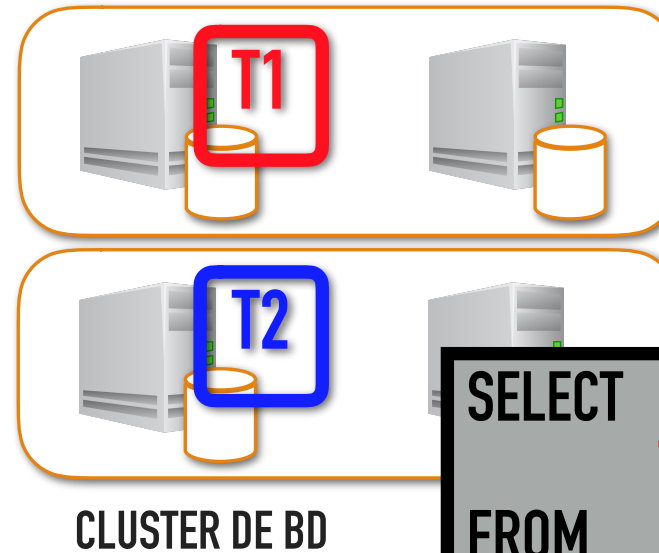


TABELA1

TABELA2

<u>id</u>	<u>cliente. nome_c</u>	<u>cesta. dt_compra</u>	<u>compra. isbn</u>	<u>id</u>
1	André	11/10/17	1234	1
2	Cido	11/10/17	1121	2
3	Siloé	20/10/17	1234	3
4	Siloé	20/10/17	2222	4
5	Siloé	20/10/17	2223	5
6	Juliana	20/10/17	1234	6
7	Juliana	20/10/17	2224	7
8	Rosana	22/10/17	2225	8
9	Rosana	22/10/17	1234	9
10	Rosana	22/10/17	1121	10
11	Rosana	22/10/17	2222	11

Básico (manual)



```

SELECT  T1.*, T2.*
FROM    T1, T2
WHERE   T1.ID = T2.ID;

```

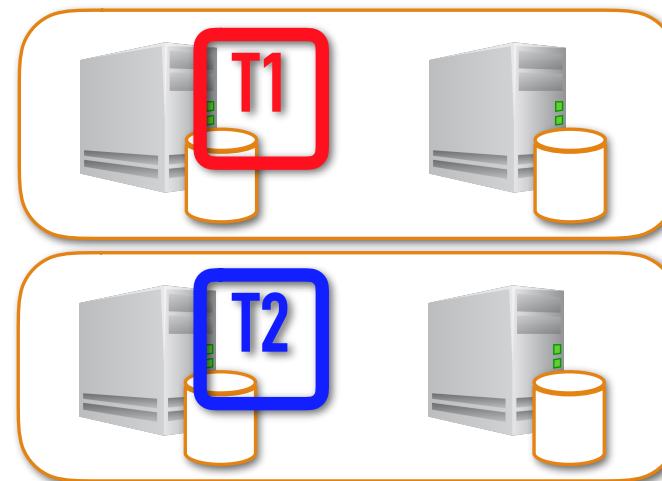
TABELA1

<u>id</u>	<u>cliente.</u> <u>nome_c</u>	<u>cesta.</u> <u>dt_compra</u>	<u>compra.</u> <u>isbn</u>	<u>id</u>
1	André	11/10/17	1234	1
2	Cido	11/10/17	1121	2
3	Siloé	20/10/17	1234	3
4	Siloé	20/10/17	2222	4
5	Siloé	20/10/17	2223	5
6	Juliana	20/10/17	1234	6
7	Juliana	20/10/17	2224	7

TABELA2

<u>id</u>	<u>cliente.</u> <u>nome_c</u>	<u>cesta.</u> <u>dt_compra</u>	<u>compra.</u> <u>isbn</u>	<u>id</u>
8	Rosana	22/10/17	2225	8
9	Rosana	22/10/17	1234	9
10	Rosana	22/10/17	1121	10
11	Rosana	22/10/17	2222	11

Básico (manual UNION ALL



CLUSTER DE BD

TABELA1

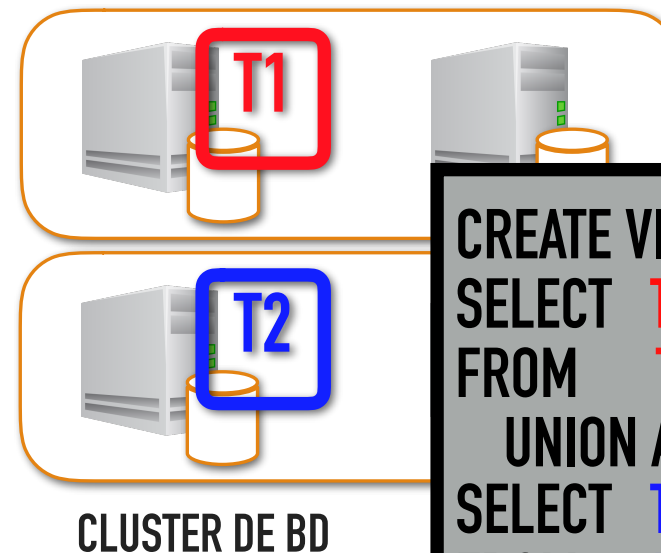
<u>id</u>	<u>cliente.</u> <u>nome_c</u>	<u>cesta.</u> <u>dt_compra</u>	<u>compra.</u> <u>isbn</u>	<u>id</u>
1	André	11/10/17	1234	1
2	Cido	11/10/17	1121	2
3	Siloé	20/10/17	1234	3
4	Siloé	20/10/17	2222	4
5	Siloé	20/10/17	2223	5
6	Juliana	20/10/17	1234	6
7	Juliana	20/10/17	2224	7

TABELA2

<u>id</u>	<u>cliente.</u> <u>nome_c</u>	<u>cesta.</u> <u>dt_compra</u>	<u>compra.</u> <u>isbn</u>	<u>id</u>
8	Rosana	22/10/17	2225	8
9	Rosana	22/10/17	1234	9
10	Rosana	22/10/17	1121	10
11	Rosana	22/10/17	2222	11

Básico (manual UNION ALL

TABELA



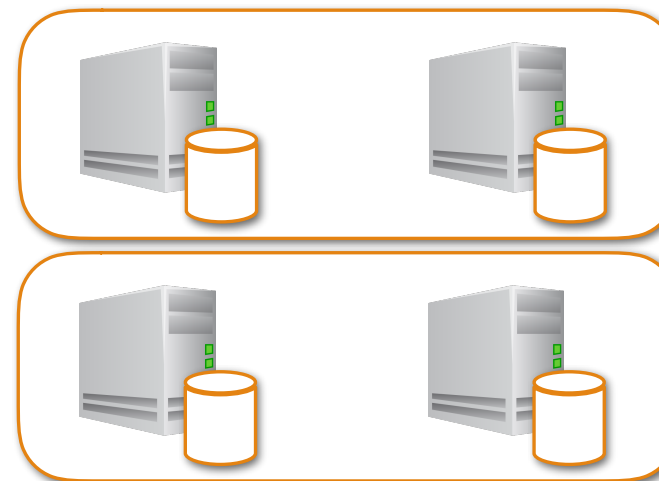
```
CREATE VIEW TABELA(  
SELECT TABELA1.*  
FROM TABELA1  
UNION ALL  
SELECT TABELA2.*  
FROM TABELA2;)
```

```
SELECT TABELA.*  
FROM TABELA;
```

<u>cliente. nome_c</u>	<u>cesta. dt_compra</u>	<u>compra. isbn</u>
André	11/10/17	1234
Cido	11/10/17	1121
Siloé	20/10/17	1234
Siloé	20/10/17	2222
Siloé	20/10/17	2223
Juliana	20/10/17	1234
Juliana	20/10/17	2224
Rosana	22/10/17	2225
Rosana	22/10/17	1234
Rosana	22/10/17	1121
Rosana	22/10/17	2222

CHAVE DE PARTIC.

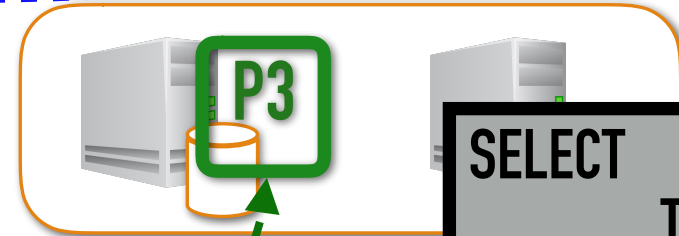
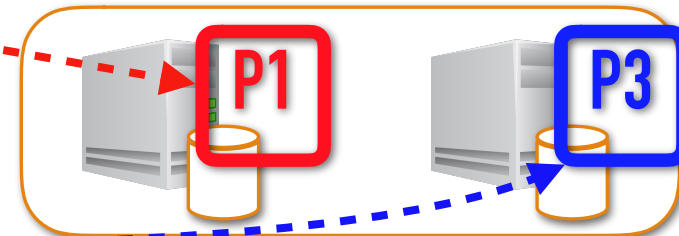
Partic. Horizontal



CLUSTER DE BD

Partic. Horizontal

	<u>cliente.</u> <u>nome_c</u>	<u>cesta.</u> <u>dt_compra</u>	<u>compra.</u> <u>isbn</u>
P1	Andre	11/10/17	1234
	Cido	11/10/17	1121
P2	Siloé	20/10/17	1234
	Siloé	20/10/17	2222
	Siloé	20/10/17	2223
	Juliana	20/10/17	1234
	Juliana	20/10/17	2224
P3	Rosana	22/10/17	2225
	Rosana	22/10/17	1234
	Rosana	22/10/17	1121
	Rosana	22/10/17	2222

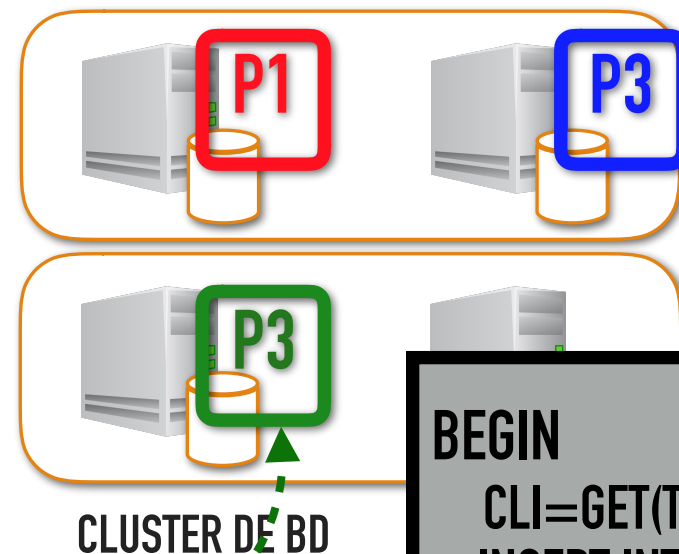


CLUSTER DE BD

```
SELECT
  TABELA.*
FROM
  TABELA;
```

<u>cliente. nome_c</u>	<u>cesta. dt_compra</u>	<u>compra. isbn</u>
André	11/10/17	1234
Cido	11/10/17	1121
Siloé	20/10/17	1234
Siloé	20/10/17	2222
Siloé	20/10/17	2223
Juliana	20/10/17	1234
Juliana	20/10/17	2224
Rosana	22/10/17	2225
Rosana	22/10/17	1234
Rosana	22/10/17	1121
Rosana	22/10/17	2222
P3 Rosana	22/10/17	2223
P3 Rosana	22/10/17	2224

Partic. Horizontal

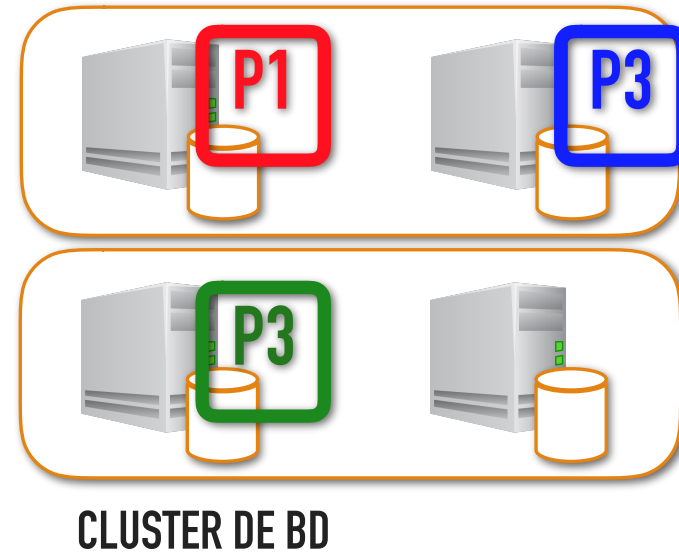


BEGIN

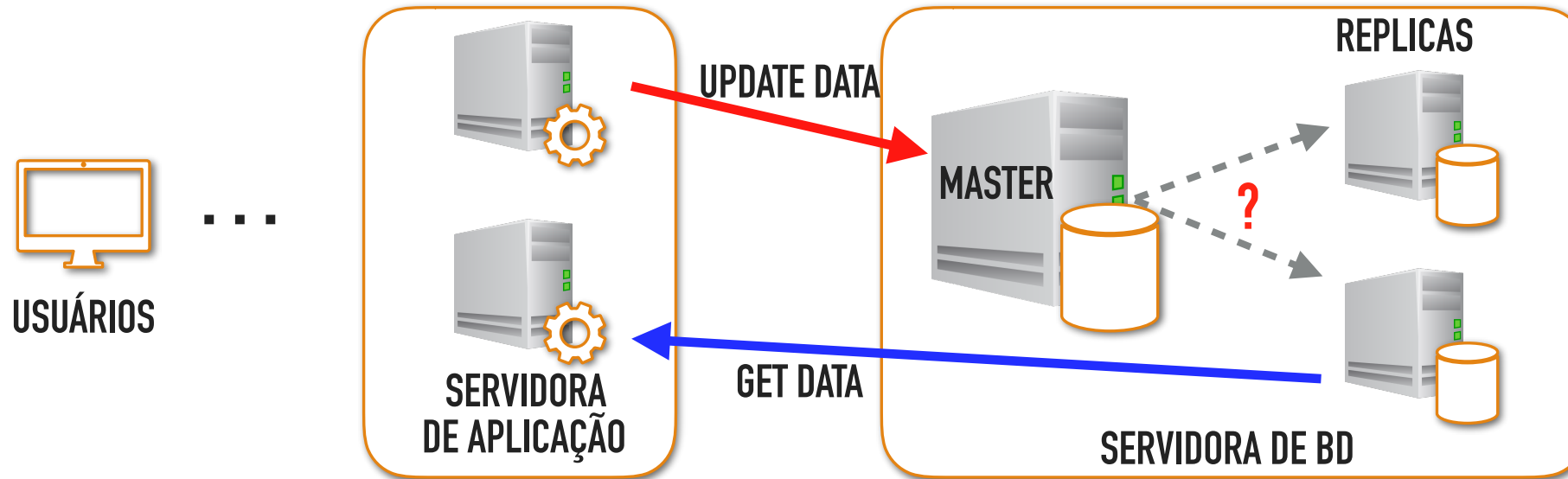
```
CLI=GET(T.NOME_C="ROSANA");
INSERT INTO VENDA(CLI,"22/20/17",2223);
INSERT INTO VENDA(CLI,"22/20/17",2224);
COMMIT
```

P1	P2	P3
<u>cliente.</u> <u>nome_c</u>	<u>cesta.</u> <u>dt_compra</u>	<u>compra.</u> <u>isbn</u>
André	11/10/17	1234
Cido	11/10/17	1121
Siloé	20/10/17	1234
Siloé	20/10/17	2222
Siloé	20/10/17	2223
Juliana	20/10/17	1234
Juliana	20/10/17	2224
Rosana	22/10/17	2225
Rosana	22/10/17	1234
Rosana	22/10/17	1121
Rosana	22/10/17	2222

Partic. Vertical (colunar)



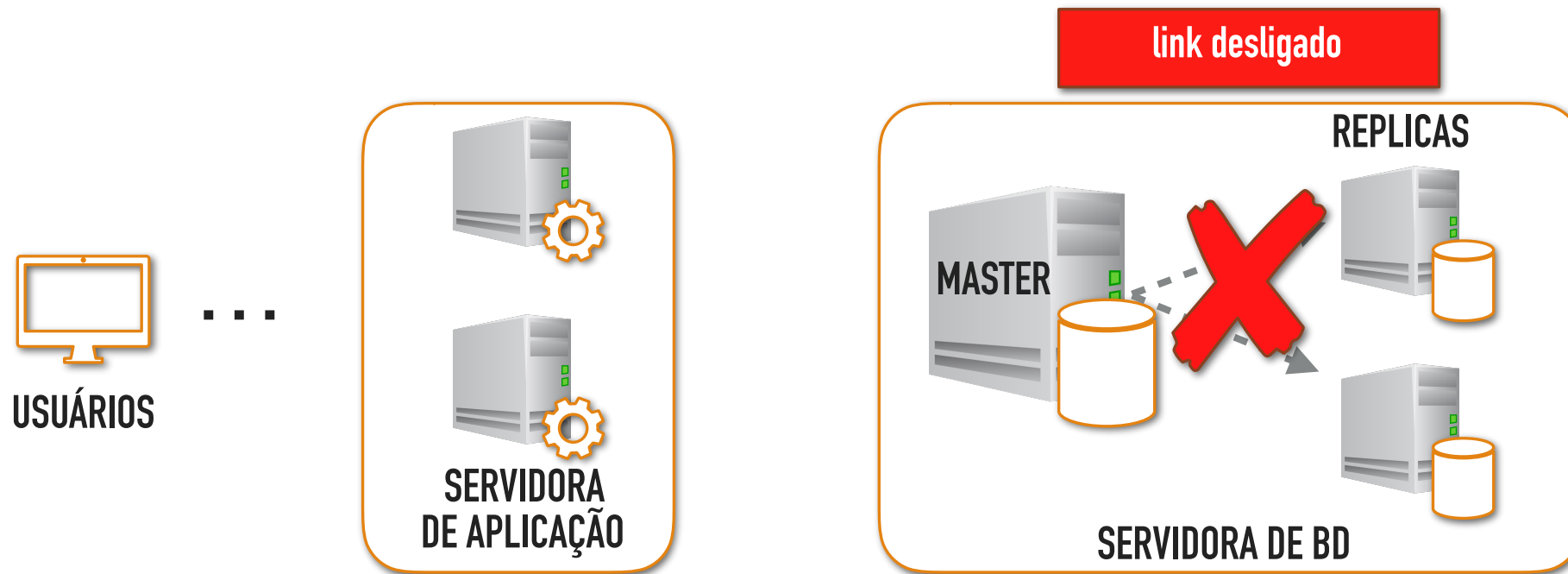
Ideia 6: Desistir do ACID



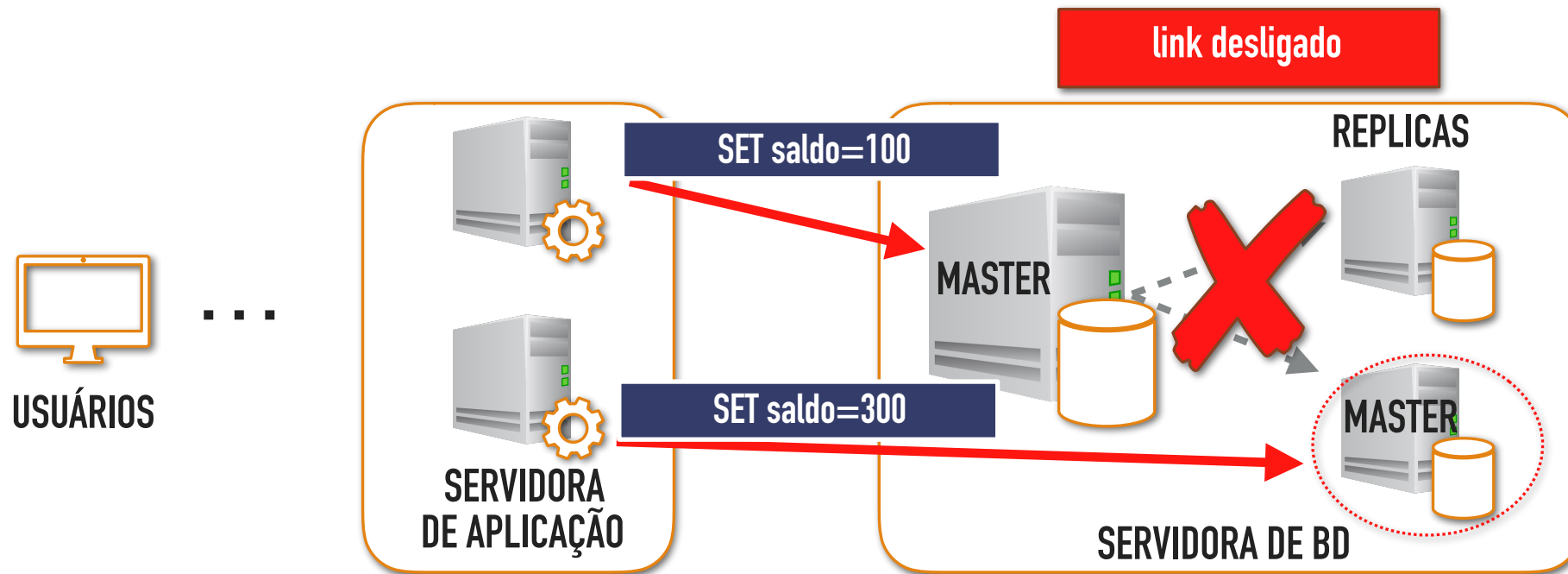
(+) ESCALABILIDADE
(+) PARALELISMO

(-) PÉSSIMA IDEIA
(-) POSSÍVEL PERDA DE ATUALIZAÇÕES
(-) DISCIPLINA DOS PROGRAMADORES

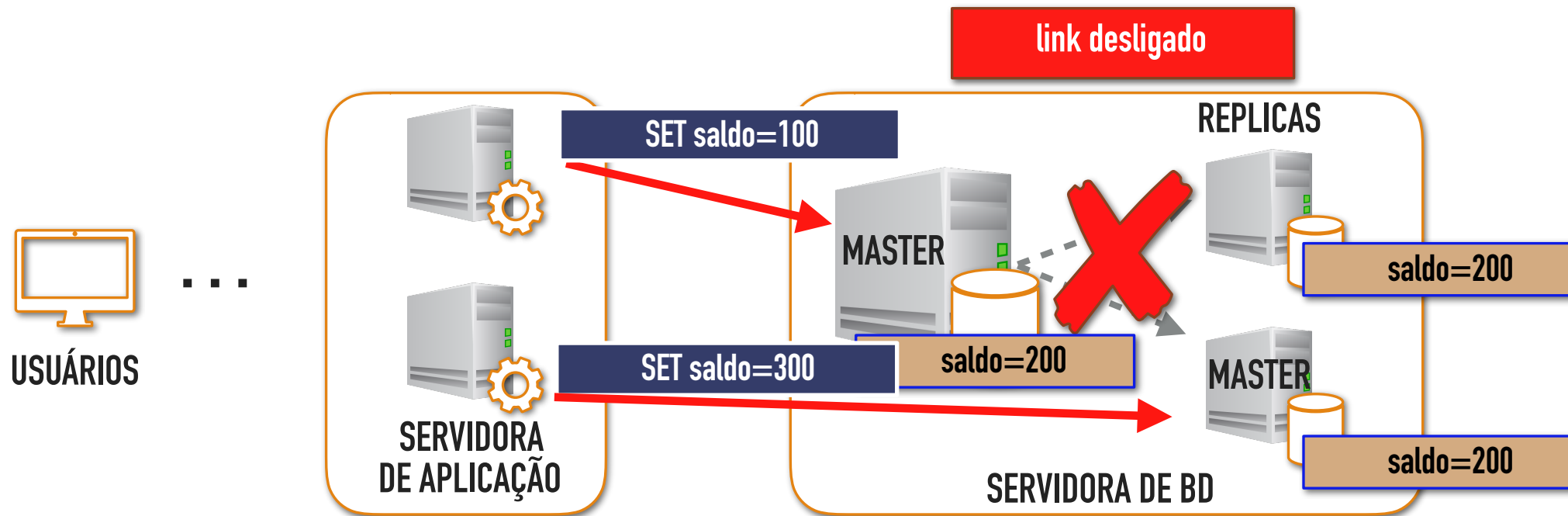
Ideia 6: Desistir do ACID



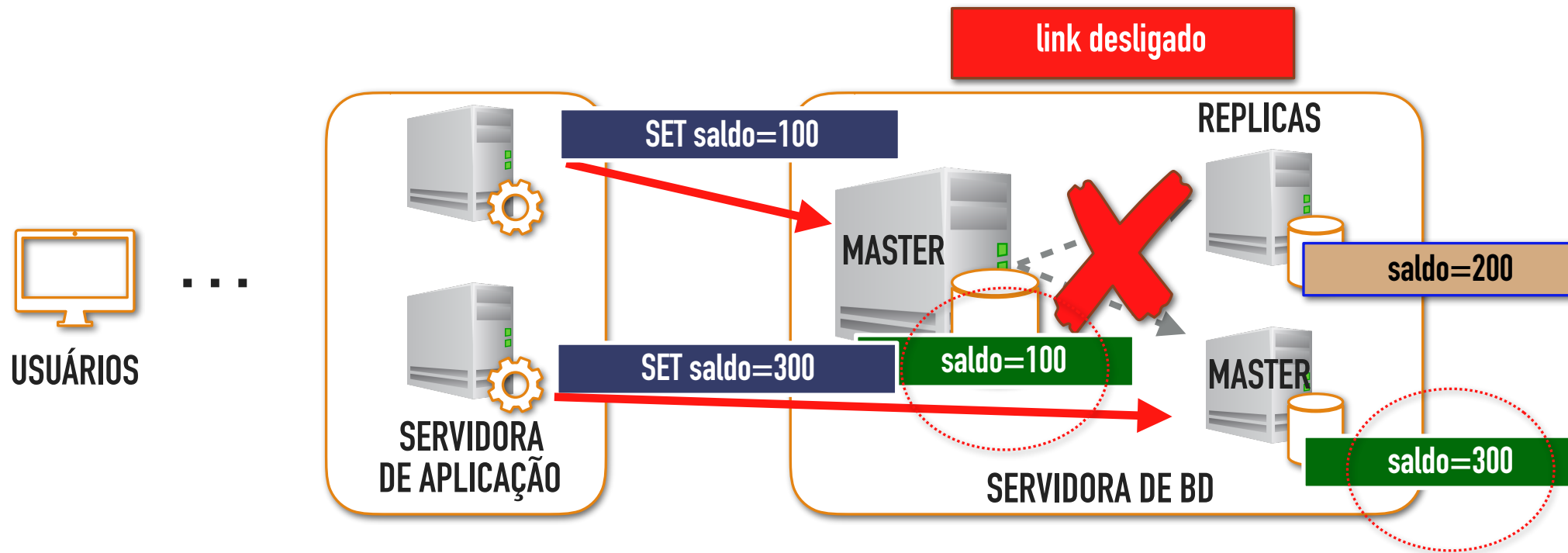
Ideia 6: Desistir do ACID



Ideia 6: Desistir do ACID



Ideia 6: Desistir do ACID



Processamento de Big Data



AVENTUREIROS DO BAIRRO PROIBIDO, FONTE: GOOGLE

Aula #9 - Transações em escala (BigData??)

EDUARDO CUNHA DE ALMEIDA