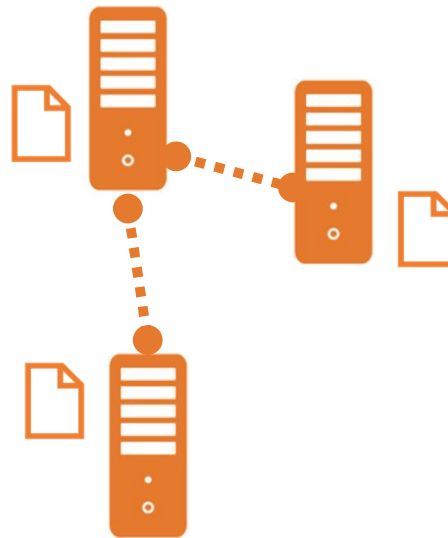


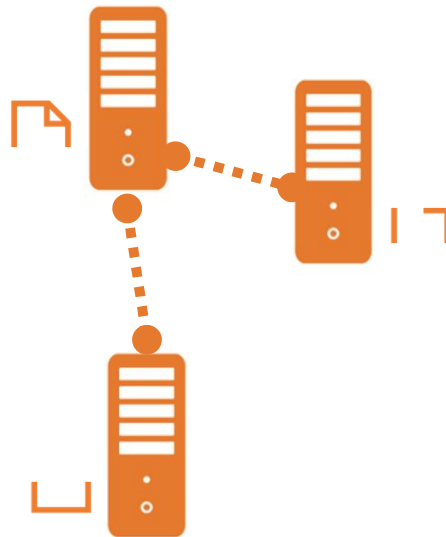
# Replicação

- Dados são copiados para diferentes nós



# Partição

- ◆ Dados são divididos



# Porque partição?

- Dividindo os dados, temos consultas otimizadas!

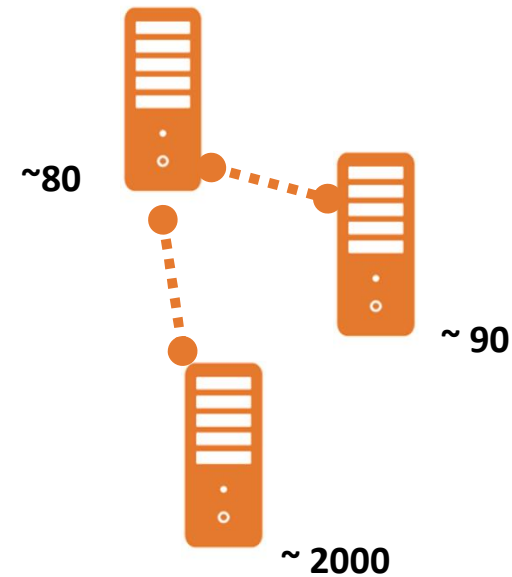


**PROF.  
FERNANDO  
AMARAL**  
[www.datascientist.com.br](http://www.datascientist.com.br)

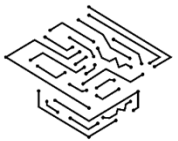
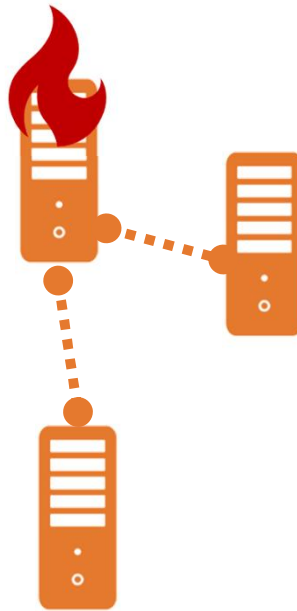
# Sharding

- Um tipo de particionamento: horizontal

Data
01/12/1980
...
21/08/1997
...
30/04/20012
...
26/12/2021



# Failover



# Pipeline



# Batch VS Streaming VS Interativo



Batch: dados são coletados (extraídos, tratados, armazenados) e processados. Pode levar horas, semanas, meses



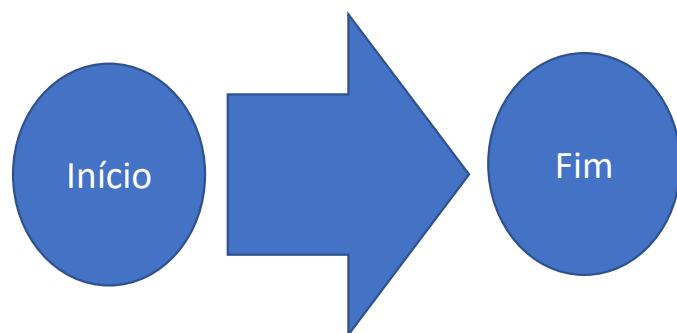
Streaming: dados processados em tempo real, a medida que são produzidos. Saída em tempo real ou próximo ao tempo real. Processamento constante (Fluxo)



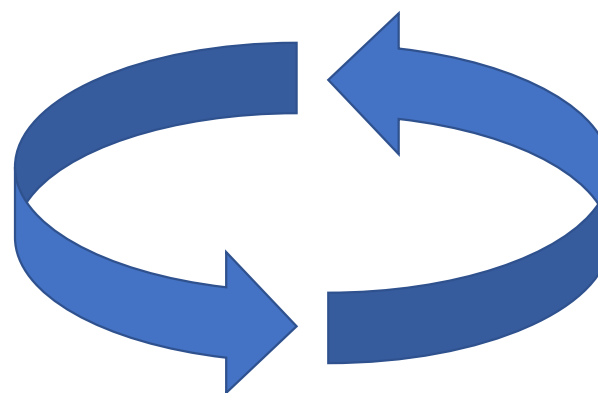
Interativo (Ad Hoc): Interface onde processamento é requerido e o resultado é analisado

# Batch VS Streaming

Batch



Streaming





# Batch VS Streaming VS Interativo

Batch:	Sistema de busca de produtos de varejo on-line: Milhões de termos pesquisados por dia Diariamente, os dados são processados para criar um ranking das palavras mais buscadas
Streaming:	Cercamento Eletrônico de Veículos
Interativo (Ad Hoc):	Rodar uma consulta para ver as vendas do dia de determinado produto



# ETL e ELT

ETL: EXTRACT,  
TRANSFORM  
AND LOAD

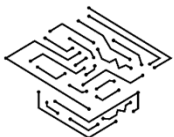
ELT: EXTRACT,  
LOAD AND  
TRANSFORM



**PROF.  
FERNANDO  
AMARAL**  
[www.datascientist.com.br](http://www.datascientist.com.br)

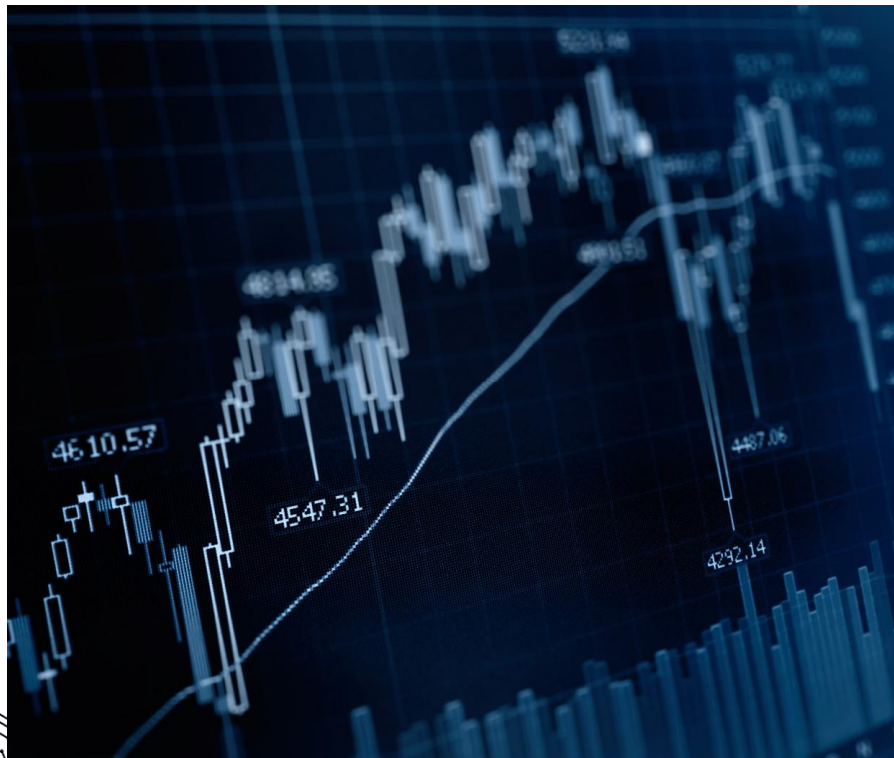
# ETL

- Fontes de Dados digitais nas organizações eram essencialmente estruturados (Relacionais)
- Custo alto de armazenamento



**PROF.  
FERNANDO  
AMARAL**  
[www.datascientist.com.br](http://www.datascientist.com.br)

# ELT



- Custo de armazenamento muito menor
- Fontes de Dados digitais são também não estruturados e semiestruturados