

## Data Quote

“Se a estatística é chata, você tem os números errados”

Edward Tufte

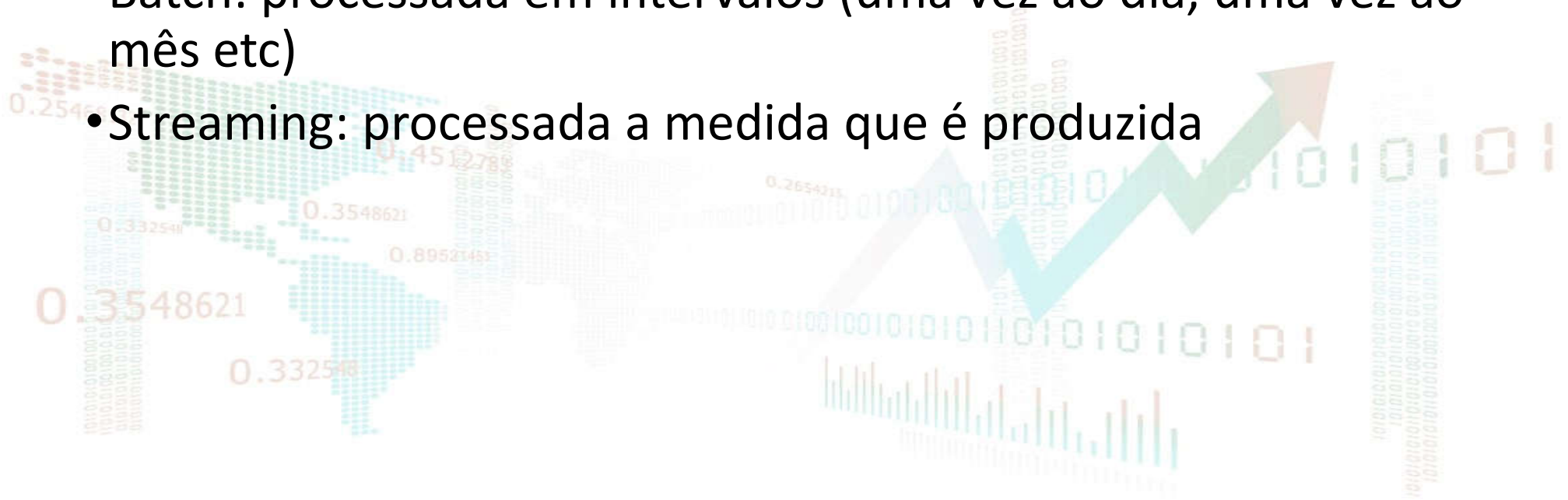
# Formação Analista de BI

ETL II



# Batch vs Streaming

- Batch: processada em intervalos (uma vez ao dia, uma vez ao mês etc)
- Streaming: processada a medida que é produzida



# Batch

- Conjunto de processos, normalmente para carregar um único Data Mart



# Latência

- Tempo em que o dado deve estar disponível para o usuário
- Influi diretamente em:
  - Arquitetura (servidores, redes etc)
  - Custos

# Latência

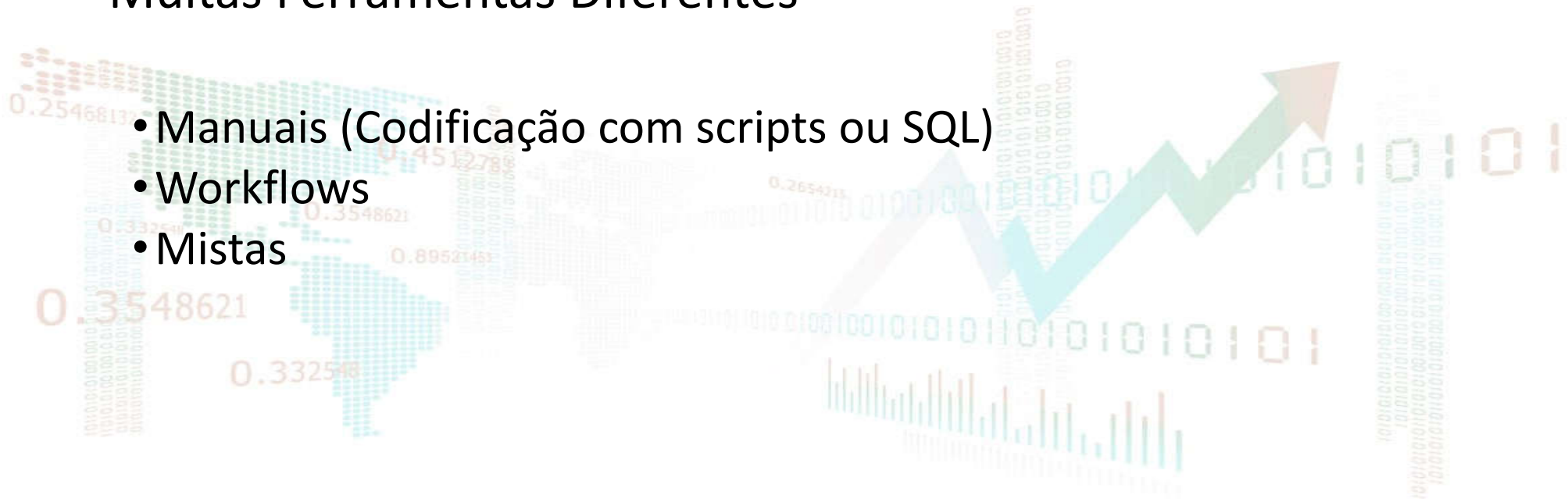
- Informações Críticas precisam de baixa latência
- Informações não Críticas podem estar disponíveis com latência mais alta



# Ferramentas de ETL

- Muitas Ferramentas Diferentes

- Manuais (Codificação com scripts ou SQL)
- Workflows
- Mistas



# Recuperação

- Reinicia o processo todo
- Continua do ponto de falha

