Formação Engenheiro de Dados

Conceitos Parte II

PROF. FERNANDO AMARAL
www.datascientist.com.br

Change Data Capture

- Capturar mudanças nos dados transacionais a partir do log de transação
  - Incremental: carga de mudanças
    - Síncrono: captura a medida que as mudanças ocorrem
    - Assíncronos: captura em intervalos
  - Bulk: carga de todos os dados

PROF. FERNANDO AMARAL
www.datascientist.com.br

Disponibilidade

- Riscos:
  - Falhas de Software
  - Falhas de Hardware
  - Falhas de Rede
  - Ataques
  - Erros Humanos
  - Upgrades

PROF. FERNANDO AMARAL
www.datascientist.com.br

Fault Tolerance

- Tolerância a Falhas
- Garantias que um sistema, em caso de falha, continuará operando

- Executer Recover: tolera falha em slave, que é replicado
- Full

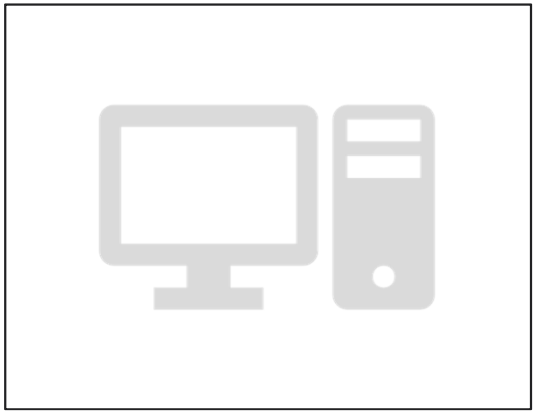
PROF. FERNANDO AMARAL
www.datascientist.com.br

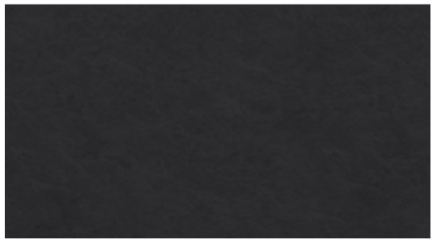
State Management

- Em caso de falha, o sistema manterá o estado na recuperação
  - Sem manter estado: o processo (ex. ingestão de dados) é reiniciado
  - Com manutenção de estado, ele continua de onde parou

Cloud e On-primese

- Cloud pode trazer custos menores (TCO)





PROF. FERNANDO AMARAL
www.datascientist.com.br

ETL e ELT

ETL: EXTRACT, TRANSFORM AND LOAD

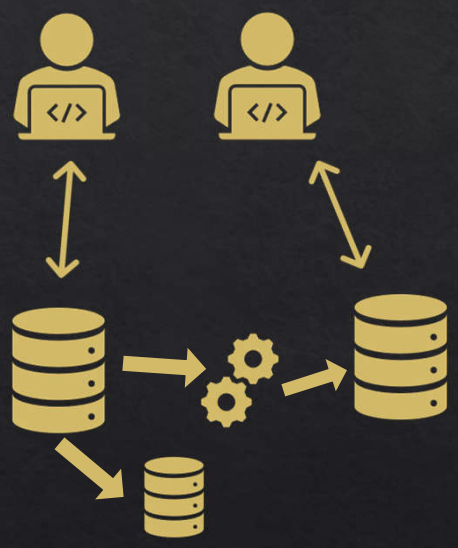
ELT: EXTRACT, LOAD AND TRANSFORM

Metadados

- Dados de Dados
  - Tipo
  - Precisão
  - Restrições

PROF. FERNANDO AMARAL
www.datascientist.com.br

Data at Rest VS Data at Wire



PROF. FERNANDO AMARAL
www.datascientist.com.br

Orientado a Coluna ou Linha

- Orientado a linha comprime por grupos de linhas (blocos)
- Mais eficiente para escrita e leitura
  - Desvantagem: baixa taxa de compressão
  - Leitura de algumas colunas: precisa ler todas

- Orientado a Colunas
  - Armazena-se colunas separadamente
  - Maior taxa de compressão
  - Melhor para leitura

PROF. FERNANDO AMARAL
www.datascientist.com.br