



# Formação Engenheiro de Dados

ETL

## ETL

**ETL:** Extract Transform Load (*Extração Transformação Carga*), são ferramentas de software cuja função é a extração de dados de diversos sistemas, transformação desses dados conforme regras de negócios e a carga dos dados;



# ETL

◆ Etapa mais difícil, longa e cara na construção de um DW



# Estratégia de ETL

- ◇ Conectar em horários apropriados
- ◇ Extrair os dados necessários
- ◇ Desconectar
- ◇ Conexões somente leitura





# ETL



- ◊ No uso clássico, ETL extrai dados de sistemas Transacionais, transforma o modelo de dados, e carrega no Armazém de Dados

# Ferramentas de ETL

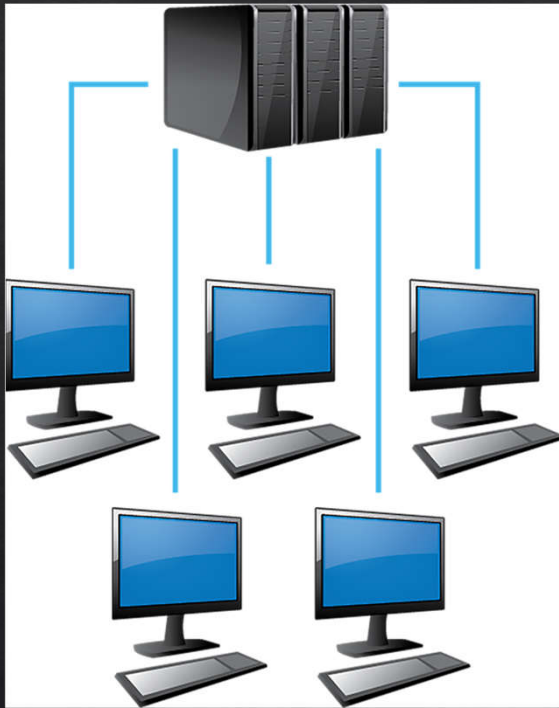
- ◆ Muitas Ferramentas Diferentes

- ◆ Manuais (Codificação com scripts ou SQL)

- ◆ Workflows

- ◆ Mistas

# Tarefas



- ◊ Qualidade de Dados:
  - ◊ Correto, Não ambíguo, Consistente, Completo
- ◊ Limpar (duplicados, regras de negócio etc)
- ◊ Alteração de codificação (EDCDIC para ASCII)
- ◊ Transformar para modelo dimensional



# Staging

- ◇ Armazenamento intermediário dos dados:
  - ◇ Após Extração
  - ◇ Antes de Transformar e Carregar
- ◇ Pode ser processada em memória ou em disco
- ◇ Alguns projetos guardam dados de Staging, outros, o apagam depois do carga



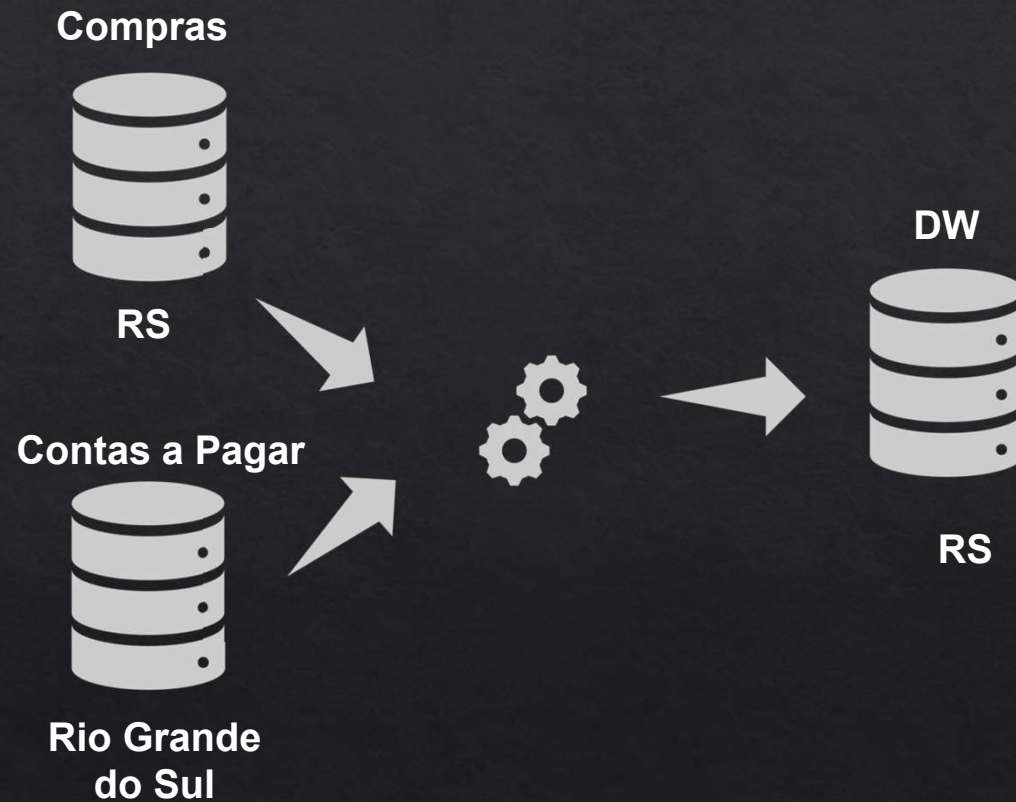


# Staging

- ◆ Dados não devem ser acessados por usuários



# Consolidação de Dados



# Consolidação de Dados

Compras



ID 030 - Maria

Contas a Pagar



ID 2345 - Maria



DW



Maria