

The background is a dark blue-grey gradient. In the top-left corner, there are two overlapping geometric shapes: a blue parallelogram and a light green parallelogram. In the bottom-left corner, there is a circular inset showing a detailed, high-magnification view of a circuit board with various electronic components. In the top-right corner, there is a faint, grey, 3D-rendered pattern of interconnected lines and squares, resembling a circuit or data flow diagram.

# Apresentação de BIG DATA

## Data Lake

É um verdadeiro ecossistema de informações. Um ambiente que contém desde arquivos diversos até registros de navegação na web e informações de IoT, visando a agilidade, disponibilizando dados em vários formatos sem estruturação prévia. é uma reunião de dados brutos em um repositório centralizado, mas demandando uma estratégia sólida de governança para garantir usabilidade pois esse conjunto de dados pode ser um tesouro ou uma armadilha se não for bem gerenciado.

## Data Mesh

Enfatiza a automação e a padronização, promovendo a descentralização. as organizações acumulavam dados em um único repositório, como um Data Lake ou Data Warehouse. No entanto, extrair valor real desses dados se provou uma tarefa complexa. É aqui que entra o Data Mesh. Cada unidade de negócios (domínio) se torna responsável por seus próprios dados, desde sua qualidade até sua governança e disponibilidade. Para que os dados sejam compreendidos e interpretados em toda a organização, é necessário padronizar metadados, semântica e interfaces.

## Data Lake

um repositório de armazenamento centralizado mas altamente flexível que armazena uma quantidade de dados muito grandes que podem ser estruturados ou não armazenados de forma bruta original sem nenhum formar consolidando os dados e a flexibilidade de dados mas falta de confiabilidade e segurança e pode ter desempenho insatisfatório para business intelligence e análise de dados.

## Data Mesh

Enfatiza a automação e a padronização, promovendo a descentralização. Aqui, os dados se transformam em produtos menores, como tabelas, relatórios, APIs e modelos analíticos, produzidos por “Data Owners”. Tendo uma facilidade de acesso e escalabilidade mas implementação pode ser complexa e governança de dados pode ser desafiadora

## Data Warehouse

um repositório de dados Unificado para armazenar **grandes quantidades** de informações de **várias Fontes** dentro de uma organização melhorando a padronização qualidade e consistência dos dados como as organizações fornecer inteligência de negócios aprimorada aumenta o poder e a velocidade das cargas de trabalho de análise de dados inteligência de negócios mas a uma falta de flexibilidade e podem ser caros de implementar e manter



## ETL

É abordagens de processamento de dados para análise, usa um conjunto de regras de negócios para processar dados de várias fontes antes da integração centralizada, tem uma etapa adicional antes de carregar dados no destino, que é difícil de escalar e desacelera o sistema à medida que o tamanho dos dados aumenta.

## ELT

Também é abordagens de processamento de dados para análise, carrega os dados como estão e os transforma em um estágio posterior, dependendo do caso de uso e dos requisitos de análise, carrega dados diretamente no sistema de destino e os transforma em paralelo, ele usa o poder de processamento e a paralelização que os data warehouses em nuvem