



IMP. DW/DL

ARQUITETURA DW

INTRODUÇÃO

A Arquitetura de um sistema de DW é fundamental para a **organização e a análise eficiente** de grandes volumes de dados.

Ela envolve uma série de **características e ferramentas** específicas para coleta, armazenamento, processamento e análise de dados.



CARACTERÍSTICAS

Dados de Origem

Devem ser selecionados de modo a fornecer **subsídio para as análises** que se deseja oferecer através do DW.

Podem advir de **SGDBs, planilhas, arquivos csv, plataformas em cloud**, dentre outras fontes.



CARACTERÍSTICAS

Armazenamento

Os dados coletados na origem são armazenados em um **repositório centralizado**.

Este repositório caracteriza o **DW**, estruturado em um **SGDB otimizado para consultas e análises de alto desempenho**.



CARACTERÍSTICAS

ETL

Processo de extração, transformação e carga que deve ocorrer para realizar a limpeza e a integração dos dados.

A etapa de carga é responsável por transferir os dados da área de estágio para o DW.



DIAGRAMA DE ARQUITETURA

Fontes de Dados



ETL (Pentaho Data Integration)



Data Warehouse (Vertica)



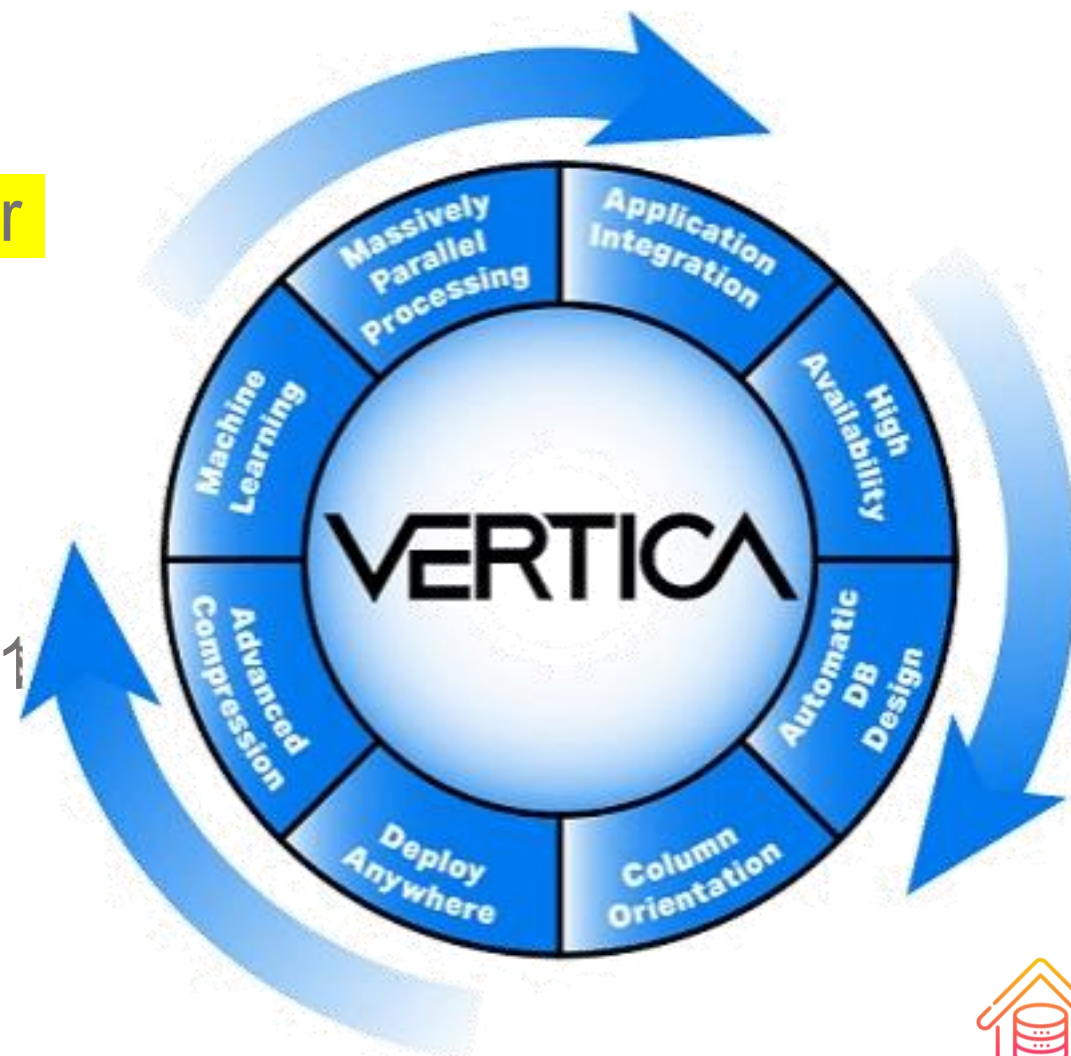
Ferramentas de Análise (ex: Tableau, Power BI)



VERTICA

O Vertica é um **banco de dados colunar de alto desempenho** projetado para análise de dados em larga escala.

Inicialmente desenvolvido pela Vertica Systems, foi adquirido pela **HP** em 2011 que o incorporou ao seu portfólio de produtos.



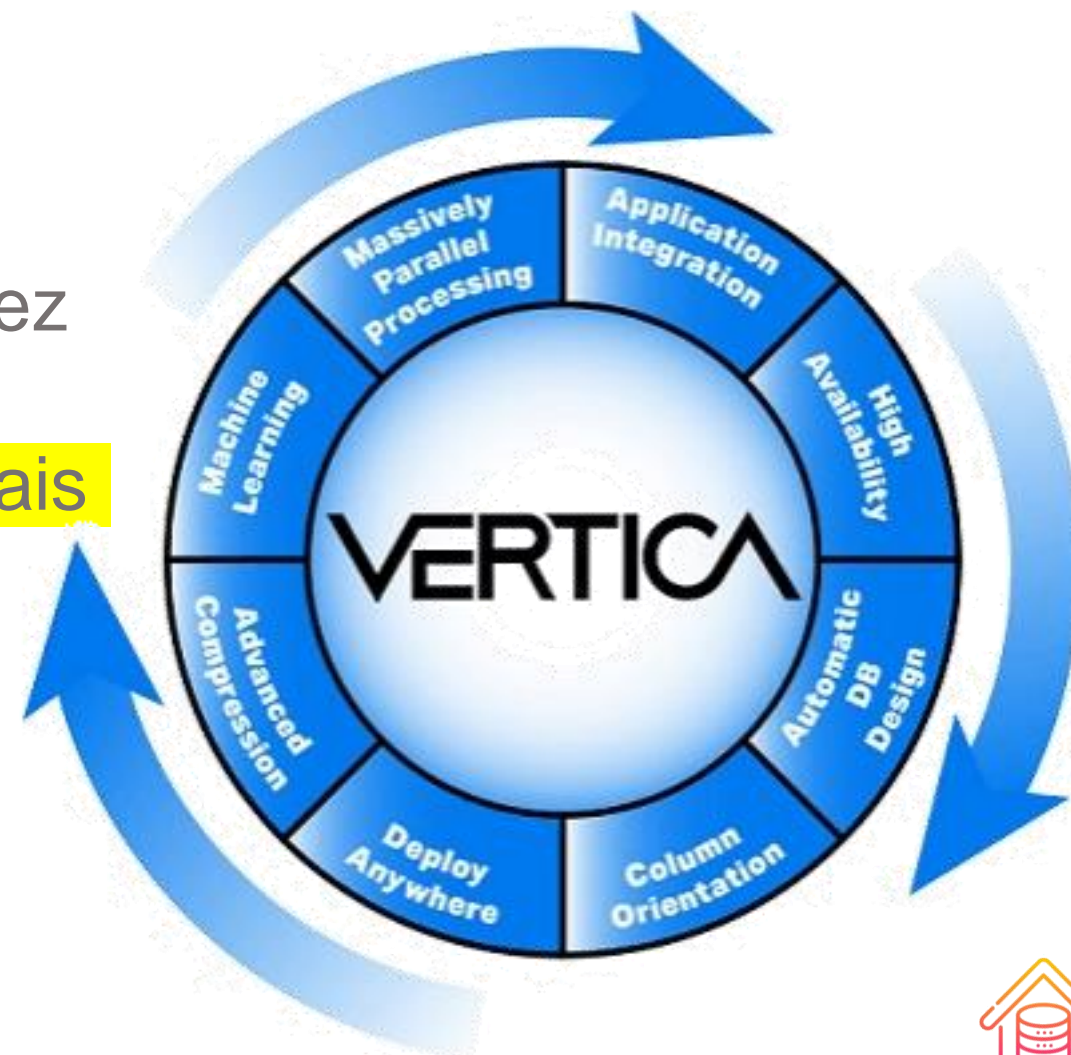
CARACTERÍSTICAS TÉCNICAS

Arquitetura Colunar

Dados armazenados em **colunas** em vez de linhas, o que permite uma **compreensão eficiente e um acesso mais rápido** para as consultas analíticas.

Processamento Massivamente Paralelo (MPP)

Escalabilidade linear que permite ao Vertica lidar com grandes volumes de dados.



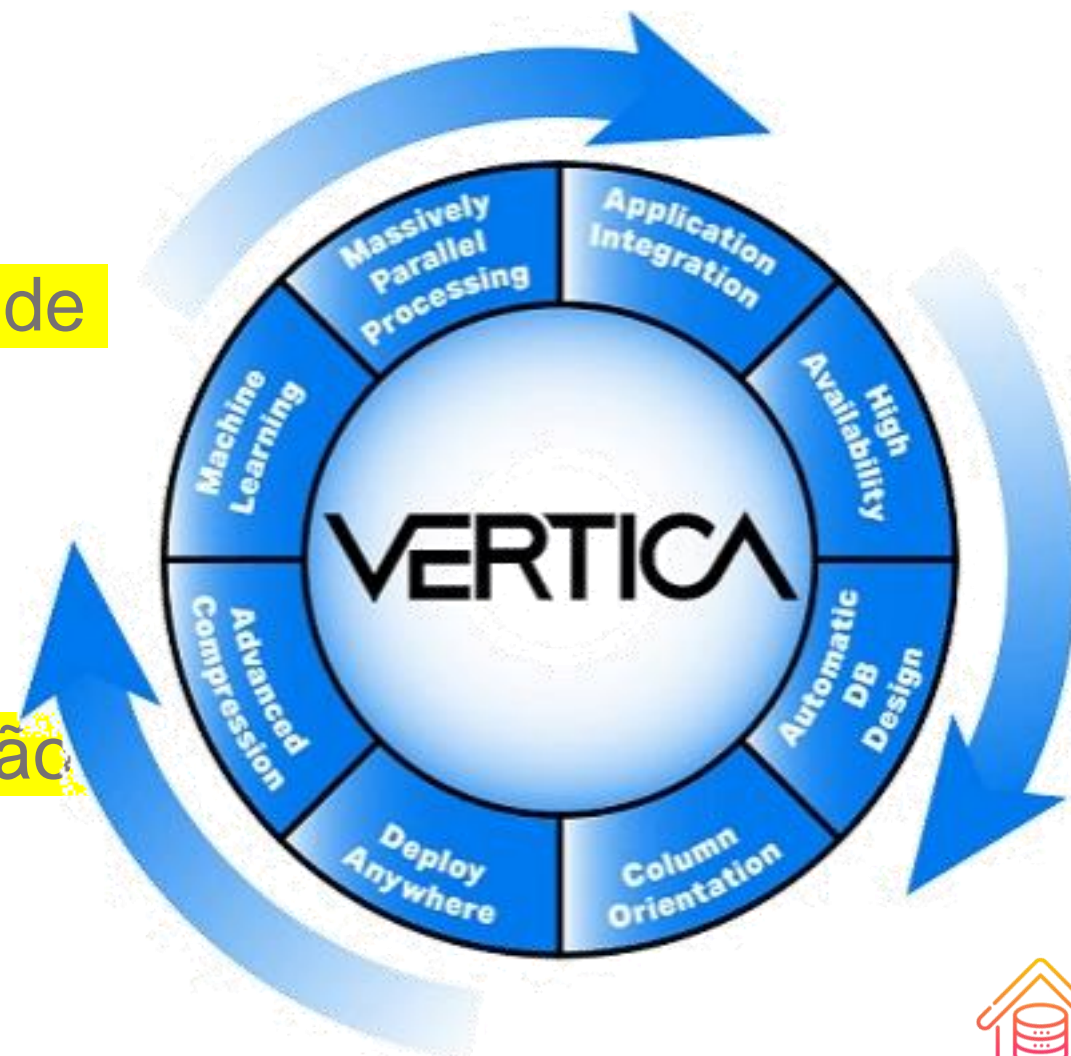
CARACTERÍSTICAS TÉCNICAS

Compressão Avançada

Reduz significativamente os requisitos de armazenamento e melhora o desempenho das consultas.

Optimizador de Consultas

Analisa o SQL e cria planos de execução eficientes, levando em consideração estatísticas e distribuição dos dados.

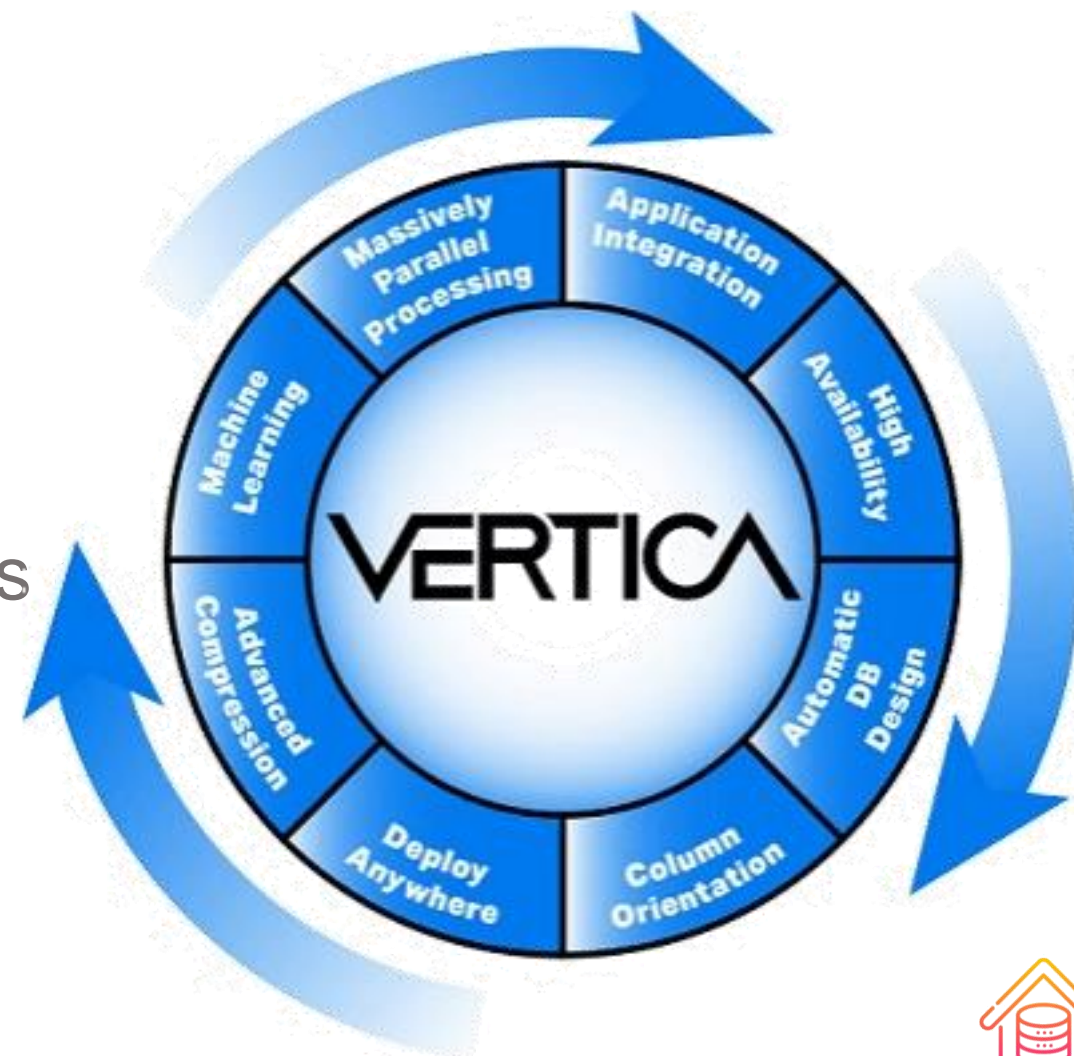


CARACTERÍSTICAS TÉCNICAS

Particionamento e Segurança

Organização e distribuição dos dados conforme necessidades do usuário.

Controle de acesso baseado em papéis criptografia.



IDEAL PARA DW

Desempenho, escalabilidade e capacidade em lidar com consultas analíticas complexas.

Arquitetura colunar e técnicas de compressão.



ARQUITETURAS ALTERNATIVAS

Fontes de Dados



ETL (AWS Glue)



Data Warehouse (Amazon Redshift)



Ferramentas de Análise



ARQUITETURAS ALTERNATIVAS

Fontes de Dados



ETL (Azure Data Factory)



Data Warehouse (Azure Synapse)



Ferramentas de Análise



ARQUITETURAS ALTERNATIVAS

Fontes de Dados



ETL (Cloud Dataflow)



Data Warehouse (BigQuery)



Ferramentas de Análise



Google Cloud



atividade

documentação





Obrigado!

Professor Gustavo Dias
luizdias@univas.edu.br