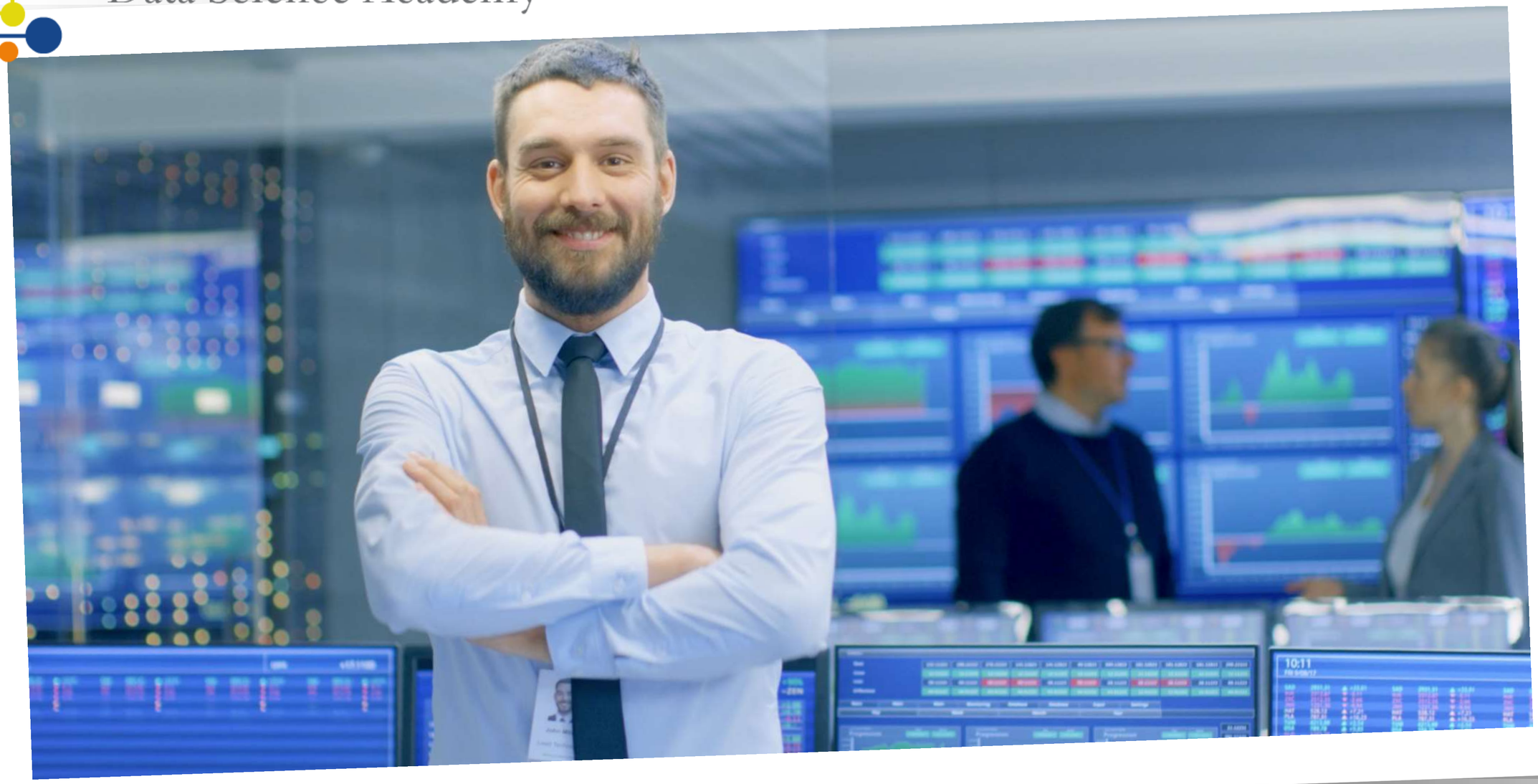
Data Science Academy

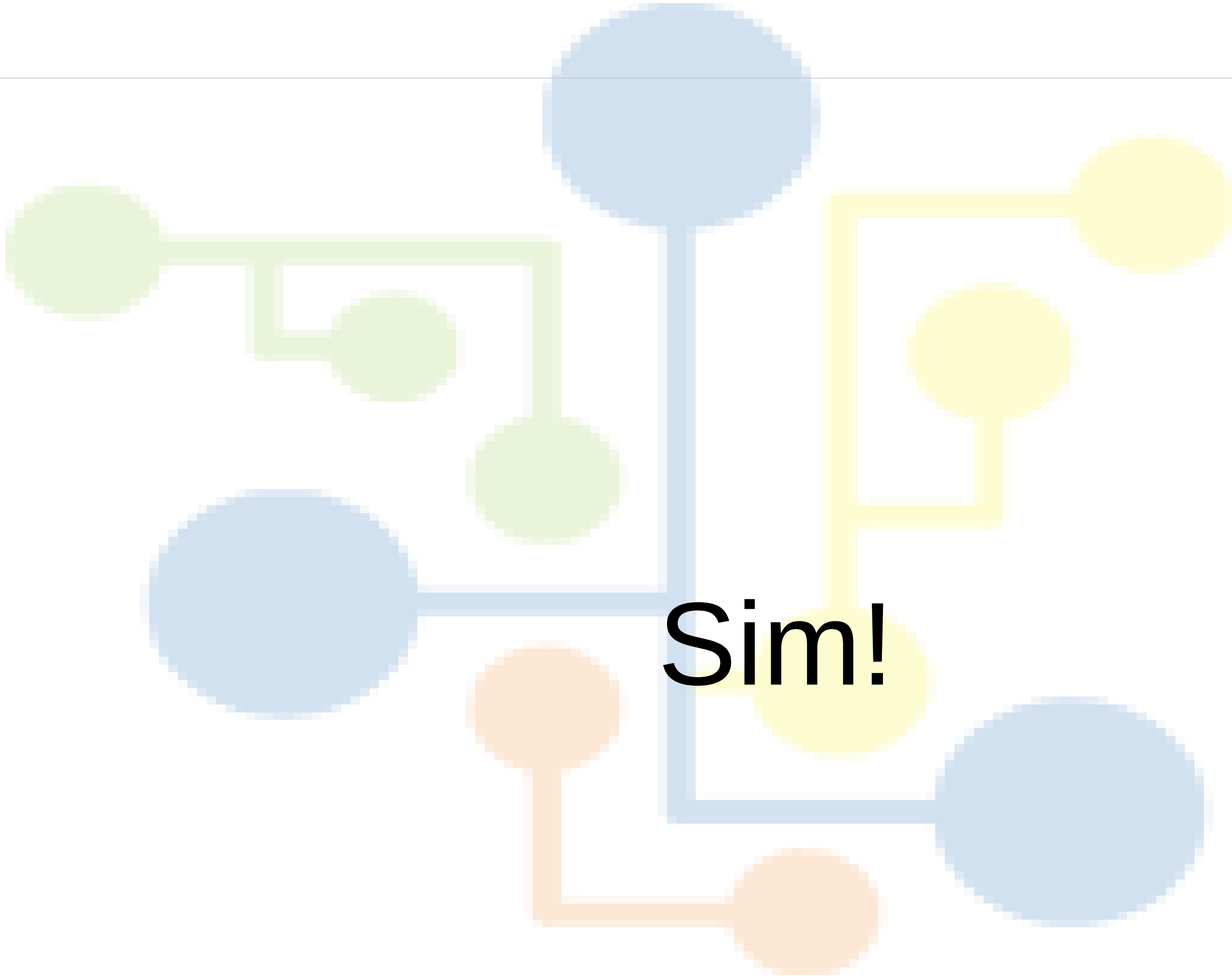


Design e Implementação de Data Warehouses

Data Science Academy

Implementando Data Warehouse em Nuvem - Parte 1

Seria possível montar um Data Warehouse para armazenamento e processamento de Petabytes de dados em apenas alguns cliques?



Sim!


amazon
REDSHIFT

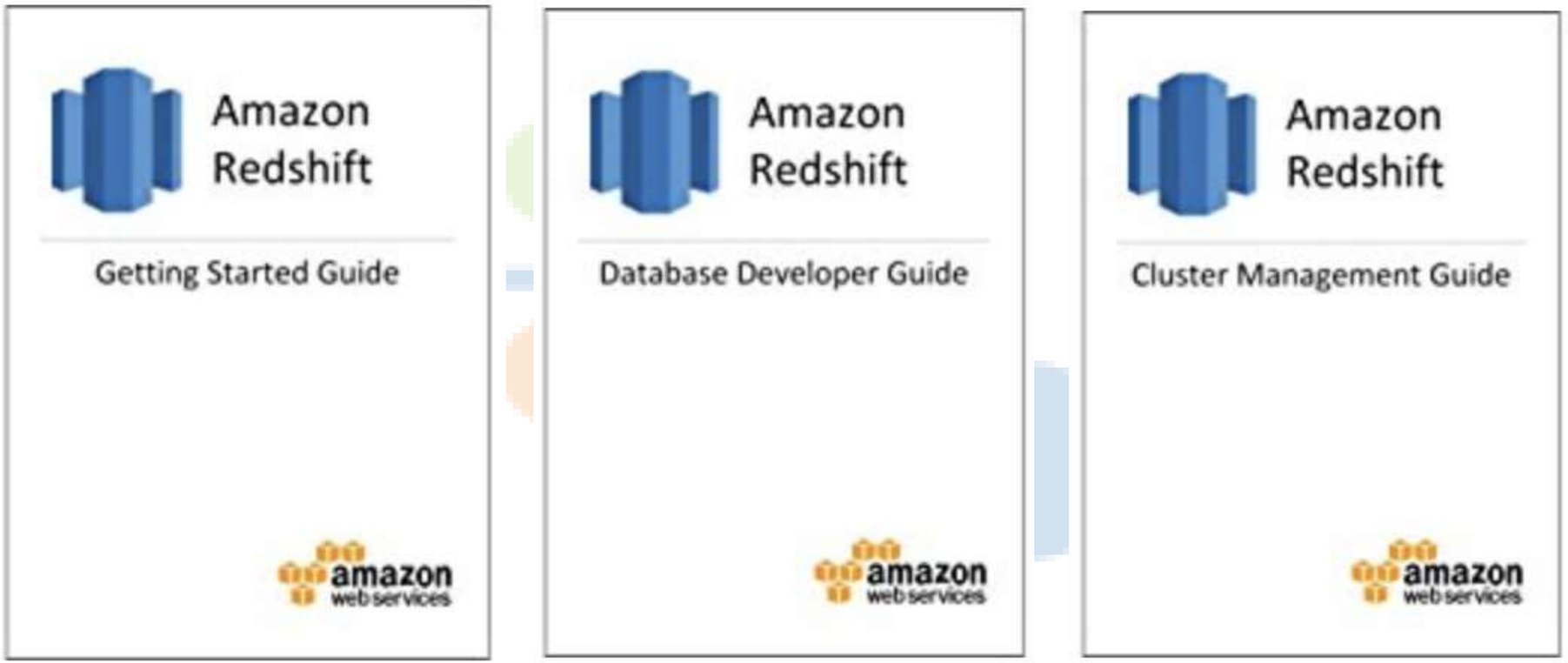
Implementação de DW em Nuvem

Parte 1

Parte 2

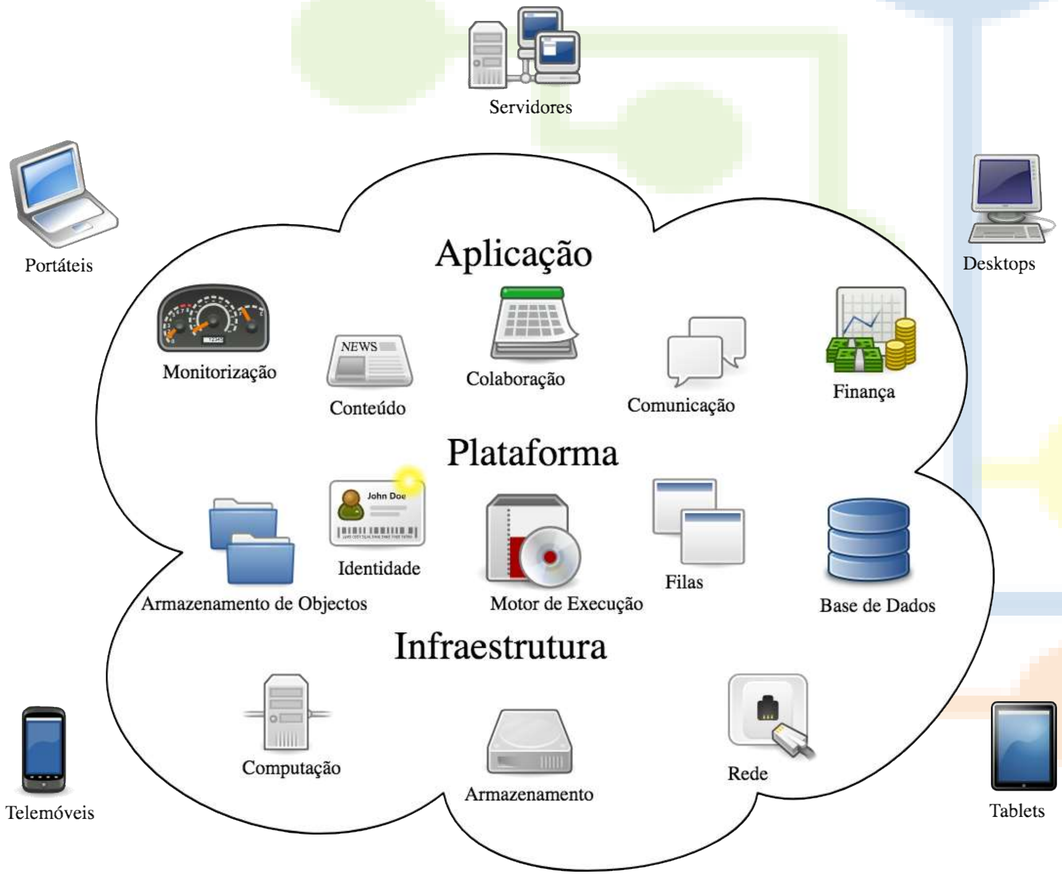
Alinhando as expectativas...

- Este não é um curso sobre AWS.
- Para fazer o cadastro gratuito na AWS é necessário um cartão de crédito.
- O Amazon Redshift pode ser testado gratuitamente por 2 meses.
- O cadastro na AWS é gratuito e diversos serviços podem ser usados sem custo por 12 meses (ou limite de horas especificado por serviço).
- Neste segundo projeto o foco será maior no Amazon Redshift, do que na especificação (como fizemos no projeto 1).


Amazon Redshift
Getting Started Guide
amazon web services
Amazon Redshift
Database Developer Guide
amazon web services
Amazon Redshift
Cluster Management Guide
amazon web services

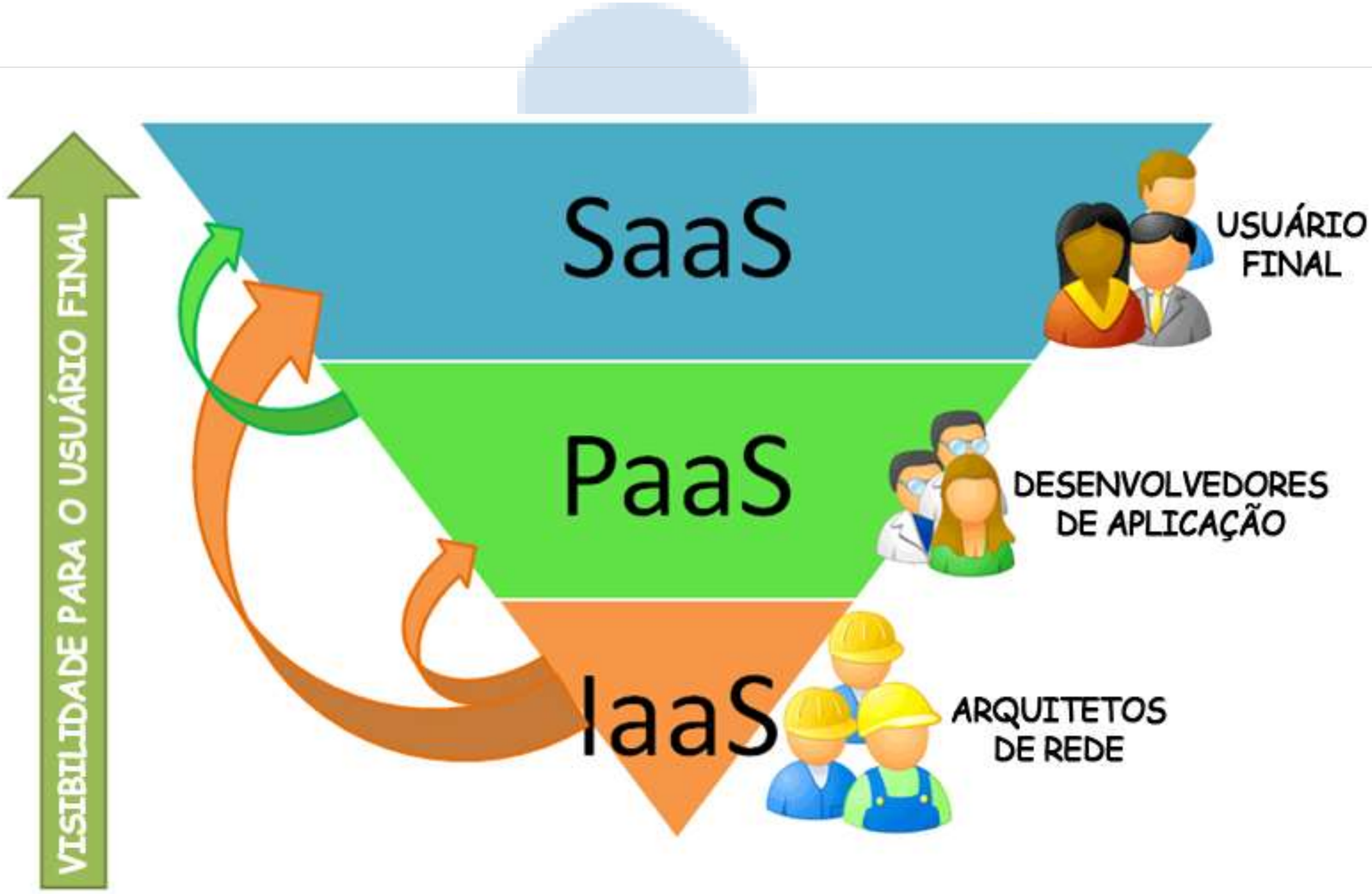
Data Science Academy

A Importância do Cloud Computing



Computação em nuvem

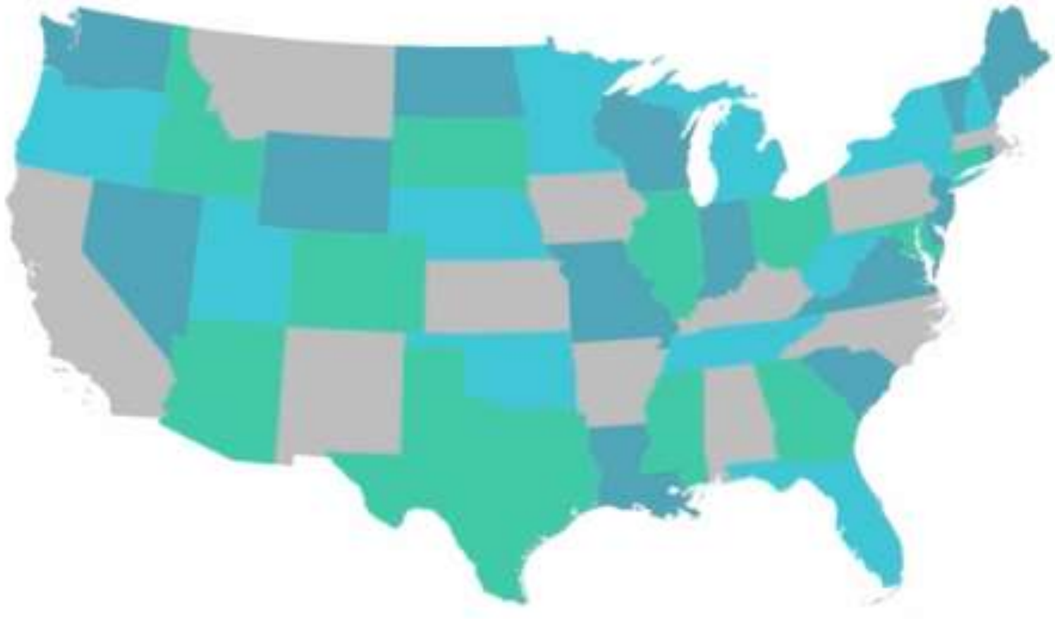
Computação em nuvem é o futuro da área de TI.



Serviços, Plataformas e Infraestrutura
já são oferecidos em nuvem!

Data Science Academy

Empresas dos EUA:



85%

JÁ USAM

ALGUM SERVIÇO

NA NUVEM

Data Science Academy

Benefícios de um Data Warehouse em Nuvem

Principais **benefícios** de um DW em nuvem:

- Ambiente descentralizado
- Sem custo de infraestrutura
- TCO reduzido
- Escalabilidade para Petabytes de dados
- Segurança
- Acesso via internet de qualquer local
- Gestão simplificada

Principais **desvantagens** de um DW em nuvem:

- Falta de flexibilidade
- Dados sensíveis estarão sob a gestão de terceiros (Amazon ou Microsoft)
- Segurança
- Integração de dados

Data Science Academy

História do Amazon Redshift

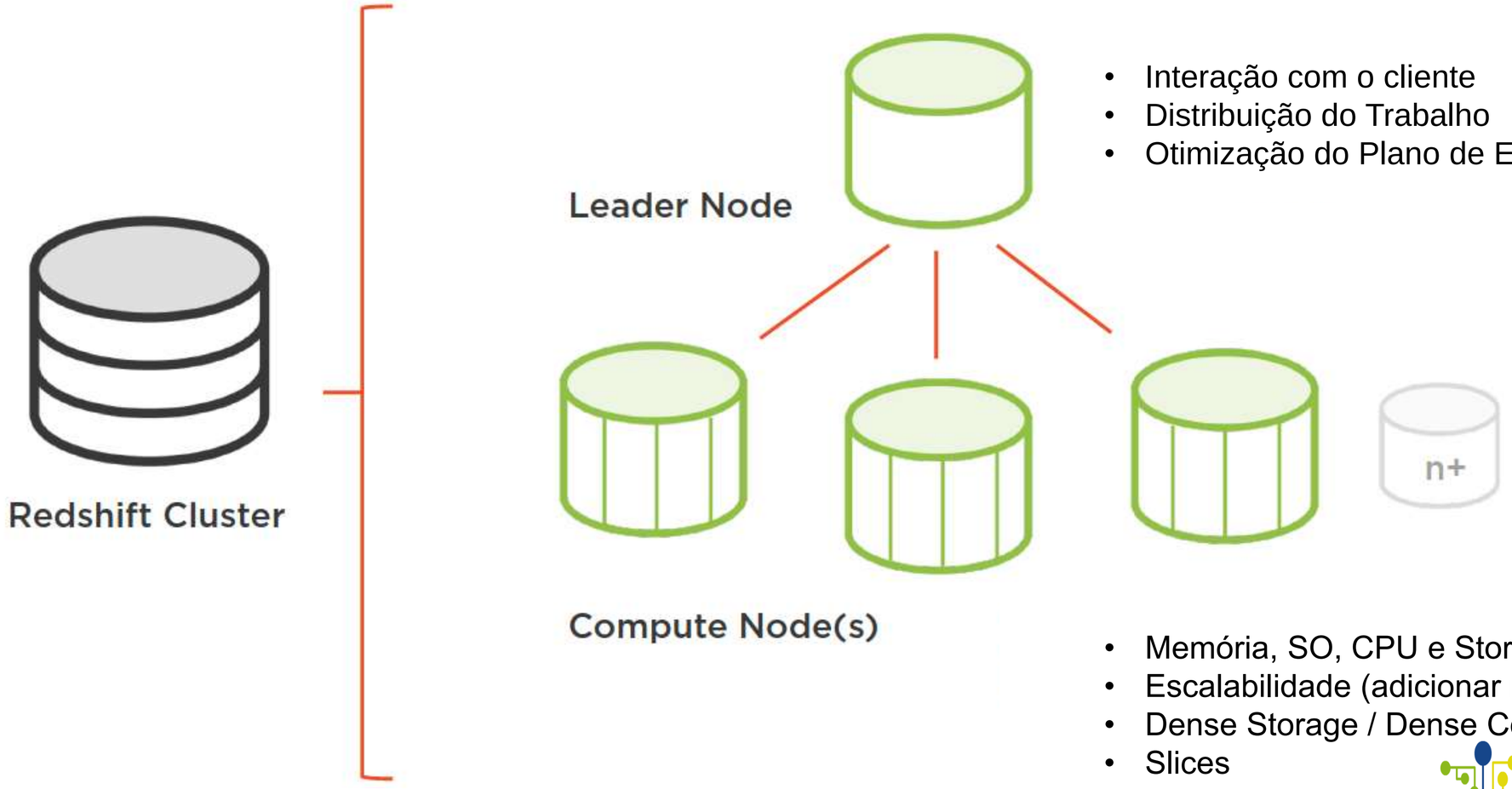

PostgreSQL

Desenvolveu uma versão do PostgreSQL para processamento paralelo massivo de dados em DW

Amazon comprou a ParAccel e criou o Amazon Redshift a partir do produto criado pela empresa, baseado no PostgreSQL

Data Science Academy

Amazon Redshift Cluster



- Interação com o cliente
- Distribuição do Trabalho
- Otimização do Plano de Execução

- Memória, SO, CPU e Storage dedicados
- Escalabilidade (adicionar mais nodes)
- Dense Storage / Dense Compute
- Slices

Data Science Academy

Principais Características do Amazon Redshift

Principais Características do Amazon Redshift

Query optimization

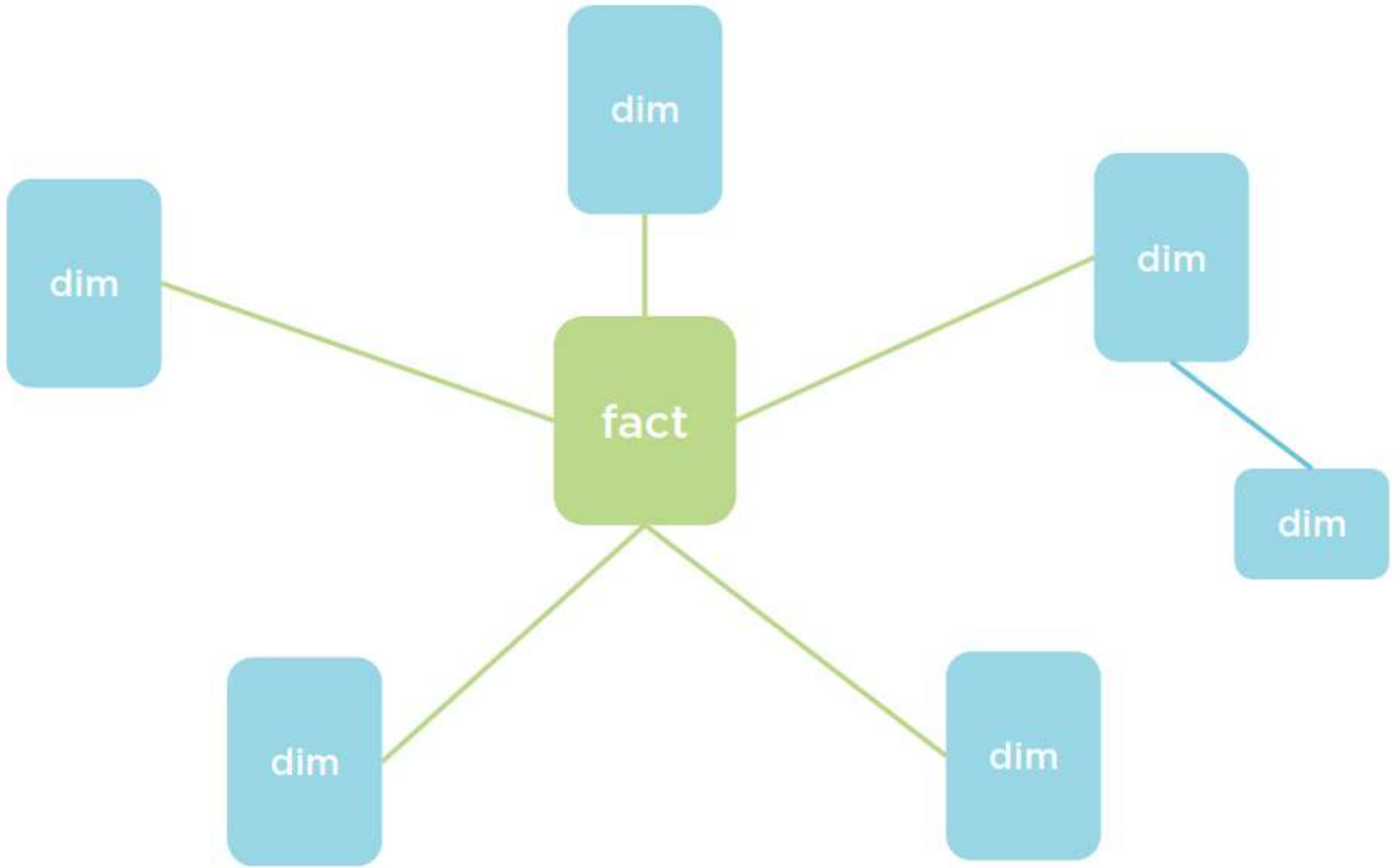
Internal statistics and heuristics

ACID compliance

Data distribution

Estilos de Distribuição

Dados são distribuídos de um node para outro em cada consulta.


dim
dim
dim
fact
dim
dim
dim

Even distribution

Distribui os valores igualmente, sendo o estilo de distribuição padrão. Oferece boa performance.

Key distribution

Os registros são distribuídos por chave e os relacionamentos conhecidos são distribuídos nos mesmos nodes. Reduz o overhead de redistribuição de registros.

All distribution

Todos os registros são copiados para todos os nodes, aumentando o custo de armazenamento e de operações DML. Usado em tabelas com poucas mudanças e muitos joins.

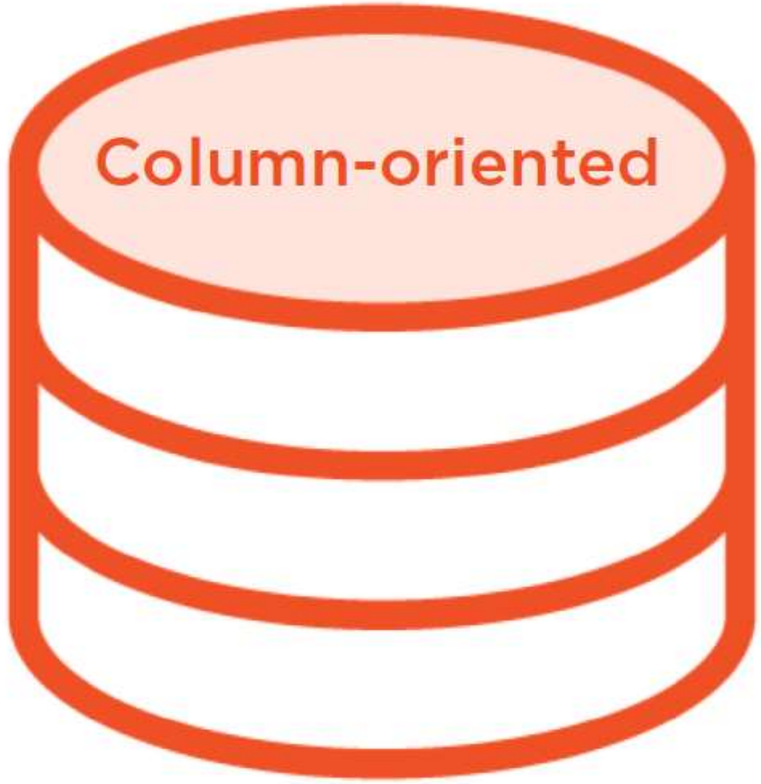
O objetivo em definir o estilo de distribuição é reduzir o impacto dos passos de redistribuição dos registros, mantendo os dados no melhor local possível antes da execução da query.

Data Science Academy

Armazenamento Orientado a Coluna e a Velocidade do Redshift

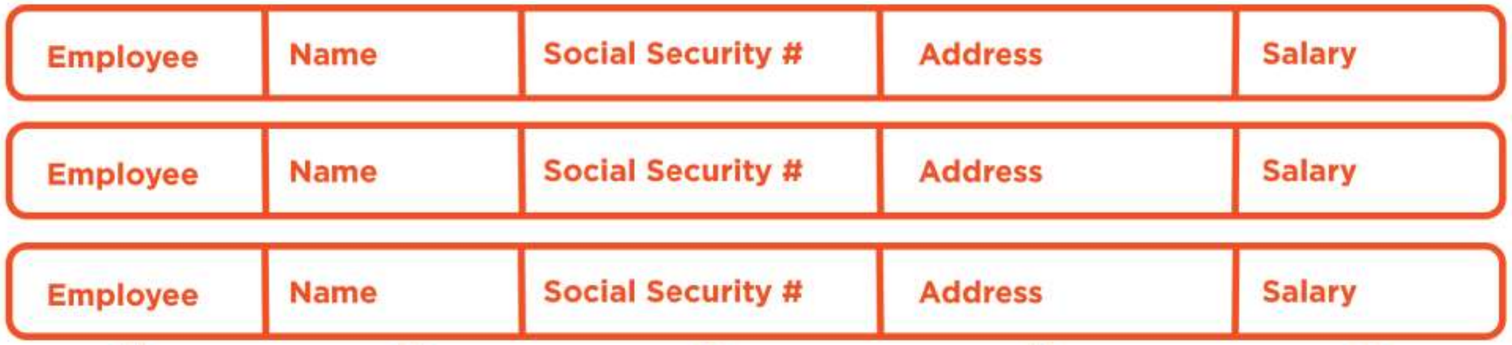
Data Science Academy

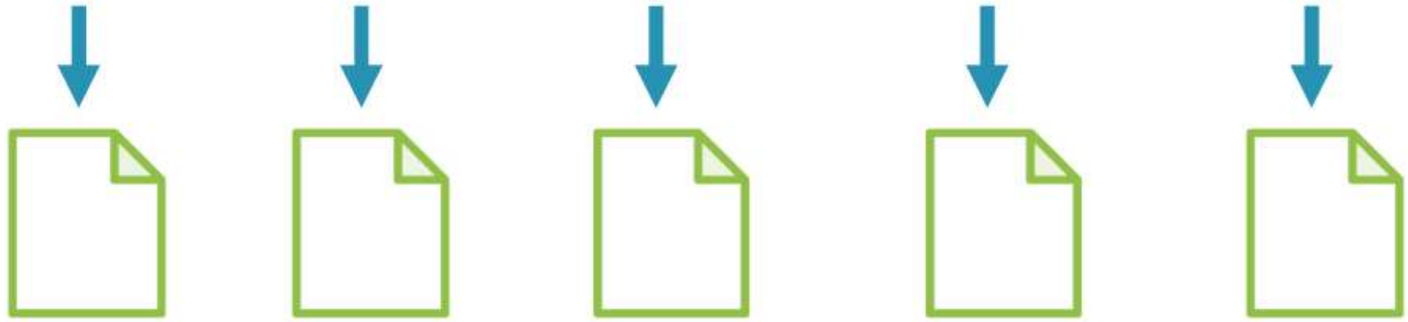
Amazon Redshift é um banco de dados orientado a coluna.

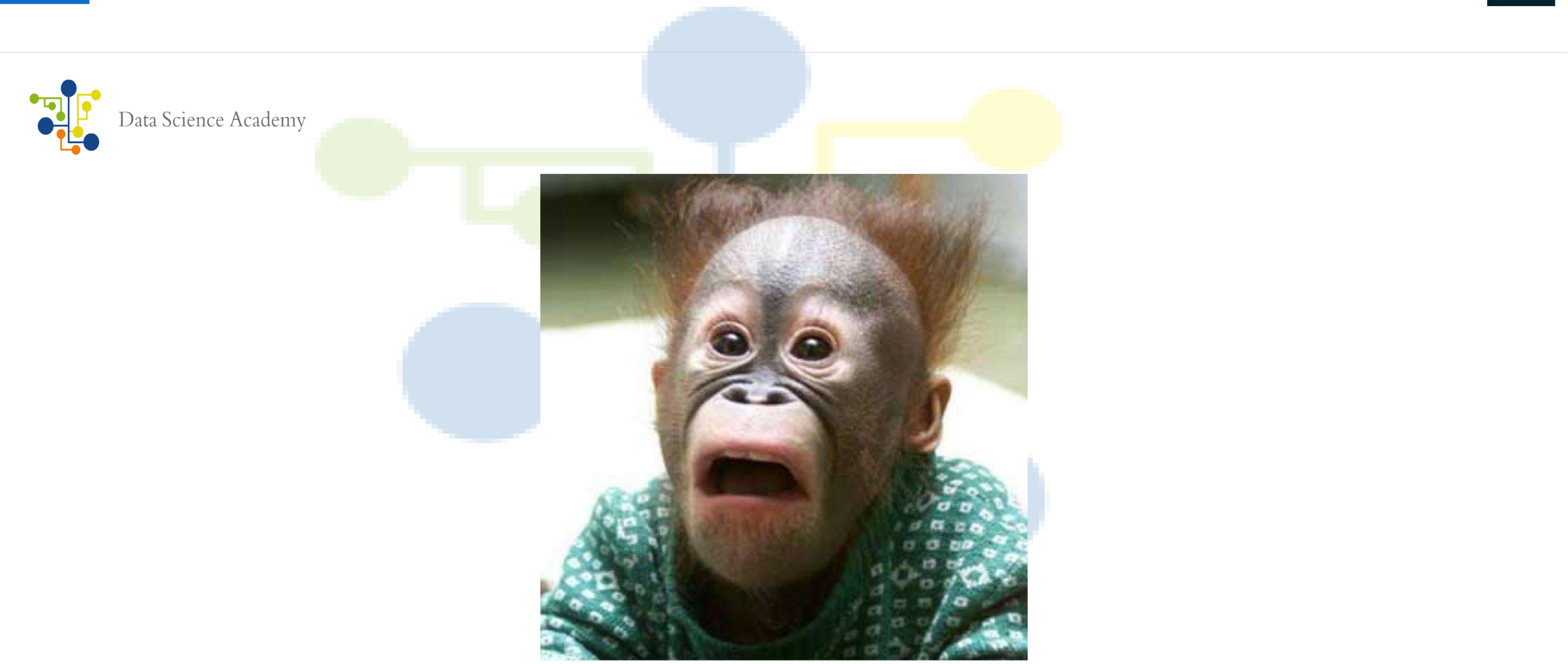


Row-oriented Storage

Column-oriented Storage







Não há clustered ou non-clustered index no Redshift.

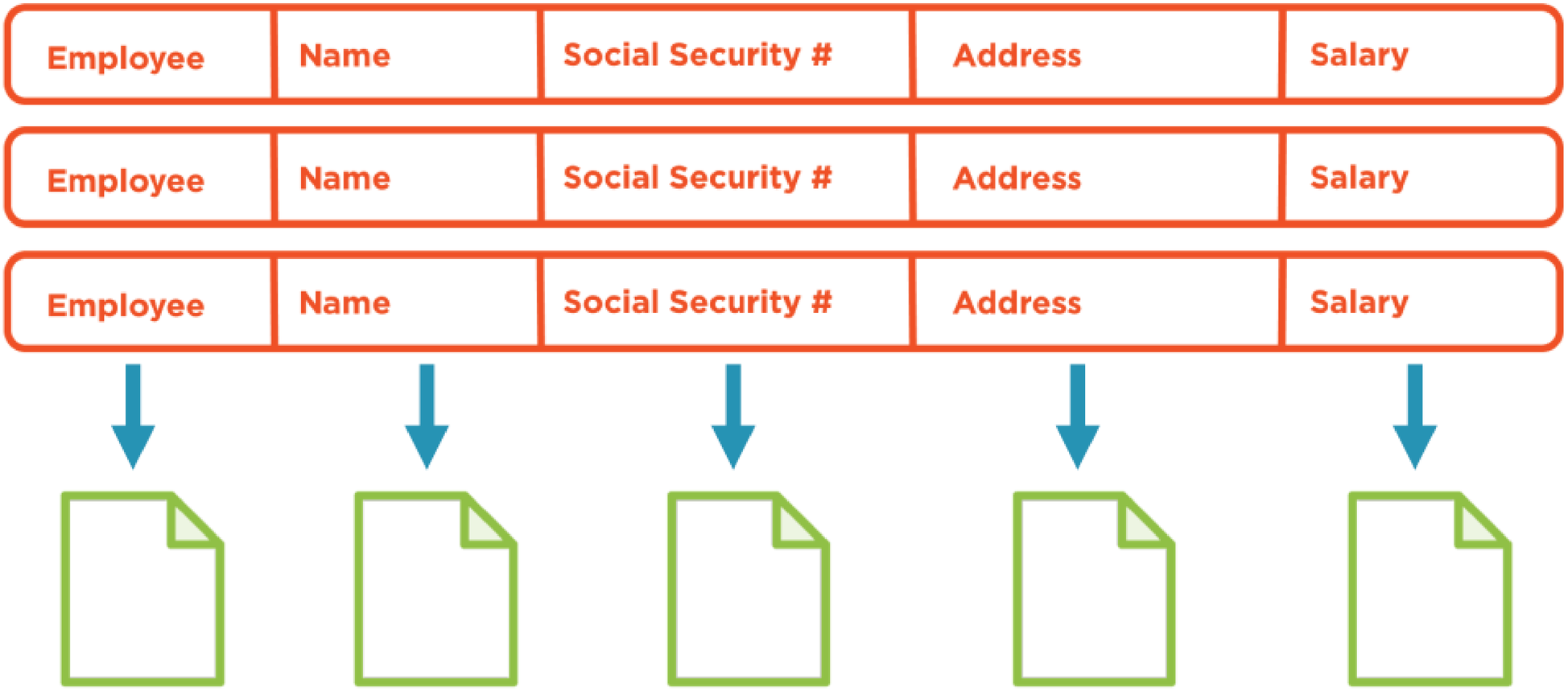
O Redshift possui arquitetura orientada a coluna, massivamente paralela.

O Redshift não deve ser usado para ambientes com muitas operações CRUD.

Data Science Academy

Compressão de Dados

Data Science Academy



- Otimiza a compressão para um único tipo de dado, reduzindo espaço de armazenamento.
- Acelera leitura de dados.
- Reduz o overhead de redistribuição de dados.

Data Science Academy

Muito Obrigado!

É um prazer ter você aqui.

Tenha uma excelente jornada de aprendizagem.

Oportunidade

Disponibilidade

Conhecimento