**Comparação de Desempenho: AWS Glue vs. EMR**

Apresentação de Evidências de Desempenho e Custos:

**Introdução**

Objetivo:  
- Apresentar a comparação de desempenho entre Glue e EMR.  
- Avaliar impacto em custo, tempo e escalabilidade para processos de ETL.

**Metodologia**

Configuração do Ambiente:

- Cluster EMR:

1 máquina m5.xlarge para Master e 50 máquinas m5.xlarge para Cores

Cada máquina m5.xlarge é equivalente à **4vCPU** e **16GB RAM** .

Capacidade do cluster (Somente cores) **200vCPU** e **800GB RAM**

- Glue:

Workers configurados com 50 DPUs G1.X.

Cada DPU **4vCPU** e **16GB RAM.**

Capacidade do cluster **200vCPU** e **800GB RAM**

**Tipo de Trabalho Executado**

Leitura de 4,9TB de dado bruto armazenado no S3, transformação de mapeamento e conversão de tipo de dados e escrita em formato parquet armazenando em S3.

**Critérios de Avaliação**

- Glue levou [Tempo] custando [Custo].  
- EMR levou [Tempo] custando [Custo].

**Facilidade de Implementação**  
- Glue: Simplicidade no código.  
- EMR: Maior complexidade, mas mais personalizável.

**Análise de Vantagens e Desvantagens**

AWS Glue  
- Vantagens: Serviço gerenciado, sem necessidade de provisionamento manual.  
- Desvantagens: Custos mais elevados para trabalhos grandes, limitações em personalização.  
  
AWS EMR  
- Vantagens: Controle total do cluster, custo escalável com **Savings Plans** e instâncias Spot.  
- Desvantagens: Maior complexidade operacional e necessidade de gerenciamento de infraestrutura.

**Recomendações**  
- Usar Glue para jobs pequenos ou testes.  
- Usar EMR para cargas maiores e trabalhos customizados.

**Próximos Passos**  
- Explorar Savings Plans para clusters EMR.