

Por que monitoramento?

01

Debugging Proativo

02

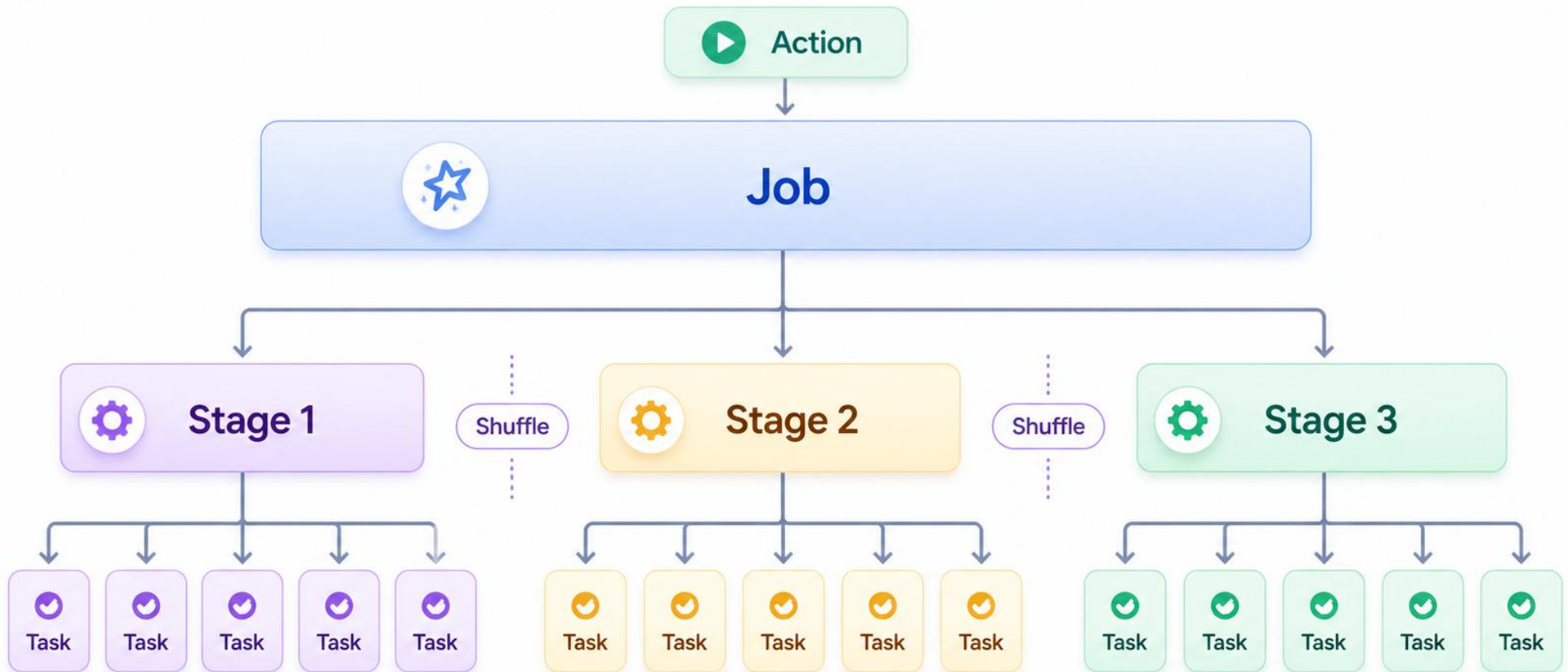
Impacto no Custo

03

Escalabilidade

A anatomia da execução no Spark

• Jobs, Stages e Tasks •



Indicadores de Performance na Spark UI

Métricas-chave para identificar gargalos



**Executor
Computing Time**



**Shuffle
Read/Write**



Spill (Memory/Disk)



**Task
Deserialization**

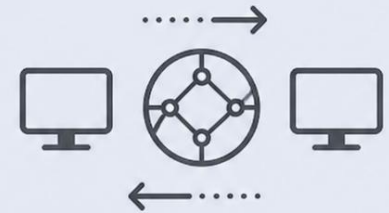


Identificando Gargalos de Spark

Data skew



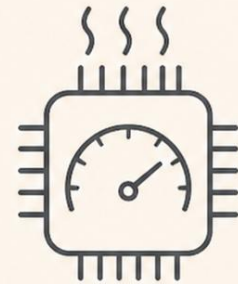
Tráfego de Rede



Disparidade de Métricas



Sobrecarga de CPU



Investigar Jobs

Método sistemático na Spark UI: identifique gargalos e optimize jobs.

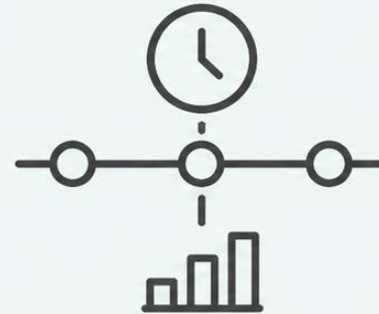
Identificar Stage



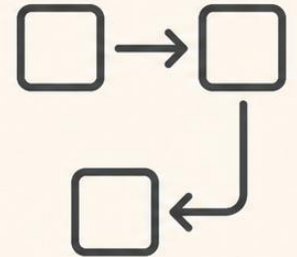
Analisar Aba SQL



Analisar Timeline



Transformações



Logs do Databricks

Analise logs para diagnosticar erros em clusters Databricks.

Driver Logs



Executor Logs



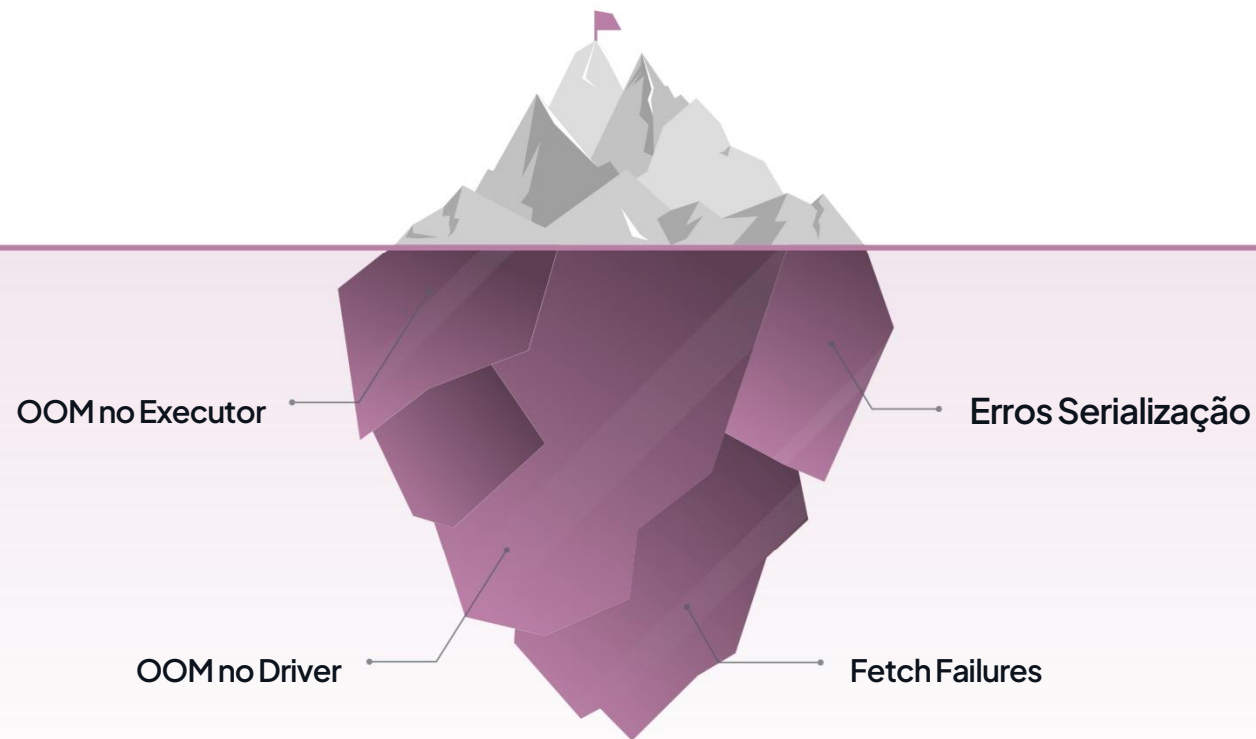
Configuração Log4j



Spark Event Logs

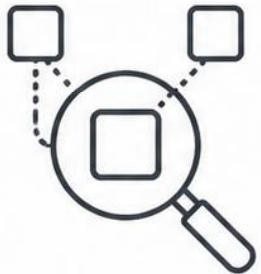


Falhas comuns: OOM e erros de serialização



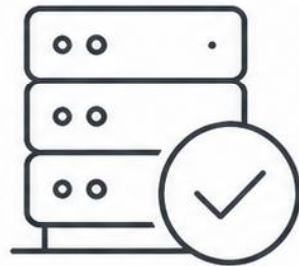
01

Isolar o Problema



02

Checamagem de Infra



03

Otimizar Processos



04

Práticas de Logging



Navegação prática na Spark UI via Databricks

Não visível



Acesso ao Cluster

Vá até a aba Compute, selecione seu cluster específico e clique na aba Spark UI dedicada.



Histórico de Jobs

Para pipelines já finalizados, acesse a aba Workflows para visualizar todo o histórico.



Métricas de Tasks

Observe a barra de progresso das Tasks para identificar se o processamento está lento.



DAG Visualization

Use o gráfico DAG para confirmar se as tabelas e filtros estão sendo otimizados aqui.