



Plano trimestral de testes de carga

Critérios e cenários de leitura e escrita em massa para funcionalidades com impacto potencial de escalabilidade.

Objetivo e cadência

Os testes de carga são realizados trimestralmente para verificar a capacidade da solução sob volumes representativos. Funcionalidades novas ou alteradas com impacto potencial de escalabilidade são incluídas no ciclo aplicável e identificadas na evidência de execução.

Modalidade	Quando executar	Escopo mínimo	Evidência
Ciclo trimestral	uma vez por trimestre	jornadas críticas, consultas analíticas e comparação com o ciclo anterior	plano, execução, métricas, conclusão e ações
Mudança relevante	no ciclo que antecede a homologação	funcionalidade com impacto em volume, concorrência, consulta, ingestão ou armazenamento	vínculo com a mudança e resultado do cenário afetado
Reexecução	após ajuste com impacto material	cenário que ultrapassou o critério ou sofreu regressão	comparativo antes e depois

Cenários de referência

Os cenários cobrem consulta analítica de grande porte e gravação concorrente de eventos de aprendizagem. A leitura em massa usa a referência do relatório institucional de redes; a escrita em massa simula estudantes registrando respostas, progresso e eventos no banco de dados.

Dimensão	Volume de referência	Uso no teste
Leitura - escolas	3.500	linhas agregadas da consulta institucional
Leitura - estudantes	1.000.000	população vinculada às escolas e filtros
Leitura - respostas	235.000.000	massa de interações usada nas agregações
Leitura - escala ampliada	300 milhões a 1 bilhão de registros	avaliação de crescimento sem perda significativa de desempenho
Escrita simultânea	200.000 estudantes	gravação concorrente de respostas, progresso e eventos no banco de dados

Cenários de leitura em massa

Cenário	Execução	Critério e métricas
Consulta-base	recuperar os indicadores agregados das 3.500 escolas, sem filtros	registrar latência, taxa de erro e volume retornado
Data dinâmica	executar ranges de 1, 7, 30, 90 e 365 dias	P95 menor que 500 ms até 30 dias; P95 menor que 1.000 ms entre 30 e 365 dias
Carga simultânea	100 requisições concorrentes com filtros diferentes	P95, P99, throughput em requisições por segundo e degradação sob carga
Crescimento	adicionar blocos de 100 milhões de registros	latência, custo e necessidade de ajuste ou upgrade

Cenários de escrita em massa

Cenário	Execução	Critério e métricas
Escrita simultânea	simular 200.000 estudantes gravando respostas, progresso e eventos em janelas concorrentes	P95 e P99 de gravação, throughput de escrita, taxa de erro, conexões, locks e saturação do banco
Idempotência	reenviar uma amostra de requisições e eventos após timeout ou falha transitória	nenhuma duplicidade de resposta, progresso ou evento; reconciliação consistente
Leitura durante escrita	executar consultas institucionais e de acompanhamento durante a carga de gravação	latência de leitura, taxa de erro e degradação comparadas à linha de base
Recuperação e backpressure	reduzir capacidade ou introduzir falha transitória controlada durante a escrita	fila, retry e limitação de taxa preservam integridade e tornam a recuperação observável

Crítérios complementares - leitura em massa

Dimensão	O que deve ser verificado	Registro esperado
Filtros dinâmicos	períodos arbitrários, escola, tipo de interação e faixa de atividade	resultado sem dependência exclusiva de pré-agregações rígidas
Consistência analítica	agregações comparadas ao PostgreSQL de origem e deltas de materialização	diferença aceitável de 0% entre OLTP e OLAP
Operação	métricas, logs, monitoramento, alertas, replicação e recuperação	esforço operacional classificado e observações do ciclo
Armazenamento	tamanho em disco, compressão, custo e velocidade de leitura	comparativo por solução ou configuração

Crítérios complementares - escrita em massa

Dimensão	O que deve ser verificado	Registro esperado
Ingestão e throughput	gravações por segundo, duração da janela, tempo de confirmação e comportamento da fila	volume aceito, P95/P99, throughput e taxa de erro
Concorrência	pool de conexões, transações em paralelo, locks, deadlocks e saturação de CPU, memória e I/O	ponto de saturação, recursos consumidos e medidas de contenção
Consistência e idempotência	duplicidade, perda, ordem de eventos, retries e reconciliação de respostas e progresso	diferença aceitável de 0% após reconciliação
Leitura concorrente	impacto das escritas nos relatórios e acompanhamentos executados durante a carga	comparativo de latência e erro com a linha de base

Execução e rastreabilidade



- Planejar: identificar as funcionalidades com impacto de escalabilidade, os responsáveis, o ambiente, a janela, os critérios e o vínculo com mudanças relevantes.
- Preparar: registrar versão da aplicação, configuração, dataset, parâmetros de consulta, massa de leitura, massa de escrita e isolamento do ambiente.
- Executar: coletar latência P50, P95 e P99, throughput de leitura e escrita, erros, consumo de recursos, conexões, locks, duração e comportamento durante a degradação.
- Analisar: comparar com os critérios e com o trimestre anterior, classificar regressões e registrar limitações do experimento.
- Concluir: registrar resultados, gráficos, logs, configuração do teste, decisão técnica, responsáveis e ações corretivas com prazo.
- Reexecutar: repetir os cenários afetados após correções materiais e preservar o comparativo antes e depois.