

Experimental Performance Evaluation Among Cloud Infrastructure Providers Under Different Load Levels



**Avaliação Experimental de Desempenho Entre Provedores de
Infraestrutura em Nuvem Sob Diferentes Níveis de Carga**



Denis B. Oliveira

Instituto Federal do Sul de
Minas Gerais
IFSULDEMINAS



Ricardo R. de Oliveira

Instituto Federal do Sul de
Minas Gerais
IFSULDEMINAS



Ricardo F. Vilela

Universidade Federal de
Santa Catarina
UFSC



Victor H. S. C. Pinto

Universidade Federal do Pará
UFPA



Roberto N. Ungarelli

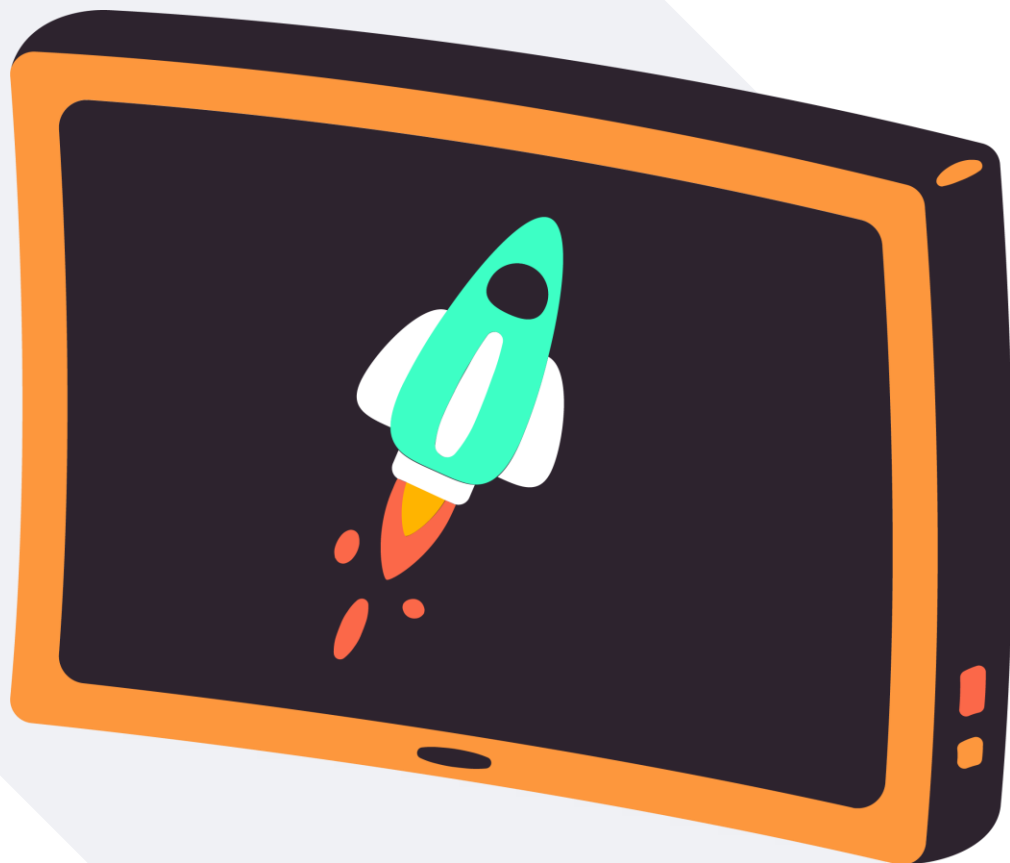
Zello

MOTIVAÇÃO




Como definir um provedor?

Considerando as possíveis adversidades para o provimento de uma aplicação sob o modelo de nuvem, um dos desafios é a escolha adequada de um provedor **IaaS** para melhor atender as necessidades de uma determinada nova aplicação ou migrar um projeto já existente.





Objetivos

-  **Realizar um estudo experimental** para avaliar o comportamento de provedores IaaS submetidos a diferentes níveis de carga.
-  **Compartilhar os resultados preliminares** para auxiliar a tomada de decisão na escolha de um provedor.
-  **Compartilhar uma abordagem de avaliação** que pode ser adaptada e replicada de modo a atender outros provedores, aplicações e tecnologias.

Questão 1

Como cada provedor de infraestrutura em nuvem avaliado se comporta em relação a diferentes números de usuários realizando requisições simultâneas?



Questão 2

Como esses provedores entregam satisfação ao usuário em relação ao desempenho das requisições?

Provedores



Microsoft Azure



Google Cloud



IBM Cloud



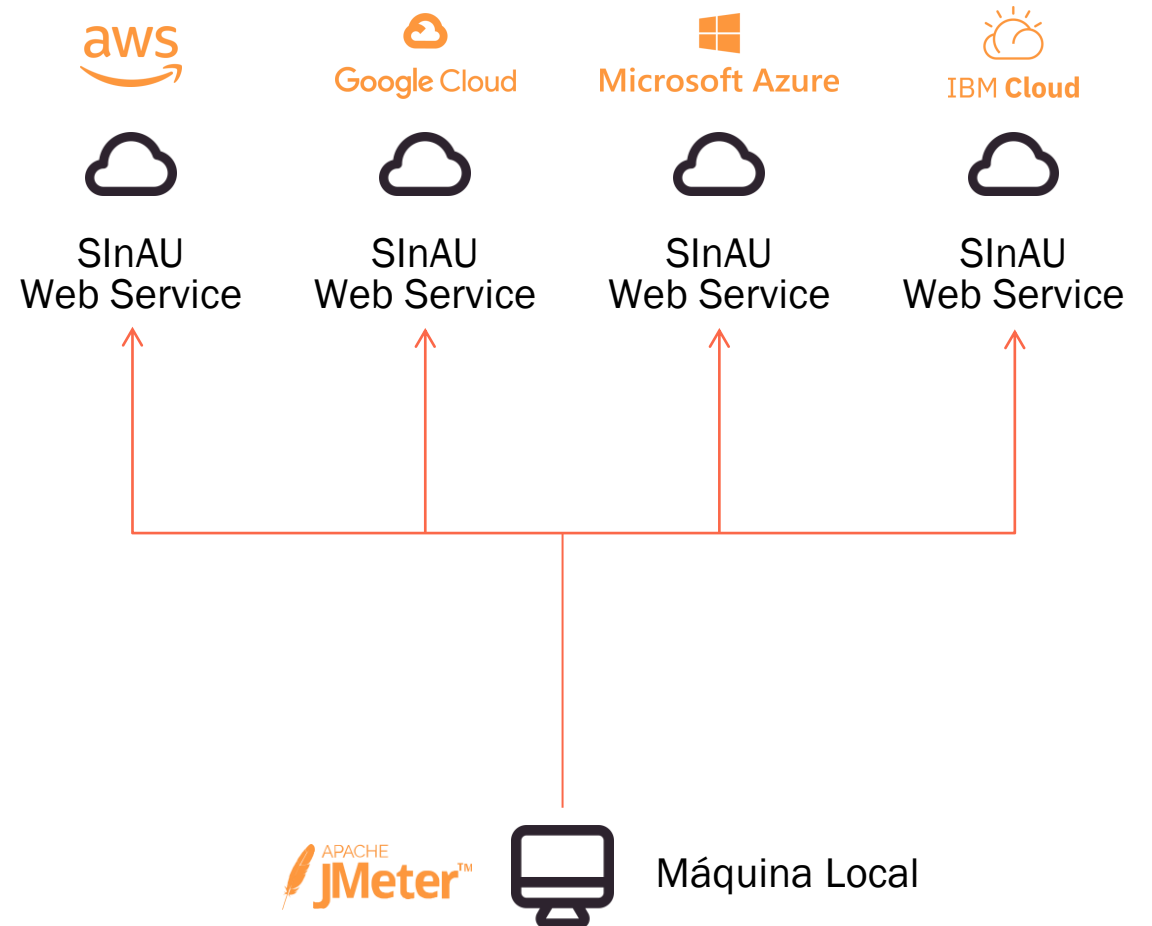
Estudo realizado

Aplicação SinAU

- RESTFul
- Desenvolvida em Java Web
- Possui 56 endpoints únicos

JMeter

- Desenvolvido pela Apache
- Ferramenta de realização de testes de carga e estresse



Configurações

Configuração	Local	Nuvem
Sistema Operacional	Ubuntu 18.04 LTS	Ubuntu 16.04 LTS
Localização	Brasil (Sudeste)	EUA (Oeste)
vCPU	4	2
Memória (GB)	8gb	4gb
Java JRE	1.8	1.8
MySQL	-	5.6
Apache Tomcat	-	7.0.92
Apache JMeter	5.2.1	-

O banco de dados **MySQL** e o **Apache Tomcat** estão presentes somente no nos ambientes nuvem pois são necessários para o funcionamento da **aplicação SinAU**.

O **Java JRE** é necessário para a execução do **JMeter** na máquina local e da aplicação desenvolvida em Java Web.

Fatores

Foram definidos

5 cenários de teste

Cada cenário foi executado

10x em cada provedor

Houve um total de

200 testes

Threads	Ramp-ups	Tempos
100	10	50
200	20	100
300	30	150
400	40	200
500	50	250

- Número de usuários que serão simulados.
Ex: 100.
- Tempo em que todos os usuários ficarão ativos.
Ex: 10 segundos, ou seja, a cada 1 segundo 10 novos usuários ficarão ativos.
- Tempo de duração total do teste.
Ex: 50 segundos, ou seja, dentro desse tempo todas as requisições feitas pelos usuários ativos será avaliada.

Rotina de Teste

24 ações

Cada ação possui uma ou mais requisições, totalizando

61 requisições

Existe um intervalo randômico entre 0 e 5 segundos entre cada requisição

Ação	Nome
001	Incluir disciplina
002	Visualizar disciplina
003	Incluir aluno
004	Visualizar aluno
005	Incluir professor
006	Visualizar professor
007	Incluir universidade
008	Visualizar universidade
009	Incluir departamento
010	Visualizar departamento
011	Incluir curso
012	Visualizar curso
013	Alterar universidade
014	Remover universidade

Ação	Etapa	Método	URI
001	01	GET	/MWS/rest/disciplinas
	02	POST	/MWS/rest/disciplinas
002	01	GET	/MWS/rest/disciplinas
	02	GET	/MWS/rest/disciplinas/x
	03	GET	/MWS/rest/disciplinas/x/alunos
	04	GET	/MWS/rest/disciplinas/x/curso
	05	GET	/MWS/rest/disciplinas/x/professor

Métricas

🕒 Latência

Intervalo de tempo entre o envio da requisição do cliente e o início do recebimento da resposta da aplicação.

⚠ Taxa de Erro

É a porcentagem de requisições que falharam.

👤 Vazão de Dados

Número de requisições por segundo que o servidor consegue processar.

⌚ Tempo de Resposta

Tempo gasto pela aplicação para responder a solicitação do usuário.

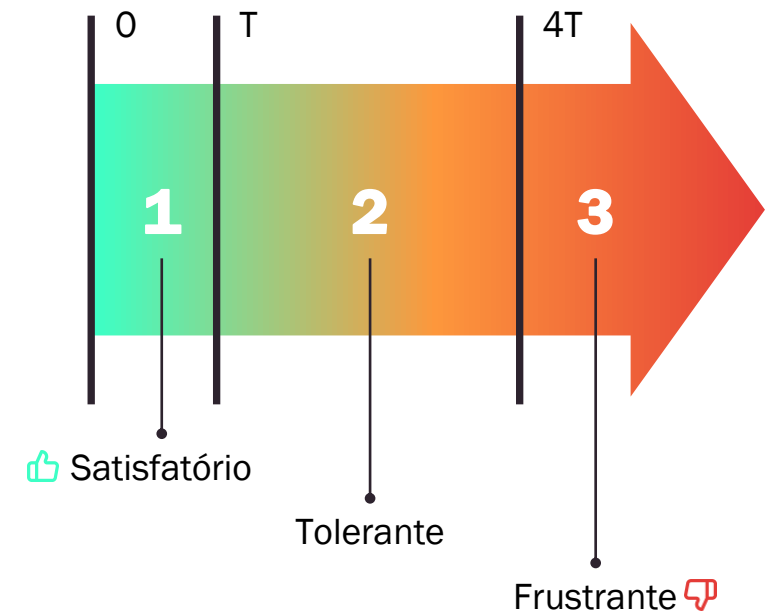


Apdex Application Performance Index

É uma forma numérica de se avaliar a satisfação do usuário em relação ao desempenho do serviço, onde um tempo **T** é definido como o limite aceitável e **4T** como limite tolerado, e qualquer requisição maior que isso se torna frustrante.

Para este estudo **T** foi definido como **2 segundos**

O **Apdex** pode ser dado em um valor entre **0** em **1**, onde **1** representa uma satisfação total.



$$\text{Apdex}_T = \frac{\text{Satisfatória} + \frac{\text{Tolerantes}}{2}}{\text{Total de requisições}}$$

Respondendo às questões



Questão 1

Como cada provedor de infraestrutura em nuvem avaliado se comporta em relação a diferentes números de usuários realizando requisições simultâneas?

 **Latência**

 **Taxa de Erro**

 **Vazão de Dados**

 **Tempo de Resposta**

Questão 2

Como esses provedores entregam satisfação ao usuário em relação ao desempenho das requisições?

Apdex



Resultados



Latência





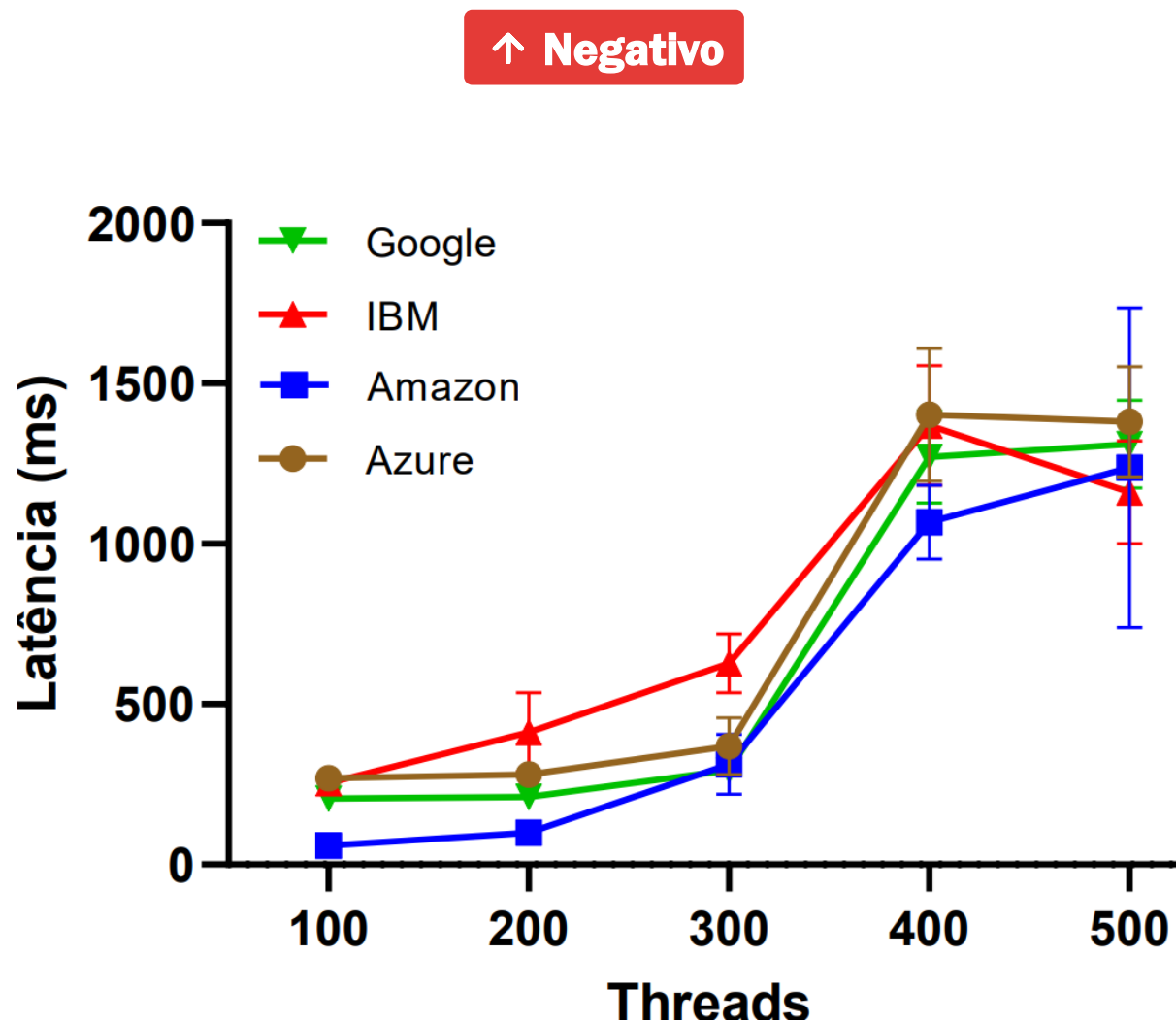
Latência

Em todos os provedores há uma considerável estabilidade até **300 usuários**.

Os níveis de latência atingem seu pico nos **400 usuários**.

A **Amazon** se destaca positivamente por atender as requisições mais rapidamente em praticamente todos os testes.

A **IBM** por outro lado é a primeira a sofrer com os impactos das mudanças de tráfego.





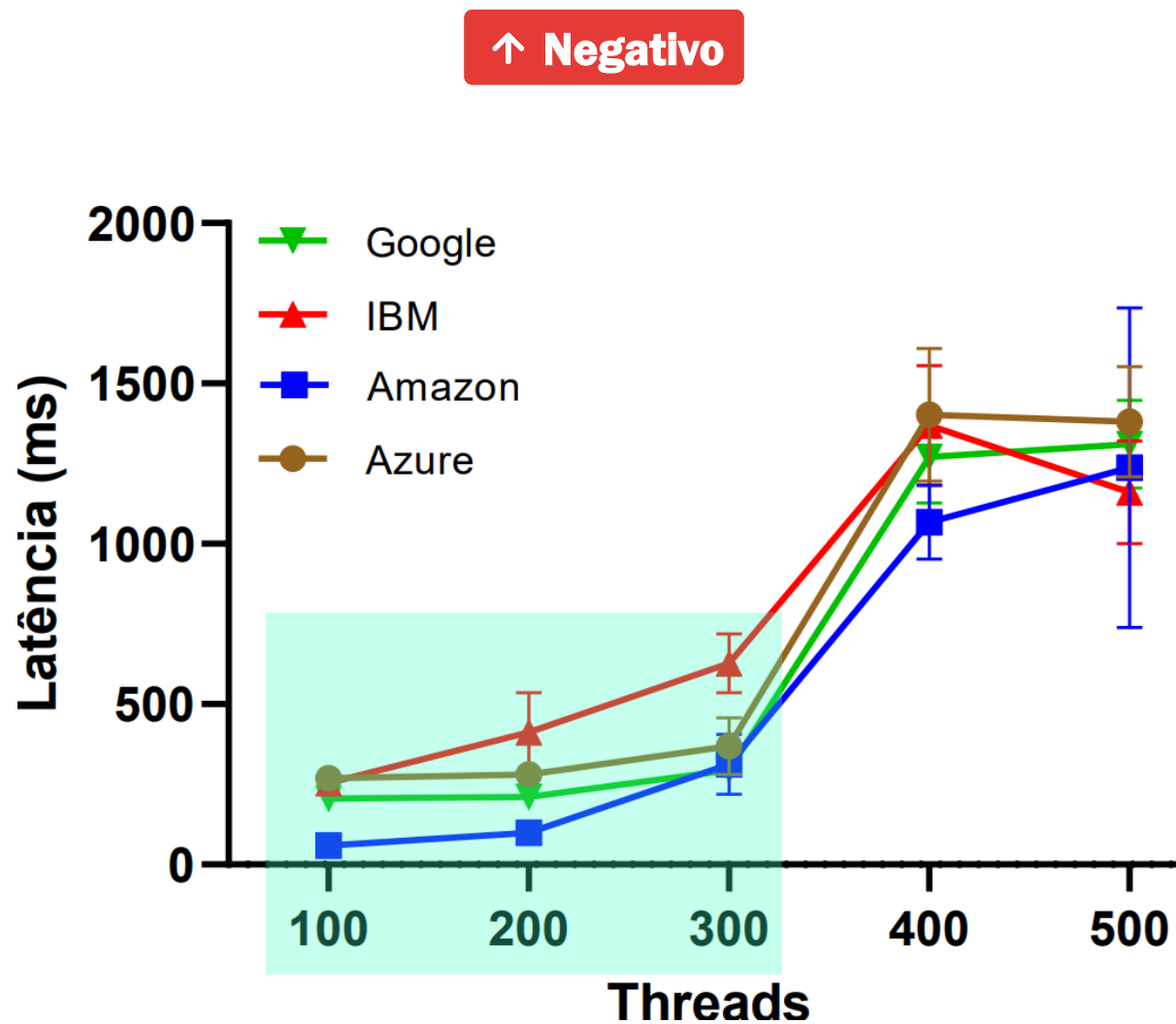
Latência

Em todos os provedores há uma considerável estabilidade até **300 usuários**.

Os níveis de latência atingem seu pico nos **400 usuários**.

A **Amazon** se destaca positivamente por atender as requisições mais rapidamente em praticamente todos os testes.

A **IBM** por outro lado é a primeira a sofrer com os impactos das mudanças de tráfego.





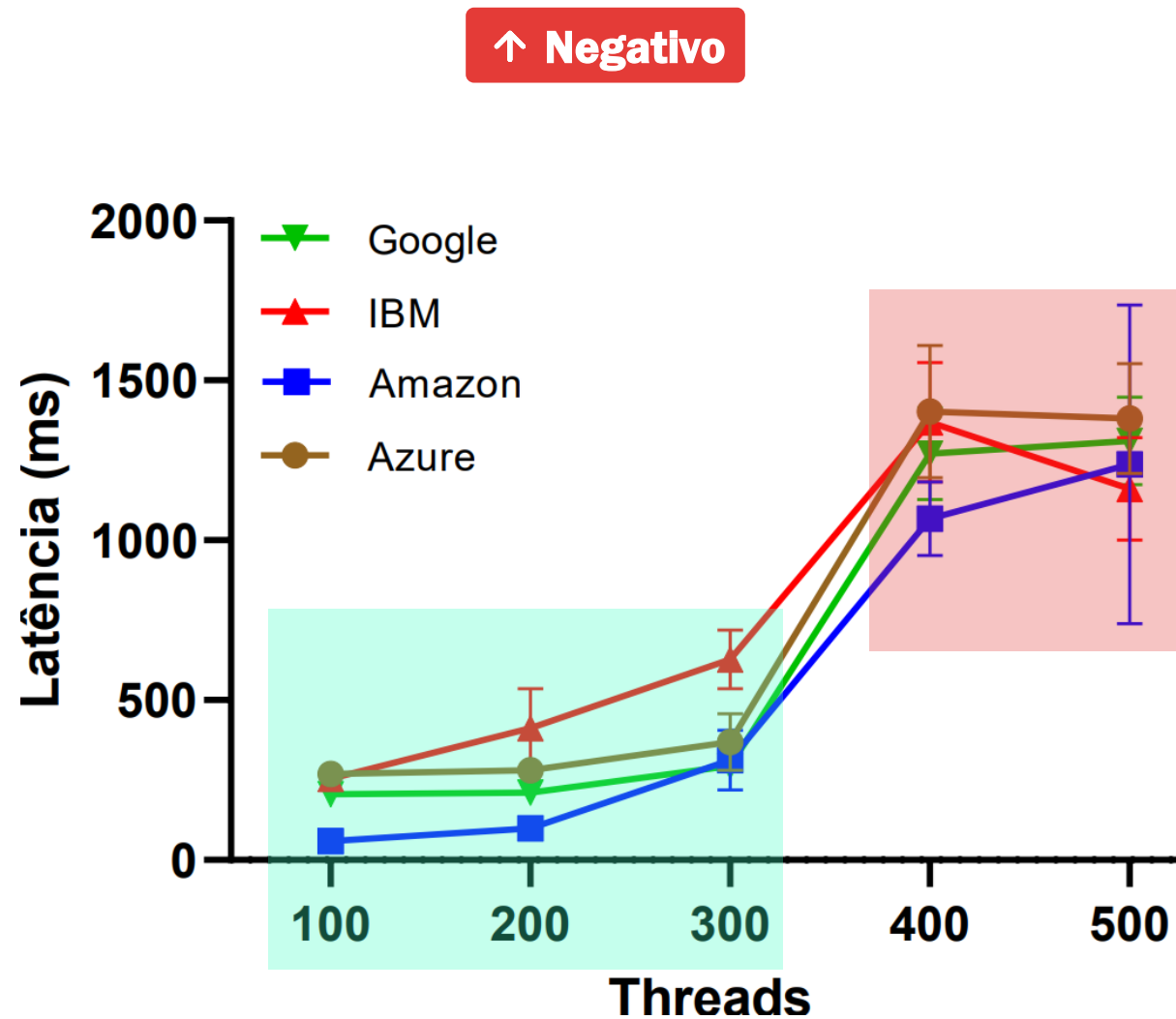
Latência

Em todos os provedores há uma considerável estabilidade até **300 usuários**.

Os níveis de latência atingem seu pico nos **400 usuários**.

A **Amazon** se destaca positivamente por atender as requisições mais rapidamente em praticamente todos os testes.

A **IBM** por outro lado é a primeira a sofrer com os impactos das mudanças de tráfego.





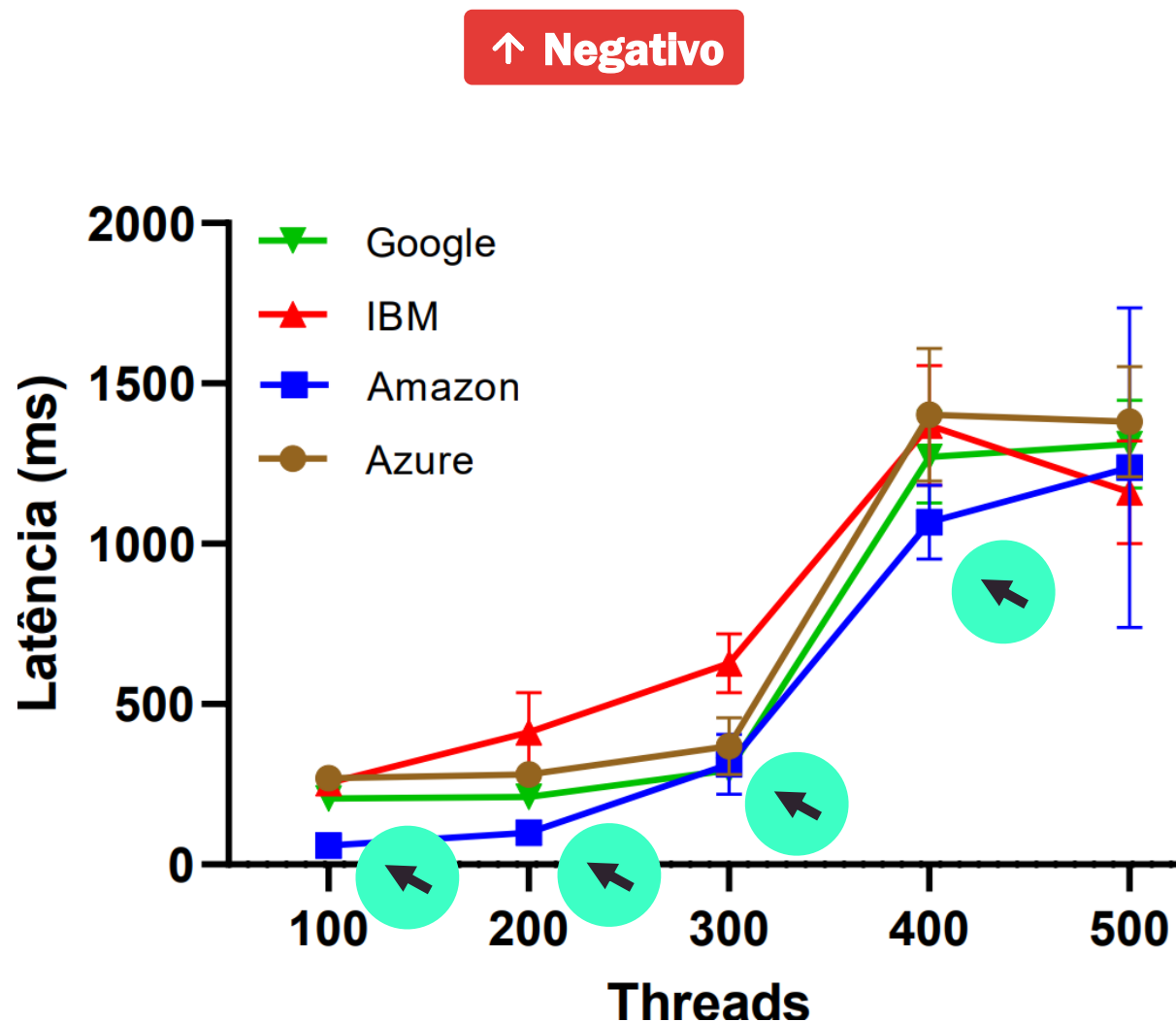
Latência

Em todos os provedores há uma considerável estabilidade até **300 usuários**.

Os níveis de latência atingem seu pico nos **400 usuários**.

A **Amazon** se destaca positivamente por atender as requisições mais rapidamente em praticamente todos os testes.

A **IBM** por outro lado é a primeira a sofrer com os impactos das mudanças de tráfego.





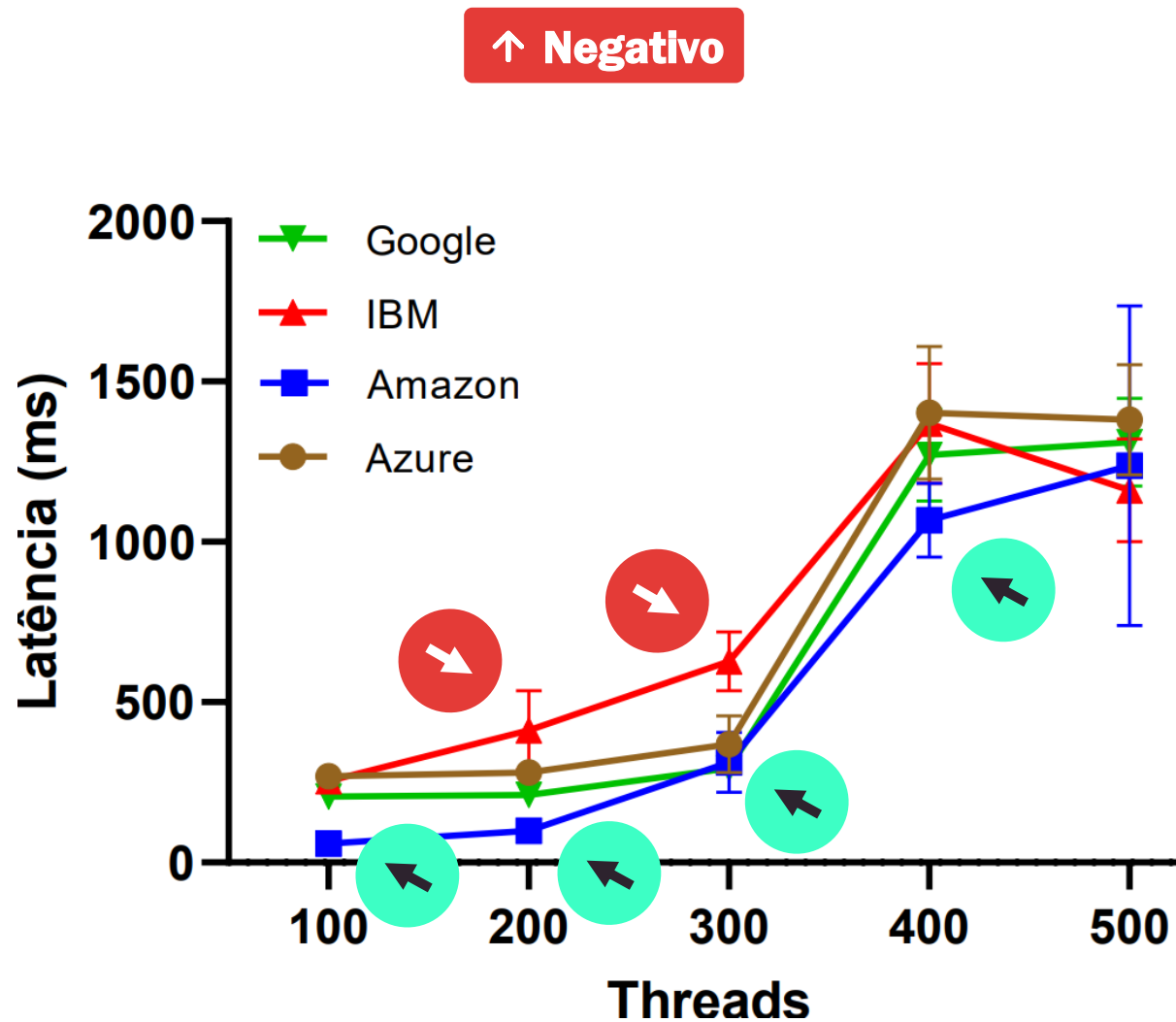
Latência

Em todos os provedores há uma considerável estabilidade até **300 usuários**.

Os níveis de latência atingem seu pico nos **400 usuários**.

A **Amazon** se destaca positivamente por atender as requisições mais rapidamente em praticamente todo os testes.

A **IBM** por outro lado é a primeira a sofrer com os impactos das mudanças de tráfego.





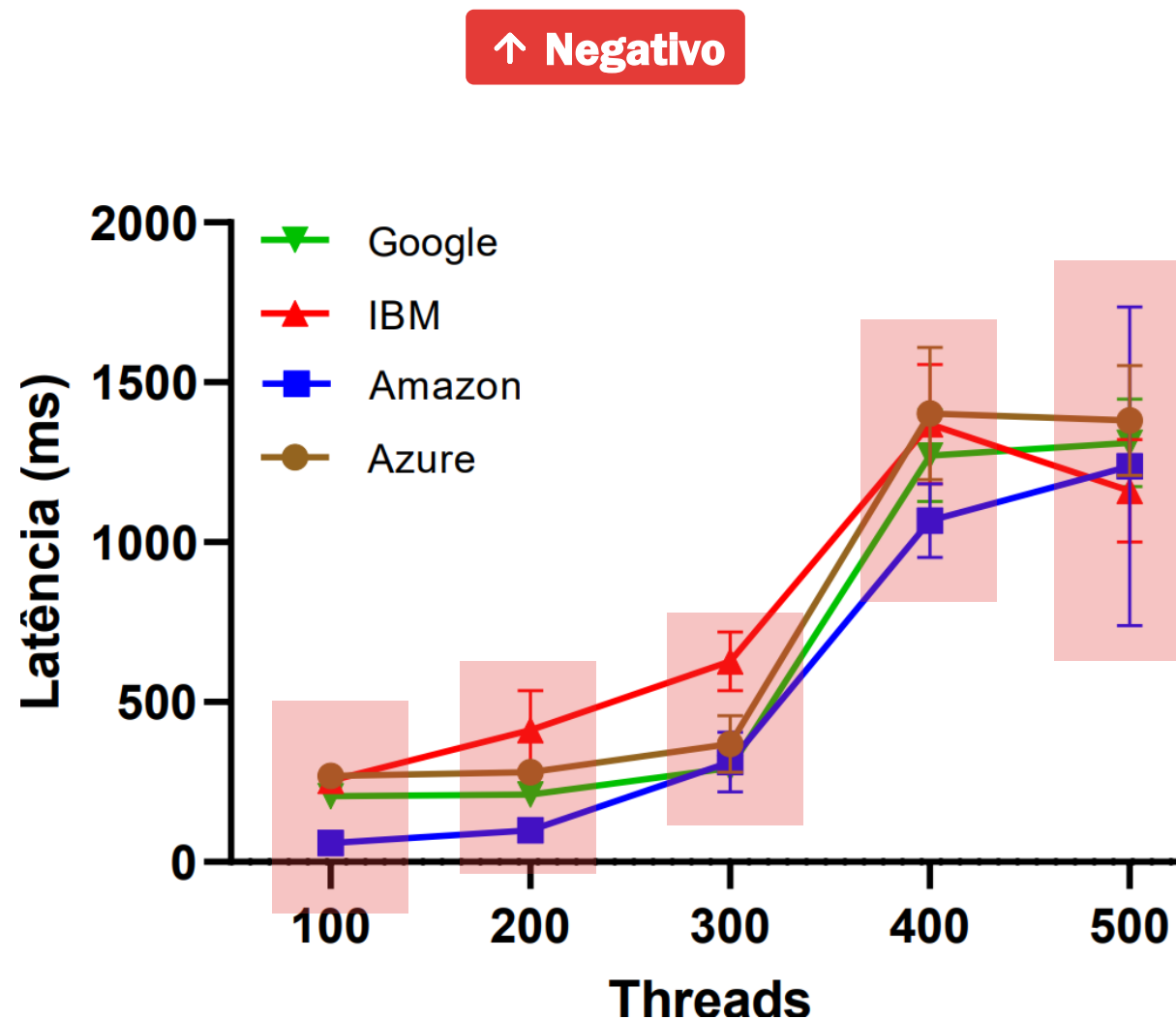
Latência

Questão 1

Como cada provedor de infraestrutura em nuvem avaliado se comporta em relação a diferentes números de usuários realizando requisições simultâneas?

Conclusão

Existem diferenças entre o comportamento dos diferentes provedores, mesmo demonstrando a tendência de convergir para comportamentos semelhantes no teste de estresse máximo.





Taxa de Erro





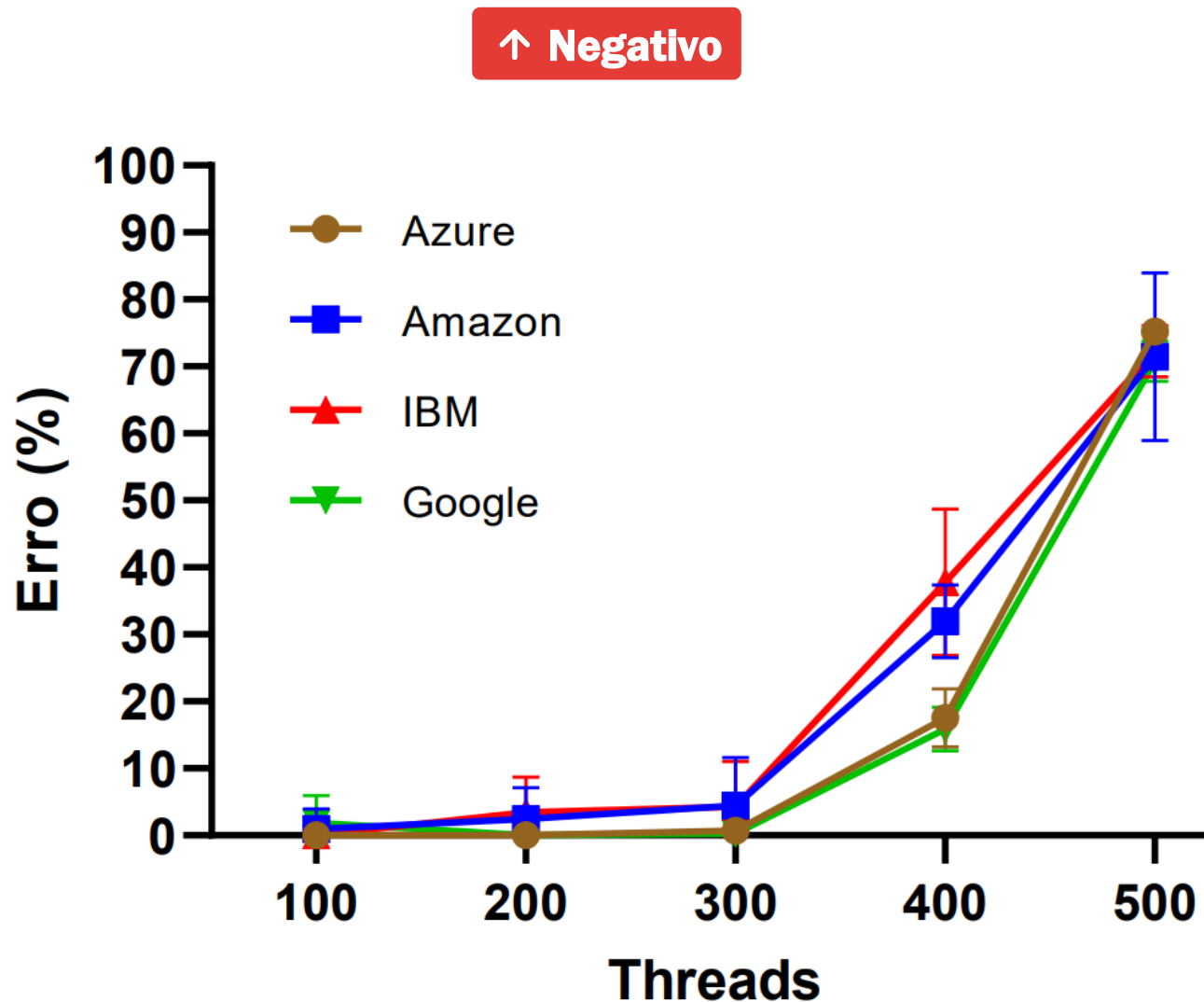
Taxa de Erro

Em todos os provedores há estabilidade até **300 usuários**.

Desse ponto em diante a taxa cresce com uma **tendência linear**.

Azure e **Google** se destacam **positivamente** por, praticamente da mesma forma, demonstrarem uma maior consistência em relação ao número de requisições atendidas.

Amazon e **IBM** não se saem muito diferente, exceto no teste com **400 usuários**, onde apresentam mais de **15% de erros** em comparação às outras.





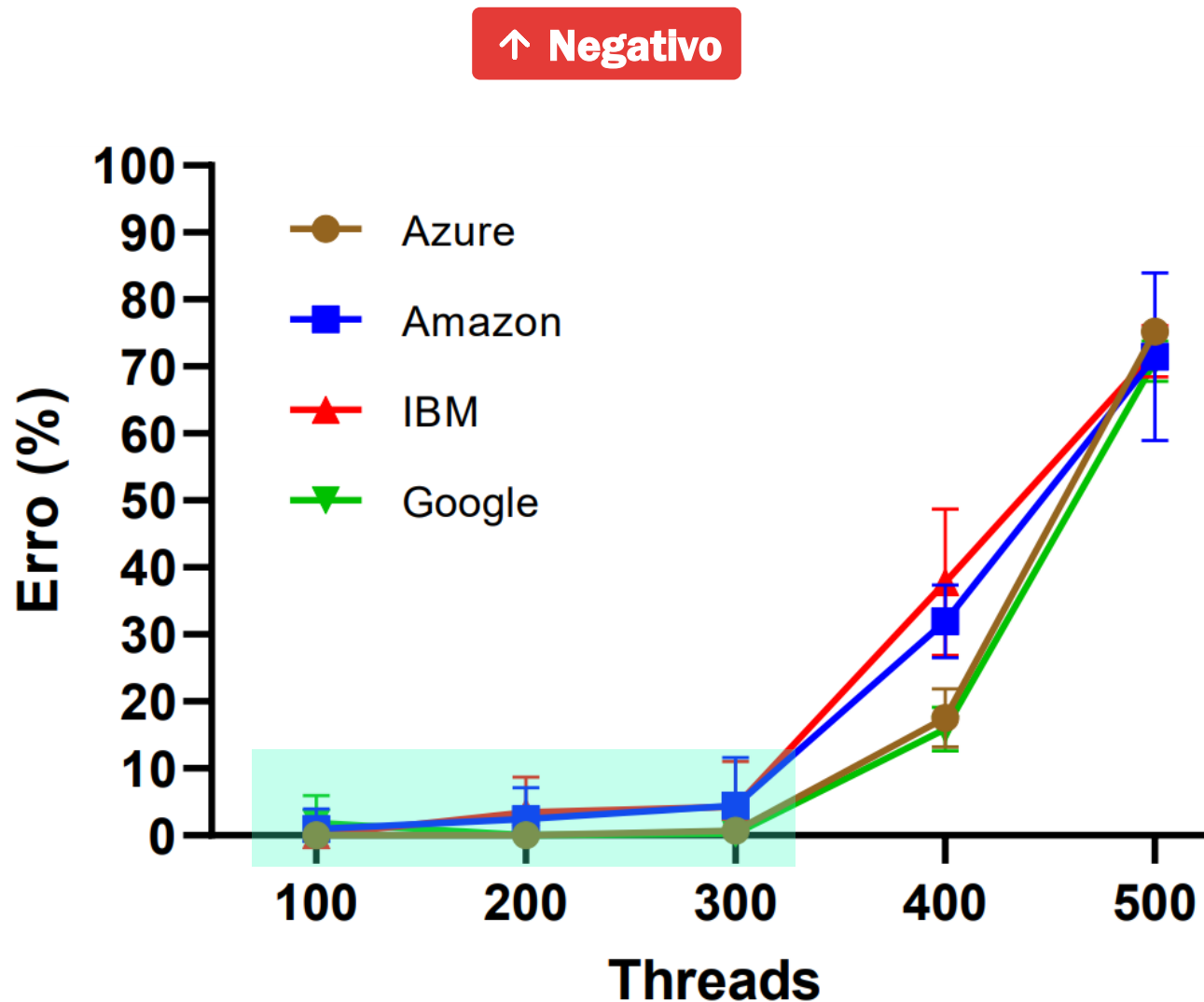
Taxa de Erro

Em todos os provedores há estabilidade até **300 usuários**.

Desse ponto em diante a taxa cresce com uma **tendência linear**.

Azure e **Google** se destacam **positivamente** por, praticamente da mesma forma, demonstrarem uma maior consistência em relação ao número de requisições atendidas.

Amazon e **IBM** não se saem muito diferente, exceto no teste com **400 usuários**, onde apresentam mais de **15% de erros** em comparação às outras.





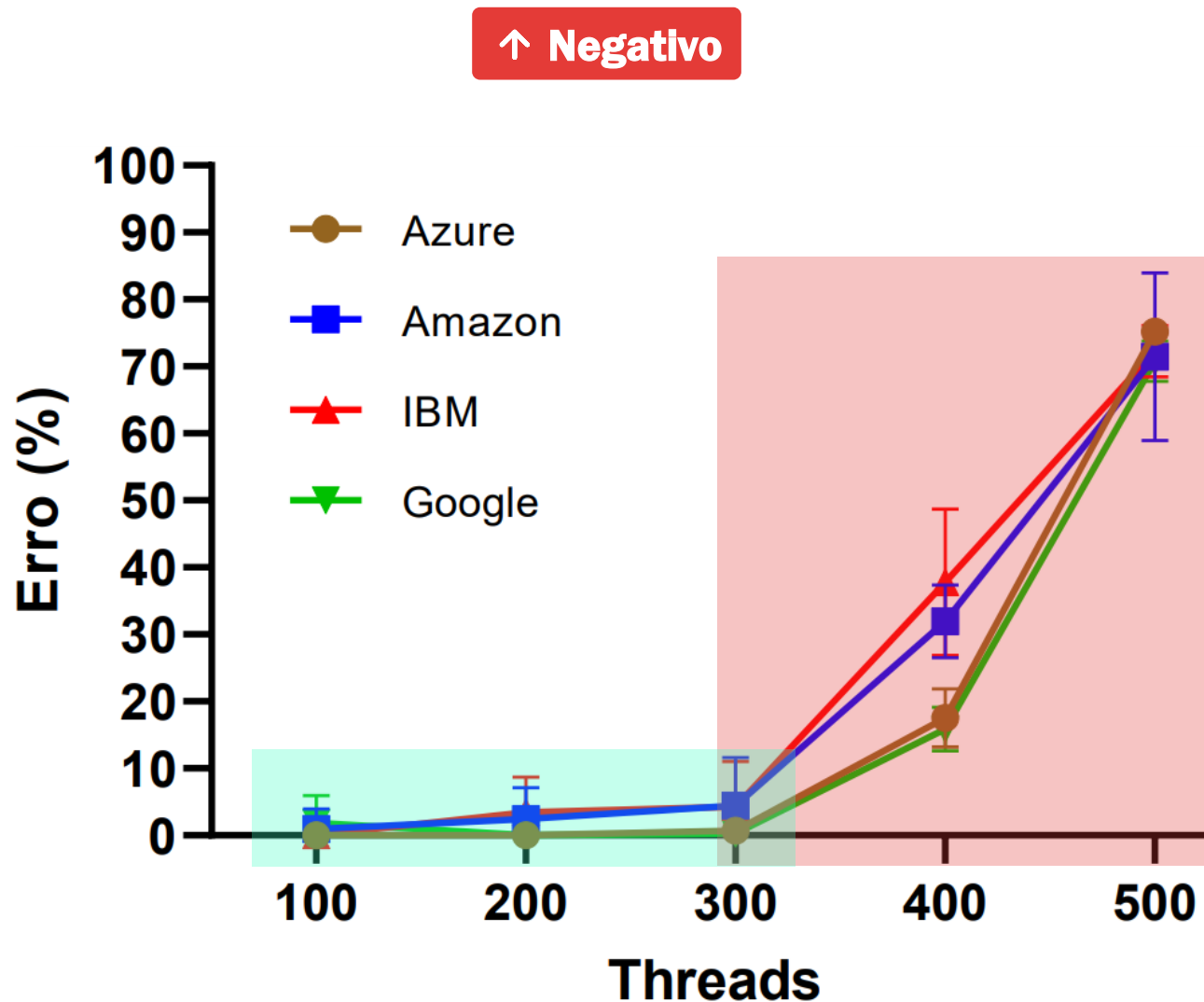
Taxa de Erro

Em todos os provedores há estabilidade até **300 usuários**.

Desse ponto em diante a taxa cresce com uma **tendência linear**.

Azure e **Google** se destacam **positivamente** por, praticamente da mesma forma, demonstrarem uma maior consistência em relação ao número de requisições atendidas.

Amazon e **IBM** não se saem muito diferente, exceto no teste com **400 usuários**, onde apresentam mais de **15% de erros** em comparação às outras.





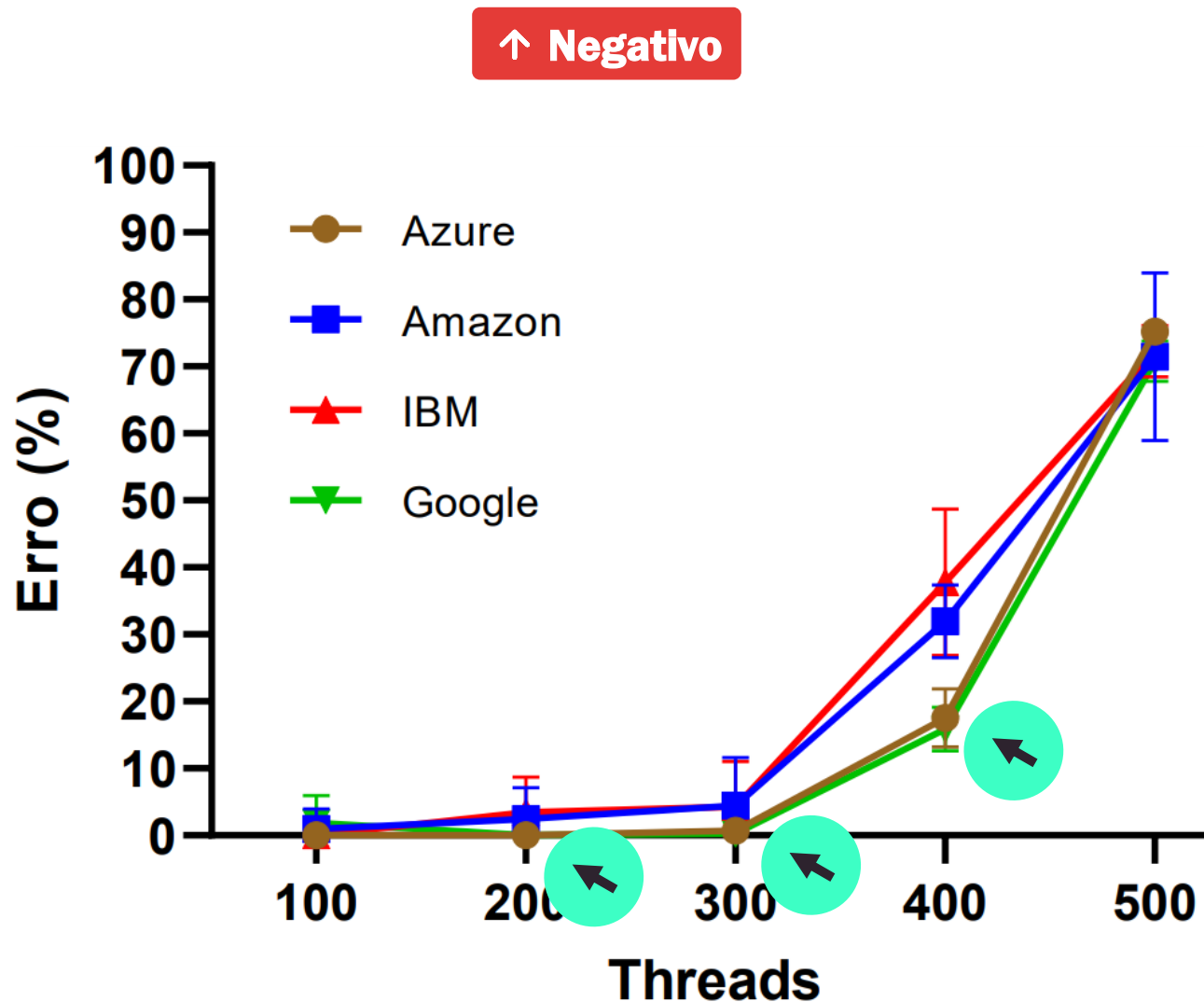
Taxa de Erro

Em todos os provedores há estabilidade até **300 usuários**.

Desse ponto em diante a taxa cresce com uma **tendência linear**.

Azure e **Google** se destacam **positivamente** por, praticamente da mesma forma, demonstrarem uma maior consistência em relação ao número de requisições atendidas.

Amazon e **IBM** não se saem muito diferente, exceto no teste com **400 usuários**, onde apresentam mais de **15% de erros** em comparação às outras.





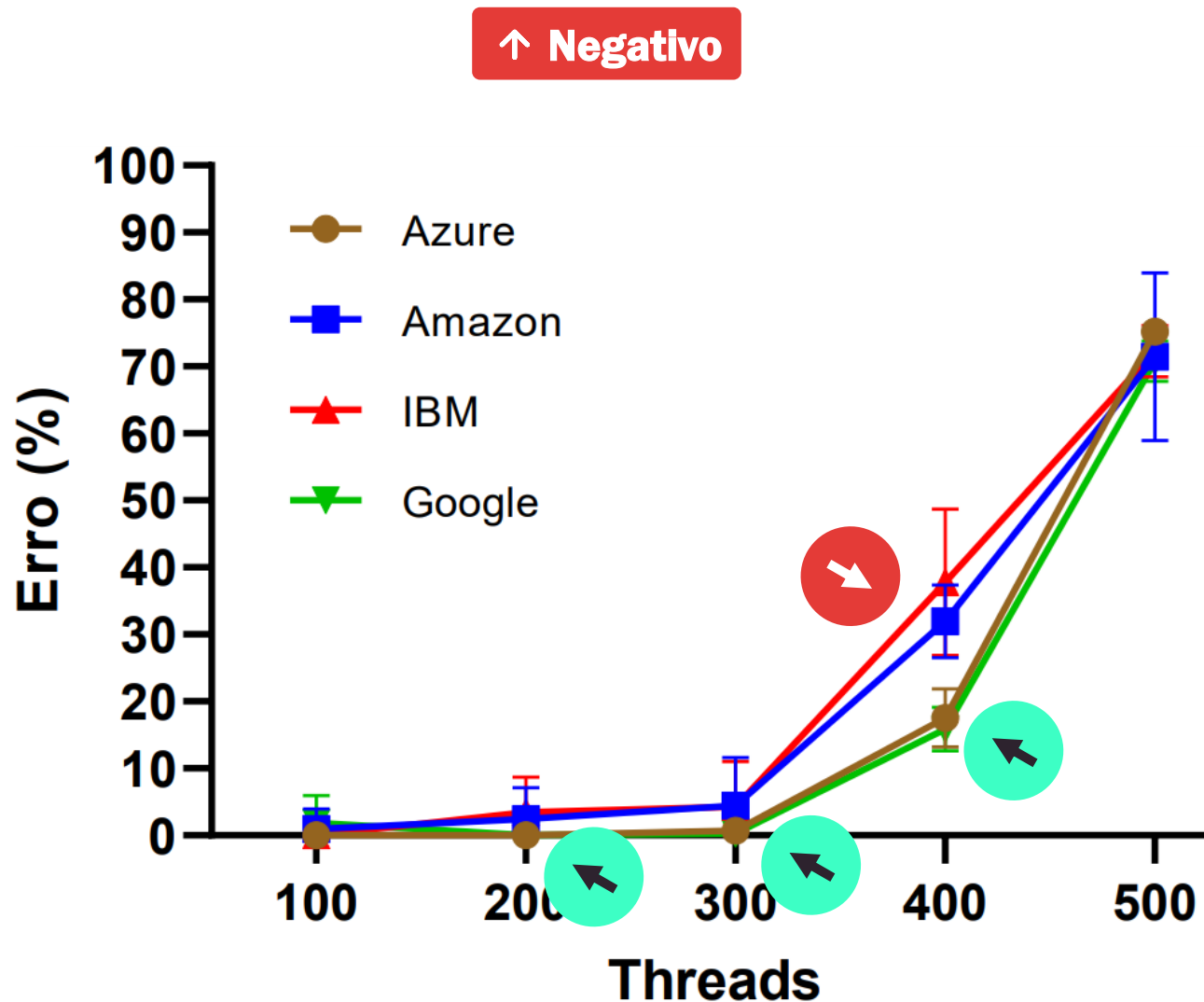
Taxa de Erro

Em todos os provedores há estabilidade até **300 usuários**.

Desse ponto em diante a taxa cresce com uma **tendência linear**.

Azure e **Google** se destacam **positivamente** por, praticamente da mesma forma, demonstrarem uma maior consistência em relação ao número de requisições atendidas.

Amazon e **IBM** não se saem muito diferente, exceto no teste com **400 usuários**, onde apresentam mais de **15% de erros** em comparação às outras.





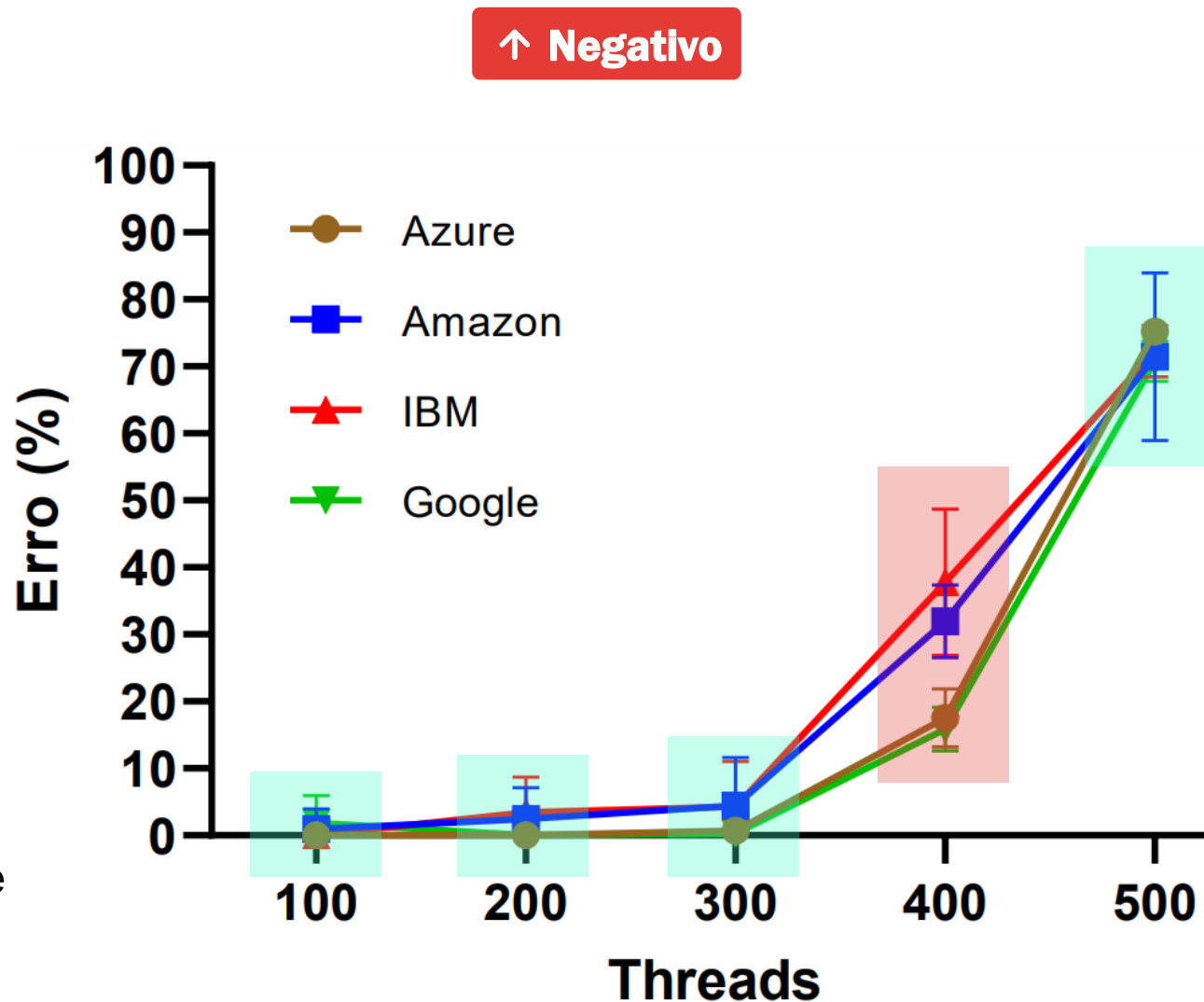
Taxa de Erro

Questão 1

Como cada provedor de infraestrutura em nuvem avaliado se comporta em relação a diferentes números de usuários realizando requisições simultâneas?

Conclusão

Apesar de se comportarem de maneira semelhante na maior parte do tempo, no teste de 400 usuários houve diferenças no comportamento dos provedores.





Vazão de Dados



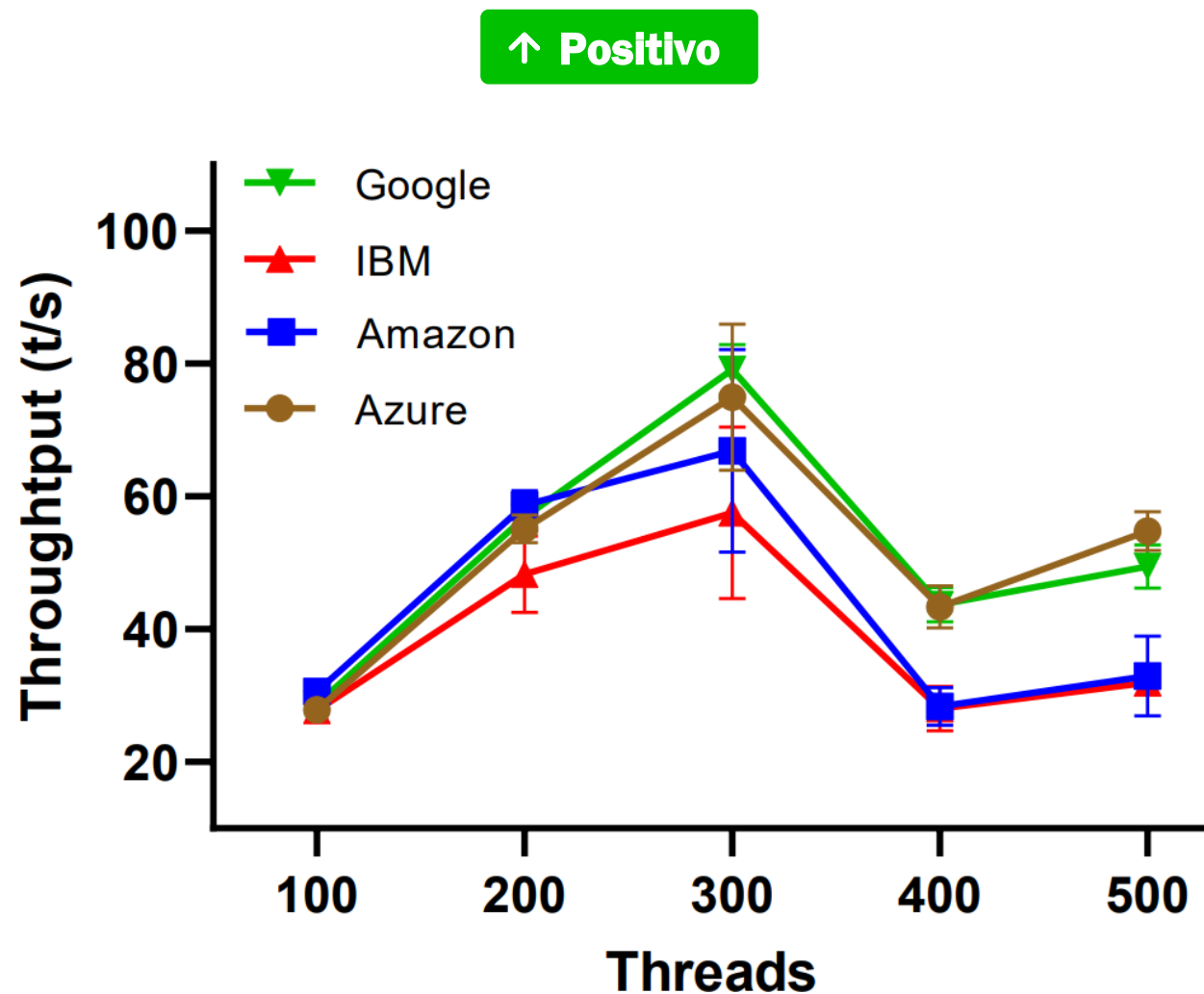


Vazão de Dados

Em um cenário ideal, a vazão deve-se manter crescente de forma linear, assim como a quantidade de usuários. É isso o que acontece até **300 usuários**.

Em seguida, **entre 300 e 400 usuários** acontece uma queda, fortemente **influenciada pela taxa de erros**.

Comparado a **400 usuários**, o teste de **500** foi capaz de gerar uma maior transferência de dados, porém níveis de aproximadamente 80% de erro foram atingidos nesse momento.



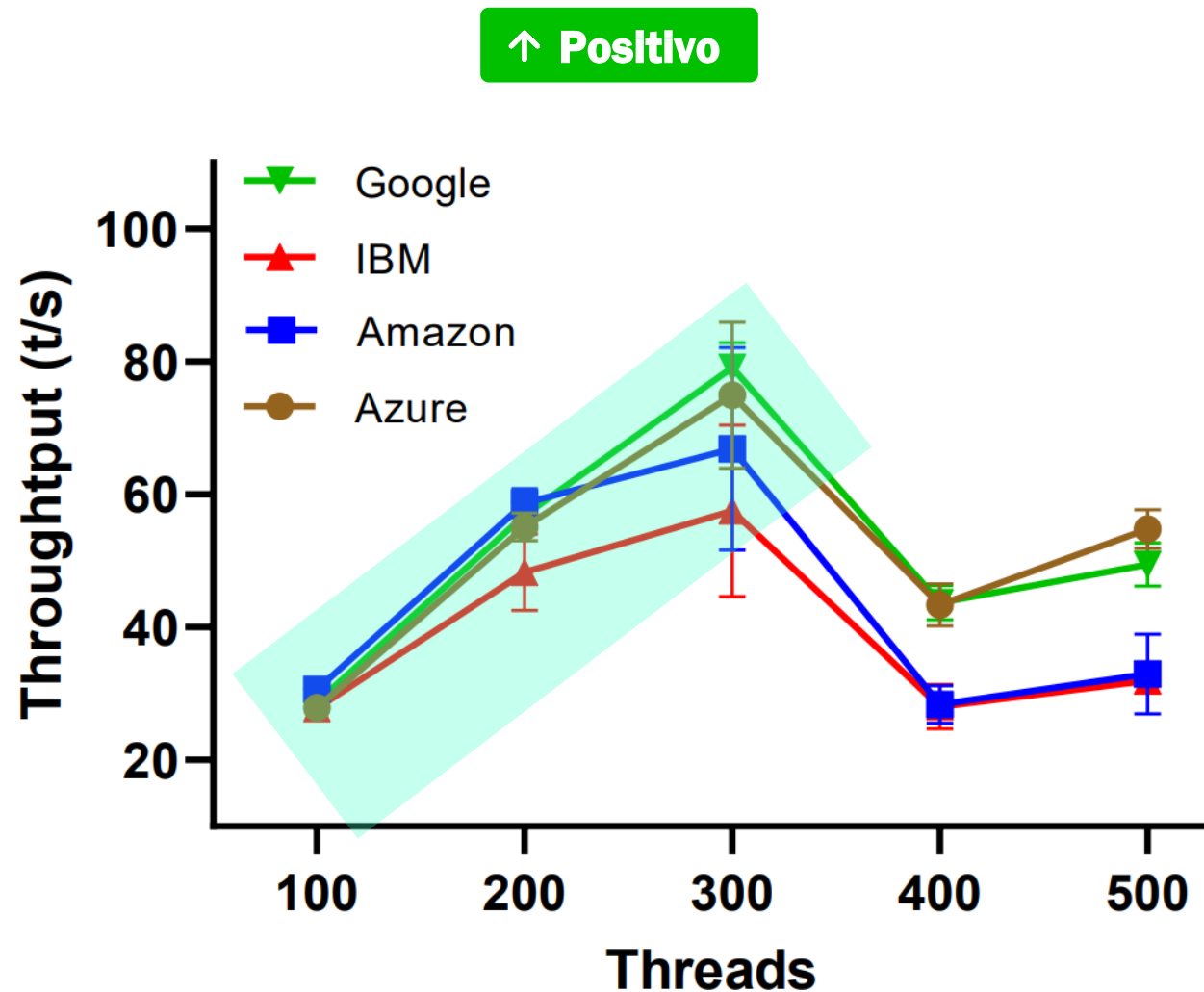


Vazão de Dados

Em um cenário ideal, a vazão deve-se manter crescente de forma linear, assim como a quantidade de usuários. É isso o que acontece até **300 usuários**.

Em seguida, **entre 300 e 400 usuários** acontece uma queda, fortemente **influenciada pela taxa de erros**.

Comparado a **400 usuários**, o teste de **500** foi capaz de gerar uma maior transferência de dados, porém níveis de aproximadamente 80% de erro foram atingidos nesse momento.



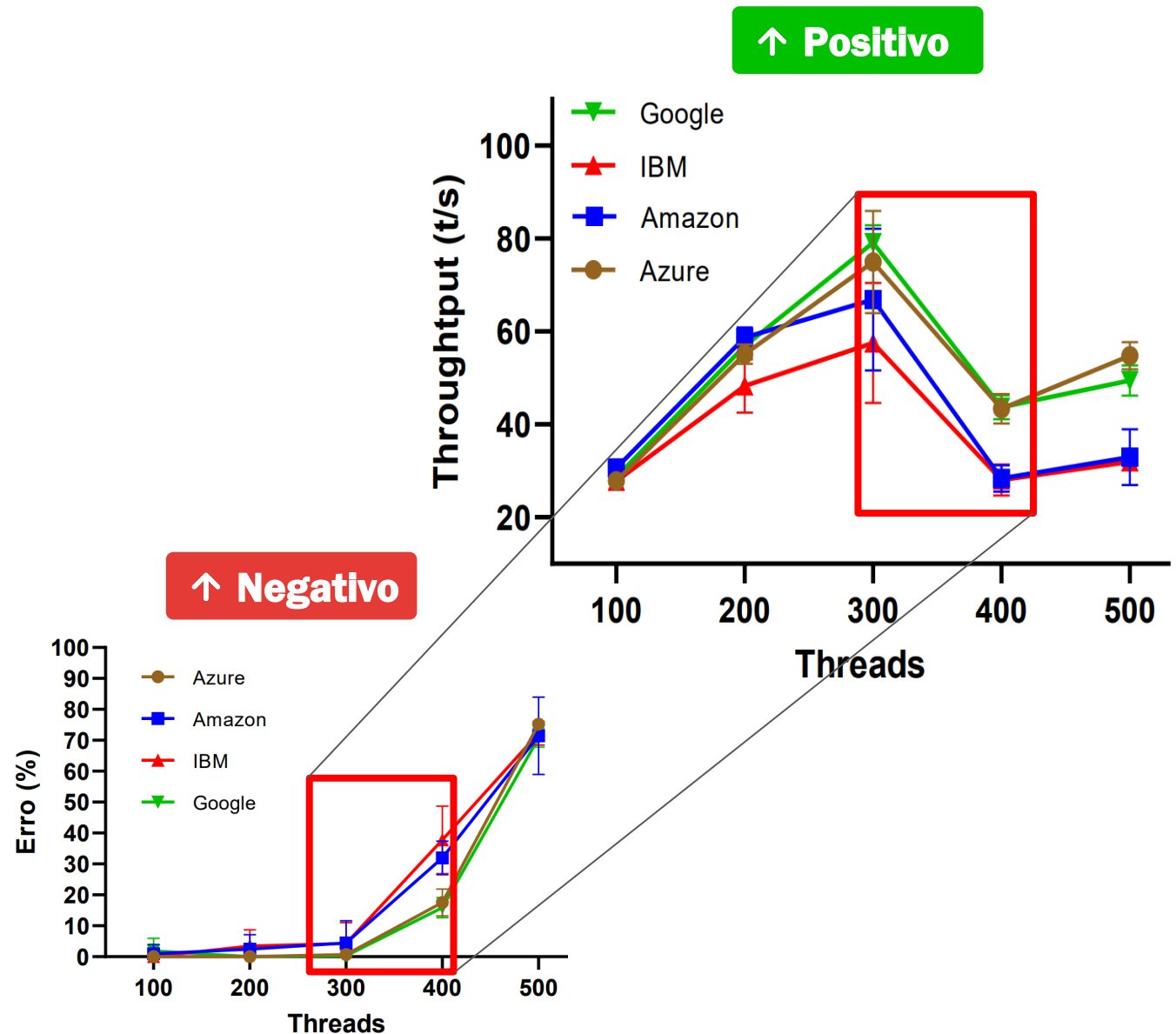


Vazão de Dados

Em um cenário ideal, a vazão deve-se manter crescente de forma linear, assim como a quantidade de usuários. É isso o que acontece até **300 usuários**.

Em seguida, **entre 300 e 400 usuários** acontece uma queda, fortemente **influenciada pela taxa de erros**.

Comparado a **400 usuários**, o teste de **500** foi capaz de gerar uma maior transferência de dados, porém níveis de aproximadamente 80% de erro foram atingidos nesse momento.



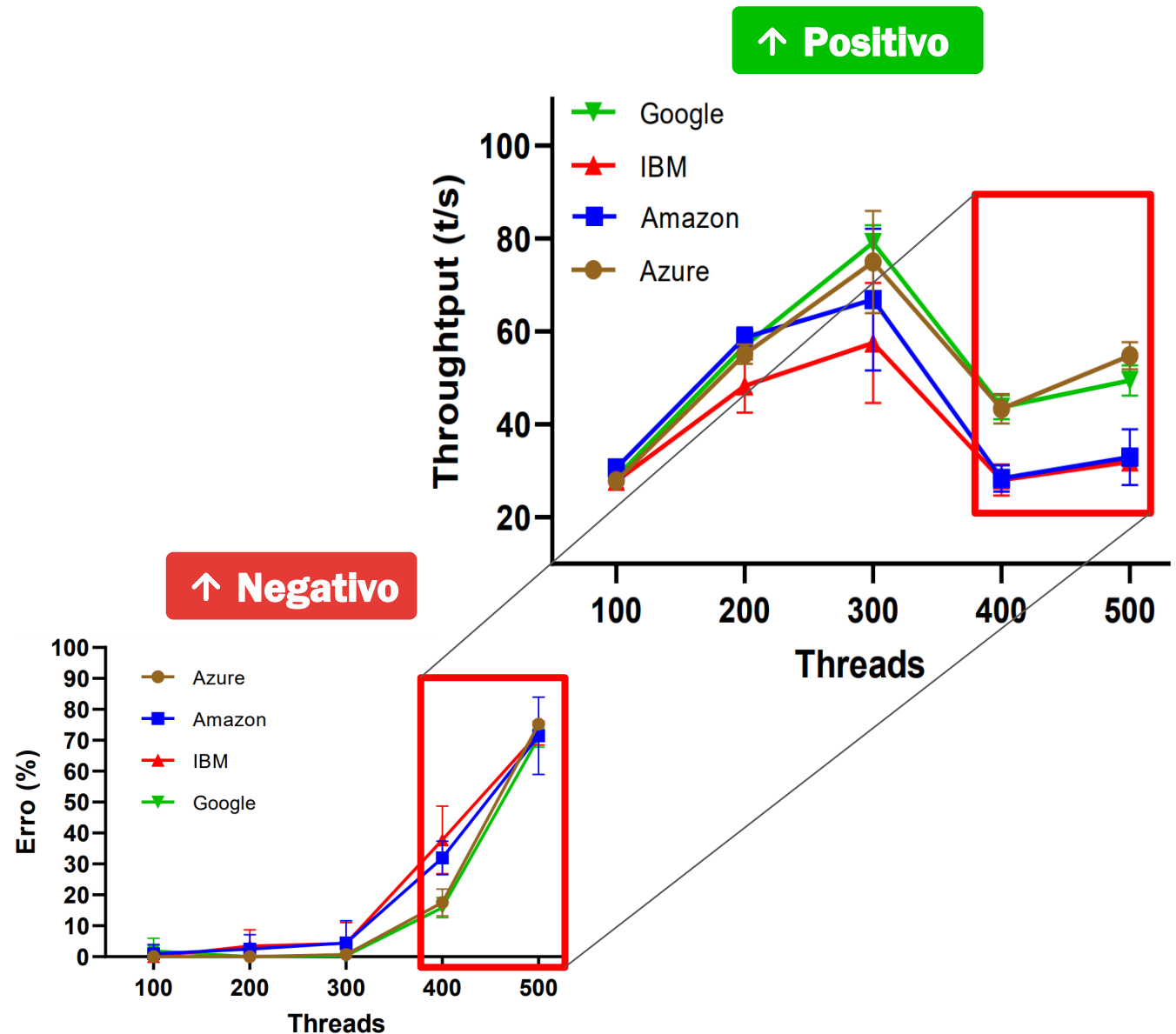


Vazão de Dados

Em um cenário ideal, a vazão deve-se manter crescente de forma linear, assim como a quantidade de usuários. É isso o que acontece até **300 usuários**.

Em seguida, **entre 300 e 400 usuários** acontece uma queda, fortemente **influenciada pela taxa de erros**.

Comparado a **400 usuários**, o teste de **500** foi capaz de gerar uma maior transferência de dados, porém níveis de aproximadamente 80% de erro foram atingidos nesse momento.

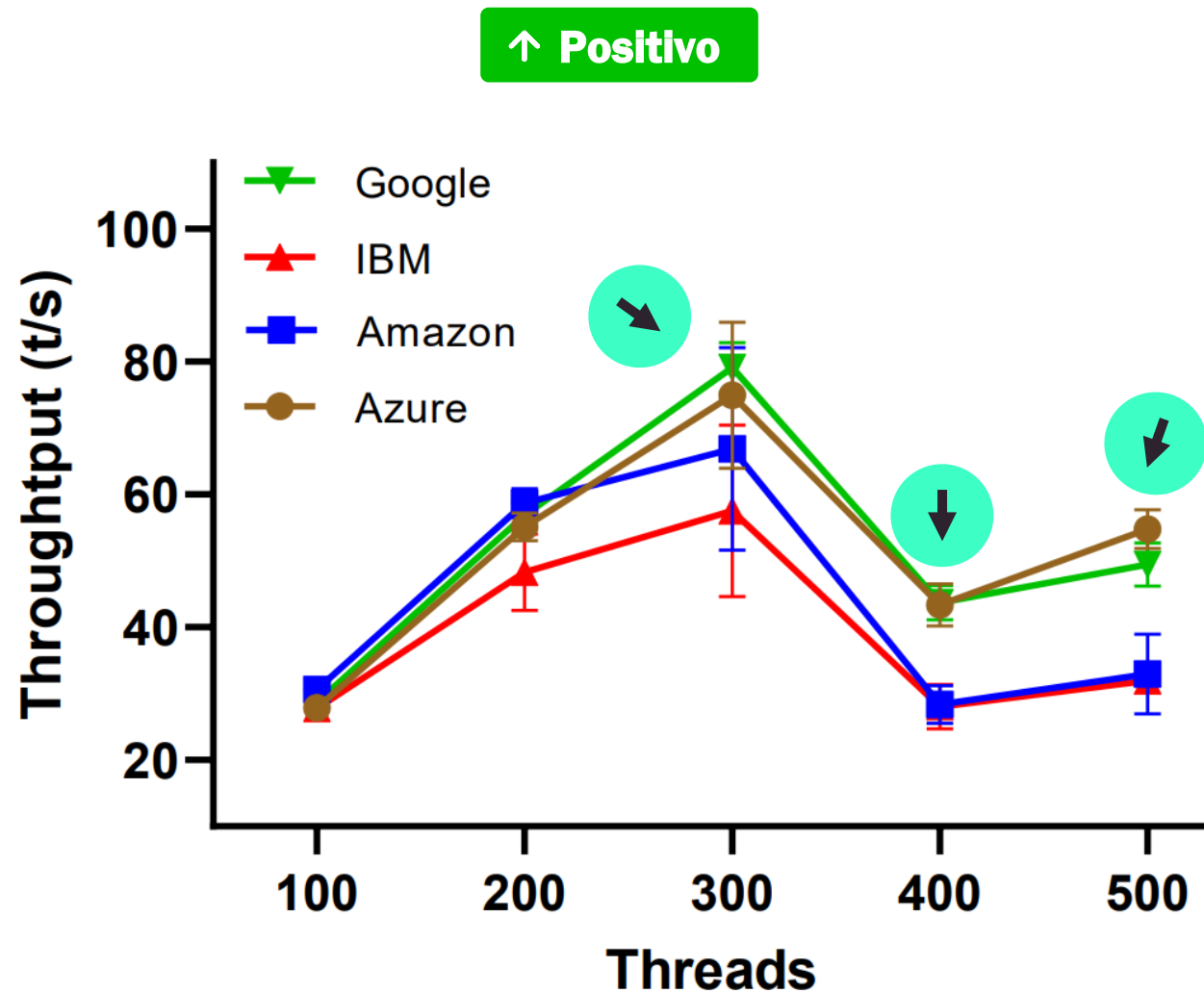




Vazão de Dados

Azure e **Google** se destacam **positivamente** ao, novamente, praticamente da mesma forma, lidarem de melhor forma com o estresse simulado.

A **IBM** sofre precocemente em comparação às outras, e a partir do cenário de **400 usuários** passa a ser acompanhado pela **Amazon**, que até então estava se comportando similarmente às outras.

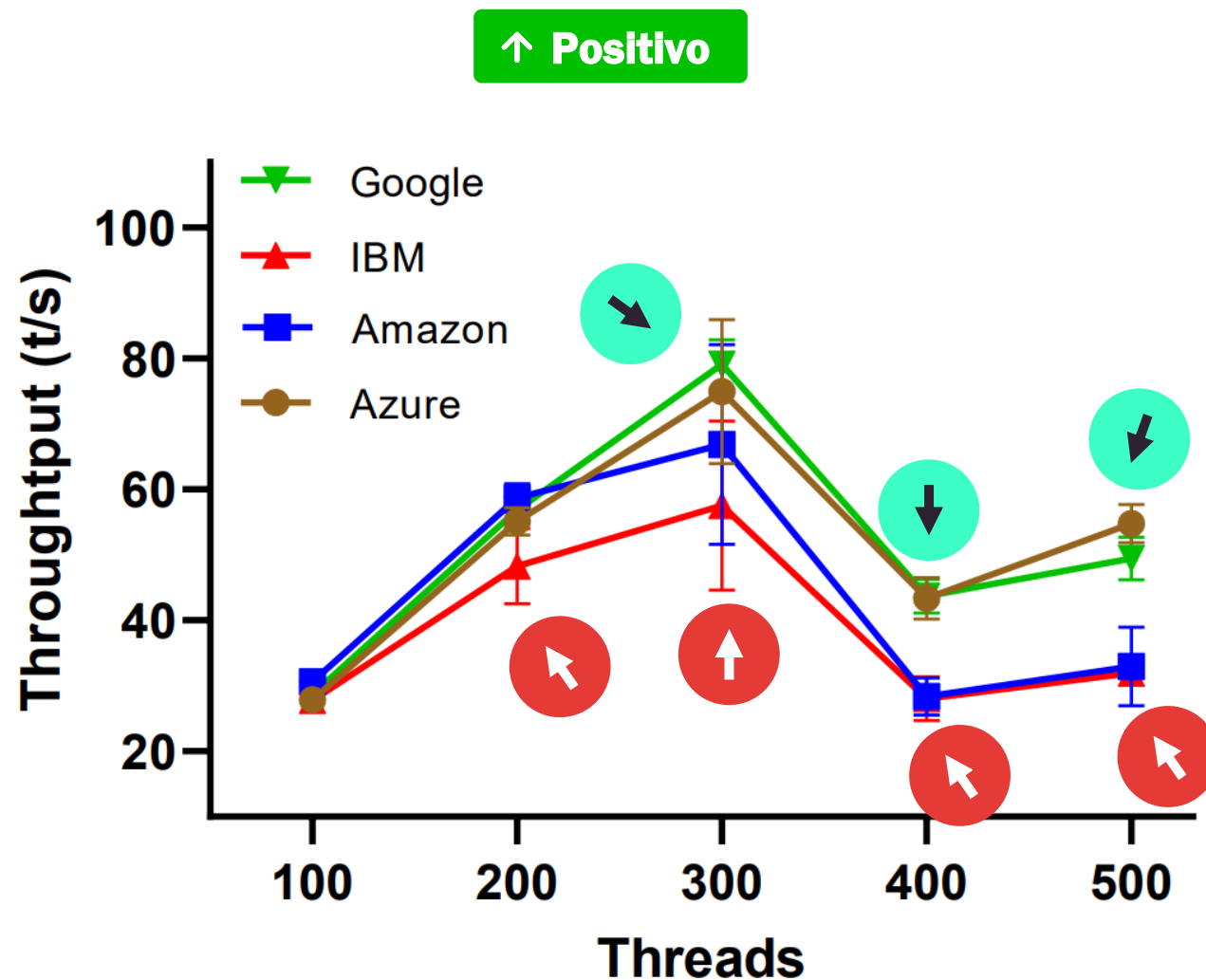




Vazão de Dados

Azure e **Google** se destacam **positivamente** ao, novamente, praticamente da mesma forma, lidarem de melhor forma com o estresse simulado.

A **IBM** sofre precocemente em comparação às outras, e a partir do cenário de **400 usuários** passa a ser acompanhado pela **Amazon**, que até então estava se comportando similarmente às outras.





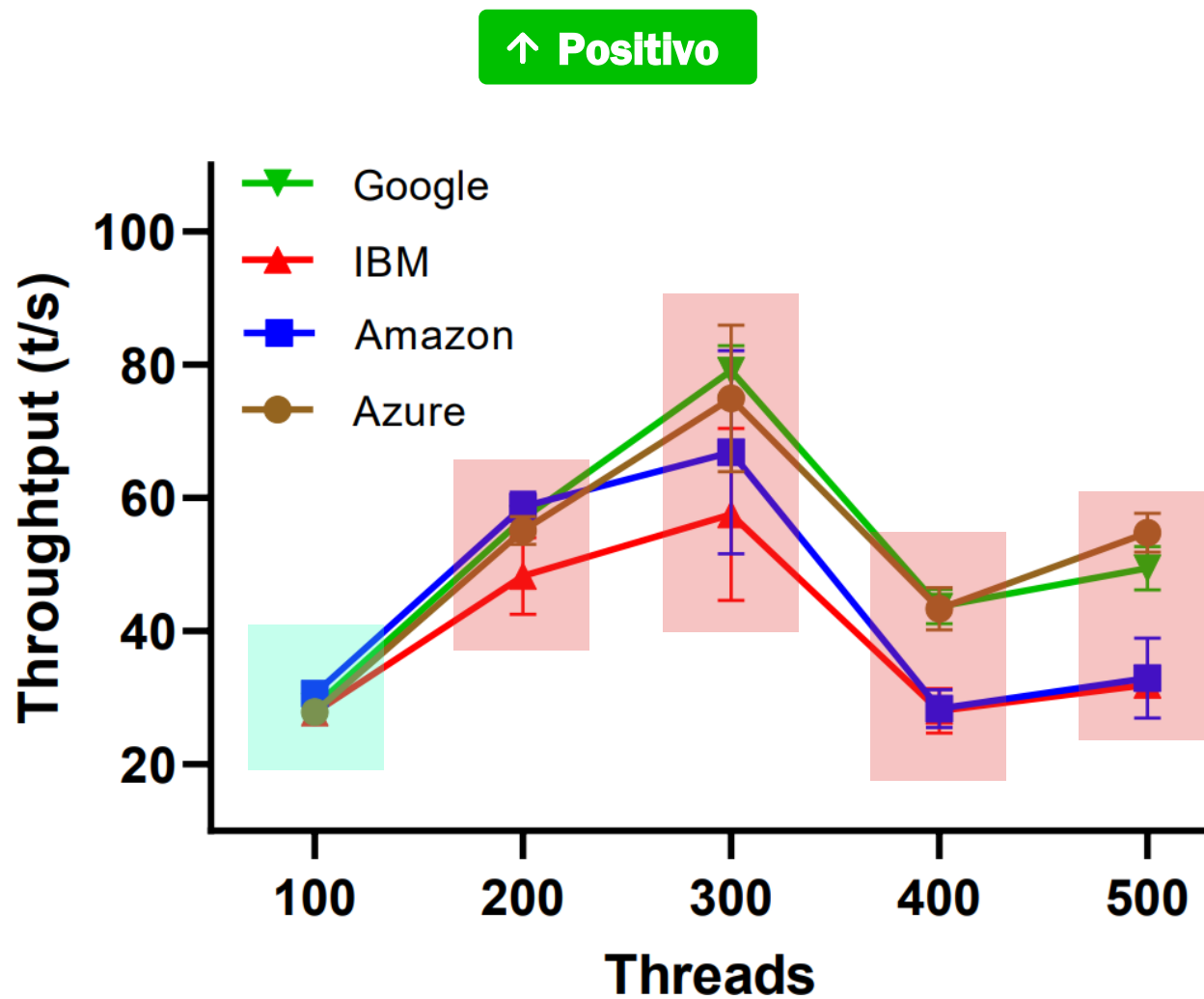
Vazão de Dados

Questão 1

Como cada provedor de infraestrutura em nuvem avaliado se comporta em relação a diferentes números de usuários realizando requisições simultâneas?

Conclusão

Apesar de se comportarem da mesma forma com 100 usuários, é possível identificar diferentes comportamentos ao decorrer dos testes.





Tempo de Resposta





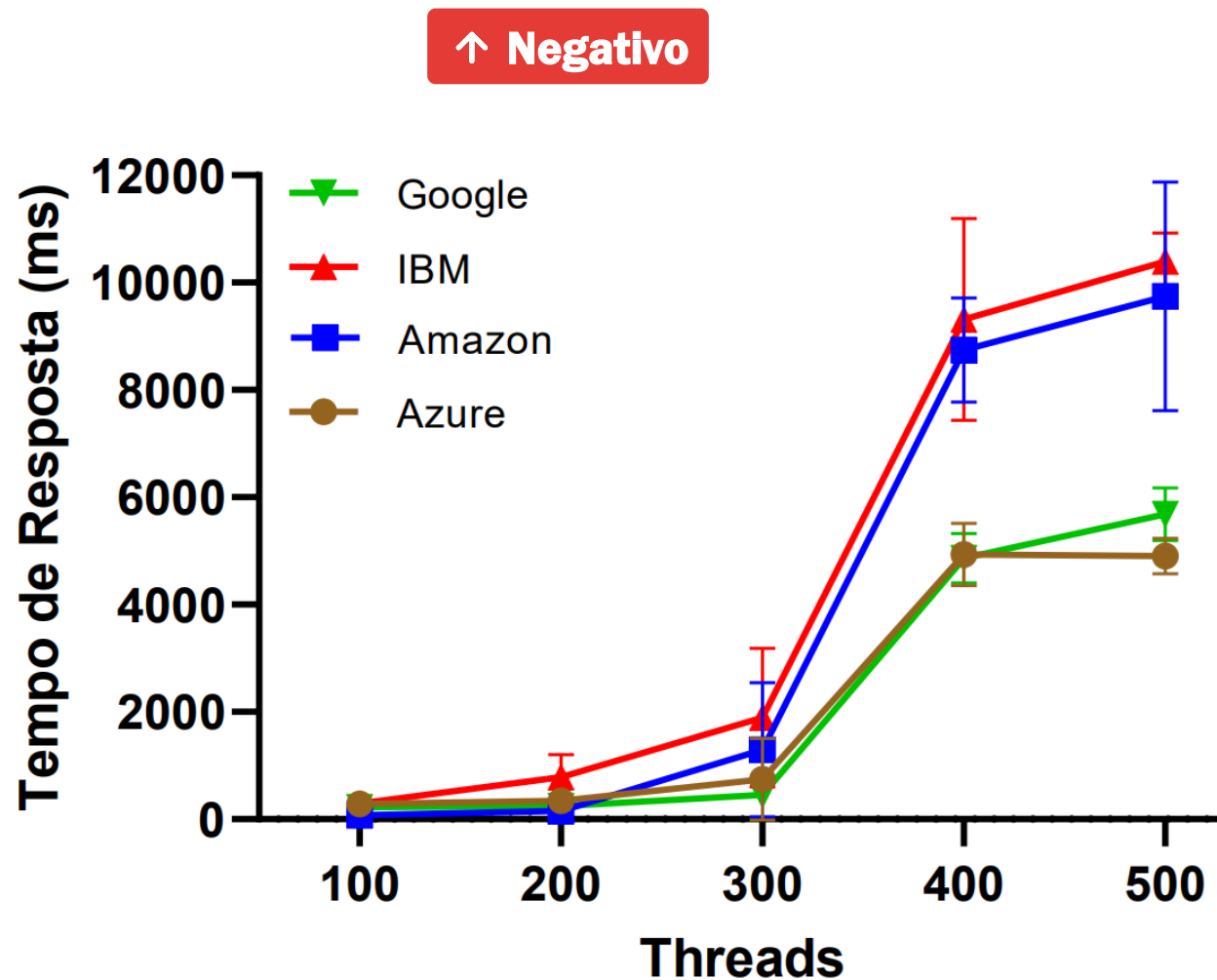
Tempo de Resposta

Até 300 usuários o tempo de resposta se mantém similar.

A partir de 300 usuários os níveis o tempo de resposta cresceu como o esperado

Novamente **Azure** e **Google** obtém um melhor resultado nos cenários mais extremos.

Assim como **Amazon** e **IBM** tem seu desempenho mais afetado ao decorrer dos testes.





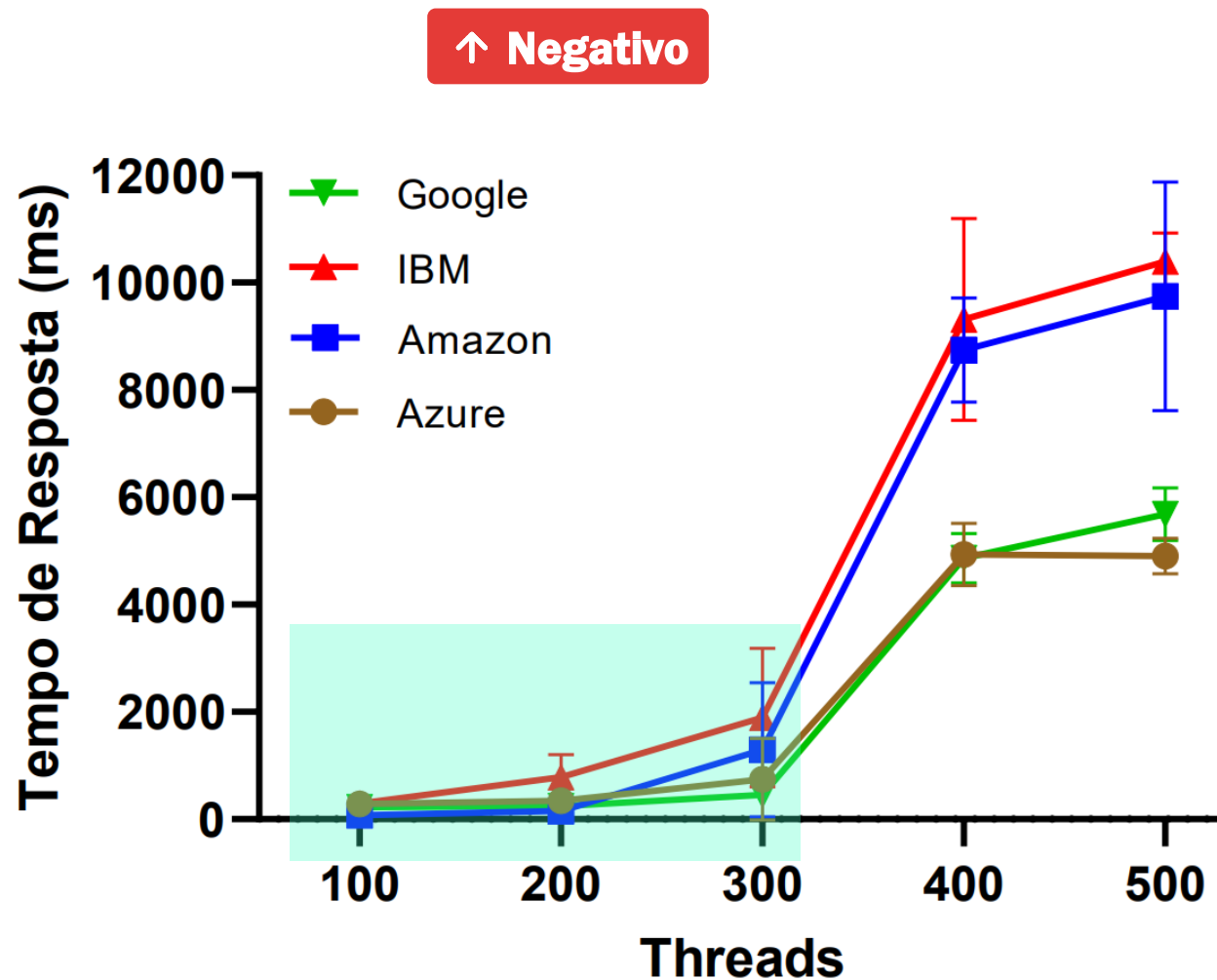
Tempo de Resposta

Até 300 usuários o tempo de resposta se mantém similar.

A partir de 300 usuários os níveis o tempo de resposta cresceu como o esperado

Novamente **Azure** e **Google** obtém um melhor resultado nos cenários mais extremos.

Assim como **Amazon** e **IBM** tem seu desempenho mais afetado ao decorrer dos testes.





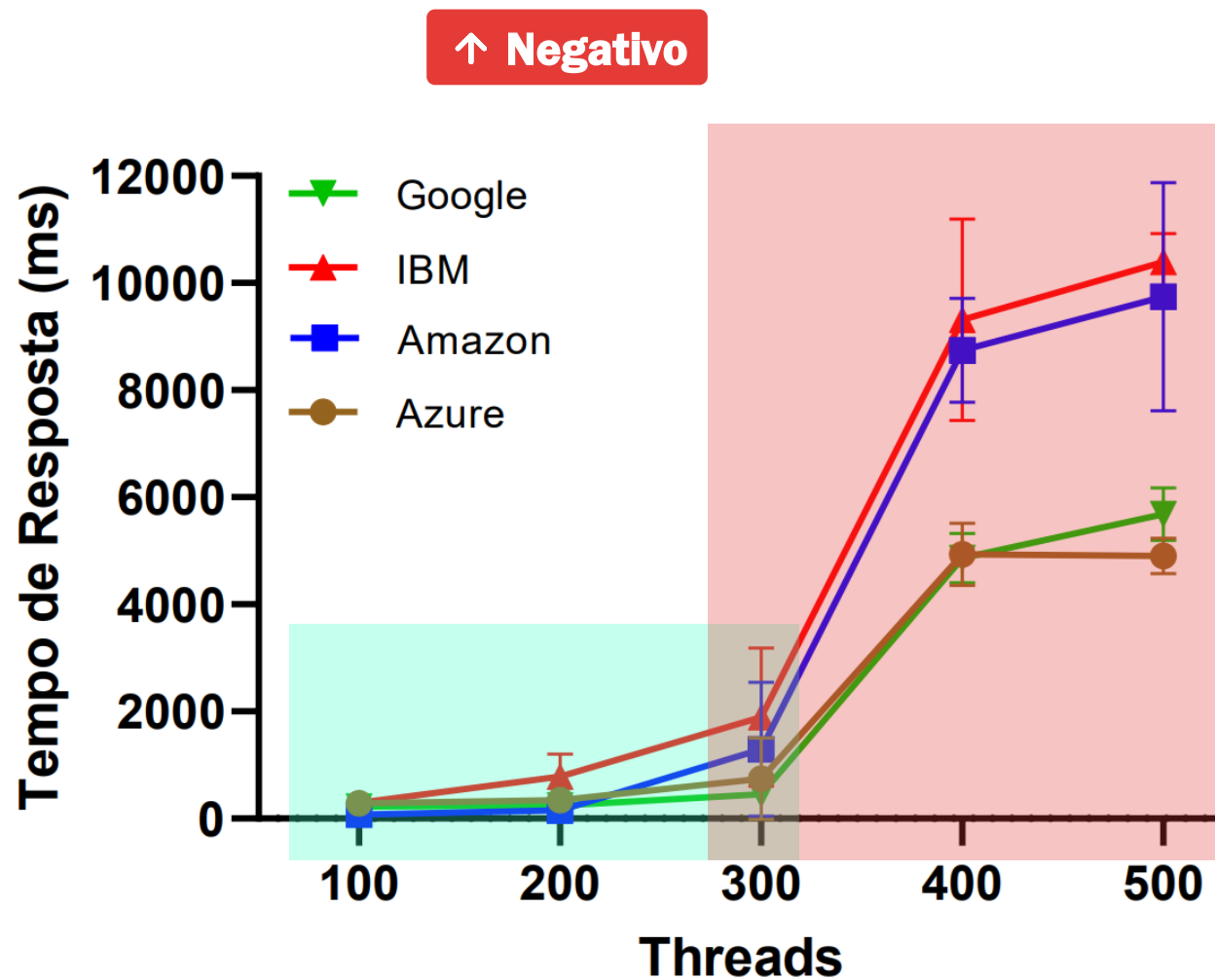
Tempo de Resposta

Até 300 usuários o tempo de resposta se mantém similar.

A partir de 300 usuários os níveis o tempo de resposta cresceu como o esperado

Novamente **Azure** e **Google** obtém um melhor resultado nos cenários mais extremos.

Assim como **Amazon** e **IBM** tem seu desempenho mais afetado ao decorrer dos testes.





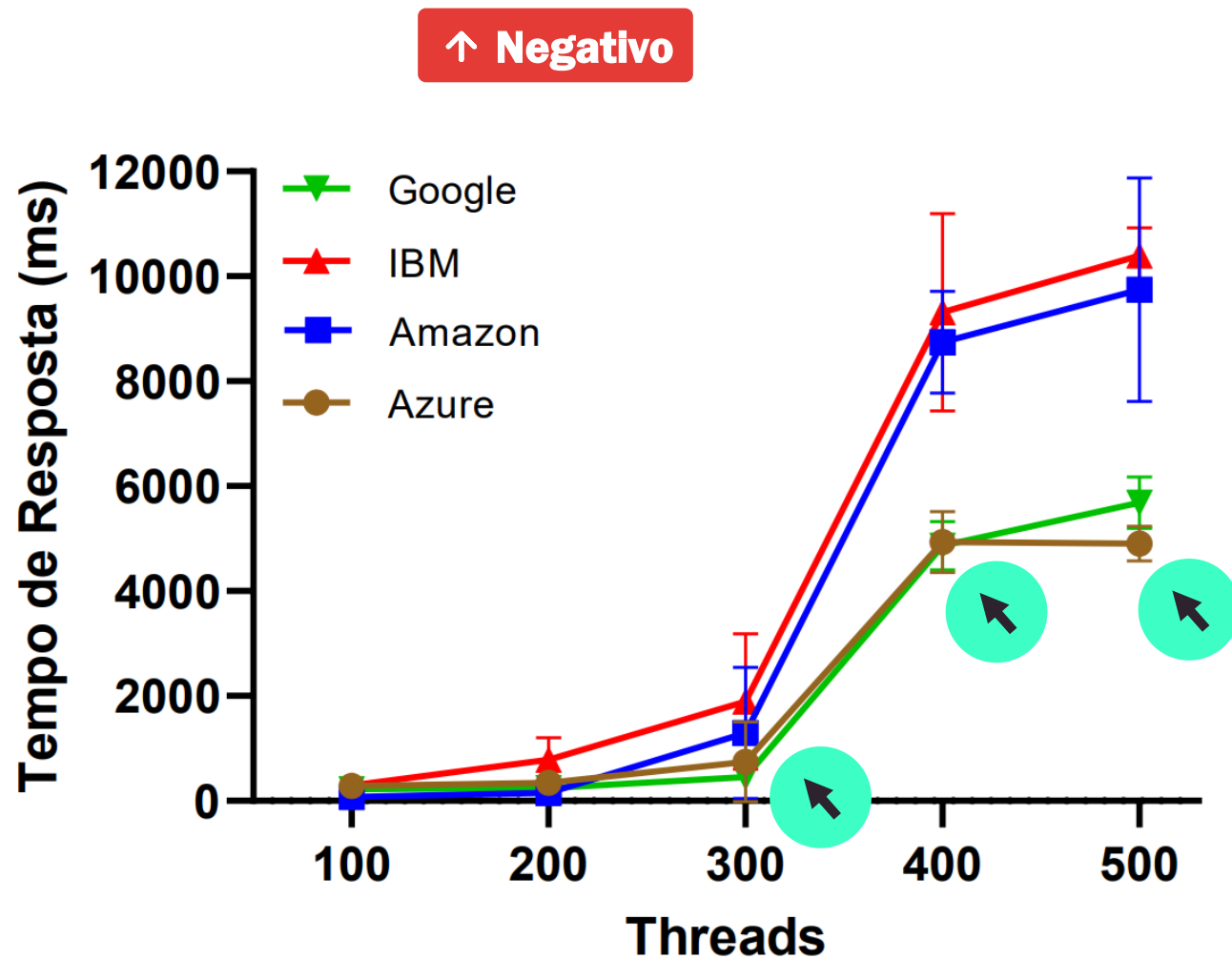
Tempo de Resposta

Até 300 usuários o tempo de resposta se mantém similar.

A partir de 300 usuários os níveis o tempo de resposta cresceu como o esperado

Novamente **Azure** e **Google** obtém um melhor resultado nos cenários mais extremos.

Assim como **Amazon** e **IBM** tem seu desempenho mais afetado ao decorrer dos testes.





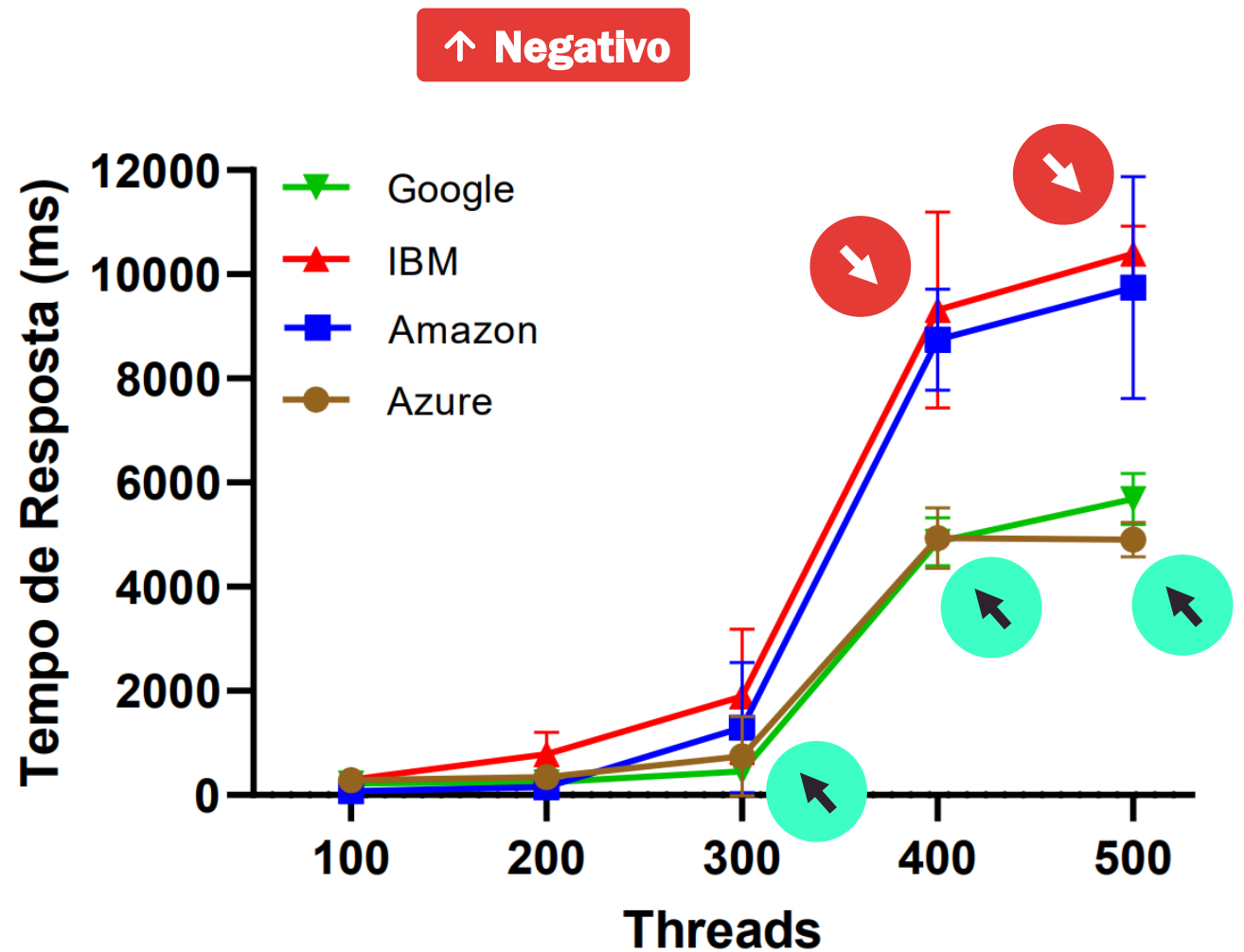
Tempo de Resposta

Até 300 usuários o tempo de resposta se mantém similar.

A partir de 300 usuários os níveis o tempo de resposta cresceu como o esperado

Novamente **Azure** e **Google** obtém um melhor resultado nos cenários mais extremos.

Assim como **Amazon** e **IBM** tem seu desempenho mais afetado ao decorrer dos testes.





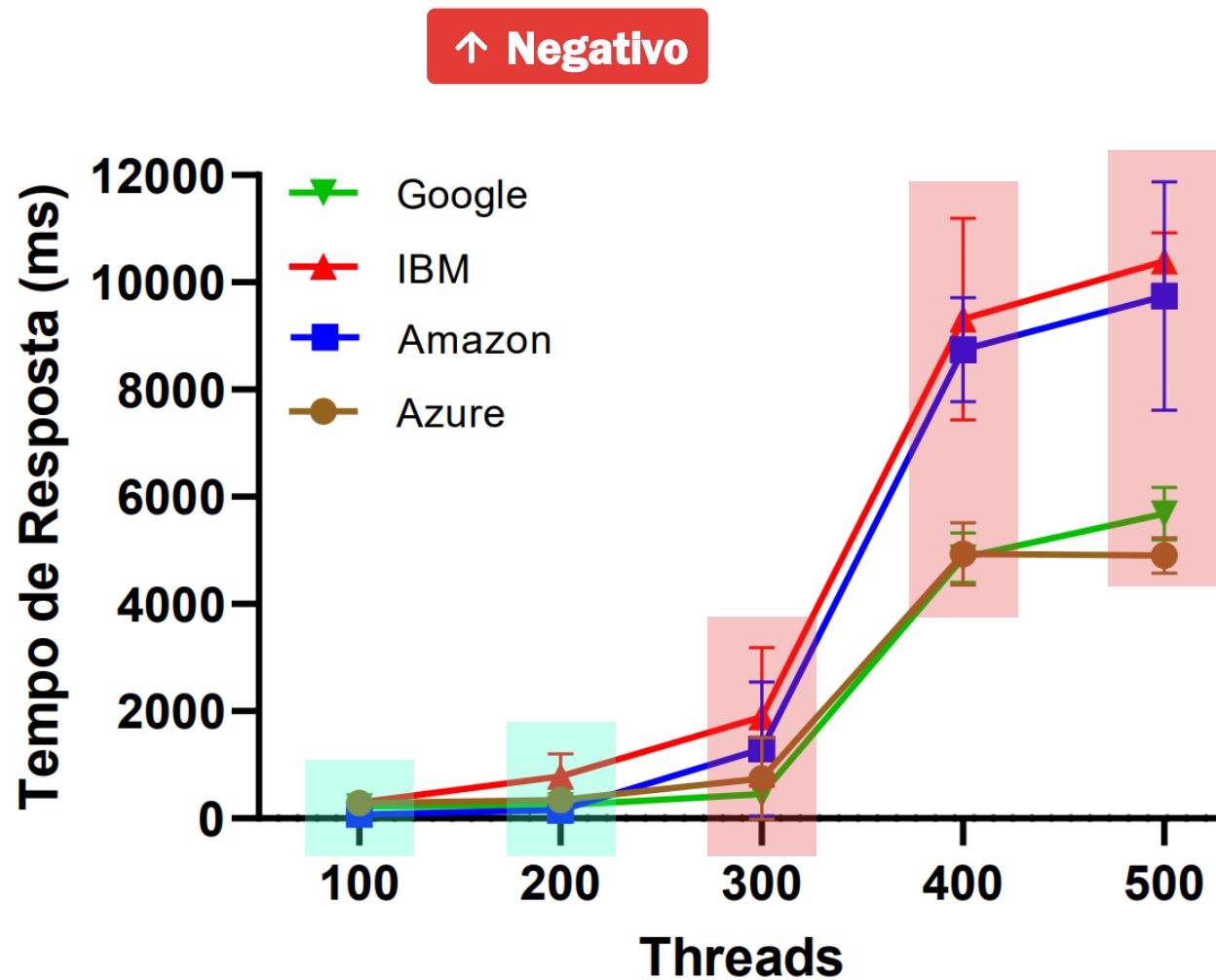
Tempo de Resposta

Questão 1

Como cada provedor de infraestrutura em nuvem avaliado se comporta em relação a diferentes números de usuários realizando requisições simultâneas?

Conclusão

Inicialmente todas as provedores se comportam da mesma forma, porém ao decorrer dos testes os se dividem em grupos de diferentes comportamentos.





Apdex



Resultados



Apdex

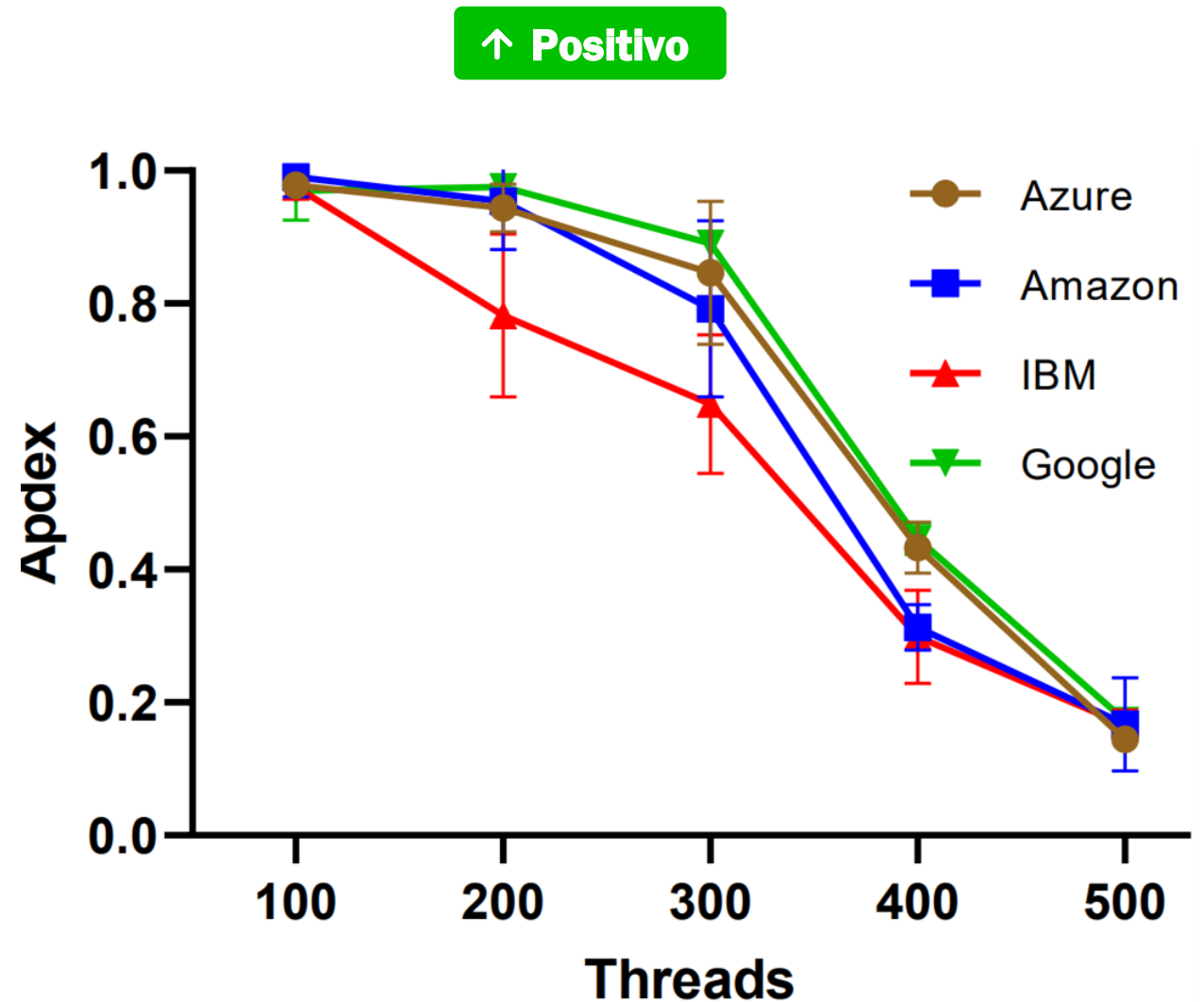
Para T=2

Até 300 usuários a maioria das plataformas manteve uma satisfação acima de 80%.

A partir de 300 usuários os níveis de satisfação se tornaram cada vez mais negativos.

Como o **Apdex** mede a qualidade geral do serviço, a **Azure** e **Google** obtém um melhor resultado geral como reflexo do resultado obtido nas outras métricas.

A **IBM** apresenta um menos Apdex, inclusive nos testes iniciais, também refletindo sua dificuldade em lidar com as cargas com quantidades de usuários relativamente baixas.



Resultados



Apdex

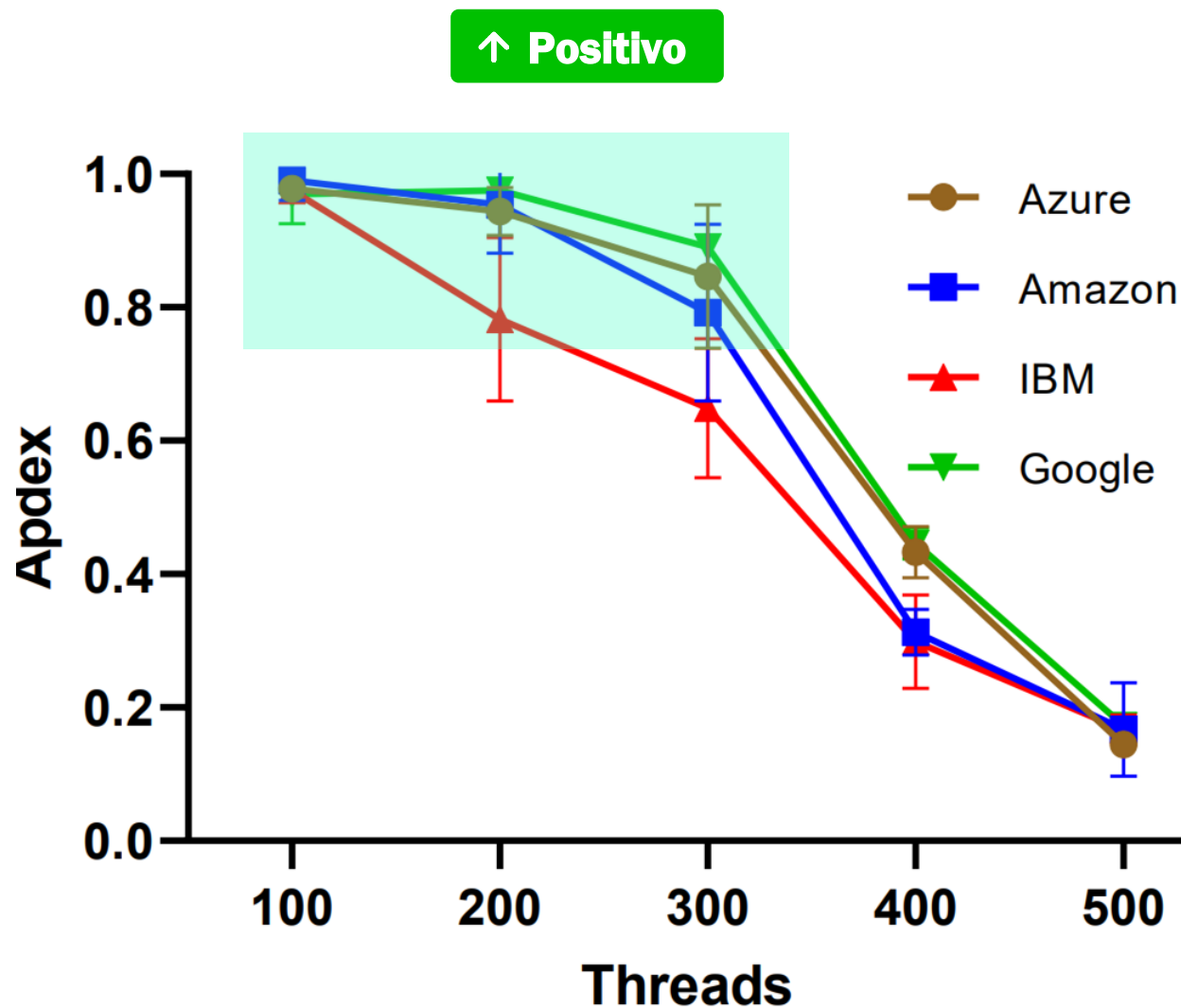
Para T=2

Até 300 usuários a maioria das plataformas manteve uma satisfação acima de 80%.

A partir de 300 usuários os níveis de satisfação se tornaram cada vez mais negativos.

Como o **Apdex** mede a qualidade geral do serviço, a **Azure** e **Google** obtém um melhor resultado geral como reflexo do resultado obtido nas outras métricas.

A **IBM** apresenta um menos Apdex, inclusive nos testes iniciais, também refletindo sua dificuldade em lidar com as cargas com quantidades de usuários relativamente baixas.



Resultados



Apdex

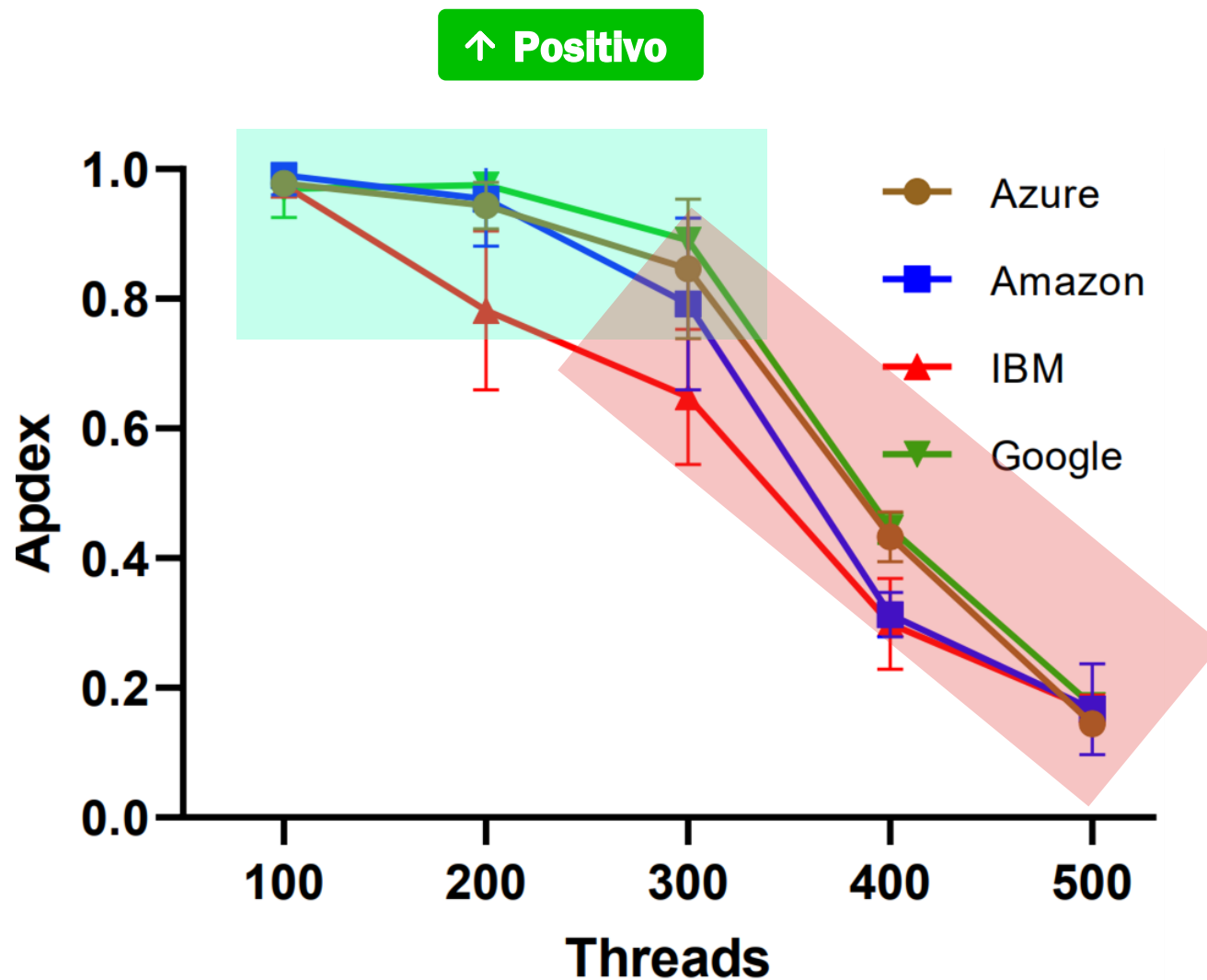
Para T=2

Até 300 usuários a maioria das plataformas manteve uma satisfação acima de 80%.

A partir de 300 usuários os níveis de satisfação se tornaram cada vez mais negativos.

Como o **Apdex** mede a qualidade geral do serviço, a **Azure** e **Google** obtém um melhor resultado geral como reflexo do resultado obtido nas outras métricas.

A **IBM** apresenta um menos Apdex, inclusive nos testes iniciais, também refletindo sua dificuldade em lidar com as cargas com quantidades de usuários relativamente baixas.



Resultados



Apdex

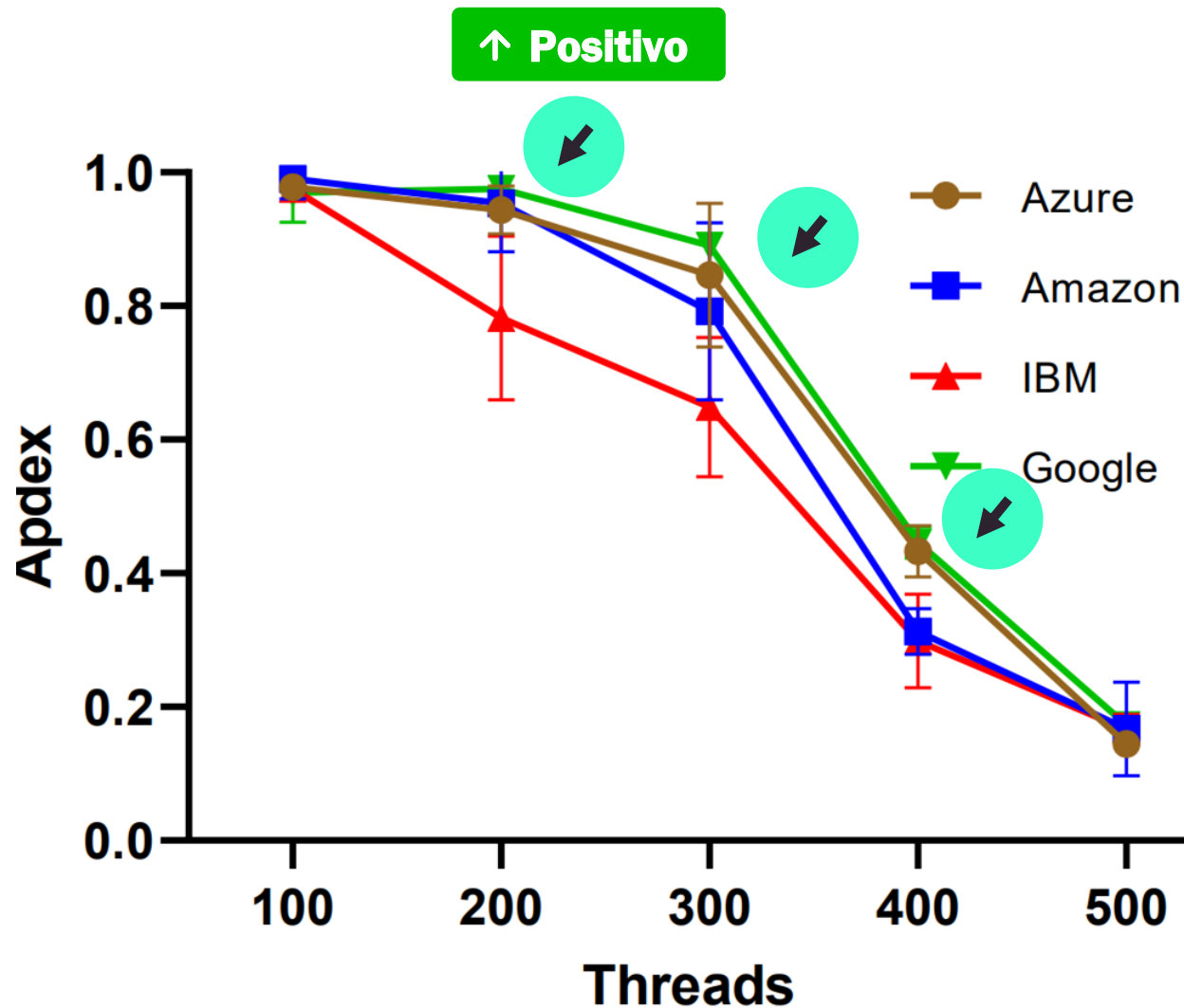
Para T=2

Até 300 usuários a maioria das plataformas manteve uma satisfação acima de 80%.

A partir de 300 usuários os níveis de satisfação se tornaram cada vez mais negativos.

Como o **Apdex** mede a qualidade geral do serviço, a **Azure** e **Google** obtém um melhor resultado geral como reflexo do resultado obtido nas outras métricas.

A **IBM** apresenta um menos Apdex, inclusive nos testes iniciais, também refletindo sua dificuldade em lidar com as cargas com quantidades de usuários relativamente baixas.



Resultados



Apdex

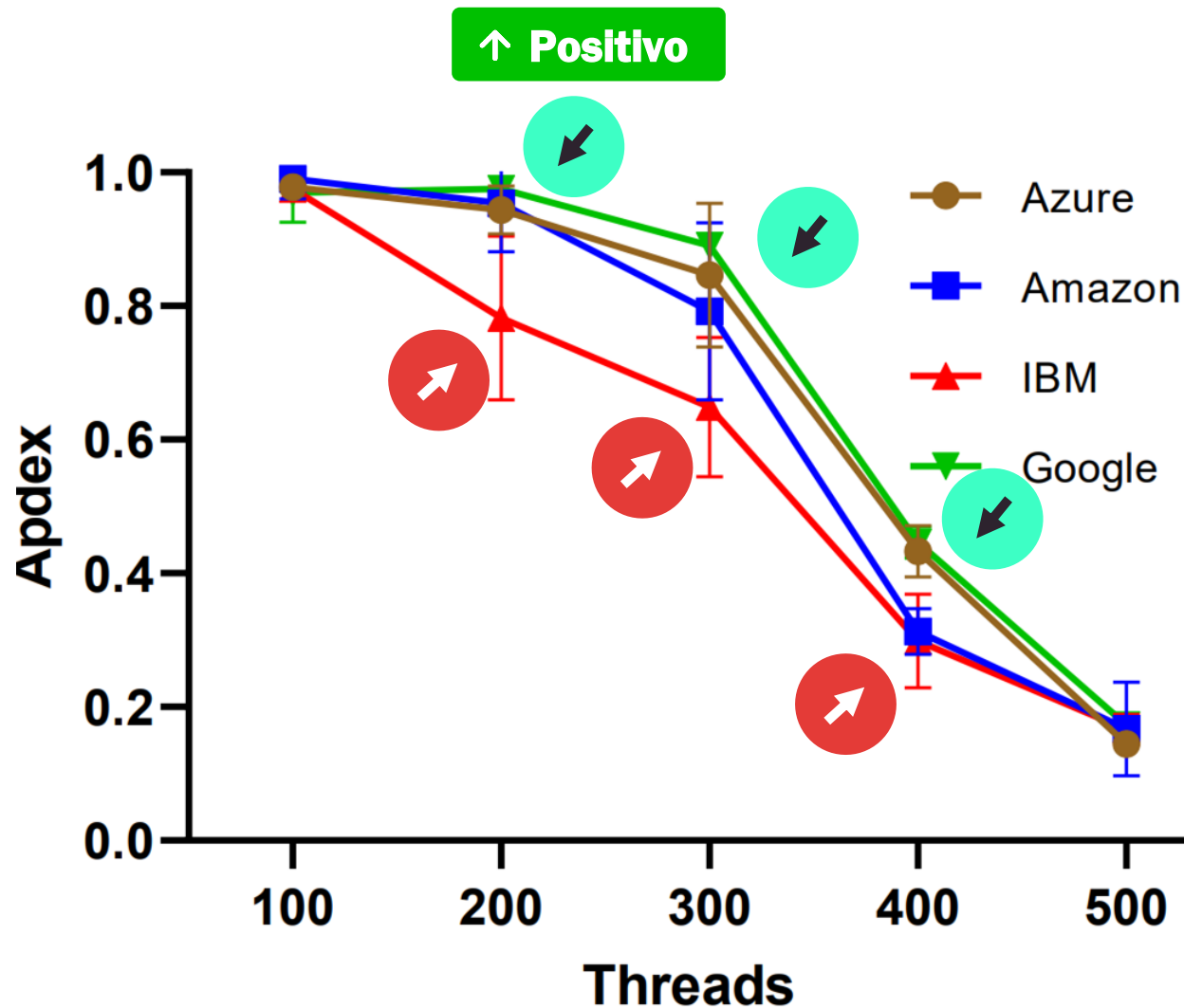
Para T=2

Até 300 usuários a maioria das plataformas manteve uma satisfação acima de 80%.

A partir de 300 usuários os níveis de satisfação se tornaram cada vez mais negativos.

Como o **Apdex** mede a qualidade geral do serviço, a **Azure** e **Google** obtém um melhor resultado geral como reflexo do resultado obtido nas outras métricas.

A **IBM** apresenta um menos Apdex, inclusive nos testes iniciais, também refletindo sua dificuldade em lidar com as cargas com quantidades de usuários relativamente baixas.



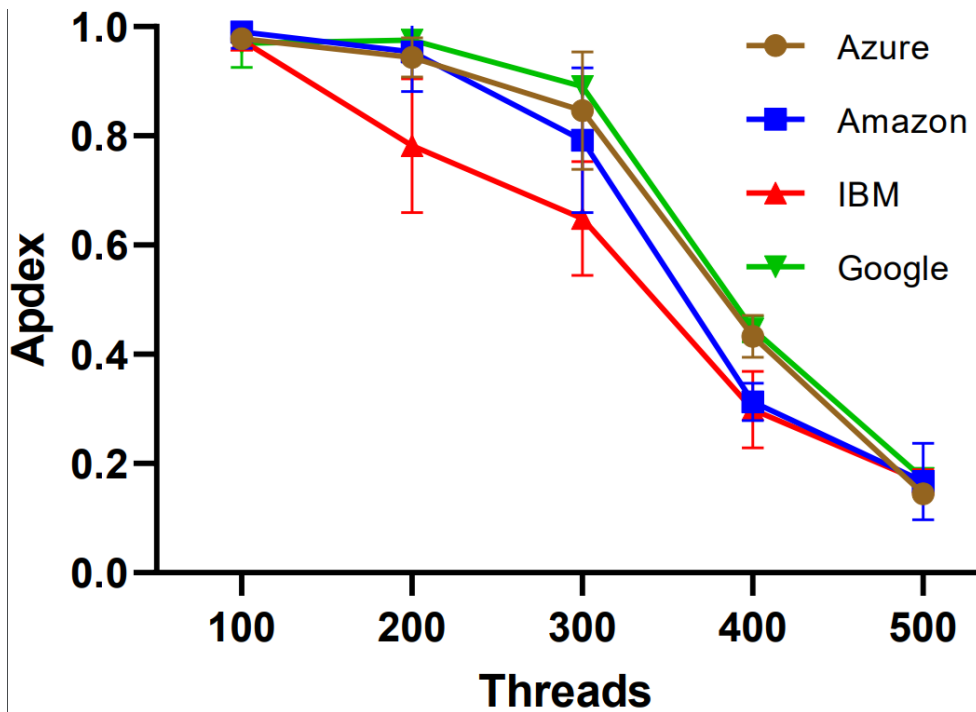
Resultados



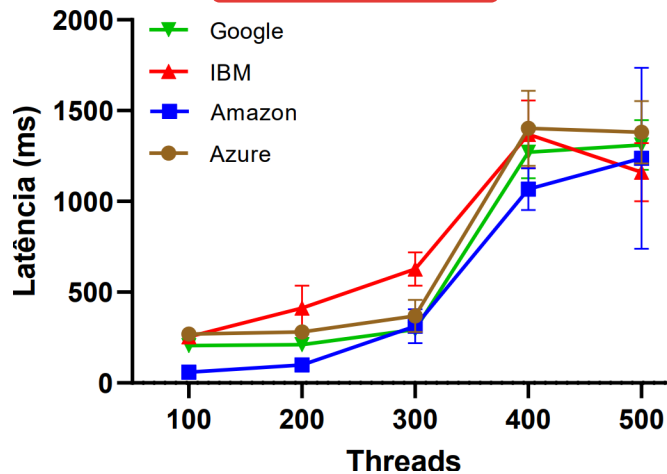
Apdex

Para T=2

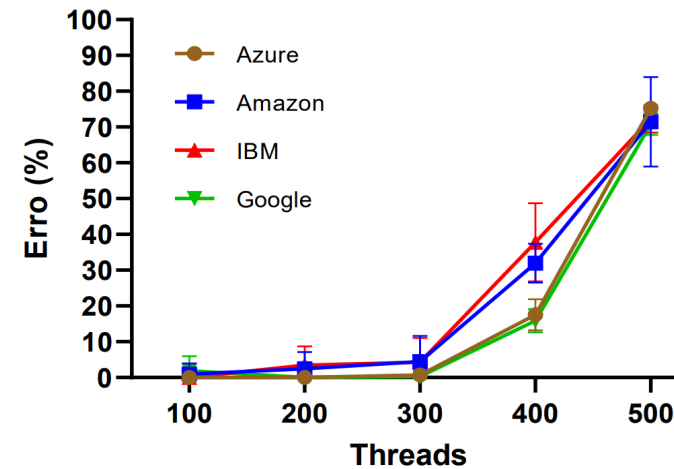
↑ Positivo



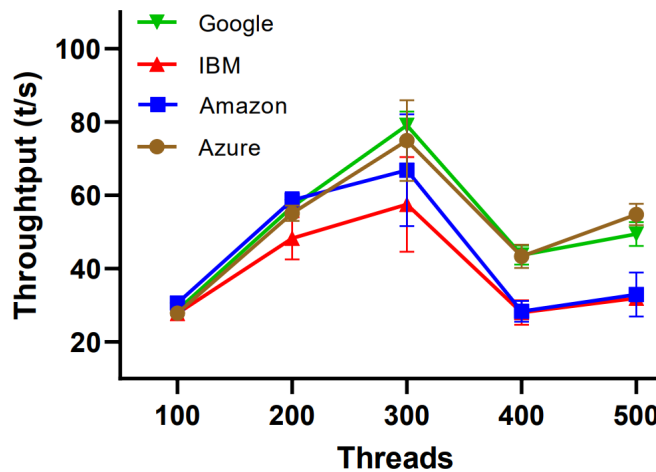
↑ Negativo



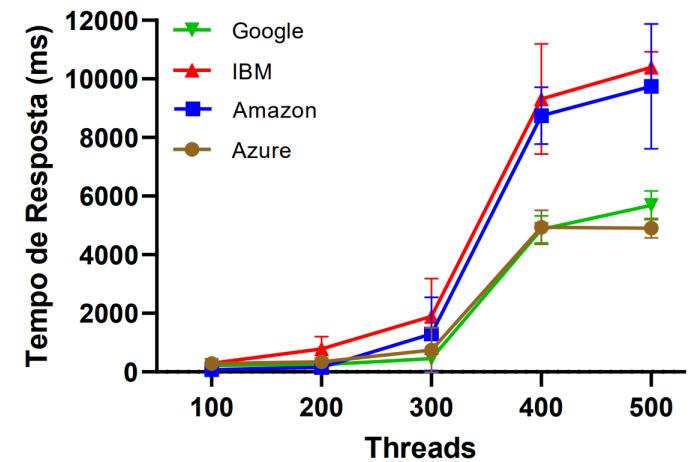
↑ Negativo



↑ Positivo



↑ Negativo



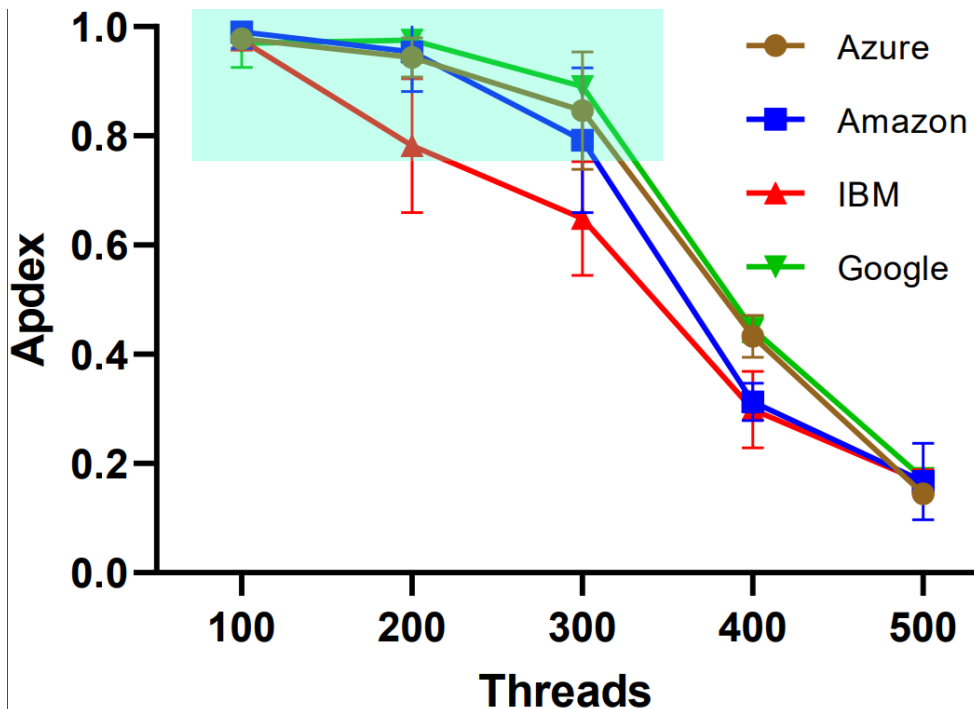
Resultados



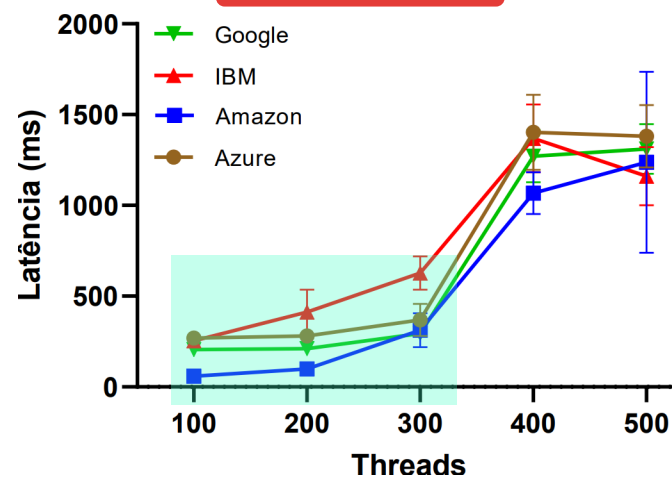
Apdex

Para T=2

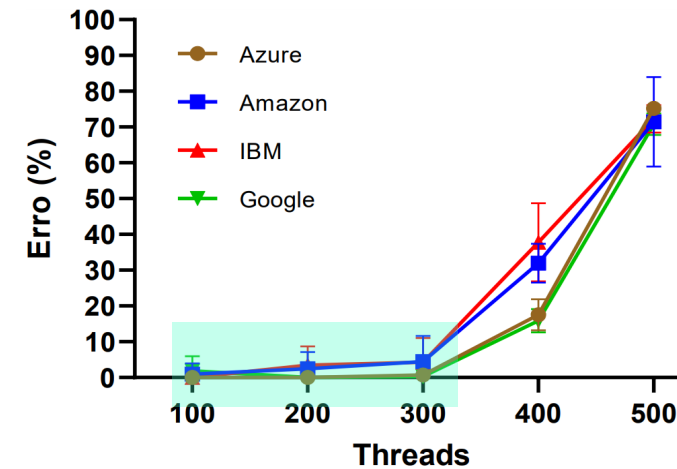
↑ Positivo



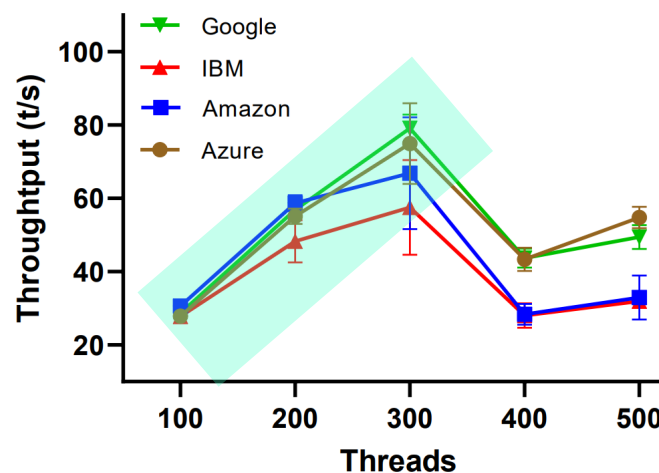
↑ Negativo



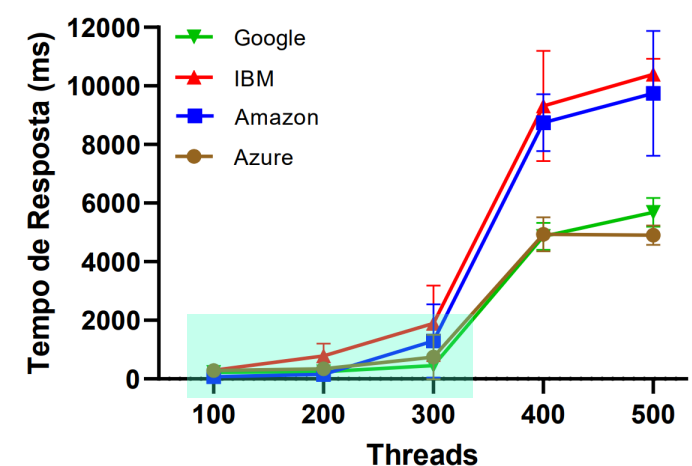
↑ Negativo



↑ Positivo



↑ Negativo



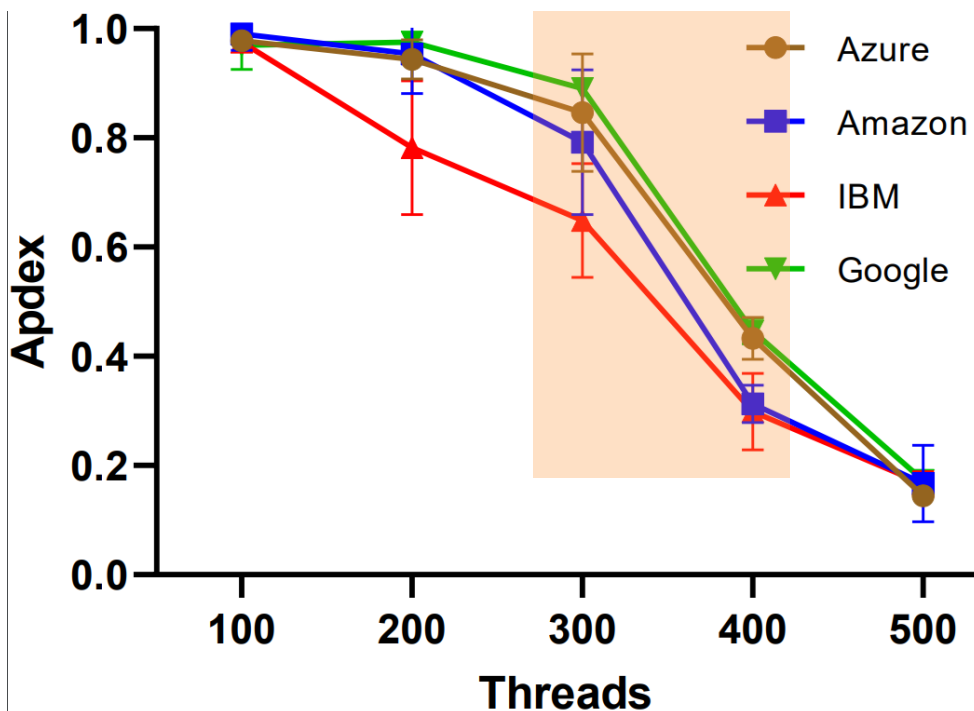
Resultados



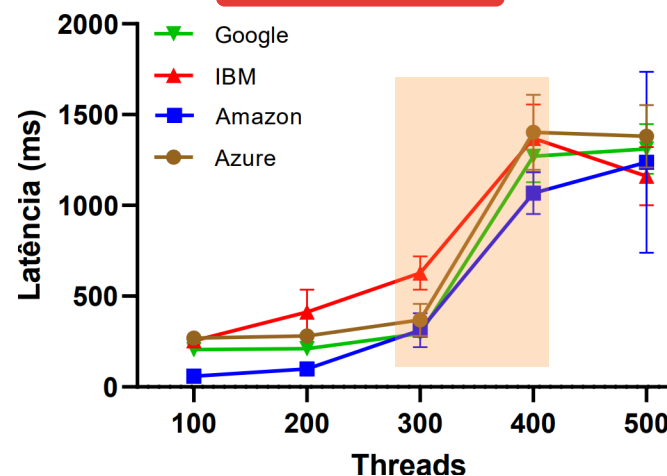
Apdex

Para T=2

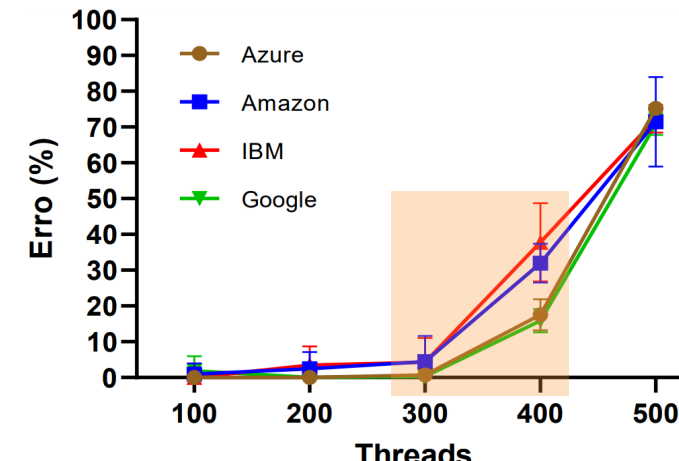
↑ Positivo



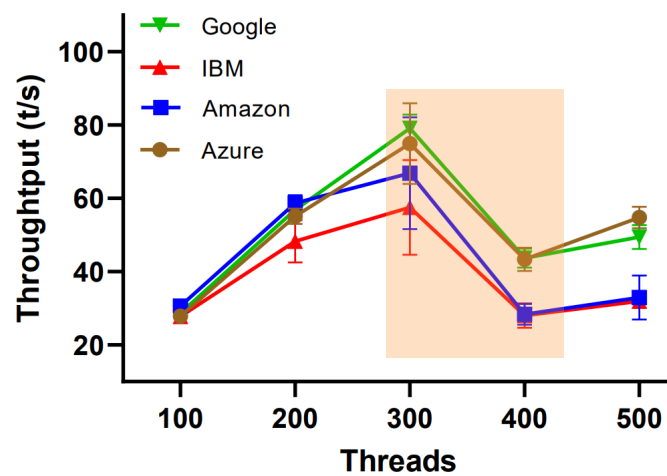
↑ Negativo



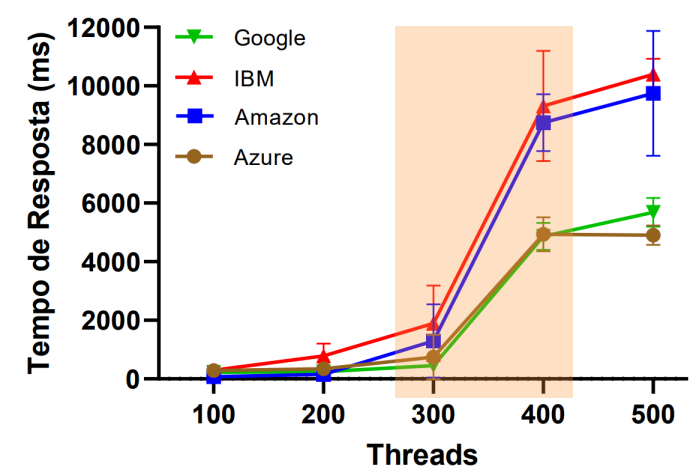
↑ Negativo



↑ Positivo



↑ Negativo



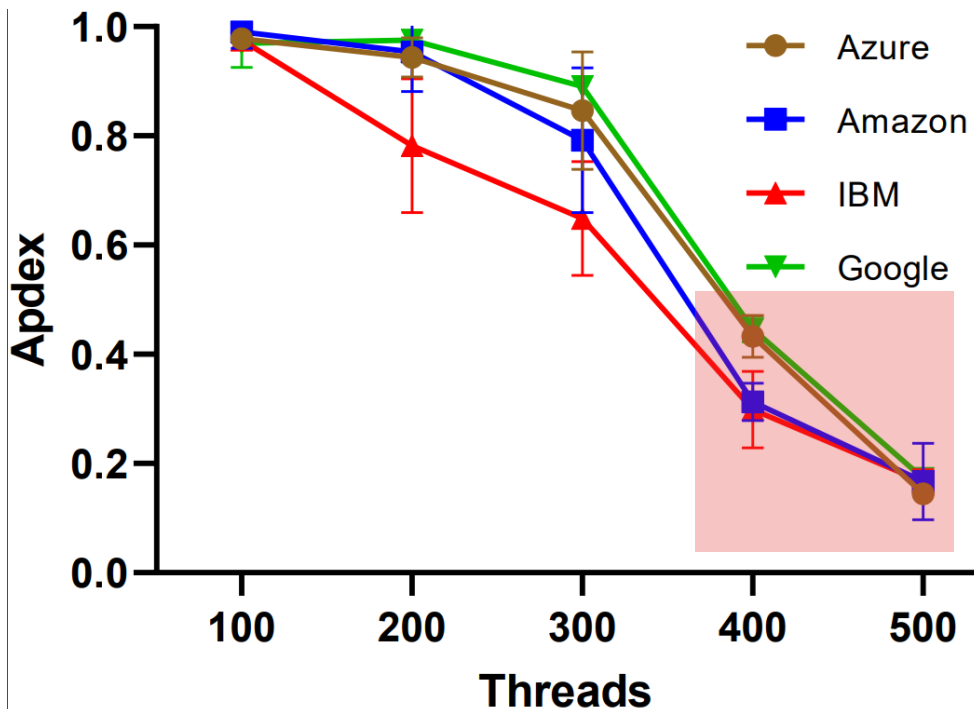
Resultados



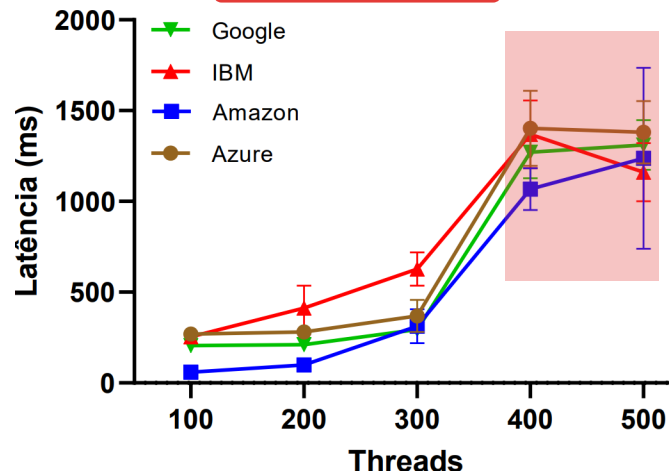
Apdex

Para T=2

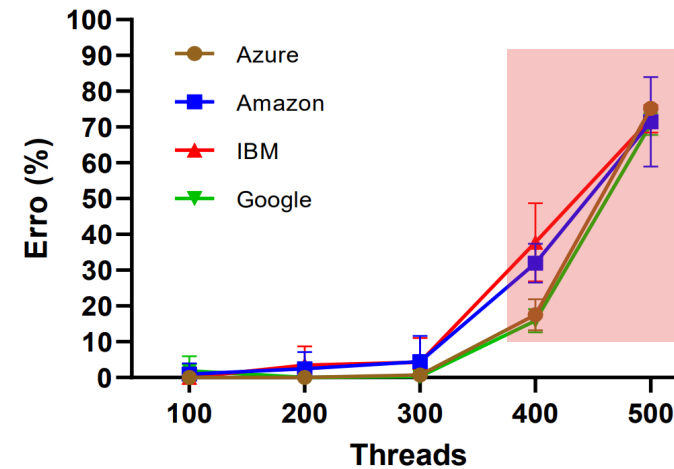
↑ Positivo



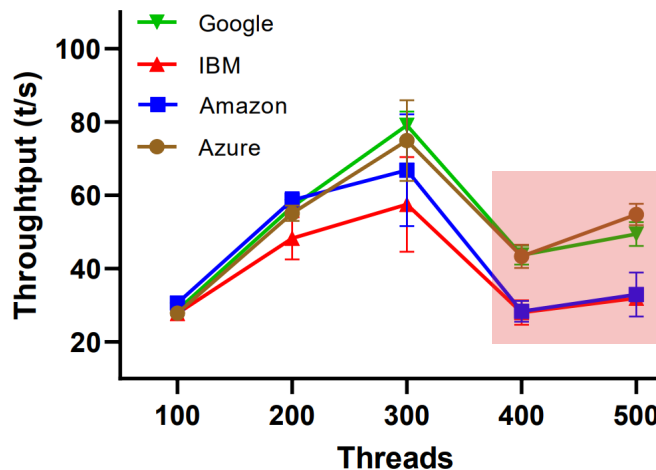
↑ Negativo



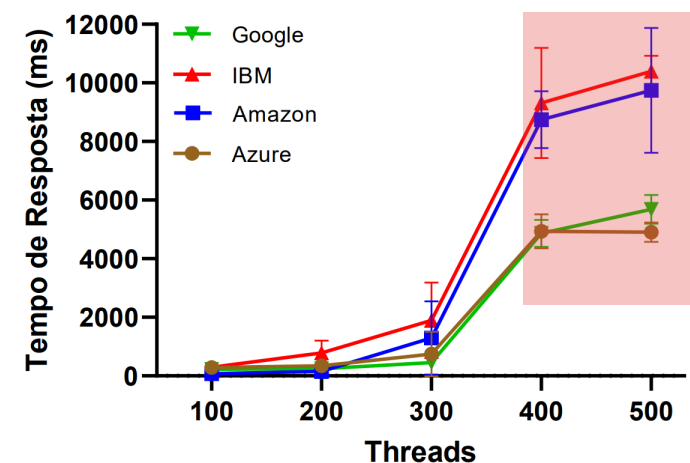
↑ Negativo



↑ Positivo



↑ Negativo





Conclusão





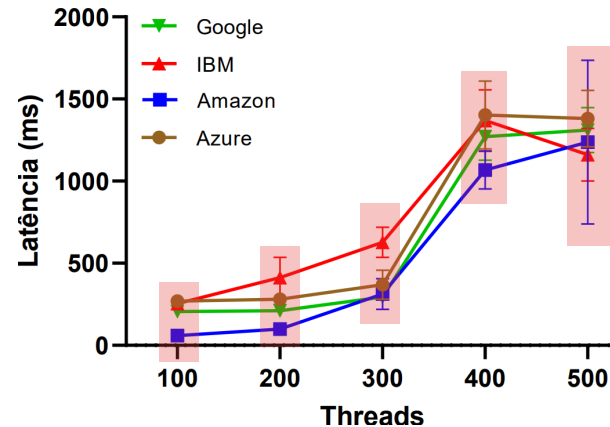
Questão 1

Como cada provedor de infraestrutura em nuvem avaliado se comporta em relação a diferentes números de usuários realizando requisições simultâneas?

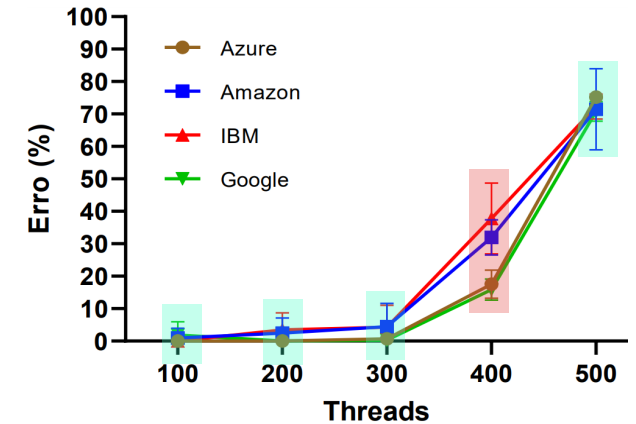
Conclusão

Das **20 representações** das combinações entre métricas e níveis de usuário pode-se observar que apenas **35%**, ou seja, **7** delas, foram similares entre todas os provedores, sendo **4** delas relativos à métrica de **taxa de erro**, logo pode-se dizer que os provedores se comportam de maneiras distintas para a configuração utilizada.

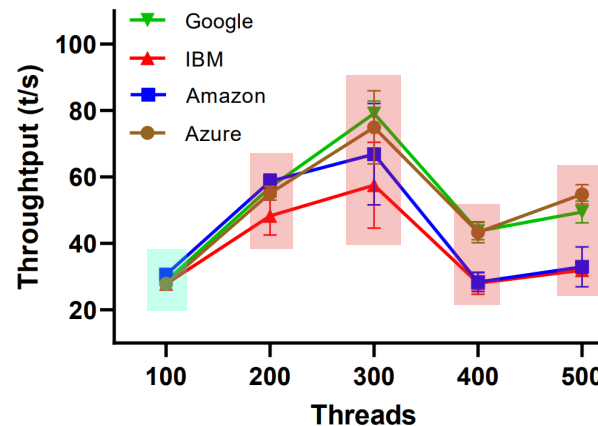
↑ Negativo



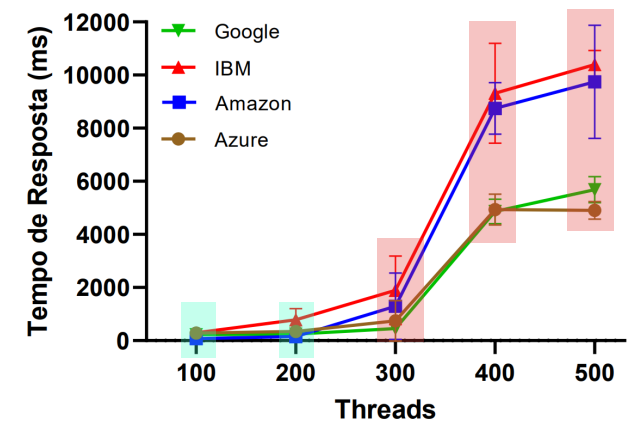
↑ Negativo



↑ Positivo



↑ Negativo



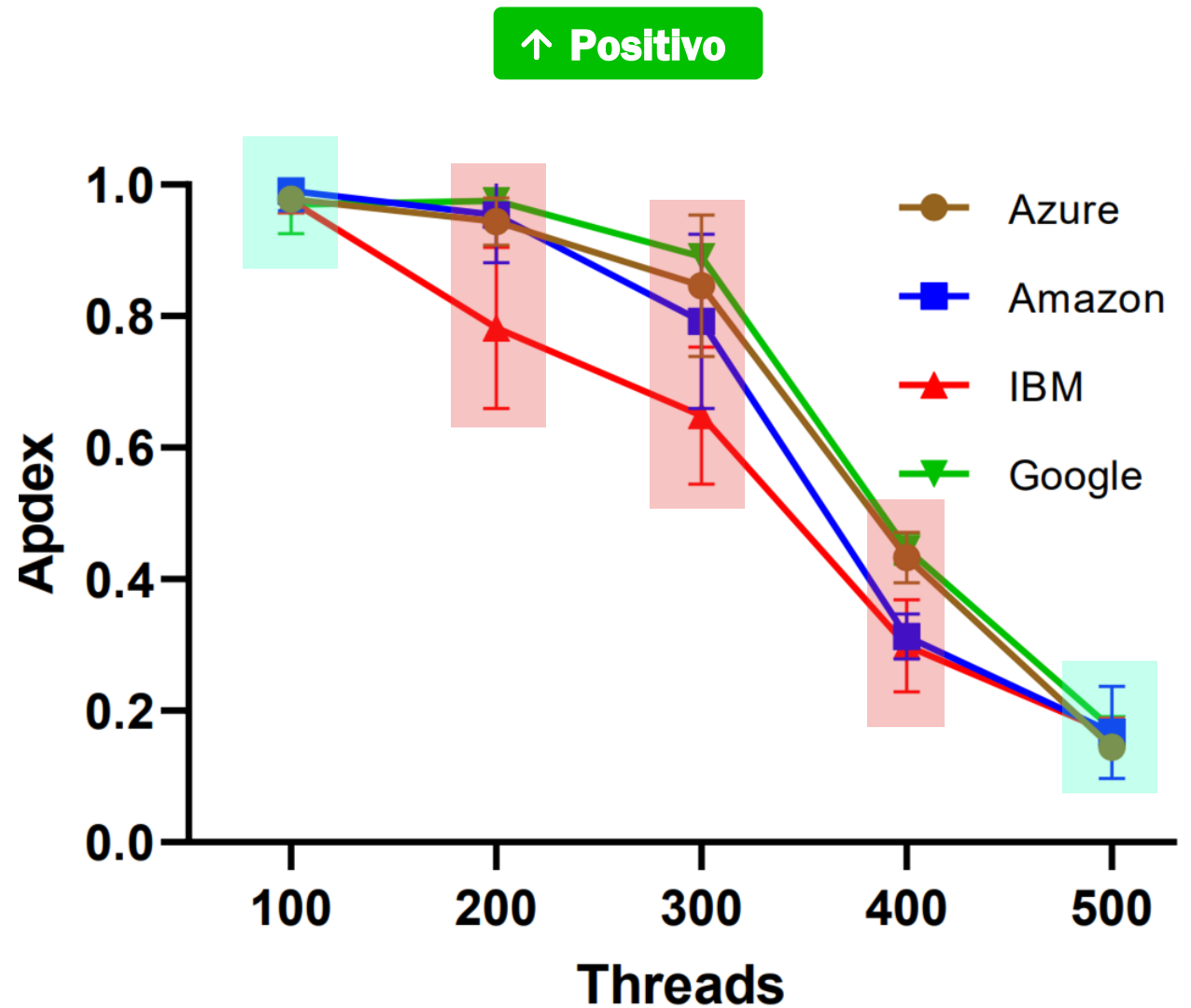


Questão 2

Como esses provedores entregam **satisfação ao usuário** em relação ao desempenho das requisições?

Conclusão

Nas 5 variações de quantidades de usuário pode-se observar que somente no teste de **100** e **500** todos os provedores se comportaram da mesma forma, logo, na maioria dos casos, apresentaram **diferentes níveis de satisfação aos usuários**.





Perspectivas futuras

Esse estudo pode ser **replicado sobre novas aplicações** que trabalhem com diferentes escopos e tecnologias, e demandem maiores níveis de qualidade em métricas específicas como **latência** ou **tempo de resposta**.

A replicação desse estudo também pode **avaliar novos provedores** que se destaquem no mercado ou que sejam interessantes para fornecer um ambiente em nuvem para determinada aplicação.

Identificar quais configurações dos servidores causam maior impacto em seu desempenho.

Avaliar o impacto da **distância geográfica** entre o servidor e a máquina na qual a rotina de testes é executada.

Obrigado! Thank you! ¡Gracias!

DÚVIDAS?



Denis B. Oliveira

Instituto Federal do Sul de
Minas Gerais
IFSULDEMINAS

denisb.oliveira@live.com



Ricardo R. de Oliveira

Instituto Federal do Sul de
Minas Gerais
IFSULDEMINAS

ricardo.ramos@ifsuldeminas.edu.br



Ricardo F. Vilela

Universidade Federal de
Santa Catarina
UFSC

ricardo.vilela@ufsc.br



Victor H. S. C. Pinto

Universidade Federal do Pará
UFPA

victor.santiago@ufpa.br



Roberto N. Ungarelli

Zello

roberto.ungarelli@gmail.com