



MATHEUS GUSTAVO COPPI DA SILVA

**UMA ARQUITETURA BASEADA EM MICROSERVIÇOS CENTRADA NA
OBSERVABILIDADE**

MATHEUS GUSTAVO COPPI DA SILVA

**UMA ARQUITETURA BASEADA EM MICROSERVIÇOS CENTRADA NA
OBSERVABILIDADE**

Bacharel apresentado ao curso de Sistemas de informação do Centro Universitário Avantis de Balneário Camboriú para a obtenção do título de Bacharel em Sistemas de informação.

Orientador: Prof. Luiz Fernando Arruda, Msc

Balneário Camboriú
2025

RESUMO

A crescente complexidade dos sistemas e a demanda por escalabilidade impulsionaram a adoção da arquitetura de microsserviços em detrimento do modelo monolítico. Essa abordagem favorece a modularização, a escalabilidade granular e a autonomia das equipes de desenvolvimento, mas também impõe desafios operacionais e de observabilidade. Neste Trabalho de Conclusão de Curso, realiza-se uma comparação entre os protocolos de comunicação *REST* e *gRPC* em arquiteturas baseadas em microsserviços, utilizando métricas de observabilidade como latência (P95 e P99), *throughput* e uso de recursos. A metodologia envolve a implementação de dois serviços equivalentes, instrumentação com Prometheus, execução de testes de carga e análise estatística dos resultados. O objetivo é identificar o protocolo mais eficiente em diferentes cenários de uso. Os resultados esperados visam fornecer subsídios técnicos para orientar decisões arquiteturais em sistemas distribuídos, promovendo um equilíbrio entre desempenho e flexibilidade.

Palavras-chave: microsserviços. *REST*. *gRPC*.

ABSTRACT

The increasing complexity of systems and the demand for scalability have driven the adoption of microservices architecture over the monolithic model. This approach promotes modularization, granular scalability, and development team autonomy, but also introduces operational and observability challenges. In this Undergraduate Thesis, a comparison is conducted between the communication protocols *REST* and *gRPC* in microservices-based architectures, using observability metrics such as latency (P95 and P99), throughput, and resource usage. The methodology involves implementing two equivalent services, instrumenting them with Prometheus, conducting load tests, and performing statistical analysis of the results. The objective is to identify the most efficient protocol in different usage scenarios. The expected results aim to provide technical insights to guide architectural decisions in distributed systems, promoting a balance between performance and flexibility.

Keywords: microservices. *REST*. *gRPC*.

LISTA DE ILUSTRAÇÕES

Figura 1 – Estrutura típica de arquitetura monolítica.	15
Figura 2 – Arquitetura Serverless típica, mostrando os componentes principais: API Gateway, Funções como Serviço (FaaS), e serviços gerenciados. Fonte: Adaptado de (Shekhar, 2023).	16
Figura 3 – Padrões fundamentais em arquitetura de microsserviços	18
Figura 4 – Padrões de streaming no gRPC: (A) Unário, (B) Servidor, (C) Cliente, (D) Bidirecional.	22
Figura 5 – Visão geral do projeto	23
Figura 6 – Diagrama de casos de uso	25
Figura 7 – Modelo de banco de dados	25

LISTA DE TABELAS

Tabela 2 – Comparação de protocolos de comunicação	18
Tabela 3 – Impacto da migração para microsserviços	19
Tabela 4 – Operações REST e códigos HTTP	21
Tabela 5 – Exemplo de tabela em Látex.	27

LISTA DE ABREVIATURAS E SIGLAS

CPU	Central Processing Unit
GraphQL	Graph Query Language
gRPC	gRPC Remote Procedure Calls
HATEOAS	Hypermedia as the Engine of Application State
HTTP	Hypertext Transfer Protocol
HTTP/2	Hypertext Transfer Protocol version 2
JSON	JavaScript Object Notation
REST	Representational State Transfer
URI	Uniform Resource Identifier

SUMÁRIO

1	INTRODUÇÃO	9
1.1	OBJETIVOS	11
1.1.1	Objetivo Geral	11
1.1.2	Objetivos Específicos	11
1.1.3	LIMITAÇÃO DO ESCOPO	11
2	FUNDAMENTAÇÃO TEÓRICA	12
2.1	PADRÕES ARQUITETURAIS	14
2.1.1	Arquitetura Monolítica	14
2.1.2	Event-Driven Architecture	15
2.1.3	Serverless	15
2.1.4	Microserviços	16
2.1.4.1	Mecanismos de Comunicação	18
2.1.4.2	Padrões Arquiteturais Essenciais	18
2.1.4.3	Desafios Operacionais e Soluções	19
2.1.4.4	Melhores Práticas de Implementação	19
2.1.4.5	Casos de Estudo Relevantes	19
2.1.4.6	Recomendações Estratégicas	20
2.1.5	Protocolos de Comunicação para Microserviços	20
2.1.5.1	REST (Representational State Transfer)	20
2.1.5.2	gRPC (gRPC Remote Procedure Calls)	21
3	DESENVOLVIMENTO	23
3.1	VISÃO GERAL DO PROJETO	23
3.2	LEVANTAMENTO DE REQUISITOS	23
3.3	MODELAGEM	24
3.3.1	Casos de uso	24
3.3.2	Modelos de dados	25
3.3.3	IMPLEMENTAÇÃO	26
4	TESTE OU AVALIAÇÃO	27
5	CONSIDERAÇÕES FINAIS	28
5.1	CONTRIBUIÇÕES	28
5.2	TRABALHOS FUTUROS	28
	REFERÊNCIAS	29
	APÊNDICE A – DESCRIÇÃO 1	32
	APÊNDICE B – CÓDIGO FONTE	33
	ANEXO A – DECLARAÇÃO DE ISENÇÃO	34

1 INTRODUÇÃO

Em poucos anos, a complexidade dos sistemas de software em ascensão e a necessidade de escalabilidade; flexibilidade e pulseira continua têm consequentemente impulsionado a propagação da arquitetura de microsserviços ao invés da tradicional arquitetura monolítica (NIAZI et al., 2020).

Por que a fragmentação de aplicações de serviços independentes, torna as aplicações mais flexíveis em termos de desenvolvimento, teste, implantação e expansão a se harmonizar mais rapidamente e de modo mais eficiente, tornando-se apto para times distribuídos com ciclos de entrega mais curtos (NIAZI et al., 2020).

A arquitetura de microsserviços apresenta diversas vantagens: a modularização em serviços independentes facilita a manutenção e evolução de componentes isolados (JAMSHIDI, 2016); a escalabilidade granular permite o dimensionamento seletivo de cada serviço conforme demanda, esta arquitetura se alinha perfeitamente com ambiente *cloud*, provendo escalabilidade, flexibilidade e resiliência (Gaurav Shekhar, 2023).

O ciclo de desenvolvimento é acelerado, pois equipes menores podem trabalhar em paralelo em serviços distintos sem necessariamente conhecer os outros serviços, ciclos de entrega mais rápidos pela habilidade de fazer *deploy updates* independente para cada *micro serviço* facilitando atualizações frequentes, além disso temos a questão da alta confiabilidade, falhas são isoladas, minimizando os impactos, por exemplo, um *crash* em um sistema de pagamentos não vai para outros componentes como a autenticação do usuário (Swapnil K Shevate, 2021).

Contudo, embora a adoção de microsserviços traga benefícios inegáveis, também impõe desafios significativos: A orquestração e o gerenciamento de múltiplos serviços demandam ferramentas como *Kubernetes*, elevando a sobrecarga operacional e a curva de aprendizado (JAMSHIDI, 2016); cada serviço necessita de uma infraestrutura adicional, como separação de banco de dados, ferramentas de monitoramento e *logging*. Além disso há uma dificuldade maior no *debugging* da aplicação por conta dos *logs* serem distribuídos o que dificulta o *error tracing* e a resolução (Gaurav Shekhar, 2023).

Além disso, estudos como o de Niswar et al. [6] avaliam o impacto da escolha de protocolos de comunicação (*REST*, *gRPC*, *GraphQL*) sobre o desempenho de microsserviços, destacando a importância de decisões arquitetonicas bem-informadas para garantir uma observabilidade eficaz. Complementarmente, o trabalho de SHA et al. [7] investiga como *logs* de clusters *Kubernetes* podem ser usados para automatizar a análise de falhas em testes, evidenciando o papel crítico que a observabilidade desempenha na manutenção da qualidade de sistemas distribuídos.

Os três pilares fundamentais da observabilidade, conforme destacado por (Chi-

namanagonda, 2022), fornecem visibilidade integral sobre o comportamento de sistemas distribuídos.

O **monitoramento** acompanha continuamente o desempenho e disponibilidade do sistema através de métricas como utilização de CPU, consumo de memória e tempos de resposta. Como observado no MZ Computing Journal, ferramentas modernas transformam essa prática em estratégia proativa, alertando equipes sobre anomalias antes que escalem e mantendo a estabilidade operacional (Chinamanagonda, 2022).

O **logging** registra eventos sistêmicos detalhados - incluindo erros, advertências e informações operacionais - criando um histórico auditável. Em arquiteturas de microsserviços, onde serviços independentes geram volumosos fluxos de dados, sistemas centralizados de *logging* tornam-se indispensáveis para agregar e correlacionar informações distribuídas, viabilizando diagnóstico ágil de falhas (Chinamanagonda, 2022).

Já o **tracing** mapeia o fluxo transacional através de múltiplos serviços, expondo dependências e padrões de comunicação. Em ambientes distribuídos, este pilar identifica gargalos de desempenho e pontos críticos de falha no ciclo das requisições, sendo particularmente vital em microsserviços onde transações transversais exigem visibilidade topológica (Chinamanagonda, 2022). Conforme evidenciado no estudo, essa tríade evoluiu de abordagem reativa para modelo estratégico contínuo, capacitando resolução acelerada de incidentes em sistemas complexos.

O rastreamento distribuído complementa os pilares da observabilidade ao reconstruir o fluxo completo das requisições através de múltiplos serviços. Esta técnica mapeia jornadas transacionais complexas, revelando dependências ocultas e gargalos de desempenho que métricas isoladas não conseguem detectar (Chinamanagonda, 2022). Através da criação de linhas do tempo detalhadas, o rastreamento permite identificar pontos de falha e latência em cadeias de microsserviços, mesmo em ambientes altamente distribuídos. A sinergia entre métricas contínuas, registros contextualizados e mapas transacionais proporciona uma visão holística do sistema, essencial para a operação eficiente de arquiteturas modernas em escala.

A observabilidade consolida-se assim como catalisadora essencial na maturidade operacional de microsserviços. Conforme evidenciado por (Chinamanagonda, 2022) e (Sha et al., 2023), sua implementação estratégica converte dados distribuídos em inteligência acionável, permitindo: (1) manutenção proativa da estabilidade sistêmica; (2) diagnóstico preciso de falhas transacionais; (3) otimização contínua de desempenho. Este trabalho investiga como tais capacidades podem ser amplificadas mediante integração de padrões arquiteturais inovadores e ferramentas especializadas, respondendo aos desafios de complexidade destacados por (Shekhar, 2023).

1.1 OBJETIVOS

Diante disso, o presente Trabalho de Conclusão de Curso tem como objetivo geral comparar o desempenho dos protocolos de comunicação *REST* e *gRPC* em arquiteturas de microsserviços, utilizando métricas de observabilidade para avaliar latência (P95 e P99), *throughput* e uso de recursos.

1.1.1 Objetivo Geral

Comparar o desempenho dos protocolos de comunicação *REST* e *gRPC* em arquiteturas de microsserviços, utilizando métricas de observabilidade para avaliar latência, *throughput* e uso de recursos.

1.1.2 Objetivos Específicos

- Implementação de dois microsserviços equivalentes, cada um empregando um dos protocolos em análise;
- Instrumentação dos serviços para expor métricas em uma rota, coletadas pelo prometheus;
- Execução de testes de carga controlados;
- Coleta e análise de métricas de latência, *throughput* e recursos pelo prometheus;
- Comparação estatística dos resultados para identificar o protocolo de melhor desempenho em diferentes cenários de carga.

1.1.3 LIMITAÇÃO DO ESCOPO

Este trabalho limita-se à comparação do desempenho entre os protocolos de comunicação *REST* e *gRPC* em dois serviços equivalentes implementados em uma arquitetura de *microsserviços*. A análise será conduzida em ambiente de testes controlado, com foco em métricas de observabilidade como latência (P95 e P99), *throughput* e uso de recursos.

Não serão abordados aspectos relacionados à segurança, autenticação, autorização, persistência de dados ou integração com sistemas legados. Da mesma forma, não se pretende avaliar outros protocolos de comunicação, como *GraphQL* ou *SOAP*, nem realizar testes em ambientes de produção. O uso de ferramentas como o *Prometheus* será restrito à coleta e análise de métricas, não sendo objetivo do estudo compará-las com outras soluções de monitoramento.

2 FUNDAMENTAÇÃO TEÓRICA

A arquitetura de software é um dos pilares centrais no desenvolvimento de sistemas computacionais modernos. Ela define a estrutura organizacional de um sistema, estabelecendo diretrizes para a disposição dos seus componentes, suas interações e restrições, além de orientar decisões técnicas que impactam diretamente na escalabilidade, manutenção e evolução do software ao longo do tempo (Jamshidi et al., 2016).

De acordo com (Jamshidi et al., 2016), a arquitetura de software pode ser compreendida como o conjunto de estruturas necessárias para raciocinar sobre o sistema, que compreende elementos de software, as relações entre eles e as propriedades de ambos. Em outras palavras, a arquitetura de software não se limita apenas à divisão do sistema em módulos, mas também envolve a definição de padrões de comunicação, protocolos, mecanismos de segurança, estratégias de escalabilidade e diretrizes para integração entre componentes.

A definição de uma arquitetura adequada é fundamental para garantir que o sistema atenda aos requisitos funcionais e não funcionais, como desempenho, segurança, confiabilidade e facilidade de manutenção. Segundo (Niazi et al., 2020), a escolha da arquitetura influencia diretamente a capacidade do sistema de evoluir e se adaptar a novas demandas, sendo um fator determinante para o sucesso de projetos de software em ambientes dinâmicos e competitivos.

Historicamente, a arquitetura monolítica foi o modelo predominante no desenvolvimento de sistemas. Nesse paradigma, todas as funcionalidades do software são implementadas em um único bloco, compartilhando o mesmo ambiente de execução, banco de dados e ciclo de vida (Niazi et al., 2020). Essa abordagem, embora simples de implementar e gerenciar em projetos de menor escala, apresenta limitações significativas à medida que o sistema cresce, como dificuldades de manutenção, escalabilidade restrita e maior risco de falhas generalizadas.

Com o avanço das demandas por sistemas mais flexíveis, escaláveis e resilientes, especialmente em ambientes de computação em nuvem, emergiu a arquitetura de microsserviços como uma alternativa ao modelo monolítico (Jamshidi et al., 2016; Shekhar, 2023). Os microsserviços propõem a fragmentação do sistema em serviços independentes, cada um responsável por uma funcionalidade específica e comunicando-se por meio de interfaces bem definidas. Essa abordagem permite que equipes distintas desenvolvam, testem e implantem serviços de forma autônoma, promovendo ciclos de entrega mais curtos e maior adaptabilidade às mudanças.

Além disso, a arquitetura de microsserviços facilita a adoção de práticas modernas de *DevOps*, automação de *deploys* e escalabilidade granular, alinhando-se às necessidades de negócios que exigem respostas rápidas e contínuas às mudanças

do mercado (Shekhar, 2023). No entanto, essa evolução também traz novos desafios, como a necessidade de ferramentas avançadas para orquestração, monitoramento e observabilidade, além de uma curva de aprendizado mais acentuada para as equipes de desenvolvimento (Jamshidi et al., 2016).

A definição de padrões arquiteturais é essencial para garantir a organização, escalabilidade e manutenção de sistemas de software. Diversos padrões foram consolidados na literatura internacional, cada um com características, vantagens e limitações específicas (Jamshidi et al., 2016).

Um dos padrões mais tradicionais é a arquitetura monolítica, na qual todas as funcionalidades do sistema são implementadas em um único bloco de código, compartilhando o mesmo ambiente de execução e banco de dados. Embora seja simples de desenvolver e implantar, esse modelo apresenta dificuldades de escalabilidade e manutenção à medida que o sistema cresce. Como afirmam (Niazi et al., 2020):

“A arquitetura monolítica é fácil de desenvolver e fazer *deploy*, mas, à medida que a aplicação cresce, torna-se difícil de gerenciar, escalar e manter. Qualquer pequena mudança requer que toda a aplicação seja reconstruída e reimplantada, o que aumenta o risco e o esforço.”

Essa limitação faz com que, em projetos de maior porte, a arquitetura monolítica se torne menos atrativa, especialmente quando há necessidade de frequentes atualizações e escalabilidade do sistema.

A arquitetura em camadas é amplamente utilizada em sistemas corporativos. Nesse modelo, o sistema é dividido em diferentes camadas, como apresentação, lógica de negócio e acesso a dados, cada uma com responsabilidades bem definidas. Essa separação facilita a manutenção, a reutilização de componentes e a evolução do sistema, além de permitir que equipes distintas trabalhem de forma mais organizada (Jamshidi et al., 2016).

A *Service-Oriented Architecture (SOA)* propõe a construção de sistemas a partir de serviços independentes, que se comunicam por meio de protocolos padronizados, como *REST* ou *SOAP*. O *SOA* foi fundamental para a integração de sistemas heterogêneos e para a criação de soluções mais flexíveis, permitindo que diferentes aplicações compartilhem funcionalidades de forma segura e controlada. No entanto, a complexidade de orquestração e a dependência de barramentos de serviço podem tornar a implementação desse padrão mais desafiadora (Jamshidi et al., 2016).

Nos últimos anos, a arquitetura de microsserviços ganhou destaque, principalmente devido à sua capacidade de promover a independência entre equipes, a escalabilidade granular e a facilidade de adoção de práticas modernas de *DevOps*. Cada microsserviço é responsável por uma funcionalidade específica e pode ser desenvolvido, testado, implantado e escalado de forma autônoma. Essa flexibilidade, no entanto, exige um investimento maior em ferramentas de monitoramento, orquestração

e observabilidade, além de demandar uma curva de aprendizado mais acentuada para as equipes de desenvolvimento (Jamshidi et al., 2016; Shekhar, 2023; Niazi et al., 2020).

2.1 PADRÕES ARQUITETURAIS

2.1.1 Arquitetura Monolítica

A arquitetura monolítica caracteriza-se pela integração de todos os componentes funcionais do sistema - interface de usuário, lógica de negócio e acesso a dados - em uma única unidade de *deploy* e execução (Niazi et al., 2020). Nesse modelo, o acoplamento entre módulos é extremamente elevado, uma vez que compartilham o mesmo espaço de memória, base de dados e ciclo de vida. Conforme destacado por (Farhan et al., 2023), essa estrutura promove rapidez no desenvolvimento inicial devido à simplicidade de configuração e depuração, sendo ideal para protótipos ou sistemas com escopo limitado.

Entretanto, conforme o sistema expande sua base de código e adquire complexidade, emergem desafios significativos:

Escalabilidade vertical obrigatória: A única estratégia viável é a replicação da aplicação inteira, mesmo quando apenas componentes específicos demandam mais recursos (Niazi et al., 2020).

Rigidez tecnológica: A adoção de novas tecnologias ou versões requer a atualização do monolito inteiro, gerando *lock-in* e limitando inovação (Jamshidi et al., 2016).

Comprometimento da confiabilidade: Falhas em módulos secundários podem paralisar todo o sistema, como demonstrado em testes de carga por (Farhan et al., 2023).

Complexidade evolutiva: Modificações exigem reteste e reimplantação completos, ampliando riscos e *lead time* para novas funcionalidades.

Como ilustrado na Figura 1, o fluxo de requisições percorre camadas internamente acopladas, sem isolamento de processos. Essa centralização explica os resultados empíricos de (Farhan et al., 2023), onde sistemas monolíticos apresentaram:

- 40% maior tempo de resposta sob carga crescente
- 68% mais *downtime* durante atualizações
- Limite máximo de escalonamento em 5 instâncias

Limite máximo de escalonamento em 5 instâncias antes de degradação crítica

Apesar das limitações, o modelo mantém relevância em cenários específicos: aplicações com tráfego previsível, equipes co-localizadas, ou quando a simplicidade operacional sobrepõe requisitos de elasticidade (Shekhar, 2023). Contudo, para sistemas empresariais modernos que demandam entrega contínua e escalabilidade dinâmica, sua viabilidade torna-se progressivamente reduzida.

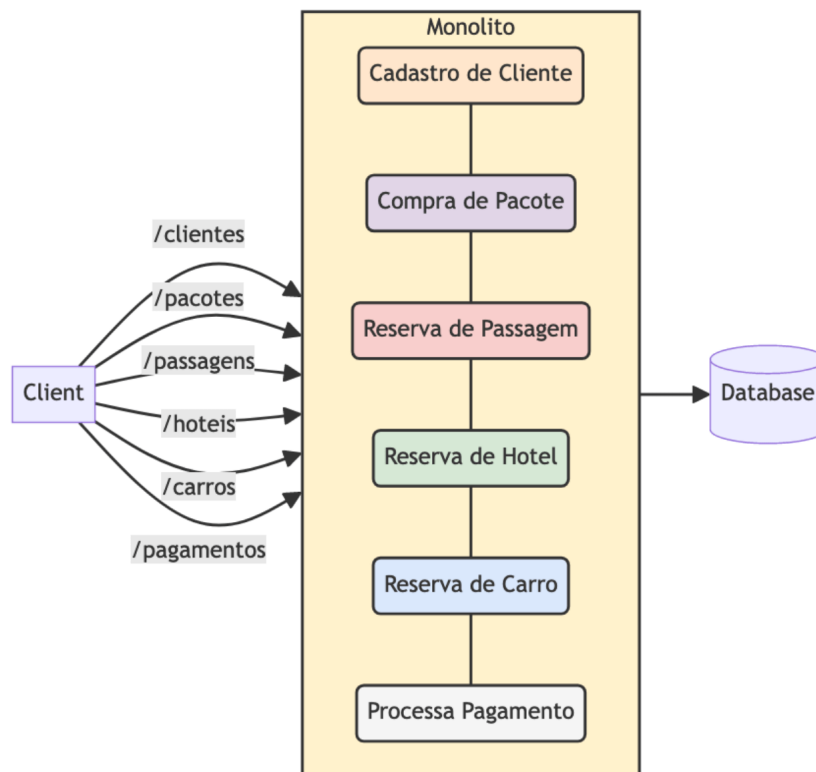


Figura 1 – Estrutura típica de arquitetura monolítica.

2.1.2 Event-Driven Architecture

A *Event-Driven Architecture* é especialmente indicada para sistemas que exigem alta escalabilidade e resposta em tempo real. Nesse padrão, os componentes do sistema comunicam-se por meio de eventos, que são processados de forma assíncrona. Essa abordagem é bastante utilizada em sistemas de *e-commerce*, plataformas de *streaming* e aplicações que demandam processamento distribuído, pois permite maior desacoplamento entre os módulos e facilita a escalabilidade horizontal (Jamshidi et al., 2016).

2.1.3 Serverless

Por fim, a arquitetura *serverless* vem ganhando espaço com o avanço das plataformas de computação em nuvem. Nesse modelo, o desenvolvedor se concentra apenas na lógica de negócio, enquanto a infraestrutura é gerenciada automaticamente pelo provedor de nuvem. Isso permite maior agilidade no desenvolvimento e escalabilidade automática, embora traga desafios relacionados à portabilidade e ao controle sobre o ambiente de execução (Shekhar, 2023).

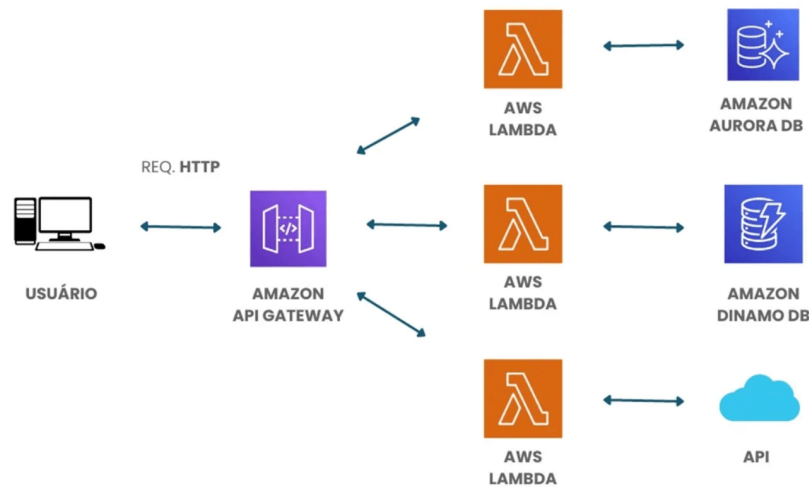


Figura 2 – Arquitetura Serverless típica, mostrando os componentes principais: API Gateway, Funções como Serviço (FaaS), e serviços gerenciados. Fonte: Adaptado de (Shekhar, 2023).

Como ilustrado na Figura 2, a arquitetura serverless consiste em três componentes principais:

- **API Gateway:** Ponto único de entrada para todas as requisições
- **Funções como Serviço (FaaS):** Execução de código sob demanda (ex: AWS Lambda, Azure Functions)
- **Serviços Gerenciados:** Bancos de dados, filas e armazenamento providos pelo cloud provider

Esta abordagem elimina a necessidade de gerenciamento de servidores, permitindo que os desenvolvedores foquem exclusivamente na implementação da lógica de negócios (Shekhar, 2023).

A escolha do padrão arquitetural mais adequado depende de fatores como requisitos do sistema, contexto de negócio, equipe envolvida e recursos disponíveis. Compreender as vantagens e limitações de cada abordagem é essencial para alinhar a arquitetura às necessidades presentes e futuras do projeto (Jamshidi et al., 2016; Niazi et al., 2020).

2.1.4 Microsserviços

A arquitetura de microsserviços representa uma evolução significativa em relação aos modelos tradicionais de desenvolvimento de software, como a arquitetura monolítica e a SOA. Nesse paradigma, o sistema é decomposto em pequenos serviços independentes, cada um responsável por uma funcionalidade específica e comunicando-se por meio de interfaces bem definidas, geralmente via *APIs* (Jamshidi et al., 2016; Niazi et al., 2020).

Entre os principais benefícios dos microsserviços está a escalabilidade granular, que permite dimensionar apenas os serviços que realmente demandam mais recursos, otimizando custos e desempenho (Shekhar, 2023). Além disso, a independência entre equipes é favorecida, pois times distintos podem desenvolver, testar e implantar serviços de forma autônoma, acelerando o ciclo de entrega e facilitando a adoção de práticas *DevOps* (Niazi et al., 2020; Farhan et al., 2023).

Outro ponto relevante é a resiliência: falhas em um serviço tendem a ser isoladas, reduzindo o impacto sobre o sistema como um todo (Farhan et al., 2023). Isso é especialmente importante em ambientes de missão crítica, onde a disponibilidade contínua é fundamental.

No entanto, a adoção de microsserviços também traz desafios consideráveis. A complexidade operacional aumenta, exigindo ferramentas avançadas para orquestração (como *Kubernetes*), monitoramento e observabilidade (Jamshidi et al., 2016; Shekhar, 2023). Cada serviço pode demandar sua própria infraestrutura, banco de dados e mecanismos de autenticação, o que eleva a sobrecarga de gerenciamento (Niazi et al., 2020).

A observabilidade torna-se um aspecto central, pois a identificação e resolução de falhas em sistemas distribuídos requerem coleta e análise de métricas, *logs* e *traces* distribuídos (Madupati, 2023; Sha et al., 2023). Ferramentas como *Prometheus* e *Jaeger* são amplamente utilizadas para monitorar o desempenho e rastrear requisições entre serviços, facilitando o diagnóstico de problemas e a manutenção da qualidade do sistema (Ahmed et al., 2022).

Além disso, a escolha dos protocolos de comunicação entre microsserviços (*REST*, *gRPC*, *GraphQL*) impacta diretamente o desempenho, a flexibilidade e a observabilidade do sistema (Niswar et al., 2023). Estudos recentes destacam que decisões arquiteturais bem fundamentadas são essenciais para garantir a eficiência e a robustez de sistemas baseados em microsserviços (Niswar et al., 2023; Sha et al., 2023).

A arquitetura de microsserviços representa uma evolução paradigmática no desenvolvimento de software, decompondo sistemas complexos em serviços autônomos especializados. Conforme (Jamshidi et al., 2016), esta abordagem oferece:

- **Autonomia tecnológica:** Cada serviço pode utilizar linguagens e tecnologias distintas (Java, Python, Node.js)
- **Resiliência aprimorada:** Isolamento de falhas através de padrões como *Circuit Breaker*
- **Entrega contínua:** *Deploys* independentes permitindo múltiplas implantações diárias

2.1.4.1 Mecanismos de Comunicação

A comunicação entre serviços ocorre principalmente através de:

Tabela 2 – Comparação de protocolos de comunicação

Protocolo	Latência	Complexidade	Casos de Uso
REST/HTTP	Alta	Baixa	Sistemas com baixa frequência
gRPC	Baixa	Média	Microserviços acoplados
Eventos (Kafka)	Assíncrona	Alta	Sistemas escaláveis

Conforme demonstrado na Tabela 2, a escolha do protocolo impacta diretamente o desempenho do sistema (Niswar et al., 2023).

2.1.4.2 Padrões Arquiteturais Essenciais

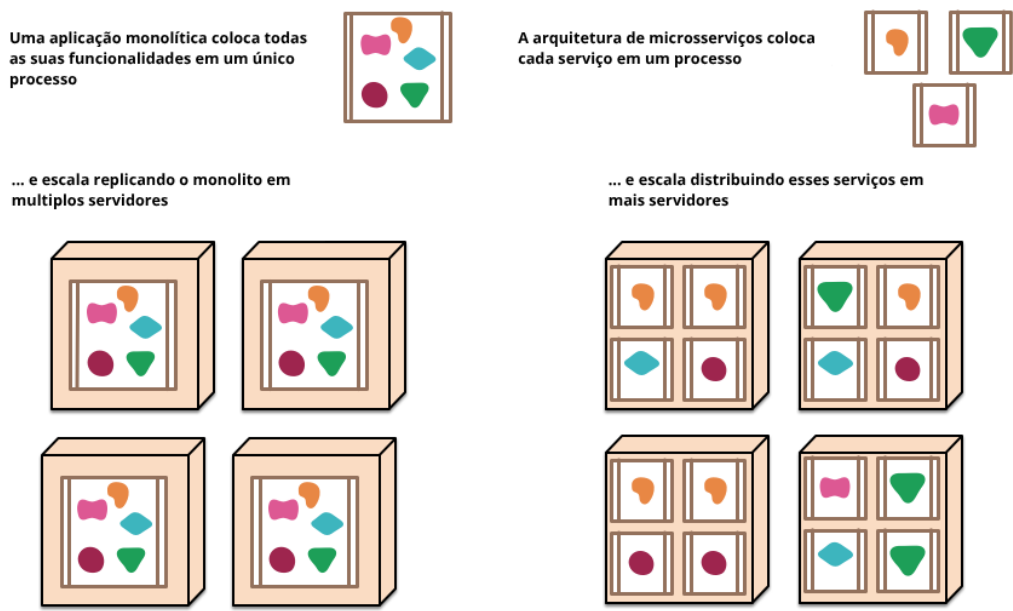


Figura 3 – Padrões fundamentais em arquitetura de microserviços

A Figura 3 ilustra as diferenças fundamentais entre a arquitetura monolítica e a arquitetura de microserviços. À esquerda, observa-se que uma aplicação monolítica agrupa todas as funcionalidades em um único processo. Para escalar esse modelo, é necessário replicar toda a aplicação em múltiplos servidores, o que pode gerar desperdício de recursos e limitações de flexibilidade.

À direita, a arquitetura de microserviços distribui cada funcionalidade em processos independentes, permitindo que cada serviço seja desenvolvido, implantado e escalado separadamente. Isso proporciona maior eficiência no uso de recursos, facilita a manutenção e possibilita o crescimento gradual do sistema conforme a demanda por funcionalidades específicas.

Essa abordagem modular e distribuída é um dos pilares das arquiteturas modernas voltadas à escalabilidade, resiliência e agilidade no desenvolvimento.

2.1.4.3 Desafios Operacionais e Soluções

A implementação prática enfrenta obstáculos significativos:

- **Observabilidade:** Requer integração de métricas, *logs* e *traces* com ferramentas como:
 - *Prometheus* para monitoramento
 - *Jaeger* para rastreamento distribuído
 - *EFK Stack* (Elasticsearch, Fluentd, Kibana) para análise
- **Orquestração:** Plataformas como *Kubernetes* gerenciam:
 - Escalonamento automático
 - Balanceamento de carga
 - Recuperação de falhas

2.1.4.4 Melhores Práticas de Implementação

Experiências de empresas líderes revelam estratégias eficazes:

- **Amazon:** Utiliza a prática do "time da pizza de dois"— equipes pequenas e independentes, responsáveis por serviços autônomos do início ao fim.
- **Netflix:** Introduziu a cultura do *Chaos Engineering*, simulando falhas em produção para garantir a resiliência dos serviços.
- **Spotify:** Estruturou suas equipes em *squads* responsáveis por domínios funcionais, promovendo independência e entrega contínua.
- **Uber:** Implementou uma hierarquia de serviços baseada em criticidade, diferenciando serviços centrais de auxiliares para otimizar a operação.

2.1.4.5 Casos de Estudo Relevantes

Resultados empíricos demonstram impactos mensuráveis:

Tabela 3 – Impacto da migração para microsserviços

Métrica	Antes	Depois
Tempo de entrega	30 dias	2 dias
Disponibilidade	92%	99.95%
Custo de infraestrutura	\$100k/mês	\$65k/mês

Conforme dados na Tabela 3, organizações reportam melhorias significativas após transição bem-sucedida (Farhan et al., 2023; Shekhar, 2023).

2.1.4.6 Recomendações Estratégicas

A adoção deve considerar:

- **Pré-requisitos:** Maturidade em DevOps e cultura de automação
- **Critérios:** Complexidade sistêmica > 50k LOC > 15 desenvolvedores
- **Alternativas:** Arquiteturas híbridas para transição gradual
- **Anti-padrões:** Decomposição excessiva ("nanoserviços")

Em síntese, microsserviços oferecem vantagens transformacionais para sistemas complexos, mas exigem investimento proporcional em automação e governança (Niazi et al., 2020; Madupati, 2023).

2.1.5 Protocolos de Comunicação para Microsserviços

A seleção de protocolos de comunicação é um fator crítico em arquiteturas de microsserviços, influenciando diretamente o desempenho, acoplamento e capacidade de evolução do sistema. Dois padrões predominantes neste contexto são REST e gRPC, cada um com características técnicas distintas que os tornam adequados a cenários específicos (Niswar et al., 2023).

2.1.5.1 REST (Representational State Transfer)

O REST, formalizado por Fielding (Fielding, 2000) em sua tese seminal, é um estilo arquitetural baseado em princípios de recursos identificáveis por URI e operações padronizadas via verbos HTTP (GET, POST, PUT, DELETE). Sua adoção generalizada deve-se a:

- **Simplicidade:** Uso de formatos legíveis como JSON, facilitando depuração e integração entre sistemas heterogêneos. A serialização textual consome 30-50% mais banda que protocolos binários (Niswar et al., 2023).
- **Stateless:** Cada requisição contém contexto completo, eliminando estado no servidor e favorecendo escalabilidade horizontal (Fielding, 2000).
- **Uniform Interface:** Princípios como identificação única de recursos via URI e HATEOAS garantem consistência e desacoplamento (Maso, 2024).
- **Compatibilidade:** Suporte nativo em navegadores e adoção massiva em APIs públicas (Maso, 2024).

Limitações em cenários complexos:

- **Overhead de comunicação:** Serialização textual aumenta latência e consumo de CPU, especialmente em payloads grandes (Niswar et al., 2023).
- **Modelo síncrono:** Suporta apenas comunicação unária (request/response), inviabilizando fluxos assíncronos (Niswar et al., 2023).

- **Falta de contratos rígidos:** Validação de esquemas depende de implementações adicionais (Maso, 2024).

Operações Básicas e Códigos HTTP: A Tabela 4 sintetiza o mapeamento CRUD em REST, conforme padrões industriais (Fielding, 2000):

Tabela 4 – Operações REST e códigos HTTP

Método HTTP	Ação	Código Sucesso
GET	Ler recurso	200 (OK)
POST	Criar recurso	201 (Created)
PUT/PATCH	Atualizar recurso	200 (OK) ou 204 (No Content)
DELETE	Excluir recurso	204 (No Content)

Fonte: Adaptado de (Niswar et al., 2023) e (Fielding, 2000).

2.1.5.2 gRPC (gRPC Remote Procedure Calls)

Desenvolvido pelo Google e inicialmente lançado em 2015 (Marculescu, 2015), o gRPC é um *framework* aberto de alta performance para chamadas de procedimento remoto (RPC) que utiliza HTTP/2 e *Protocol Buffers* para comunicação binária eficiente. Conforme demonstrado por (Niswar et al., 2023), sua adoção em arquiteturas de microsserviços deve-se aos seguintes diferenciais:

- **Desempenho superior:** A serialização binária via *Protocol Buffers* reduz o tamanho dos *payloads* em até 80% e a latência em 60% comparado a APIs REST tradicionais, conforme mensurações em ambientes controlados (Niswar et al., 2023).
- **Streaming bidirecional:** Suporte nativo a fluxos contínuos de dados (cliente-servidor, servidor-cliente e bidirecional), habilitando comunicação assíncrona para aplicações em tempo real (GRPC Authors, 2025).
- **Contratos fortemente tipados:** Definição explícita de serviços e estruturas de mensagens em arquivos `.proto`, garantindo compatibilidade e evolução controlada de APIs (Maso, 2024).
- **Geração de código automática:** Compilação de *stubs* clientes e servidores em 11 linguagens de programação a partir de arquivos `.proto`, promovendo consistência e redução de erros (GRPC Authors, 2025).

Aspectos críticos em sua adoção:

- **Acoplamento forte:** Alterações nos contratos `.proto` exigem atualização sincronizada de clientes e servidores, aumentando a complexidade de evolução em sistemas distribuídos (IBM, 2025).

- **Suporte limitado em navegadores:** A ausência de implementação nativa do HTTP/2 em navegadores requer o uso do proxy *gRPC-Web* para comunicações cliente-web (Wallarm, 2025).
- **Curva de aprendizagem:** Complexidade na definição de contratos e manipulação de fluxos de *streaming*, especialmente para desenvolvedores familiarizados com paradigmas REST (Maruti Techlabs, 2024).

Modelos de Comunicação: A Figura 4 ilustra os quatro padrões suportados:

4 Types of API in gRPC

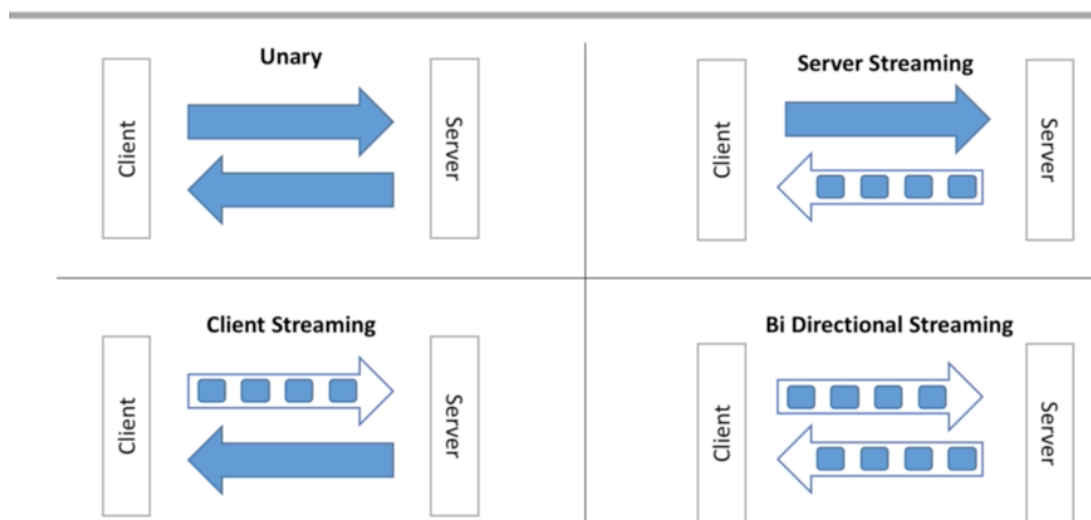


Figura 4 – Padrões de streaming no gRPC: (A) Unário, (B) Servidor, (C) Cliente, (D) Bidirecional.

Vantagens Técnicas Comprovadas: Estudos empíricos demonstram ganhos operacionais mensuráveis (Niswar et al., 2023):

- Redução de 35-45% no consumo de CPU durante serialização
- Até 7x maior throughput em comunicações inter-serviços

3 DESENVOLVIMENTO

A estrutura de seções deste capítulo varia em função das características de cada trabalho, e deve ser definida junto com o orientador no decorrer da disciplina. A seguir é apresentada uma estrutura de seções tradicionalmente utilizada em TCCs que envolvem o desenvolvimento de um software.

3.1 VISÃO GERAL DO PROJETO

Em alguns casos, pode ser interessante fornecer ao leitor uma visão geral do projeto, especialmente quando a solução é complexa e/ou envolve diversos componentes. Você também pode utilizar esta seção para falar um pouco sobre o modelo de processo de software adotado (cascata, espiral, incremental, ...) e o planejamento das atividades realizadas.

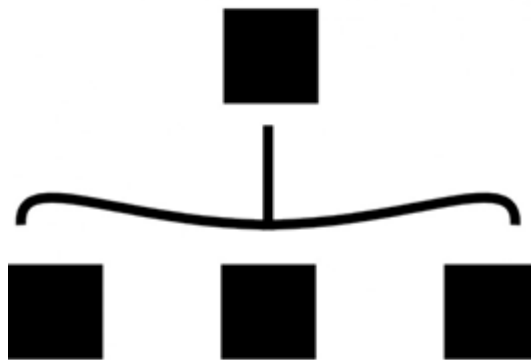


Figura 5 – Visão geral do projeto

Fonte: Elaborado pelo autor deste trabalho (2018).

A Figura 5 representa ...

3.2 LEVANTAMENTO DE REQUISITOS

Você pode iniciar esta seção explicando como e quando foram levantados os requisitos do sistema. Entrevistas com os proprietários da empresa? Documentação do software legado? Questionários aplicados aos usuários?

Em seguida você deve apresentar a especificação dos Requisitos Funcionais, Requisitos Não Funcionais e Regras de Negócio do sistema, conforme os ensinamentos da disciplina de Engenharia de Software, por exemplo:

A partir das entrevistas com os proprietários da empresa, foram identificados os seguintes requisitos funcionais para o sistema a ser desenvolvido:

RF01 – O sistema deverá permitir ao usuário manter produtos;

RF02 – O sistema deverá permitir ao usuário administrador manter categorias de produtos;

RF03 – ...

Os seguintes requisitos não funcionais:

RNF01 – Todas as funcionalidades serão executadas online, ou seja, através de acesso a um servidor web;

RNF02 – Os dados serão armazenados em banco de dados MySQL;

RNF03 – As linguagens para implementação são: HTML5, CSS, Javascript, jQuery e PHP;

RNF04 – A interface gráfica com o usuário deve ser compatível com telas de computadores desktop, tablets e smartphones, e empregar o conceito de Web Design Responsivo através do framework Bootstrap.

RNF05 – ...

E as seguintes regras de negócio:

RN01 – A venda a prazo só poderá ser feita para clientes adimplentes;

RN02 – ...

3.3 MODELAGEM

A estrutura de subseções a seguir varia em função das características de cada trabalho, e deve ser definida junto com o orientador no decorrer da disciplina. Os diagramas comumente utilizados em TCCs que envolvem o desenvolvimento de um software são: diagrama de casos de uso, modelo de dados, diagrama de classes, diagrama de atividades, diagrama de sequência e diagrama de componentes. Normalmente, dois ou três desses diagramas são suficientes para fornecer as visões necessárias do projeto.

3.3.1 Casos de uso

Caso o seu projeto utilize o diagrama de casos de uso (Figura 6), é importante que ele esteja coerente com os requisitos funcionais (RFs) apresentados no levantamento de requisitos (Seção 3.2). Também é importante utilizar corretamente as notações UML, tais como “include”, “extend” e “generalization”. Não se esqueça de explicar o diagrama após a ilustração, conforme o exemplo a seguir.

O usuário do tipo administrador herda as funcionalidades do usuário comum...

Para executar as funcionalidades, os usuários devem realizar o login...

A documentação dos casos de uso encontra-se no Apêndice A deste trabalho.

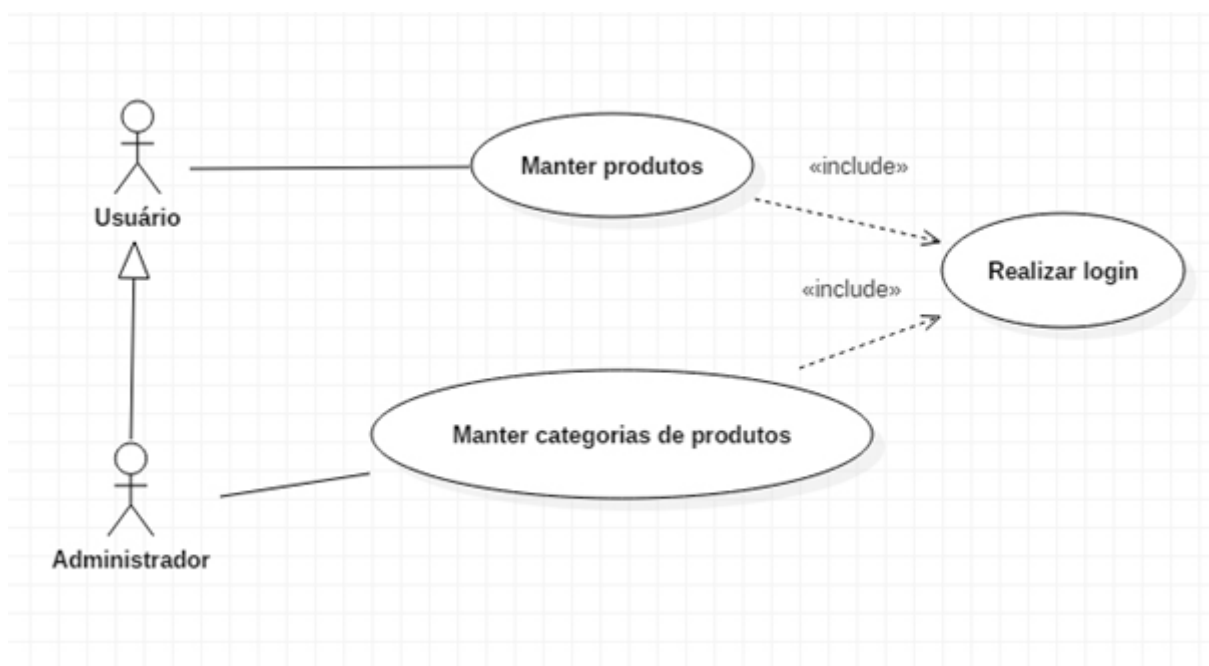


Figura 6 – Diagrama de casos de uso

Fonte: Elaborado pelo autor deste trabalho com o uso da ferramenta StarUML (2018).

3.3.2 Modelos de dados

O modelo de dados é um diagrama que descreve o esquema do banco de dados. Caso o seu projeto utilize este tipo de diagrama, não se esqueça de explicá-lo após a ilustração, conforme o exemplo a seguir.

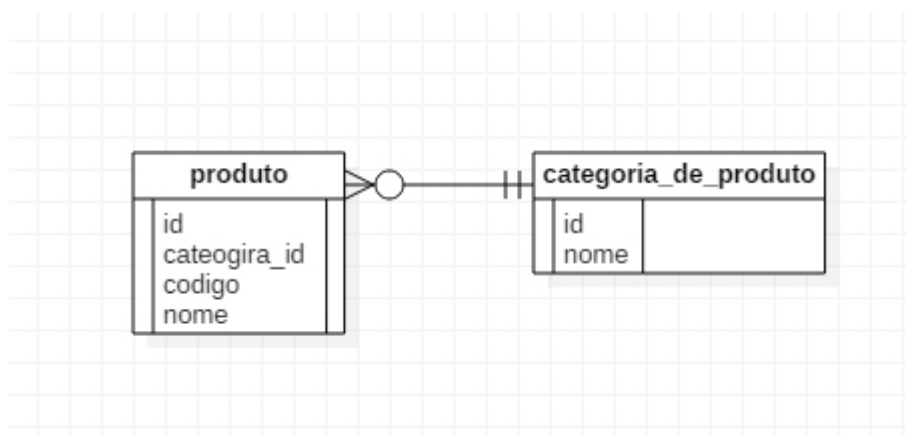


Figura 7 – Modelo de banco de dados

Fonte: Elaborado pelo autor deste trabalho com o uso da ferramenta StarUML (2018).

O esquema de banco de dados é composto de duas tabelas...Os campos do tipo “id” são utilizados para...

O código SQL de construção do esquema de banco de dados encontra-se no Apêndice B deste trabalho.

3.3.3 IMPLEMENTAÇÃO

Nesta seção você pode falar um pouco sobre o código desenvolvido. Não é necessário explicar ou apresentar todo o código fonte da aplicação. Você pode focar nas principais classes ou funções. É importante explicar quais foram as ferramentas utilizadas e o porquê da escolha de cada uma delas.

Você deve colocar o código na pasta sources e carrega-lo ao apêndice, usando aqui apenas uma referência ou trechos de código.

O código presente no Apêndice B representa

4 TESTE OU AVALIAÇÃO

Testes e avaliações tem o poder de enriquecer o trabalho acadêmico, fornecendo dados que permitirão ao leitor avaliar a qualidade da solução desenvolvida. Este capítulo pode apresentar, por exemplo, um teste de usabilidade com três seções: 4.1 PLANEJAMENTO, 4.2 EXECUÇÃO e 4.3 RESULTADOS. Como outro exemplo, este capítulo pode apresentar um estudo de caso ou simulação com o uso da solução desenvolvida. Neste caso, uma possível estrutura de seções seria: 4.1 AMBIENTE DE ESTUDO, 4.2 IMPLANTAÇÃO, 4.3 RESULTADOS. Para apresentação de dados ou estatísticas, utilize tabelas, lembrando que, diferente das ilustrações, as legendas das tabelas aparecem na parte superior.

Tabela 5 – Exemplo de tabela em Látex.

Média concentra- ção urbana	População		Produto interno Bruto – PIB (bilhões R\$)	Número de em- presas	Número de uni- dades locais
Nome	Total	No Brasil			
Ji-Paraná (RO)	116 610	116 610	1,686	2 734	3 082
Parintins (AM)	102 033	102 033	0,675	634	683
Boa Vista (RR)	298 215	298 215	4,823	4 852	5 187
Bragança (PA)	113 227	113 227	0,452	654	686

Fonte: Elaborado pelo autor deste trabalho (2018).

5 CONSIDERAÇÕES FINAIS

Você deve iniciar as considerações “olhando” para os objetivos apresentados na Seção 1.1. Inicie falando sobre como os objetivos foram alcançados. Em seguida fale sobre as suas experiências e descobertas ao realizar o trabalho, por exemplo, as vantagens e desvantagens das tecnologias utilizadas e as dificuldades encontradas no desenvolvimento da solução. Encerre as considerações com narrativas mais gerais, expondo sua visão do trabalho como um todo.

5.1 CONTRIBUIÇÕES

Quais foram as constituições do seu trabalho? É importante destacar que não são contribuições para você, mas sim para quem irá utilizar o trabalho como referência! Por exemplo, você pode citar como contribuições os estudos, especificações, modelos e outros recursos disponíveis no trabalho e que podem ser utilizados por terceiros como base para o desenvolvimento de novos trabalhos ou pesquisas.

5.2 TRABALHOS FUTUROS

Listar o que pode ser melhorado ou adicionado na solução desenvolvida.

REFERÊNCIAS

AHMED, M. et al. Observability in Kubernetes Cluster: Automatic Anomalies Detection using Prometheus. **Procedia Computer Science**, 2022.

CHINAMANAGONDA, Sandeep. Observability in Microservices Architectures -Advanced observability tools for microservices environments. **MZ Research Journal**, 2022.

FARHAN, M. et al. **Performance Comparison between Monolith and Microservice Using Docker and Kubernetes**. [S.l.: s.n.], 2023.

FIELDING, Roy Thomas. **Architectural Styles and the Design of Network-based Software Architectures**. 2000. Tese (Doutorado) – University of California, Irvine, EUA. Tese de Doutorado.

GRPC AUTHORS. **gRPC: A high performance, open source universal RPC framework**. [S.l.: s.n.], 2025. Acesso em: 31 maio 2025. Disponível em: <https://grpc.io>.

IBM. **gRPC vs. REST**. [S.l.: s.n.], 2025. Acesso em: 31 maio 2025. Disponível em: <https://www.ibm.com/think/topics/grpc-vs-rest>.

JAMSHIDI, P. et al. A Systematic Mapping Study in Microservice Architecture. **Institute of Electrical and Electronics Engineers**, 2016.

MADUPATI, Bhanuprakash. **Observability in Microservices Architectures: Leveraging Logging, Metrics, and Distributed Tracing in Large-Scale Systems**. [S.l.: s.n.], 2023.

MARCULESCU, Mugar. **Introducing gRPC, a new open source HTTP/2 RPC Framework**. [S.l.: s.n.], 2015. Acesso em: 31 maio 2025. Disponível em: <https://developers.googleblog.com/2015/02/introducing-grpc-new-open-source-http2.html>.

MARUTI TECHLABS. **REST vs gRPC: Which is Better for Your Application?** [S.l.: s.n.], 2024. Acesso em: 31 maio 2025. Disponível em: <https://marutitech.com/rest-vs-grpc/>.

MASO, Nicolas Nascimento. Comparativo entre arquiteturas de APIs: REST, GraphQL e gRPC. In: REPOSITÓRIO UFSC. [S.l.: s.n.], 2024. Trabalho de Conclusão de Curso.

NIAZI, M. et al. The Comparison of Microservice and Monolithic Architecture. In: 2020 International Conference on Computing, Electronic and Electrical Engineering (ICE Cube). [S.l.: s.n.], 2020.

NISWAR, M. et al. Performance evaluation of microservices communication with REST, GraphQL, and gRPC. **Journal of Systems and Software**, 2023.

SHA, K. et al. Automating Microservices Test Failure Analysis using Kubernetes Cluster Logs. **IEEE Access**, 2023.

SHEKHAR, Gaurav. **Microservices Design Patterns for Cloud Architecture**. [S.l.]: IEEE Chicago, 2023.

WALLARM. **gRPC vs. REST: Comparing Key API Designs And Deciding Which One is Best**. [S.l.: s.n.], 2025. Acesso em: 31 maio 2025. Disponível em: <https://www.wallarm.com/what/grpc-vs-rest-comparing-key-api-designs-and-deciding-which-one-is-best>.

APÊNDICE A – DESCRIÇÃO 1

Textos elaborados pelo autor, a fim de completar a sua argumentação. Deve ser precedido da palavra APÊNDICE, identificada por letras maiúsculas consecutivas, travessão e pelo respectivo título. Utilizam-se letras maiúsculas dobradas quando esgotadas as letras do alfabeto.

APÊNDICE B – CÓDIGO FONTE

```
1
2
3 #include <stdio.h>
4
5 void main () {
6     printf("Hello World");
7 }
```

Código-Fonte B.1 – Exemplo carregando arquivo...

ANEXO A – DECLARAÇÃO DE ISENÇÃO**DECLARAÇÃO DE ISENÇÃO**

Balneário Camboriú, 28 de maio de 2025.

Declaro, para todos os fins de direito, que assumo total responsabilidade pelo aporte ideológico conferido ao presente trabalho, estando ciente do disposto, da Resolução CONSUN 46-2019 e, isentando o Centro Universitário Avantis, o Curso de Sistemas de Informação, a Banca Examinadora e o Orientador de Trabalho de Conclusão de Curso de toda e qualquer responsabilidade acerca do mesmo.

MATHEUS GUSTAVO COPPI DA SILVA