



UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO TECNOLÓGICO
CURSO DE GRADUAÇÃO EM CIÊNCIAS DA COMPUTAÇÃO

Rafael Begnini de Castilhos

**Análise da estrutura e formatação de dados para comunicação eficiente entre
ambientes de computação em nuvem**

Florianópolis, Santa Catarina
2022

Rafael Begnini de Castilhos

Análise da estrutura e formatação de dados para comunicação eficiente entre ambientes de computação em nuvem

Proposta de Trabalho de Conclusão de Curso do
Curso de Graduação em Ciências da
Computação do Centro Tecnológico da
Universidade Federal de Santa Catarina para a
obtenção do título de Bacharel em Ciências da
Computação.

Orientador: Prof. Carlos Becker Westphall, Dr.

Florianópolis, Santa Catarina
2022

FOLHA DE APROVAÇÃO DE PROPOSTA DE TCC

Acadêmico(s)	Rafael Begnini de Castilhos
Título do trabalho	Análise da estrutura e formatação de dados para comunicação eficiente entre ambientes de computação em nuvem
Curso	Ciências da Computação INE/UFSC
Área de Concentração	Arquitetura de Sistemas de Computação

Instruções para preenchimento pelo ORIENTADOR DO TRABALHO:

- Para cada critério avaliado, assinale um X na coluna SIM apenas se considerado aprovado. Caso contrário, indique as alterações necessárias na coluna Observação.

Critérios	Aprovado				Observação
	Sim	Parcial	Não	Não se aplica	
1. O trabalho é adequado para um TCC no CCO/SIN (relevância / abrangência)?					
2. O título do trabalho é adequado?					
3. O tema de pesquisa está claramente descrito?					
4. O problema/hipóteses de pesquisa do trabalho está claramente identificado?					
5. A relevância da pesquisa é justificada?					
6. Os objetivos descrevem completa e claramente o que se pretende alcançar neste trabalho?					
7. É definido o método a ser adotado no trabalho? O método condiz com os objetivos e é adequado para um TCC?					
8. Foi definido um cronograma coerente com o método definido (indicando todas as atividades) e com as datas das entregas (p.ex. Projeto I, II, Defesa)?					
9. Foram identificados custos relativos à execução deste trabalho (se houver)? Haverá financiamento para estes custos?					
10. Foram identificados todos os envolvidos neste trabalho?					
11. As formas de comunicação foram definidas (ex: horários para orientação)?					
12. Riscos potenciais que podem causar desvios do plano foram identificados?					
13. Caso o TCC envolva a produção de um software ou outro tipo de produto e seja desenvolvido também como uma atividade realizada numa empresa ou laboratório, consta da proposta uma declaração (Anexo 3) de ciência e concordância com a entrega do código fonte e/ou documentação produzidos?					

Avaliação	<input type="checkbox"/> Aprovado <input type="checkbox"/> Não Aprovado		
Professor Responsável	Prof. Dr. Carlos Becker Westphall	(Data)	(Assinatura)

RESUMO

A computação em nuvem teve um crescimento exponencial, principalmente para aplicativos da web comerciais. De acordo com o Instituto Nacional de Padrões e Tecnologia (NIST), “a computação em nuvem é um modelo para permitir acesso de rede onipresente, conveniente e sob demanda a um conjunto compartilhado de recursos de computação configuráveis (por exemplo, redes, servidores, armazenamento, aplicativos e serviços) que podem ser rapidamente provisionados e liberados com o mínimo de esforço de gerenciamento e/ou interação com o provedor de serviços”. Por essas razões, os aplicativos cada vez mais frequentemente realizam trocas de informações entre serviços com finalidade de compartilhar, replicar, armazenar e entre outras funcionalidades características de um sistema distribuído, e, com isso, se faz necessário definição da estrutura e formatação por parte de desenvolvedores, arquitetos e engenheiros. A escolha deste formato e protocolo implica diretamente no tamanho, na latência e eficiência da comunicação. Preambularmente, será realizado o desenvolvimento de uma aplicação que atuará como emitente que enviará dados fictícios de diferentes tamanhos de um comércio eletrônico para um destinatário. Posteriormente, será efetuado monitoramento dos dados trafegados, sendo possível extrair métricas e avaliar os resultados obtidos de maneira comparativa e exploratória. Logo, esse trabalho visa analisar, comparar e encontrar as opções de formato de dados e protocolos já existentes que são satisfatórias para este cenário, potencializando operações escaláveis e utilizando menos recursos.

Palavras-chave: Computação em nuvem. Serialização. Desserialização. Sistemas distribuídos. Redes.

LISTA DE ILUSTRAÇÕES

Figura 1 - Modelo de arquitetura Cliente/Servidor.....	11
Figura 2 - Fluxograma das etapas utilizadas na metodologia CRISP-DM.....	12

LISTA DE TABELAS

Tabela 2.2 - Gráfico de Gantt.....	13
Tabela 2.3 - Custos.....	14
Tabela 2.4 - Recursos Humanos.....	14
Tabela 2.5 - Fluxo de comunicação.....	15
Tabela 2.6 - Riscos ativos.....	15

LISTA DE ABREVIATURAS E SIGLAS

AWS	Amazon Web Services
GCP	Google Cloud Plataform
VM	Máquina Virtual

SUMÁRIO

1	INTRODUÇÃO	9
1.1	OBJETIVOS.....	10
2	PLANEJAMENTO	11
2.1	MATERIAIS E MÉTODOS.....	11
2.2	CRONOGRAMA.....	13
2.3	CUSTOS.....	14
2.4	RECURSOS HUMANOS.....	14
2.5	COMUNICAÇÃO.....	15
2.6	RISCOS.....	15
	REFERÊNCIAS.....	16

1. INTRODUÇÃO

A computação em nuvem é amplamente adotada na indústria como uma tecnologia que permite acesso fácil e barato ao processamento e armazenamento de dados. A nuvem é apoiada por servidores físicos que hospedam máquinas virtuais (VMs) que são fornecidas aos usuários (JINDAL, *et al.*, 2019). Além disso, a computação em nuvem permite provisionar serviços através da Internet, de forma escalável e sob demanda (EMEAKAROHA, *et al.*, 2016).

Durante as últimas décadas, muitas aplicações de *software* que até então operavam de maneira *standalone*, como por exemplo: sistemas de reserva, sistema bancário e comércio eletrônico e entre outras, passaram a ser distribuídas. De acordo com Abdullah, *et al.* (2022), um sistema distribuído é uma coleção de sistemas autônomos, estações de trabalho e servidores conectados para fornecer serviços ao cliente ou usuário virtualmente usando um protocolo chamado *middleware*. O objetivo do sistema distribuído é compartilhar recursos, aumentar a disponibilidade, rendimento, eficiência e superar falhas.

A comunicação entre os serviços é dada através de mensagens que são geradas por um serviço emissor, trafegadas e recebidas por um serviço destinatário. Para isso, é necessário que o emissor serialize o dado que será transacionado, estruture de acordo com o protocolo que será utilizado e por fim que o destinatário desserialize. Entretanto, diferentes mecanismos podem ser aplicados para realizar a serialização e desserialização, além de que variados protocolos podem ser utilizados.

Existem muitas aplicações cuja funcionalidade principal é o envio e recebimento de dados pela rede. O aumento de dados intercambiáveis pela Internet fez a seleção de um formato de serialização de dados adequado cada vez mais importante (SUMARAY; MAKKI. 2012). A serialização e desserialização de objetos é um recurso essencial para a comunicação eficiente entre nós de computação distribuídos com ambientes de execução potencialmente não uniformes (JANG, *et al.*, 2020). Desse modo, é imperioso destacar a relevância desse assunto diante a quantidade diária de troca de informação em ambientes distribuídos em nuvem.

Este trabalho propõe desenvolver uma aplicação responsável por emitir e receber informações utilizando diferentes abordagens, a fim de obter um aglomerado de dados para extrair métricas, analisar e comparar os variados mecanismos

responsáveis. Com isso será possível identificar os métodos mais satisfatórios, objetivando aumentar a eficiência, reduzir os custos e a latência, e também identificar os que devem ser evitados. Logo, este trabalho visa concluir as melhores abordagens para estruturar e formatar os dados na comunicação entre dispositivos, potencializando operações escaláveis utilizando menos recursos.

1.1. OBJETIVOS

O objetivo geral deste trabalho é desenvolver uma aplicação para trafegar mensagens visando eficiência entre máquinas em ambiente de nuvem. Para isso, diferentes métodos de serialização, desserialização e protocolos serão considerados e comparados entre si.

Objetivos específicos:

- Examinar a literatura referente a comunicação eficiente em ambiente de nuvem.
- Analisar estrutura e formatação na serialização e desserialização de objetos.
- Pesquisar diferentes técnicas e bibliotecas para serializar e desserializar objetos, a fim de enriquecer análises e comparações.
- Analisar protocolos para comunicação entre dispositivos na internet.
- Comparar os dados obtidos nas análises, a fim de expor resultados satisfatórios e insatisfatórios.

2. PLANEJAMENTO

As próximas seções discutirão aspectos relacionados ao desenvolvimento do trabalho, incluindo material e métodos de pesquisa; cronograma, custos e recursos humanos esperados e, por fim, possíveis riscos.

2.1. MATERIAL E MÉTODOS

O desenvolvimento da aplicação se dará por intermédio do ambiente de desenvolvimento integrado *IntelliJ IDEA*, utilizando a linguagem de programação Java. Para executar o envio e recebimento das mensagens, serão utilizadas duas máquinas virtuais localizadas na região do Norte da Virgínia, Estados Unidos. A primeira será instanciada na provedora Amazon Web Services (AWS) e será do modelo *t2.micro*, e a segunda será instanciada na provedora Google Cloud Platform (GCP) e terá o modelo *e2-micro*.

Pretende-se definir o escopo do conteúdo dos dados trafegados como informações fictícias de diferentes tamanhos de compras de produtos em um comércio eletrônico imaginário, contendo informações como: produto(s) comprado, valores parciais e total, comprador, endereço de entrega e transportador.

Uma das máquinas virtuais irá desempenhar o papel de cliente, e irá emitir as mensagens, enquanto a outra máquina virtual terá como responsabilidade ser o receptor das mensagens, simulando um servidor, conforme ilustrado na Figura 1.

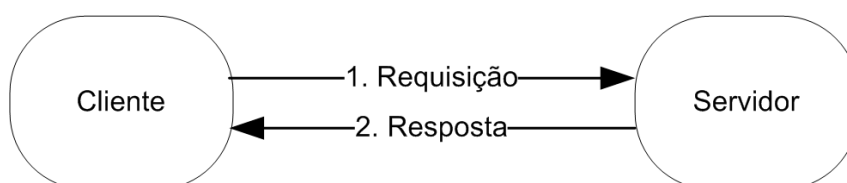


Figura 1 - Modelo de arquitetura Cliente/Servidor

Os dados utilizados neste trabalho serão obtidos a partir do monitoramento das ações do cliente e do servidor, e serão manipulados utilizando Shell Script e Python. A metodologia empregada para obtenção de dados e tratamento dos mesmos será *Cross-Industry Standard Process for Data Mining* (CRISP-DM) (SCHRÖER, *et al.*, 2021), na qual é uma modelagem de processos para Data Mining. A Figura 2 apresenta os passos definidos na metodologia CRISP-DM.

A metodologia CRISP-DM é dividida em seis fases iterativas e não estritas, sendo elas:

1. *Business understanding*: busca estudos referentes aos recursos e literaturas requeridas e obtidas.
2. *Data understanding*: propõe a coleta de dados.
3. *Data preparation*: consiste em inclusão e exclusão de dados por meio de definições de critérios.
4. *Modeling*: corresponde a classificação dos dados, além de avaliar e comparar os resultados.
5. *Evaluation*: se refere a verificação se os resultados obtidos correspondem com os objetivos propostos.
6. *Deployment*: diz respeito a finalização do projeto, sendo a entrega do relatório.

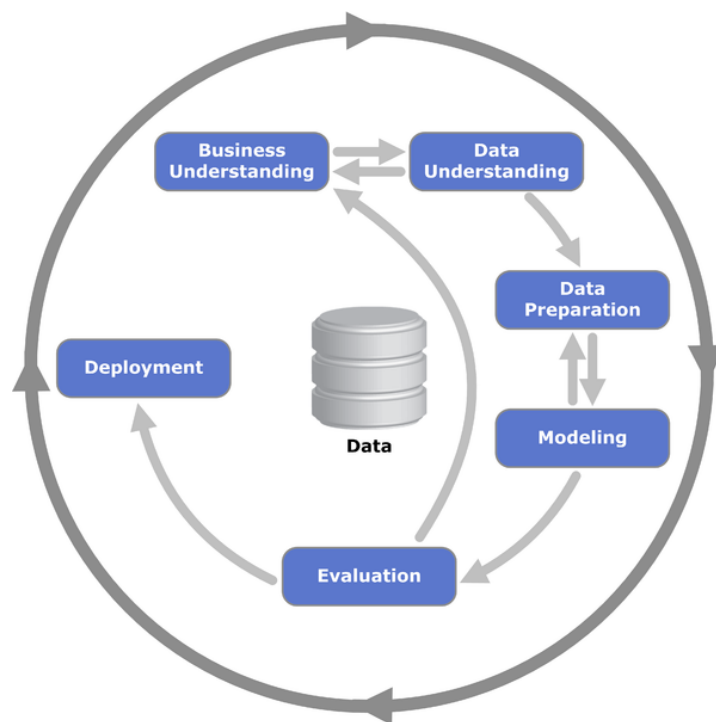


Figura 2 - Fluxograma das etapas utilizadas na metodologia CRISP-DM

2.2. CRONOGRAMA

O gráfico de Gantt do cronograma é apresentado na tabela 2.2.

Etapas	Meses											
	ago.	set.	out.	nov.	dez.	jan.	fev.	mar.	abr.	mai.	jun.	jul.
Revisão do estado da arte												
Preparação e provisionamento do ambiente												
Desenvolvimento do sistema de análise												
Relatório TCC I												
Análise e extração dos dados												
Avaliação dos resultados												
Redação do rascunho do TCC												
Relatório TCC II												
Preparação da defesa pública												
Defesa pública												
Ajustes no relatório final do TCC												

Tabela 2.2 - Gráfico de Grantt.

2.3. CUSTOS

A tabela almeja listar os custos indiretos e diretos previstos para o projeto, demais custos como softwares, licenças e literatura não serão contabilizados pois o autor possui acesso gratuito. Além disso, também não serão contabilizadas as horas de trabalhadas do autor e dos professores envolvidos. O autor não recebe bolsa de pesquisa, portanto o projeto também não possui esse orçamento.

Item	Quantidade	Valor unitário (R\$)	Valor Total (R\$)
Material permanente			
Internet Residencial e 4G	12	100,00	1.200,00
Material de consumo			
Folhas impressas	100	0,20	20,00
Outros recursos e serviços			
Aluguel máquina virtual t2.micro AWS	12	50,00	600,00
Aluguel máquina virtual e2-micro GCP	12	60,00	720,00
TOTAL			2.540,00

Tabela 2.3 - Custos

2.4. RECURSOS HUMANOS

Os recursos humanos do projeto e seus papéis estão descritos na tabela 2.4.

Nome	Função
Rafael Begnini de Castilhos	Autor
Carlos Becker Westphall	Orientador/Professor Responsável
A definir	Membro da banca
A definir	Membro da banca

Tabela 2.4 - Recursos Humanos

2.5. COMUNICAÇÃO

O fluxo de comunicação está descrito na tabela 2.5

O que precisa ser comunicado	Por quem	Para quem	Melhor forma de comunicação	Quando e com que frequência
Ante-projeto	Autor	Coordenador de projetos	Via site de projetos	Singular
Reuniões com orientador	Autor	Orientador	Videoconferência ou presencial	Semanalmente ou quinzenalmente
Dúvidas	Autor	Orientador	E-mail e/ou videoconferência	Conforme necessidade
Relatório de TCC I	Autor	Orientador e membros da banca	Papel impresso e/ou via site de projetos	Singular, próximo da conclusão
Relatório de TCC II	Autor	Orientador e membros da banca	Papel impresso e/ou via site de projetos	Singular, próximo da conclusão

Tabela 2.5 - Fluxo de comunicação

2.6. RISCOS

Os riscos mais prováveis e perigosos levantados foram apresentados na tabela 2.6.

Risco	Probabilidade	Impacto	Prioridade	Estratégia de resposta	Ações de prevenção
Perda de dados	baixa	alta	alta	Utilizar ferramenta de versionamento (GitHub)	Geração de Backups diariamente
Alteração no tema	baixa	alto	alta	Buscar novo tema ou modificar escopo atual	Manter interação constante com orientador
Problemas de saúde	baixa	alto	alta	Realizar tratamento	Manter bons hábitos e ser prudente
Alteração no cronograma	média	alto	alta	Realizar as adaptações necessárias	Manter-se dentro do cronograma e atentar aos prazos.
Indisponibilidade de serviços	baixa	alto	alta	Verificar gravidade e/ou realizar migração para outros serviços de nuvem pública	Manter-se atualizado das novidades e monitoramento dos serviços

Tabela 2.6 - Riscos ativos

REFERÊNCIAS

M. Hogan and A. Sokol. NIST Cloud Computing Standards Roadmap Version 2. NIST Cloud Computing Standards Roadmap Working Group, NIST Special Publications 500-291, NIST, Gaithersburg, MD, 2013, p.1-113

Schroer, Christoph; KRUSE, Félix; GÓMEZ, Jorge Marx. A Systematic Literature Review on Applying CRISP-DM Process Model. [S.l.: s.n.], 2021. v. 181, p. 526–534. DOI: 10.1016/j.procs.2021.01.199.

Abdullah, Khalid & Alzubaidi, Abdulaziz & Dauda, Muhammed & Al-Yahya, Mohammed & Al-Haddad, Mohammed. (2022). Distributed Systems: Concepts, Principles, Models and Algorithms. Journal of Early Modern Studies. 6. 256-269.

Sumaray, Audie & Makki, S.. (2012). A comparison of data serialization formats for optimal efficiency on a mobile platform. 10.1145/2184751.2184810.

Emeakaroha, Vincent & Healy, Philip & Fatema, Kaniz & Morrison, John. (2013). Analysis of Data Interchange Formats for Interoperable and Efficient Data Communication in Clouds. 10.1109/UCC.2013.79.

Jindal, Anshul & Podolskiy, Vladimir & Gerndt, Michael. (2019). Performance Modeling for Cloud Microservice Applications. 25-32. 10.1145/3297663.3310309.

Emeakaroha, M. Bullman & J. P. Morrison, Towards Automated Cost-Efficient Data Management for Federated Cloud Services. 2016 5th IEEE International Conference on Cloud Networking (Cloudnet), 2016, pp. 158-163, doi: 10.1109/CloudNet.2016.37.