



**UNIVERSIDADE FEDERAL DE CAMPINA GRANDE
CENTRO DE ENGENHARIA ELÉTRICA E INFORMÁTICA
CURSO DE BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO**

WESLEY HENRIQUE ARAÚJO MONTE

**REDASH: UMA SOLUÇÃO DE MONITORAMENTO E ANÁLISE DE DADOS PARA
GERENCIAMENTO DE RECURSOS EM TEMPO REAL EM NUVENS PRIVADAS**

CAMPINA GRANDE - PB

2023

WESLEY HENRIQUE ARAÚJO MONTE

**REDASH: UMA SOLUÇÃO DE MONITORAMENTO E ANÁLISE DE DADOS PARA
GERENCIAMENTO DE RECURSOS EM TEMPO REAL EM NUVENS PRIVADAS**

**Trabalho de Conclusão Curso
apresentado ao Curso Bacharelado em
Ciência da Computação do Centro de
Engenharia Elétrica e Informática da
Universidade Federal de Campina
Grande, como requisito parcial para
obtenção do título de Bacharel em
Ciência da Computação.**

Orientador : Thiago Emmanuel Pereira da Cunha Silva

CAMPINA GRANDE - PB

2023

WESLEY HENRIQUE ARAÚJO MONTE

**REDASH: UMA SOLUÇÃO DE MONITORAMENTO E ANÁLISE DE DADOS PARA
GERENCIAMENTO DE RECURSOS EM TEMPO REAL EM NUVENS PRIVADAS**

**Trabalho de Conclusão Curso
apresentado ao Curso Bacharelado em
Ciência da Computação do Centro de
Engenharia Elétrica e Informática da
Universidade Federal de Campina
Grande, como requisito parcial para
obtenção do título de Bacharel em
Ciência da Computação.**

BANCA EXAMINADORA:

Thiago Emmanuel Pereira da Cunha Silva

Orientador – UASC/CEEI/UFCG

Carlos Wilson Dantas de Almeida

Examinador – UASC/CEEI/UFCG

Francisco Vilar Brasileiro

Professor da Disciplina TCC – UASC/CEEI/UFCG

Trabalho aprovado em: 28 de Junho de 2023.

CAMPINA GRANDE - PB

REDASH: A MONITORING AND DATA ANALYSIS SOLUTION FOR REAL-TIME RESOURCE MANAGEMENT IN PRIVATE CLOUDS

ABSTRACT

The Laboratório de Sistemas Distribuídos (LSD) is responsible for managing the private cloud at UASC/UFCG. In order to maintain this private cloud, the LSD faced challenges in managing the cloud resources, especially in visualizing the data and generating new reports for those responsible for the projects that use these resources. The tool used previously was Shylock, but it presented difficulties in generating reports on the amount and consumption of cloud resources because it only generated one report per day. In addition, reporting was a complex process that required the creation of a new Jinja model and the coding of variables. Given the shortcomings identified in the current system, Redash was chosen as the platform to address them. Redash solves the difficulty of visualizing the analyzed data by creating dashboards for real-time visualization and monitoring of cloud resources. By incorporating the information from the dashboards created in Redash into the cloud management and monitoring workflow, the ease of making decisions based on actual, up-to-date data has enabled better decision making in the management of cloud resources.

Redash: Uma solução de monitoramento e análise de dados para gerenciamento de recursos em tempo real em nuvens privadas

Wesley Henrique Araújo Monte*
wesley.monte@ccc.ufcg.edu.br
Departamento de Sistemas e Computação
Universidade Federal de Campina Grande
Campina Grande, Paraíba - Brasil

Thiago Emmanuel Pereira da Cunha Silva
Departamento de Sistemas e Computação
Universidade Federal de Campina Grande
Campina Grande, Paraíba - Brasil
temmanuel@computacao.ufcg.edu.br

RESUMO

O Laboratório de Sistemas Distribuídos (LSD) é responsável por gerenciar a nuvem privada do UASC/UFCG. Para manter essa nuvem privada, o LSD enfrentou desafios no gerenciamento dos recursos da nuvem, especialmente na visualização de dados e na geração de novos relatórios para os responsáveis pelos projetos que usam esses recursos.

Anteriormente, a ferramenta utilizada era o Shylock, mas ela apresentava dificuldades na geração de relatórios sobre a quantidade e o consumo de recursos da nuvem, pois gerava apenas um relatório por dia. Além disso, a criação de relatórios era um processo complexo que exigia a criação de um novo modelo Jinja e a codificação de variáveis.

Dado os déficits encontrados no sistema atual, o Redash foi a plataforma selecionada para supri-los. O Redash resolve a dificuldade de visualização dos dados analisados através da criação de dashboards para a visualização e monitoramento dos recursos da nuvem em tempo real.

Ao incluir as informações dos dashboards criados no Redash no fluxo de trabalho de gerenciamento e monitoramento da nuvem, a facilidade de tomar decisões com base em dados reais e atualizados permitiu uma melhor tomada de decisões sobre o gerenciamento dos recursos da nuvem.

1 INTRODUÇÃO

O advento da computação em nuvem revolucionou a maneira como as empresas operam, oferecendo uma solução flexível, dimensionável e econômica para armazenamento e processamento de dados. No entanto, gerenciar recursos em um ambiente de nuvem pode ser uma tarefa complexa, especialmente em nuvens privadas.

As nuvens privadas são exclusivas de uma única organização e não têm faturamento, proporcionando um maior nível de controle e privacidade, mas muitas vezes às custas de maiores responsabilidades de gestão e requisitos de infraestrutura. Portanto, é preciso controlar o uso de recursos para evitar desperdícios. A Universidade Federal de Campina Grande (UFCG) é um exemplo de organização que opera sua própria nuvem privada, a nuvem da Unidade Acadêmica de Sistemas e Computação (UASC). Essa infraestrutura, mantida pelo Laboratório de Sistemas Distribuídos (LSD), oferece recursos de nuvem adaptados às necessidades de seus funcionários e estudantes, mas também levanta desafios únicos em termos de gestão e atribuição eficientes desses recursos.

Nesse contexto, o Laboratório de Sistemas Distribuídos (LSD) abordou o desafio do gerenciamento de recursos inicialmente com um sistema conhecido como Shylock. Embora o Shylock tenha fornecido uma solução para o gerenciamento de recursos, ele tinha

várias limitações. O sistema tinha dificuldade para apresentar os dados coletados de forma clara e compreensível, seus relatórios eram enviados apenas uma vez por dia, o que os tornava desatualizados para o resto do dia, e criar novos relatórios era um processo complexo que exigia a criação de um novo modelo Jinja e a codificação de variáveis.

Para enfrentar esses desafios, propomos uma nova solução que aproveita o poder do Redash, uma ferramenta de visualização e análise de dados que simplifica a geração de relatórios e fornece acesso em tempo real aos dados. Nosso sistema visa simplificar as tarefas de gerenciamento de recursos na nuvem no LSD, tornando-as menos dispendiosas e mais eficientes. Essa solução agrega dados automaticamente, trata-os com eficiência, fornece uma visualização clara da análise e é capaz de notificar eventos relevantes. Além disso, ela é facilmente acessível aos usuários na nuvem.

2 FUNDAMENTAÇÃO

Apesar da funcionalidade inerente do sistema de geração de relatórios e consulta de métricas de recursos da nuvem do Laboratório de Sistemas Distribuídos (LSD), conhecido como Shylock, a eficiência desse processo pode ser aprimorada. O Shylock enfrenta desafios com a apresentação clara dos dados coletados, causando dificuldades na compreensão dos padrões de uso de recursos e resultando em decisões subótimas.

O Shylock requer um esforço significativo para adicionar novos relatórios devido à complexidade de desenvolver novos templates Jinja e codificar as variáveis para serem renderizadas nesses templates. Além disso, os relatórios se tornam rapidamente obsoletos; após a coleta e inserção dos dados no relatório, eles podem já não representar a situação atual de maneira precisa, levando a possíveis equívocos na tomada de decisões ou à perda de oportunidades valiosas.

Outra limitação do Shylock é que ele envia relatórios apenas uma vez por dia, tornando-os desatualizados para o resto do dia. Este cenário limita a capacidade dos operadores de tomar decisões em tempo real com base nas informações mais recentes. A criação de novos relatórios também é uma tarefa complexa, envolvendo a elaboração de um novo modelo Jinja e a codificação de variáveis, prejudicando a flexibilidade e adaptabilidade da ferramenta para atender a diversas necessidades de relatórios.

O Shylock falha em oferecer uma visualização adequada dos dados, o que poderia facilitar a identificação de padrões e tendências, tornando o processo de tomada de decisões mais rápido e menos complexo. Este cenário se agrava quando gerentes de projetos não conseguem acompanhar o consumo ao longo do mês por falta de informações precisas, levando a possíveis decisões equivocadas.

A proposta deste artigo é superar essas limitações através da implementação do Redash, uma plataforma de consulta de data-sources que adiciona uma camada user-friendly sobre o banco de dados, permitindo registro e reutilização de consultas. Este sistema utiliza a linguagem SQL universal para consultar dados do InfluxDB e do PostgreSQL, possibilitando a criação de painéis personalizados para diferentes grupos de usuários. Além disso, o Redash oferece funcionalidades de visualização de dados em diversos formatos, organizando essas visualizações em dashboards. Através do monitoramento do desempenho dos dados pelo Redash, possíveis problemas podem ser identificados de forma mais rápida e eficiente.

Em essência, apresentamos o Redash como uma solução abrangente para análise de dados, oferecendo uma ampla gama de funcionalidades para o gerenciamento de recursos em nuvens privadas do LSD. Esta proposta tem como objetivo abordar as lacunas existentes em termos de eficiência e clareza na visualização de dados, buscando aprimorar o processo de tomada de decisões baseado em dados.

3 OPENSTACK

O OpenStack é uma poderosa plataforma de computação em nuvem de código aberto que fornece infraestrutura como serviço (IaaS). Ela é amplamente usada em nuvens privadas devido à sua flexibilidade, escalabilidade e robustez. Para entender como o OpenStack pode ser aproveitado para o gerenciamento de recursos em tempo real, é essencial compreender alguns conceitos e componentes principais.

O OpenStack opera com vários conceitos fundamentais que definem sua arquitetura e funcionalidade:

- **Imagem:** Este é um template de uma máquina virtual. Ela contém um sistema operacional e, possivelmente, software adicional. As imagens são usadas para criar novas instâncias.
- **Grupo de segurança:** É um conjunto de regras de firewall que define qual tráfego de rede de entrada é permitido às instâncias.
- **Par de chaves:** É um par de chaves criptográficas (uma pública e uma privada) usado para fazer login nas instâncias de forma segura.
- **Instância:** É uma máquina virtual criada a partir de uma imagem. As instâncias são os recursos de computação na nuvem.
- **Flavor:** É uma configuração de hardware predefinida para uma instância. Ela define a capacidade de computação, memória e armazenamento de uma instância.
- **Volume:** É um dispositivo de armazenamento em bloco removível, semelhante a um disco rígido USB. Você pode anexar um volume a apenas uma instância de cada vez.
- **Usuário:** É um indivíduo que usa os serviços do OpenStack. Um usuário é associado a um nome de usuário exclusivo e autenticado usando uma senha ou um par de chaves.
- **Projeto:** É uma unidade organizacional que permite que usuários ou grupos gerenciem seu próprio conjunto de recursos dentro da infraestrutura de nuvem do OpenStack.
- **Domínio:** É uma unidade administrativa no OpenStack. Ela agrupa projetos e usuários, fornecendo um mecanismo para gerenciar o acesso.

Os principais componentes do OpenStack trabalham juntos para fornecer um conjunto abrangente de ferramentas e serviços para o gerenciamento de uma nuvem privada. Sua arquitetura modular permite flexibilidade e escalabilidade, enquanto componentes como o Monasca oferecem informações valiosas para o gerenciamento de recursos em tempo real. À medida que nos aprofundamos no uso do Redash, veremos como essas capacidades podem ser aproveitadas para um gerenciamento eficaz de recursos em nuvens privadas.

O OpenStack é modular, com cada componente oferecendo um serviço distinto. Aqui estão alguns dos principais componentes:

- **Keystone:** Este é o serviço de identidade no OpenStack. É responsável pelo gerenciamento de usuários e catálogo de serviços. Ele também lida com a autenticação de clientes de API e autorização multi-inquilino.
- **Nova:** Este é o serviço de computação. Ele é responsável pelo gerenciamento do ciclo de vida das instâncias em um ambiente OpenStack.
- **Cinder:** Este é o serviço de armazenamento em bloco. Ele fornece dispositivos de armazenamento em bloco persistentes para uso com instâncias OpenStack.
- **Glance:** Este é o serviço de imagens. Ele permite que os usuários descubram, registrem e recuperem imagens de máquinas virtuais.
- **Neutron:** Este é o serviço de rede. Ele garante que a rede não seja um gargalo ou fator limitante em uma implantação em nuvem e dá aos usuários a capacidade de autoatendimento sobre as configurações de rede.
- **Monasca:** Este é o serviço de monitoramento e diagnóstico. É usado para coleta, armazenamento e análise de métricas, fornecendo uma visão unificada dos recursos no ambiente OpenStack. Monasca pode monitorar e acompanhar métricas como uso de CPU, taxas de E/S de disco, tráfego de rede e uso de memória, tornando-o uma ferramenta vital para o gerenciamento de recursos em tempo real.

3.1 Monasca e comunicação entre componentes: Coleta e gerenciamento de métricas

O Monasca se destaca entre os componentes do OpenStack como uma solução de monitoramento como serviço multilocatário, altamente dimensionável, de alto desempenho e tolerante a falhas. Ele é usado para coleta, armazenamento e análise de métricas, fornecendo uma visão unificada dos recursos no ambiente OpenStack. O Monasca pode monitorar e rastrear métricas como uso da CPU, taxas de E/S de disco, tráfego de rede e uso de memória, o que o torna uma ferramenta essencial para o gerenciamento de recursos em tempo real.

A comunicação entre os componentes do OpenStack é crucial para a operação contínua da nuvem, e o Monasca desempenha um papel fundamental nesse processo. Quando um usuário interage com o ambiente do OpenStack, por exemplo, ao iniciar uma nova instância, vários componentes se comunicam entre si por meio de APIs. O componente Nova, responsável pelo gerenciamento do ciclo de vida das instâncias, comunica-se com o componente Glance para recuperar a imagem necessária, com o componente Neutron para configurar a rede e com o componente Cinder se for necessário anexar um volume de armazenamento em bloco à instância.

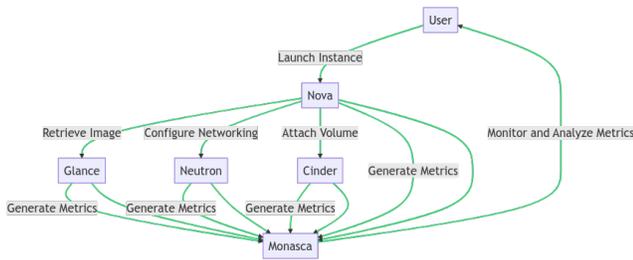


Figura 1: Fluxo de dados entre o Monasca e os componentes do OpenStack

À medida que essas operações ocorrem, são geradas métricas relacionadas ao uso de recursos e ao desempenho. O Monasca coleta essas métricas dos vários componentes, fornecendo monitoramento e diagnóstico em tempo real. Essa comunicação entre componentes e a coleta de métricas são fundamentais para manter a integridade e o desempenho da nuvem OpenStack. Concluindo, a função do Monasca na coleta e no gerenciamento de métricas, combinada com a comunicação entre componentes no OpenStack, fornece uma estrutura robusta para o gerenciamento de recursos em tempo real. Isso se torna ainda mais avançado quando integrado a ferramentas de análise e visualização de dados como o Redash, conforme exploraremos nas seções a seguir. Na **Figura 1** está um diagrama que visualiza o processo de comunicação entre componentes e coleta de métricas:

Neste diagrama:

- O usuário inicia o processo ao iniciar uma instância.
- O componente Nova recupera a imagem necessária do Glance, configura a rede com o Neutron e anexa um volume, se necessário, com o Cinder.
- À medida que essas operações ocorrem, o Nova, o Glance, o Neutron e o Cinder geram métricas relacionadas ao uso de recursos e ao desempenho.
- O Monasca coleta essas métricas, fornecendo monitoramento e diagnóstico em tempo real.
- O usuário pode então monitorar e analisar essas métricas por meio do Monasca.

4 REDASH

4.1 Introdução ao Redash

O Redash surge como uma plataforma poderosa de visualização e exploração de dados que pode potencialmente enfrentar os desafios identificados no Shylock. Ele oferece recursos de visualização de dados em tempo real, permitindo que os operadores compreendam os padrões de uso de recursos com mais clareza e tomem decisões oportunas com base nos dados mais recentes.

Ao integrar o Redash ao Shylock, podemos potencialmente superar os problemas identificados, melhorando a eficiência da gestão de recursos na nuvem LSD. As seções subsequentes deste artigo aprofundaram o processo de integração e avaliaram sua eficácia.

Em conclusão, embora o mecanismo de fluxo de dados do Shylock teoricamente permita um monitoramento eficaz do uso de recursos, sua implementação prática na nuvem LSD tem sido menos eficiente.

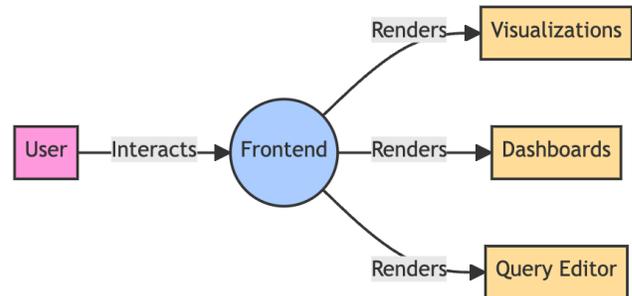


Figura 2: Frontend

A integração do Redash apresenta uma solução promissora para esses desafios, potencialmente melhorando a eficácia da gestão de recursos em nuvens privadas.

4.2 Arquitetura

O Redash é uma solução robusta de análise de dados projetada para ser fácil de usar e flexível, permitindo que os usuários consultem, visualizem e compartilhem dados com facilidade. Sua arquitetura é modular, permitindo extensibilidade e adaptabilidade, o que o torna uma ferramenta avançada para a tomada de decisões orientada por dados, especialmente no contexto do gerenciamento de recursos em tempo real em nuvens privadas. Esta seção apresenta uma visão geral detalhada da arquitetura do Redash.

4.2.1 Overview. O Redash é um aplicativo da Web de página única que consiste em um frontend JavaScript e um backend Python. O front-end é responsável por todas as visualizações, painéis e editor de consultas, que são os componentes com os quais os usuários comuns mais interagem. O back-end, por outro lado, é um aplicativo Flask que usa o Celery Distributed Task Queue como seu mecanismo de Task Worker, com os Workers do Celery responsáveis pela execução da consulta.

4.2.2 Frontend. O front-end do Redash foi originalmente escrito em AngularJS. No entanto, desde a versão 5, ele está fazendo a transição para o React. Essa transição é um indicativo da natureza evolutiva do Redash, que é constantemente atualizado para incorporar as tecnologias mais recentes e as práticas recomendadas de desenvolvimento da Web. O front-end é responsável por criar e gerenciar visualizações e painéis, além de fornecer uma interface amigável para o editor de consultas.

4.2.3 Backend. O backend do Redash é um aplicativo Flask, uma estrutura leve e flexível para desenvolvimento da Web em Python. O backend usa o Celery Distributed Task Queue como seu Task Worker Engine. Os Workers do Celery são responsáveis pela execução

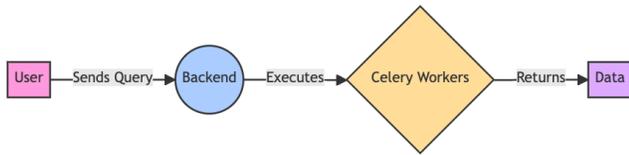


Figura 3: Backend

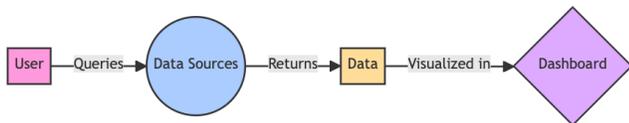


Figura 4: Data Source

das consultas criadas no frontend. Essa separação de preocupações entre o frontend e o backend permite o gerenciamento eficiente de recursos e o dimensionamento, o que é crucial para o gerenciamento de recursos em tempo real em nuvens privadas.

4.2.4 Data Sources. Um dos principais recursos do Redash é a capacidade de consultar todas as fontes de dados do usuário em um único local, sem mover os dados para qualquer lugar. Isso é possível graças à API de fonte de dados extensível do Redash, que tem suporte nativo para uma longa lista de bancos de dados e plataformas comuns. Essa capacidade de combinar consultas e visualizações de diferentes fontes de dados em um único painel oferece uma visão geral abrangente dos dados, facilitando o gerenciamento eficiente e eficaz dos recursos.

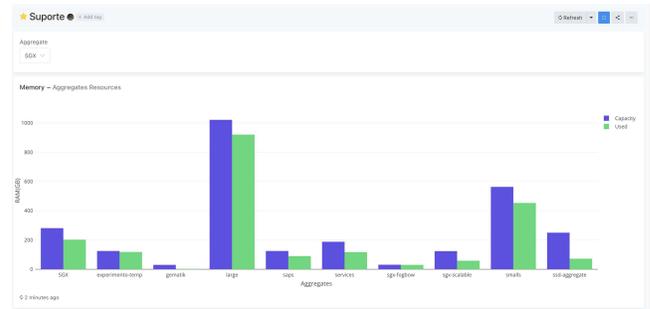


Figura 5: Gráfico em barras com a memória total e memória usada por cada Aggregate

4.3 Monitoramento com o Redash

4.3.1 Aggregate. Um Aggregate é uma entidade que representa um conjunto de computadores físicos agrupados para formar uma unidade coesa com o objetivo de gerenciamento de recursos. Nesse contexto, o gerenciamento eficiente dos recursos é crucial, uma vez que permite a otimização do desempenho e a prevenção de desperdícios e gargalos.

Monitorar o uso de recursos em um Aggregate é essencial para assegurar sua utilização eficiente. Métricas como uso de CPU, memória e armazenamento são vitais para identificar possíveis problemas e otimizar a alocação de recursos. As métricas específicas a serem monitoradas podem variar conforme o tipo de recurso em questão. Por exemplo, no caso de recursos de computação, é importante observar o uso de CPU, memória e armazenamento. Quando se trata de recursos de armazenamento, a utilização de espaço em disco e a taxa de leitura e gravação são métricas relevantes.

O Redash é uma ferramenta extremamente útil para monitorar Aggregates. Ele permite a criação de vários tipos de visualizações, como gráficos de barras e gráficos de linhas, que ajudam a monitorar as métricas de interesse de maneira clara e intuitiva. Por exemplo, gráficos de barras podem ser usados para mostrar a memória total e a memória utilizada de um Aggregate, permitindo que os operadores identifiquem rapidamente se a memória está sendo usada de forma eficiente ou se existem gargalos. As Figuras 5 e 6 mostram exemplos de visualização de métricas dos Aggregates no Redash.

Além disso, o Redash oferece a capacidade de criar dashboards personalizados que reúnem as métricas mais importantes em um único lugar. Isso não só facilita a visualização dos dados, mas também contribui para uma identificação mais ágil de problemas. Um exemplo seria um dashboard que exiba o uso de recursos de todos os Aggregates em uma única tela, permitindo que os operadores detectem rapidamente se algum Aggregate está consumindo recursos excessivamente ou apresentando problemas de desempenho.

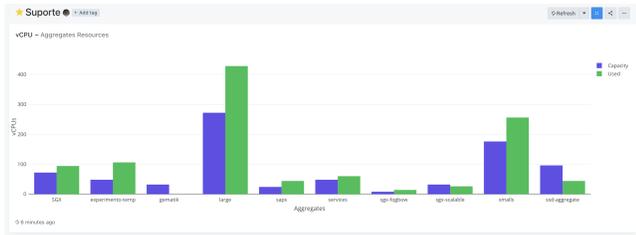


Figura 6: Gráfico em barras com a vCPU total e vCPU usada do Aggregate

5 CONCLUSÃO

Este trabalho buscou enfrentar os desafios da orquestração de recursos em tempo real em nuvens privadas, com ênfase na nuvem privada da UFCG/UASC. A implementação do Redash sobre o LSD conseguiu superar com sucesso as limitações inerentes ao uso do Shylock para governança de recursos. O Redash se tornou uma ferramenta indispensável, especialmente devido à sua habilidade em visualizar dados. Isso permitiu que os administradores da nuvem identificassem facilmente padrões de uso de recursos, refinando assim o processo de tomada de decisões.

É imperativo reconhecer que, apesar dos avanços realizados, o caminho para a perfeição é interminável. A ampliação dos Dashboards para incorporar dados de recursos, que estavam fora do alcance desta pesquisa, é uma preocupação. Esse aprimoramento forneceria uma visão panorâmica do ecossistema, reforçando assim a administração de recursos.

Além disso, a combinação de notificações com os canais de comunicação à disposição dos supervisores de nuvem é um avanço fundamental. Isso garantiria a transmissão imediata de dados vitais por meio de comunicados estruturados, acelerando a compreensão dos dados e as ações consequentes.

Também é significativa a necessidade de refinar a eficácia das consultas SQL no Redash. O ajuste fino dessas consultas é fundamental para evitar a pressão indevida sobre o sistema, o que poderia culminar no desperdício de recursos ou em um desempenho lento.

Ao longo desta busca acadêmica, um tesouro de percepções foi descoberto. Acima de tudo, foi cultivada uma profunda compreensão da arquitetura do OpenStack, de seus componentes e do papel fundamental da visualização de dados. Essa perspicácia é o alicerce para direcionar habilmente os recursos em um ambiente de nuvem privada.

Posteriormente, o estudo esclareceu as complexidades envolvidas na operação de uma nuvem privada. Compreender os detalhes desse ecossistema é vital para identificar e enfrentar os diferentes desafios que podem surgir.

Além disso, o estudo destacou a importância da comunicação convincente com todas as partes envolvidas. Nesse cenário, a LSD assumiu o papel de cliente e, por meio da colaboração com os administradores da infraestrutura de nuvem, obteve-se uma compreensão clara dos problemas em jogo. Isso preparou o caminho para a obtenção de requisitos e a elaboração de uma solução personalizada.

Em resumo, a integração do Redash como instrumento de análise de dados reforçou consideravelmente a governança de recursos em tempo real na nuvem privada da UFCG/UASC. Embora a solução

tenha enfrentado desafios fundamentais, é evidente que o refinamento constante é indispensável para o desempenho máximo. Essa investigação estabelece a base para futuros esforços acadêmicos e inovações, promovendo assim o amadurecimento dos paradigmas de governança de recursos em domínios de nuvem privada.