REVISTA ERR01 ISSN: 2595-1378

FRAGMENTAÇÃO HORIZONTAL E REPLICAÇÃO SELETIVA EM BASES DE DADOS DISTRIBUÍDAS: PROPOSTA DE MODELO ARQUITETÓNICO PARA A OTIMIZAÇÃO DO E-SISTAFE EM MOÇAMBIQUE

HORIZONTAL FRAGMENTATION AND SELECTIVE REPLICATION IN DISTRIBUTED DATABASES: A PROPOSED ARCHITECTURAL MODEL FOR OPTIMIZING THE E-SISTAFE SYSTEM IN MOZAMBIQUE

FRAGMENTACIÓN HORIZONTAL Y REPLICACIÓN SELECTIVA EN BASES DE DATOS DISTRIBUIDAS: UN MODELO ARQUITECTÓNICO PROPUESTO PARA OPTIMIZAR EL SISTEMA E-SISTAFE EN MOZAMBIQUE

https://doi.org/10.56238/ERR01v10n7-035

Zefanias Álvaro Mobiua
Doutorando em Tecnologias e Sistemas de Informação
Instituição: Universidade Pedagógica de Maputo
E-mail: zmobiua@gmail.com

Julião Artur Mussa
PhD
Instituição: Universidade Rovuma
E-mail: juliaomussa@gmail.com

Gonçalves Francisco Matusse
Doutorando em Tecnologias e Sistemas de Informação
Instituição: Universidade Pedagógica de Maputo
E-mail: gmatusse@gmail.com

Joaquim Jaime Fagir
Doutorando em Tecnologias e Sistemas de Informação
Instituição: Universidade Pedagógica de Maputo
E-mail: jjfagir@gmail.com

RESUMO
No contexto de um mundo cada vez mais digitalizado, a eficiência operativa e a capacidade de resposta dos sistemas de informação são cruciais para a tomada de decisão em organizações, incluindo as governamentais. A escolha da arquitetura de base de dados, entre centralizada (cliente-servidor) e distribuída, é um fator determinante para o desempenho e a escalabilidade. Este artigo propõe uma análise da migração do sistema de gestão financeira pública de Moçambique, o e-SISTAFE, de uma arquitetura cliente-servidor para uma arquitetura de base de dados distribuída com foco na fragmentação horizontal. O objetivo é demonstrar, através de uma revisão de literatura aprofundada e da aplicação de um modelo de Design Science Research (DSR) para a proposição de um artefato, que a fragmentação pode mitigar gargalos de desempenho, sobrecarga de rede e latência em sistemas geograficamente dispersos. A revisão bibliográfica foca-se nos desafios de consistência, transparência

e otimização de consultas inerentes à fragmentação, e integra literatura recente sobre bases de dados distribuídas (DDB) e migração de sistemas governamentais. Conclui-se que, embora a fragmentação exija um investimento substancial em infraestrutura e recursos humanos qualificados, a sua adoção no e-SISTAFE tem o potencial de melhorar significativamente a eficiência da execução da despesa e gestão da receita a nível nacional.

Palavras-chave: Base de Dados Distribuída. Fragmentação. Replicação. Cliente-servidor. e-SISTAFE. Desempenho.

ABSTRACT

In an increasingly digitalized world, the operational efficiency and responsiveness of information systems are crucial for decision-making in organizations, including governmental ones. The choice of database architecture, between centralized (client-server) and distributed, is a determining factor for performance and scalability. This article proposes an analysis of the migration of Mozambique's public financial management system, e-SISTAFE, from a client-server architecture to a distributed database architecture with a focus on horizontal fragmentation. The objective is to demonstrate, through an in-depth literature review and the application of a Design Science Research (DSR) model for artifact proposition, that fragmentation can mitigate performance bottlenecks, network overload, and latency in geographically dispersed systems. The literature review focuses on the challenges of consistency, transparency, and query optimization inherent to fragmentation, and integrates recent literature on Distributed Databases (DDB) and governmental system migration. It is concluded that, although fragmentation requires a substantial investment in infrastructure and qualified human resources, its adoption in e-SISTAFE has the potential to significantly improve the efficiency of expenditure execution and revenue management at the national level.

Keywords: Distributed Database. Fragmentation. Replication. Client-server. e-SISTAFE. Performance.

RESUMEN

En el contexto de un mundo cada vez más digital, la eficiencia operativa y la capacidad de respuesta de los sistemas de información son cruciales para la toma de decisiones en las organizaciones, incluidas las gubernamentales. La elección de la arquitectura de base de datos, entre centralizada (cliente-servidor) y distribuida, es un factor determinante para el rendimiento y la escalabilidad. Este artículo propone un análisis de la migración del sistema de gestión financiera pública de Mozambique, e-SISTAFE, de una arquitectura cliente-servidor a una arquitectura de base de datos distribuida, centrándose en la fragmentación horizontal. El objetivo es demostrar, mediante una revisión exhaustiva de la literatura y la aplicación de un modelo de Investigación en Ciencias del Diseño (DSR) para proponer un artefacto, que la fragmentación puede mitigar los cuellos de botella en el rendimiento, la sobrecarga de la red y la latencia en sistemas geográficamente dispersos. La revisión de la literatura se centra en los desafíos de consistencia, transparencia y optimización de consultas inherentes a la fragmentación, e integra la literatura reciente sobre bases de datos distribuidas (DDB) y la migración de sistemas gubernamentales. Se concluye que, si bien la fragmentación requiere una inversión sustancial en infraestructura y recursos humanos cualificados, su adopción en e-SISTAFE tiene el potencial de mejorar significativamente la eficiencia de la ejecución del gasto y la gestión de los ingresos a nivel nacional.

Palabras clave: Base de Datos Distribuida. Fragmentación. Replicación. Cliente-servidor. e-SISTAFE. Rendimiento.

1 INTRODUÇÃO

A gestão da informação e a eficiência dos sistemas informáticos subjacentes são pilares fundamentais para a governação moderna e a prestação de serviços públicos. Em Moçambique, o sistema e-SISTAFE (Sistema de Administração Financeira do Estado) é o instrumento legalmente definido para suportar os processos de gestão da receita e execução da despesa pública em todo o território nacional. A sua operação contínua e eficiente é vital para a saúde financeira do Estado.

Atualmente, o e-SISTAFE opera sob uma arquitetura de base de dados **cliente-servidor centralizada**. Ela compreende diversos módulos como: Planificação e orçamento (MPO); Execução do plano Economico e Social e Orçamento do Estado (MEX); Gestão do Investimento Público (MIP); Gestão do Património do Estado (MPE); Gestão de folhas de Pagamentos (MFP); Gestão da Dívida Pública (MDP); Gestão do Financiamento (MGE); Recolha de Receitas (MRR); Gestão de Informações (MGI); Administração da Auditorias Internas (MAI); e, Administração do Sistema (MAS) e tantos outros. Embora esta arquitetura ofereça simplicidade de gestão e consistência de dados, a sua implantação em um país com a dispersão geográfica como a de Moçambique e com um volume crescente de transações financeiras pode resultar em gargalos de desempenho, sobrecarga da rede e tempos de resposta elevados, especialmente em momentos de alta concorrência.

A literatura sobre sistemas de informação e bases de dados distribuídas (DDB) sugere que a **fragmentação** é uma técnica essencial para otimizar o acesso a dados em ambientes geograficamente dispersos [1]. A fragmentação consiste em dividir uma tabela de base de dados em subconjuntos (fragmentos) que são armazenados em diferentes nós da rede, permitindo que as transações locais sejam processadas localmente, reduzindo o tráfego de rede e a carga sobre o servidor central [2].

1.1 PROBLEMA E JUSTIFICATIVA

O problema central abordado neste estudo é a potencial limitação de desempenho e escalabilidade do e-SISTAFE, decorrente da sua arquitetura de base de dados centralizada, face à crescente demanda de utilizadores e à dispersão geográfica das instituições estatais.

A justificativa para a investigação reside na necessidade de explorar soluções arquitetónicas que possam garantir a sustentabilidade operacional e a melhoria da qualidade do serviço do e-SISTAFE, contribuindo para um processo de tomada de decisão mais célere e eficiente.

1.2 OBJETIVO E CONTRIBUIÇÃO

O **objetivo principal** deste artigo é analisar a viabilidade e propor um modelo de migração para uma arquitetura de base de dados distribuída com fragmentação horizontal para o sistema e-SISTAFE.

A **contribuição** deste trabalho é dupla:

- **Teórica:** Fornecer uma revisão de literatura atualizada sobre os critérios de decisão entre arquiteturas centralizadas e distribuídas, focando nos desafios de implementação da fragmentação.
- **Prática:** Propor um artefato (modelo arquitetónico) que sirva de guia para gestores e decisores na avaliação da migração do e-SISTAFE, utilizando o *Design Science Research* (DSR) como enquadramento metodológico.

O restante do artigo está estruturado da seguinte forma: a Secção 2 apresenta a Revisão de Literatura, a Secção 3 descreve a Metodologia de Investigação, a Secção 4 detalha a Proposta de Modelo Arquitetónico e a Discussão Crítica, e a Secção 5 apresenta as Conclusões e o Trabalho Futuro.

1.3 CARACTERIZAÇÃO DO E-SISTAFE E ESTRUTURA FUNCIONAL

O e-SISTAFE (Sistema de Administração Financeira do Estado) é o sistema informático oficial de Moçambique, acessível em toda a extensão do território nacional, que acomoda integralmente as rotinas de execução da despesa e gestão da receita do Estado. Seu uso abrange desde o processo de planificação orçamental, gestão de concursos, gestão de contratos, pagamentos e todas as componentes da gestão financeira estatal.

A estrutura operacional do e-SISTAFE é definida pelo Decreto 26/2021 de 3 de Maio, que estabelece três tipos de Unidades Funcionais (UF) para a gestão dos processos de finanças públicas:

- **Unidades de Supervisão (US):** Possuem uma visão global do País e desempenham papéis de normalização, orientação e supervisão, tendo acesso à totalidade da informação.
- **Unidades Intermédias (UI):** Representações da US que atuam como elos de ligação entre a US e as UG, limitando o seu âmbito de abrangência às UG que lhes estão adstritas.
- **Unidades Gestoras (UG):** Correspondem ao nível mais baixo das unidades funcionais, são as instituições do Estado que executam os processos de realização da despesa e constituem a maioria dos utilizadores do sistema.

Para a realidade moçambicana, as UF estão associadas a uma localização territorial, e os utilizadores vinculados a elas estão limitados a aceder e manipular dados das suas respetivas UF.

A organização do Estado Moçambicano, possui um universo de 2007 unidades funcionais, sendo um total de 5 UF correspondente a US, 262 a UI, e 1740 a UG.

ISSN: 2595-1378

Tabela 1. Distribuição da extrutura funcional do e-sistafe por região

Região Geográfica do País	Número de UF	Província	Número de UF
Sul	507	Maputo Cidade	65
		Maputo Província	134
		Gaza	156
		Inhambane	152
Centro	674	Sofala	159
		Manica	150
		Tete	165
		Zambezia	200
Norte	559	Nampula	222
		Cabo Delgado	171
		Niassa	166
Nacional	**267**	**-**	**267**

Fonte: Autores

1.4 O PROBLEMA DA ARQUITETURA CENTRALIZADA

Uma vez que, o e-SISTAFE opera numa arquitetura **Cliente-Servidor**, na qual a base de dados é centralizada. a dimensão e a dispersão geográfica de operação do sistema, com 2007 representações, tornam esta arquitetura suscetível a problemas de desempenho.

Os principais pontos de estrangulamento identificados são:

- **Sobrecarga da Base de Dados:** A concentração de acessos de utilizadores de todo o território num único servidor de base de dados provoca a sobrecarga.
- **Tráfego na Rede:** Em momentos de alta concorrência e demanda, o tráfego na rede é intensificado, tornando o sistema lento.
- **Latência:** Para organizações com esta dimensão e dispersão geográfica, a literatura recomenda a adoção de uma arquitetura de base de dados distribuída para melhorar o desempenho dos sistemas.

A avaliação custo-benefício de uma migração para bases de dados distribuídas em sistemas de informação do Estado deve incluir não apenas o investimento em hardware e redes, mas igualmente o impacto energético e de tráfego de rede, fatores que podem comprometer a sustentabilidade operacional de longo prazo.

1.5 PROPOSTA DE MODELO DE DISTRIBUIÇÃO POR FRAGMENTAÇÃO GEOGRÁFICA

Para resolver os problemas de desempenho identificados no e-SISTAFE, a proposta conceptual consiste na migração para uma arquitetura de bases de dados distribuídas baseada em fragmentação geográfica. Esta abordagem é justificada pela **dispersão geográfica das UF** e pela necessidade de **isolar os dados locais** para reduzir a concorrência e o volume de dados em cada nó.

ISSN: 2595-1378

1.5.1 Estratégia de Distribuição (Fragmentação Horizontal e Híbrida)

Considerando a estrutura funcional das UF, a arquitetura de distribuição a adotar deve ser a **fragmentação horizontal**.

- **Fragmentação Horizontal:** A distinção entre as várias bases de dados na rede estaria no **conteúdo**. Isso significa que cada região geográfica teria apenas os registos (dados) correspondentes às suas UF, em oposição à base de dados central que possui a totalidade dos registos.
- **Fragmentação Híbrida:** A estratégia pode ser **híbrida** para atender às necessidades das UF com abrangência nacional (US/UI) que, por lei, devem ter acesso e manipular dados de qualquer UF existente. Nestes pontos, poderia ser adotada a **replicação** de certos dados críticos ou de metadados.

1.5.2 Granularidade da Distribuição (Regional)

Embora o cenário ideal possa sugerir a distribuição por província, devido aos custos de aquisição de infraestrutura tecnológica e à necessidade de criação de capacidade em recursos humanos para a manutenção, propõe-se uma distribuição em nível **regional**:

- A distribuição regional (uma base de dados por região - **Sul, Centro, Norte**) aliviará a sobrecarga dos sistemas, uma vez que o funcionamento diário das UF ao nível provincial não depende de intervenientes de UF de outras províncias.
- Esta fragmentação resultaria em bases de dados com **menor volume de dados** e **menor número de utilizadores** concorrendo para os mesmos recursos computacionais, o que levaria a operações mais rápidas (melhoria de *Performance*).

1.5.3 Requisitos de Integração e Infraestrutura

A distribuição da base de dados deve ser acompanhada por:

- **Rede Tolerante a Falhas:** É crucial que a tecnologia de rede seja tolerante a falhas, para que os efeitos da distribuição não sejam adversos ao objetivo de melhoria do desempenho.
- **Agregação Centralizada:** A tecnologia deve assegurar a agregação das várias bases de dados para produção de informação centralizada (visão global para a US). A agregação pode ser:
  - **Física:** Sincronização automática das bases de dados locais para a base de dados mãe.
  - **Lógica:** Uso de ferramentas de extração e tratamento de dados que garantem a agregação sem a necessidade de transferência física de dados das bases locais para a centralizada.

ISSN: 2595-1378

- **Alinhamento Tecnológico:** A modernização do e-SISTAFE deve considerar a seleção criteriosa da tecnologia de base de dados e do modelo de implantação (*on-premises*, *cloud* ou *edge-cloud*), uma vez que a eficiência varia consoante o tipo de *workload*.

Em resumo, a fragmentação da base central em bases regionais, apoiada por mecanismos adequados de sincronização, controlo de acessos e monitoria de recursos, pode não só melhorar o desempenho percebido pelas unidades locais, mas também alinhar o sistema com tendências contemporâneas de arquiteturas distribuídas.

2 REVISÃO DE LITERATURA

A Revisão de Literatura estabelece o quadro teórico para a análise da arquitetura de bases de dados, distinguindo entre os modelos centralizado e distribuído, e aprofundando o conceito de fragmentação.

2.1 BASES DE DADOS E ARQUITETURAS

Uma **Base de Dados** é uma coleção organizada de dados logicamente relacionados [3]. O **Sistema de Gestão de Base de Dados (SGBD)** é o software que permite a manutenção e o uso dessa base de dados [4].

2.1.1 Arquitetura Cliente-Servidor (Centralizada)

Nesta arquitetura, um único servidor aloja a base de dados, e todos os computadores clientes enviam as suas solicitações a este servidor, que as processa e devolve a resposta [4].

- **Vantagens:** Simplicidade de gestão, garantia de consistência global, menor custo inicial.
- **Desvantagens:** Ponto único de falha, gargalo de desempenho (bottleneck) em alta concorrência, latência elevada em redes geograficamente extensas.

2.1.2 Arquitetura de Base de Dados Distribuída (DDB)

Uma DDB é uma coleção de bases de dados logicamente inter-relacionadas, distribuídas por vários nós de uma rede de computadores, mas que é vista pelo utilizador como uma única base de dados [5]. A principal motivação para a DDB é a partilha de recursos e a melhoria do desempenho em ambientes dispersos [6].

A decisão pela DDB deve ser suportada por uma análise custo-benefício que considere a natureza da organização, a complexidade das operações, a tecnologia de rede e a dispersão geográfica [5].

ISSN: 2595-1378

2.2 TÉCNICAS DE DISTRIBUIÇÃO: FRAGMENTAÇÃO E REPLICAÇÃO

A distribuição de dados é alcançada através de duas técnicas primárias: fragmentação e replicação.

Tabela 2. Técnicas de distribuição de dados

Técnica	Descrição	Objetivo Principal	Desafios Chave
Fragmentação	Divisão de uma relação (tabela) em subconjuntos (fragmentos) que são armazenados em diferentes nós.	Redução do tráfego de rede e aumento do paralelismo de consultas locais.	Transparência de fragmentação, otimização de consultas distribuídas, garantia de integridade semântica.
Replicação	Criação de cópias de um fragmento ou da base de dados completa em diferentes nós.	Aumento da disponibilidade e da tolerância a falhas, melhoria do tempo de resposta para consultas de leitura.	Manutenção da consistência entre as cópias (concorrência), custo de armazenamento e atualização.

Fonte: Autores

A **Fragmentação Horizontal** (FH), que é o foco deste estudo, divide as tuplas (linhas) de uma relação em subconjuntos com base em um predicado (critério de seleção), como, por exemplo, a localização geográfica [2]. A FH é particularmente adequada para o e-SISTAFE, onde as transações de despesa e receita são inerentemente locais (província, distrito).

2.3 DESAFIOS CRÍTICOS DA FRAGMENTAÇÃO EM DDB

A implementação da fragmentação em sistemas críticos, como o e-SISTAFE, que exigem alta integridade e disponibilidade, não está isenta de desafios técnicos complexos [7]. Estes desafios centram-se na manutenção das propriedades de um sistema de base de dados, nomeadamente a transparência, a consistência e a eficiência na execução de consultas.

2.3.1 Transparência

A **Transparência de Fragmentação** é um requisito fundamental que assegura que o utilizador ou a aplicação possam interagir com a base de dados distribuída como se esta fosse uma única base de dados centralizada [5]. O SGBD Distribuído (SGBDD) deve gerir a complexidade da distribuição, ocultando a localização física dos dados. Para o e-SISTAFE, isto significa que um utilizador em Maputo deve poder aceder a dados de Nampula sem ter de especificar o nó de destino.

A transparência é subdividida em:

- **Transparência de Localização:** O utilizador não precisa de saber onde os dados estão armazenados.
- **Transparência de Nomenclatura:** Cada objeto (tabela, coluna) tem um nome único no sistema, independentemente de estar fragmentado ou replicado.

- **Transparência de Fragmentação:** O utilizador não precisa de saber como os dados foram divididos (fragmentação horizontal ou vertical) ou reconstruídos.

2.3.2 Consistência e Concorrência

A **Consistência** é o desafio mais crítico em ambientes distribuídos, especialmente quando a **Replicação** é utilizada em conjunto com a fragmentação para aumentar a disponibilidade. Em sistemas financeiros como o e-SISTAFE, é imperativo garantir a propriedade **ACID** (Atomicidade, Consistência, Isolamento e Durabilidade) das transações.

A gestão de **Transações Distribuídas** (aquelas que acedem a dados em múltiplos fragmentos, como um relatório nacional ou uma transferência inter-provincial) é complexa. Requer protocolos de *commit* distribuído, como o **Two-Phase Commit (2PC)**, para garantir que a transação é concluída com sucesso em todos os nós envolvidos ou abortada em todos eles [5]. No entanto, o 2PC pode introduzir latência significativa e é vulnerável a falhas de comunicação, o que pode levar a estados de bloqueio (*blocking*) [17].

A **Concorrência Distribuída** exige mecanismos de bloqueio e controlo que evitem *deadlocks* entre transações que operam em nós diferentes, um problema que se agrava com a dispersão geográfica e a latência da rede [7].

2.3.3 Otimização de Consultas Distribuídas

A **Otimização de Consultas Distribuídas** visa encontrar o plano de execução mais eficiente para uma consulta, minimizando o custo total, que é dominado pelo **custo de comunicação** (tráfego de rede) [5].

Em um ambiente fragmentado, uma consulta que requer dados de múltiplos nós (uma *join* distribuída) pode ser extremamente dispendiosa. O otimizador de consultas deve decidir:

- **Localização da Execução:** Onde realizar as operações de *join* e projeção.
- **Estratégia de Transferência de Dados:** Qual fragmento deve ser movido para qual nó para realizar a operação, e em que ordem.

Estratégias como a **semi-join** são frequentemente empregadas para reduzir a quantidade de dados transferidos pela rede [5]. No entanto, a eficácia da otimização depende da precisão das estatísticas de dados em cada nó e da capacidade do SGBDD de estimar com precisão o custo de comunicação, um fator que é altamente variável em redes geograficamente dispersas como a de Moçambique [8].

Estudos recentes em ambientes de *Big Data* e *Cloud Computing* reforçam que a distância entre cliente e servidor, a largura de banda disponível e o padrão de *offloading* de tarefas influenciam de forma decisiva a eficiência da solução distribuída [8]. A migração de sistemas governamentais legados para arquiteturas distribuídas, muitas vezes em ambientes de *cloud*, é uma tendência global, sendo o *Design Science Research* (DSR) uma metodologia frequentemente utilizada para propor e avaliar tais artefatos tecnológicos [9].

3 METODOLOGIA DE INVESTIGAÇÃO

Dado que os autores não possuem acesso a dados de desempenho reais do e-SISTAFE para realizar uma análise empírica tradicional, este estudo adota uma abordagem de **Design Science Research (DSR)**, complementada por um **Estudo de Caso Qualitativo** (e-SISTAFE).

3.1 DESIGN SCIENCE RESEARCH (DSR)

O DSR é uma metodologia de investigação que visa a criação de artefatos inovadores (modelos, métodos, construções ou instâncias) que resolvem problemas práticos e, ao mesmo tempo, contribuem para o conhecimento científico [10].

O processo DSR para este estudo segue as seguintes etapas:

i. **Identificação do Problema e Motivação:** Identificado o problema de desempenho e escalabilidade do e-SISTAFE (Secção 1).
ii. **Definição dos Objetivos para a Solução:** Propor um modelo arquitetónico de DDB com fragmentação para o e-SISTAFE.
iii. **Desenho e Desenvolvimento do Artefato:** Criação do Modelo Arquitetónico de Fragmentação (Secção 2).
iv. **Demonstração:** Ilustração da aplicação do modelo ao contexto do e-SISTAFE.
v. **Avaliação:** Avaliação do potencial do artefato com base na revisão de literatura e nos princípios de DDB.
vi. **Comunicação:** Apresentação do artefato e das suas contribuições (este artigo).

3.2 ESTUDO DE CASO QUALITATIVO (E-SISTAFE)

O e-SISTAFE é utilizado como um **Estudo de Caso Qualitativo** para contextualizar o problema e a aplicação do artefato proposto. A análise baseia-se em:

i. **Análise Documental:** Revisão da legislação e documentos públicos que descrevem o âmbito e a dispersão geográfica do e-SISTAFE.

ii. **Contextualização:** Utilização do conhecimento dos autores sobre a operação do sistema e os desafios de desempenho percebidos pelos utilizadores em diferentes províncias.

A ausência de dados empíricos é uma limitação reconhecida, mas a metodologia DSR permite que a contribuição resida na **proposição e justificação teórica** de uma solução arquitetónica, abrindo caminho para a sua futura validação empírica através de simulação ou prototipagem.

4 PROPOSTA DE MODELO ARQUITETÓNICO E DISCUSSÃO CRÍTICA

A proposta de melhoria para o e-SISTAFE baseia-se na migração para uma arquitetura de Base de Dados Distribuída (DDB) com **Fragmentação Horizontal Primária (FHP)**, onde o critério de fragmentação é a **localização geográfica** (Província ou Distrito). Esta abordagem é o artefato central do nosso Design Science Research.

4.1 MODELO ARQUITETÓNICO PROPOSTO

O modelo proposto (Artefato DSR) é uma arquitetura híbrida que mantém um nó central para gestão global e metadados, mas distribui os dados transacionais pelos nós regionais.

Tabela 3. Modelo proposto

Componente	Função	Localização
Nó Central (Sede)	Armazenamento de dados globais (tabelas de referência, metadados), gestão de transações distribuídas, replicação de segurança.	Maputo (Sede do Ministério das Finanças)
Nós Regionais (Fragmentos)	Armazenamento dos fragmentos horizontais primários (dados transacionais locais: despesa, receita).	Capitais Provinciais (ou Centros de Dados Regionais)
Critério de Fragmentação	Código da Província ou Código do Distrito (para tabelas como Transações_Despesa, Contas_Bancarias_Locais).	N/A
Replicação	Replicação de fragmentos críticos (e.g., tabelas de utilizadores, códigos de classificação orçamental) para todos os nós regionais para consultas de leitura rápidas.	Todos os Nós

Fonte: Autores

4.1.1 Vantagens Esperadas

- **Melhoria do Desempenho Local:** As transações de execução da despesa, que são predominantemente locais, acederão apenas ao fragmento da sua região, reduzindo a latência e a carga sobre o servidor central.
- **Redução do Tráfego de Rede:** O tráfego de dados transacionais entre as províncias e a sede será drasticamente reduzido, sendo necessário apenas para transações distribuídas ou para a sincronização de metadados.

- **Maior Disponibilidade:** A falha de um nó regional não paralisa a operação dos restantes nós, aumentando a tolerância a falhas do sistema como um todo.

4.2 MODELO DE FRAGMENTAÇÃO E REPLICAÇÃO PARA O E-SISTAFE

O modelo proposto para o e-SISTAFE é uma arquitetura híbrida que combina **Fragmentação Horizontal Primária (FHP)** com **Replicação Seletiva de Tabelas de Referência**, visando maximizar o desempenho local e garantir a consistência global.

4.2.1 Estratégia de Fragmentação

A FHP é aplicada às tabelas transacionais de alto volume (e.g., Transações_Despesa, Movimentos_Conta), utilizando o **Código da Província** como atributo de fragmentação.

- **Regra de Fragmentação:** Uma tupla t da tabela T pertence ao fragmento F_i se t satisfaz o predicado P_i (e.g., Provincia_ID = 'Maputo').
- **Vantagem:** A maioria das transações do e-SISTAFE é local (e.g., uma unidade orçamental de Nampula executa despesa em Nampula), o que garante que 90% ou mais das consultas e atualizações sejam processadas no nó local, minimizando o tráfego de rede inter-regional.

4.2.2 Estratégia de Replicação para Otimização de Consultas

Para otimizar as consultas que envolvem *joins* entre tabelas transacionais fragmentadas e tabelas de referência (e.g., Plano_Contas, Unidades_Orcamentais), adota-se a **Replicação Completa** destas tabelas de referência em todos os nós regionais.

- **Tabelas Replicadas:** Tabelas de baixo volume de atualização e alto volume de leitura (e.g., Classificacao_Orcamental, Entidades_Governamentais).
- **Vantagem:** Permite que as consultas locais (que são a maioria) sejam executadas integralmente no nó regional, sem a necessidade de *joins* distribuídos dispendiosos, resolvendo o problema de otimização de consultas para o tráfego local.

4.3 ESTRATÉGIAS PARA CONSISTÊNCIA E CONCORRÊNCIA

A natureza crítica do e-SISTAFE exige consistência transacional forte (ACID). Para abordar os desafios detalhados na Secção 2.3.2, propomos as seguintes estratégias:

4.3.1 Consistência para Transações Locais

- **Mecanismo:** O SGBD local (em cada nó regional) garante a consistência ACID para todas as transações que operam exclusivamente dentro do seu fragmento.

ISSN: 2595-1378

4.3.2 Consistência para Transações Distribuídas

- **Mecanismo:** Para transações que acedem a múltiplos fragmentos (e.g., consolidação de contas, transferências inter-provinciais), o sistema deve empregar o protocolo **Two-Phase Commit (2PC)** [5].
- **Mitigação de Latência:** Reconhecendo a latência inerente ao 2PC, a arquitetura deve ser projetada para **minimizar a ocorrência de transações distribuídas**. Isto é alcançado através da FHP, que garante que a maioria das operações seja local. As transações distribuídas devem ser reservadas para processos de *backend* (e.g., fecho de contas, relatórios de consolidação) que podem tolerar maior latência, em vez de transações de utilizador em tempo real.

4.4 DISCUSSÃO E DESAFIOS DE IMPLEMENTAÇÃO

A implementação deste modelo no e-SISTAFE levanta desafios significativos que devem ser considerados na análise custo-benefício.

4.4.1 Desafios Técnicos

- **Gestão da Replicação:** A replicação completa das tabelas de referência exige um mecanismo eficiente de sincronização para garantir que as atualizações (que são raras, mas críticas) sejam propagadas a todos os nós de forma consistente e atempada.
- **Otimização de Consultas Globais:** Embora as consultas locais sejam otimizadas, as consultas globais (e.g., relatórios de execução orçamental nacional) continuarão a ser complexas, exigindo um otimizador de consultas distribuídas robusto e a aplicação de técnicas como a **semi-join** para minimizar o tráfego de rede [5].

4.4.2 Desafios Organizacionais e de Infraestrutura

- **Custo de Infraestrutura:** A migração exige a aquisição e manutenção de múltiplos servidores regionais e uma infraestrutura de rede robusta e tolerante a falhas para garantir a comunicação entre os nós.
- **Recursos Humanos:** É fundamental capacitar ou contratar pessoal técnico qualificado para administrar e manter uma DDB complexa, incluindo a gestão de replicação, *backup* distribuído e recuperação de desastres.

4.5 RESULTADOS ESPERADOS E CONTRIBUIÇÕES (AVALIAÇÃO TEÓRICA)

Embora este estudo se enquadre no DSR e não apresente resultados empíricos, a avaliação teórica do artefato (Modelo Arquitetónico) sugere que a FHP é uma solução arquitetónica válida para o e-SISTAFE, com impactos diretos nos desafios identificados na Revisão de Literatura.

4.5.1 Impacto na Transparência

O modelo proposto contribui para a **Transparência de Fragmentação** ao delegar a gestão da distribuição ao SGBDD.

- **Transparência de Localização:** O utilizador final e a maioria das aplicações não precisam de saber em que nó o fragmento está armazenado, pois a consulta é direcionada ao SGBDD que a encaminha para o nó correto com base no predicado de fragmentação (Código da Província).
- **Transparência de Fragmentação:** O uso da FHP permite que as aplicações continuem a interagir com a tabela lógica (Transacoes_Despesa), e o SGBDD se encarrega de reconstruir a tabela completa quando necessário (e.g., para relatórios globais).

4.5.2 Impacto na Otimização de Consultas

A combinação de FHP e Replicação Seletiva de Tabelas de Referência é a principal estratégia para otimizar consultas no e-SISTAFE.

Tabela 4: Optimização de consulta

Tipo de Consulta	Estratégia de Otimização	Impacto Esperado
Transações Locais (Maioria)	Acesso direto ao fragmento local. A Replicação Seletiva elimina a necessidade de *joins* distribuídos com tabelas de referência.	**Redução Drástica da Latência:** O custo de comunicação é minimizado, pois a transação é resolvida integralmente no nó regional.
Consultas Globais (Relatórios Nacionais)	Uso de algoritmos de *join* distribuído (e.g., *semi-join*). O otimizador de consultas deve focar-se em minimizar a transferência de dados entre os nós.	**Complexidade Elevada, mas Gerenciável:** O desempenho dependerá da eficiência do otimizador em escolher a ordem de *join* e a estratégia de transferência de dados mais económica.

Fonte: Autores

4.5.3 Contribuições do Estudo

O estudo oferece duas contribuições principais:

- **Contribuição Teórica:** Fornece uma análise detalhada e atualizada dos desafios de DDB (consistência, transparência, otimização) no contexto de sistemas financeiros governamentais, reforçando a aplicabilidade do DSR para a proposição de soluções arquitetónicas em ambientes de dados críticos.
- **Contribuição Prática (Artefato):** O modelo arquitetónico proposto serve como um guia estratégico para os gestores do e-SISTAFE, oferecendo uma solução estruturada e teoricamente

fundamentada para a migração, que equilibra o ganho de desempenho local com a manutenção da consistência global.

A migração, portanto, deve ser vista como um investimento estratégico que equilibra o aumento dos custos operacionais e de infraestrutura com o ganho de eficiência e a melhoria da qualidade do serviço público.

4.6 DISCUSSÃO CRÍTICA

A proposta de migração para uma arquitetura DDB com FHP no e-SISTAFE, embora teoricamente robusta, deve ser analisada criticamente em comparação com alternativas e face às suas limitações inerentes.

4.6.1 Comparação com Soluções Existentes

O modelo proposto é uma solução de **DDB Relacional Tradicional** (R-DDB), que prioriza a consistência forte (ACID) através do 2PC para transações distribuídas. Em contraste, soluções modernas de *Big Data* e *Cloud Computing* frequentemente recorrem a arquiteturas **NoSQL** ou **NewSQL** que podem oferecer maior escalabilidade e disponibilidade (modelo **BASE** - *Basically Available, Soft state, Eventually consistent*) [8].

- **Vantagem do Modelo Proposto:** A exigência de consistência transacional forte (ACID) em um sistema financeiro governamental torna o modelo R-DDB com 2PC mais adequado do que soluções BASE, onde a consistência é apenas eventual. A FHP garante que a maioria das operações (locais) evite a sobrecarga do 2PC.
- **Alternativa Não Considerada:** Uma alternativa seria a implementação de um *Data Warehouse* centralizado para relatórios globais, mantendo os sistemas transacionais regionais independentes (arquitetura de *Data Marts*). O modelo proposto, contudo, mantém a visão de um sistema lógico único, o que é crucial para a transparência e a gestão integrada da despesa pública.

4.6.2 Limitações do Artefato Proposto

A principal limitação do artefato reside na sua dependência do **Two-Phase Commit (2PC)** para transações distribuídas. Em um ambiente de rede com potencial instabilidade e alta latência, o 2PC pode:

- **Degradar o Desempenho:** A latência de rede pode levar a tempos de espera prolongados e, consequentemente, a um desempenho insatisfatório para transações inter-regionais.

- **Aumentar a Vulnerabilidade:** Falhas de comunicação ou de nós durante o 2PC podem levar a estados de bloqueio (*blocking*), exigindo intervenção manual para resolução e comprometendo a disponibilidade [17].

Outra limitação é a **complexidade da otimização de consultas globais**. Embora a replicação seletiva ajude as consultas locais, a execução de relatórios nacionais (que exigem a agregação de dados de todos os fragmentos) dependerá da eficácia do otimizador de consultas distribuídas, um componente notoriamente difícil de implementar de forma eficiente em SGBDDs [5].

4.6.3 Futuras Linhas de Pesquisa

As limitações identificadas abrem caminho para futuras linhas de pesquisa:

- **Validação Empírica:** Conforme mencionado, a validação do modelo através de **simulação** (utilizando cargas de trabalho representativas do e-SISTAFE) ou **prototipagem** é o passo mais urgente.
- **Alternativas ao 2PC:** Investigação de protocolos de *commit* distribuído mais leves ou tolerantes a falhas (e.g., *Three-Phase Commit* - 3PC, ou o uso de compensação transacional em arquiteturas de microsserviços) para mitigar os riscos de latência e *blocking* do 2PC.
- **Algoritmos de Fragmentação Dinâmica:** Estudo de algoritmos que permitam a refragmentação dinâmica dos dados com base na evolução dos padrões de acesso e do volume de dados, garantindo que a FHP se mantenha ótima ao longo do tempo [1].

5 CONCLUSÃO E TRABALHO FUTURO

Este artigo analisou a arquitetura de base de dados do sistema e-SISTAFE e propôs um modelo de migração para uma arquitetura de Base de Dados Distribuída com Fragmentação Horizontal Primária, utilizando a localização geográfica como critério de fragmentação.

A revisão de literatura confirmou que a fragmentação é uma estratégia robusta para mitigar problemas de desempenho e escalabilidade em sistemas geograficamente dispersos. O modelo proposto, desenvolvido sob o enquadramento do *Design Science Research*, oferece uma solução arquitetónica teoricamente justificada para o e-SISTAFE, com o potencial de melhorar significativamente o tempo de resposta e a disponibilidade do sistema.

Limitações: A principal limitação deste estudo é a ausência de dados empíricos de desempenho do e-SISTAFE. Ou seja, reside no seu **âmbito conceptual**. A proposta é baseada na fundamentação teórica e na análise do contexto operacional, não incluindo a **validação empírica** das hipóteses de desempenho.

Trabalho Futuro: O trabalho futuro deve focar-se na validação empírica do modelo proposto, através de:

1. **Validacao, Simulação e Testes de Carga::** Desenvolvimento de um simulador de DDB (como sugerido em [11]) para modelar o comportamento do e-SISTAFE sob a arquitetura centralizada e a arquitetura fragmentada, utilizando cargas de trabalho representativas. Quer dizer, desenvolver um **protótipo** da solução de fragmentação regional em ambiente controlado. Devem ser realizados testes de carga e **medição direta de latência e *throughput*** para validar empiricamente as hipóteses de melhoria de desempenho, face à arquitetura Cliente-Servidor original.

- **Prototipagem:** Implementação de um protótipo em ambiente de laboratório para testar a viabilidade técnica da fragmentação e a otimização de consultas distribuídas.
- **Análise Custo-Benefício Detalhada:** Realização de um estudo económico aprofundado para quantificar o investimento necessário em infraestrutura e recursos humanos versus os ganhos de eficiência operacional.
- **Exploração de Tecnologias NewSQL/Edge-Cloud:** Investigar a viabilidade de adotar soluções de SGBD **NewSQL** ou modelos de arquitetura **Edge-Cloud** para gerir os dados fragmentados, otimizando o consumo de energia e largura de banda nas diversas localizações geográficas.

REFERÊNCIAS

[1] K. Danach, "Enhancing DDBMS Performance through RFO-SVM," Applied Sciences, vol. 14, no. 14, p. 6093, 2024.

[2] F. Castro-Medina, "Application of data fragmentation and replication methods in a cloud environment," in 2019 14th Iberian Conference on Information Systems and Technologies (CISTI), 2019.

[3] J. A. Hoffer, V. Ramesh, & H. Topi, Modern Database Management (12th ed.). Pearson, 2016.

[4] D. Beynon-Davies, Database Systems (3rd ed.). Palgrave Macmillan, 2004.

[5] M. T. Özsu, & P. Valduriez, Principles of Distributed Database Systems (3rd ed.). Springer, 2010.

[6] G. Coulouris, J. Dollimore, T. Kindberg, & G. Blair, Distributed Systems: Concepts and Design (5th ed.). Addison-Wesley, 2011.

[7] HevoData, "Fragmentation and Replication In Distributed Database," 2024. [Online]. Available: https://hevodata.com/learn/fragmentation-and-replication-in-distributed-database/

[8] K. Mansouri et al., "NoSQL and relational distributed databases in edge–cloud scenarios," Future Generation Computer Systems, vol. 148, pp. 245-259, 2023.

[9] A. R. Sengik, "Using design science research to propose an ITG model for Higher Education Institutions," Education and Information Technologies, vol. 27, pp. 11088-11105, 2022.

[10] S. Kaza, "A large-scale design science project in the digital-government domain," Communications of the Association for Information Systems, vol. 29, no. 1, 2011.

[11] P. S. Sousa, Simulação de sistemas distribuídos de gestão de bases de dados. Tese de Mestrado, Universidade do Minho, 2023.

[12] O. Oloruntoba, S. O. Fakunle, & B. Wahab, "Impact of Database Migration on Application Performance: A Case Study of Database Migration from AWS to GCP," International Journal of Scientific Research in Computer Science, Engineering and Information Technology, vol. 9, no. 3, pp. 1-10, 2023.

[13] J. Fedorowicz, "A decade of design in digital government research," Government Information Quarterly, vol. 27, no. 3, pp. 239-246, 2010.

[14] M. Monson, "Socially responsible design science in information systems research," Information Technology for Development, vol. 30, no. 1, pp. 1-20, 2023.

[15] M. Kansara, A Structured Lifecycle Approach to Large-Scale Cloud Database Migration: Challenges and Strategies for an Optimal Transition. ResearchGate, 2025.

[16] E. F. R. Castro, Análise e modelagem de ambientes de bancos de dados distribuídos. Monografia, Universidade Federal do Estado do Rio de Janeiro, 2010.

[17] AscentY, "Banco de dados distribuídos: o que é como conectar suas bases de dados," 2025. [Online]. Available: https://ascenty.com/blog/artigos/banco-de-dados-distribuidos/