

# 1 Introdução

Os sistemas estão cada vez mais ubíquos, produzindo grandes quantidades de dados e com um maior consumo destes. Com isso, tais sistemas possuem uma demanda por maior desempenho do Sistema de Banco de Dados (SBD), ou seja, menor tempo de resposta e maior vazão.

## 1.1. Motivação

Com o avanço das tecnologias, as cargas de trabalho que são submetidas aos SBDs (Elmasri, 2011) (Garcia-Molina et al, 2001) e as execuções das máquinas de consultas presentes nesses sistemas estão cada vez mais complexas. Além disso, as aplicações e os usuários que acessam as bases de dados estão demandando cada vez mais que os SBDs respondam às suas requisições no menor tempo de resposta possível e que tenham uma maior vazão. Entende-se por vazão o número de transações que são finalizadas por unidade de tempo (Shasha et al, 2003). Para atingir essa necessidade por menor tempo de resposta e maior vazão, Administradores de Banco de Dados (DBAs) utilizam estratégias de sintonia fina sobre o projeto físico do banco de dados (Shasha et al, 2003) (Bruno, 2011).

Realizar a sintonia fina de um banco de dados não é uma tarefa trivial. A sintonia fina pode envolver diversas estratégias relacionadas ao projeto físico, por exemplo, através de manutenções de índices, visões materializadas, particionamentos horizontais ou verticais de tabelas, replicação de dados, reescrita de consultas, desnormalização de tabelas, entre outras.

Quando um DBA é contratado para realizar a sintonia fina da base de dados de uma empresa, ele apresenta uma solução de acordo com o comportamento do sistema e da base de dados. Essa solução pode ser, por exemplo, um conjunto de índices recomendados e/ou uma estratégia de particionamento de tabelas. Dada essa solução, na maioria das vezes, ela não vem acompanhada de razões que levaram o DBA a escolhê-la entre outras alternativas de solução disponíveis, bem

como o porquê dessas outras alternativas terem sido descartadas. Dessa forma, a empresa não tem como saber se a solução indicada pelo DBA não poderia ter sido melhor. Em um trabalho de sintonia fina eficaz, não é suficiente que a solução apresente uma pequena melhoria para o desempenho da base de dados, mas sim que ela seja a melhor possível entre as alternativas de solução disponíveis dentro dos requisitos do ambiente em que se encontra a base de dados. Por exemplo, imagine um *site* de vendas. É normal que os usuários submetam consultas sobre a base de dados, buscando por produtos através do seu nome. Além disso, os usuários costumam ordenar o resultado por menor preço. Suponha que esse tipo de consulta está com um tempo de resposta muito longo, gerando a insatisfação do usuário. A empresa responsável pelo *site* decide solicitar uma consultoria ao DBA para agilizar o tempo de resposta. O DBA contratado utiliza uma ferramenta de auto-sintonia e indica a criação de um índice sobre uma das tabelas existentes. Diante dessa sugestão, a empresa pode ficar insegura, ou seja, sem saber se o DBA realmente analisou as possíveis consequências dessa decisão, se ele realmente avaliou todas as possibilidades de solução e se a solução indicada realmente é a melhor. Então, diante de uma solução tão simples, que é a criação de um índice, a empresa questiona o DBA para saber em quais alternativas de solução ele pensou para a base de dados do site, considerando a carga de trabalho que é submetida à base. O DBA analisa o *log* da ferramenta de sintonia e tenta identificar todos os índices pensados pela ferramenta, por comando da carga de trabalho, e os motivos pelos quais alguns deles foram descartados.

Para cada comando da carga de trabalho capturado pela ferramenta, o DBA analisa os índices que foram pensados e tenta inferir uma justificativa para os índices que foram descartados.

Assim, após um período de análises, o DBA elabora o relatório do seu trabalho e entrega ao responsável da empresa como uma forma de justificar a eficácia do seu trabalho.

Por outro lado, já do ponto de vista do DBA, ele pode ficar inseguro com a configuração sugerida por uma determinada ferramenta de auto-sintonia e querer verificar o comportamento de outra ferramenta sobre a mesma carga de trabalho. Nesse caso, se ele usasse duas ferramentas diferentes, ele teria de repetir os passos de análise sobre a outra ferramenta para complementar o relatório.

Adicionalmente, o DBA precisa entender e conhecer os termos usados por cada ferramenta em seu registro de *log* ou metadados para poder buscar as informações necessárias a fim de criar o seu relatório de justificativas.

Dessa forma, o DBA não possui uma ferramenta integrada que o auxilie com argumentos suficientes para justificar a sua decisão. Faz parte da justificativa de um DBA, demonstrar todas ou, se inviável, a maioria das soluções disponíveis e avaliadas para a melhoria de desempenho da base de dados.

As ferramentas que o DBA possui para auxiliá-lo na tarefa de sintonia fina são destinadas a sugerir configurações para a base de dados de acordo com a sua carga de trabalho. Essas ferramentas podem ser automáticas ou semi-automáticas. São consideradas automáticas aquelas que tomam as suas decisões e as aplicam na base de dados, sem qualquer interferência humana. Já as ferramentas semi-automáticas são aquelas que possuem algum tipo de interação com o usuário especialista ou não. Essa interação pode ocorrer, por exemplo, durante a análise da carga de trabalho. As ferramentas de auxílio à sintonia fina são baseadas em estimativas e suposições. É com base nessas suposições que as ferramentas julgam se determinada configuração é ou não a solução que traz melhor desempenho para a base de dados. Para elaborar essas supostas configurações na base de dados, não existe uma fórmula ou algoritmo único, ou seja, são usadas heurísticas para gerar valores aproximados e maneiras de configurar as estruturas auxiliares disponíveis em um SBD.

Por serem baseadas em estimativas, os cálculos, funções de custos e heurísticas usados podem gerar discordâncias e conflitos com atitudes e opiniões de DBAs experientes. Tais discordâncias podem ser originadas dos seguintes cenários:

- O DBA não concorda com a função de custo utilizada pela heurística da ferramenta. Nesse caso, deve haver alguma flexibilidade para que o DBA possa alternar sobre heurísticas disponíveis para estimativas desses custos;
- Há informações adicionais que o DBA conhece e que poderiam passar a ser consideradas pela ferramenta, gerando novas práticas de sintonia fina. Por exemplo: o conhecimento prévio de que um determinado comando gera um custo alto para a base de dados.

Outro cenário motivador para o desenvolvimento da presente tese vem do fato de muitos DBAs deixarem de usar as ferramentas de apoio à sintonia fina pelo simples motivo de não conhecerem o raciocínio que está por trás das decisões e/ou sugestões de configurações para a base de dados, quando a ferramenta não gera um relatório capaz de ser interpretado pelo DBA. Conhecendo o raciocínio (regras) implementado pela ferramenta e as razões pelas quais as decisões foram tomadas, o DBA sente-se mais confiante em usar a ferramenta, e se for o caso, alterá-la no caso de discordâncias.

## **1.2. Caracterização do Problema**

Embora existam diversas ferramentas no mercado sendo desenvolvidas para automatizar a maioria das tarefas relacionadas à sintonia fina de banco de dados, ainda há resistência de alguns DBAs e usuários em aceitar a solução de configuração sugerida pela ferramenta. Esse problema não ocorre apenas com as ferramentas de auto-sintonia, mas também com os profissionais da área de banco de dados. Ao ser contratado por uma empresa, é interessante que o DBA comprove, de alguma forma, que o seu trabalho foi o melhor possível. Dessa forma, tanto o DBA quanto a empresa que o contratou terão mais segurança de que a solução adotada foi derivada de uma análise de um conjunto de alternativas, com suas devidas justificativas de descartes, aumentando a confiança sobre o trabalho do DBA ou da ação da ferramenta de auto-sintonia. Uma das maneiras de comprovar a eficácia desse tipo de trabalho é através da documentação de todas as alternativas testadas no ambiente e a justificativa pela qual ele optou pela solução implementada. No entanto, isso adicionaria um custo e tempo considerável ao trabalho de sintonia fina.

Uma alternativa de solução seria adicionar explicações às ferramentas de auto-sintonia. Mas, para isso, surge a necessidade de que o DBA tenha o conhecimento prévio do algoritmo implementado. Tendo o conhecimento, o DBA pode discordar de alguma regra usada, possuir algum argumento novo ou alguma nova alternativa de solução que o algoritmo não contempla. No caso de discordâncias, o DBA vai precisar alterar o algoritmo implementado com os

ajustes desejados, sem conseguir ter uma visão global do impacto de sua alteração.

Diante disso, esta tese aborda o seguinte problema de pesquisa:

Seja uma carga de trabalho  $W$  (conjunto de comandos DML - *Data Manipulation Language*), um conjunto de características  $F$  (objetos e parâmetros do SBD), um espaço físico de armazenamento disponível  $D$ . Supondo como pré-condição para a análise do problema não ser possível alterar o *hardware* (por exemplo, memória RAM ou secundária) disponível:

*Como se pode **comprovar a eficácia** e aumentar a confiança sobre o trabalho de sintonia fina, considerando  $F$  de forma que se reduza o tempo de resposta (ou aumente a vazão) para  $W$  enquanto se usa um espaço em disco que não ultrapasse  $D$ ?*

### 1.3. Objetivo e Escopo da Tese

O objetivo da presente tese não é mudar as soluções propostas por heurísticas já existentes, mas sim melhorar a eficácia da prática de sintonia fina. Busca-se prover mais instrumentos para o DBA tomar a sua decisão entre as alternativas disponíveis, de uma forma melhor fundamentada e aumentar a confiança em relação ao uso de ferramentas automáticas.

Uma forma do DBA humano aceitar uma proposta de solução de sintonia fina automática é adicionando explicações à ferramenta. No entanto, isso requer um conhecimento prévio do código fonte da mesma e se torna uma tarefa complexa por não se ter uma visão consolidada sobre o impacto global de alterações no código fonte. Sendo assim, surge a necessidade de uma ferramenta que cumpra os seguintes requisitos:

- ✓ Explicar todas as decisões que são tomadas pela ferramenta de sintonia fina;
- ✓ Adquirir um retorno (*feedback*) do DBA especialista em sintonia fina para saber se o mesmo concordou ou discordou da decisão

sugerida ou tomada pela ferramenta de sintonia fina e sua justificativa;

- ✓ Adicionar semântica ao raciocínio provido pela ferramenta;
- ✓ Apresentar os motivos de descartes de alternativas de sintonia fina;
- ✓ Possibilitar a extensão da ferramenta com novos conhecimentos do DBA especialista sobre a área de sintonia fina (novas técnicas) ou até mesmo sobre o domínio em que o banco de dados está situado.

Diante dos requisitos levantados, surge a proposta dessa tese em desenvolver um *framework*, chamado de outer-tuning em uma analogia com o *outer-join*, onde as tuplas de uma tabela que não correspondem com as tuplas da outra tabela, que se está realizando uma junção, são apresentadas no resultado final, as alternativas de sintonia fina não utilizadas na solução final, também são apresentadas, bem como as justificativas do seu não uso.

Como são necessárias explicações sobre o que está acontecendo, é necessário agregar semântica ao raciocínio das heurísticas que propõem configurações dos elementos do banco de dados, ou seja, o que esses elementos fazem e como eles fazem. A ontologia de tarefa descreve conceitos e relacionamentos entre esses conceitos com o objetivo de atingir determinada tarefa (Guarino, 1998). Sendo assim, é necessária uma ontologia de tarefa que formalize e represente o conhecimento sobre as heurísticas. A ontologia de tarefa permite descrever os conceitos e relacionamentos dos elementos do banco de dados de forma a atingir o objetivo da tarefa de executar heurísticas de sintonia fina que fazem uso desses elementos. Além disso, para poder raciocinar em cima das configurações propostas pelas heurísticas, é necessário descrever o domínio de sintonia fina, ou seja, os elementos do banco de dados, que nessa tese, serão descritos em OWL (Web Ontology Language), estabelecendo uma ontologia de domínio. Com o apoio de uma ontologia de aplicação (ontologia de tarefa e ontologia de domínio), o *framework* proposto pela presente tese tem como principais objetivos:

- ✓ Prover argumentos e justificativas através do uso de consultas sobre as instâncias da ontologia. Dessa forma, a solução fica mais bem fundamentada, ajudando na comprovação da eficácia do trabalho de sintonia fina;

- ✓ Combinar heurísticas através do uso de uma máquina de regras, buscando ampliar o escopo de solução de sintonia fina bem como gerar novas práticas de sintonia fina.

Com a semântica das decisões explícitas, o DBA não só tem como justificar suas escolhas como também usar as regras dos raciocínios usados pela ferramenta para agregar ao seu conhecimento e tomar decisões diferentes das sugeridas.

É importante ressaltar que esse trabalho de pesquisa lida com o escopo de melhoria de desempenho do SBD, no sentido de diminuir o tempo de resposta e aumentar a vazão da sua carga de trabalho, considerando apenas os elementos disponibilizados pelo SBD, ou seja, uso de estruturas auxiliares e configurações de parâmetros, sem envolver melhorias nos itens de *hardware*. A intenção é esgotar as possibilidades de melhorias de desempenho do SBD, utilizando apenas o *hardware* disponível, ou seja, sem alternativa de aumento de memória RAM, por exemplo.

Os resultados esperados desta pesquisa são:

- Uma ontologia que represente o domínio de sintonia fina de banco de dados que relacione os conceitos envolvidos na técnica de índices, com a finalidade de possibilitar a justificativa da sintonia fina de banco de dados e flexibilizar, em alto nível, as alternativas a serem aplicadas no raciocínio de sintonia. Embora a ontologia também esteja prevendo as técnicas de visões materializadas e particionamento, o foco dessa tese será sobre a estrutura de índices de banco de dados;
- Um metamodelo do *framework* para execução de heurísticas de sintonia fina de banco de dados. Ele deve ser capaz de ser estendido ou parametrizado de maneira a ser aplicado em cenários de uso por um DBA humano, uma ferramenta semi-automática e/ou automática. Esse metamodelo deve ter a capacidade de explicitar as regras aplicadas por cada heurística que resultaram em determinadas ações na base de dados, de forma a **comprovar a eficácia** do raciocínio implementado e **justificar** a solução final escolhida.

#### 1.4. Principais Contribuições

A principal contribuição dessa tese é proporcionar um avanço na área de sintonia fina de banco de dados através do uso de um *framework* que apoie o DBA no registro de seu raciocínio com um mínimo de impacto. Esse registro de ações e raciocínios sendo realizado, com semântica associada, proporciona uma forma de justificar as suas decisões (incluindo descartes de soluções). Além disso, também proporciona uma maneira de alterar o raciocínio da heurística no momento das análises para a sintonia fina de banco de dados. Essa alteração pode ser, por exemplo, fazer com que ela passe a considerar novas regras e novas práticas de soluções. Usando o *framework* proposto, o DBA conseguiria, através do uso da ontologia instanciada, informações para comprovar a eficácia do trabalho de sintonia fina, demonstrando que qualquer outra solução analisada teria um ganho de desempenho inferior à solução indicada. Além disso, uma ferramenta de auto-sintonia que utilize a ontologia e o *framework* proposto também se beneficia no ganho de credibilidade, pois consegue apresentar de maneira explícita o seu raciocínio e explicá-lo de forma automática.

Uma segunda contribuição importante da presente tese é a definição de uma ontologia de tarefa para a execução de heurísticas de sintonia fina, bem como de uma ontologia do domínio de sintonia fina de banco de dados. É válido ressaltar que a ontologia é uma consequência do objetivo dessa tese. Para que o *framework* seja genérico o suficiente e proporcione semântica à tarefa de sintonia fina faz-se necessária a definição da ontologia.

A seguir, são citadas as contribuições adicionais dessa tese:

- Extensão do domínio e metamodelo do *framework*, acrescentando novos conhecimentos à ontologia de tarefa;
- Extração de informações úteis, do domínio de sintonia fina, para serem usadas como argumentos e justificativas;
- Forma sistemática de gerar novas práticas de sintonia fina. Essas novas práticas podem ser derivadas de combinações de heurísticas existentes ou novas regras e conceitos que venham a surgir no futuro;

- Possibilitar a comparação de heurísticas por meio de um vocabulário comum;
- Possibilitar uma análise mais detalhada sobre o raciocínio de heurísticas de sintonia fina.

## 1.5. Estrutura da Tese

Essa tese encontra-se organizada da seguinte forma: no capítulo 2 são apresentados os conceitos básicos necessários ao desenvolvimento e entendimento do restante deste trabalho. Entre os conceitos descritos neste capítulo tem-se a sintonia fina de banco de dados, a sintonia automática de banco de dados e a ontologia.

No capítulo 3 descreve-se a ontologia de aplicação que foi criada, composta por um modelo de representação de conhecimento para o domínio de sintonia fina e um metamodelo do *framework* para a execução de heurísticas de sintonia fina de banco de dados. Apresenta-se o vocabulário da ontologia e suas propriedades.

O metamodelo é genérico de forma a proporcionar combinações de heurísticas diferentes e também extensível, no caso de necessidade.

O capítulo 4 detalha como o *framework* pode ser usado como apoio para justificar e comprovar a eficácia de um trabalho de sintonia fina. Além disso, demonstram-se como as regras de heurísticas diferentes podem ser combinadas, fazendo uso de uma sistemática para isso. Através de combinações de regras, o DBA pode chegar a novas práticas de sintonia fina.

O capítulo 5 discute os trabalhos relacionados, destacando os diferenciais destes para a abordagem proposta e explicitando a contribuição inovadora proporcionada por este trabalho de tese.

O capítulo 6 conclui este trabalho, indicando suas principais contribuições e sugestões de trabalhos futuros.