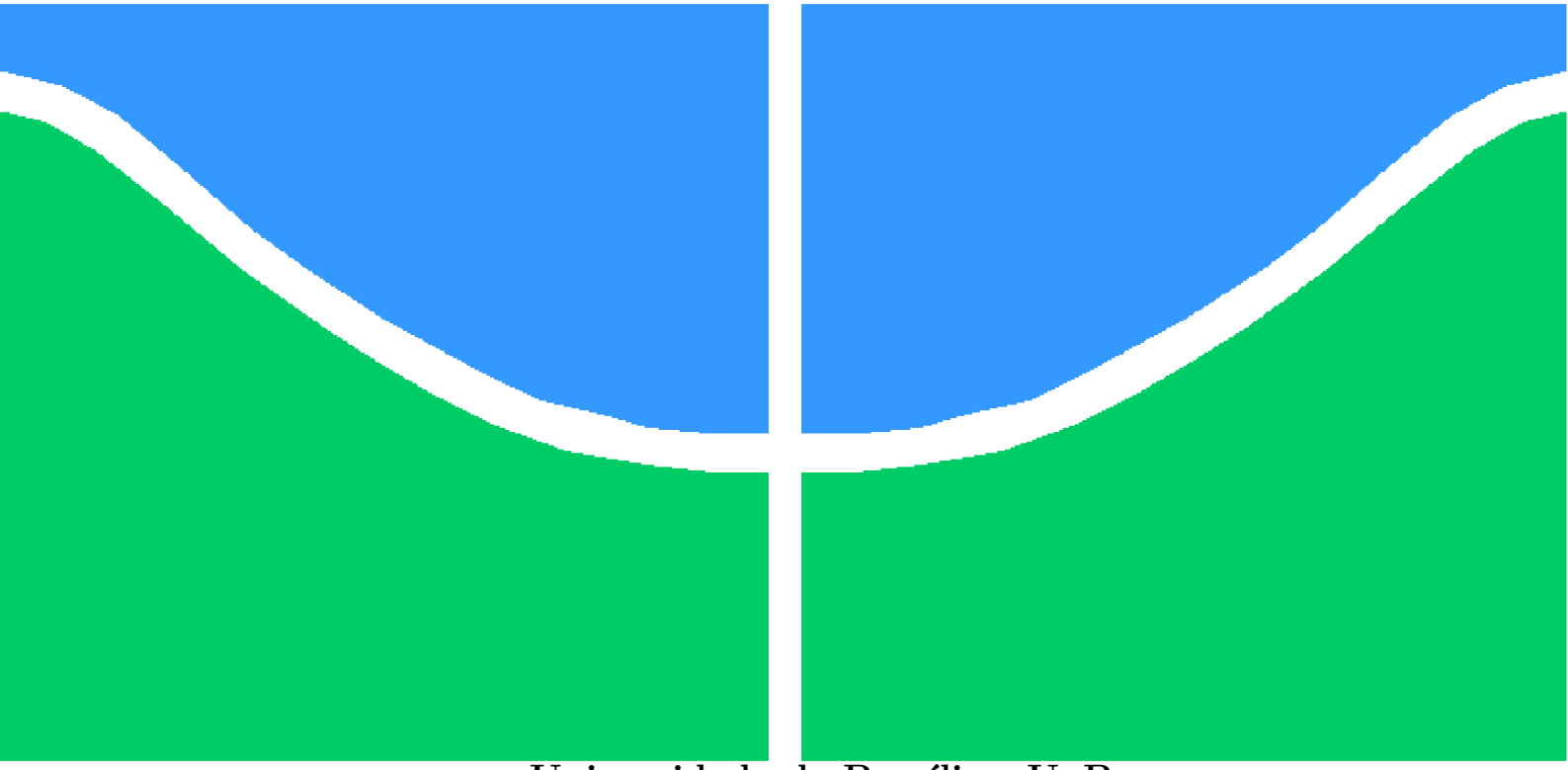

Universidade de Brasília - UnB

Faculdade UnB Gama - FGA

Engenharia de Software

Estudo da Eficácia e Eficiência do Uso de um Ambiente de Data *Warehousing* para Aferição da Qualidade Interna de *Software*: Um Estudo de Caso em uma Autarquia Pública

Autor: Nilton Cesar Campos Araruna

Orientador: Prof. Msc. Hilmer Rodrigues Neri

Brasília, DF

2014

Nilton Cesar Campos Araruna

Estudo da Eficácia e Eficiência do Uso de um Ambiente de Data *Warehousing* para Aferição da Qualidade Interna de *Software*: Um Estudo de Caso em uma Autarquia Pública

Monografia submetida ao curso de graduação em Engenharia de Software da Universidade de Brasília, como requisito parcial para obtenção do Título de Bacharel em Engenharia de Software.

Universidade de Brasília - UnB

Faculdade UnB Gama - FGA

Orientador: Prof. Msc. Hilmer Rodrigues Neri

Brasília, DF

2014

Nilton Cesar Campos Araruna

Estudo da Eficácia e Eficiência do Uso de um Ambiente de Data *Warehousing* para Aferição da Qualidade Interna de *Software*: Um Estudo de Caso em uma Autarquia Pública/ Nilton Cesar Campos Araruna. – Brasília, DF, 2014-

63 p. : il. (algumas color.) ; 30 cm.

Orientador: Prof. Msc. Hilmer Rodrigues Neri

Trabalho de Conclusão de Curso – Universidade de Brasília - UnB
Faculdade UnB Gama - FGA , 2014.

1. Métricas de Código-Fonte. 2. *Data Warehousing*. I. Prof. Msc. Hilmer Rodrigues Neri. II. Universidade de Brasília. III. Faculdade UnB Gama. IV. Estudo da Eficácia e Eficiência do Uso de um Ambiente de Data *Warehousing* para Aferição da Qualidade Interna de *Software*: Um Estudo de Caso em uma Autarquia Pública

CDU a obter

Nilton Cesar Campos Araruna

Estudo da Eficácia e Eficiência do Uso de um Ambiente de Data *Warehousing* para Aferição da Qualidade Interna de *Software*: Um Estudo de Caso em uma Autarquia Pública

Monografia submetida ao curso de graduação em Engenharia de Software da Universidade de Brasília, como requisito parcial para obtenção do Título de Bacharel em Engenharia de Software.

Trabalho aprovado. Brasília, DF, a obter:

Prof. Msc. Hilmer Rodrigues Neri
Orientador

---

a obter
Convidado 1

---

a obter
Convidado 2

Brasília, DF
2014

Este trabalho é dedicado ao meu pai, Aurenilton Araruna, minha mãe, Cláudia Araruna e ao meu irmão, Gustavo Campos. Estas inseriram na minha vida as virtudes do esforço.

Agradecimentos

Agradeço primeiramente aos meus pais, Aurenilton e Claudia, por sempre acreditarem em mim e no meu potencial, por me darem apoio, força e amor sempre que necessário, por me proporcionar uma vida excelente cheia de conforto e por não medirem esforços na hora de me amparar. Obrigado por me ensinarem os grandes valores da vida, do valor do esforço, do tempo e do dinheiro. Vocês sempre serão meus heróis e minhas grandes referências.

Agradeço ao meu irmão, Gustavo Henrique, que sempre me mostrou o que é ser um homem intelectual e de caráter, mesmo com tamanhas diferenças entre nós, sempre me serviu como modelo. Se não fossem suas conquistas não acreditaria que as minhas seriam possível.

Agradeço a minha namorada, Giovana Giugliani, que esteve ao meu lado em toda essa fase da minha vida, do cursinho até a graduação, vivenciando todo meu esforço, minhas dúvidas e e minha ansiedade. Agradeço pelo apoio, pela paciência, pela ajuda, pela companhia em noites de estudo e as tornando bem melhor, também peço desculpa pelo estresse e pelo tempo que tive que me ausentar do seu lado.

Agradeço ao meu orientador Hilmer Rodrigues Neri pela confiança depositada em mim, pelo aprendizado dentro e fora de sala e por ter me motivado ao longo da graduação. Agradeço aos meus amigos e colegas Matheus e Pedro Tomioka pela colaboração e ajuda não só durante a execução desse trabalho mas também em toda a graduação.

Agradeço aos meus amigos de infância, vulgo frangos, por entenderem a minha ausência durante essa fase da minha vida e por sempre desejarem a minha graduação e um belo futuro profissional para poder pagar a conta do bar.

Agradeço a equipe PEDeS da CAIXA, onde tive o prazer de estagiar, em especial ao meu supervisor, Diego Costa, pela oportunidade, pelo aprendizado que levarei pelo resto da vida, por me prepararem para o mercado de trabalho, por me ensinarem tudo o que sei sobre ser um profissional e pelos momentos de alegria. Sei que aprendi com os melhores.

Agradeço ao grupo kanburn, Thiago Kairala, Bruno Contessoto, Rafael Fazzolino, Eduardo Brasil, Thabata Granja e em especial ao meu primo João Araruna que ingressou na universidade junto comigo e participou da minha luta, pelos encontros regados de estudos, pizza e alegria. Os melhores e mais chatos amigos que não escolhi viver mas fui obrigado e hoje a obrigação é manter cada um na minha vida.

"Learn from yesterday, live for today, hope for tomorrow. The important thing is not to stop questioning."

(Albert Einstein)

Resumo

A qualidade do software depende da qualidade do código-fonte, um bom código-fonte é um bom indicador de qualidade interna do produto de *software* (ISO/IEC 25023, 2011). Portanto o monitoramento de métricas de código-fonte de um *software* significa melhorar a sua qualidade. Existem diversas soluções e ferramentas para se obter um monitoramento de métricas de *software* e que conseguem extrair valores de métricas de código com facilidade. Porém a decisão sobre o que fazer com os dados extraídos ainda esbarram na dificuldade relacionada à visibilidade e interpretação dos dados. Este trabalho se propõe a analisar a eficácia e eficiência do uso de um ambiente de *Data Warehousing* para facilitar a interpretação, visibilidade e avaliação das métricas de código-fonte, associando-as a cenários de limpeza. Os cenários de limpeza buscam apoiar as tomadas de decisão que reflitam na alteração do código-fonte e um *software* com mais qualidade. Para um melhor entendimento da solução DW proposta e dos elementos que dizem respeito a sua arquitetura e seus requisitos de negócio foram apresentas as fundamentações teóricas necessárias. Com o objetivo de preparar uma pesquisa sobre sua eficácia e eficiência, foi elaborado um projeto para a realização de uma investigação empírica através da técnica do estudo de caso, que visa responder questões qualitativas e quantitativas a respeito da eficácia e eficiência na utilização da solução citada em um órgão público.

Palavras-chaves: Métricas de Código-Fonte. *Data Warehousing*. *Data Warehouse*

Abstract

Key-words: Source Code Metrics, Data Warehousing, Data Warehouse

Lista de ilustrações

*

Lista de tabelas

*

Lista de abreviaturas e siglas

ACC	*Afferent Connections per Class*
ACCM	*Average Cyclomatic Complexity per Method*
AMLOC	*Average Method Lines of Code*
ANPM	*Average Number of Parameters per Method*
CBO	*Coupling Between Objects*
CSV	*Comma-Separated Values*
DER	Diagrama Entidade Relacionamento
DIT	*Depth of Inheritance Tree*
DW	*Data Warehouse*
ETL	*Extraction-Transformation-Load*
FTP	*File Transfer Protocol*
GQM	*Goal-Question-Metric*
IEC	*International Electrotechnical Commission*
ISO	*International Organization for Standardization*
JSON	*JavaScript Object Notation*
LCOM4	*Lack of Cohesion in Methods*
LOC	*Lines of Code*
NPA	*Number of Public Attributes*
NOC	*Number of Children*
NOM	*Number of Methods*
OLAP	*On-Line Analytical Processing*
OLTP	*Online Transaction Processing*
RFC	*Response For a Class*

SCAM	*IEEE International Working Conference on Source Code Analysis and Manipulation*
SGBD	Sistema de Gerenciamento de Bancos de Dados
SICG	Sistema Integrado de Conhecimento e Gestão
XML	*Extensible Markup Language*
YAML	*YAML Ain't Markup Language*

Sumário

*

1 Introdução

1.1 Contexto

Segundo(WILLCOCKS; LACITY, 2001) a terceirização é uma prática cada vez mais adotada por organizações. A terceirização de atividades, ou seja, o ato de transferir para fora da organização uma parte do seu processo produtivo, não é uma prática recente, desde há muitos anos que as atividades e processos que eram muito específicos dentro de uma organização foram transferidos, de uma forma parcial ou total, para outras organizações ou agentes externo (LEITE, 1997).

Uma das motivações para a terceirização é a qualidade do serviço prometida pelas empresas fornecedoras. No entanto, existem vários riscos associados à decisão pela terceirização, que podem comprometer a qualidade esperada, como: as expectativas de serviço e a resposta rápida não serem atendidas adequadamente; o serviço prestado apresentar qualidade inferior ao existente anteriormente; e as tecnologias utilizadas não corresponderem ao esperado(WILLCOCKS; LACITY, 2001). Segundo o (SOFTEX, 2013a) a aquisição de um *software* é um processo complexo, principalmente no que diz respeito à caracterização dos requisitos necessários ao *software* e às condições envolvidas na contratação como a qualidade esperada.

Acompanhando o ritmo da terceirização o cenário com empresas terceirizadas contratadas para o desenvolvimento de *software* está cada vez maior em organizações públicas. Tais organizações não são diretamente responsáveis pelo desenvolvimento do *software* mas são responsáveis pelo processo de verificação de sua qualidade conforme a norma Brasil (2014) que será explicada no capítulo 3 deste trabalho.

1.2 Justificativa

Segundo (BECK, 1999)(FOWLER, 1999) a qualidade de *software* é medida pela qualidade de seu código-fonte. Conforme a (ISO/IEC 15939, 2002) medição é uma ferramenta primordial para avaliar a qualidade dos produtos e a capacidade de processos organizacionais, portanto o Órgão Público contratante pode fazer uso do monitoramento de métricas de código-fonte para assistir ao processo de aferição de qualidade do *software* desenvolvido pela contratada.

(RêGO, 2014) propôs uma solução para o monitoramento de métricas de código-fonte com suporte de um ambiente de *Data Warehousing* que será explicada no capítulo 5. Uma das principais contribuições desse trabalho será evidenciar a eficácia e a eficiência

da proposta por Rêgo (2014) no aferimento de qualidade do software adquirido de uma empresa terceirizada para equipe responsável do Órgão público e seus demais envolvidos.

1.3 Problema

Com foco na avaliação de controles gerais de Tecnologia da Informação nos órgãos públicos, em 2011, o TCU detectou, por meio de auditorias de governança de TI, em diversos órgãos uma considerável frequência de irregularidades relacionadas à inexistência, deficiências e a falhas de processos de *software* que comprometem a eficácia e eficiência das contratações de desenvolvimento de sistemas (BRASIL, 2011).

A inexistência de parâmetros de aferição de qualidade para contratação de desenvolvimento de sistemas e a deficiência no processo de contratação, decorrente da inexistência de metodologia que assegure boa contratação de desenvolvimento de sistemas foram listados nas auditorias como consequências da inexistência e falha de processos de *software* nos órgãos públicos. Os critérios indicados pelo TCU no acórdão (BRASIL, 2011) foram:Constituição Federal, art. 37, caput; Instrução Normativa 4/2008, SLTI/MPOG, art. 12, inciso II; Lei 8666/1993, art. 6º, inciso IX; Norma Técnica - ITGI - Cobit 4.1, PO8.3 - Padrões de desenvolvimento e de aquisições; Norma Técnica - NBR ISO/IEC - 12.207 e 15.504;e Resolução 90/2009, CNJ, art. 10.

A auditoria do acordão (BRASIL, 2011) reforça o problema da falta de capacidade da administração pública de aferir a qualidade interna dos produtos de software desenvolvido por terceirizadas. A partir desse problema e da solução para monitoramento de métricas de código-fonte com suporte de um ambiente de *Data Warehousing* proposta por Rêgo (2014) , foi formulada a questão de pesquisa geral deste trabalho que é:

O uso de DW para o monitoramento de métricas de código fonte para assistir ao processo de aferição de qualidade interna dos produtos de software desenvolvido por terceirizadas, do ponto de vista da equipe de qualidade no desenvolvimento de software na CAIXA, é eficaz e eficiente?

1.4 Objetivos

O objetivo geral deste trabalho é realizar um estudo de caso onde será analisado a eficácia e a eficiência no uso de DW para o monitoramento de métricas de código fonte para assistir ao processo de aferição de qualidade interna dos produtos de software desenvolvido por uma empresa terceirizada para uma organização pública brasileira a partir das métricas e cenários de limpeza coletados pelo código-fonte do *software* desenvolvido, de questionário aos principais envolvidos no processo de verificação de qualidade e da observação em campo. Dentre os objetivos específicos deste trabalho estão:

- Avaliar a eficácia e eficiência da Solução de DW no processo de aferição de qualidade interna dos produtos de software desenvolvido por terceirizadas.
- Definir, projetar e caracterizar o estudo de caso.
- Coletar métricas e cenários de limpeza a partir código-fonte do *software* adquirido pela organização selecionada;
- Realizar análise dos dados coletados.
- Relatar os resultados obtidos.

1.5 Metodologia de pesquisa

Nessa seção apresenta-se a metodologia de pesquisa adotada neste trabalho. Para isso, foram definidos: a natureza da pesquisa; o tipo de metodologia de pesquisa; o tipo de abordagem de pesquisa; os métodos de procedimentos de pesquisa e os tipos de técnicas de coletas de dados.

Os procedimentos de pesquisa selecionados foram pesquisa bibliográfica, documental, levantamento e estudo de caso. As técnicas de coleta de dados selecionadas foram entrevistas, questionários e registro de observação na vida real. A seleção metodológica é apresentada na Figura 1.

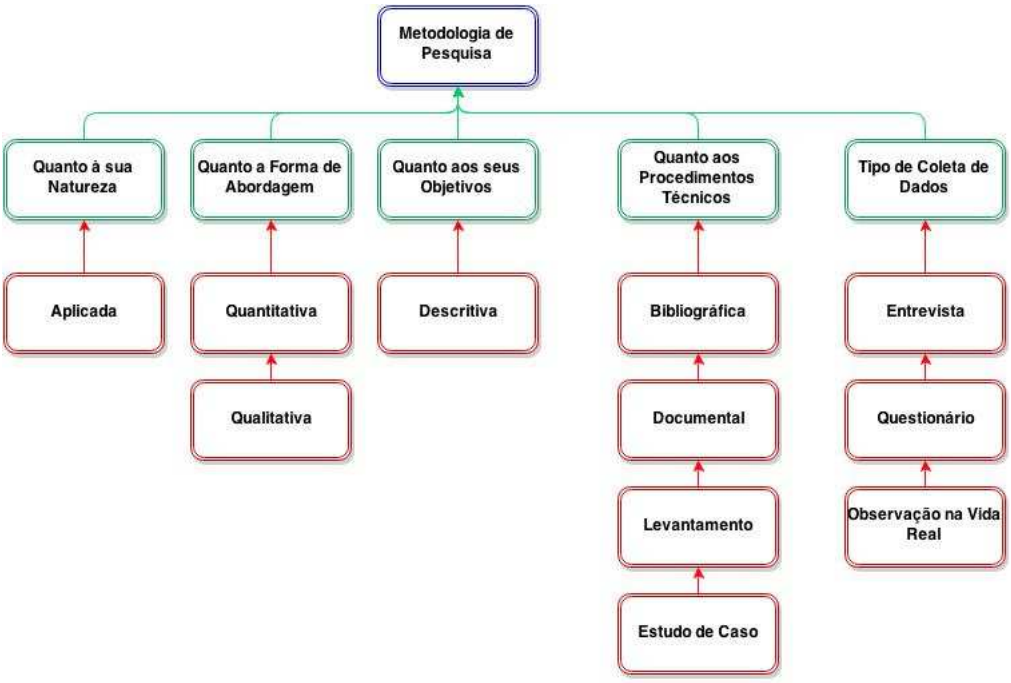

Figura 1 – Metodologia de Pesquisa

Segundo (YIN, 2001) o estudo de caso é um conjunto de procedimentos pré-especificados para se obter uma investigação empírica que investiga um fenômeno con-

temporâneo dentro de seu contexto da vida real, especialmente quando os limites entre o fenômeno e o contexto não estão claramente definidos. Uma grande vantagem do estudo de caso é a sua capacidade de lidar com uma ampla variedade de evidências - documentos, artefatos, entrevistas e observações - além do que pode estar disponível no estudo histórico convencional. Além disso, em algumas situações, como na observação participante, pode ocorrer manipulação informal.

A engenharia de *software* envolve o processo, desenvolvimento, operação e manutenção de *software* e artefatos relacionados. A maior parte da pesquisas na engenharia de software teve como objetivo investigar como o processo, desenvolvimento, operação e manutenção são conduzidos por engenheiros de *software* e outras partes interessadas em diferentes condições. Indivíduos, grupos e organizações, questões sociais e políticas são importantes para esse desenvolvimento. Isto é, a engenharia *software* é uma disciplina multidisciplinar que envolve áreas onde os estudos de caso são realizadas, como na psicologia, sociologia, ciência política, serviço social. Isto significa que muitas questões de pesquisa na engenharia de software são adequados para estudo de caso.(WOHLIN et al., 2012).

(WOHLIN et al., 2012) fraciona o estudo de caso em cinco passos. Nesta pesquisa, o estudo de caso compreende os passos: Planejar o Estudo de Caso; Coletar Dados; Analisar Dados Coletados e Compartilhar os Resultados. Portanto, os passos Projetar o Estudo de Caso e Preparar a Coleta de Dados definidas por Wohlin et al. (2012) foram agrupadas no passo Planejar o estudo de caso como na Figura 2.

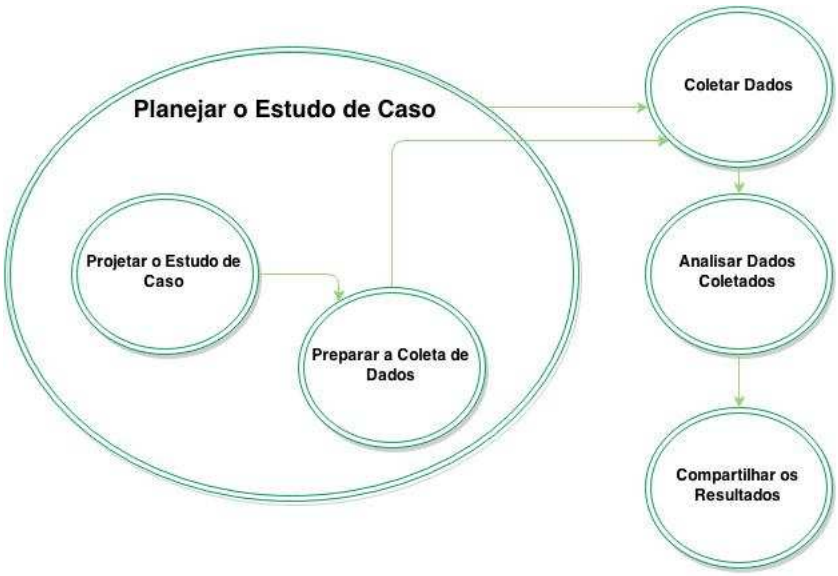

Figura 2 – Passos do Estudo de Caso

O passo Planejar o Estudo de Caso consiste na determinação do objetivo e da questão de pesquisa, da escolha da metodologia de pesquisa, da definição das fases da pesquisa, da definição do procedimentos de pesquisa, do protocolo, das técnicas de coleta

de dados e da proposta do trabalho final.

No passo Coletar Dados são executados os procedimentos de pesquisa e as técnicas de coletas de dados a seguir:

- Pesquisa Bibliográfica: pesquisa realizada a partir de livros, dissertações e trabalhos relacionados à área de pesquisa;
- Pesquisa Documental: pesquisa realizada a partir de documentos publicados por organizações públicas;
- Estudo de Caso: utilizar um estudo de caso real de uma organização pública brasileira;
- Entrevistas: dados serão coletados por meio de entrevistas informais, além de questionário, para incremento do estudo de caso;
- Documentos: coleta de dados dos documentos dos processos fornecidos pelo órgão público do estudo de caso será realizada para coleta de dados para análise;
- Observação na Vida Real: coleta de dados a partir da observação no campo de estudo;

O passo Analizar Dados Coletados d é onde os dados coletados serão analisados e interpretados. A análise compreende tanto a análise quantitativa quanto a análise qualitativa.

Por fim o passo Compartilhar os resultados diz respeito expor os resultados de forma adequada para o leitor alvo.

1.6 Organização do Trabalho

Esse trabalho está dividido em 5 capítulos:

- **Capítulo 1 - Introdução:** Esse capítulo tem como objetivo apresentar o contexto que esse trabalho está inserido, o problema sobre o qual ele buscará resolver, qual a justificativa, os objetivos da sua realização e metodologia de pesquisa adotada.
- **Capítulo 2 - Métricas de Software:** Capítulo responsável pela explicação teórica a respeito do que são métricas de código e como elas foram utilizadas no desenvolvimento da solução que esse trabalho busca analisar.
- **Capítulo 3 - Contratações de Fornecedores de Desenvolvimento de Software:** apresenta-se as principais informações referentes à Contratação de Fornecedores de Desenvolvimento de Software. Para isso, o capítulo é iniciado com uma visão geral sobre a importância das contratações e suas principais características. Poste-

riormente, é apresentado um resumo das legislações relacionadas à Contratação de Soluções de Tecnologia da Informação.

- **Capítulo 4 - Data Warehouse:** Nesse capítulo serão apresentados conceitos teóricos sobre *Data Warehousing*, assim como a maneira como foi desenvolvido e como funciona o ambiente de *Data Warehouse* para armazenamento de métricas de código fonte.

- **Capítulo 5 - Projeto de estudo de caso:** é apresentado o projeto do estudo de caso resultante do passo planejar Estudo de Caso, buscando demonstrar o escopo e elaborar um protocolo para o estudo de caso que será realizado. Elementos de pesquisa serão identificados e explicados como o problema a ser resolvido, os objetivos a serem alcançados no estudo de caso, quais os métodos de coleta e a análise dos dados.

- **Capítulo 6 - Conclusão:** Além das considerações finais dessa primeira parte do trabalho, serão descritos objetivos para o trabalho de conclusão de curso dois.

2 Métricas de Software

2.1 Processo de Medição

Segundo o (SOFTEX, 2013b) o processo Medição é um processo que apoia os processos de gerência e melhoria de processo e de produto, sendo um dos processos principais para gerenciar as atividades do ciclo de vida do trabalho e avaliar a viabilidade dos planos de trabalho. O propósito da Medição é coletar e analisar os dados relativos aos produtos desenvolvidos e aos processos implementados na organização e em seus trabalhos, de forma a apoiar o efetivo gerenciamento dos processos e demonstrar objetivamente a qualidade dos produtos(ISO/IEC 15939, 2008).

Ainda segundo o (SOFTEX, 2013b) Entende-se por método de medição uma sequência lógica de operações, descritas genericamente, usadas para quantificar um atributo com respeito a uma escala especificada. Esta escala pode ser nominal, ordinal ou de razão (de proporção), bem como definida em um intervalo.

- **Nominal:** A ordem não possui significado na interpretação dos valores (MEIRELLES, 2013)
- **Ordinal:** A ordem dos valores possui significado, porém a distância entre os valores não. (MEIRELLES, 2013)
- **Intervalo:** A ordem dos valores possui significado e a distância entre os valores também. Porém, a proporção entre os valores não necessariamente possui significado. (MEIRELLES, 2013)
- **Racional:** Semelhante a a medida com escala do tipo intervalo, porém a proporção possui significado. (MEIRELLES, 2013)

A ISO/IEC 15939 (2002) também divide o processo de medição em dois métodos diferentes, que se distinguem pela natureza do que é quantificado:

- **Subjetiva:** Quantificação envolvendo julgamento de um humano
- **Objetiva:** Quantificação baseada em regras numéricas. Essas regras podem ser implementadas por um humano.

2.2 Definição das métricas de software

As métricas de software são medidas resultantes da medição do produto ou do processo do *software* pelo qual é desenvolvido, sendo que o produto de *software* deve

ser visto como um objeto abstrato que se desenvolveu a partir de uma declaração inicial da necessidade de um sistema para um software finalizado, incluindo o código-fonte e as várias formas de documentação produzida durante o desenvolvimento Mills (1999). Estas medidas resultantes podem ser estudas para serem utilizadas para medir a produtividade e a qualidade do produto.

(MILLS, 1999) classifica as métricas de software como métricas de produtos ou métricas de processo, essa divisão pode ser vista na figura 3.

- **Métricas de produtos:** são as medidas do produto de software em qualquer fase do seu desenvolvimento, a partir dos requisitos do sistema. Métricas de produto podem medir a complexidade da arquitetura do software, o tamanho do programa(código-fonte), ou o número de páginas de documentos produzidos.

- **Métricas de processo:** são as medidas do processo de desenvolvimento de software, como o tempo de desenvolvimento global, o tipo de metodologia utilizada, ou o nível médio da experiência da equipe de programação.

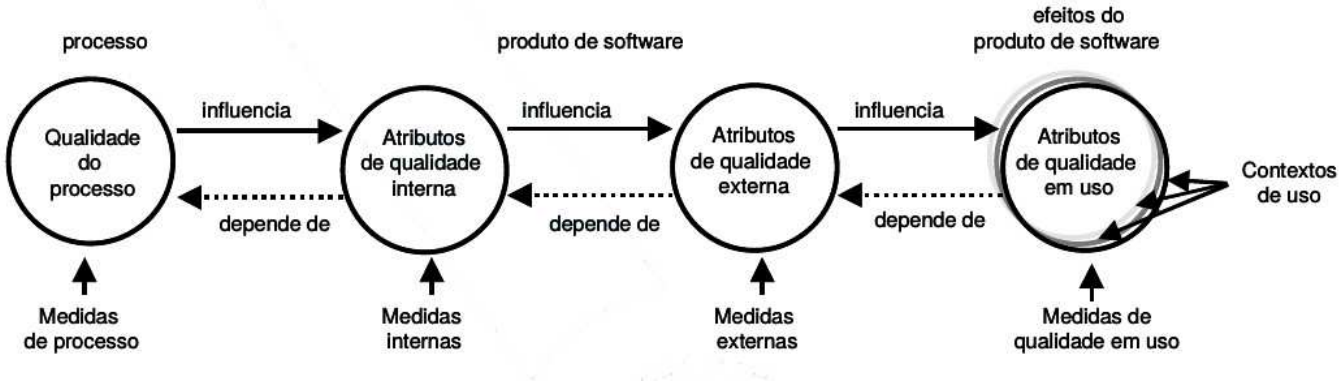

Figura 3 – Modelo de Qualidade do Produto da ISO/IEC 9126 (2001)

A figura 3 também pode ser notado a classificação das métricas de acordo com os diferentes tipos de medição, refletindo como as métricas influenciam nos contextos em que elas estão envolvidas, A qualidade do produto de software pode ser avaliada medindo-se os atributos internos (tipicamente medidas estáticas de produtos intermediários), os atributos externos (tipicamente pela medição do comportamento do código quando executado) ou os atributos de qualidade em uso (ISO/IEC 9126, 2001).

- **Métricas internas:** Aplicadas em um produto de software não executável, como código fonte. Oferecem aos usuários, desenvolvedores ou avaliadores o benefício de poder avaliar a qualidade do produto antes que ele seja executável.

- **Métricas externas:** Aplicadas a um produto de software executável, medindo o comportamento do sistema do qual o software é uma parte através de teste, operação ou mesmo obervação. Oferecem aos usuários, desenvolvedores ou avaliadores o

benefício de poder avaliar a qualidade do produto durante seu processo de teste ou operação.

- **Métricas de qualidade em uso:** Aplicadas para medir o quanto um produto atende as necessidades de um usuário para que sejam atingidas metas especificadas como eficácia, produtividade, segurança e satisfação.

2.3 Métricas de código fonte

A definição de código-fonte segundo a SCAM é qualquer descrição executável de um sistema de software sendo incluído código de máquina, linguagens de alto nível e por representações gráficas executáveis (HARMAN, 2010).

Segundo (MILLS, 1999) a maior parte do trabalho inicial em métricas de produto analisaram as características do código-fonte. A partir da experiência com métricas e modelos, tornou-se cada vez mais evidente que as informações de métricas obtidas anteriormente do ciclo de desenvolvimento pode ser de grande valor no controle do processo e dos resultados, o que sucedeu uma série de trabalhos tratando sobre o tamanho ou complexidade do software.

Neste capítulo será evidenciado em duas categorias as métricas de código-fonte que serão utilizadas neste trabalho de conclusão de curso, métricas de tamanho e complexidade e métricas de orientação a objetos. Estas métricas são objetivas e serão calculadas a partir da análise estática do código-fonte de um software.

2.3.1 Métricas de tamanho e complexidade

Cada produto do desenvolvimento de software é uma entidade física, como tal, pode ser descrito em termos de tamanho. Desde outros objetos físicos são facilmente mensuráveis (em comprimento, volume, massa, ou outra medida padrão), medir o tamanho do software deve ser relativamente simples e coerente de acordo com os princípios da teoria da medição. No entanto, na prática, a medição de tamanho apresenta grandes dificuldades(FENTON; PFLEEGER, 1998).

Segundo (HONGLEI; WEI; YANAN, 2009) as métricas de complexidade *software* pertencem as principais medições de software e é o principal método para assegurar a qualidade do *software*. Quanto menor a complexidade dos programas, melhor eles são, as métricas de complexidade também podem ser usadas para prever os defeitos ou erros. A seguir são apresentadas algumas métricas de tamanho e complexidade.

- **LOC** (*Lines of Code*): Métrica simples em que são contadas as linhas executáveis de um código, desconsiderando linhas em branco e comentários. (KAN, 2002)

- **ACCM** (*Average Cyclomatic Complexity per Method*): Mede a complexidade do programa, podendo ser representada através de um grafo de fluxo de controle. (MCCABE, 1976)

- **AMLOC** (*Average Method Lines of Code*): Indica a distribuição de código entre os métodos. Quanto maior o valor da métrica, mais pesado é o método. É preferível que haja muitos métodos com pequenas operações do que um método grande e de entendimento complexo. (MEIRELLES, 2013)

2.3.2 Métricas de Orientação a Objetos

Segundo (BUDD, 2002) a programação orientada a objetos tornou-se extremamente popular nos últimos anos, inúmeros livros e edições especiais de revistas acadêmicas e comerciais têm surgido sobre o assunto. A julgar por essa atividade frenética, programação orientada a objeto está sendo recebido com ainda mais entusiasmo do que vimos em ideias revolucionárias mais antigas, tais como programação estruturada "ou sistemas especialistas."

Há uma série de razões importantes pelas quais nas últimas duas décadas a programação orientada a objetos tornou-se o paradigma de programação dominante. A programação orientada a objeto possui uma ótima escala, desde o mais trivial dos problemas para a maioria das tarefas complexas. Ele fornece uma forma de abstração que reflete em técnicas de como pessoas usam para resolver problemas em sua vida cotidiana. E para a maioria da linguagens orientadas a objeto dominante há um número cada vez maior de bibliotecas que auxiliam no desenvolvimento de aplicações para muitos domínios (BUDD, 2002).

Serão adotadas neste trabalho as seguintes métricas de orientação a objetos:

- **ACC** (*Afferent Connections per Class* - Conexões Aferentes por Classe): Mede a conectividade entre as classes. Quanto maior a conectividade entre elas, maior o potencial de impacto que uma alteração pode gerar. (MEIRELLES, 2013)

- **ANPM** (*Average Number of Parameters per Method* - Média do Número de Parâmetros por Método): Indica a média de parâmetros que os métodos possuem. Um valor muito alto para quantidade de parâmetros pode indicar que o método está tendo mais de uma responsabilidade. (BASILI; ROMBACH, 1987)

- **CBO** (*Coupling Between Objects* - Acoplamento entre Objetos): Essa é uma métrica que diz respeito a quantas outras classes dependem de uma classe. É a conta das classes às quais uma classe está acoplada. Duas classes estão acopladas quando métodos de uma delas utilizam métodos ou variáveis de outra. Altos valores dessa métrica aumentam a complexidade e diminuem a manutenibilidade. (LAIRD, 2006).

- **DIT** (*Depth of Inheritance Tree* - Profundidade da Árvore de Herança): Responsável por medir quantas camadas de herança compõem uma determinada hierarquia de classes (LAIRD, 2006). Segundo Meirelles (2013), quanto maior o valor de DIT, maior o número de métodos e atributos herdados, portanto maior a complexidade.

- **LCOM4** (*Lack of Cohesion in Methods* - Falta de Coesão entre Métodos): A coesão de uma classe é indicada por quão próximas as variáveis locais estão relacionadas com variáveis de instância locais. Alta coesão indica uma boa subdivisão de classes. A LCOM mede a falta de coesão através dissimilaridade dos métodos de uma classe pelo emprego de variáveis de instância.(KAN, 2002). A métrica LCOM foi revista e passou a ser conhecida como LCOM4, sendo necessário para seu cálculo a construção de um gráfico não-orientado em que os nós são os atributos e métodos de uma classe. Para cada método deve haver uma aresta entre ele e outro método ou variável. O valor da LCOM4 é o número de componentes fracamente conectados a esse gráfico (MEIRELLES, 2013)

- **NOC** (*Number of Children* - Número de Filhos): É o número de sucessores imediatos, (portanto filhos) de uma classe. Segundo Laird (2006), altos valores indicam que a abstração da super classe foi diluída e uma reorganização da arquitetura deve ser considerada.

- **NOM** (*Number of Methods* - Número de Métodos): Indica a quantidade de métodos de uma classe, medindo seu tamanho. Classes com muitos métodos são mais difíceis de serem reutilizadas pois são propensas a serem menos coesas. (MEIRELLES, 2013)

- **NPA** (*Number of Public Attributes* - Número de Atributos Públicos): Mede o encapsulamento de uma classe, através da medição dos atributos públicos. O número ideal para essa métrica é zero (MEIRELLES, 2013)

- **RFC** (*Response For a Class* - Respostas para uma Classe): Kan (2002) define essa métrica como o número de métodos que podem ser executados em respostas a uma mensagem recebida por um objeto da classe.

2.4 Configurações de qualidade para métricas de código fonte

Meirelles (2013) apresentou uma abordagem para a observação das métricas de código-fonte em seu doutorado, onde estas métricas foram estudadas através de suas distribuições e associações. Foram avaliados a distribuição e correlações dos valores das métricas de 38 projetos de software livre, sendo coletado e analisado valores para cada métrica em mais de 344.872 classes e módulos. Segundo o próprio Meirelles (2013) entre as principais contribuições de sua tese foi a análise detalhada, em relação ao comportamento,

valores e estudos de caso, de 15 métricas de código-fonte.

Dentre os objetivos científicos da tese de Meirelles (2013) o mais crucial para este trabalho de conclusão de curso é o objetivo OC4 que trata das distribuições estatísticas dos valores de métricas em 38 projetos de software livre, a fim de compreender qual a abordagem estatística mais informativa para o monitoramento dessas métricas, bem como observar os valores frequentes para essas métricas, de forma a servirem de referência para projetos futuros.

(MEIRELLES, 2013) para conduzir o seu estudo qualitativo utilizou-se da técnica de estatística descritiva: percentil. O percentil separa os dados em cem grupos que apresentam o mesmo número de valores, sendo a definição do percentil de ordem px100(0<p<1), em um conjunto de dados de tamanho n, é o valor da variável que ocupa a posição px(n+1) do conjunto de dados ordenados. O percentil de ordem p (ou p-quantil) deixa px100% das observações abaixo dele na amostra ordenada. O Percentil 25 recebe o nome de primeiro quartil, o percentil 50 de segundo quartil ou mediana, e o percentil 75 de terceiro quartil. Uma das hipóteses que Meirelles (2013) utiliza é que só a partir do terceiro quartil será possível obter dados informativos sobre as métricas.

Após a aplicação da técnica percentil na série de dados das métricas de código-fonte obtidos da análise estática do código-fonte dos projetos de software livre, tabelas apresentando os valores percentis de cada métrica nos 38 projetos avaliados foram criadas, como a Tabela 1, que mostra os percentis para a métrica NOC.

	Mín	**1%**	**5%**	**10%**	**25%**	**50%**	**75%**	**90%**	**95%**	**99%**	**Máx**
Eclipse	0,0	0,0	0,0	0,0	0,0	0,0	0,0	2,0	3,0	10,0	243,0
Open JDK8	0,0	0,0	0,0	0,0	0,0	0,0	0,0	1,0	3,0	9,0	301,0
Ant	0,0	0,0	0,0	0,0	0,0	0,0	0,0	1,0	3,0	12,0	162,0
Checkstyle	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	2,0	5,0	144,0
Eclipse Metrics	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	2,0	11,0	24,0
Findbugs	0,0	0,0	0,0	0,0	0,0	0,0	0,0	1,0	2,0	5,0	55,0
GWT	0,0	0,0	0,0	0,0	0,0	0,0	0,0	39,0	1,0	8,0	398,0
Hudson	0,0	0,0	0,0	0,0	0,0	0,0	0,0	1,0	2,0	7,0	23,0
JBoss	0,0	0,0	0,0	0,0	0,0	0,0	0,0	1,0	2,0	6,0	256,0
Kalibro	0,0	0,0	0,0	0,0	0,0	0,0	0,0	1,0	2,0	9,0	101,0
Log4J	0,0	0,0	0,0	0,0	0,0	0,0	0,0	1,0	3,0	9,0	23,0
Netbeans	0,0	0,0	0,0	0,0	0,0	0,0	0,0	1,0	2,0	6,0	989,0
Spring	0,0	0,0	0,0	0,0	0,0	0,0	0,0	2,0	3,0	7,0	175,0
Tomcat	0,0	0,0	0,0	0,0	0,0	0,0	0,0	1,0	2,0	8,0	82,0

Tabela 1 – Percentis para métrica NOC em projetos Java extraídos de Meirelles (2013)

Através dos resultados obtidos para cada métrica, Meirelles (2013) observou que era possível identificar valores frequentes analisando os percentis. Na tabela 1, por exemplo, considerando o projeto **Open JDK8** como referência foram observados os intervalos

de valores de 0, como muito frequente, 1 e 2 como frequente, 3 como pouco frequente e acima de 3 como não frequente (MEIRELLES, 2013).

(RêGO, 2014) observando o trabalho de análise de percentis de (MEIRELLES, 2013) percebeu que é possível utilizar os intervalos de frequência obtidos como uma evidência empírica de qualidade do código-fonte. (RêGO, 2014) também renomeou os intervalos de frequência obtidos por (MEIRELLES, 2013), como na Tabela 2, a fim de facilitar a interpretação de métricas de código-fonte.

Intervalo de Frequência	Intervalo Qualitativo
Muito Frequente	Excelente
Frequente	Bom
Pouco Frequente	Regular
Não Frequente	Ruim

Tabela 2 – Nome dos Intervalos de Frequência extraídos de Rêgo (2014)

(RêGO, 2014) observando a diferença dos valores das métricas entre os projetos analisados e na tentativa de diminuir tamanha diferença, considerou dois cenários. Foi utilizado dois *softwares* como referências para cada uma das métricas na linguagem de programação Java. Em um primeiro cenário, foi analisado o *Open JDK8*, software que contia os menores valores percentis para as métricas. Já em um segundo cenário, foi considerado o Tomcat contendo os valores percentis mais altos. A tabela 3 mostra o resultado desta análise:

Métrica	Intervalo Qualitativo	OpenJDK8 Metrics	Tomcat Metrics
LOC	Excelente	[de 0 a 33]	[de 0 a 33]
	Bom	[de 34 a 87]	[de 34 a 105]
	Regular	[de 88 a 200]	[de 106 a 276]
	Ruim	[acima de 200]	[acima de 276]
ACCM	Excelente	[de 0 a 2,8]	[de 0 a 3]
	Bom	[de 2,9 a 4,4]	[de 3,1 a 4,0]
	Regular	[de 4,5 a 6,0]	[de 4,1 a 6,0]
	Ruim	[acima de 6]	[acima de 6]
AMLOC	Excelente	[de 0 a 8,3]	[de 0 a 8]
	Bom	[de 8,4 a 18]	[de 8,1 a 16,0]
	Regular	[de 19 a 34]	[de 16,1 a 27]
	Ruim	[acima de 34]	[acima de 27]
ACC	Excelente	[de 0 a 1]	[de 0 a 1,0]
	Bom	[de 1,1 a 5]	[de 1,1 a 5,0]
	Regular	[de 5,1 a 12]	[de 5,1 a 13]
	Ruim	[acima de 12]	[acima de 13]
ANPM	Excelente	[de 0 a 1,5]	[de 0 a 2,0]
	Bom	[de 1,6 a 2,3]	[de 2,1 a 3,0]
	Regular	[de 2,4 a 3,0]	[de 3,1 a 5,0]
	Ruim	[acima de 3]	[acima de 5]
CBO	Excelente	[de 0 a 3]	[de 0 a 2]
	Bom	[de 4 a 6]	[de 3 a 5]
	Regular	[de 7 a 9]	[de 5 a 7]
	Ruim	[acima de 9]	[acima de 7]
DIT	Excelente	[de 0 a 2]	[de 0 a 1]
	Bom	[de 3 a 4]	[de 2 a 3]
	Regular	[de 5 a 6]	[de 3 a 4]
	Ruim	[acima de 6]	[acima de 4]
LCOM4	Excelente	[de 0 a 3]	[de 0 a 3]
	Bom	[de 4 a 7]	[de 4 a 7]
	Regular	[de 8 a 12]	[de 8 a 11]
	Ruim	[acima de 12]	[acima de 11]
NOC	Excelente	[0]	[1]
	Bom	[1 a 2]	[1 a 2]
	Regular	[3]	[3]
	Ruim	[acima de 3]	[acima de 3]
NOM	Excelente	[de 0 a 8]	[de 0 a 10]
	Bom	[de 9 a 17]	[de 11 a 21]
	Regular	[de 18 a 27]	[de 22 a 35]
	Ruim	[acima de 27]	[acima de 35]
NPA	Excelente	[0]	[0]
	Bom	[1]	[1]
	Regular	[de 2 a 3]	[de 2 a 3]
	Ruim	[acima de 3]	[acima de 3]
RFC	Excelente	[de 0 a 9]	[de 0 a 11]
	Bom	[de 10 a 26]	[de 12 a 30]
	Regular	[de 27 a 59]	[de 31 a 74]
	Ruim	[acima de 59]	[acima de 74]

Tabela 3 – Configurações para os Intervalos das Métricas para Java extraídas de Rêgo (2014)

2.5 Cenários de limpeza

(MACHINI et al., 2010) apresenta uma maneira de interpretar os valores das métricas através de cenários problemáticos e suas possíveis melhorias, para que as métricas possam ser mais facilmente incorporadas no cotidiano dos programadores. Machini et al. (2010) também apresenta um estilo de programação baseado no paradigma da Orientação a Objetos que busca o que denominamos de "Código Limpo", concebido e aplicado por renomados desenvolvedores de software como Robert C. Martin (MARTIN, 2008) e Kent Beck (BECK, 2007)

Segundo (BECK, 2007) um código limpo está inserido em um estilo de programação que busca a proximidade a três valores: expressividade, simplicidade e flexibilidade.

- **Expressividade:** Um código se expressa bem quando alguém que o lê é capaz de compreendê-lo e modificá-lo. Beck (2007) destaca que quando foi necessário modificar um código, ele gastou muito mais tempo lendo o que já havia sido feito do que escrevendo sua modificação.

- **Simplicidade:** Um código é expressivo quando pode ser facilmente lido nas diferentes camadas de abstração, havendo uma redução da quantidade de informação que o leitor deve compreender para fazer alterações. Eliminar o excesso de complexidade faz com que aqueles que estejam lendo o código o entenda mais rapidamente.

- **Flexibilidade:** Capacidade de estender a aplicação alterando o mínimo possível a estrutura já criada.

(MACHINI et al., 2010) apresenta em seu trabalho diversas técnicas para a obtenção de um código limpo, um resumo destas técnicas são apresentadas nas tabelas 4, 5, 6 e 7.

Técnicas de Limpeza	Descrição	Contribuições	Consequências
Composição de Métodos	Compor os métodos em chamadas para outros rigorosamente no mesmo nível de abstração abaixo.	• Facilidade de entendimento de métodos menores • Criação de métodos menores com nomes explicativos	• Menos Operações por Método • Mais Parâmetros de Classe • Mais Métodos na Classe
Métodos Explicativos	Criar um método que encapsule uma operação pouco clara geralmente associada a um comentário	• O código cliente do método novo terá uma operação com nome que melhor se encaixa no contexto.	• Mais métodos na classe.
Métodos como Condicionais	Criar um método que encapsule uma expressão booleana para obter condicionais mais claros.	• Facilidade na leitura de condicionais no código cliente. • Encapsulamento de uma expressão booleana.	• Mais métodos na classe.
Evitar Estruturas Encadeadas	Utilizar a composição de métodos para minimizar a quantidade de estruturas encadeadas em cada método (if, else).	• Facilidade para a criação de testes. • Cada método terá estruturas mais simples e fáceis de serem compreendidas.	• Menos Estruturas encadeadas por método (if e else) • Benefícios do Uso de Composição de Métodos

Tabela 4 – Conceitos de Limpeza levantados por Machini et al. (2010) e adaptados por Rêgo (2014) parte 1.

Técnicas de Limpeza	Descrição	Contribuições	Consequências
Cláusulas Guarda	Criar um retorno logo no inicio de um método ao invés da criação de estruturas encadeadas com if sem else.	• Estruturas de condicionais mais simples. • Leitor não espera por uma contrapartida do condicional(ex:if sem else).	• Menos estruturas encadeadas na classe.
Objeto Método	Criar uma classe que encapsule uma operação complexa simplificando a original(cliente).	• O código cliente terá um método bastante simples. • Nova classe poderá ser refatorada sem preocupações com alterações no código cliente. • Nova classe poderá ter testes separados.	• Menos operações no método cliente. • Menos responsabilidades da classe cliente. • Mais classes. • Mais acoplamento da classe cliente com a nova classe.
Evitar *Flags* como Argumentos	Ao invés de criar um método que recebe uma flag e tem diversos comportamentos, criar um método para cada comportamento.	• Leitor não precisará entender um método com muitos condicionais. • Testes de unidade independentes.	• Mais métodos na classe.

Tabela 5 – Conceitos de Limpeza levantados por Machini et al. (2010) e adaptados por Rêgo (2014) parte 2.

Técnicas de Limpeza	Descrição	Contribuições	Consequências
Maximizar a Coesão	Quebrar uma classe que não segue o Princípio da Responsabilidade Única: as classes devem ter uma única responsabilidade, ou seja, ter uma única razão para mudar.	• Cada classe terá uma única responsabilidade. • Cada classe terá seus testes independentes. • Sem interferências na implementação das responsabilidades.	• Mais Classes • Menos Métodos em cada Classe • Menos Atributos em cada Classe
Objeto como Parâmetro	Localizar parâmetros que formam uma unidade e criar uma classe que os encapsule.	• Menor número de parâmetros facilita testes e legibilidade. • Criação de uma classe que poderá ser reutilizada em outras partes do sistema.	• Menos Parâmetros sendo passados para Métodos • Mais Classes
Parâmetros como Variável de Instância	Localizar parâmetro muito utilizado pelos métodos de uma classe e transformá-lo em variável de instância.	• Não haverá a necessidade de passar longas listas de parâmetro através de todos os métodos.	• Menos Parâmetros passados pela Classe • Possível diminuição na coesão
Uso de Exceções	Criar um fluxo normal separado do fluxo de tratamento de erros utilizando exceções ao invés de valores de retornos e condicionais.	• Clareza do fluxo normal sem tratamento de erros através de valores de retornos e condicionais.	• Menos Estruturas encadeadas.

Tabela 6 – Conceitos de Limpeza levantados por Machini et al. (2010) e adaptados por Rêgo (2014) parte 3.

Técnicas de Limpeza	Descrição	Contribuições	Consequências
Delegação de Tarefa	Transferir um método que utiliza dados de uma classe "B"para a "B".	• Redução do acoplamento entre classes. • Proximidade dos métodos e dados sobre os quais trabalham.	• Menos métodos na classe inicial. • Mais métodos na classe que recebe o novo método.
Objeto Centralizador	Criar uma classe que encapsule uma operação com alta dependência entre classes.	• Simplificação da classe cliente. • Redução do acoplamento da classe cliente com as demais. • Nova classe poderá receber testes e melhorias independentes.	• Menos operações no método cliente. • Menos responsabilidades da classe cliente. • Mais classes. • Mais acoplamento da classe cliente com a nova classe.
Uso Excessivo de Herança	Localizar uso excessivo de herança e transformá-lo em agregação simples.	• Redução do do acoplamento entre as classes.	• Maior Flexibilidade de Adição de Novas Classes. • Menor Acoplamento entre as classes.
Exposição Pública Excessiva	Localizar uso excessivo de parâmetros públicos e transformá-lo em parâmetros privados.	• Menor Acoplamento entre as classes.	• Maior Encapsulamento de Parâmetros.

Tabela 7 – Conceitos de Limpeza levantados por Machini et al. (2010) e adaptados por Rêgo (2014) parte 4.

Após apresentar as técnicas para obtenção de um código limpo Machini et al. (2010) adota uma abordagem baseada em cenários para identificar trechos de código com características indesejáveis, onde estes cenários devem possuir um contexto criado a partir

de poucos conceitos de código limpo no qual um pequeno conjunto de métricas é analisado e interpretado através da combinação de seus valores. A ideia principal desta abordagem é facilitar melhorias de implementação e a procura por problemas quanto a limpeza de código através da aproximação dos valores das métricas com os esperados nos contextos de interpretação (MACHINI et al., 2010).

(RêGO, 2014) utiliza a mesma abordagem de cenários que Machini et al. (2010) onde algumas técnicas de limpeza de código apresentadas nas tabelas 4, 5, 6 e 7 são correlacionadas com as métricas da Seção 2.3 e utiliza a configuração do *Open JDK8* considerando como valores altos os valores obtidos pelos intervalos Regular e Ruim tal como mostrados na Tabela 3.

Aproveitando inicialmente os cenários **Classe pouco coesa** e **Interface dos métodos** extraídos de Machini et al. (2010), Rêgo (2014) elaborou mais alguns cenários de limpeza. O resultado dessa atividade pode ser visto na tabela 8:

Cenário de Limpeza	Conceito de Limpeza	Características	Recomendações	Forma de Detecção pelas Métricas de Código-Fonte	Padrões de Projeto Associados
Classe Pouco Coesa	Maximização da Coesão	Classe Subdivida em grupos de métodos que não se relacionam	Reduzir a subdvisão da Classe	Intervalos Regulares e Ruins de LCOM4, RFC.	*Chain of Reponsabilities, Mediator, Decorator.*
Interface dos Métodos	Objetos como Parâmetro e Parâmetro como Variáveis de Instância	Elevada Média de parâmetros repassados pela Classe	Minimizar o número de Parâmetros.	Intervalos Regulares e Ruins de ANPM.	*Facade, Template Method, Strategy, Command, Mediator, Bridge.*
Classes com muitos filhos	Evitar Uso Excessivo de Herança	Muitas Sobreescritas de Métodos	Trocar a Herança por uma Agregação.	Intervalos Regulares e Ruins de NOC.	*Composite, Prototype, Decorator, Adapter.*
Classe com métodos grandes e/ou muitos condicionais	Composição de Métodos, Evitar Estrutura Encadeadas Complexas	Grande Número Efetivo de Linhas de Código	Reduzir LOC da Classe e de seus métodos, Reduzir a Complexidade Ciclomática e Quebrar os métodos.	Intervalos Regulares e Ruins de AMLOC, ACCM.	*Chain of Reponsabilities, Mediator, Flyweight* .
Classe com muita Exposição	Parâmetros Privados	Grande Número de Parâmetros Públicos	Reduzir o Número de Parâmetros Públicos.	Intervalos Regulares e Ruins de NPA.	*Facade, Singleton.*
Complexidade Estrutural	Maximização da Coesão	Grande Acoplamento entre Objetos	Reduzir a a quantidade de responsabilidades dos Métodos.	Intervalos Regulares e Ruins de CBO e LCOM4.	*Chain of Responsabilities, Mediator, Strategy.*

Tabela 8 – Cenários de Limpeza extraídos de Rêgo (2014)

2.6 Considerações Finais do Capítulo

Esse capítulo apresentou a fundamentação teórica sobre métricas de código fonte, quais são seus intervalos qualitativos e também foi apresentada a maneira como elas foram relacionadas a cenários de limpeza. O próximo capítulo será responsável por apresentar a teoria acerca de *Data Warehousing* e como seu uso pode servir no monitoramento das métricas apresentadas nesse capítulo.

3 Contratações de Fornecedores de Desenvolvimento de Software

3.1 Importância da Contratação de Fornecedores de Desenvolvimento de Software

3.2 Normas, Processos e Legislação Pertinentes à Contratação

3.3 Contrato do Órgão X

4 *Data Warehouse*

Neste capítulo será apresentada a fundamentação teórica sobre*Data Warehouse* e como utiliza-lo para o monitoramento de métricas. Também será apresentado ambiente de *Data Warehouse* utilizado na solução proposta por Rêgo (2014), cujo este trabalho busca analisar a eficácia e eficiência no monitoramento de métricas, e como ela foi desenvolvida.

4.1 Definição

Na década de 80 as organizações perceberam a importância de não apenas usar dados para propósitos operacionais, mas também para derivar a inteligência por trás deles. Essa inteligência não só justificaria na decisões passadas, mas também para ajudar na tomada de decisões para o futuro. O termo *Business Intelligence* tornou-se cada vez mais popular, e foi durante o final dos anos 1980 que pesquisadores da IBM, Barry Devlin e Paul Murphy desenvolveram o conceito de *Business data warehouse* . A partir que aplicações de *business intelligence* foram surgindo, foi rapidamente verificado que os dados de bancos transacionais tinham que primeiro ser transformados e armazenados em outros bancos de dados com um esquema específico para poderem derivar de sua inteligência. Esta base de dados poderia ser usada para arquivamento, e seria maior em tamanho do que as bases de dados transacionais, mas seu design seria ideal para executar relatórios que permitem as grandes organizações planejarem e tomarem decisões de forma proativa. Este banco de dados, normalmente armazenando as atividades realizadas no passado e no presente das organizações, foi chamado de *Data Warehouse* (SHARMA, 2011).

Para (INMON, 2002) *Data Warehouse* é uma coleção de dados que tem como característica ser orientada a assunto, integrada, não volátil e temporal. Por orientação a assunto, podemos entender como um foco em algum aspecto específico da organização. O fato do ambiente ser integrado remete ao fato dele ser alimentado com dados que têm como origem de múltiplas fontes, integrando esses dados de maneira a construir uma única orientação. Como um conjunto não volátil e temporal de dados, é entendido que a informação carregada remete a um determinado momento da aplicação, possibilitando assim acesso a diferentes intervalos de tempo, não havendo como modificá-los atualizando em tempo real.

Segundo (ROCHA, 2000) *data warehousing* é a infra-estrutura tecnológica de hardware e software para a atividade de análise gerencial. Agora que sabemos o que é o *data wharehouse* e *Data Warehousing*, será mostrado os componentes que compõem um ambiente completo de *data warehousing*. É importante entender como os componentes funcionam individualmente antes de começarem a ser combinados para se criar um *data*

warehouse. Cada componente do armazém tem uma função específica. É preciso aprender a importância estratégica de cada componente e como manuseá-los efetivamente para fazer uso do *data warehousing*. Uma das maiores ameaças à sucesso no *data warehousing* é confundir os papéis e as funções dos componentes (KIMBALL; ROSS, 2002). A Figura 4 descreve uma arquitetura geral de um ambiente de *Data Warehousing*, os componentes do ambiente serão esclarecidos no decorrer deste capítulo.

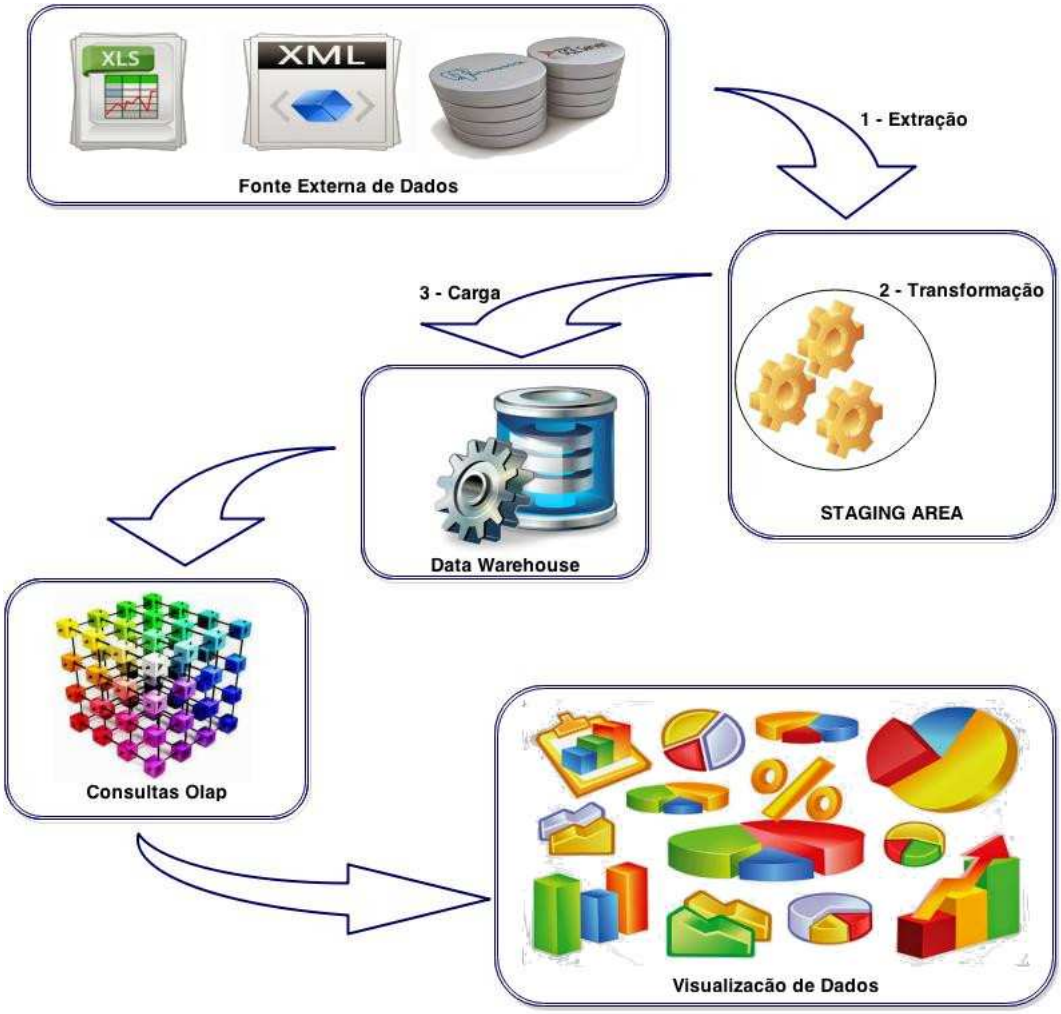

Figura 4 – Arquitetura de um ambiente de Data Warehousing

4.2 *Extraction-Transformation-Load*

A *Staging Area* é uma área de armazenamento onde acontece o processo de ETL. Na Figura 4 as etapas 1- Extração, 2- Transformação e 3- Carga formam o processo de *Extraction-Transformation-Load* (ETL). Cada uma das etapas recebe a seguinte descrição:

- **Extração:** Primeira etapa do processo de ETL, consiste na leitura e entendimento da fonte dos dados, copiando os que são necessário para futuros trabalhos (KIMBALL; ROSS, 2002).

- **Transformação:** Após a etapa de extração ter sido feita, os dados podem receber diversos tipos de transformações, que incluem correções de conflitos, conversão de formatos, remoção de campos que não são úteis, combinação entre dados de diversas fontes, entre outros (KIMBALL; ROSS, 2002).

- **Carga:** Após ter sido realizado o processo de transformação, os dados já estão prontos para serem carregados no *Data Warehouse*, tornando possível que todos os dados visualizados após esse processo reflitam a informação que passou pelos processos de extração e transformação (SHARMA, 2011).

4.3 Modelagem Dimensional

Modelagem dimensional é um novo nome para uma velha técnica para deixar os bancos de dados simples e compreensível. No início dos anos 70 as organizações de TI , consultores, usuários finais e fornecedores tiveram que migrar para uma estrutura dimensional simples que combinasse com a necessidade humana pela simplicidade (KIMBALL; ROSS, 2002).

Segundo (KIMBALL; ROSS, 2002) a modelagem dimensional tem sido amplamente aceita como a técnica dominante para a apresentação do *data warehouse*. Os profissionais e especialistas de *data warehouse* reconhecem que a apresentação do *data warehouse* deve ser fundamentado na simplicidade. A simplicidade é a chave fundamental que permite que os usuários entendam facilmente as bases de dados e naveguem de forma eficiente no bancos de dados do *software*.

Para facilitar na difusão do conceito de modelagem dimensional Kimball e Ross (2002) utiliza como exemplo um diretor geral que descreve seus negócios como "Nós vendemos produtos em várias áreas de negócio e medimos nossa desempenho ao longo do tempo". Assim designers dimensionais colocariam em ênfase o produto, as áreas de negócio e o tempo, pensando intuitivamente neste negócio como um cubo de dados, onde as bordas estariam marcadas como produto, negócio e tempo. Pontos dentro do cubo são onde as medições que combinam produtos, áreas de negócio e tempo são salvas. A capacidade de visualizar algo tão abstrato como um conjunto de dados de uma forma concreta e tangível é o segredo da compreensão.

Os conceitos básicos da modelagem dimensional são: fatos, dimensões e medidas. Um fato é uma coleção de itens de dados relacionados, que consiste em medidas e dados de contexto. Ele representa tipicamente itens de negócios ou transações de negócio. A dimensão é uma coleção de dados que descrevem uma dimensão negócio. Dimensões determinam o contextual a fundo para os fatos; eles são os parâmetros cujo queremos realizar a *On-Line Analytic Processing (OLAP)*. A medida é um atributo numérico de um fato, que representa o desempenho ou comportamento do negócio em relação às dimensões

(GUTIéRREZ; MAROTTA, 2000).

Considerando o contexto relacional, existem dois modelos básicos que são utilizados na modelagem dimensional: modelo de estrela e modelo de floco de neve. O modelo estrela é a estrutura básica de um modelo dimensional. Ele tem uma grande tabela central (tabelas fato) e um conjunto de pequenas tabelas (tabelas dimensão) dispostos em um padrão radial ao redor da tabela central, como é mostrado na Figura 5. O modelo de floco de neve é o resultado da decomposição de uma ou mais das dimensões (GUTIéRREZ; MAROTTA, 2000).

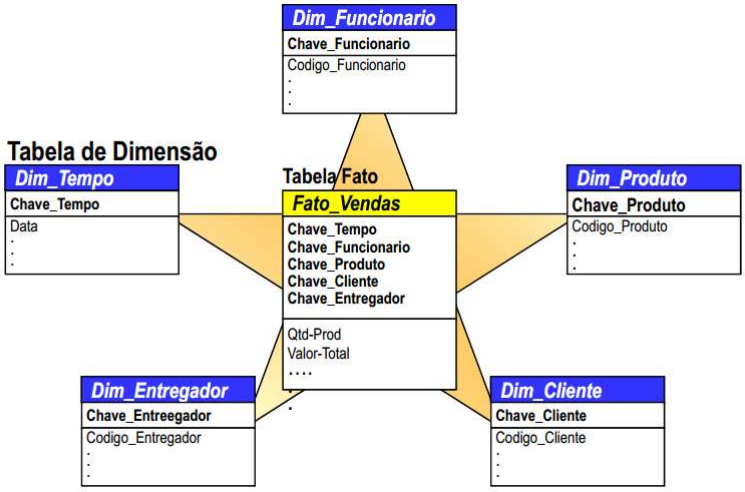

Figura 5 – Esquema estrela extraído de Times (2012)

4.3.1 OLAP *(On-Line Analytic Processing)*

A atividade que consiste em buscar e apresentar os dados de um *Data Warehouse*, sendo essa busca quase sempre baseada em um cubo multidimensional de dados, é chamada de *On-Line Analytic Processing* (OLAP) (KIMBALL; ROSS, 2002). Segundo (SHARMA, 2011) *On-Line Analytic Processing(OLAP)* refere-se a cargas onde grandes quantidades de históricos de dados são processados para gerar relatórios e executar análise de dados. Normalmente, bancos de dados que utilizam OLAP são alimentados a partir de bancos de dados *On-Line Transaction Processing* (OLTP) que são alterados para gerir cargas OLAP. Um banco de dados OLAP armazena um grande volume com os mesmos dados transacionais que um banco de dados OLTP, mas estes dados são transformados pelo processo de ETL para permitir um melhor desempenho para geração de relatórios e para facilitar as análises. Geralmente sistemas OLTP são ajustados para inserções, atualizações e exclusões bem rápidas, enquanto os sistemas OLAP estão ajustados para consultas rápidas.

A tabela 9 evidencia as diferenças entre aplicações OLTP e OLAP extraídas do trabalho de Neri (2002):

OLPT	OLAP
Controle operacional	Tomada de decisão
Atualização de dados	Análise de dados
Pequena complexidade das operações	Grande complexidade das operações
Não armazena dados históricos	Armazena dados históricos
Voltada ao pessoal operacional	Voltada aos gestores do negócio

Tabela 9 – Diferenças entre OLTP e OLAP extraídas de Neri (2002)

No modelo multidimensional, os dados são organizados em múltiplas dimensões, e cada dimensão contém vários níveis de abstração definidos pelas hierarquias. Esta organização fornece aos usuários a flexibilidade para visualizar os dados a partir de diferentes perspectivas. Existe uma série de operações de cubo de dados OLAP para materializar esses diferentes pontos de vista, permitindo consultas interativas e análises de dados. Assim, OLAP fornece um ambiente amigável para análise de dados interativos.

Entre as operações OLAP estão:

- ***Drill Down:*** Busca aumentar o nível de detalhamento, partindo de um certo nível de dados para um nível mais detalhado (SHARMA, 2011).
- ***Drill Up:*** Ao contrário da operação *Drill Down*, a *Roll Up* parte de um nível mais detalhado para um nível menos detalhado (SHARMA, 2011).

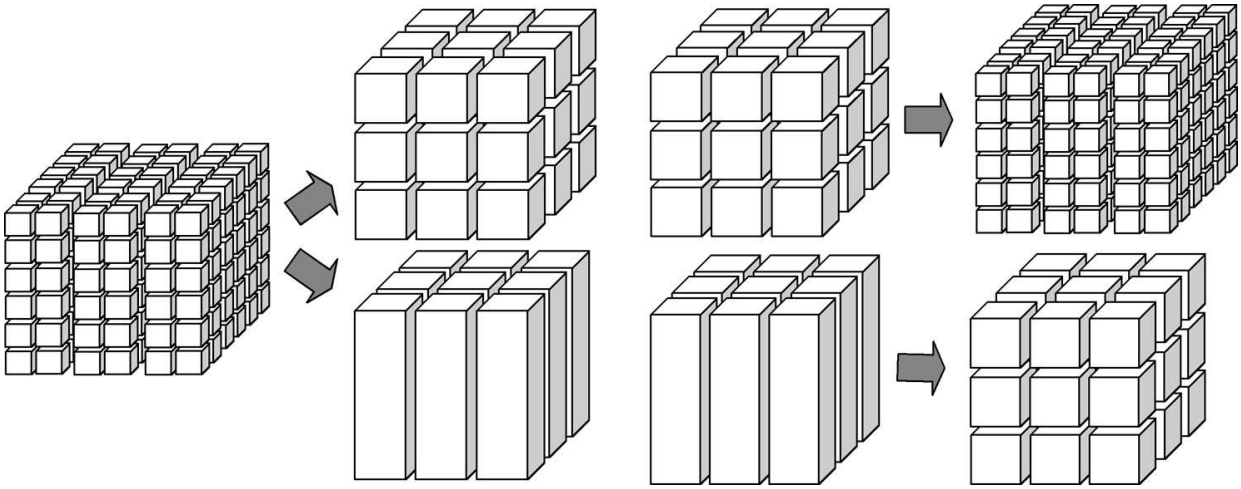

Figura 6 – Exemplo de operações *Drill Down*(direita) e *Drill up*(esquerda) extraídos de Golfarelli (2009)

- ***Slice and Dice:*** Técnica com filosofia parecida à cláusula *where* usada em *SQL*. Permite que sejam criadas restrições na análise dos dados (TIMES, 2012). O *Slice* faz restrição de um valor ao longo de uma dimensão, já o *Dice* faz restrições de valores em várias dimensões. Semelhante ao *Slice*, só que mais complexo.

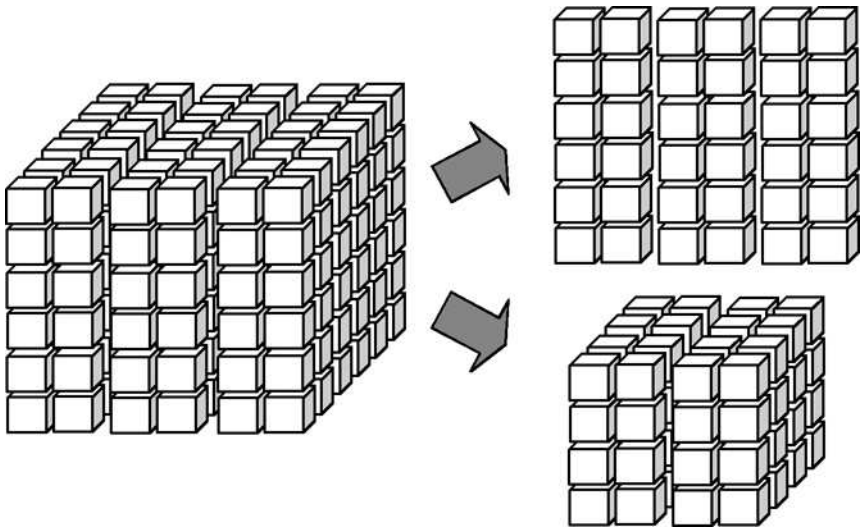

Figura 7 – Exemplo de operações *Slice*(acima) e *Dice*(embaixo) extraídos de Golfarelli (2009)

- ***Drill Across:*** Permite que diferentes cubos sejam concatenados (NERI, 2002). Uma operação do tipo *Drill Across* irá simplesmente unir diferentes tabelas fato através de dimensões correspondentes (KIMBALL, 1998).

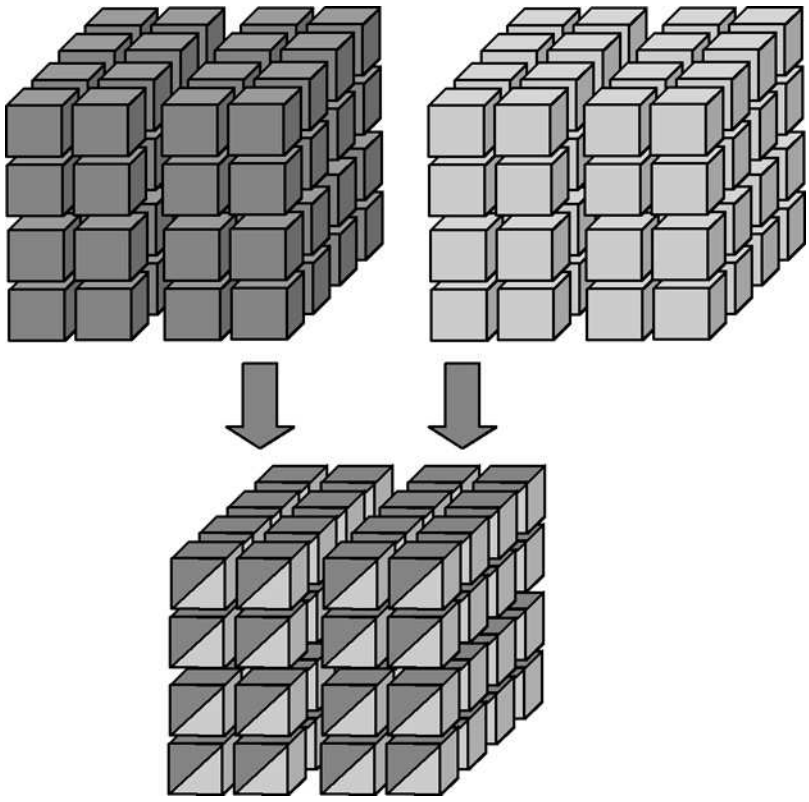

Figura 8 – Exemplo da operação *Drill Across* extraído de Golfarelli (2009)

- ***Pivoting:*** Metaforicamente, significa rotacionar o cubo. Essa técnica altera a ordenação das tabelas dimensionais (NERI, 2002).

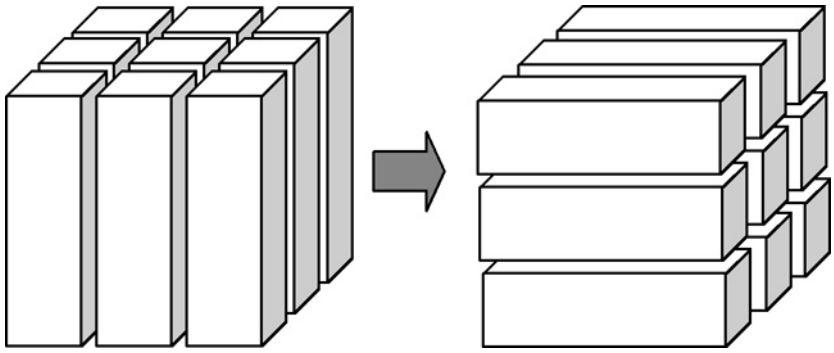

Figura 9 – Exemplo da operação *Pivoting* extraído de Golfarelli (2009)

4.4 Ambiente de *Data Warehousing* para Métricas de Código-Fonte

Para a implementação do ambiente de Data Warehousing para métricas de código-fonte, Rêgo (2014) definiu a arquitetura tal como se mostra Figura 4.

A escolha da ferramenta de análise estática de código-fonte escolhida por Rêgo (2014) foi a Analizo. A Analizo possibilita emitir saídas das métricas em CSV que detalham nome da classe e as respectivas métricas e permitiu ao a seu trabalho incorporar a análise das métricas ANPM, AMLOC, CBO, NPA.

Rêgo (2014) seguiu a metodologia de (KIMBALL; ROSS, 2002), apresentada na Figura 10, para o seu projeto de *data warehouse*. Seguindo esta metodologia Rêgo (2014) definiu os seguintes requisitos de negócios que a sua solução deveria suportar:

- **Requisito 1:** Visualizar o intervalo qualitativo obtido para cada métrica de código-fonte em uma determinada *release* do projeto para a configuração O*pen JDK8 Metrics*.

- **Requisito 2:** Comparar o intervalo qualitativo obtido para cada métrica de código-fonte ao longo de todas as *releases* de um projeto para a configuração *Open JDK8 Metrics*.

- **Requisito 3:** Visualizar o o valor percentil obtida para cada métrica de código-fonte em uma determinada *release* do projeto para a configuração *Open JDK8 Metrics*.

- **Requisito 4:** Comparar o o valor percentil a para cada métrica de código-fonte ao longo de todas as *releases* para a configuração *Open JDK8 Metrics*.

- **Requisito 5:** Visualizar o intervalo qualitativo obtido para cada métrica de código-fonte em uma determinada *release* do projeto para a configuração *Tomcat Metrics*.
- **Requisito 6:** Comparar o intervalo qualitativo obtido para cada métrica de código-fonte ao longo de todas as *releases* de um projeto para a configuração *Tomcat Metrics*.
- **Requisito 7:** Visualizar a medida obtida para cada métrica de código-fonte em uma determinada *release* do projeto para a configuração *Tomcat Metrics*.
- **Requisito 8:** Comparar o valor percentil obtido para cada métrica de código-fonte ao longo de todas as *releases* para a configuração *Tomcat Metrics*.
- **Requisito 9:** Visualizar a quantidade de cenários de limpeza identificados por tipo de cenários de limpeza de código-fonte em cada classe ao longo de cada *release* de um projeto.
- **Requisito 10:** Comparar a quantidade de cenários de limpeza por tipo de cenários de limpeza de código-fonte em uma *release* de um projeto.
- **Requisito 11:** Visualizar o total de cenários de limpeza em uma determinada *release* de um projeto.
- **Requisito 12:** Visualizar cada uma das classes com um determinado cenário de limpeza de código-fonte ao longo das *releases* do projeto.
- **Requisito 13:** Visualizar as 10 classes de um projeto com menor número de cenários de limpeza identificados.
- **Requisito 14:** Visualizar as 10 classes de um projeto com maior número de cenários de limpeza identificados.
- **Requisito 15:** Acompanhar a Taxa de Aproveitamento de Oportunidades de Melhoria de Código-Fonte que é a divisão do total de cenários de limpeza identificados em uma *release* e o o número total de classes da mesma *release* de um projeto

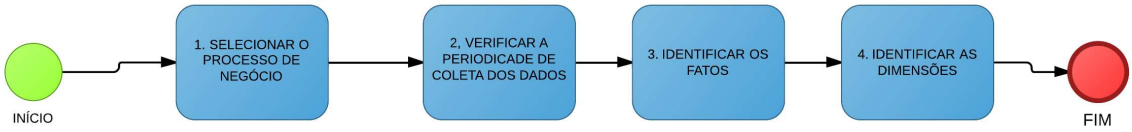

Figura 10 – Metodologia de Projeto de *Data Warehouse* proposta por Kimball e Ross (2002) extraída de Rêgo (2014)

Ainda seguindo a metodologia de (KIMBALL; ROSS, 2002), Rêgo (2014) identificou fatos e dimensões no contexto de monitoramento de métricas, como mostra a Tabela

10

Fato	Dimensões
Valor Percentil	• Projeto • Métrica • Configuração • Qualidade • *Release* • Tempo
Quantidade de Cenários de Limpeza	• Projeto • Cenário de Limpeza • Classe • *Release*
Taxa de Aproveitamento de Oportunidade de Melhoria de Código-Fonte	• Projeto • *Release*

Tabela 10 – Fatos e dimensões identificadas por Rêgo (2014)

Após a identificação dos fatos e das dimensões, Rêgo (2014) construiu o projeto físico do *Data Warehouse* que pode ser visto na Figura 11. A Tabela 11 facilita a interpretação do projeto físico.

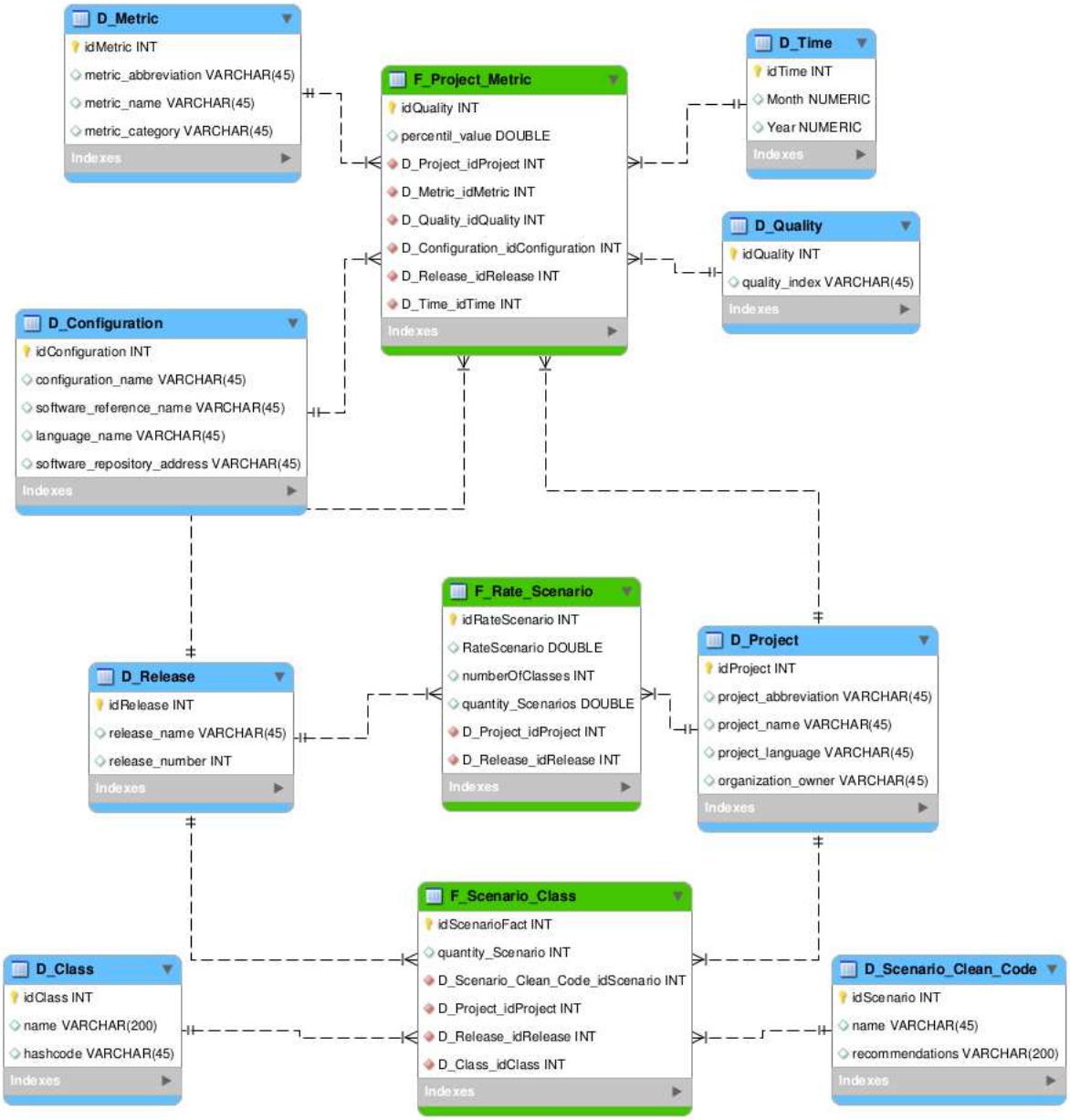

Figura 11 – Projeto físico do *Data Warehouse* extraído de Rêgo (2014)

Tabela Fato	Tabelas Dimensões
F_Project_Metric	• D_Project • D_Metric • D_Configuration • D_Quality • D_Release • D_Time
F_Scenario_Class	• D_Project • D_Scenario_Clean_Code • D_Class • D_Release
F_Rate_Scenario	• D_Project • D_Release

Tabela 11 – Tabelas fatos e tabelas dimensões elaboradas por Rêgo (2014)

Visando facilitar o processo de ETL principalmente na etapa da transformação, Rêgo (2014) criou uma área de metadados mostrada na Figura 12 com intuito de representar os dados que representam os próprios dados dos processos de negócio. A Tabela 12 descreve as tabelas contidas a área de metadados do *Data Warehouse*.

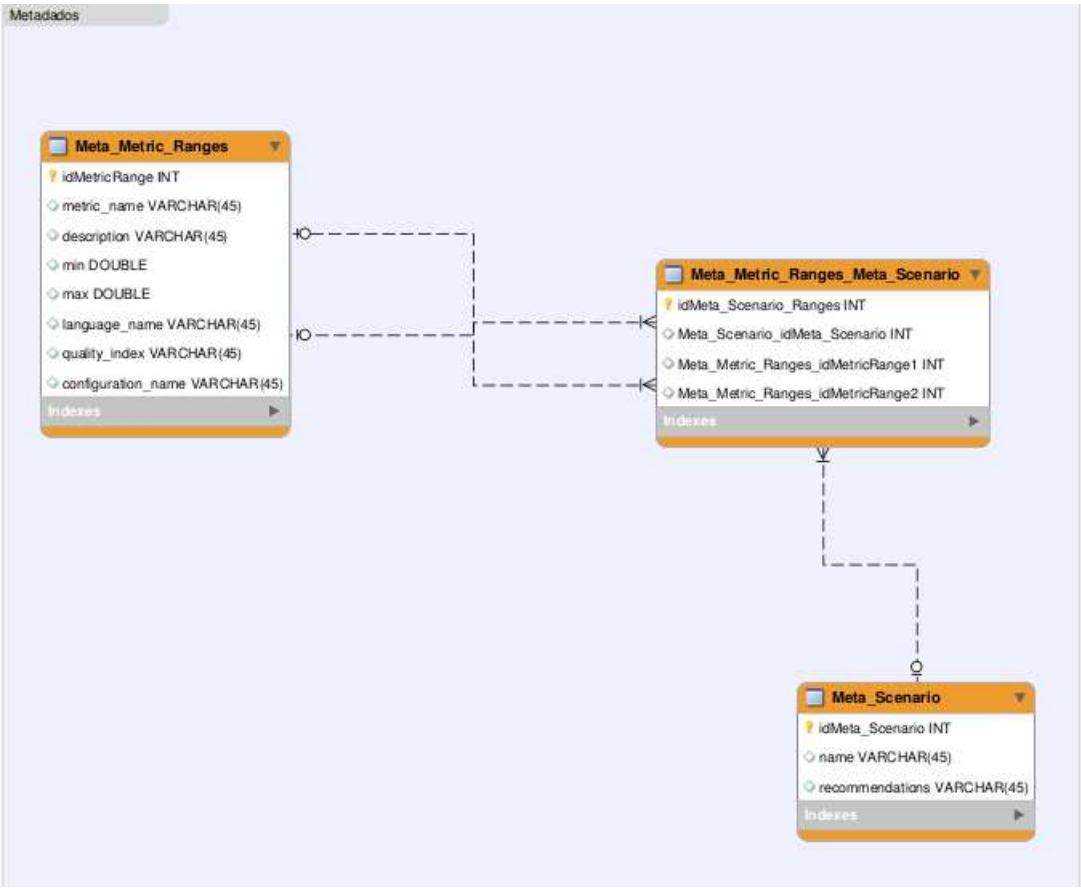

Figura 12 – Projeto físico do *Data Warehouse* extraído de Rêgo (2014)

Tabelas	Descrição
Meta_Metric_Ranges	Contém cada configuração de intervalo qualitativo para cada métrica de código fonte.
Meta_Scenario	Contém os cenários de limpeza de código e suas recomendações.
Meta_Metric_Ranges_Meta_Scenario	Como a cardinalidade entre as tabelas *Meta_Scenario* e *Meta_Metric_Ranges* seria de n para n, essa tabela foi criada contendo os registros únicos dessas duas tabelas.

Tabela 12 – Descrição das Tabelas do Metadados do *Data Warehouse*

4.5 Considerações finais do capítulo

Nesse capítulo foi apresentada a solução proposta no trabalho de Rêgo (2014), bem como a base teórica para sua compreensão. No próximo capítulo será apresentado o projeto de estudo de caso que visa a análise da eficácia e eficiência da solução de DW proposta nesse capítulo.

5 Projeto de estudo de Caso

Este capítulo irá apresentar o projeto do estudo de caso que foi realizado no passo planejar estudo de caso, este passo consiste na elaboração de um protocolo para o estudo de caso, na identificação do problema e na definição das questões e objetivos de pesquisa. O método para coleta dos dados e como eles serão analisados também serão expostos neste capítulo.

5.1 Definição

Segundo (YIN, 2001) o estudo de caso é um conjunto de procedimentos pré-especificados para se obter uma investigação empírica que investiga um fenômeno contemporâneo dentro de seu contexto da vida real, especialmente quando os limites entre o fenômeno e o contexto não estão claramente definidos. Uma grande vantagem do estudo de caso é a sua capacidade de lidar com uma ampla variedade de evidências - documentos, artefatos, entrevistas e observações - além do que pode estar disponível no estudo histórico convencional. Além disso, em algumas situações, como na observação participante, pode ocorrer manipulação informal.

Buscando maior entendimento a respeito do estudo de caso proposto, foram criadas algumas perguntas que são fundamentais para o seu entendimento:

- Qual o escopo do estudo de caso?
- Qual o problema a ser tratado?
- Qual a questão de pesquisa relacionada a esse problema?
- Quais são os objetivos a serem alcançados nessa pesquisa?
- Como foi a seleção do estudo de caso?
- Qual é fonte dos dados coletados nessa pesquisa e qual o método de coleta?

Com o objetivo da elucidação do escopo do estudo de caso proposto neste trabalho apresentado na Figura 13,foi apresentado nos capítulos antecedentes, um estudo teórico relacionado à métricas de software, contratação de serviços de TI em administrações públicas brasileiras e de Data *Warehouse*. Também foi apresentado uma solução para o monitoramento de Métricas de Código-Fonte com suporte de um ambiente de Data *Warehousing*.

Portanto o fundamento da proposta deste trabalho está ligado à análise e coleta de

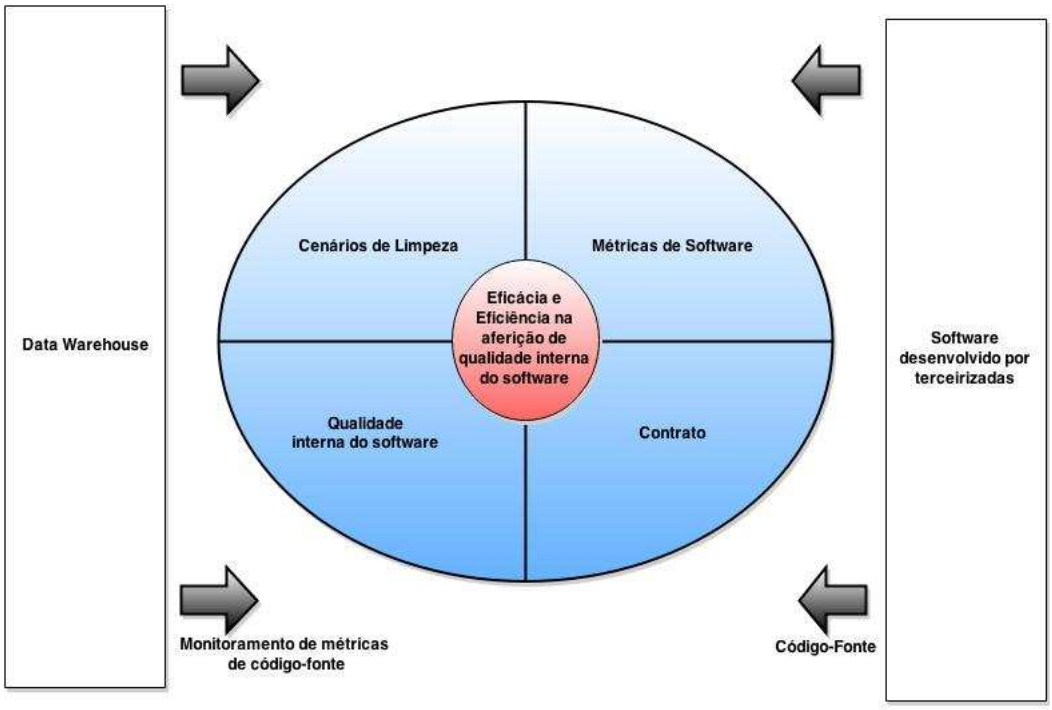

Figura 13 – Escopo do Estudo de Caso

métricas de código-fonte e de cenários de limpeza de um *software*, desenvolvido por uma empresa terceirizada contratada por um órgão público brasileiro, utilizando uma solução de DW. Onde o foco será será evidenciar a eficácia e a eficiência da proposta por Rêgo (2014) no aferimento de qualidade do software adquirido de uma empresa terceirizada pela equipe responsável do Órgão público e seus demais envolvidos.

Para as demais perguntas serem respondidas, foi estruturado neste trabalho um problema, uma questão de pesquisa, objetivos e questões específicas conforme ilustrado na Figura 14.

5.1.1 Problema

O problema refere-se ao problema de pesquisa identificado na seção "Problema"do capítulo um.

A falta de capacidade da administração pública de aferir a qualidade interna dos produtos de software adquiridos e desenvolvidos por empresas contratadas.

5.1.2 Questão de Pesquisa

A questão de pesquisa refere-se a questão de pesquisa identificada na seção "Questão de Pesquisa"do capítulo um.

O uso de DW para o monitoramento de métricas de código fonte para

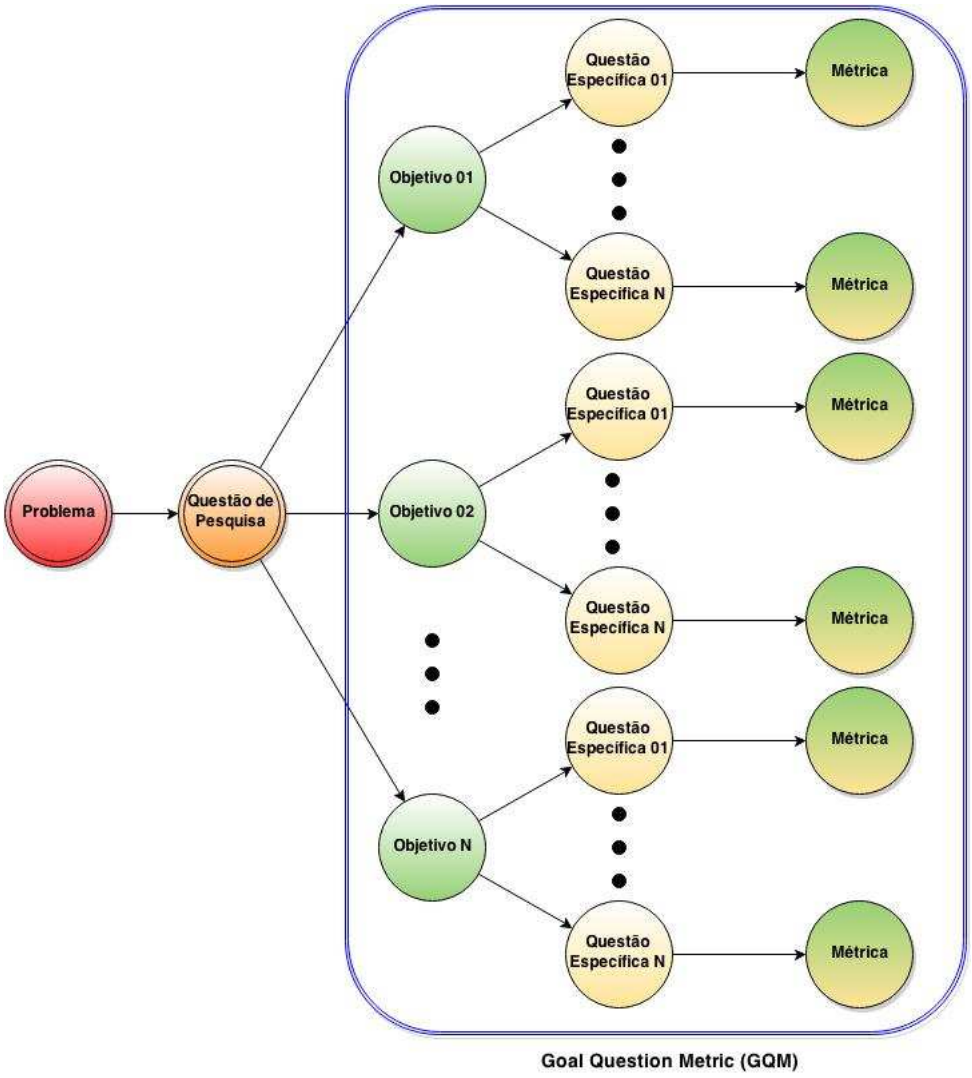

Figura 14 – Estrutura do Estudo de Caso

assistir ao processo de aferição de qualidade interna dos produtos de software desenvolvido por terceirizadas, do ponto de vista da equipe de qualidade no desenvolvimento de software no órgão público, é eficaz e eficiente?

Para atender a questão de pesquisa foi utilizado o mecanismo *goal-question-metrics*. Segundo (BASILI; CALDIERA; ROMBACH, 1996) o resultado da aplicação do GQM é a especificação de um sistema de medida destinada a um conjunto especifico de questões e um conjunto de regras para a interpretação dos dados da medição. O GQM é uma estrutura hierárquica que começa com um objetivo que especifica o propósito da medição, objeto a ser medido, questão a ser medida e o ponto de vista de que a medida seja tomada. O objetivo é refinado em várias questões, cada questão é transformada em métrica(objetiva ou subjetiva)(BASILI; CALDIERA; ROMBACH, 1996). A estrutura do GQM está a seguir.

Objetivo 01: Avaliar a eficácia e eficiência do uso de *Data Warehouse* para mo-

nitoramento de métricas de código fonte.

QE01: Quantas tomadas de decisão foram realizadas pela equipe baseando-se no uso da solução desenvolvida em um <período de tempo>?

Fonte: Registro de observação em campo.

Métrica: Número de decisões tomadas/tempo.

QE02: Quantas tomadas de decisão ao todo foram realizadas pela equipe baseando-se no uso da solução desenvolvida?

Fonte: Registro de observação em campo.

Métrica: Número de decisões tomadas.

QE03: Qual a avaliação da equipe de qualidade quanto a detecção de cenários de limpeza de código?

Fonte: Questionário com equipe de qualidade.

Métrica: muito bom, bom, regular, ruim, muito ruim.

QE04: Com que frequência a equipe de qualidade encontra falhas relacionados à utilização da ferramenta em um determinado intervalo de tempo?

Fonte: Registro de observação em campo.

Métrica: Quantidade de falha / tempo (release, sprint).

Interpretação da métrica: Quanto mais próximo de zero melhor 0=<X

QE05: Qual a quantidade total de falhas encontradas pela equipe de qualidade relacionadas à utilização da ferramenta?

Fonte: Registro de observação em campo (áudio/vídeo).

Métrica: Quantidade de falhas.

QE06: Qual a proporção do uso da ferramenta para tomada de decisões?

Fonte: Questionário com equipe de qualidade.

Métrica: Número de decisões tomadas / número de vezes que a solução foi usada.

QE07: Qual a quantidade de cenários que foram corrigidos após utilização da

solução?

Fonte: Código fonte.

Métrica: Números de cenários corrigidos / número de cenários encontrados.

QE08: Qual o nível de satisfação do uso da solução em comparação à solução anterior?

Fonte: Equipe de qualidade.

Métrica: Muito satisfeito, Satisfeito, Neutro, Insatisfeito, Muito Insatisfeito.

QE09: Qual a taxa de oportunidade de melhoria de código em uma intervalo de tempo (sprint, release)?

Fonte: Código fonte.

Métrica: Taxa de oportunidade de melhoria de código:

$$T_r = \frac{\sum_{i=1}^{n} Ce_i}{\sum_{i=1}^{n} Cl_i}$$

Ce é o total de cenários de limpezas identificados e Cl é o total de classes em um intervalo de tempo (sprint, release).

Interpretação da métrica:

- Número de classes crescente e constante, Número de oportunidade de melhoria estável: Projeto cresceu mais dos que o cenários, cenários podem ou não estar sendo tratados.
- Número de classes crescente e constante, Número de oportunidade de melhoria crescendo: Projeto cresce junto com cenários que não são tratados com eficiência ou não são tratados
- Número de classes crescente ou não, mas constante, número de oportunidade de melhoria diminuindo: Projeto cresce e cenários são tratados com eficiência.

5.2 *Background*

Referente ao contexto deste trabalho, o principal *background* é o trabalho desenvolvido por Rêgo (2014), no qual utilizaremos sua solução para monitorar as métricas de código-fonte utilizando *Data Warehouse*. Novello (2006) utilizou *Data Warehouse* para monitoramento de métricas no processo de desenvolvimento de software, porém são métricas voltadas para a qualidade do processo de desenvolvimento de software mas são diretamente relacionadas ao código-fonte. Portanto o levantamento bibliográfico realizado, não

foram encontrados estudos que utilizassem *Data Warehouse* para aferir qualidade interna do *software* aplicando coleta de métricas de código-fonte e geração de cenários de limpeza assim como foi feito por Rêgo (2014).

5.3 Seleção

Os dados foram coletados na CAIXA porque ??????????????????????????????

5.4 Fonte dos dados coletados e método de coleta

Os dados deste estudo de caso,serão coletados através de registros de observações em campo, questionários, entrevistas e resultados gerados pela própria solução utilizando o código-fonte de um ou mais sistemas desenvolvidos por empresas contratadas pelo órgão público.

Os registros oriundos de observações em campo são gerados a partir da equipe responsável pela tomada de decisões de qualidade e são coletados durante as visitas realizadas no órgão público. Nessas visitas, a solução proposta é utilizada e qualquer atitude relacionada ao seu uso pela equipe é registrada e será do ponto de vista qualitativo e quantitativo.

A adoção de questionários também foi utilizada tanto para dados qualitativos quanto para dados quantitativos á respeito da solução de DW vista pela equipe responsável e os demais envolvidos na qualidade de *software* no órgão público, no que diz respeito à tomadas de decisões, detecção de cenários de limpeza de código, falhas relacionados à utilização da ferramenta adotada na solução de DW, ao nível de satisfação no uso da solução e as taxas de oportunidades de melhoria de código.

Os gráficos e resultados gerados pela solução de DW, serão coletados a medida que a solução for utilizada pela equipe de qualidade ao longo das releases dos projetos adotados.

5.5 Ameaças a validade do estudo de caso

Yin (2001) descreve como principais ameaças relacionadas à validade do estudo de caso as ameaças relacionadas à validade do constructo, à validade interna, à validade externa e à confiabilidade. As quatro ameaças definidas por ele, bem como a forma usada nesse trabalho para preveni-las, são descritas da seguinte maneira:

- **Validade do Constructo:** A validade de construção está presente na fase de coleta de dados, quando deve ser evidenciado as múltiplas fontes de evidência e a coleta de um conjunto de métricas para que se possa saber exatamente o que medir e quais

dados são relevantes para o estudo, de forma a responder as questões de pesquisa (YIN, 2001). Buscou-se garantir a validade de construção ao definir objetivos com evidências diferentes. Estas, por sua vez, estão diretamente relacionadas com os objetivos do estudo de caso e os objetivos do trabalho.

- **Validade interna:** Para Yin (2001) o uso de várias fontes de dados e métodos de coleta permite a triangulação, uma técnica para confirmar se os resultados de diversas fontes e de diversos métodos convergem. Dessa forma é possível aumentar a validade interna do estudo e aumentar a força das conclusões. A triangulação de dados se deu pelo resultado da solução de *Data Warehouse* que utiliza o código-fonte (explicada no capítulo 4), pela análise de questionários e pelos dados coletados através de entrevistas.

- **Validade externa:** Por este ser um caso único, a generalização do estudo de caso se dá de maneira pobre (YIN, 2001). Assim é necessária a utilização do estudo em múltiplos casos para que se comprove a generalidade dos resultados. Como este trabalho é o primeiro a verificar a eficácia e eficiência da solução para o estudo de caso no órgão, não há como correlacionar os resultados obtidos a nenhum outro estudo.

- **Confiabilidade:** Com relação a confiabilidade, Yin (2001) associa à repetibilidade, desde que seja usada a mesma fonte de dados. Nesse trabalho o protocolo de estudo de caso apresentado nessa seção garantem a repetibilidade desse trabalho e consequentemente a validade relacionada à confiabilidade.

5.6 Processo de análise dos dados

Análise dos dados coletados durante o estudo de caso a ser realizado no TCU? será feita através de 4 etapas:

- **Categorização:** Organização dos dados em duas categorias - qualitativos e quantitativos. Os dados qualitativos referem-se aos questionários realizados. Os dados quantitativos, por sua vez, referem-se aos valores numéricos da solução de DW para monitoramento de métricas.

- **Exibição:** Consiste na organização dos dados coletados para serem exibidos através de gráficos, tabelas e texto para poderem ser analisados.

- **Verificação:** Atestar padrões, tendências e aspectos específicos dos significados dos dados. Procurando assim gerar uma discussão e interpretação de cada dado exibido.

- **Conclusão:** Agrupamento dos resultados mais relevantes das discussões e interpretações dos dados anteriormente apresentados.

5.7 Considerações finais do capítulo

Esse capítulo teve como objetivo apresentar o protocolo de estudo de caso que será adotado na continuação deste trabalho. No trabalho de conclusão de curso dois serão detalhados o órgão público onde será aplicado o projeto de estudo de caso desenhado neste trabalho, os termos do contrato do órgão e a empresa contratada referentes a qualidade de *software* e serão apresentados os resultados obtidos para as questões específicas apresentadas neste capítulo assim como a interpretação dos mesmos possibilitando responder a questão de pesquisa abordada.

6 Conclusão

Referências

BASILI, V. R.; CALDIERA, G.; ROMBACH, H. D. *The Goal Question Metric Approach*. [S.l.]: Encyclopedia of Software Engineering, 1996. Citado na página 53.

BASILI, V. R.; ROMBACH, H. D. *TAME: Integrating Measurement into Software Environments*. 1987. Disponível em: <http://drum.lib.umd.edu//handle/1903/7517>. Citado na página 24.

BECK, K. *Extreme Programming Explained*. [S.l.]: Addison Wesley, 1999. Citado na página 15.

BECK, K. *Implementation Patterns*. 1. ed. [S.l.]: Addison-Wesley Professional, 2007. Citado na página 29.

BRASIL. *Acórdão-381/2011-TCU-Plenário*. [S.l.], 2011. Disponível em: <https://contas.tcu.gov.br/juris/SvlHighLight?key=ACORDAO-LEGADO-89657&texto=2b4e554d41434f5244414f2533413338312b414e442b2b4e55RELACAO-LEGADO;DECISAO-LEGADO;SIDOC;ACORDAO-RELACAO-LEGADO;>. Citado na página 16.

BRASIL. Instrução normativa número 4. 2014. Citado na página 15.

BUDD, T. *An introduction to object-oriented programming*. 3rd ed. ed. Boston: Addison-Wesley, 2002. ISBN 0201760312. Citado na página 24.

FENTON, N. E.; PFLEEGER, S. L. *Software Metrics: A Rigorous and Practical Approach*. 2 edition. ed. [S.l.]: Course Technology, 1998. 656 p. Citado na página 23.

FOWLER, M. *Refactoring: improving the design of existing code*. [S.l.]: Addison-Wesley Professional, 1999. Citado na página 15.

GOLFARELLI, M. *Data warehouse design: modern principles and methodologies*. New York: McGraw-Hill, 2009. ISBN 9780071610391. Citado 3 vezes nas páginas 42, 43 e 44.

GUTIéRREZ, A.; MAROTTA, A. An overview of data warehouse design approaches and techniques. 2000. Citado na página 41.

HARMAN, M. Why source code analysis and manipulation will always be important. In: IEEE. *Source Code Analysis and Manipulation (SCAM), 2010 10th IEEE Working Conference on*. [S.l.], 2010. Citado na página 23.

HONGLEI, T.; WEI, S.; YANAN, Z. The research on software metrics and software complexity metrics. In: . IEEE, 2009. ISBN 978-1-4244-5422-8, 978-0-7695-3930-0. Disponível em: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5385114>. Citado na página 23.

INMON, W. H. *Building the Data Warehouse*. 3rd. ed. New York, NY, USA: John Wiley & Sons, Inc., 2002. Citado na página 38.

ISO/IEC 15939. *ISO/IEC 15939: Software Engineering - Software Measurement Process*. [S.l.], 2002. Citado 2 vezes nas páginas 15 e 21.

ISO/IEC 15939. *ISO/IEC 12207: System and Software Engineering - Software Life Cycle Processes*. [S.l.], 2008. Citado na página 21.

ISO/IEC 25023. *ISO/IEC 25023: Systems and software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - Measurement of system and software product quality*. [S.l.], 2011. Citado na página 7.

ISO/IEC 9126. *ISO/IEC 9126-1: Software Engineering - Product Quality*. [S.l.], 2001. Citado na página 22.

KAN, S. H. *Metrics and models in software quality engineering*. [S.l.]: Addison Wesley, 2002. Citado 2 vezes nas páginas 23 e 25.

KIMBALL, R. *The data warehouse lifecycle toolkit: expert methods for designing, developing, and deploying data warehouses*. [S.l.]: Wiley. com, 1998. Citado na página 43.

KIMBALL, R.; ROSS, M. *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling*. 2nd. ed. New York, NY, USA: John Wiley & Sons, Inc., 2002. ISBN 0471200247, 9780471200246. Citado 5 vezes nas páginas 39, 40, 41, 44 e 45.

LAIRD, M. C. B. L. M. *Software measurement and estimation: A practical approach*. [S.l.]: Wiley-IEEE Computer Society Press, 2006. Citado 2 vezes nas páginas 24 e 25.

LEITE, J. C. Terceirização em informática sob a ótica do prestador de serviços. *Revista de Administração de Empresas*, v. 37, n. 4, 1997. Disponível em: <http://www.scielo.br/pdf/rae/v37n4/a08v37n4>. Citado na página 15.

MACHINI, J. a. et al. *Código Limpo e seu Mapeamento para Métricas de Código-Fonte*. [S.l.]: Universidade de São Paulo, 2010. Citado 6 vezes nas páginas 29, 30, 31, 32, 33 e 34.

MARTIN, R. C. *Clean Code: A Handbook of Agile Software Craftsmanship*. [s.n.], 2008. 464 p. ISBN 9780132350884. Disponível em: <http://portal.acm.org/citation.cfm?id=1388398>. Citado na página 29.

MCCABE, T. J. A Complexity Measure. *IEEE Transactions Software Engineering*, v. 2, n. 4, p. 308–320, December 1976. Citado na página 24.

MEIRELLES, P. R. M. *Monitoramento de métricas de código-fonte em projetos de software livre*. Tese (Doutorado) — Instituto de Matemática e Estátistica – Universidade de São Paulo (IME/USP), 2013. Citado 5 vezes nas páginas 21, 24, 25, 26 e 27.

MILLS, E. E. Metrics in the software engineering curriculum. *Ann. Softw. Eng.*, J. C. Baltzer AG, Science Publishers, Red Bank, NJ, USA, v. 6, n. 1-4, abr. 1999. ISSN 1022-7091. Citado 2 vezes nas páginas 22 e 23.

NERI, H. R. *Análise, Projeto e Implementação de um Esquema MOLAP de Data Warehouse utilizando SGBD-OR Oracle 8.1*. Universidade Federal da Paraíba - UFPB: [s.n.], 2002. Citado 4 vezes nas páginas 41, 42, 43 e 44.

NOVELLO, T. C. *Uma abordagem de Data Warehouse para análise de processos de desenvolvimento de software*. phdthesis — Pontifícia Universidade Católica do Rio Grande do Sul, 2006. Disponível em: <http://tardis.pucrs.br/dspace/handle/10923/1570>. Citado na página 55.

RêGO, G. B. Monitoramento de métricas de código-fonte com suporte de um ambiente de data warehousing: um estudo de caso em uma autarquia da administração pública federal. 2014. Disponível em: <http://bdm.unb.br/handle/10483/8069>. Citado 21 vezes nas páginas 15, 16, 27, 28, 30, 31, 32, 33, 34, 35, 38, 44, 45, 46, 47, 48, 49, 50, 52, 55 e 56.

ROCHA, A. B. *Guardando Históricos de Dimensões em Data Warehouses*. Dissertação (Mestrado), Universidade Federal da Paraíba - Centro de Ciências e Tecnologia, 2000. Citado na página 38.

SHARMA, N. *Getting started with data warehousing*. [S.l.]: IBM Redbooks, 2011. Citado 4 vezes nas páginas 38, 40, 41 e 42.

SOFTEX. *MPS. BR-Guia de Aquisição*. [S.l.], 2013. Disponível em: <http://www.softex.br/wp-content/uploads/2013/07/MPS.BR_Guia_de_Implementacao_SV_Parte_2_20132.pdf>. Citado na página 15.

SOFTEX. MPS BR-guia de implementação – parte 2: Fundamentação para implementação do nível f do mr-mps-sv:2012. 2013. Citado na página 21.

TIMES, V. C. *Sistemas de DW*. 2012. Universidade Federal de Pernambuco - UFPE. Disponível em: <www.cin.ufpe.br/~if695/bda_dw.pdf>. Citado 2 vezes nas páginas 41 e 42.

WILLCOCKS, L.; LACITY, M. *Global information technology outsourcing*. [S.l.: s.n.], 2001. Citado na página 15.

WOHLIN, C. et al. *Experimentation in software engineering*. [S.l.]: Springer, 2012. Citado na página 18.

YIN, R. *Estudo de caso: planejamento e métodos*. [S.l.]: Bookman, 2001. Citado 4 vezes nas páginas 17, 51, 56 e 57.

APÊNDICE A – ?