

Dimensionando *Data Marts*: Uma Adequação de uma Métrica Funcional

Angélica Toffano Seidel
Calazans
Caixa Econômica Federal
angelica.calazans@caixa.gov.br

Káthia Marçal de Oliveira,
Univ. Católica de Brasília
kathia@ucb.br,

Rildo Ribeiro dos Santos
Univ. Católica de Brasília
Rildo@ucb.br

Resumo

Estimar o tamanho de um projeto para permitir definir prazos, custos e recursos é uma necessidade contínua das empresas. Nesse sentido, várias abordagens surgiram com o objetivo de estimar o tamanho do software, destacando-se entre elas, a Análise por Pontos de Função como uma das abordagens mais utilizadas pelo mercado atualmente. Com o surgimento da tecnologia de *Data Mart* e o aumento da demanda de desenvolvimento desses sistemas, as empresas passaram a exigir também estimar o tamanho desses produtos para permitir uma melhor gerência na produção dos mesmos. Sistemas de *Data Mart*, no entanto, possuem características próprias e particularidades no desenvolvimento diferentes dos sistemas tradicionais. Dessa forma, se faz necessário a adequação de uma das abordagens de medição de sistemas tradicionais para sistemas de *Data Mart*. Neste artigo, propomos uma adequação da Análise por Pontos de Função e apresentamos os resultados da aplicação desta proposta em projetos reais da Caixa Econômica Federal.

Palavras Chaves: Métricas funcionais, Data Mart, estimativa de tamanho, Análise por pontos de Função.

Abstract

Estimate the size of a project to define time, cost and resources is a continuous necessity of the companies. In this direction, several boardings had appeared with the objective of estimate the size of software, like the Analysis for Points of Function - one of boardings more used by the market currently. With the appearing of Data Mart technology and the increase of the demand of development of these systems, the companies had started to also demand estimate of these products to allow a better management in the production of them. Data Mart systems, however, have proper characteristics and particularities in the development different of the traditional systems. So, it makes necessary the adequacy of one of the boardings of measurement of traditional systems for Data Mart systems. In this article, we consider one adequacy of the Analysis for Points of Function and present results of application of this proposal in real projects of Caixa Econômica Federal.

Key words: Functional measurement, Data Mart, size estimation, *Function Point Analysis*.

1 Introdução

A medição na engenharia de *software* é um dos fatores importantes para a geração de um produto com qualidade. Segundo [3] mensura-se para entender, controlar e aperfeiçoar o processo de produção de *software*. Medidas ajudam a visualizar o processo de desenvolvimento e manutenção de software e proporcionam o estabelecimento de *baselines* para ajudar o desenvolvimento futuro. Permitem um melhor controle dos projetos, possibilitam a incrementação de mudanças e proporcionam o aperfeiçoamento contínuo deste processo e do produto de *software*.

A mensuração do tamanho do *software* na gestão de projetos, além de proporcionar um melhor entendimento, controle e aperfeiçoamento do processo de construção de software, também está vinculada à necessidade de gerar expectativas mais realistas para o usuário, avaliar e medir resultados, conhecer melhor o patrimônio de *software*, obter e/ou melhorar estimativas de prazo, custo e recursos, gerar indicadores para tomada de decisão, avaliar o impacto da introdução de novas tecnologias e obter vários indicadores de desempenho (produtividade, qualidade do código) [12]. Dessa forma, é essencial que a mensuração de tamanho seja o mais aproximada possível da realidade, pois ela é um fator de impacto nas demais variáveis.

Desde a década de 80, várias abordagens têm sido propostas e aperfeiçoadas com o objetivo de estimar funcionalmente o tamanho de um *software*, entre elas destacam-se: a Análise por Pontos de Função (APF), proposta em 1979 e atualmente na versão 4.1 [6]; o MKII FPA, abordagem proposta em 1984 com base na APF que se encontra na versão 1.3.1, com uma visão totalmente independente da APF [13]; o Feature Points (1986), método experimental voltado para mensurar *features* de tipos especiais de software (sistemas de tempo real, com grande complexidade, CAD), o *Full Function Points*, abordagem proposta em 1997 como uma adequação da APF [1] e que evoluiu para o COSMIC Full Functions Points (1999) uma abordagem totalmente independente da APF que se propõe a mensurar qualquer tipo de software [1].

A maior parte destas propostas se propõe a medir o tamanho de qualquer tipo de *software*, independente da tecnologia. Sistemas de *Data Warehouse/Data Mart*, no entanto, são um tipo especial de software, com características particulares como, por exemplo, o fato dos usuários utilizarem o software somente para consultas e não atualização dos dados, o desenvolvimento baseado em dados existentes de outros sistemas sem gerar novos, e um processo de desenvolvimento diferente de sistemas tradicionais [11]. É necessário, portanto, adequar as abordagens definidas para sistemas tradicionais para que considerem as características específicas de *Data Warehouse/Data Mart* e com isso gerem estimativas mais reais.

Neste artigo analisamos o processo de mensuração para sistemas de *Data Mart* com relação à proposta de APF. A escolha dessa abordagem se deve ao fato da APF ser uma das abordagens funcionais mais utilizadas pelo mercado atualmente, tendo na última década 15 livros e mais de 100 artigos publicados [4] o que mostra sua maturidade.

Nas seções seguintes apresentamos, inicialmente, uma breve descrição sobre *Data Warehouse/Data Mart*, no que se refere a sua definição e processo de construção (seção 2); e sobre medição e análise por pontos de função, no que se refere a sua forma de medição (seção 3). Na seção 4, será apresentada a adequação da APF para *Data Mart*. Na seção 5 mostramos os resultados da aplicação desta adequação em alguns projetos da Caixa Econômica Federal. Finalmente, na seção 6, apresentamos as conclusões deste trabalho.

2 Características de *Data Warehouse/Data Mart*

2.1 Definição

Segundo [5], um “*Data Warehouse* é uma coleção de dados orientada por assuntos, integrada, variante no tempo, e não volátil, que tem por objetivo dar suporte aos processos de tomada de decisão”.

A tecnologia utilizada tanto no *Data Warehouse* como no *Data Mart* é a mesma, sendo que as variações que ocorrem são mínimas, mais voltadas para o volume de dados, abrangência da arquitetura e o foco [2]. Os *Data Marts* são voltados somente para uma determinada área referenciando um escopo menor, a uma unidade de negócio, a um departamento, ou a um conjunto específico dos usuários, já o *Data Warehouse* é voltado para os assuntos de toda empresa.

As principais características de um *Data Warehouse/Data Mart* são [8]: informações facilmente acessadas e consistentes, adaptabilidade e flexibilidade às mudanças, proteção destas informações e utilização das informações como base para tomada de decisões.

Existem quatro componentes separados e distintos no ambiente de *Data Warehouse* (Figura 1) [8]:

- ✓ sistemas operacionais de origem - sistemas que capturam as transações da empresa;
- ✓ *data staging area* - área de armazenamento de dados e de conjunto de processos que preparam os dados de origem para serem utilizados;
- ✓ área de apresentação de dados - local onde os dados ficam armazenados e disponíveis ao usuário final;
- ✓ e ferramentas de acesso a dados - ferramentas OLAP e de mineração de dados que permitem aos usuários utilizar os dados de uma maneira rápida, interativa, de forma fácil para executar análises mensuráveis.

Os *Data Mart* servem como fonte de dados para estas ferramentas e devem assegurar consistência, integração e precisão.

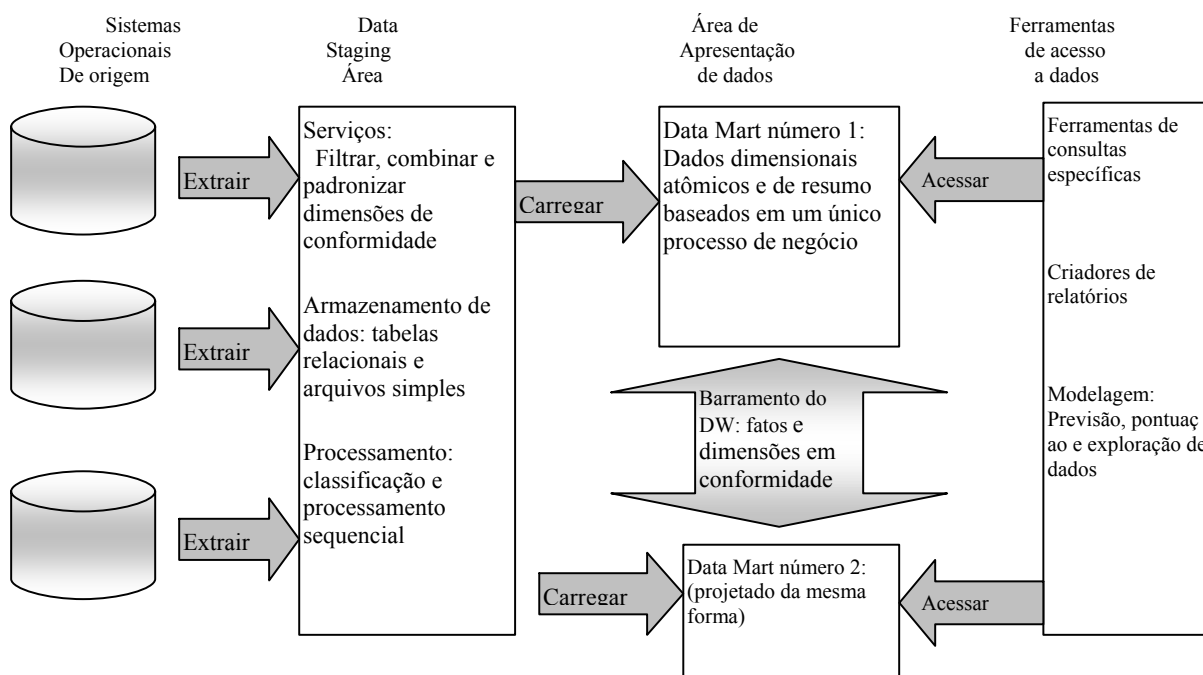


Figura 1 – Elementos básicos do *Data Warehouse* [8]

2.2 Processo de Construção

As fases básicas para se criar e atualizar um *Data Warehouse* são [8]: (i) extração, (ii) transformação e (iii) carga dos dados (*ETL – Extraction, Transformation, Load*).

O processo de extração (i) envolve a leitura e compreensão dos dados de origem e cópia destes dados na *staging area* para serem manipulados posteriormente. Normalmente, cada sistema de origem é uma aplicação independente e que possui pouco compartilhamento de dados comuns como produto, cliente e geografia com outros sistemas operacionais da empresa. A integração destes dados é uma das tarefas que geram mais esforço no projeto de um *Data Warehouse*. A quantidade de sistemas transacionais envolvidos, suas estruturas de dados¹ e o nível de documentação (o *Data Mart* necessita apresentar todos os conceitos e as origens dos dados) interferem diretamente na dimensão do sistema de *Data Mart*. O processo de extração pode ser realizado de forma automatizada através de ferramenta de ETL. A existência ou não desta ferramenta também impacta o tamanho do produto seja na geração de um maior número de funcionalidades para a extração destes dados ou na exigência de conhecimento profundo, por parte dos desenvolvedores do *Data Mart*, das regras de negócio dos sistemas transacionais e definição de formas de extração.

Na fase de transformação (ii) modifica-se a estrutura do armazenamento de dados. Nesta fase ocorrem ”transformações em potencial, como filtragem dos dados (correções de erros de digitação, solução de conflitos de domínio, tratamento de elementos ausentes ou a divisão em formatos padrão), combinação de dados de várias origens, cancelamento de dados duplicados e atribuições de chaves” [8]. Nesta fase também podem ser aplicados níveis de desnормalização e renormalização², combinação³, auditoria no conteúdo de dados⁴ e agregações necessárias para melhorar o desempenho das consultas para o usuário final (considerando a previsão de volume de dados). Toda esta transformação ocorre na *staging area* ou *Operational data storage* (ODS) (se a arquitetura da solução envolver este componente) e também impacta no tamanho de um projeto de *Data Mart*.

A fase de carga (iii) é um processo iterativo, pois o *Data Warehouse* tem que ser povoado continuamente e refletir de forma incremental as mudanças dos sistemas operacionais. Manutenções que possam ocorrer nas fontes de dados interferem diretamente na dimensão do projeto, pois além das transformações precisarem ser re-definidas e aplicadas, a carga também é alterada a cada modificação das fontes de dados das origens. A carga é a última etapa do processo de ETL e é realizada no banco de dados do DW, na área de apresentação de dados.

Neste banco de dados (que pode ser desenvolvido em uma tecnologia de banco de dados multidimensional ou relacional) os dados são armazenados em cubos. Um modelo multidimensional possui três elementos básicos: fatos, que são definidos como a coleção de itens de dados, composta de dados de medidas e de contexto, representando um item de negócio, uma transação de negócio ou um evento de negócio; dimensões que são os

¹ Definição da estrutura em que estão os dados de origem: VSAM, Banco de Dados Relacional (DB2, Sybase, Oracle, etc), Banco de dados hierárquico (IDMS), etc.

² Reunificação das hierarquias de dados, separadas pela normalização dentro de uma tabela desnормalizada.

³ Realizada quando fontes de dados possuem os mesmos valores de chaves representando registros iguais ou complementares ou atributos de chaves não iguais, incluindo equivalência textual de códigos de sistemas de legados distintos.

⁴ O processo de transformação deve realizar constantes verificações de somas, contagens de linhas e testes.

elementos que participam de um fato e determinam o contexto de um assunto de negócios e medidas que são atributos numéricos que representam um fato [9].

3 Medição de Software

Medição é o processo através do qual números ou símbolos são atribuídos a entidades do mundo real de forma a tornar possível caracterizar cada entidade através de regras claramente definidas [3], ou seja, é o processo de obtenção de uma medida para uma entidade do mundo real. Uma medida fornece uma indicação de quantidade, dimensão, capacidade ou tamanho de algum produto de software ou de um processo. Em outras palavras uma medida refere-se a um valor de uma métrica. Segundo [7], métrica é a composição de métodos para medição e escalas de medição.

Para se chegar a uma medida de software, existem técnicas de estimativas que avaliam as variáveis de tamanho, esforço e prazo. Estas técnicas podem ser classificadas basicamente em Analógicas (baseada na experiência de quem faz estimativas), Modelos Algoritmos (que considera modelos matemáticos, por exemplo, o LOC que pontua o número de instruções fontes (número de linhas de código), regressão linear, Halstead) e Análise de Funcionalidade (baseada nas funcionalidades do software, por exemplo, a APF) [12].

Algumas das principais abordagens utilizadas para análise de funcionalidade são a Análise por Pontos de Função, definida desde 1979 e que vem continuamente sendo utilizada e melhorada desde então, e a COSMIC-FFP [1] proposta em 1999, tornou-se um padrão em 2003, mas ainda sem grande experimentação prática.

A seguir serão descritas as principais características da APF por ser essa a métrica adequada nesse trabalho.

3.1 Análise por Pontos de Função

A Análise por Pontos de Função (APF) mede o tamanho do *software* pela quantificação de suas funcionalidades, baseadas no projeto lógico ou a partir do modelo de dados segundo a visão e os requisitos do usuário final [6]. Suas principais características são: medir o que foi requisitado e recebido do usuário, medir independente da tecnologia, ser aplicável desde o início do sistema, apoiar a análise de produtividade e qualidade e estimar o tamanho do *software* com uma unidade de medida padrão.

Os seguintes passos devem ser observados para mensuração de tamanho do *software* utilizando esta abordagem [6]:

- i) **Estabelecer o objeto da contagem** (projetos de desenvolvimento, projetos de manutenção ou contagem de uma aplicação);
- ii) **Determinar a fronteira de medição** (a fronteira de medição deve ser sempre determinada sob o ponto de vista do usuário);
- iii) **Contar as funções de dados**, divididos em Arquivos Lógicos Internos (ALIs - que são grupos lógicos de dados mantidos dentro da fronteira da aplicação) e Arquivos de Interface Externa (AIEs – arquivos somente referenciados pela aplicação);
- iv) **Contar as funções transacionais**, divididos em Entradas Externas (EEs), Saídas Externas (SEs) e Consultas Externas (CEs);
- v) **Determinar o Fator de Ajuste** (conjunto de 14 características que influenciarão a complexidade do *software*. São elas: comunicação de dados, processamento

distribuído, performance, utilização de equipamento, volume de transações, entrada de dados on-line, eficiência do usuário final, atualização on-line, processamento complexo, reutilização de código, facilidade de implantação, facilidade operacional, múltiplos locais, facilidade de mudanças); e,

- vi) **Determinar o tamanho do projeto** (considera as funções de dados, transacionais, fatores de ajuste e tipo de projeto).

Cada função de dado ou transacional terá um peso diferente dependente de sua complexidade. Diversas tabelas baseadas na quantidade de elementos de dados, de registros e de arquivos referenciados são utilizadas para determinar a complexidade de cada função em Baixa, Média ou Alta.

A Tabela 1 mostra o número de pontos de função atribuídos a cada tipo de função conforme o grau de complexidade.

Tipo de função	Baixa	Média	Alta
EE	3	4	6
SE	4	5	7
CE	3	4	6
ALI	7	10	15
AIE	5	7	10

Tabela 1 – Quantidade de pontos de função x Tipo de função x complexidade [6]

O resultado da contagem de funções de dados e transacionais é uma medida chamada de contagem não ajustada (NoPF não ajustado), pois não considera ambiente ou plataforma tecnológica que, entre outros, são detalhes que afetam o produto e sua construção. O ajuste na mensuração é efetuado através do Fator de Ajuste determinado.

A determinação do Fator de Ajuste considera a avaliação de cada característica numa escala de 0 (nenhuma influência) a 5 (grande influência). A determinação deste fator de ajuste (FA) é baseada na equação: $FA = 0,65 + (0,01 \times \text{Soma das características gerais do sistema})$.

Para determinar o tamanho do projeto consideram-se fórmulas específicas como por exemplo, para medir aplicação ou projetos de desenvolvimento, a seguinte: $\text{NoPFajustado} = \text{NoPF não ajustado} \times FA$.

4 Medição de Tamanho de Data Marts

Existem diferenças substanciais entre a construção de um *software* transacional e a construção de um produto de *Data Warehouse/Data Mart*. Além do processo ser bastante diferenciado, o resultado, o tratamento dos dados, a visão do usuário possuem características muito diferentes quando comparados a um sistema tradicional. Existem, no caso deste domínio, funcionalidades que não são visualizadas pelo usuário final e que impactam no tamanho do sistema que esta sendo mensurado, como por exemplo todas as funcionalidades geradas para tratar os dados na *staging area*.

A mensuração de um projeto de *Data Warehouse* com a abordagem APF fica prejudicada considerando que este processo, na visão do usuário, recebe dados já armazenados por outros sistemas e disponibiliza-os de forma que ferramentas adquiridas possam ser utilizadas para minerar ou consultar historicamente estes dados. A métrica APF

examinada não trata exemplos específicos para contagem de tamanho de sistemas de *Data Warehouse/Data Mart*.

Dessa forma é necessário analisar cada um dos itens que são considerados na APF e adequá-los às características de *Data Mart* apresentadas na seção 3. Apresentaremos, a seguir como adaptamos a APF considerando os seis (6) passos de medição apresentados na seção anterior.

No que se refere a **(i) estabelecer o objeto da contagem** não é necessário nenhuma adequação, pois identificar o objeto da contagem segue os mesmos padrões de um desenvolvimento de um sistema transacional.

Com relação a **(ii) Determinar a fronteira de medição** do aplicativo é necessário verificar se os dados de origem são fornecidos pelos sistemas de origem e neste caso a fronteira de medição fica restrita as tabelas internas do projeto de *Data Mart*. Não serão computadas funções de dados para os arquivos dos sistemas de origens. Se os dados de origem não são fornecidos pelos sistemas de origem, caberá a equipe responsável pelo sistema de *Data Mart* gerar a extração, e neste caso os arquivos dos sistemas de origens serão pontuados como AIE, tornando a fronteira de medição do *Data Mart* mais ampla. No caso da utilização de uma ferramenta de ETL, que automatiza todo o processo de extração, sugere-se que a fronteira da medição fique restrita as tabelas internas do projeto de *Data Mart*.

No que se refere a **(iii) contar as funções de dados** deve ser considerado que o usuário possui a visão das dimensões que necessitará para suas pesquisas, e que estas visões estão muito interrelacionadas aos fatos, todos os fatos e dimensões devem ser pontuados como ALIs. Como AIE serão pontuados os dados corporativos utilizados no projeto.

O usuário também possui a visão de que estes dados deverão ser tratados, limpos, agregados, sumarizados em uma área antes de serem disponibilizados para consulta. Com base nesta visão e considerando também a necessidade de se computar os dados da *staging área*, sugerimos que para todos os ALI computados inicialmente sejam também computados ALI para a *staging área*. Os dados da *staging área* são dados que permanecem e que são utilizados constantemente para as novas cargas e atualizações. Normalmente, podem ser criados mais de um arquivo na *staging area* para cada dimensão e fato, mas inicialmente será inferido somente a existência de um para cada ALI.

A definição da complexidade para cada função de dado será aplicada conforme a proposta da APF.

Para **(iii) contar as funções transacionais** deve-se para cada ALI considerar uma EE, pois eles são atualizados a partir dos dados de sistemas operacionais que funcionam como uma EE. Na realidade o processo de carga é muito mais complexo e gera muito mais processos do que apenas um como está sendo sugerido, mas considerando a visão do usuário sugerimos a definição de uma EE para cada ALI. Com relação a SE ou CE sugere-se que sejam computadas qualquer solicitação de relatórios/consultas/view pré formatadas para facilitar a consulta do usuário final, respeitando-se a distinção entre SE e CE da proposta AFP. A definição da complexidade para cada função transacional será aplicada conforme a proposta da APF.

O passo referente a **(iv) Determinar os Fatores de ajuste** implicou numa análise cuidadosa dos fatores de ajuste propostos na APF no que se refere à *Data Marts*. Como resultado dessa análise percebemos que dos 14 fatores de ajuste:

- 4 fatores são aplicáveis a este tipo de software: Processamento distribuído de dados, Desempenho, Reusabilidade de Código e Facilidade Operacional.

- 2 fatores poderiam ser adaptados para *Data Warehouse/Data Mart*, são eles:
 - Eficiência do usuário final⁵ que poderia ser adequado para considerando a Quantidade de agregação⁶ onde a definição dos níveis de agregação necessários tem como objetivo proporcionar eficiência para o usuário final.
 - Processamento complexo⁷ que poderia ser adequado para Qualidade dos dados⁸ onde a quantidade de tratamento (de dados e das exceções) necessário ao projeto pode ser comparada como um nível de complexidade do processo.

Os demais fatores (no total de 8) estão intrinsecamente relacionados a sistemas transacionais o que implica em quando analisados no contexto da *Data Warehouse/ Data Mart* sempre receberiam o valor 0 (nenhuma influência). Por exemplo, entrada de dados on-line⁹ e atualização on-line¹⁰ pois no contexto de *Data Warehouse/Data Mart* não existem entradas nem atualizações on-line; e ainda múltiplos locais¹¹ considerando que o *Data Warehouse/Data Mart* não é instalado em nenhum local.

Baseados nessa análise resolvemos considerar os 4 fatores realmente aplicáveis, substituir os 2 fatores possíveis de adequar por nomes mais pertinentes, e propor novos fatores de ajustes que representem as características de *Data Mart*. Para cada um dos fatores, foram definidos os níveis de influência numa escala de 0-5 conforme proposto na APF. A proposta final de adequação dos fatores de ajuste para *Data Mart* pode ser visualizada na Tabela 2.

Finalmente para **(vi) determinar o tamanho do projeto** utilizamos as fórmulas propostas na APF.

Fatores de Ajuste e comentários sobre aplicabilidade	Níveis de influência
1 Processamento distribuído de dados Aspectos relacionados com processamento e funções distribuídas. Comentários: Leitura via client ou via Internet ou Intranet pode receber o valor 2 a 4. Aplicável. O processo de <i>Data Mart</i> prepara dados para leitura do usuário final em outra ferramenta, no caso uma ferramenta OLAP, na intranet.	0. O aplicativo não auxilia na transferência de dados ou funções entre os processadores envolvidos; 1. O aplicativo prepara dados para que o usuário final os utilize em outro processador (planilhas de calculo, por exemplo); 2. O aplicativo prepara dados e os transfere para que outros equipamentos os utilizem; 3. O processamento é distribuído e a transferência de dados acontece de forma on-line apenas em uma direção; 4. O processamento é distribuído e a transferência de dados acontece de forma on-line em ambas as direções; 5. As funções de processamento são dinamicamente executadas no equipamento mais apropriado.
2 Desempenho/Performance Aspectos relacionados a parâmetros estabelecidos pelo usuário quanto a	0. Nenhum requerimento especial de performance foi solicitado pelo usuário; 1. Requerimentos de performance foram estabelecidos e revistos, mas nenhuma ação especial foi requerida;

⁵ Aspectos relacionados com a eficiência do aplicativo na interação com o usuário.

⁶ Definição dos níveis de agregação necessários de forma a melhorar o desempenho das consultas do usuário final.

⁷ Aspectos relacionados com a complexidade do processamento.

⁸ Descreve o grau previsto para tratamento de exceções identificadas inicialmente com relação à qualidade de dados, dados rejeitados, erro de conteúdo, etc.

⁹ Aspectos relacionados com a quantidade de entrada de dados on-line do aplicativo

¹⁰ Aspectos relacionados com a quantidade de atualização on-line dos arquivos lógicos internos.

¹¹ Aspectos relacionados à arquitetura do aplicativo e a necessidade de instalação em vários lugares.

Fatores de Ajuste e comentários sobre aplicabilidade	Níveis de influência
<p>tempos de resposta.</p> <p>É aplicável ao projeto de <i>Data Mart</i> que trabalha com volumes altos de transações, mesmo que sejam transações efetuadas por ferramentas OLAP.</p> <p>Dependendo do projeto seriam indicados os níveis 4 ou 5.</p>	<ol style="list-style-type: none"> 2. Tempo de resposta e volume de processamento são itens críticos durante horários de pico de processamento. Porém, nenhuma determinação especial foi estabelecida quanto à utilização do processador. A data limite para a disponibilidade do processamento é sempre o próximo dia útil; 3. Tempo de resposta e volume de processamento são itens críticos durante todo o horário comercial. Não há determinação especial para a utilização do processador. A data limite para a comunicação com outros aplicativos é um item importante e deve ser considerado. 4. quando, além do descrito no item 3, os requisitos de performance estabelecidos requerem tarefas de análise de performance na fase de análise e desenho do aplicativo. 5. quando, além do descrito no item 4, ferramentas de análise de performance precisam ser usadas nas fases de desenho, desenvolvimento ou mesmo na fase de implementação para que os requisitos do usuário sejam atendidos plenamente.
<p>3</p> <p>Utilização de ferramenta apropriada para extração e carga</p> <p>Indica o nível de automatização do processo de <i>Data Mart</i></p>	<ol style="list-style-type: none"> 0. quando é utilizada ferramenta para 100% do processo de extração e/ou carga; 1. quando é utilizada ferramenta para 80% do processo de extração e/ou carga; 2. quando é utilizada ferramenta para 60% do processo de extração e/ou carga; 3. quando é utilizada ferramenta para 40% do processo de extração e/ou carga; 4. quando é utilizada ferramenta para 20% do processo de extração e/ou carga; 5. quando nenhuma ferramenta é utilizada para extração e carga dos dados transacionais
<p>4</p> <p>Quantidade de sistemas transacionais envolvidos no projeto Descreve o grau em que a quantidade de interfaces com outros sistemas influenciará o desenvolvimento da aplicação.</p> <p>Quando a quantidade de sistemas transacionais é alto, influencia o projeto, desenvolvimento, implantação e suporte da aplicação.</p>	<ol style="list-style-type: none"> 1. quando o projeto envolve 1 sistema transacional 2. quando o projeto envolve de 2 a 3 sistemas transacionais; 3. quando o projeto envolve de 4 a 5 sistemas transacionais 4. quando o projeto envolve de 6 a 7 sistemas transacionais; 5. quando o projeto envolve mais de 8 sistemas transacionais.
<p>5</p> <p>Documentação dos sistemas transacionais de origem (Existência de Metadados dos sistemas de origem)</p> <p>Nível de documentação dos sistemas de origem, de forma a identificar a existência de metadados dos dados de origem.</p>	<ol style="list-style-type: none"> 0. quando todos os sistemas de origem possuem metadados; 1. quando 90% dos sistemas de origem possuem meta dados; 2. quando 70% dos sistemas de origem possuem metadados; 3. quando 50% dos sistemas de origem possuem metadados; 4. quando 30% dos sistemas de origem possuem metadados; 5. quando nenhum dos sistemas de origem possui metadados;

Fatores de Ajuste e comentários sobre aplicabilidade	Níveis de influência
<p>6 Quantidade de agregação substituindo Eficiência do usuário final Definição dos níveis de agregação necessários de forma a melhorar o desempenho das consultas do usuário final</p>	<ol style="list-style-type: none"> 0. nenhum nível de agregação identificado; 1. Um nível de agregação identificado; 2. De dois a três quantidades de agregação identificadas; 3. De quatro a cinco quantidades de agregação identificadas; 4. Seis ou sete quantidades de agregação identificadas; 5. Oito ou mais quantidades de agregação identificadas
<p>7 Frequência de atualização das fontes de dados Descreve o grau em que os sistemas transacionais são alterados implicando em constantes alterações nas aplicações de extração e carga.</p>	<ol style="list-style-type: none"> 0. quando não houver previsão; 1. quando houver atualizações de 10% a 20% dos arquivos de extração /carga; 2. quando houver atualizações de 20% a 30% dos arquivos de extração /carga; 3. quando houver atualizações de 30% a 40% arquivos de extração /carga; 4. quando houver atualizações de 40% a 50% dos arquivos de extração /carga; 5. quando houver atualizações de mais de 50% arquivos de extração /carga.
<p>8 Qualidade dos dados em substituição ao Processamento complexo Descreve o grau previsto para tratamento de exceções identificadas inicialmente com relação à qualidade de dados, dados rejeitados, erro de conteúdo, etc.</p>	<p>Ao considerar as características da aplicação verificar a necessidade de aplicabilidade dos seguintes itens:</p> <ul style="list-style-type: none"> - Integração – envolve a geração de chaves substitutas para cada registro, de modo a evitar a dependência de chaves definidas no sistema legado; - Limpeza – correção de códigos e caracteres especiais, resolvendo problemas de domínios, tratando dados perdidos e corrigindo valores duplicados ou errados; - Eliminação – eliminar campos e dados provenientes dos sistemas legados que não serão úteis ao DW. - Combinação – realizada quando fontes de dados possuem os mesmos valores de chaves representando registros iguais ou complementares ou atributos de chaves não iguais, incluindo equivalência textual de códigos de sistemas legados distintos; - Verificação de integridade referencial – significa verificar se os dados de uma tabela são iguais aos dados correspondentes em outra tabela - Desnormalização e renormalização – consiste em reunificar as hierarquias de dados, separadas pela normalização dentro de uma tabela desnormalizada; - Conversão de tipo de dados – envolve transformação de baixo nível de forma a converter um tipo de dado em outro formato - Cálculos, derivação e alocação - são transformações a serem aplicadas sobre as regras de negócio identificadas durante o processo de levantamento de requisitos; - Auditoria no conteúdo dos dados – o processo de transformação deve realizar constantes verificações de somas, contagem de linhas e testes. Tem-se: <ol style="list-style-type: none"> 0. quando não ocorrer nenhuma das características acima; 1. quando ocorrer de uma a duas das características acima; 2. quando ocorrer de três a quatro das características acima; 3. quando ocorrer cinco a seis das características acima; 4. quando ocorrer sete a oito das características acima; 5. quando ocorrer todas as características acima;

Fatores de Ajuste e comentários sobre aplicabilidade	Níveis de influência
<p>9 Reusabilidade de código Aspectos relacionados à reutilização do código do aplicativo. Aplicável</p>	<p>0. Nenhuma preocupação com reutilização de código. 1. Reutilização de código apenas no aplicativo. 2. Menos de 10% do código do aplicativo foi projetado para ser utilizado em outros aplicativos. 3. 10% ou mais do código do aplicativo foi escrito para ser utilizado em outros aplicativos. 4. O código do aplicativo foi projetado para ser utilizado em outros aplicativos. A customização deve ser realizada em nível de código-fonte. 5. O código do aplicativo pode ser reutilizado em outros aplicativos com alto grau de parametrização. É apenas necessário que o usuário altere determinados parâmetros.</p>
<p>10 Estrutura dos dados de origem Definição da estrutura em que estão os dados de origem, VSAM, Relacional(DB2, Sybase, Oracle), Hierárquico – IDMS).</p>	<p>1. quando existir uma única estrutura dos dados de origem; 2. quando existir duas estruturas dos dados de origem; 3. quando existir três estruturas dos dados de origem; 4. quando existir quatro estruturas dos dados de origem; 5. quando existir mais de quatro estruturas.</p>
<p>11 Facilidade operacional Aspectos relacionados com a facilidade de operação do aplicativo. Avalia procedimentos operacionais automáticos e mecanismos de iniciação, salvamento e recuperação de dados. Aplicável quando direcionado aos procedimentos batch de extração e carga.</p>	<p>0 Nenhuma consideração especial de operação além do processo normal de salvamento de dados; 1 a 4: quando um ou todos os itens seguintes se aplicarem (selecionar todos os que se aplicam; cada item soma um ponto): Foram desenvolvidos procedimentos de iniciação, salvamento e recuperação, mas a intervenção do operador é necessária; Foram desenvolvidos procedimentos de iniciação, salvamento e recuperação, sem a necessidade de intervenção do operador (vale dois itens); O aplicativo minimiza a necessidade de montagem de fita magnética; O aplicativo minimiza a necessidade de manuseio de papel; 5 O aplicativo foi desenhado para trabalhar sem operador. Nenhuma intervenção do operador é necessária além de iniciar e encerrar o aplicativo porque este já contém rotinas automáticas de recuperação de erros.</p>
<p>12 Volume de dados Previsão do volume de dados do projeto O volume de dados interfere no tamanho e deve ser previsto visando garantir performance</p>	<p>1 Baixo 3 Médio 5 Alto</p>
<p>13. Nível de conhecimento exigido pela equipe de <i>Data Mart</i> da base de dados/regras de negócio dos sistemas transacionais de origem (Vinculada à existência de ferramenta ETL, pois a existência obriga a equipe de <i>Data Mart</i> a conhecer todas as regras de negócio transacional e definir formas de extração).</p>	<p>1 Pouco conhecimento da equipe de <i>Data Mart</i> das regras de negócio dos sistemas transacionais 3 Médio conhecimento da equipe de <i>Data Mart</i> das regras de negócio dos sistemas transacionais 5 Alto conhecimento da equipe de <i>Data Mart</i> das regras de negócio dos sistemas transacionais</p>

Tabela 2 - Fatores propostos para sistemas de *Data Mart*

5 Aplicação da proposta de adequação e da APF

Nos últimos anos a Caixa Econômica Federal tem utilizado a Análise de Ponto de Função para mensurar o tamanho de seus sistemas. Esta métrica tem características próprias, conforme poderá ser verificado no escopo deste trabalho, e tem sido utilizada na medição de Sistemas de Informações operacionais, *Data Mart*, *Workflows*.

Foram escolhidos três projetos de *Data Mart* que já estão em produção e possuem dados com relação ao tempo de construção e quantidade de recursos envolvidos, de forma a poder se efetuar uma comparação entre as duas abordagens: a APF como ela é proposta, e a adequação da APF apresentada na seção 5.

Foram pontuados os seguintes sistemas: SIST. 1, sistema de informações gerenciais dos negócios da empresa; SIST. 2, um sistema de informações gerenciais de determinada área negocial da empresa; e, SIST. 3, um sistema de informações gerenciais de Recursos Humanos. Não foi contada nenhuma SE nem CE para os projetos em questão, pois, a empresa utiliza ferramentas *OLAP* para geração de relatórios e consultas. Estas ferramentas são utilizadas pelo usuário final que define suas próprias consultas dinamicamente. Os resultados são apresentados na Tabela 3.

Sistemas	APF									Proposta de Adequação								
	ALI	AIE	EE	SE	CE	IF	PFN	FA	PFA	ALI	AIE	EE	CE	CE	IF	PFN	FA	PFA
SIST. 1	507	7	234			14	748	0,79	590,92	1014	7	234			43	1255	1,08	1355,40
SIST. 2	252	24	116			13	392	0,78	305,76	504	24	116			39	644	1,04	669,76
SIST. 3	252	5	110			13	367	0,78	286,26	504	5	110			40	619	1,05	649,95

Legenda: IF – Itens de influencia
 PFN – Pontos de função não ajustados
 FA – Fator de ajuste
 PFA – Pontos de função ajustados

Tabela 3 – Contagem dos sistemas de *Data Mart*

Para analisar a contagem de pontos de função decidimos definir a estimativa de tempo considerando o número de recursos (tamanho da equipe) que participou do desenvolvimento. Para isso, foi considerado um fator de produtividade que é utilizado pela empresa, ou seja, o esforço (horas/pf) de 16,00 com a quantidade de 22 dias por mês, com a carga horária de 8 horas.

Na tabela 4 são apresentados os resultados da aplicação da APF e da proposta de adequação com relação ao tempo real de construção e à quantidade de recursos alocados para cada um dos sistemas.

Sistema	Qtd recurso	Tempo real	APF		Proposta de Adequação	
			Ponto Função	Tempo estimado	Ponto Função	Tempo estimado
SIST. 1	5	19 meses	590,92	10,74 meses	1355,40	24,64 meses
SIST. 2	7	8 a 9 meses	305,76	3,97 meses	669,76	8,69 meses
SIST. 3	6	10 meses	286,26	4,33 meses	649,95	9,84 meses

Tabela 4 – Comparação da abordagem APF e da proposta de adequação com relação ao tempo real

Como pode ser observada na Tabela 4, a proposta de adequação se aproxima mais ao tempo real do projeto do que a abordagem APF. O único sistema que ficou com um tempo estimado da Proposta de adequação menos aproximado ao real foi o SIST.1. Quando consultada, a equipe informou que este sistema foi construído por uma equipe altamente especializada de consultores, e inferimos que, talvez neste caso, o fator de esforço tenha sido superestimado.

6 Conclusões

Mensuração na engenharia de *software* é uma prática essencial para entender, controlar e melhorar o processo e a construção de um produto de *software* adicionando qualidade. A mensuração do porte de um sistema possibilita gerar estimativas de custo, esforço e produtividade, verificar a adequabilidade dos processos e produtos e evoluir estes processos e produtos agregando qualidade.

Uma das maiores dificuldades encontradas pela gestão de projetos é estimar o porte do que está sendo construído. Existem muitas abordagens para mensurar o tamanho de um *software* e não há uma abordagem que seja melhor que outra, sob todos os aspectos, em qualquer situação. A abordagem de mensuração de tamanho deve ser escolhida e/ou adequada dependendo das características particulares do sistema que se pretenda desenvolver ou do problema que se pretenda solucionar.

A partir da observação de que os sistemas de *Data Warehouse/Data Mart* possuem diferenças substanciais da construção de um *software* transacional iniciamos uma pesquisa de como adaptar uma abordagem de medição de tamanho (a análise por pontos de função) para esse contexto. O resultado desta investigação demonstrou que uma adequação da abordagem APF é necessária para uma mensuração de tamanho de *software* mais adequada e confiável para este domínio. A adequação e a medição de três projetos reais nos mostrou resultados promissores. Sabemos que são necessárias mais aplicações da abordagem proposta para confirmar a sua melhor adequabilidade para projetos de *Data Warehouse* e *Data Mart*, mas este pode ser o caminho inicial para a solução do problema de mensurar tamanho de *softwares* para domínios de *Data warehouse/Data Mart*.

Referências Bibliográficas

- [1] ABRAN, A., DESHARNAIS, J., OLIGNY, S., ST-PIERRE, D., SYMONS C. **Cosmic FFP Measurement Manual**, version 2.2, Ed. S.Oligny, Software Engineering Management Research Laboratory, Université du Quebec a Montreal, Canada, 1999. 81 p.
- [2] BARBIERI, C. **BI-Business Intelligence – Modelagem & tecnologia**. Rio de Janeiro: Axcel Books do Brasil Editora, 2001.
- [3] FENTON,N., PFLEEGER, S. **Software Metrics A Rigorous & Practical Approach**. Boston: PWS Publishing Company, 1997. 638 p.
- [4] GARMUS,D. **Function Points Analysis – Measurement Practices for Successful Software Projects**. Estados Unidos: Addison Wesley,2001.363 p.
- [5] INMON, W.H. , **Definition of a Data Warehouse. 1999**. Disponível em: <www.billinmon.com/library/articles/dwedef.asp. Acesso em 05 Mai 2003.
- [6] IFPUG. International Function Point Users Group. **Function Point Counting Practices Manual: Release 4.1**. Ohio: IFPUG. 2000. 1 v.
- [7] ISO/IEC 9126:2001.**Software engineering – Product quality**.2001.
- [8] KIMBALL, R., ROSS, M. **Data warehouse toolkit: o guia completo para modelagem multidimensional**.Rio de Janeiro: Campus, 2002. 494 p.
- [9] MACHADO, F. **Projeto de Data Warehouse – uma visão multidimensional**. São Paulo: Erica, 2000.
- [10] OLIGNY, S., ABRAN, A. **On the compatibility between full function points and IFPUG function points**. In: PROCEEDINGS OF THE 10TH EUROPEAN SOFTWARE CONTROL AND METRIC CONFERENCE – ESCOM, 1999. Herstmonceux Castle, Inglaterra. p.1-9.
- [11] SANTILLO, L. **Size & estimation of data warehouse systems**. In: THE EUROPEAN SOFTWARE MEASUREMENT CONFERENCE – FESMA – DASMA, 2001. Heidelberg, Alemanha .p.12.
- [12] SIMÕES, C. Sistemática de Métricas, qualidade e produtividade.**Developers’ Magazine**, Brasil, 1999. 7p.
- [13] UK. UKSMA Metrics Practices Committee. **MKII Function Point Analysis Counting Practices Manual. Version 1.3.1**. UK, 1998. 100 p.