

Uma Metodologia *Agile* ROLAP para Implantação de Ambientes de Inteligência de Negócios

Elielson B. de Souza¹, André L. Andrade Menolli¹, Ricardo G. Coelho¹

¹Centro de Ciências Tecnológicas – Universidade Estadual do Norte do Paraná (UENP)
Caixa Postal 261 – 86.360-000 – Bandeirantes – PR – Brasil

{elielson,menolli,rgcoelho}@uenp.edu.br

Abstract. *Deploying business intelligence environments allow organizations to make decisions that help their development. These environments are constantly changing due to trends in the market, which makes traditional technologies having difficulties in meeting the needs of business. This way, it was developed a methodology that enables the implementation of such environment in a quickly way, allowing to apply adjustments that comes from the market.*

Resumo. *A implantação de ambientes de inteligência de negócio permitem que as organizações tomem decisões que auxiliem no seu desenvolvimento. Estes ambientes sofrem constantes mudanças devido as evoluções sofridas pelo mercado, o que faz com que tecnologias tradicionais tenham dificuldades em atender as necessidades das empresas. Assim, foi desenvolvida uma metodologia que propicia a implantação de tal ambiente de forma rápida, permitindo que adaptações oriundas do mercado sejam aplicadas.*

1. Introdução

Inteligência de Negócio ou *Business Intelligence* (BI) é o conjunto de tecnologias orientadas a disponibilizar informação e conhecimento para as empresas, permitindo uma melhor visualização do que está ocorrendo, o que contribui para tomadas de decisões mais assertivas [Machado 2010]. As informações disponibilizadas por estas tecnologias são oriundas dos dados que as próprias empresas geram na realização de tarefas corriqueiras, sendo que esses dados são organizados permitindo a análise mais eficiente da informação gerada.

Os principais objetivos do BI é permitir que os dados possam ser acessados de forma interativa, proporcionando assim a sua manipulação e fornecendo aos gerentes e analistas de negócio a capacidade de realizar a análise adequada das informações geradas pela própria organização [Turban et al. 2009]. Dentre as ferramentas e tecnologias utilizadas pelos sistemas de BI destacam-se os *Data Warehouse* (DW) e as ferramentas *On-Line Analytical Processing* (OLAP), que surgiram no começo dos anos 90, como novas ferramentas que foram a base dos sistemas de BI [Shim et al. 2002].

Segundo [Kimball 2004] DW é o processo de transformação dos dados obtidos de sistemas legados e de bancos de dados transacionais que ficam organizados sob um formato compreensível ao usuário, auxiliando na tomada de decisão. Um DW é considerado um armazém de dados que tem como objetivo auxiliar a tomada de decisão dentro de uma empresa. Esse armazém de dados possui algumas características que o diferencia das outras formas de análise dos dados, eles são: orientados por assunto,

variantes no tempo, integrados e não voláteis [Machado 2010].

Os dados que são armazenados no DW seguem a modelagem dimensional, pois ela possibilita a criação de um modelo mais simples sendo entendido facilmente pelos analistas de negócio. O modelo dimensional possui em sua estrutura dois tipos de tabelas, uma representa as dimensões e a outra representa os fatos, os fatos se relacionam com as dimensões e assim permitem obter informações que auxiliam na tomada de decisão.

A concepção de um ambiente de BI ocorre por meio de execução de algumas etapas, mas dependendo da arquitetura utilizada pode-se eliminar algumas etapas que deixam de serem necessárias. Os dados provenientes ao ambiente de BI são obtidos de fontes externas como de bases de dados *On-line Transaction Processing* (OLTP), esses dados são então armazenados em uma nova base de dados, necessitando a execução do processo de *Extract, Transform and Load* (ETL). Segundo [Kimball 2004] um sistema de ETL devidamente projetado extrai os dados dos sistemas de origem, reforça a qualidade dos dados e padrões de consistência, conforma os dados de forma que fontes distintas possam ser usadas juntas e, finalmente, oferece dados em um formato de apresentação pronto para que os desenvolvedores de aplicativos possam construir aplicações e possibilita que os usuários finais tomem decisões.

A execução do processo de ETL manipula os dados de origem armazenando-os no DW, em um formato que permite a análise dos dados pelos analistas de negócio por meio de ferramentas OLAP. OLAP é uma forma de se analisar grandes volumes de dados sobre múltiplas perspectivas, sendo amplamente utilizado em ambientes de BI por possibilitar a análise rápida dos dados gerados pela implantação de tal ambiente. De acordo com [Machado 2010] ferramentas OLAP são aplicações às quais os usuários finais têm acesso para extrair os dados de suas bases e construir os relatórios capazes de responder às suas questões gerenciais.

No entanto, a implantação de um ambiente de BI, por meio da construção de um DW é um processo que demanda muito tempo e recurso para sua elaboração. As empresas desejam obter respostas rápidas aos seus investimentos, por isso, novas metodologias, tais como *Agile BI* foram propostas para tentar facilitar esta implantação. Entretanto, essas metodologias possuem diversas restrições, tais como utilizar ferramentas OLAP específicas. Assim, visando facilitar a implantação do BI, e ao mesmo tempo permitir que ferramentas de BI já existentes sejam utilizadas, é proposta uma nova metodologia, que visa agilizar o processo e reduzir os custos.

O restante deste artigo está organizado como segue: na Seção 2 são abordadas as diferentes arquiteturas disponíveis para a implantação de um ambiente de BI, bem como suas vantagens e desvantagens; na Seção 3 é apresentada a metodologia proposta para implantação de ambientes de BI; na Seção 4 são descritos os resultados obtidos com a nova proposta; na Seção 5 as conclusões do trabalho são apresentadas.

2. Trabalhos Relacionados

É possível projetar um ambiente de BI seguindo diferentes arquiteturas, existem vários fatores que determinam qual a melhor tecnologia a escolher, como o volume de dados, o capital a ser investido, o tempo de implantação, entre outros.

Um DW global é uma arquitetura que visa atender todas as necessidades da empresa, sua capacidade de suporte a tomada de decisão fica disponível para todos os

setores da empresa. Essa arquitetura é composta por *Data Marts* (DMs) que podem ser acessados por toda a empresa, eles podem ficar unidos em uma única instalação física ou serem distribuídos. A arquitetura global está dividida em 5 fases, como mostrado na Figura 1, e apresenta a vantagem de permitir o acesso à visão corporativa dos dados, porém sua construção tem alto custo de implementação e um longo tempo de desenvolvimento [Inmon 1997].

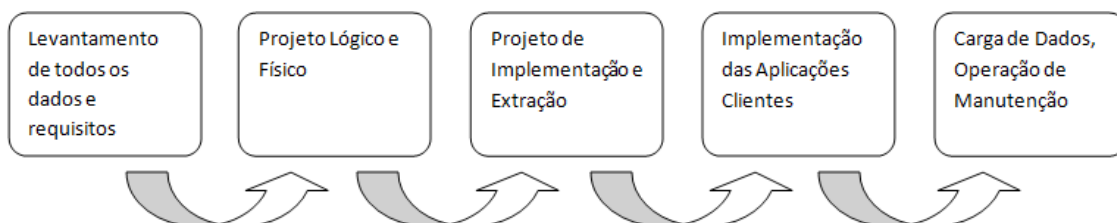


Figura 1: Fases da Arquitetura de Data Warehouse Global

A arquitetura de um DW com *Data Marts* Independentes tem características diversas da arquitetura Global. Esta não possui foco corporativo, pois os DMs são construídos sem apresentarem relação uns com os outros, atendendo a cada setor específico da organização, como ilustrado na Figura 2. Essa arquitetura implica em *Data Marts stand alone* controlados por um grupo específico de usuários atendendo somente às necessidades específicas e departamentais, é a preferida entre os fornecedores de software para consulta de informações por ser isolada [Machado 2010]. Os DMs independentes requerem as mesmas técnicas para implementação de uma arquitetura global, sua implementação é mais rápida, porém não possui muita integração corporativa o que acaba por não permitir uma visão global do negócio [Machado 2010].

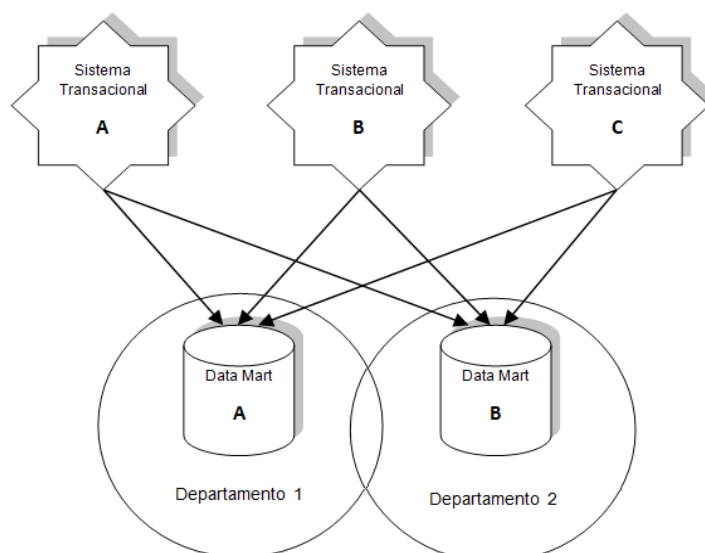


Figura 2: Arquitetura de Data Marts Independentes

Os DMs também podem ser construídos de forma incremental, neste tipo de arquitetura tem-se as características de um DW global, onde sua construção é feita de modo incremental. Segundo [Clemes 2001], nesta abordagem, os requisitos são levantados de forma global, os DMs que serão construídos são identificados, e é definida a maneira como serão integrados. A partir deste momento, cada DM é

implementado completamente até que todos os DM tenham sido implementados, constituindo o DW global da organização. Suas principais vantagens são a apresentação dos resultados de forma rápida, os mecanismos de extração são projetados uma única vez e sua arquitetura permite uma visão global dos dados, porém sua implementação exige a criação de políticas que determinam a sequência de implementação dos DMs e um maior controle no nível de granularidade dos dados [Sell 2001]. A Figura 3 mostra o processo de desenvolvimento dessa abordagem que se constitui de vários ciclos até o completo desenvolvimento do ambiente de BI.

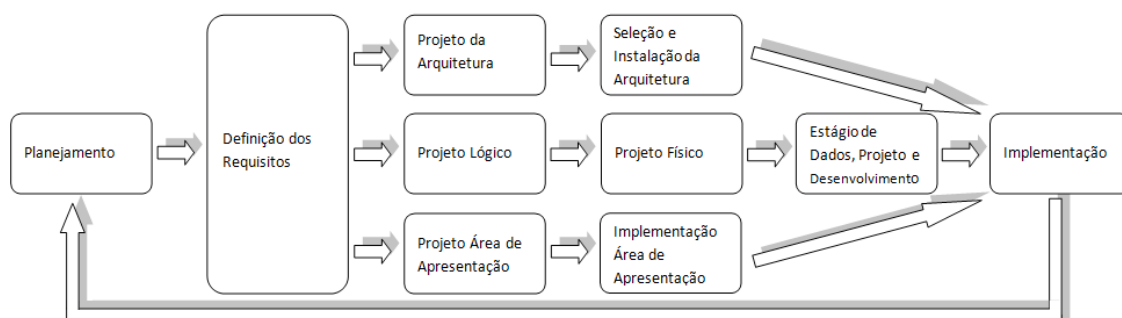


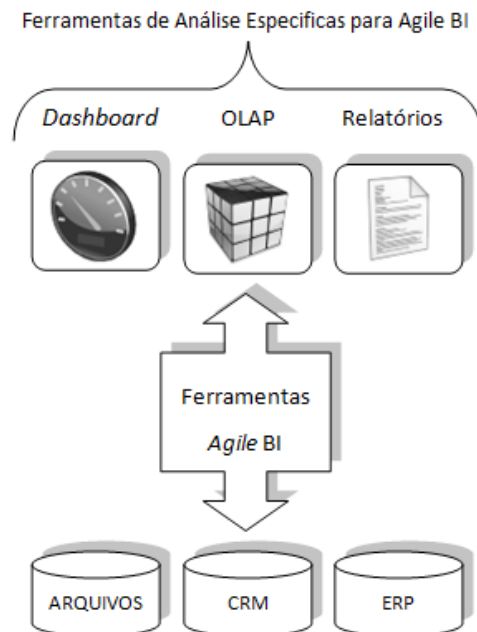
Figura 3: Fases da Arquitetura de Incremental

Um DW construído seguindo a arquitetura de DMs integrados e incrementais permite que seja adicionado um novo DM ao ambiente de BI, ou até mesmo que um DM já existente seja modificado para atender a necessidade da organização. Segundo [Menolli 2006], a principal característica desta arquitetura é a substituição de fontes de vários formatos para uma base de dados padronizada, facilitando a integração de dados. Essa arquitetura está dividida em cinco camadas, desse modo é possível ter independência entre as camadas, ocultando detalhes de implementação entre uma camada e outra. As cinco camadas definidas pela arquitetura são [Menolli 2006]:

- Camada de Fonte de Dados
- Camada de ETL
- Camada de Área de *Staging*
- Camada de *Data Warehouse*
- Camada de Área de Análise

Além das arquiteturas convencionais de desenvolvimento de projetos de BI, recentemente surgiu a arquitetura ágil, que procura oferecer um método mais simples para a elaboração de ambientes de BI. Essa arquitetura ganhou o nome de *Agile BI* e como o próprio nome diz, ela se trata de uma metodologia rápida, de baixo custo e flexível, que visa reduzir o tempo de construção de um ambiente de BI tradicional [Logix, 2012].

As ferramentas *Agile BI* permitem que os dados sejam consultados diretamente das bases transacionais da empresa (Figura 4), eliminando grande parte do processo de implementação efetuado por um ambiente convencional.



**Figura 4: Arquitetura Agile BI
Adaptado de Sandhill (2014)**

Essa abordagem defende, que como principais características o projeto de DW tradicional é lento, caro e inflexível e com a implementação do *Agile BI*, as organizações passam a ser capazes de responder às rápidas mudanças do mercado, obtendo informações em um prazo mais curto [Yellowfin 2010]. No entanto, o grande problema desta metodologia é que as ferramentas de análise de dados precisam ser construídas junto ao projeto, ou seja, não é possível utilizar as ferramentas de análise de dados para ambientes de DW tradicionais, pois o formato físico dos dados nesta arquitetura difere do formato dimensional.

3. Metodologia Proposta

Neste trabalho é proposta uma nova metodologia que visa amenizar alguns dos problemas encontrados na tecnologia *Agile BI*, principalmente no que se refere a utilização de ferramentas OLAP construídas para ambientes de BI tradicionais. Estas ferramentas dispõem de inúmeros recursos que auxiliam na análise dos dados e que poderiam contribuir para um desempenho ainda melhor da tecnologia ágil.

O método denotado de *Agile ROLAP* tem por objetivo permitir a implantação de forma ágil de ambientes de BI que utilizem bancos de dados relacionais, e ao mesmo tempo permitam a utilização de ferramentas OLAP projetadas para ambientes tradicionais por meio de um servidor ROLAP.

A metodologia *Agile ROLAP* visa permitir que empresas de pequeno porte possam utilizar dessa tecnologia para auxiliar na tomada de decisões pertinentes ao seu negócio, essas empresas devem possuir seus dados integrados em um único banco de dados, ou seja, os dados gerados pelos sistemas operacionais da organização devem concentrar-se em um único repositório.

Essa metodologia dispensa o processo de ETL realizado nas metodologias supracitadas, pois não é necessária a construção de uma nova base de dados. Os dados

são consultados diretamente das bases transacionais da organização. Estima-se que mais de 1/3 do custo e do tempo são gastos com o processo de ETL, gerados por problemas na extração, transformação ou na limpeza dos dados e que podem consumir até 80% do tempo gasto no desenvolvimento de um projeto de BI [Menolli 2006]. A possibilidade de se utilizar ferramentas OLAP convencionais permite que a organização tenha a possibilidade de escolher ferramentas que se adequem as suas necessidades, não sendo necessária sua implementação, o que agiliza o processo de implantação da tecnologia.

3.1. Arquitetura

A arquitetura do *Agile ROLAP* procura utilizar em sua implementação os recursos já disponíveis no mercado, procurando reduzir os custos e principalmente o tempo necessário para que seja possível sua utilização pelos analistas de negócio. A Figura 5 ilustra a arquitetura, que possui alguns elementos que são essenciais para o seu funcionamento, como as bases de dados OLTP, o servidor Mondrian e as ferramentas de análise dos dados.

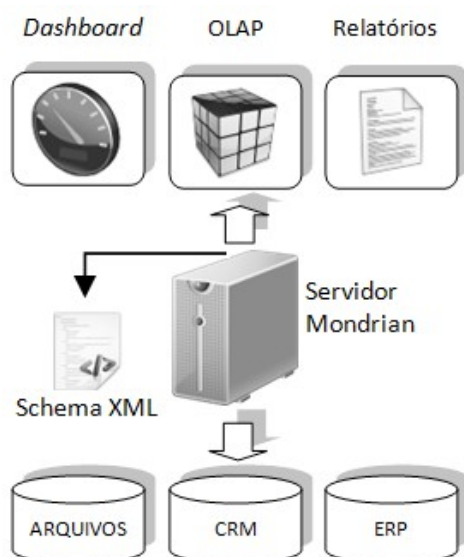


Figura 5: Arquitetura *Agile ROLAP*

Os dados utilizados pelo ambiente de BI são obtidos das fontes de dados da própria empresa, essas fontes permitirão obter os dados necessários para prover aos analistas de negócios as informações que auxiliarão na toma de decisão. Os dados são acessados pelo servidor Mondrian que realizará as consultas necessárias para responder as questões levantadas pelos analistas de negócios, feitas por meio das ferramentas de análise dos dados.

O Mondrian é um servidor OLAP escrito em JAVA que recebe consultas na linguagem *multi-dimensional expressions* (MDX) e converte essas consultas para a linguagem SQL, fazendo assim a obtenção dos dados através de bases de dados que são gerenciadas por SGBDs relacionais, por fim o servidor retorna a consulta solicitada em um formato multidimensional permitindo a análise dos dados sobre vários ângulos [Hyde 2006].

Para que o servidor consiga converter as consultas em MDX para SQL é necessário que o mesmo leia um arquivo conhecido como *schema XML*, esse arquivo

realiza o mapeamento dos dados que estão armazenados na forma relacional para os dados que devem ser mostrados na forma dimensional.

O *schema xml* trabalha com a representação de dois modelos, o lógico que trata da representação multidimensional utilizada pelas consultas em MDX como os cubos, dimensões, hierarquias, níveis e membros e o modelo físico que trata da representação dos dados presentes no banco de dados, realizando assim o relacionamento entre o modelo físico e o lógico [Hyde 2011].

3.2. Ambiente *Agile* ROLAP

O processo de implementação de um ambiente de BI seguindo a metodologia *Agile* ROLAP se diferencia das arquiteturas mencionadas anteriormente, pois neste caso os dados são acessados diretamente da base OLTP da organização por intermédio de um servidor ROLAP.

Os dados que serão utilizados não passam pelo processo de limpeza, eles permanecem da forma com que os sistemas operacionais da organização os gerou, o ambiente de BI somente usufruirá desses dados, da forma com que estão armazenados e organizados, desse modo para que seja possível realizar a análises sob uma perspectiva temporal é necessário que a base de dados mantenha dados históricos gerados pela organização.

O processo de implantação do ambiente *Agile* ROLAP é constituído de 6 fases ilustradas na Figura 6, que permitirão a utilização dos dados da organização por meio das ferramentas de análise dos dados utilizados por ambientes tradicionais.

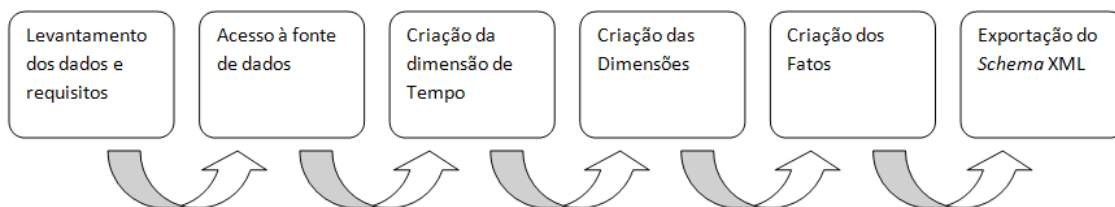


Figura 6: Fases da arquitetura *Agile* ROLAP

O primeiro passo a ser realizado é o levantamento dos dados e dos requisitos, assim como nas outras arquiteturas é necessário conhecer o que a organização pretende com a utilização da tecnologia, quais as áreas que ela deseja estudar, além de outras informações, assim é possível implementar um ambiente que atenda às necessidades da empresa.

Conhecendo as necessidades da empresa é necessário obter o acesso as fontes de dados que serão utilizadas pelo *Agile* ROLAP, isso permitirá uma análise mais detalhada dos dados que serão utilizados, sendo possível identificar quais os requisitos solicitados poderão ser atendidos.

Assim como no processo de DW, é necessário criar uma ou mais dimensões tempo, que conterà todas as informações necessárias para permitir a análise temporal. A dimensão de tempo possui uma importância acentuada em todo o modelo de dados de um ambiente de BI, estando presente em qualquer que seja ele, pois quando se deseja analisar uma fato, não se interessa guardar informações sobre cada transação, mas sobre as transações em um espaço de tempo definido [Machado 2010]. Nessa dimensão é necessário que se faça a criação de uma nova tabela na base OLTP da organização, essa

nova tabela armazenará informações temporais que consigam permitir a análise temporal sob uma perspectiva diária.

O próximo passo desse processo é a criação das outras dimensões que serão utilizadas na análise dos dados, elas fornecerão informações descritivas sobre as atividades da organização. Diferente das dimensões de tempo, essas dimensões são organizadas seguindo a estrutura da base de dados de modo que se consiga obter as informações necessárias a dimensão.

Nesta fase é necessário definir qual será a tabela principal que a dimensão utilizará e quais informações ela dispõe. Caso não seja possível extrair os dados necessários de uma única tabela, são realizadas junções (*joins*) entre a tabela principal e outras tabelas que relacionam com esta. Este relacionamento é feito de forma lógica por meio do mondrian *schema*, mantendo a estrutura física original, mas sendo visualizada na ferramenta de BI de forma desnormalizada.

Os fatos, assim como a dimensão de tempo, precisam de uma estrutura própria para que armazenem as atividades da organização. Porém os fatos como são originados das tabelas das bases transacionais não necessitam a criação de uma nova tabela, neste caso se utiliza a criação de *views*. Cada *view* criada corresponde a um fato, lá são armazenados os relacionamentos com as dimensões e as medidas referentes ao fato, sua criação ocorre por meio da execução de uma consulta constituída de *joins* que acessam as tabelas da base OLTP e recuperam as informações necessárias que permitirão o acesso as dimensões e para a criação das medidas.

A criação dos fatos por meio de *views* permitem que o tempo de execução de uma consulta seja reduzido, pois os fatos são originados de diversas tabelas o que acaba exigindo um trabalho extremo do SGBD, pois já são realizados *joins* na constituição das dimensões.

O último passo desse processo é a criação do *schema* XML que será utilizado pelo servidor Mondrian para fazer o mapeamento do modelo lógico para o físico, são lá que a estrutura das dimensões e fatos ficam definidas bastando apenas serem interpretadas pelo servidor.

Finalizado esse processo, o ambiente de BI já pode ser utilizado pelos analistas de negócio por meio de ferramentas de análise dos dados disponíveis no mercado que a organização identificou como sendo a melhor para atender suas necessidades.

4. Resultados

A utilização da metodologia *Agile* ROLAP proporciona os benefícios já consagrados pelas tecnologias ágeis, como o *Agile* BI, mas além disso proporcionam a possibilidade da utilização de ferramentas tradicionais, o que aumenta consideravelmente a variedade de ferramentas disponíveis no mercado.

Um estudo realizado em uma base de dados relativamente simples mostrou que a implantação de um ambiente utilizando a metodologia *Agile* ROLAP é atrativa para empresas de pequeno porte, que não possuem uma base de dados volumosa e que deseja utilizar dessa tecnologia para auxiliar no desenvolvimento da organização.

A base de dados utilizada, chamada de Pagila, retrata os dados provenientes de um sistema OLTP de locação de filmes, ela está disponível sob a licença *Berkely Software Distribution* (BSD). O estudo de caso desenvolvido procurou analisar as

locações de filmes registradas na base, procurando responder a questões como o número total de locações nos fins de semana, os gêneros dos filmes preferidos pelos clientes, entre outros.

Essa base foi desenvolvida por Mike Hillyer membro da equipe de desenvolvimento do MySQL com o objetivo de fornecer um esquema padrão para ser usados como exemplo em livros, tutoriais, artigos, entre outros, sua base contém mais de 45.000 registros e é composta por 21 tabelas que procuram armazenar dados como os autores dos filmes, as locações realizadas, os pagamentos, etc [PgFoundry 2008].

Foram criadas seis dimensões convencionais, duas dimensões de tempo e um fato como na Figura 7, o tempo de estudo da base não foi levado em consideração, pois é o mesmo para qualquer arquitetura utilizada.

Enquanto que na criação do ambiente de BI tradicional provida de um DW foi utilizado a ferramenta kettle, sendo possível realizar o processo de ETL construindo-se uma nova base de dados, no ambiente *Agile* ROLAP a definição das dimensões foi feita direto no *schema* XML e a criação do fato no próprio SGBD.

Neste estudo pode-se utilizar a mesma ferramenta de análise em ambos os ambientes, cumprindo o objetivo da tecnologia. Além disso, a metodologia permite que por exemplo, seja possível utilizar conceitos como *role play dimensions* que é a capacidade de se utilizar uma dimensão várias vezes em um fato, de modo a permitir que por exemplo sejam necessário a existência de somente uma dimensão de tempo em um fato que necessite representar os dados sob duas perspectivas temporais.

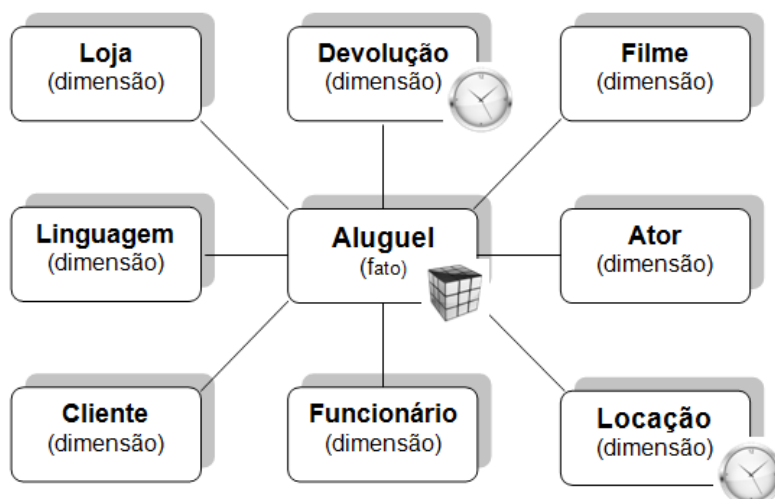


Figura 7: Dimensões e fato criados no estudo realizado

5. Conclusões

As constantes mudanças que ocorrem no mercado fazem com que ambientes de BI tenham que se adaptar para atender as necessidades da organização. Os ambientes tradicionais possuem dificuldade em se adaptarem, pois são complexos, o que não acontece com a metodologia *Agile* ROLAP, que permite que o processo de implantação de tal ambiente seja mais simples, rápido e adaptável.

Uma nova alternativa para implantação de ambientes de BI dentro das organizações permite que um maior número de empresas possam utilizar desta tecnologia a fim de auxiliar na tomada de decisões, além disso pretendesse construir um componente para a ferramenta Kettle com o intuito de automatizar o processo do *Agile* ROLAP na implantação de ambientes de BI.

Referências

- Clemes, M. (2001) “Data warehouse como suporte ao sistema de informações gerenciais em uma instituição de ensino superior: estudo de caso na UFSC.”, Dissertação de Mestrado – Universidade Federal de Santa Catarina, Programa de Pós-Graduação em Engenharia de Produção, Florianópolis.
- Hyde, J. (2006) “Mondrian Documentation: Arquiteture”, Pentaho, Disponível em: <<http://mondrian.pentaho.com/documentation/architecture.php>>.
- Hyde, J. (2011) “Mondrian Documentation: MDX Specification”, Pentaho, Disponível em: <<http://mondrian.pentaho.com/documentation/mdx.php>>.
- Inmon, W. H. (1997) “Como Construir o Data Warehouse”, Rio de Janeiro.
- Kimball, R., Caseta, J. (2004) “The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data”, John Wiley & Sons.
- Logix. (2012) “What is Agile BI”, Disponível em <<http://www.nathean.com/index.php/what-is-agile-bi/>>.
- Machado, F. N. R. (2010) “Tecnologia e Projeto de Data Warehouse: uma visão multidimensional”, 5 ed., Érica, São Paulo.
- Menolli, A. L. A. (2006) “A Data Warehouse Architecture in Layers for Science and Technology”, Proceedings of the Eighteenth International Conference on Software Engineering Knowledge Engineering (SEKE'2006), San Francisco, CA, USA.
- PgFoundry. (2008) “A collection of sample databases for PostgreSQL”, Disponível em: <<http://pgfoundry.org/projects/dbsamples/>>
- Sandhill. (2014) “BI Ready Data Warehouse Automation”, Disponível em: <<http://www.sandhillconsultants.com/BIReadyProduct.asp>>
- Sell, D. (2001) “Uma arquitetura para distribuição de componentes tecnológicos de sistemas de informações baseado em data warehouse”, Dissertação de Mestrado – Universidade Federal de Santa Catarina, Programa de Pós-Graduação em Engenharia de Produção, Florianópolis.
- Shim, J. P., Warkentin, M., Courtney, J., Power, D. J., Sharda, R. e Carlsson, C. (2002) “Past, Present, And Future Of Decision Support Technology”, Decision Support System, V. 33, N. 2, P. 111-126.
- Turban, E., Sharda, R., Aronson, J. E. and King, D. (2009) “Business Intelligence: um enfoque gerencial para a inteligência do negócio”, Grupo A.
- Yellowfin. (2010) “Making Business Intelligence Easy: Write Paper Agile Business Intelligence”, Disponível em: <<http://www.yellowfinbi.com/Document.i4?DocumentId=97458>>.