



UNIVERSIDADE FEDERAL DO PAMPA
CIÊNCIA DA COMPUTAÇÃO

MAICON CONCATTO ISOTON

**USO DE INTEGRAÇÃO DE DADOS PARA UMA APLICAÇÃO DE
WEBCONFERÊNCIA**

Trabalho de Conclusão de Curso

Alegrete
2012

MAICON CONCATTO ISOTON

**USO DE INTEGRAÇÃO DE DADOS PARA UMA APLICAÇÃO DE
WEBCONFERÊNCIA**

Trabalho de Conclusão de Curso apresentado
como parte das atividades para obtenção do
título de bacharel em Ciência da Computação
na Universidade Federal do Pampa.

Orientador: Prof. Dr. Sérgio Mergen

Co-orientador: Me. Patric da Silva Ribeiro

Alegrete

2012

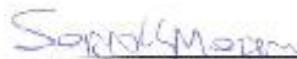
MAICON CONCATTO ISOTON

USO DE INTEGRAÇÃO DE DADOS PARA UMA APLICAÇÃO DE
WEBCONFERÊNCIA

Trabalho de Conclusão de Curso apresentado
como parte das atividades para obtenção do
título de bacharel em Ciência da Computação
na Universidade Federal do Pampa.

Trabalho apresentado e aprovado em: 12 de Julho de 2012.

Banca Examinadora:



Prof. Sergio Mergen
Orientador

Ciência da Computação - UNIPAMPA



Patrícia da Silva Ribeiro
Coorientador

NTIC - UNIPAMPA



Prof. Fábio Natanael Kepler
Ciência da Computação - UNIPAMPA



Prof. Aline Vieira de Mello
Ciência da Computação - UNIPAMPA

RESUMO

Sistemas de Webconferência permitem realizar reuniões virtuais com pessoas situadas em diversos locais, reduzindo consideravelmente custos com viagens e o tempo de deslocamento. A UNIPAMPA é uma universidade composta por múltiplos campos e, portanto, tem interesse especial no uso de sistemas de Webconferência. A UNIPAMPA utiliza o sistema de Webconferência chamado Abe Connect. Os dados estatísticos gerados pelo Adobe Connect ainda não satisfazem as demandas da UNIPAMPA. Este sistema provê dados estatísticos que não respondem a questões como: Qual a economia da UNIPAMPA na realização de uma reunião virtual? Esta economia é maior do que o custo do Adobe Connect? Os recursos do Adobe Connect estão sendo utilizados adequadamente? Estas questões podem ser respondidas através do acesso a duas bases de dados distintas: a base do Adobe Connect e a base do Sistema de Abertura de Chamados da UNIPAMPA, onde são agendadas as reuniões virtuais. Diferentes abordagens de integração de dados podem ser utilizadas para (i) combinar os dados destas duas bases distintas e (ii) realizar consultas sob estes dados combinados. Desse modo, este trabalho apresenta as diferentes abordagens de integração de dados e implementa a mais adequada ao contexto da UNIPAMPA.

Palavras-chave: Integração de Dados, Banco de Dados, *Wrappers*.

ABSTRACT

Conferencing systems allow virtual meetings with individuals located in various locations, significantly reducing travel costs and travel time. The UNIPAMPA is composed of a university multiple fields and, therefore, has particular interest in the use Conferencing system. The UNIPAMPA use Web Conferencing system called Adobe Connect. Statistical data generated by Adobe Connect does not satisfy the demands of UNIPAMPA. This system provides statistical data that do not respond to questions like: What's economy UNIPAMPA in conducting a virtual meeting? This saving is greater than the cost of Adobe Connect? The feature of Adobe Connect is being used properly? These questions can be answered by accessing two separate databases: the base and the base of the Adobe Connect System Opening Recalled UNIPAMPA where virtual meetings are scheduled. Different approaches to data integration can be used to (i) combining the data from these two different bases and (ii) consultations under these combined data. Thus, this paper presents different approaches to data integration and implements the most appropriate to the context of UNIPAMPA.

Keywords: Data Integration, Database, Wrappers.

LISTA DE ILUSTRAÇÕES

- Figura 1. Arquitetura clássica de integração de dados.
- Figura 2. Arquitetura de integração baseada em uma abordagem virtual.
- Figura 3. Arquitetura de integração baseada em uma abordagem materializada.
- Figura 4. Exemplo demonstrando o casamento entre dois esquemas.
- Figura 5. Arquitetura de um data warehouse.
- Figura 6. Arquitetura básica usada para ambientes de replicação de dados.
- Figura 7. Snapshot da Ferramenta Adobe Connect.
- Figura 8. Snapshot da parte do Sistema de Abertura de Chamados referente ao agendamento de Webconferências.
- Figura 9. Tabelas relevantes extraídas do modelo de dados do Sistema Adobe Connect.
- Figura 10. Esquema das tabelas “ocorrencias” e “usuarios”, usadas pelo Sistema de Abertura de Chamados.
- Figura 11. Arquitetura escolhida para o desenvolvimento do trabalho proposto.
- Figura 12. Esquema da base de dados materializada do trabalho.
- Figura 13. Ordem de execução dos extratores.
- Figura 14. Mapeamento para tabela global “usuario”.
- Figura 15. Consulta SQL para extração dos usuários.
- Figura 16. Mapeamento para a tabela global "sala".
- Figura 17. Consulta SQL para extração das salas.
- Figura 18. Mapeamento para a tabela global "reserva".
- Figura 19. Consulta SQL para extração das reservas.
- Figura 20. Expressão regular para localizar o solicitante da descrição.
- Figura 21. Algumas solicitações que o extrator conseguiu identificar.
- Figura 22. Ilustração das tabelas de exemplo.
- Figura 23. Consulta SQL para extração dos participantes.
- Figura 24. Mapeamentos para a carga total da “reserva”.
- Figura 25. Visão da tela do administrador.
- Figura 26. Resultados para a primeira consulta.
- Figura 27. Resultados para a segunda consulta.
- Figura 28. Resultados para a terceira consulta.
- Figura 29. Resultados para a quarta consulta.
- Figura 30. Resultados para primeira consulta, utilizando a segunda base de testes.
- Figura 31. Resultados para segunda consulta, utilizando a segunda base de testes.
- Figura 32. Resultados para terceira consulta, utilizando a segunda base de testes.
- Figura 33. Resultados para quarta consulta, utilizando a segunda base de testes.

LISTA DE TABELAS

Tabela 1. Resultado de uma consulta utilizando a abordagem virtual ilustrada na Figura 2.

Tabela 2. Resultados gerados na tabela global “reserva”.

LISTA DE ABREVIATURAS E SIGLAS.

UNIPAMPA: Universidade Federal do Pampa.

NTIC: Núcleo de Tecnologia da Informação e da Comunicação.

STIC: Setores de Tecnologia da Informação da Comunicação dos Campi.

XML: *Extensible Markup Language*.

GAV: *Global-as-View*.

LAV: *Local-as-View*.

ETL: *Extract, Transform and Load*.

OLAP: *On-line Analytical Processing*.

GPL: *General Public License*.

SGBD: Sistema de Gerenciamento de Banco de Dados.

LDAP: *Lightweight Directory Access Protocol*.

URL: *Uniform Resource Locator*.

SUMÁRIO

Resumo	4
ABSTRACT	5
Lista de tabelas	7
Lista de abreviaturas e siglas	8
Sumário.....	9
Introdução.....	10
Trabalhos Relacionados.....	12
Aplicação de Integração de Dados	25
Resultados.....	45
Conclusão	52
Referências	53
Apêndice.....	55

1. INTRODUÇÃO

Sistemas de Webconferência permitem realizar reuniões virtuais com diversas pessoas de diversos lugares. Sistemas desse tipo são extremamente úteis quando os participantes estão distantes, pois economizam recursos gastos com transporte e hospedagem, que seriam necessários em caso de reuniões presenciais. Como a UNIPAMPA possui uma estrutura descentralizada, composta por diversos campi espalhados em áreas geograficamente distantes, as Webconferências podem diminuir consideravelmente os gastos nesse tipo de atividade. O custo para adquirir este software de Webconferência é inferior aos gastos da instituição no sentido de locomoção, diárias e manutenção da frota ao realizar encontros reais ao invés de encontros virtuais.

Já existe um software que está sendo usado na UNIPAMPA para esse fim. Trata-se do Adobe Connect (CARVALHO, 2007), um produto proprietário que realiza Webconferências de forma simples, exigindo apenas que o usuário tenha um computador com acesso a Internet e o plugin do Flash instalado.

Através da ferramenta Adobe Connect é possível obter dados estatísticos referentes às Webconferências realizadas. No entanto, os dados fornecidos são limitados, e não satisfazem completamente as demandas da UNIPAMPA. Por exemplo, a ferramenta fornece informações como o número de usuários diários por sala. Por outro lado, não é possível descobrir o número de usuários presentes em uma reunião específica. Se forem marcadas 10 reuniões em um dia, não será possível saber quantos participantes estavam em cada reunião, e isso impede que se possa determinar com precisão quantos recursos foram economizados com uma reunião por Webconferência.

Limitações como essa ocorrem porque os dados estão divididos em duas bases de dados distintas, uma no próprio Adobe Connect e a outra no Sistema de Abertura de Chamados da UNIPAMPA. O Sistema de Abertura de Chamados foi desenvolvido para aperfeiçoar o processo de suporte técnico na instituição, através de um programa informatizado de abertura, acompanhamento e gerenciamento de chamados de suporte ao NTIC (Núcleo de Tecnologia da Informação e da Comunicação) e STIC (Setor de Tecnologia da Informação da Comunicação dos Campi), e seu uso pode ser feito por qualquer pessoa que tenha vínculo com a UNIPAMPA. Esse sistema armazena as informações como número de participantes, data e hora de todas as reuniões agendadas para Webconferência. Já o Adobe Connect armazena informações complementares, como nome da sala, link da sala, uma descrição e o tipo de acesso que é aplicado à sala.

Como os dados de interesse estão distribuídos em duas bases distintas, pretende-se combinar os dados dessas duas fontes e aplicar consultas sob os dados combinados para resgatar informações mais completas. Tais consultas permitirão descobrir informações úteis, como por

exemplo, o número de pessoas que participaram de uma reunião específica. Em última análise, as consultas sobre os dados combinados permitirão verificar o real ganho obtido pelo uso do Adobe Connect na UNIPAMPA.

A combinação de dados mencionada no parágrafo anterior costuma ser realizada através de técnicas de integração de dados. O problema de integração de dados pode ser estudado por diversos ângulos. Por exemplo, uma divisão básica considera a abordagem como virtual ou materializada (SALGADO, 2010). As abordagens virtual e materializada apresentam vantagens e desvantagens que podem variar de acordo com o tipo de problema que se busca resolver. É importante destacar que dentro da grande área de integração diversos outros problemas precisam ser resolvidos, como a extração de dados das fontes (LAENDER, 2002), limpeza de dados para eliminar inconsistências (OLIVEIRA, 2004) e a criação de mapeamentos de forma automática, utilizando técnicas de casamento de esquemas (MERGEN, 2007).

Desse modo, o corrente trabalho visa escolher um método de integração, estudando as soluções possíveis e averiguando qual delas é mais vantajosa para o problema de integração proposto. O método escolhido será utilizado para implementar o sistema, buscando permitir que consultas, que possuem grande valor para a instituição, possam ser respondidas.

Este trabalho está organizado da seguinte forma: no capítulo 2 são apresentados trabalhos relacionados, visando demonstrar a arquitetura geral de sistemas de integração de dados e seus principais componentes, como os mediadores, *wrappers* e extratores. Além disso, outros conceitos relacionados ao trato de dados são descritos, como sincronização, replicação, conversão e mineração. No capítulo 3 são apresentados detalhes referentes à proposta de trabalho, como os modelos de dados necessários, as consultas que se respondeu e a arquitetura de integração escolhida. Adicionalmente é relatado o processo que resultou no conjunto de extratores capazes de resgatar todas as informações das fontes. O capítulo 4 será apresentado os resultados obtidos da integração aplicada neste trabalho. Por fim as conclusões obtidas são apresentadas no capítulo 5.

2. TRABALHOS RELACIONADOS

Este capítulo aborda um dos assuntos principais deste trabalho de conclusão de curso, a integração de dados. Além disso, também são apresentados outros temas relacionados, como sincronização, *data warehouse*, replicação, tipos de integração, arquitetura da integração, mapeamentos e conversão de dados. Os tópicos abordados são bastante relevantes para aplicações que objetivam trabalhar com cenários como o proposto neste trabalho, e conhecer seus conceitos e funcionamento serve de embasamento para as escolhas arquiteturais feitas.

2.1 Integrações de Dados

O objetivo da integração de dados é conseguir gerar uma interface na qual o usuário consiga as respostas às suas consultas sem precisar interagir diretamente com as bases de dados.

A Figura 1 traz a arquitetura clássica de um sistema de integração de dados (CAVALCANTE, 2005). Dentre os componentes principais da arquitetura destacam-se o mediador, onde se encontra um esquema de dados chamado de esquema global e os *wrappers* que encapsulam o acesso às fontes de dados. O esquema global é associado com as fontes de dados através de mapeamentos. O esquema global é o ponto de partida por onde o usuário realiza pesquisas. Uma das vantagens de sistemas desse tipo é a transparência, ou seja, o usuário é capaz de acessar dados das fontes sem se preocupar com detalhes particulares a essas fontes, como sua localização e seu esquema. Outra vantagem diz respeito à autonomia. Em alguns sistemas de integração, as fontes integradas são independentes em relação às demais fontes, e tem completa liberdade para definir como a fonte deve ser mantida.

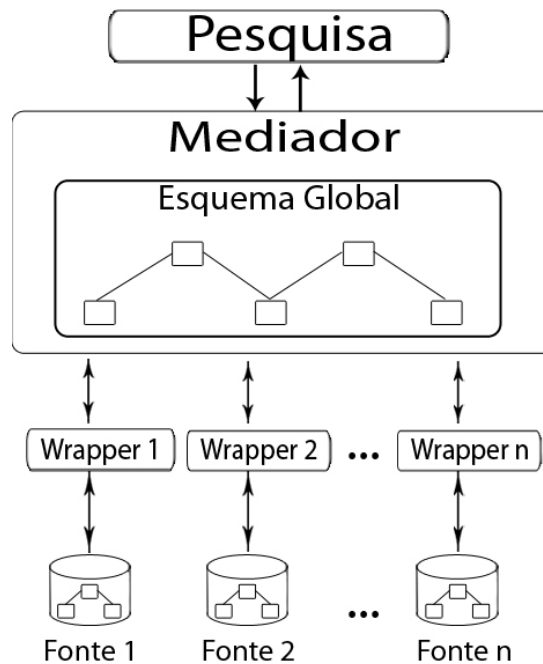


Figura 1. Arquitetura clássica de integração de dados.

As seções seguintes trazem mais detalhes referentes aos componentes básicos de uma arquitetura de integração de dados.

2.1.1 Mediadores

Os mediadores (BARBOSA, 2001) tem uma visão de todo o mapeamento das fontes, sabendo o que é e onde fica cada fonte dos dados. A camada de mediação é responsável pelo transporte das informações das fontes até o esquema global. A sua principal função é gerar uma informação integrada com base nas informações obtidas das fontes.

Ao receber uma consulta feita sobre o esquema global, o mediador analisa sua estrutura e realiza uma divisão em sub-consultas, as quais são enviadas para as fontes de dados. Essa divisão é necessária porque normalmente o esquema global difere dos esquemas das fontes. Assim, a consulta original precisa ser adaptada para que seja compatível com o esquema das fontes. A geração das sub-consultas é realizada com base nas informações de mapeamento que associam cada um dos elementos do esquema global com cada um dos elementos dos esquemas das fontes.

Após processadas as sub-consultas, o mediador analisa cada uma das respostas obtidas e forma uma resposta única que é devolvida para o usuário. Além do processo de subdivisão

da consulta global, a unificação das respostas é considerada um dos maiores desafios dentro da área de integração de dados.

Para que um mediador seja criado, é importante ter em mente qual será sua finalidade, pois a sua implementação pode sofrer variações de acordo com o tipo de aplicação a que ele se destina. Por exemplo, sistemas heterogêneos, que permitem acesso a diversos tipos de fontes de dados, são mais difíceis de construir do que sistemas homogêneos, que limitam o acesso a um tipo particular de fonte de dados. Sistemas abertos, que permitem que novas fontes de dados sejam adicionadas a qualquer momento, também são naturalmente mais difíceis de construir do que sistemas fechados, que são destinados a um conjunto pré-definido e constante de fontes de dados.

2.1.2 *Wrappers*/Extratores

Os *wrappers* (BIANCARDI, 2005) são componentes fundamentais que auxiliam na integração de dados. Eles são mecanismos de tradução que encapsulam as fontes de dados e realizam a intermediação das comunicações entre o mediador e a fonte encapsulada. Estes mecanismos são responsáveis por receber as solicitações que os mediadores realizam e converter estas solicitações em consultas na linguagem que a fonte de dados compreenda. Após realizar a consulta, o *wrapper* devolve os resultados no formato de dados esperado pelo mediador.

Como exemplo de uso de *wrappers* considere a existência de duas fontes de dados, uma construída no modelo XML e outra com Modelo Relacional. Considere também que o esquema global utiliza o modelo relacional. A fonte de dados relacional possui um relacionamento mais natural com o esquema global, pois ambas estão descritas no mesmo formato. No entanto, a fonte de dados XML deverá ter seus dados convertidos para o modelo relacional para que o mediador consiga acessar os dados. Essa comunicação entre o mediador e fontes de dados de modelos heterogêneos pode ser alcançada através do uso de um *wrapper*.

Por sua vez, os extratores são componentes que realizam a extração das informações que estão contidas nas fontes. Geralmente são usados para a conversão/tradução de dados, onde os dados são primeiramente extraídos de uma fonte de origem e depois convertidos para o formato de uma fonte de destino (LEE, 2005). O problema de conversão/tradução de dados normalmente ocorre em situações onde entidades distintas necessitam trocar informações entre si, como aplicações de *Business-to-Government* (B2G), onde empresas precisam enviar dados para órgãos ligados ao governo (MONTARDO, 2006).

Especificamente na área de integração de dados, os extratores podem ser usados para extrair informações para uma base de conhecimento, que é posteriormente acessada para a realização de consultas. Por exemplo, a extração de dados é utilizada para obter informações sobre áreas de interesse dos currículos que estão armazenados na plataforma LATTES, e que são recuperados em formato XML (RIBEIRO, 2005).

Considerando a área de integração, é importante não confundir *wrappers* com extratores. Na verdade, pode-se considerar que um extrator seja parte de um *wrapper*, pois além da extração, um *wrapper* é responsável por toda a intermediação da comunicação entre uma fonte de dados e o sistema que deseja comunicar-se com ela.

2.1.3 Abordagens de Integração de Dados

Na integração de dados, existem dois tipos de abordagens clássicas: materializada e virtual (SALGADO, 2010). A diferença fundamental é que na abordagem materializada os dados das fontes são carregados para o esquema global, enquanto na abordagem virtual os dados permanecem nas fontes. Considerando a primeira alternativa, quando uma consulta é disparada apenas o esquema global precisa ser acessado, visto que ele já possui todas as informações necessárias. Já na segunda alternativa, se torna necessário acessar as fontes de dados para descobrir as informações necessárias para resolver a consulta.

A abordagem virtual pode ser comparada com as visões de tabelas relacionais (ELMASRI, 2011), que trabalham da mesma forma. Uma visão é uma tabela que se forma a partir de uma consulta que acessa outras tabelas. Uma visão é virtual quando seus dados não residem fisicamente dentro dela, mas sim obtidos a partir de outras tabelas no momento que a visão for consultada. Essa característica das visões virtuais limita as ações que se pode realizar sobre elas. Por exemplo, algumas visões não permitem operações de atualizações, pois o banco de dados é incapaz de determinar como as tabelas por trás da visão devem ser atualizadas. Visões são úteis para dados que são requisitados com bastante frequência, pois elas simplificam o acesso a esses dados além de conseguir que esse acesso seja otimizado.

Já a abordagem materializada pode ser comparada com visões materializadas dos bancos de dados relacionais (SILBERSCHATZ, 2006). Esse tipo de visão possui dados redundantes, pois os dados das tabelas originais são também armazenados dentro da visão. Essa redundância torna o acesso aos dados mais eficiente. Contudo, os dados residentes na visão podem estar desatualizados em relação às tabelas originais.

A Figura 2 mostra um exemplo aplicado em que a abordagem virtual é utilizada em uma arquitetura de integração de dados. Os dados existentes nas fontes são dados

complementares, onde uma fonte tem informações sobre a data de nascimento do cliente e a outra sobre o endereço deste cliente. Como se pode ver, o esquema global não possui nenhum registro. Quando o usuário faz uma pesquisa, o mediador i) recebe essa consulta, ii) verifica quais serão as fontes de dados que serão necessárias para responder esta consulta, e iii) gera uma ou mais sub-consultas, para cada fonte de dados distribuída. Por último, o mediador processa as respostas das sub-consultas e repassa para o usuário o que ele tinha pesquisado.

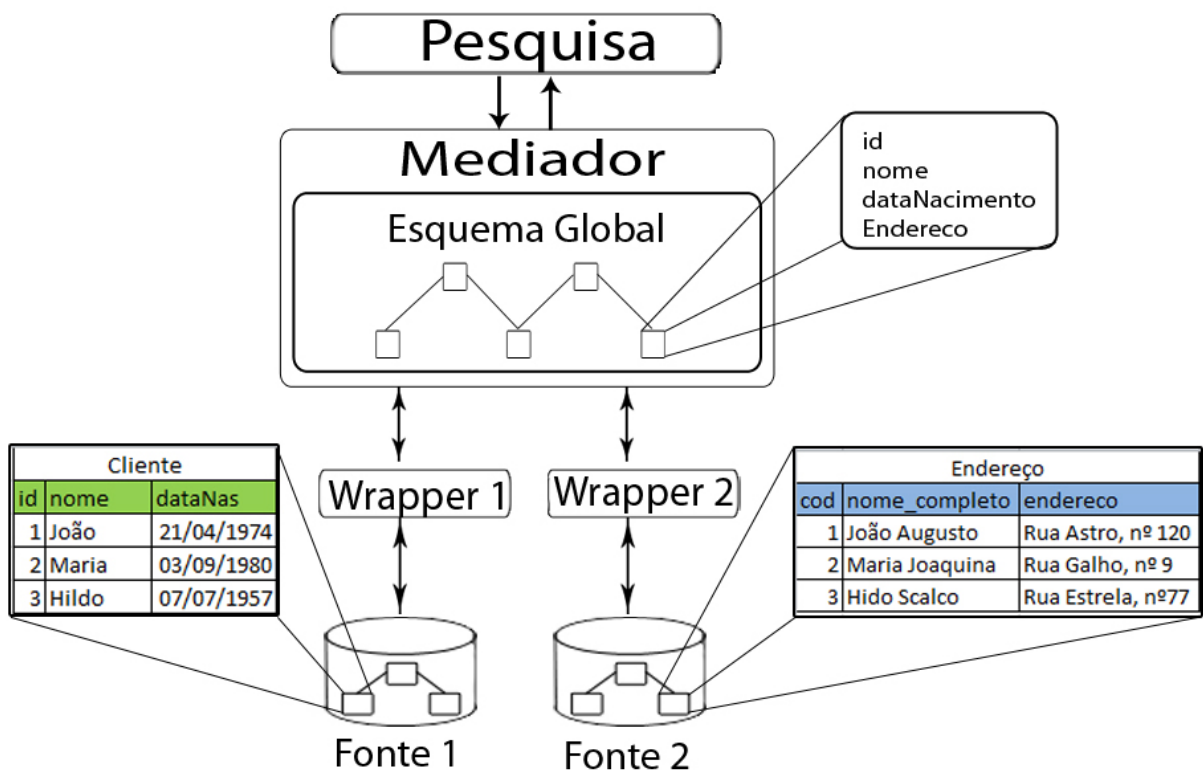


Figura 2. Arquitetura de integração baseada em uma abordagem virtual.

Ainda de acordo com o exemplo, o usuário poderia pesquisar por dados de clientes nascidos em 1974, e obter a resposta conforme demonstrado na Tabela 1. Observe que o mediador foi capaz de reconhecer registros de fontes distintas que equivalem à mesma informação, mesmo estando essa informação armazenada de forma diferente. No caso, os dados de “João Augusto” aparecem na Fonte 1 sob o atributo “nome” e escrito como “João”, e na Fonte 2 sob o atributo “nome_completo” e escrito como “Joao Augusto”. Isso demonstra uma das habilidades desejáveis em sistemas de integração, e um dos principais desafios deste tipo de sistema. Para resolver problemas como esse, é comum o uso de técnicas de similaridade, capazes de reconhecer equivalência de textos escritos de forma diferente (LIMA,

2011), além de técnicas de casamento de registros, capazes de reconhecer a equivalência de estruturas de dados distintas (NUNES, 2009).

id	nome	endereço	dataNasciemento
1	João Augusto	Rua Astro, nº120	21/04/1974

Tabela 1. Resultado de uma consulta utilizando a abordagem virtual ilustrada na Figura 2.

A Figura 3 demonstra um exemplo baseado em arquiteturas materializadas. Observe que, ao contrário do esquema anterior, os dados foram previamente carregados no esquema global. Quando o usuário realizar pesquisas, deixa de ser necessário o acesso às fontes de dados, uma vez que os dados já estão à disposição no repositório central.

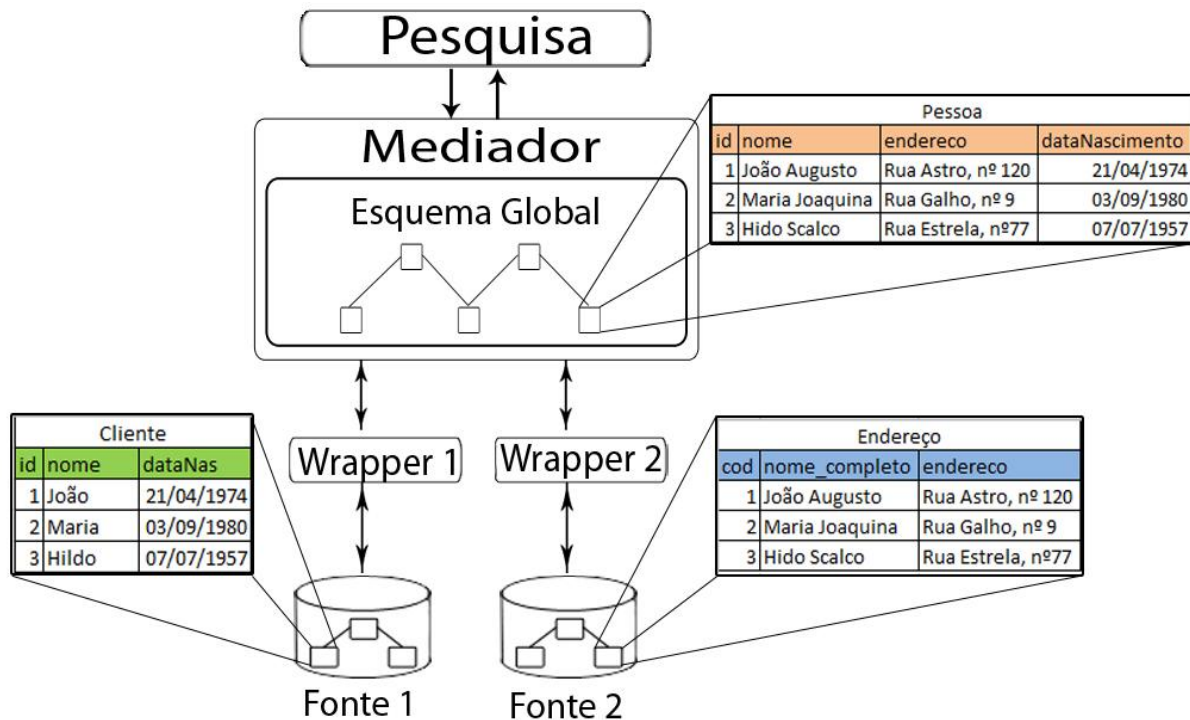


Figura 3. Arquitetura de integração baseada em uma abordagem materializada.

Nessa abordagem, o desafio é realizar a carga desses dados. Para que isso seja feito, antes da pesquisa ser realizada é necessário que a integração carregue os dados das diversas fontes para o esquema global. Para realizar esse processo de carregamento, é necessário que em um determinado momento o sistema execute a integração, onde o mediador solicita as fontes que lhe enviem todas as informações contidas nelas. Essas informações são extraídas e convertidas pelos *wrappers* para que os mediadores possam compreender o que foi extraído. O mediador se encarrega de filtrar e organizar as informações obtidas e colocá-las no seu devido lugar dentro de uma nova tabela no esquema global.

É importante observar que os problemas de registros similares também ocorrem nesta arquitetura. No entanto, os registros similares devem ser mesclados para que possam ser carregados no esquema global, diferentemente da virtual que retorna os dados diretamente para o usuário.

As duas abordagens mostradas anteriormente apresentam vantagens e desvantagens. Por um lado, a abordagem materializada é mais rápida, uma vez que a consulta é respondida sem ter que acessar a fontes de dados, utilizando somente o esquema global. Por outro lado, os dados materializados podem estar desatualizados, dependendo de quando foi realizada a atualização. Já a abordagem virtual é mais lenta, pois as fontes precisam ser acessadas para a recuperação dos dados. Além disso, esta abordagem está sujeita a falhas de comunicação, caso as conexões de redes com as fontes de dados apresentem alguma espécie de problema. No entanto, é garantido que os dados obtidos representam o estado atual da informação.

2.1.4 Mapeamentos

Os mapeamentos indicam como as informações de um esquema de dados podem ser armazenadas em outro tipo de esquema. Esse conceito é semelhante ao conceito de casamento de esquemas. A diferença é que o casamento indica somente quais elementos de um esquema correspondem a elementos do outro esquema. Já os mapeamentos são mais completos e descrevem como os dados podem ser traduzidos de um esquema para o outro.

A Figura 4 mostra como foi feito o casamento das informações do exemplo citado nas abordagens virtual e materializada. As ligações mostram onde foram feitas as junções de valores para formar a tabela resultado (Pessoa).

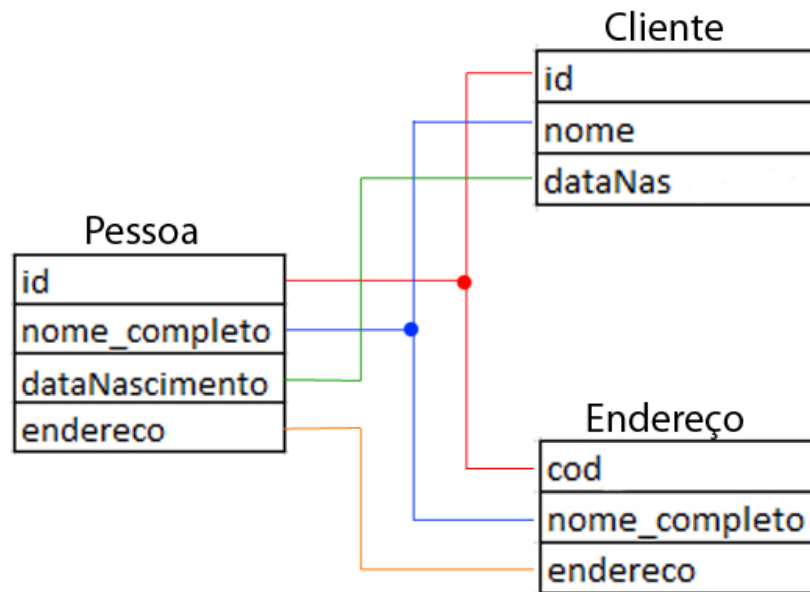


Figura 4. Exemplo demonstrando o casamento entre dois esquemas.

Tanto a criação de casamentos quanto de mapeamentos pode ser automatizada através do uso de técnicas especializadas (ORTIZ, 2002). A propósito, a etapa de criação de casamentos pode servir de entrada para o processo de criação de mapeamentos, que irá complementar as informações de casamento com informações que indiquem como os dados representados em um esquema são convertidos em outro esquema.

Basicamente, os mapeamentos podem ser de dois tipos: *global as view* (GAV) e *local as view* (LAV). Na abordagem GAV, para cada estrutura existente no esquema global são definidas as formas de mapeá-la com as fontes. Em outras palavras, as estruturas do esquema global são visões sobre as fontes. Para exemplificar, considere a Figura 3. Um possível mapeamento GAV seria criar uma consulta SQL sobre a tabela “Cliente” e indicar que o resultado dessa consulta corresponde aos atributos da tabela global “Pessoa”.

Na abordagem LAV, para cada estrutura existente nas fontes são definidas as formas de mapeá-la com o esquema global. Em outras palavras, as estruturas das fontes são visões sobre o esquema global. Um possível mapeamento LAV seria criar uma consulta SQL sobre a tabela “Pessoa” e indicar que o resultado dessa consulta corresponde aos atributos da tabela local “Cliente”.

2.2 *Data Warehouse*

O *data warehouse* (MACHADO, 2004) é um repositório de dados que armazena informações enriquecidas obtidas das fontes de dados. Essas informações são usadas basicamente para obter conhecimento implícito, o que costuma ser usado na tomadas de decisões gerenciais.

Data warehouse é utilizado, por exemplo, em sites de comércio eletrônico, onde é necessário saber o comportamento dos clientes de modo a tratá-lo de uma forma mais personalizada. Identificar esse comportamento com base nos bancos que armazenam as transações comerciais pode ser difícil, dado o volume de informações disponíveis e a quantidade de relacionamentos entre essas informações. Para essa situação, um *data warehouse* poderia sintetizar as informações das transações comerciais e o histórico de navegação, de modo a manter dados estatísticos preservados. Acessando o *data warehouse*, seria mais fácil descobrir o padrão de comportamento e os interesses dos clientes, e utilizar os dados obtidos para, por exemplo, exibir produtos compatíveis com o perfil do cliente logo que ele entra na página do site de vendas.

Existe uma grande diferença entre integração de dados e de *data warehouse*. A integração de dados serve para criar uma interface transparente entre os usuários e as fontes, mesmo estas fontes sendo locais ou distribuídas em vários lugares. Já o *data warehouse* é utilizado em um determinado sistema para obter informações importantes e privilegiadas do comportamento dos usuários, permitindo o melhor atendimento deste usuário.

A Figura 5 mostra um esquema clássico de *data warehouse*. A parte à esquerda da Figura 5 exibe as fontes de dados, de onde serão extraídas as informações. O ETL (*Extract, Transform and Load*) (CIELO, 2002) é utilizado para popular às tabelas do *data warehouse*. O processo ETL inicia-se pela extração dos dados de suas fontes, podendo ser de diversos formatos (*Extract*). Somente são extraídos os dados mais significativos que podem ser utilizados para obter informações chaves. Feito isso, é necessário transformar esses dados, filtrando e adaptando essas informações (*Transform*). Após isso os dados são gravados no repositório (*Load*). A carga é feita em tabelas fato e tabelas dimensão, que representam a estrutura típica dos dados de um *data warehouse* (OLIVEIRA, 2002). Uma tabela fato é aquela que armazena informações referentes a valores detalhados de medidas. Já a tabela dimensão é aquela que armazena informações referentes ao nome específico da dimensão.

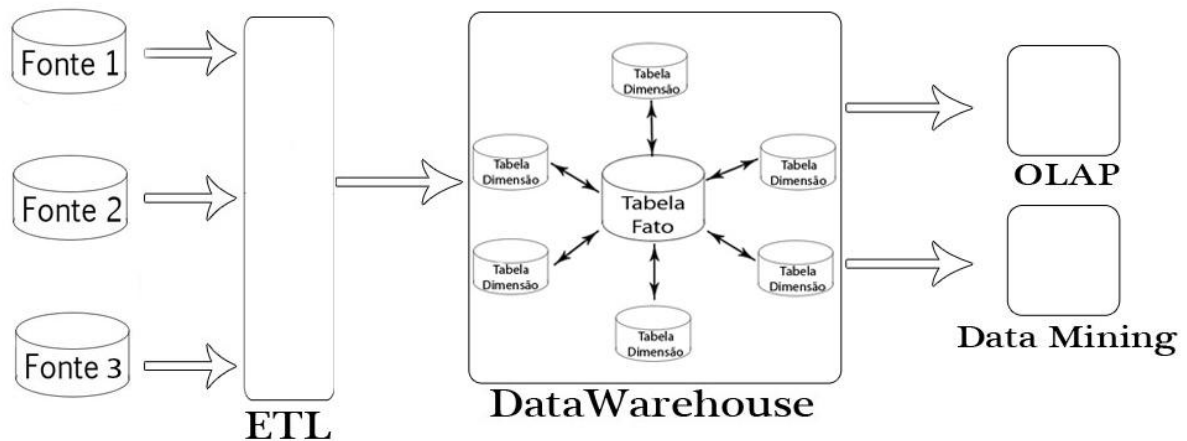


Figura 5. Arquitetura de um data warehouse.

Com os dados armazenados no *data warehouse*, pode-se utilizar ferramentas que consomem esses dados para produzir informações gerenciais. Por exemplo, ferramentas OLAP (*On-line Analytical Processing*) (ANZANELLO, 2007) possibilitam o processamento analítico online de dados. Com elas é possível visualizar as informações de várias perspectivas diferentes, deixando sua estrutura mais eficiente. Já as ferramentas de *Data Mining* (NAVEGA, 2002) buscam identificar nas informações armazenadas dados que contenham algum padrão válido, novo ou valioso.

Um dos grandes problemas do uso do *data warehouse* é manter sempre atualizado o repositório, garantindo que somente as informações mais importantes estarão contidas nele, e preservando sua integridade e consistência. Existem dois métodos para realizar este procedimento: por visão e manutenção incremental.

O método de manutenção incremental (ANGER, 2000) realiza uma análise de todo esse repositório e faz uma comparação de integridade entre os dados deste repositório e os dados das fontes. Somente os dados que forem mais recentes na fonte serão atualizados no repositório. Isso pode ser feito através de técnicas inteligentes, que utilizam algoritmos complexos de verificação.

O método por visão (SALGADO, 2001) visa substituir todo o repositório, removendo o conteúdo antigo e inserindo as informações mais atualizadas das fontes de dados. Este método é mais custoso e pode ser inviável quando o volume de dados a ser carregado for muito grande.

As diferentes técnicas usadas em *data warehouse* para atualização do repositório também podem ser usadas em sistemas de integração de dados, especialmente quando os dados são materializados no esquema global. Isso demonstra que sistemas criados para

propósitos distintos podem compartilhar técnicas em comum. É importante reconhecer quando essas técnicas podem ser usadas como forma de aprimorar o sistema que está sendo desenvolvido.

2.3 Replicações de Dados

A replicação de dados (LORÊDO, 2004) é a cópia de dados de uma fonte de dados principal para uma ou mais fontes auxiliares. A replicação possui dois principais objetivos: garantir a disponibilidade dos dados e aumentar o desempenho do sistema. Por exemplo, quando a fonte de dados principal para de funcionar, ou o limite de usuários é extrapolado, a base auxiliar pode ser acessada.

Hoje em dia existem diversas aplicações que utilizam replicação de dados, algumas famosas como os serviços Gmail e o Docs da Google. A Computação em Nuvens (*Cloud Computing*) (RUSCHEL, 2010) utiliza constantemente a replicação para poder garantir a disponibilidade de serviços na Internet. Por exemplo, quando o cliente acessa um serviço e o servidor que o provê falha, esta falha não é percebida pelo cliente graças à replicação.

A replicação de dados tem um funcionamento diferente da integração de dados. A replicação é feita para garantir que os dados ou serviços sejam mantidos disponíveis. Por outro lado, na integração de dados tem como papel principal o acesso aos dados de forma transparente, sem o usuário se preocupar quanto à localização das informações.

A Figura 6 mostra como funciona uma arquitetura simples de replicação de dados. Considere uma aplicação qualquer que possua um banco de dados, chamado de base Primária. Esta base primária é então replicada, o, que implica na realização de uma cópia da base primária para as cópias de backup. Essa cópia é realizada periodicamente para manter as cópias mais atualizadas possíveis.

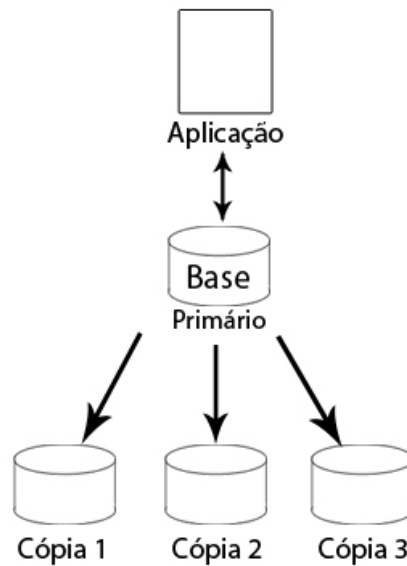


Figura 6. Arquitetura básica usada para ambientes de replicação de dados.

Em caso de falhas da base primária, deve ser feito um redirecionamento automático para uma das réplicas dos dados. Um esquema de replicação bem implementado deve ter no mínimo duas réplicas, aumentando a disponibilidade caso a falha atinja uma das réplicas. A maior dificuldade na replicação de dados é conseguir manter as réplicas confiáveis, garantindo a integridade e a consistência das informações contidas nelas.

A replicação de dados pode ser dividida em réplica primária e réplica ativa. Na réplica primária, apenas uma réplica recebe, processa e responde as requisições feitas. As outras réplicas são passivas, realizando apenas o backup, e têm seus estados atualizados periodicamente a partir da réplica primária. A vantagem desse tipo de replicação é o baixo consumo de processamento. Em caso de falhas na réplica primária, uma réplica passiva se tornará a primária, mas pode ocorrer perda de dados quando a réplica passiva não está atualizada em relação à réplica primária.

Na réplica ativa, todas as réplicas receberão as requisições, realizarão o processamento de forma paralela e produzirão as mesmas saídas. Esta técnica de replicação também pode ser chamada de Máquina de Estados (SCHNEIDER, 1990). Quando for feita uma atualização, ela vai ser distribuída por todas as réplicas. Caso alguma réplica falhe, é possível obter o resultado de outra réplica instantaneamente, já que todas possuem a mesma resposta. Além disso, este tipo de réplica é capaz de tolerar a maioria dos tipos de falhas. O seu uso é indicado para aplicações que necessitam que seus serviços sejam ininterruptos.

Os sistemas de replicação também podem ser classificados como síncronos ou assíncronos. Em uma replicação de dados síncrona (SIPRIANO, 2009), todas as bases de dados vão permanecer com as mesmas informações sincronizadas e consistentes. Quando uma

base é atualizada, essas atualizações são imediatamente propagadas para todas as suas réplicas. Este tipo de replicação somente é utilizado quando se necessita que todas as réplicas sejam alteradas imediatamente ao mesmo tempo. Já em uma replicação de dados assíncrona (LORÊDO, 2004), as réplicas ficam momentaneamente desatualizadas. Quando uma base é atualizada, esta nova informação vai ser distribuída para as réplicas em um segundo plano, o que pode demorar alguns segundos ou até mesmo horas, dependendo da forma de como a arquitetura foi configurada. Este método é bastante utilizado quando a conexão com o banco de dados não é permanente.

3. APLICAÇÃO DE INTEGRAÇÃO DE DADOS

Este trabalho de conclusão de curso tem como objetivo o uso da integração de dados para conseguir obter informações combinadas a partir de bases de dados distintas. Neste capítulo serão abordadas as questões relacionadas à forma como a integração será realizada. Primeiramente serão apresentados detalhes a respeito dos sistemas de onde os dados serão integrados. Em seguida, serão apresentadas as consultas que se pretende responder ao término da integração. Para concluir o capítulo, será descrita a arquitetura de integração que foi selecionada para resolver o problema, juntamente com as justificativas para o seu uso. Outro ponto a salientar é que o usuário final deste trabalho será um usuário administrador capaz de tomar decisões adequadas sobre o resultado de cada uma das consultas propostas.

3.1 Sistema Adobe Connect

O Sistema do Adobe Connect começou a ser desenvolvido pela Macromedia com o nome Breeze. A empresa Adobe comprou este sistema e transformou-o no novo e melhorado Adobe Connect, incrementando algumas funcionalidades que o deixaram com uma aparência mais elegante, garantindo assim uma experiência profissional de comunicação via Webconferência. Trata-se de um sistema seguro e flexível de comunicação via Internet, que permite gerar soluções de comunicação inteligentes entre empresas de grande porte e suas filiais, ou marcar reuniões com várias pessoas que se encontram em pontos geograficamente distantes.

Atualmente, o contrato de licença da UNIPAMPA com a Adobe permite que o sistema de Webconferência suporte cerca de 1000 pessoas online ao mesmo tempo, que podem ser divididas em até 10 salas de 100 pessoas cada. Esta divisão de salas pode ser variada, o importante é que não ultrapasse o número total de pessoas simultaneamente conectadas (1000). Este limite de pessoas foi estipulado contratualmente com a Adobe, devido, em parte, a limitações de *hardware* onde o software está instalado. Caso seja necessário aumentar o limite de número de pessoas, seria necessário também trocar tanto o contrato de licença quanto o *hardware* implantado. Como este software é pago, isso acarretaria em custos para a UNIPAMPA.

3.1.1 Funcionamento

O Sistema de Webconferência do Adobe Connect é de fácil utilização, possibilitando que qualquer pessoa com pouco conhecimento em informática consiga usá-lo para se reunir com outras pessoas. Para poder usufruir de uma sala, é necessário configurar um *host*, ou seja, uma pessoa que será o administrador da sala e terá a sala a sua disposição na data e período agendados para uma reunião. Essa configuração é realizada a partir de dados fornecidos através do Sistema de Abertura de Chamados, que é descrito na Seção 3.2.

No dia e na hora marcados para reunião, o *host* entra na sala reservada e faz o login com nome de usuário e senha e espera pela entrada dos participantes. Se os participantes possuem nome de usuário e senha, eles podem se conectar diretamente na sala. Caso contrário, podem entrar como convidado, preenchendo o campo “nome” e clicando em “entrar”. Em ambos os casos, é necessário esperar que o *host* permita a entrada.

A Figura 7 apresenta a tela principal do sistema. A parte central é reservada para a visualização do material apresentado. No painel à direita é possível ver todos os participantes da sala, e no menu superior são disponibilizadas as opções de uso do sistema. Todos que entram na sala são postos como convidados, que tem como funções liberadas apenas a opção de levantar a mão, para chamar a atenção do *host*, e escrever no chat. Para se tornar apresentador da sala, é preciso levantar a mão e pedir ao *host* permissão para falar. Se a permissão for concedida, o convidado passará a ser um apresentador e poderá falar no microfone e compartilhar a webcam e arquivos do seu computador. Quando o *host* sair da sala - o que a princípio só deverá ocorrer ao término da reunião - os convidados serão redirecionados para a página inicial da sala de Webconferência.

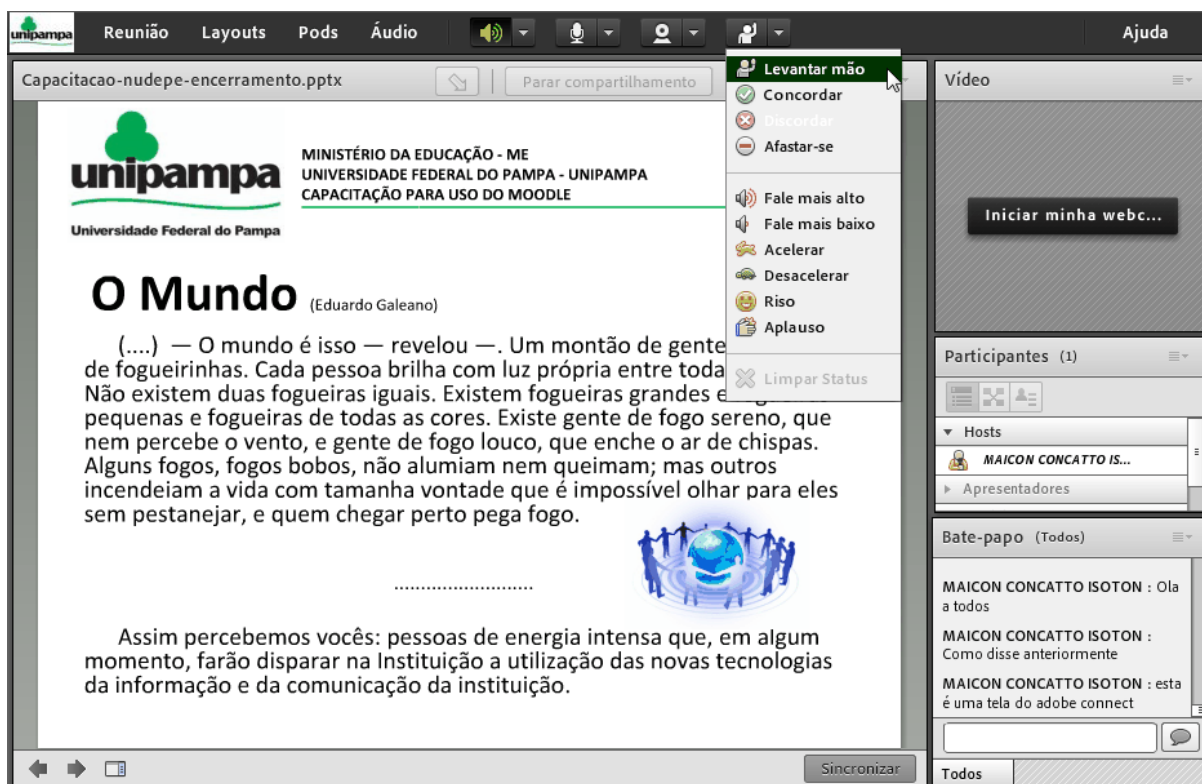


Figura 7. Snapshot da Ferramenta Adobe Connect.

3.2 Sistema de Abertura de Chamados da UNIPAMPA

O Sistema de Abertura de Chamados é um sistema que auxilia nas atividades da UNIPAMPA, visando atender alguns chamados de manutenção e reparos, assim como pedido de empréstimos de materiais e agendamento de Vídeo e Webconferências. Vídeo Conferências são realizadas utilizando equipamentos próprios para este fim enquanto Webconferências são realizadas utilizando um computador com acesso a Internet.

O Sistema de Abertura de Chamados original se chamava Ocomon. Um sistema criado em março de 2002 como projeto pessoal de Franque Custodio (FERNANDES, 2010), tendo como principais atividades o cadastro, acompanhamento, controle e consulta de atividades ou ocorrências de suporte técnico. Em 2003 surgiu uma versão mais consistente com o nome de Invmon. Em 2004 foi adicionado mais alguns módulos ao sistema, cujo nome foi trocado em definitivo para OcoMom, que é o software que vem sendo utilizado pela UNIPAMPA. Este software tem a licença GPL (*General Public License*), o que o torna um software livre para ser modificado.

3.2.1 Funcionamento

Como dito anteriormente, o Sistema de Abertura de Chamados possui diversos recursos. Este trabalho concentra-se na parte responsável pelos agendamentos de Webconferência, visto que esse é o objeto de estudo do trabalho.

Para agendar uma sala, é necessário que a pessoa acesse o Sistema de Abertura de Chamados, selecione a opção respectiva a agendamento de Webconferência, preencha os campos requisitados e aguarde pela resposta do sistema.

A Figura 8 apresenta a parte do sistema onde pode ser realizada uma reserva de sala. A fim de facilitar o agendamento da sala, é de extrema importância que o solicitante preencha o campo “Descrição” com o maior número de informações possíveis, como data/hora, nome da sala, tempo de duração e nome do *host*. É importante destacar que esses dados são úteis no contexto da integração que se pretende realizar.

Abertura de Ocorrências:

Área Principal*: Conferência On-Line

Área Responsável*: Conferência Web

Problema*: Agendamento

PROBLEMA	SLA	Categoria 1	Categoria 2	Categoria 3
<input checked="" type="radio"/> Agendamento	2 dias			

Este espaço é destinado para solicitação de Conferência Web para eventos que envolvam todas as unidades em conjunto. Para aulas, reuniões e projetos de pesquisa, extensão e ensino devem solicitar o agendamento conferência web por meio do suporte da própria unidade (Ex: Área Principal (Suporte) -> Área Responsável (Unidade) -> Problema (Conferência Web)). O agendamento deve ser solicitado com antecedência de 48hs. Informe o nome e e-mail do solicitante, o horário de início, a duração prevista e o número de participantes; descreva a finalidade da conferência.

Descrição do problema e local*:

Nº do Patrimônio do equipamento:

Número de Série:

Data de abertura: 05/01/2012 15:30:52

Status: Aguardando atendimento

Anexar arquivo: Escolher arquivo Nenhum arquivo selecionado Anexar outro

Anexar arquivo: Escolher arquivo Nenhum arquivo selecionado Anexar outro

Anexar arquivo: Escolher arquivo Nenhum arquivo selecionado Anexar outro

Anexar arquivo: Escolher arquivo Nenhum arquivo selecionado Anexar outro

Anexar arquivo: Escolher arquivo Nenhum arquivo selecionado Anexar outro

Figura 8. Snapshot da parte do Sistema de Abertura de Chamados referente ao agendamento de Webconferências.

Geralmente a resposta do Sistema de Abertura de Chamados para esse tipo de solicitação ocorre em torno de 48 horas após o pedido. A confirmação da sala é recebida através de um e-mail, sendo que o corpo deste e-mail possui os dados de confirmação da reserva que foi efetuada com sucesso. De posse disso, o usuário que fez a requisição se torna *host*. O *host* pode enviar para as outras pessoas que irão participar o link de acesso à sala e também a data/hora de início da reunião.

Hoje este processo de agendamento de salas ainda é manual. No momento que é confirmada a reserva, o analista que atendeu ao pedido vincula o papel de *host* ao solicitante. Para um futuro próximo, existe um projeto de automatização deste serviço, de modo que o solicitante se torne *host* em um prazo mais curto.

O processo de desligamento de *host* também é manual. Todo dia os analistas examinam as Webconferências agendadas para o dia anterior e removem o *host* respectivo. Este procedimento pode causar alguns inconvenientes, como o de pessoas ficarem além do tempo estipulado em suas salas. Existem planos de correção e automatização desse processo, o que irá diminuir o tempo de permanência. Esses dois problemas visam ser eliminados após o projeto de automatização ser colocado em prática, o que irá facilitar o controle de salas e *hosts*.

3.3 Consultas Elaboradas

As consultas são a parte chave deste trabalho. Ao determinar as consultas que deverão ser atendidas, será possível conceber um modelo de integração adequado, que seja apto a responder essas consultas de forma satisfatória. Cabe ressaltar que o objetivo não é construir um modelo que fique restrito a alguns casos isolados, e sim um modelo flexível que possa ser estendido para atender a novos tipos de requisições, quando elas surgirem. Ou seja, as consultas propostas nesse trabalho tem o papel de ajudar na concepção de um modelo de integração inicial, que pode ser evoluído ao longo do tempo.

A lista inicial de consultas foi elaborada com base em entrevistas realizadas com especialista tanto no uso do Adobe Connect quanto no Sistema de Abertura de Chamados. As consultas definidas procuram combater algumas deficiências que foram identificadas durante as entrevistas. O objetivo de cada consulta está descrito abaixo:

1. Verificar o número de participantes de cada reunião solicitada e realizada pelo Adobe Connect: Hoje é possível ter um número de participantes diários de todas as salas,

sem conseguir dividi-las por salas e reuniões. Por exemplo, pode-se descobrir que 1000 pessoas utilizaram o sistema de Webconferências em um dia específico, mas não se consegue descobrir quantas utilizaram para a reunião “A” ou mesmo quantos participantes integraram uma aula introdutória ao Moodle.

Para realizar esta consulta é necessário acessar as duas fontes de dados (Sistema do Adobe Connect e Sistema de Abertura de Chamados), e obter informações das reuniões que foram reservadas e realizadas. Esta consulta é interessante para a UNIPAMPA porque permite monitorar todas as reuniões feitas e verificar como estão sendo distribuídos os participantes entre elas. Com esta consulta, o analista também pode detectar o horário onde existem mais participantes online ao mesmo tempo, identificando assim se há necessidade de limitar o número de reuniões nos horários de maior fluxo, garantindo assim maior qualidade do serviço.

2. Verificar se a reserva de salas pelo *host* está sendo realizada de maneira adequada: Utilizando somente o Sistema de Abertura de Chamados é possível saber quantas reuniões foram solicitadas por certo *host*, mas não as que foram realmente realizadas. Ou seja, não existe um controle efetivo sobre as desistências de agendamentos. Quando um usuário desiste da sala, os analistas retiram desse usuário as permissões de *host*. Caso a sala não seja permanente, ou seja, caso ela tenha sido criada para uso exclusivo do solicitante, ela é removida. O problema ocorre quando o usuário não avisa a respeito da desistência. Nesse caso, as salas excedentes e permissões de *host* só são removidas no dia posterior, deixando recursos alocados desnecessariamente.

Hoje em dia ainda não se tem uma política própria para o caso de desistências. Mas com a implantação do projeto de agendamento automático podem-se criar políticas para o melhor uso do sistema evitando assim inconvenientes como o descrito acima. A consulta concebida neste tópico pode ajudar a identificar os casos problemáticos, e com isso elaborar políticas de controle mais eficazes.

Para responder esta consulta e saber quais das reuniões que foram solicitadas realmente aconteceram, vai ser necessário cruzar informações com as bases dos dois sistemas. Sendo assim, serão extraídas do Sistema de Abertura de Chamados as informações iniciais como nome do solicitante, data e hora em que foram marcadas as reuniões, e do sistema do Adobe Connect serão extraídas informações que possam ser comparadas. Tendo posse desses dados, é possível realizar o cruzamento e obter a certificação que naquele horário determinado o *host* esteve online e verificar também se o período em que ocorreu a reunião está correto.

3. Verificar se o tempo estipulado pelo *host* está sendo utilizado da maneira adequada: Em alguns casos, o *host* usa a sala por um período de tempo superior ao que foi realmente solicitado. Isso pode gerar conflitos, caso outro *host* esteja esperando para usar a sala. O uso de salas exclusivas por usuário elimina esse problema. Contudo, algumas salas de uso comunitário costumam ser utilizadas com frequência, como as salas dos campi. Até o momento, nenhum problema de conflito foi identificado, mas isso não se deve a obediência aos horários solicitados, e sim ao intervalo de tempo existente entre as reuniões. Caso as reuniões passem a ter intervalos mais curtos, é possível que esse problema passe a ocorrer. Com a consulta aqui proposta, os *hosts* que não respeitarem o limite de tempo estipulado podem ser identificados e alertados quanto a esse problema.

4. Verificar se o número de participantes solicitado pelo *host* está sendo utilizado da maneira adequada: O sistema do Webconferências possui um limite de conexões simultâneas. Se este limite for atingido, o sistema pode ser prejudicado por efeitos colaterais como latência alta nas Webconferências, o que gera falhas de conexão durante as reuniões. Para evitar problemas desse gênero, a memória necessária pelo Adobe Connect é gerenciada com base nas informações de demanda de salas e número de participantes. Se essas informações estiverem equivocadas, recursos insuficientes podem ser alocados. Ou seja, se mais pessoas do que o previsto acessarem o sistema, isso pode gerar um gargalo que provoque falhas de conexão. Sabendo quais *hosts* não estão respeitando a configuração requisitada, podem-se tomar as medidas necessárias para que eventos como esse não tornem a acontecer, como por exemplo, alertando-os a respeito desse problema.

3.4 Modelo de Dados do Adobe Connect

O Sistema de Webconferências (Adobe Connect) utiliza um SGBD relacional baseando em SQL Server. Seu banco de dados é bastante complexo e extenso, e possui informações espalhadas em 116 tabelas. O fato de as informações estarem divididas, aliado a falta de documentação, dificultou a descoberta das tabelas mais significativas para o trabalho proposto. Além disso, existem inúmeras tabelas e atributos de tabela com nomes pouco intuitivos, o que tornou o processo de reconhecimento dos dados relevantes ainda mais trabalhoso.

Com base nas consultas demonstradas na seção anterior, foi possível identificar quatro tabelas que contém atributos relevantes como nome, data de início e data de fim das reuniões,

nomes de salas, entre outras informações. A Figura 9 mostra essas tabelas e seus respectivos campos. Os campos em preto são os mais promissores para a realização deste trabalho.

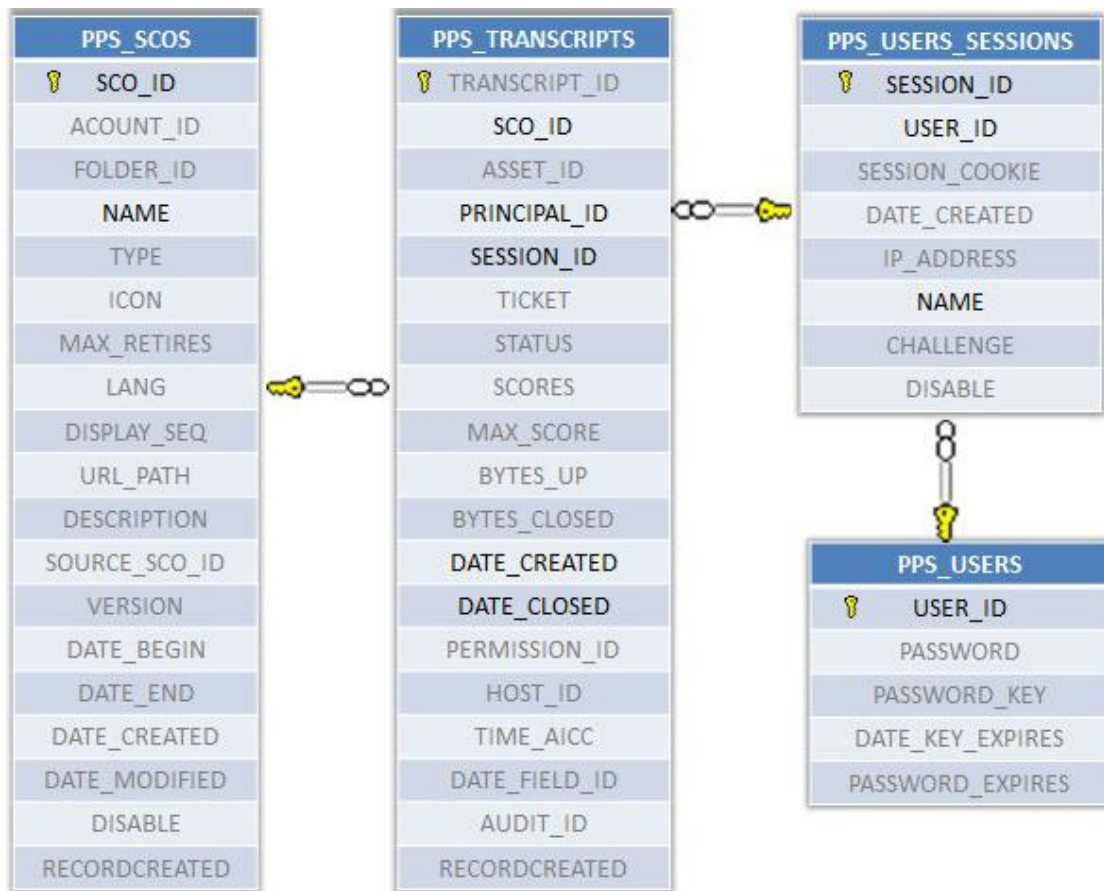


Figura 9. Tabelas relevantes extraídas do modelo de dados do Sistema Adobe Connect.

A identificação dos usuários (“USER_ID”) é obtida através da tabela “PPS_USERS”. Curiosamente, o nome do usuário fica armazenado em outra tabela, conforme será mostrado mais adiante. Não foi possível determinar o motivo dessa distribuição. Esta tabela contém um total de 21432 registros.

A tabela “PPS_USER_SESSIONS” contém informações como o nome do usuário (atributo “NAME”) e os códigos de identificação do usuário e da sessão (atributos “USER_ID” e “SESSION_ID”). Isso significa que ficam armazenados todos os logs de Webconferências que os usuários participaram. Esta tabela contém um total de 23005 registros.

A tabela “PPS_TRANSCRIPTS” possui as informações de entrada e saída de usuários na sala (atributos “DATE_CREATED” e “DATE_CLOSED”). Esta tabela também armazena informações como o código de identificação da sala em que o usuário acessou (atributo “SCO_ID”), e os códigos de sessão e usuário (atributos “PRINCIPAL_ID” e “SESSION_ID”), que servem para realizar o cruzamento das informações. Esta tabela contém um total de 36883 registros.

A tabela “PPS_SCOS” possui informações como o código de identificação de uma sala (atributo “SCO_ID”) e o nome desta sala (atributo “NAME”). Também pode ser observado que esta tabela contém a URL relativa da sala. Para obter a URL completa, é necessário adicionar a URL relativa no final do link institucional do sistema de Webconferência. Esta tabela contém um total de 1744 registros.


Relacionando as informações dessas quatro tabelas, é possível descobrir quais *hosts* estão cadastrados, que sessões eles iniciaram, quando eles iniciaram e finalizaram essas sessões. Outra informação que se pode extrair são os participantes de cada sessão, conforme será visto mais adiante.

3.5 Modelo de Dados do Sistema de Abertura de Chamados

O Sistema de Abertura de Chamados possui um banco de dados relacional implementado em um SGBD MySQL, contendo 82 tabelas. Como esse sistema foi originado do OcoMom, algumas tabelas desse banco não são utilizadas pelo Sistema de Abertura de Chamados na UNIPAMPA.

Inicialmente foram identificadas duas tabelas com dados relevantes ao trabalho: “Ocorrências” e “Assentamentos”. A primeira traz informações sobre as solicitações feitas para a equipe de suporte. Para o caso das reservas de sala para Webconferência, a tabela traz atributos como nome do host, número de participantes e data da reunião. Já a segunda tabela traz informações de resposta dos atendentes para essas solicitações. Após analisar os dados contidos nas tabelas, optou-se por ignorar os dados da tabela “Assentamento” visto que as respostas dos atendentes são muito vagas e pouco úteis para as consultas que se deseja responder. Como exemplo, existem casos em que a resposta para um pedido de sala foi “Sua sala está reservada!”.

A Figura 10 mostra o esquema das tabelas “ocorrencias” e “usuarios”. A primeira tabela possui 21 atributos, contudo, apenas dois são necessários para este trabalho: “descricao” e “contato”. Esta tabela contém um total de 7516 registros. A segunda tabela possui 11 atributos, mas apenas dois são necessários para este trabalho: “user_id” e “nome”. Esta tabela contém um total de 1085 registros. Os mesmo atributos necessários estão destacados em preto.

OCORRENCIAS	
	NUMERO
	PROBLEMA
	DESCRIÇÃO
	EQUIPAMENTO
	SISTEMA
	CONCATO
	TELEFONE
	LOCAL
	OPERADOR
	DATA_ABERTURA
	DATA_FECHAMENTO
	STATUS
	DATA_ATENDIMENTO
	INSTITUICAO
	ABERTO_POR
	OCO_SCHEDULED
	OCO_REAL_OPEN_DATE
	OCO_SCRIPT_SOL
	DATE_FIRST_QUEUED
	OCO_PRIOR
	NUMERO_SERIE


USUARIOS	
	USER_ID
	LOGIN
	NOME
	PASSWORD
	DATA_INC
	DATA_ADMIS
	EMAIL
	FONE
	NIVEL
	AREA
	USER_ADMIN

Figura 10. Esquema das tabelas “ocorrencias” e “usuarios”, usadas pelo Sistema de Abertura de Chamados.

O atributo “contato”, da tabela “ocorrências”, contém o nome do solicitante, que mais adiante se tornará o nome do *host* da sala. No momento da solicitação, o Sistema de Abertura de Chamados resgata esse nome a partir do LDAP da instituição.

O atributo “descricao”, da tabela “ocorrências”, armazena todos os dados digitados pelo usuário para o agendamento da sala. Nestes dados se encontra a breve descrição sobre o assunto da reunião, o número de participantes estimados, o dia e a hora de início e término e o e-mail do solicitante. Esse atributo pode conter o nome do solicitante quando este pedir para que alguma outra pessoa realize este agendamento. Em alguns casos os solicitantes informam se precisam de ajuda no uso da sala. Também existem casos de solicitantes que, ao invés de colocarem a hora de início e hora de término, preenchem a hora de início e o tempo de duração.

O atributo “user_id”, da tabela “usuarios”, contém um código de identificação único para cada usuário cadastrado. O atributo “nome” possui o nome completo deste usuário.

3.6 Abordagem Escolhida

Esta seção relata as decisões arquiteturais que foram estabelecidas. Para tomar essas decisões, foram levadas em consideração as características do problema proposto e as particularidades das arquiteturas de integração presentes na literatura. Com base nessas considerações, foi possível conceber a arquitetura disposta na Figura 11.

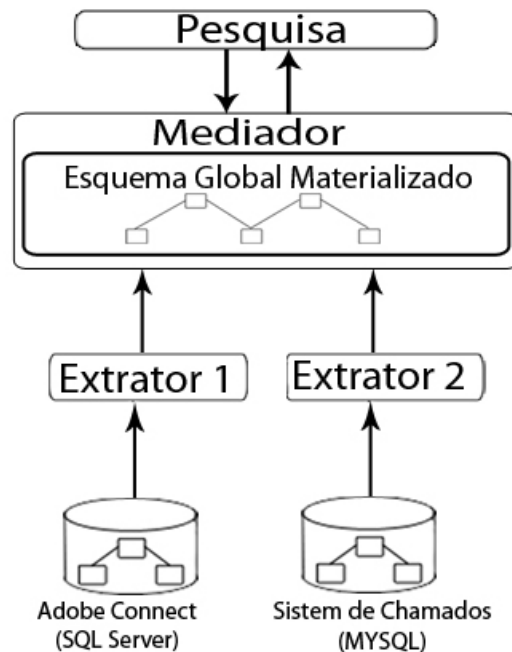


Figura 11. Arquitetura escolhida para o desenvolvimento do trabalho proposto.

Dentre as abordagens virtual e materializada, optou-se pela segunda. Ou seja, as informações das fontes de dados são todas carregadas para o esquema global. Essa decisão foi tomada porque o volume de dados a ser carregado é relativamente pequeno e, conseqüentemente, a atualização do esquema global pode ser feita com certa periodicidade sem acarretar alto custo à instituição. A implementação desta abordagem é mais simples que a virtual, o que também significa um custo menor de implantação. Além disso, o uso desta arquitetura vai proporcionar uma maior rapidez em buscar respostas para as pesquisas do que construir um algoritmo que faça essa checagem de novos dados a serem inseridos.

Como os dados serão materializados, *wrappers* não são necessários, pois não há necessidade de tradução e interpretação de dados. Serão necessários quatro extratores, um para cada tabela do esquema global que será criado, extraindo dados da base de dados do Adobe Connect e da base de dados do Sistema de Abertura de Chamados. O objetivo de cada extrator é extrair as informações relevantes para que o mediador popule o esquema global com essas informações.

3.7 Implementação da Abordagem Escolhida

A principal parte do desenvolvimento deste trabalho envolve a implementação dos extratores. Foi criado um extrator para cada tabela que constitui o esquema global. A Figura 12 mostra o esquema global que modela informações de agendamentos e reuniões. Esse esquema foi construído usando a estratégia bottom-up, pois se partiu de dois esquemas de dados existentes para derivar as tabelas e atributos que precisariam ser criados. Este é o esquema usado para responder as consultas propostas na Seção 3.3.

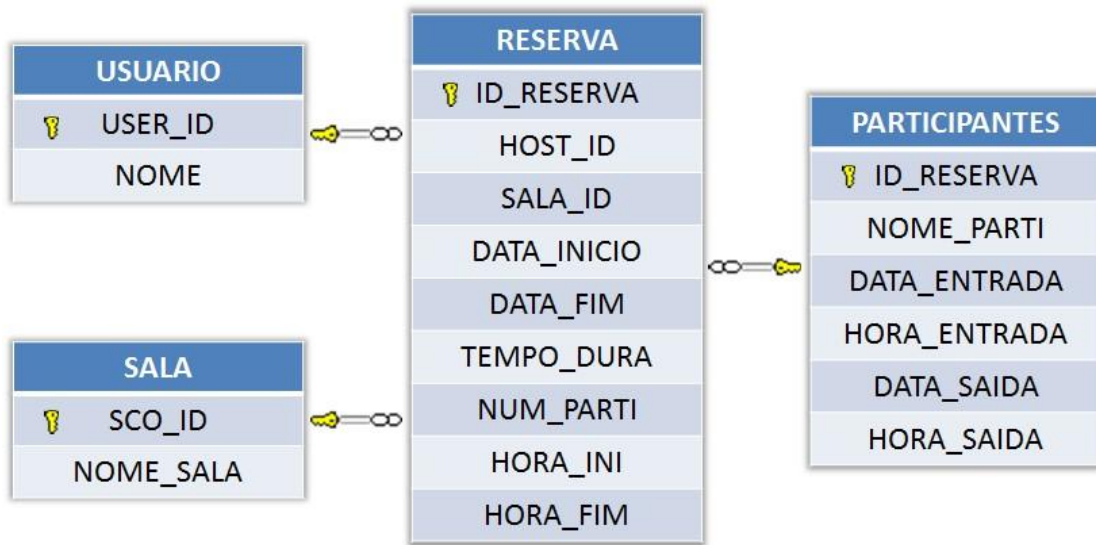


Figura 12. Esquema da base de dados materializada do trabalho.

A próxima etapa envolveu a criação de extratores, cujo propósito é gerar uma base de dados integrada consolidando informações das bases de dados do Adobe Connect e do Sistema de Abertura de Chamados. Para isso a execução dos extratores foi dividida em quatro etapas que são executadas em uma ordem específica. A Figura 13 apresenta a ordem definida para execução dos extratores.

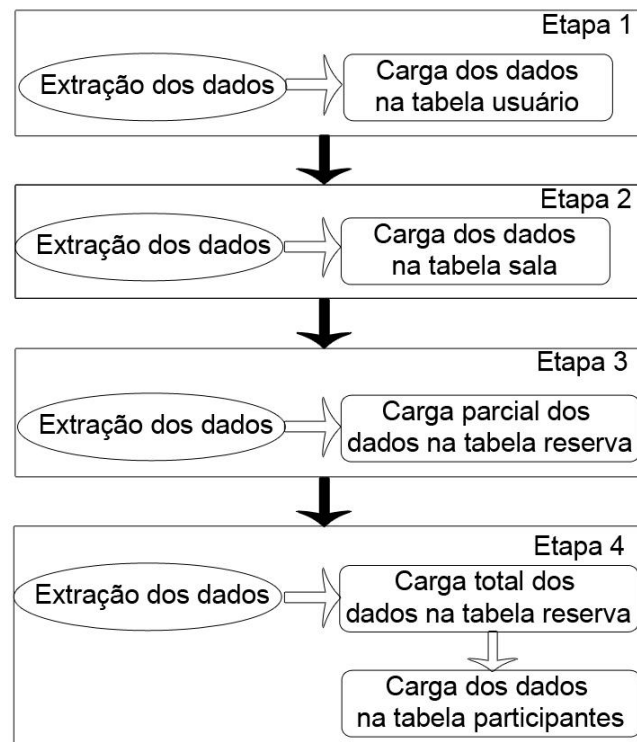


Figura 13. Ordem de execução dos extratores.

I. Primeira Etapa

A primeira etapa consiste em preencher a tabela global “usuarios” com dados vindos do Sistema de Abertura de Chamados. Conforme descrito na Seção 3.5, esses dados encontram-se na tabela local “usuários”. A Figura 14 ilustra os mapeamentos existentes entre a fonte de dados e o esquema global. Os atributos em preto são os mapeados. Campos não utilizados estão em cinza ou foram omitidos e representados na forma de três pontos. Dos atributos disponíveis na fonte, o extrator recupera dados de código de identificação e nome dos usuários e os carrega na tabela global “usuario”.

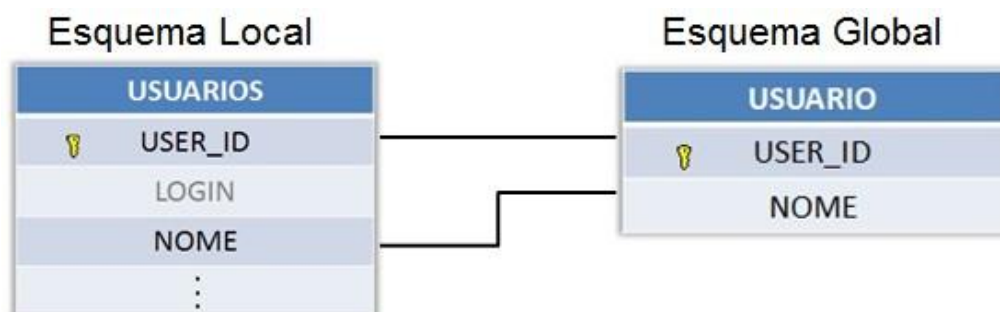


Figura 14. Mapeamento para tabela global “usuario”.

A Figura 15 descreve a consulta SQL utilizada para recuperar esses dados. A consulta recupera todos os usuários que já fizeram alguma solicitação, mesmo aqueles que não tenham

especificamente solicitado uma Webconferência. Isso se deve ao fato de que, no momento da extração, não se sabe quais chamados se referem a Webconferências. Cabe ao extrator da terceira etapa descobrir essa informação, como será visto mais adiante.

```
"SELECT `user_id` , `nome` FROM `usuarios`"
```

Figura 15. Consulta SQL para extração dos usuários.

II. Segunda Etapa

A segunda etapa consiste em preencher a tabela global “sala” com dados advindos do Adobe Connect. Conforme descrito na Seção 3.4, esses dados encontram-se na tabela local “PPS_SCOS”. A Figura 16 ilustra os mapeamentos existentes entre a fonte de dados e a tabela global. Os atributos em preto são os mapeados, como existem alguns campos que não são utilizados estão em cinza e outros foram omitidos e representados na forma de três pontos. Dos atributos disponíveis na fonte, o extrator recupera dados de código e nome da sala e os carrega na tabela global “sala”.

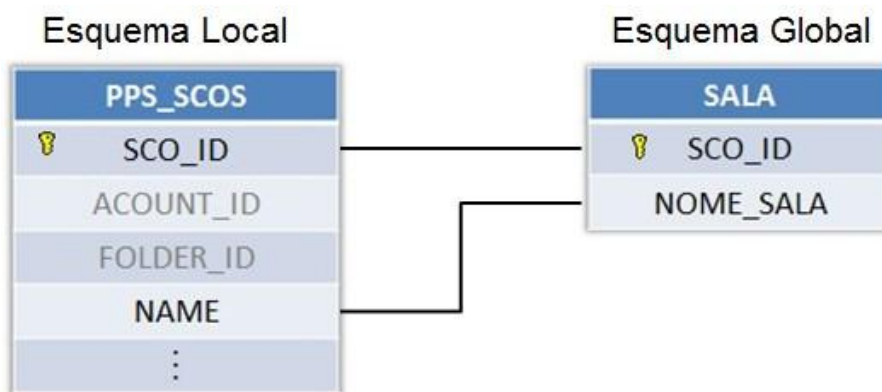


Figura 16. Mapeamento para a tabela global "sala".

A Figura 17 descreve a consulta SQL utilizada para recuperar esses dados. A consulta recupera todas as salas criadas até o momento. Muitas dessas salas nunca foram utilizadas para atender pedidos de reserva, se tratando de informações de testes ou criadas para algum evento específico. Essa informação só será descoberta pelo extrator da quarta etapa, responsável por fazer a análise das entradas e saídas de usuários no sistema de Webconferências.

```
"SELECT [SCO_ID],[NAME] FROM [breeze].[dbo].[PPS_SCOS]"
```

Figura 17. Consulta SQL para extração das salas.

III. Terceira Etapa

A terceira etapa consiste em preencher a tabela global “reserva” com dados advindos do Sistema de Abertura de Chamados. Conforme descrito na Seção 3.5, esses dados encontram-se na tabela local “ocorrencias”. A Figura 18 ilustra os mapeamentos existentes entre a fonte de dados e o esquema global. Os atributos em preto são os mapeados, como existem alguns campos que não são utilizados estão em cinza e outros foram omitidos e representados na forma de três pontos. As linhas pretas indicam mapeamentos 1:n, onde as informações do atributo descrição devem ser extraídas para diversos atributos no esquema global. Já a linha vermelha indica um mapeamento com o atributo já mapeado “host_id”. Este mapeamento deve ser utilizado caso o solicitante não seja indicado no atributo descrição. A regra serve para casos em que o usuário que realizou o agendamento não ser o mesmo que realizará a reunião.



Figura 18. Mapeamento para a tabela global "reserva".

A Figura 19 descreve a consulta SQL utilizada para recuperar esses dados. Para encontrar ocorrências relacionadas à Webconferência foi utilizado um filtro que recupera todos os registros que contenham “webcon” como parte do nome. Com base nos dados da origem, esse filtro tem uma cobertura ótima, recuperando todas as ocorrências relacionadas à Webconferência. No entanto, a precisão é penalizada, pois alguns dos registros recuperados não são de agendamento, e sim de empréstimos de materiais para a realização das Webconferências ou algum outro tipo de reservas para este fim.

```
"SELECT `descricao`, `contato` FROM `ocorrencias` WHERE `descricao` LIKE '%webcon%'"
```

Figura 19. Consulta SQL para extração das reservas.

Para realizar a extração indicada no mapeamento pintado de preto (Figura 18), foi necessário construir um algoritmo composto por expressões regulares (PRADO, 2003). Essas expressões caracterizam-se por conter um grupo de caracteres específicos, que podem ser letras, pontos ou espaços em branco, com isso o algoritmo percorre o bloco até encontrar algo que se encaixa com o padrão desse grupo. Ao encontrar algo similar é possível trabalhar e lapidar a informação encontrada podendo descobrir se ela se tornará útil ao extrator ou não.

Após análise dos textos utilizados no atributo descrição da tabela local “ocorrencias”, advinda do Sistema de Abertura de Chamados, diversas expressões regulares foram propostas, uma para cada atributo de tabela do esquema global. A Figura 20 demonstra a expressão regular usada para reconhecer dados de solicitante. As demais expressões podem ser encontradas no Apêndice I.

```
'/te: ([-.a-zA-Z_]{1,20}) ([-.a-zA-Z_]{1,20}) ([-.a-zA-Z_]{1,20})/'
```

Figura 20. Expressão regular para localizar o solicitante da descrição.

Algumas expressões propostas possuem alto grau de complexidade. Isso se deve ao fato de o atributo descrição não seguir um padrão de escrita, por este motivo a solicitação é preenchida de diversos modos diferentes. Essa diversidade é notada na Figura 21, que apresenta três solicitações distintas que foram recuperadas do Sistema de Abertura de Chamados. Em todos os casos, o mecanismo de extração conseguiu extrair todos os dados relevantes.

Gostaria de solicitar uma webconferência para o dia 27/04/11 (quarta-feira) com início previsto para 10:50 h. e término por volta de 12:30 h. com aproximadamente 10 participantes para tratar de assuntos relativos a capacitação em EAD e NUDEPE. Solicitante: Ricardo Bello Lemos - e-mail: ricardobello@unipampa.edu.br Obs: Não foi possível agendar com maior antecedência.
Solicitante: Ricardo Bello Lemos e-mail: ricardobello@unipampa.edu.br Webconferência para o dia 20/05 (sexta-feira) das 10:00 h. até 17:30 h. (duração de 07: 30 min.) com aproximadamente 15 participantes com a finalidade de tratar de assuntos relativos ao Edital da EAD sobre a Jornada em TIC na Educação entre a Reitoria, PROPG, Campus Alegrete, PROGRAD, PROEXT.
Webconferencia solicitada por Helgona Dantas helgonadantas@unipampa.edu.br Dia 18/04, das 13:30 às 15:30 Teremos 3 participantes e a pauta é implantação do gerenciamento de eventos no portal da unipampa

Figura 21. Algumas solicitações que o extrator conseguiu identificar.

Após a aplicação das expressões regulares, os dados lapidados são importados na tabela global “reserva”, obedecendo aos mapeamentos indicados na Figura 18. Os atributos

são populadas com uma simples cópia do conteúdo existente na origem, com exceção do atributo “host_id”. Para esse caso, dado o contato identificado na origem, é necessário buscar seu respectivo identificador na tabela global “usuário”, populada na primeira etapa. Caso o atributo descrição não contenha dados do contato, ou caso o contato identificado não seja encontrado na tabela usuário, o algoritmo de extração passa a utilizar o próprio atributo contato da solicitação, conforme indicado em vermelho na Figura 18. Como o atributo contato contém dados provindos do Sistema de Abertura de Chamados, pode-se garantir que o valor desse atributo será encontrado na tabela global “usuario”, que foi populada também com dados do Sistema de Abertura de Chamados na primeira etapa. A Tabela 2 exibe alguns resultados gerados na tabela global “reserva”.

id_reserva	host_id	sala_id	data_inicio	data_fim	tempo_dura	num_parti	hora_ini	hora_fim
14	258	NULL	0000-00-00	0000-00-00	00:00:00	0	00:00:00	00:00:00
37	258	NULL	2011-05-30	2011-05-30	03:30:00	8	14:00:00	17:30:00
23	546	NULL	2011-05-07	2011-05-07	00:00:00	0	10:00:00	10:00:00
25	335	NULL	2011-04-29	2011-04-29	02:00:00	0	10:00:00	12:00:00
56	311	NULL	0000-00-00	0000-00-00	01:00:00	0	14:00:00	15:00:00
53	994	NULL	0000-00-00	0000-00-00	02:00:00	20	10:00:00	12:00:00

Tabela 2. Resultados gerados da tabela global “reserva”.

Conforme indicado na Figura 12, a terceira etapa realiza uma carga parcial dos dados de reserva. Isso ocorre porque no momento da solicitação da reunião não é possível determinar a sala que será utilizada. Para descobrir essa informação, é necessário analisar dados do Sistema de Webconferência, como descrito na quarta etapa.

IV. Quarta Etapa

A quarta etapa é dividida em dois estágios. No primeiro deles é feita a carga final da tabela global “reserva”, enquanto o segundo estágio realiza a carga de dados na tabela global “participantes”.

A Figura 22 ilustra um exemplo fictício que ajuda na compreensão dos processos envolvidos nesta etapa. Nesta Figura, a “Tabela A” traz os dados presentes na tabela global “reserva”, após a carga realizada na terceira etapa. Como se pode ver ainda não se conhece o identificador da sala. Já a “Tabela B” traz os dados extraídos do Adobe Connect, referentes às movimentações de entrada e saída de usuários em salas de Webconferência.

Tabela A

ID Sala	ID Reserva	ID Host	Data Início	Data Término	Hora Início	Hora Término	Duração	Nº Part	Nome
NULL	17	231	28/01/2011	28/01/2011	18:30	20:30	2:00	30	Luiz Alberto
NULL	28	350	24/04/2011	24/04/2011	14:00	16:00	2:00	10	João da Silva
NULL	87	419	01/05/2011	01/05/2011	08:00	11:00	3:00	15	Adalberto Sampaio
NULL	12	132	30/06/2011	30/06/2011	15:00	1:00	3:00	40	Demétrio Ribeiro

Tabela B

Nome Sala	ID Sala	Nome Parti	ID Parti	Data Início	Data Término
Introdução a Computação	3291	Andressa	NULL	30/03/2011 21:23:54	NULL
Introdução a Computação	3291	Monica	NULL	30/03/2011 21:43:32	NULL
Reunião Sistemas	2324	Adalberto Sampaio	419	01/05/2011 07:59:01	01/05/2011 09:01:22
Reunião Sistemas	2324	Paulo	NULL	01/05/2011 08:00:23	NULL
Reunião Sistemas	2324	Luiz Amaral	785	01/05/2011 08:10:31	01/05/2011 09:52:54
Reunião Sistemas	2324	Adalberto Sampaio	419	01/05/2011 09:35:51	01/05/2011 09:57:32
Sala Campus	1700	Luiz Roberto	602	13/05/2011 11:10:32	13/05/2011 12:52:34
Sala Campus	1700	Maria	NULL	13/05/2011 11:03:12	NULL

Figura 22. Ilustração das tabelas de exemplo.

Com esses dados, criou-se um critério para determinar quais entradas de usuários correspondem a uma reserva em particular. Para cada reserva, delimita-se um bloco de registros, sendo que o primeiro e o último registro do bloco devem obedecer a regras pré-definidas. O primeiro registro é aquele cuja data de início da reserva (Data Início da “Tabela A”) for igual à data de entrada (Data Início da “Tabela B”) e o nome do solicitante da reserva (Nome da “Tabela A”) for igual ao nome do participante na reunião (Nome Parti da “Tabela B”). Já o último registro passa a ser o último registro consecutivo ao primeiro que possua duas informações em comum com o primeiro: ambos usam a mesma sala (ID Sala) e tem a mesma data de início (Data Início). Nota-se que os códigos de identificação das tabelas A e B (ID Host e ID Parti) são diferentes, pois na “Tabela A” foi utilizado o código do Sistema de Abertura de Chamados e os dados da “Tabela B” provém do Adobe Connect.

As movimentações de entrada e saída de usuários podem ser recuperadas através da consulta SQL da Figura 23, que realiza a junção entre a tabela de movimentações (PPS_TRANSCRIPTS), a tabela que contém as salas (PPS_SCOS) e a tabela que contém as sessões de cada usuário (PPS_USER_SESSIONS).

```

"SELECT
    sala.name, sala.sco_id,
    usuario.name,
    reserva.principal_id, reserva.date_created, reserva.date_closed
FROM
    [breeze].[dbo].[PPS_TRANSCRIPTS] reserva
    INNER JOIN
        [breeze].[dbo].[PPS_USER_SESSIONS] usuario
    ON
        reserva.session_id = usuario.session_id
    INNER JOIN
        [breeze].[dbo].[PPS_SCOS] sala
    ON
        reserva.sco_id = sala.sco_id"

```

Figura 23. Consulta SQL para extração dos participantes.

Tendo delimitadas as entradas e saídas de cada reunião, é possível realizar tanto a carga final da tabela global “reserva” quanto à carga de dados na tabela global “participantes”. Para a carga da tabela “reserva”, basta recuperar o código de identificação da sala de qualquer um dos registros do bloco (todos usam a mesma sala). Para a carga da tabela “participante”, deve-se gerar um registro na tabela global “participantes” para cada registro do bloco.

De modo geral, o processo de mapeamento descrito acima pode ser representado visualmente conforme indicado na Figura 24. Os atributos em preto são os mapeados, como existem alguns campos que não são utilizados estão em cinza e outros foram omitidos e representados na forma de três pontos. No lado esquerdo, a linha vermelha indica os dados a serem convertidos. Já as linhas tracejadas indicam as correspondências que devem ser satisfeitas para encontrar os registros corretos. Já no lado direito, os mapeamentos indicam os dados a serem convertidos para a tabela global “participantes”. Cabe ressaltar que, para preencher a tabela global “participantes”, leva-se em consideração o bloco de dados referente a cada reserva específica, conforme indicado na Figura 22. Por fim, as linhas representadas com uma chave de um lado e o sinal do infinito do outro, indicam os relacionamentos de chave estrangeira usados para gerar as consultas SQL que extraem dados das tabelas locais.

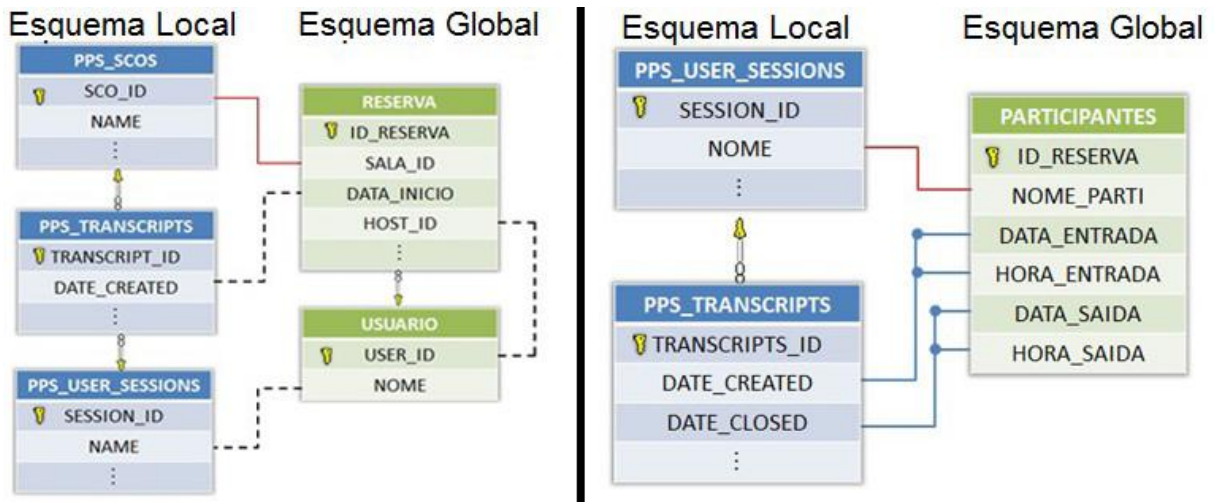


Figura 24. Mapeamentos para a carga total da “reserva”.

4. RESULTADOS

Para validar o trabalho, as consultas propostas foram executadas, e seus resultados foram analisados. A Figura 25 apresenta a interface gráfica desenvolvida, utilizada pelo usuário administrador, facilitando a análise dos resultados obtidos através da integração proposta neste trabalho.

Executar o Sctrip novamente, clique [AQUI](#) e aguarde

Consulta 1: Verificar o número de participantes de cada reunião solicitada e realizada pelo Adobe Connect:
Participantes de cada reunião:

ID Reserva	Nome do Usuário	Nome da Sala	N. Participantes	ID Sala
14		Sala 1007	18	76474
23		Sala 1006	1	75637
25		Reunião Sistemas	31	26134
37		Sala 1007	12	76474

Consulta 2: Verificar se a reserva de salas pelo host está sendo realizada de maneira adequada
Possíveis reservas não realizadas:

ID Host	Host	Data	Hora Início	Hora Fim	N. Participantes
592		0000-00-00	14:30:00	17:00:00	0
592		2011-11-09	10:00:00	12:00:00	3
592		2011-11-09	10:00:00	12:00:00	4
258		0000-00-00	14:00:00	17:00:00	0
258		0000-00-00	20:30:00	22:30:00	0
258		0000-00-00	20:30:00	22:20:00	0
258		0000-00-00	13:00:00	18:30:00	0
258		0000-00-00	20:30:00	22:30:00	0
258		0000-00-00	00:00:00	00:00:00	0

Consulta 3: Verificar se o tempo estipulado pelo host está sendo utilizado da maneira adequada:

ID Reserva	Hora Início	Hora Fechamento	Duração	Duração Solicitada	Situação
14	11:38:58	21:08:53	09:29:55	00:00:00	Tempo não determinado!
23	12:10:01	15:37:11	03:27:10	00:00:00	Tempo não determinado!
25	12:56:35	18:26:59	05:30:24	02:00:00	Estourou o tempo solicitado
37	16:50:35	00:00:00	Hora NULL!	03:30:00	Tempo não determinado!

Consulta 4: Verificar se o número de participantes solicitado pelo host está sendo utilizado da maneira adequada.

ID Reserva	Sol. Participantes	Entrou Participantes	Entrou Participantes
14	0	18	Solicitante não determinado!
23	0	1	Solicitante não determinado!
25	0	31	Solicitante não determinado!
37	8	12	Estourou as pessoas solicitadas

Figura 25. Visão da tela do administrador.

Neste capítulo, gráficos são usados para ilustrar a proporção de registros encontrados em cada uma das consultas. Para construir os resultados, foram utilizadas as duas bases de dados para conseguir responder todas as consultas. Estas bases de dados utilizadas não são as mais atuais, contendo informações referentes aos anos de 2010 e 2011.

Para a realização dos testes foi utilizado duas configurações: uma com os dados reais obtidos do NTIC e outra com estes mesmos dados, porém com a adição de oito novos registros. A segunda configuração (com dados fictícios) serve como forma de validar se o sistema desenvolvido realmente extrai e integra os registros conforme esperado.

A Figura 26 visa responder a primeira consulta proposta, utilizando a primeira configuração de dados, na qual se deseja verificar o número de participantes que consta em cada reunião solicitada e que foi realizada no Adobe Connect. Com esta verificação é possível saber quantas pessoas participaram em cada reunião. É possível observar que das quatro reservas encontradas pelo trabalho desenvolvido, foi encontrado um registro da sala “1006” (com um participante), dois registros da sala “1007” (com 18 e 12 participantes) e um registro da sala “Reunião Sistemas” (com 31 participantes). O número baixo de registros recuperados pela consulta deve-se ao fato de ter-se usado uma base de testes local e desatualizada, porém esta base contém dados reais de reuniões realizadas. Espera-se que o sistema possa ser testado sobre as bases reais, para que as informações retornadas sejam mais relevantes. Contudo, mesmo com o número baixo de resposta, já se pode perceber que algumas salas são subutilizadas, onde a mesma sala é utilizada mais de uma vez para outra reunião.

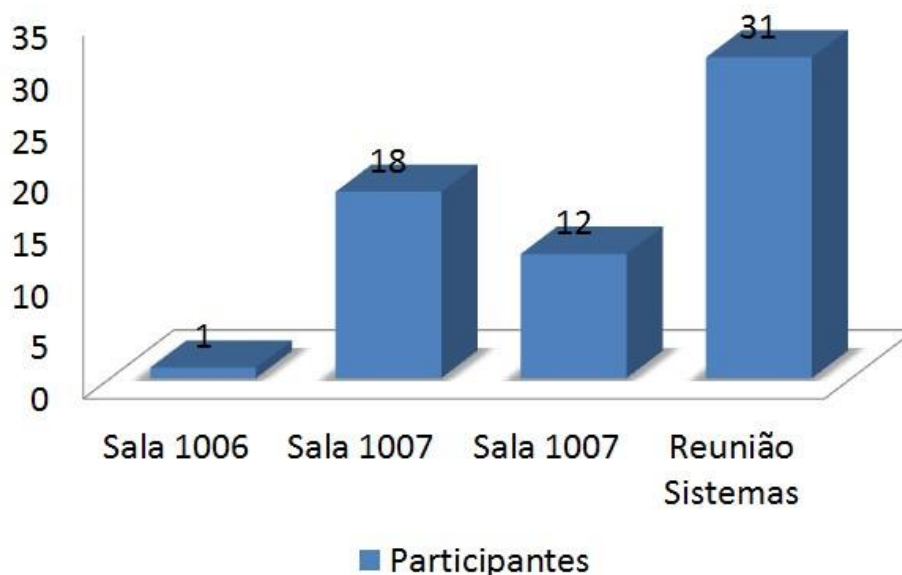


Figura 26. Resultados para a primeira consulta.

A Figura 27 visa responder a segunda consulta proposta, utilizando a primeira configuração de dados, na qual se deseja obter as reservas que foram solicitadas pelo Sistema de Abertura de Chamados e que não constam no Adobe Connect. É possível observar que das 82 reservas encontradas pelo trabalho desenvolvido, apenas quatro delas foram detectadas na base de dados do Adobe Connect, ou seja, menos de 5%. Disso conclui-se que, ou a reunião

não foi realizada ou foi realizada em outro horário não compatível com os dados informados na solicitação. Outro motivo para este número ser tão alto, das reuniões não realizadas, é que alguns pedidos capturados são de empréstimos de materiais para a realização de Webconferências. Também existem casos que o usuário solicita Webconferência pensando que é Vídeo Conferência, e acaba não usufruindo os recursos agendados. De qualquer forma, o gráfico traz indícios que podem levar o administrador do sistema de Webconferência a tomar medidas que evitem esse tipo de ocorrência.

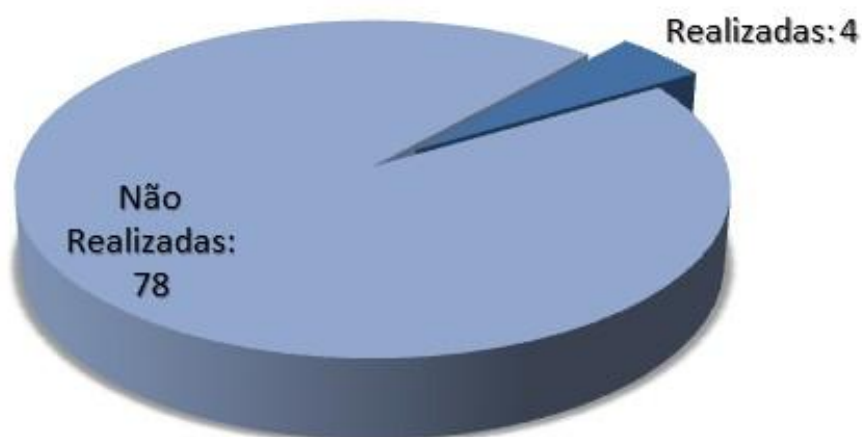


Figura 27. Resultados para a segunda consulta.

A Figura 28 visa responder a terceira consulta proposta, utilizando a primeira configuração de dados, na qual se deseja verificar se o tempo solicitado pelo Sistema de Abertura de Chamados está de acordo com o tempo de duração que consta no Adobe Connect. Com esta verificação é possível determinar como o tempo de cada reunião está sendo utilizado. É possível observar que das quatro reservas encontradas pelo trabalho desenvolvido, nenhuma está dentro do tempo estipulado: em três delas o usuário não registrou o tempo de duração e uma ultrapassou o tempo solicitado. Com isso, é possível deduzir que os usuários que estão solicitando as Webconferências não estão informando de maneira adequada as informações solicitadas no Sistema de Abertura de Chamados.



Figura 28. Resultados para a terceira consulta.

A Figura 29 visa responder a quarta consulta proposta, utilizando a primeira configuração de dados, na qual se deseja verificar se o número de participantes solicitado pelo Sistema de Abertura de Chamados está de acordo com o que foi realizado no Adobe Connect. Com esta verificação é possível determinar como está o movimento de determinadas reuniões, podendo assim determinar se o uso desta sala está sendo realizado de acordo. É possível observar que das quatro reservas encontradas pelo trabalho desenvolvido, nenhuma está dentro do número de participantes solicitados: em três delas o solicitante não informou o número de participantes e uma ultrapassou o número de participantes solicitados.



Figura 29. Resultados para a quarta consulta.

Como se pode constatar, os resultados atuais não provam totalmente que as consultas estão sendo respondidas corretamente. Por este motivo foi necessário utilizar a segunda configuração de base para testes, onde foram inseridos na base do Sistema de Abertura de Chamados oito novos registros. Esses registros foram inseridos de forma que combinassem com as informações existentes na base do Adobe Connect. Ao executar a integração foram capturados os novos registros e a partir deles foram gerados os resultados a seguir.

A Figura 30 exibe os dados referentes à primeira consulta utilizando a segunda configuração dos dados. É possível observar que as oito reservas inseridas foram capturadas, totalizando 12 reservas encontradas. Foi encontrado um registro da sala “Conselho Gestor de TIC – UNIPAMPA” (com 15 participantes), um registro da sala “Fundo da Reunião Pública” (com quatro participantes), um registro da sala “Grupo de Pesquisa em Neuromecânica aplicada (GNAP)” (com quatro participantes), um registro da sala “GT Segurança e Recursos Computacionais” (com dois registros), dois registro da sala “NTIC – CGTIC – GA CDPDTIC” (com 24 e dois participantes), um registro da sala “Obras” (com três participantes), dois registros da sala “Reunião Sistemas” (com seis e 31 participantes), um registro da sala “1006” (com um participante) e dois registros da sala “1007” (com 18 e 12 participantes).

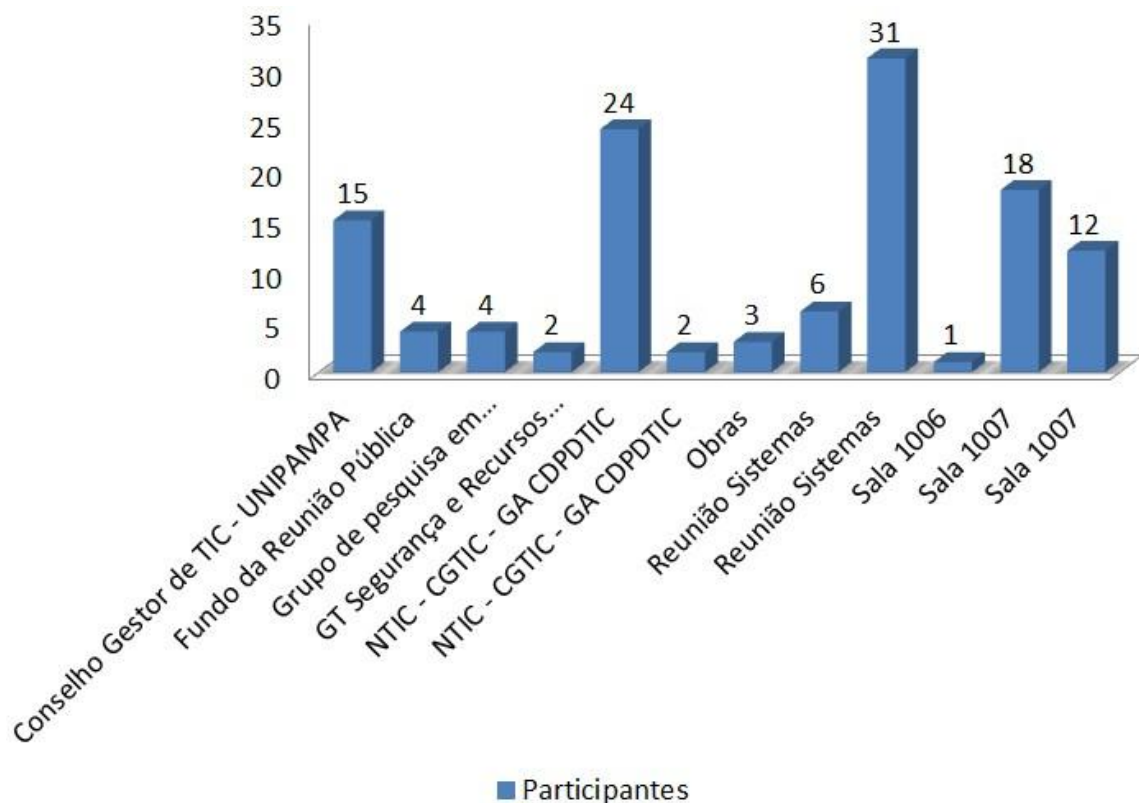


Figura 30. Resultados para primeira consulta, utilizando a segunda base de testes.

A Figura 31 exibe os dados referentes à segunda consulta utilizando a segunda configuração dos dados. É possível observar que as oito reservas inseridas foram capturadas. Totalizando 90 registros encontrados, sendo que deste valor 12 constam como realizadas e 78 não realizadas.

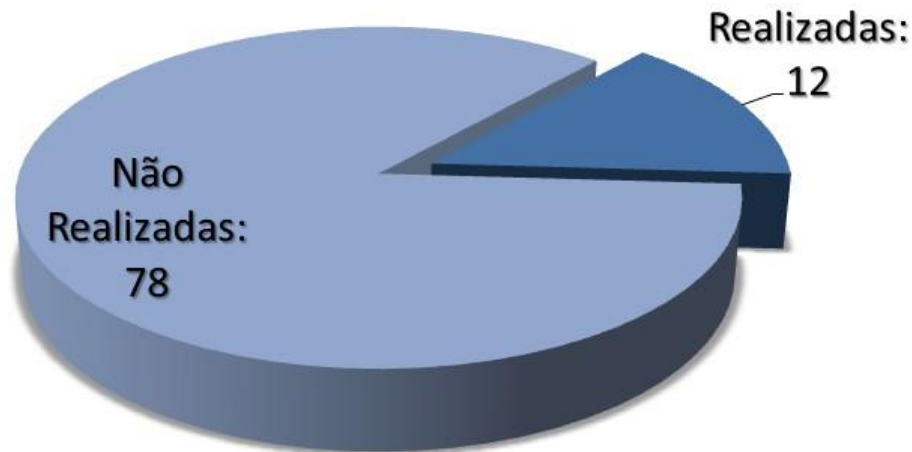


Figura 31. Resultados para segunda consulta, utilizando a segunda base de testes.

A Figura 32 exibe os dados referentes à terceira consulta utilizando a segunda configuração dos dados. É possível observar que as oito reservas inseridas foram capturadas, sendo inseridos seis reservas dentro do tempo estipulado e duas reservas que ultrapassaram o tempo determinado. Totalizando 12 registros encontrados com seis reservas dentro do tempo estipulado, três reservas que ultrapassaram este tempo e três reservas que não informaram o tempo.



Figura 32. Resultados para terceira consulta, utilizando a segunda base de testes.

A Figura 33 exibe os dados referentes à quarta consulta utilizando a segunda configuração dos dados. É possível observar que as oito reservas inseridas foram capturadas, sendo inseridos sete reservas dentro do número de participantes solicitados e uma reserva que

ultrapassou os participantes solicitados. Totalizando 12 registros encontrados com sete reservas dentro do solicitado, duas reservas que ultrapassaram este limite e três reservas que não foi informado pelo usuário este limite.



Figura 33. Resultados para quarta consulta, utilizando a segunda base de testes.

5. CONCLUSÃO

A integração de dados é uma técnica que vem sendo aplicada em diversos tipos de sistemas, tornando-os mais práticos e de acesso transparente ao usuário. Conforme demonstrado no trabalho, técnicas voltadas à integração podem ser utilizadas para resolver o problema que foi apresentado.

No decorrer do trabalho, foram vistas algumas possibilidades de integração, desde abordagens mais simples, restritas a poucas fontes de dados, até abordagens complexas voltadas para fontes de dados mais diversificadas. Com base nessas informações, e dadas as características do problema exposto, optou-se pelo uso de uma arquitetura mais simples, baseada em materializações totais e uso de extratores.

A partir dessa definição, os dois maiores desafios residiram na construção dos extratores e no processo de unificação dos registros realizado pelo mediador. Quanto aos extratores, destaca-se a complexidade na obtenção de dados a partir de atributos multivalorados, como por exemplo, o atributo “descricao” da tabela “ocorrencia”.

Quanto à unificação, os dados extraídos das fontes de dados tiveram que ser combinados no esquema global para que as informações complementares fossem vistas de modo integral. Para que o processo de unificação ocorresse, foi necessário estudar quais atributos os modelos de dados distintos possuem em comum, e analisar a possibilidade do uso de métodos de unificação, que combinem os dados com base nesses atributos.

Outra parte relevante deste trabalho foi a própria modelagem do esquema global materializado no mediador. Essa tarefa foi realizada criteriosamente, para que o modelo não ficasse por demais restritivo e oferecesse pontos de extensão.

Como trabalhos futuros, destaca-se a necessidade de utilizar extratores que recuperem um número maior de solicitações de sala. Para isso, uma possibilidade envolve o uso de técnicas de aprendizado de máquina (BATISTA, 2003). Outra possibilidade envolve a adoção de um template de solicitação de salas. Com o uso de um template, a recuperação das informações necessárias seria bastante simplificada. Além disso, pretende-se analisar a viabilidade de uso da ferramenta pelos administradores de sistemas do NTIC. Para isso, novos testes seriam necessários, utilizando dados mais atuais. Dependendo da aceitação, pode-se adaptar a ferramenta para que novos recursos sejam adicionados ou para que a arquitetura seja alterada, como por exemplo, para dar suporte a atualizações incrementais.

6. REFERÊNCIAS

- ANGER, Luciane T. W.. “**Manutenção Incremental de Visões Materializadas em Ambientes DataWarehouse**”. Universidade Federal do Paraná. Curitiba, 2000.
- ANZANELLO, Cynthia A. “**OLAP Conceitos e Utilização**”. Instituto de Informática – Universidade Federal do Rio Grande do Sul. Porto Alegre, 2007
- BATISTA, G. E. A. P. A.. **Pré-processamento de Dados em Aprendizado de Máquina Supervisionado**. Tese de Doutorado, ICMC-USP, 2003.
- CARVALHO, Grulherme B., “**Voz sobre IP: uma alternativa econômica**”.
http://www.cgeti.deti.ufc.br/monografias/GUILHERME_BARRETO_CARVALHO.pdf.
Acessado em: Outubro 2011.
- CAVALCANTE, André L. D.. “**Uma Arquitetura Aberta para Integração de Dados do Ambiente Fabril com os Níveis Superiores da Empresa**”. Dissertação (Mestrado em Telecomunicações), Universidade Federal do Rio de Janeiro, 2005.
- CIELO, Ivã Rafael. Etl: “**Extração, transformação e carga de dados**”.
http://www.angelicatoffano.pro.br/upload_arquivos/pt/etl.pdf. Acesso em: Novembro 2011.
- CIPRIANO, Otávio N.. “**Replicação de Bases de Dados Postgrsql Utilizando Pgcluster**”. Trabalho de conclusão de curso. LAVRAS, Minas Gerais. 2009.
- ELMASRI, Rames; NAVATHE, Shamkant B.. “**Sistema de Banco de Dados**”. 6ª edição, editora: Pearson. São Paulo, 2011.
- FERNANDES, Anita R.; SCHEUER, Leandro; MOREIRA, Daniela S.; CUNHA, Fernanda. “**Aplicação de Raciocínio Baseado em Casos em Service Desk**”. Apresentado em: VII SEGeT – Simpósio de Excelência em Gestão e Tecnologia. UNIVALE, São José - SC, 2010.
- LAENDER, Alberto H. F.; Ribeiro-Neto, Berthier A.; Silva, Altigran S.; Teixeira, Juliana S. “*A brief survey of web data extraction tools.*” SIGMOD, 2002.
- LEE, Huei D.. “**Seleção de Atributos Importantes para a Extração de Conhecimento de Bases de Dados**”. Tese de Doutorado Universidade de São Paulo (USP/ICMC), São Carlos – SP, 2005.
- LIMA, Debora A. “**Interface Gráfica para busca de Livros Minerva**”. Monografia – Universidade Federal do Rio de Janeiro. Rio de Janeiro, 2011.
- LORÊDO H. Q.; FERREIRA L. N.; ASSIS G. T.. “**Replicação assíncrona entre bancos de dados heterogêneos: uma abordagem prática**”. Publicado em: Congresso brasileiro de computação, IV. Itajaí, 2004.
- MACHADO, F. N. R. “**Tecnologia e projeto de data warehouse: uma visão multidimensional**”. São Paulo: Érica, 2004.

- MERGEN, Sergio L.; Heuser, Carlos A. “**Casamento de Esquemas XML e Esquemas Relacionais**”. Dissertação de Mestrado. Porto Alegre, 2007.
- MONTARDO, Sandra P.. “**A busca é a mensagem: Links Patrocinados e Marketing de Otimização de Busca (MOB)**”. Tazón Y Palabra, Campus Estado de México, 2006.
- NAVEGA, S. ; “**Princípios Essenciais do Data Mining**”. Em INFOIMAGEM 2002, CENADEM.
- NUNES, Marcos F. “**Avaliação experimental de uma técnica de padronização de escores de similaridade**”. Dissertação (Mestrado), Porto Alegre PPGC da UFRGS, 2009.
- OLIVEIRA, P., Henriques, P.; Rodrigues, F. “**Limpeza de Dados – Uma Visão Geral**”. Publicado em: Proceedings of Data Gadgets' 2004 Workshop – Bringing Up Emerging Solutions for Data Warehousing Systems. Málaga, Espanha. Novembro, 2004.
- OLIVEIRA, Wilson José de. “**Data Warehouse**”. Publicado em: Visual Books. Florianópolis, 2002.
- ORTIZ, L. C.; ORTIZ, W. A.; SILVA, S. L.da.. “**Ferramentas alternativas para monitoramento e mapeamento automatizado do conhecimento**”. Ciência da Informação, Brasília, 2002.
- PRADO, T.A.S. e Nunes, M.G.V. “**Segmentação Textual Automática: Uma Revisão Bibliográfica. Série de Relatórios Técnicos do Instituto de Ciências Matemáticas e de Computação - ICMC**”, Universidade de São Paulo, Fevereiro 2003.
- RIBEIRO, Luis C. R... “**Identificação de áreas de interesse a partir da extração de informações de currículos Lattes/XML**”. Publicado em: I Escola Regional de Banco de Dados. Porto Alegre: UFRGS, 2005.
- RUSCHEL, M., Zanotto, M. and Mota W.. “**Computação em Nuvem**”. Pontificia Universidade Católica do Paraná, Curitiba, 2010.
- SALGDO, A.C.; Lóscio, B.F. “**Integração de Dados na Web**”, Publicado em: Anais da VI Escola Regional de Informática, São Carlos, 2001.
- SCHNEIDER, F.B.. “**Implementing Fault-Tolerant Service Using the State Machine Approach: A Tutorial**”. ACM Computing Survey. 1990.
- SILVERSCHATZ, Abraham; KORTH, Henry F.; SUDARSHAM, S.. “**Sistema de Banco de Dados**” – tradução de Daniel Viera. Rio de Janeiro: Elsevier, 2006.

7. APÊNDICE I - EXPRESSÕES REGULARES

```
'/\d{1,2}\/\d{1,2}\/\d{1,4}';
'| \d{1,2}\/\d{1,2}';
'| \d{1,2} de ([-a-zA-Zç_]{1,20})/';
'/ de ([-.a-zA-Zç ]{1,20})/';
```

Expressão regular para Descrição.

```
'/([0-9]|[0-1][0-9]|[2][0-3])h([0-5][0-9])';
'| ([0-9]|[0-1][0-9]|[2][0-3])h| ([0-1][0-9]|[2][0-3]) h';
'| ([0-1][0-9]|[2][0-3]) h';
'| ([0-1][0-9]|[2][0-3])H';
'| ([0-9]|[0-1][0-9]|[2][0-3]):[0-5][0-9]/';
```

Expressão regular para Hora de Início.

```
'/te: ([-.a-zA-Z_]{1,20}) ([-.a-zA-Z_]{1,20}) ([-.a-zA-Z_]{1,20})/';
```

Expressão regular para Solicitante.

```
'/([0-9]|[0-1][0-9]|[2][0-3])h([0-5][0-9])min às ([0-9]|[0-1][0-9]|[2][0-3])h([0-5][0-9])';
'| ([0-9]|[0-1][0-9]|[2][0-3])h([0-5][0-9])min às ([0-9]|[0-1][0-9]|[2][0-3])h';
'| ([0-9]|[0-1][0-9]|[2][0-3])h([0-5][0-9])min às ([0-9]|[0-1][0-9]|[2][0-3]) h';
'| ([0-9]|[0-1][0-9]|[2][0-3])h([0-5][0-9]) e ([0-9]|[0-1][0-9]|[2][0-3])h([0-5][0-9])';
'| ([0-9]|[0-1][0-9]|[2][0-3])h e ([0-9]|[0-1][0-9]|[2][0-3])h([0-5][0-9])';
'| ([0-9]|[0-1][0-9]|[2][0-3])h([0-5][0-9]) e ([0-9]|[0-1][0-9]|[2][0-3])h';
'| ([0-9]|[0-1][0-9]|[2][0-3])h e ([0-9]|[0-1][0-9]|[2][0-3])h';
'| ([0-9]|[0-1][0-9]|[2][0-3]):([0-5][0-9]) e ([0-9]|[0-1][0-9]|[2][0-3]):([0-5][0-9])';
'| ([0-9]|[0-1][0-9]|[2][0-3])h as ([0-9]|[0-1][0-9]|[2][0-3])h';
'| ([0-9]|[0-1][0-9]|[2][0-3]):([0-5][0-9]) h. e término por volta de ([0-9]|[0-1][0-9]|[2][0-3]):([0-5][0-9])';
'| ([0-9]|[0-1][0-9]|[2][0-3])horas e ([0-9]|[0-1][0-9]|[2][0-3]) h';
'| ([0-9]|[0-1][0-9]|[2][0-3]):([0-5][0-9]) h. até ([0-9]|[0-1][0-9]|[2][0-3]):([0-5][0-9])';
'| ([0-9]|[0-1][0-9]|[2][0-3])h até às ([0-9]|[0-1][0-9]|[2][0-3])h';
'| ([0-9]|[0-1][0-9]|[2][0-3])h([0-5][0-9]) até ([0-9]|[0-1][0-9]|[2][0-3])h([0-5][0-9])';
'| ([0-9]|[0-1][0-9]|[2][0-3])h([0-5][0-9]) até ([0-9]|[0-1][0-9]|[2][0-3])h';
'| ([0-9]|[0-1][0-9]|[2][0-3])hs às ([0-9]|[0-1][0-9]|[2][0-3])h([0-5][0-9])';
'| ([0-9]|[0-1][0-9]|[2][0-3])h([0-5][0-9]) às ([0-9]|[0-1][0-9]|[2][0-3])h([0-5][0-9])';
'| ([0-9]|[0-1][0-9]|[2][0-3])h às ([0-9]|[0-1][0-9]|[2][0-3])h([0-5][0-9])';
'| ([0-9]|[0-1][0-9]|[2][0-3]) às ([0-9]|[0-1][0-9]|[2][0-3])h';
'| ([0-9]|[0-1][0-9]|[2][0-3])h às ([0-9]|[0-1][0-9]|[2][0-3])h';
'| ([0-9]|[0-1][0-9]|[2][0-3])h às ([0-9]|[0-1][0-9]|[2][0-3]) h';
'| ([0-9]|[0-1][0-9]|[2][0-3]):([0-5][0-9]) às ([0-9]|[0-1][0-9]|[2][0-3]):([0-5][0-9])/';
```

Expressão regular para Duração, primeira parte.

```
'/Duração: [0-9]';
'|duração: [0-9]';
'|prevista: ([0-9])h([0-5][0-9])';
'|prevista: [0-9]/';
```

Expressão regular para Duração, segunda parte.

```
'/Participantes: ([0-9][0-9][0-9]|[0-9][0-9]|[0-9])';
'|participantes: ([0-9][0-9][0-9]|[0-9][0-9]|[0-9])';
'|([0-9][0-9][0-9]|[0-9][0-9]|[0-9]) participantes';
'|em torno de ([0-9][0-9][0-9]|[0-9][0-9]|[0-9])';
'|até ([0-9][0-9][0-9]|[0-9][0-9]|[0-9]) pessoas';
'| aproximadamente ([0-9][0-9][0-9]|[0-9][0-9]|[0-9])/';
```

Expressão regular para Participantes, primeira parte.

```
'/([0-9][0-9][0-9]|[0-9][0-9]|[0-9])/';
```

Expressão regular para Participantes, segunda parte.