Capítulo

7

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Projeto – Estudo de Caso

* 1. **Introdução**

O Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Norte (IFRN), nova institucionalidade dada pelos termos da Lei 11.829, de 29 de dezembro de 2008, faz parte da rede federal de educação profissional e tecnológica, vincula-se ao Ministério da Educação, possui natureza jurídica de autarquia e detém autonomia administrativa, patrimonial, financeira, didático-pedagógica e disciplinar. Trata-se de uma instituição de educação superior, básica e profissional, especializada na oferta de educação profissional e tecnológica nas diferentes modalidades de ensino, conjugando conhecimento científico, técnico e tecnológico a ideais pedagógicos de fundamentação histórico-crítica.

Com estrutura multicampis, o IFRN está sediado na Reitoria, localizada na Rua Dr. Nilo Bezerra Ramalho, nº 1692, Tirol, Natal-RN. É composto, tomando-se, o ano de 2014 como referência, por vinte e um campus.

De organização pluricurricular, o IFRN oferece um ensino público laico, gratuito e de qualidade. Oferta, nesse sentido, cursos em sintonia com a função social que desempenha, visando a consolidação e, o fortalecimento, dos arranjos produtivos, culturais e sociais locais. Apresenta, para tanto, um currículo organizado a partir de quatro eixos – ciência, trabalho, cultura e tecnologia – que atuam, de modo entrelaçado e Inter complementar, como princípios norteadores da prática educativa. O Instituto desenvolve à pesquisa e a extensão, na perspectiva da produção, socialização e difusão de conhecimentos. Estimula a produção cultural e realiza processos pedagógicos que levem à geração de trabalho e renda. Em um contexto mais amplo, a Instituição visa contribuir para as transformações da sociedade, visto que esses processos educacionais são construídos nas relações sociais.

No que concerne à comunidade acadêmica, há os sujeitos sociais diretamente envolvidos com os processos pedagógicos e administrativos do IFRN. Essa comunidade é constituída por três seguimentos: estudantes, professores e técnicos-administrativos. Numa perspectiva mais abrangente, acrescenta-se, a esse coletivo, a comunidade local, composta tanto por pais dos estudantes e/ou responsáveis pelos estudantes quanto representantes da sociedade civil.

Para efeito de regulação, avaliação e supervisão da Instituição e dos cursos de educação superior, equipara-se às universidades federais. Além de se submeter à legislação federal específica, rege-se pelos seguintes instrumentos normativos: estatuto; regimento geral; regimento interno dos campis e dos demais órgãos componentes da estrutura organizacional dos institutos federais; resolução do Conselho Superior (CONSUP); deliberações do Colégio de Diretrizes (CODIR) e do Conselho de Ensino, Pesquisa e Extensão (CONSEPEX); e atos da Reitoria.

A expansão do IFRN amplia, significativamente, a atuação nas áreas de ensino, de pesquisa e de extensão; contribui, de modo mais extensivo, para a formação humana e cidadã; e estimula o desenvolvimento socioeconômico, à medida que potencializa soluções científicas, técnicas e tecnológicas, com compromisso de estender benefícios à comunidade.

Essa ampla abrangência em todo território norte-rio-grandense contribui para posicionar tanto o IFRN como uma instituição de educação, ciência e tecnologia quanto os seus campis como elos de produção de conhecimento e de desenvolvimento social. Garante, assim, a manutenção da respeitabilidade junto às comunidades nas quais os campis se inserem e da credibilidade construída ao longo da história da Instituição [Projeto Político-Pedagógico do IFRN].

* 1. **Histórico**

Para uma melhor compreensão do perfil institucional, na perspectiva de consolidar a função social, os objetivos e os princípios orientadores do instituto, faz-se necessário uma análise histórica, políticas e sociais do IFRN.

Criada pelo Decreto 7.566, de 23 de setembro de 1909, como Escola de Aprendizes Artífices, a instituição passou por diversas mudanças e recebeu várias denominações ao longo do tempo. Instalada, inicialmente, em janeiro de 1910, no prédio do antigo Hospital da Caridade, na Praça Coronel Lins Caldas, nº 678, Cidade Alta, onde hoje funciona a casa de estudantes de Natal, a Escola de Aprendizes Artífices oferecia cursos primários de desenho e oficinas de trabalhos manuais. Em 1914, o estabelecimento foi transferido para a Avenida Rio Branco, nº 743, ocupando, durante cinquenta e três anos, um edifício construído no início do século XX.

Mais tarde, ocorreu a mudança de denominação para Liceu Industrial de Natal, orientada pela reforma instituída pela Lei 378, de 13 de janeiro de 1937, do Ministério da Educação e Saúde, órgão a que a Instituição estava subordinada desde 1930. Na época, eram oferecidas oficinas de desenho, de sapataria, de funilaria, de marcenaria e de alfaiataria, inspiradas, segundo Meireles (2006, p.55), em “modelos exteriores ao Brasil, o que evidencia a influência de outros formatos culturais, educacionais, tecnológicos e produtivos na realidade brasileira do século XX”.

Designada como Escola Industrial de Natal (EIN), no ano de 1942, após a promulgação da Lei Orgânica do Ensino Industrial, a Instituição transformou as oficinas em cursos básicos de primeiro ciclo, organizados em quatro seções: Trabalhos de Metal, Indústria Mecânica, Eletrotécnica e Artes Industriais. Ademais, a Escola também estava autorizada a oferecer cursos de mestria para os professores atuantes nessas áreas.

Transformada em autarquia pela Lei Federal 3.552, de 16 de fevereiro de 1959, todas as Escolas Industriais do Brasil conseguiram autonomia administrativa, didática e financeira, transformando-se em instituições federais destinadas a ministrar cursos técnicos de nível médio. Porém, somente em 1963, EIN implantou seus primeiros cursos técnicos de nível médio, com as ofertas de Mineração e de Estradas. O novo modelo tinha equivalência ao ensino de 2º grau, o que permitia a continuidade de estudos no ensino superior para os egressos que assim o desejassem.

Em 1965, o Estabelecimento passou a nomear-se Escola Industrial Federal do Rio Grande do Norte (EIFRN). Nessa década, no dia 11 de março de 1967, ocorreu a inauguração da “nova” Escola Industrial nas recém-construídas instalações do prédio situado na Avenida Salgado Filho nº 1559, no bairro Tirol, atendendo a uma comunidade escolar de 233 servidores e cerca de 1.100 estudantes.

Na condição de Escola Técnica Federal do Rio Grande do Norte (ETFRN), mudança impetrada pela Portaria Ministerial 331, de 16 de junho de 1968, o Conselho de Representantes deliberou a extinção gradativa dos cursos industriais básicos, passando-se a ministrar somente o ensino profissional de nível técnico. Em consequência, foram criados, em 1969 e 1973, os cursos técnicos de nível médio em Eletrotécnica, em Mecânica, em Edificações, em Saneamento e em Geologia, sob a orientação da Lei 5.692/71, a qual definia a estrutura do ensino de 2º grau como ensino profissionalizante obrigatório. A partir de então, a ETFRN passou a dedicar-se, exclusivamente, ao ensino técnico profissionalizante de 2º grau.

No ano de 1994, iniciou-se outro processo de transição das escolas técnicas federais. No entanto, somente no dia 18 de janeiro de 1999, efetivou-se a mudança de ETFRN para Centro Federal de Educação Tecnológica do Rio Grande do Norte (CEFETRN). De acordo com os documentos oficiais, os centros de educação tecnológica foram implantados com a finalidade de formar e qualificar profissionais no âmbito da educação tecnológica, em diferentes níveis e modalidades de ensino, para os diversos setores da economia. Também objetivavam realizar pesquisas aplicadas e promover o desenvolvimento tecnológico de novos processos, produtos e serviços, em estreita articulação com os setores produtivos e com a sociedade.

Em 2006, o Governo Federal lançou um arrojado plano de expansão da rede federal. Com características de interiorização de educação profissional e tecnológica para todo o País, e foram implantadas mais três unidades de ensino vinculadas ao CEFET-RN: na Zona Norte da cidade de Natal, na cidade de Ipanguaçu e Currais Novos. Ainda nesse mesmo ano, devido ao lançamento do Programa Nacional de Educação de Jovens e Adultos (PROEJA), o CEFET-RN começou a atuar na educação profissional técnica de nível médio, na modalidade de educação de jovens e adultos, oferecendo também, para educadores que atuam nessa modalidade, cursos de formação em nível de pós-graduação lato sensu.

Ao limiar de um século de existência, a Instituição adquiriu nova configuração, transformando-se em Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Norte (IFRN), nos termos da Lei 11.892, de 29 de dezembro de 2008. De acordo com o estabelecido oficialmente, os institutos federais são instituições, pluricurriculares e multicampis, de educação superior, básica e profissional. São especializadas na oferta de educação profissional e tecnológica nas diferentes modalidades de ensino, ancorando-se na conjuntura de conhecimentos técnicos e tecnológicos com as práticas pedagógicas. Esse processo, uma vez criada a rede federal de educação profissional, cientifica e tecnológica, constituiu-se em elemento de redefinição do sistema de ensino brasileiro. Reforçou, por outro lado, a autonomia administrativa, patrimonial, financeira, didático-pedagógica e disciplinar dessas instituições educativas, além de imprimir a equivalência às universidades federais, no que refere às disposições que regem a regulação, a avaliação e a supervisão das instituições e dos cursos de educação superior.

A expansão da rede federal de educação profissional e tecnológica está pautada na interiorização da educação profissional, com o compromisso de contribuir, significativamente, para o desenvolvimento socioeconômico do País. Nessa perspectiva, a criação dos institutos federais responde à necessidade da institucionalização definitiva da educação profissional e tecnológica como política pública permanente de Estado.

Esse processo de interiorização da educação profissional e tecnológica contribui para o combate às desigualdades estruturais de diversas ordens, proporcionando o desenvolvimento social por meio da formação humana integral dos sujeitos atendidos. Propicia, ainda, o desenvolvimento econômico, a partir da articulação das ofertas educacionais e das ações de pesquisas e de extensão. Tal articulação vincula-se aos arranjos produtivos sociais e culturais, com possibilidades de permanência e de emancipação dos cidadãos assim como de desenvolvimento das diversas regiões do Rio Grande do Norte.

* 1. **Definição do Problema**

O Instituto Federal de Educação, Ciência e tecnologia do Rio Grande do Norte, ministra diversos cursos em diferentes níveis de ensino: médio, técnico, técnico-subsequente, tecnológico, PROEJA e cursos de pós-graduação, latu e estrito sensu. Com um corpo docente amplo e uma grande massa de alunos, como já foi falado, nos parágrafos anteriores. A movimentação diária, e constantes, do sistema acadêmico, gera uma grande massa de dados. Esses dados acadêmicos estão em bases de dados, agregando informações sobre a vida escolar do aluno, identifica quem cada estudante é (**dados demográficos** como idade, sexo, situação financeira dos pais, nível de escolaridade dos pais, região de procedência, dentre outros), e **dados comportamentais**, como o número de reprovações, a situação no curso, informações de evasão e os **dados sobre avaliações**.

Nos últimos anos, têm-se observado, por parte dos educadores e gestores educacionais, um alto índice de reprovações e evasão, em alguns cursos e até mesmo, nota-se um certo desestímulos por parte dos alunos, em determinados cursos, ministrados pelos IFS. É fato, por exemplo, que o curso de Tecnologia de Desenvolvimento de Software (**TADS**), nos últimos dois anos, as turmas de 3º e 4º períodos, chegam com média muita baixa de alunos, em torno de 20% e, isso tem como consequência, vários professores, ministrando aulas para poucos alunos, diminuindo em muito, a relação aluno professor, um dos índices, que é utilizado pelo governo federal, para avaliar o Instituto. Diante desse quadro, surgiu a necessidade de fazer uma análise mais aprofundada de tais problemas. Ou seja, deseja-se analisar as seguintes necessidades: o **índice de reprovação**, o **índice de evasão**, o **índice de aprovação**, o **índice de conclusão** nos referidos cursos, como posso melhorar o índice “relação aluno versus professor?”, tendo como finalidade, descobrir quais são as causas que implicam, diretamente ou indiretamente nesses índices. E identificadas essas causas, apresentar, através de um sistema, as soluções para os problemas.

Para tanto, será proposto o desenvolvimento de dois sistemas: um sistema desktop, que será utilizado pelos gestores do instituto, que compreenderá o **Data Warehouse** e o **Data Mining**, e um sistema Web, que contemplará, *Case based Reasonry*, o game e, os relatórios que serão disponibilizados pelo sistema. Portanto, o projeto será desenvolvido em dois módulos: um sistema de software desktop (administrativo) e, um outro módulo Web. No primeiro módulo será implementado um *Data Warehouse*, que terá como finalidade, disponibilizar para os gestores, relatórios diversos (Ad hoc), que darão suporte, com certeza, a tomada de decisões educacionais, mais eficientes e, terá também, a finalidade de tornar o processo de Mineração de Dados, mais rápido, uma vez que, os dados já estarão de uma certa forma, preparados para os algoritmos de Mineração de Dados. Após ter sido implementado o *Data Warehouse*, serão aplicadas técnicas de Mineração Dados (***Data Mining***), para tentar encontrar padrões de comportamentos úteis nos dados educacionais, que esses padrões, após serem analisados pelos especialistas do domínio, possam ser utilizados, para identificar as principais causas de evasões escolares, bem como reprovações, em nossos campis, e que, isso seja utilizado em prol do ensino aprendizagem de nossos alunos.

* 1. **Módulo Data Warehouse**

Neste módulo, será desenvolvido todo o *Data Warehouse* do projeto de pesquisa, ou seja, neste módulo serão desenvolvidos todos os modelos dimensional e multidimensional do proposto para o IFRN.

**7.4.1 O Ciclo de Vida**

Inicialmente, será definido o ciclo de vida de desenvolvimento do sistema, em inglês system development life cycle (SDLC) que, alguns autores chamam simplesmente de metodologia de desenvolvimento de sistema, em inglês system development methodology. Há duas variantes principais: o modelo em cascata e o iterativo. O modelo em cascata é também conhecido como a metodologia sequencial. A metodologia iterativa é conhecida como espiral ou incremental. Para esse projeto será utilizado a metodologia em cascata, seguindo o diagrama da Figura 7.1.

Estudo de viabilidade

Requisitos

Arquitetura

Modelagem

Desenvolvimento

Teste

Implantação

Operação

**Figura 7.1 Metologia em Cascata Fonte: adaptado de Rainardi, 2008**

**7.4.2. Estudo de viabilidade**

O instituto tem um sistema de controle acadêmico que é responsável em gerenciar toda a vida acadêmica do mesmo. Neste sistema, temos dados acadêmicos dos alunos, cadastro dos cursos e suas matrizes curriculares, cadastro dos docentes e suas capacitações e áreas de atuação, inclusive pesquisas e extensão. Temos os dados socioeconômicos dos alunos, incluindo ai, situação financeira dos pais. Resumindo, todas essas informações estão em uma base de dados, em um servidor, localizado no setor administrativo do instituto, mais precisamente no prédio da reitoria, em Natal, capital do estado. Além desses dados, temos os dados de formulários de pesquisas que, são aplicadas semestralmente, no instituto, para avaliação do desempenho escolar dos alunos e turmas, avaliação do corpo docente e avaliação da instituição, no que se diz respeito a infraestrutura, bibliotecas, e outros assuntos administrativos relacionados a educação. Todos esses dados de pesquisas, estão armazenadas em planilhas eletrônicas nos seus respectivos setores. Temos também, dados do serviço social da escola, que constam da assistência social que o instituto presta aos alunos, como por exemplos, bolsas de estudo, alimentação, assistência medica, dentária dentre outras. Estes dados, também estão em planilhas eletrônicas. Temos ainda, os dados de acesso a biblioteca, onde são registradas todas as consultas que os alunos fizeram, em forma de requisições de livros. Estes dados, estão armazenados em bases de dados e, não em planilhas. A tabela 1 a seguir, mostra um resumo dessas fontes de dados.

**Tabela 1** Bases de Dados do Instituto Natal Central - RN

|  |  |  |
| --- | --- | --- |
| Fonte de dados | Modelo | Descrição |
| Sistema acadêmico | Base de dados SQL Server | Vida acadêmica dos alunos, professores, cursos e matrizes curriculares. |
| Pedagogia | Planilhas eletrônicas | Dados de pesquisas aplicadas semestralmente, para avaliação do desempenho escolar dos alunos. Para avaliação do desempenho dos professores e pesquisas avaliativas do instituto, como por exemplo, infraestrutura. |
| Serviço social | Planilhas eletrônicas | Dados referentes a assistência social do instituto para com os alunos carentes. |
| Biblioteca | Base de dado MySql | Informações de acesso a biblioteca pelos alunos (requisições de livros). |

Diante desse cenário, pode-se perceber facilmente que, essas diversas bases de dados não estão interligadas, ou seja, elas não se comunicam entre si, dificultando se fazer uma análise estatística com esses dados. Portanto, a proposta apresentada nesse projeto é de consolidar essas diversas fontes de dados, afim de propiciar uma maneira mais palpável, de se fazer cruzamentos com esses dados, afim de tentar encontrar relacionamentos entre os mesmos, que possam ser úteis para os gestores educacionais, em tomarem decisões mais precisas, em prol do ensino aprendizagem e, como isso, desenvolver ações que possam engajar ou incentivar os alunos em nossos cursos, e dessa forma, minimizarmos os problemas da evasão e reprovação nos institutos. Portanto, a justificativa para implementar um sistema de *Data Warehouse*, é propiciar aos gestores educacionais, uma forma deles poderem analisa essas informações de diversas maneira, através de relatórios extraídos do modelo multidimensional e dos cubos **OLAP** gerados.

**7.4.3. Requisitos**

Para assegurar que os objetivos dos usuários serão alcançados, é de fundamental importância que sejam definidos os requisitos funcionais e não funcionais do sistema *Data Warehouse*, antes de começar seu desenvolvimento. Os requisitos funcionais, definem o que o sistema irá fazer, ou seja, os recursos que o sistema irá ter. Os requisitos não funcionais definem a arquitetura que será usada na construção do sistema.

**7.4.3.1. Requisitos Funcionais**

Nesta fase, foram levantadas, de forma sucinta, todas as características que cercam a educação no instituto federal. Essas informações foram coletadas de forma a elucidar os principais índices de desempenho que se queira analisar, de forma a facilitar a escolha do processo educativo, que se queira modelar. Um processo educativo, é por exemplo, analisar a evasão escolar. No geral, foram realizadas as seguintes atividades:

* Identificar os fatores “humanos” relacionados ao ambiente educacional, tais como o interesse e expectativas dos gestores educacionais em relação ao desenvolvimento e implantação do projeto em questão;
* Reuniões para se identificar os requisitos necessários para a definição do processo a modelar;
* Levantamento de dados das diversas fontes de dados, através do estudo de documentos, planilhas e relatórios e da própria base de dados do acadêmico;
* Levantamento de informações analíticas, ou seja, dos índices indicadores de desempenho escolar.

Em consequência dos levantamentos realizados nessa fase, se pode constatar os seguintes fatores, em relacionados ao projeto:

* Com a divulgação do projeto junto aos usuários da gestão educacional, criou-se uma grande expectativa dos potenciais usuários, especialmente aos pedagogos, que puderam enxergar os benefícios que o sistema trará para os mesmos.
* Os dados coletados para testes dos modelos, serão os dados do sistema controle acadêmico, dados do sistema social e dados do sistema da biblioteca do campus Natal central.
* O levantamento junto aos gestores educacionais, revelaram os seguintes índices a serem analisados: **reprovação escolar**, **evasão escolar**, **aprovação escolar**, **conclusão de cursos**, relação **aluno x professor**, **alunos desligados** e **alunos diplomados**.

Após a análise sucinta das fontes de dados, pode-se identificar os seguintes requisitos funcionais, mostrados na Tabela 7.1.

**Tabela 7.1** Requisitos Funcionais do Sistema Data Warehouse do IFRN

|  |  |  |
| --- | --- | --- |
| Nº | Requisito | Prioridade |
| 1 | Os gestores educacionais, necessitam analisar o **índice de evasão escolar** nos campis do IFRN. Isso poderá ser feito, analisando o seu desempenho escolar, suas faltas, sua participação em projetos, seu índice de satisfação com o curso, uma vez que, muitas vezes o aluno está fazendo um curso que não é o que ele gostaria de fazer e, sim seus pais. | Alta |
| 2 | Os gestores educacionais, necessitam analisar o **índice de reprovação escolar** nos campis do IFRN. Isso poderá ser feito, analisando o seu desempenho escolar, suas faltas, acessos a biblioteca, relacionamentos com professores e colegas. Situação financeira dos pais, dentre outras coisas. Essa análise deve ser feita por curso, turmas, disciplinas, períodos. | Alta |
| 3 | Precisa-se também, analisar o índice de aprovados, parasse ter, por exemplo, uma melhor projeção dos concluintes por turma e curso. Isso, implicará na formação de novas turmas e para poder definir com mais precisão as entradas nos cursos. | Alta |
| 4 | Precisa-se também, analisar o índice da relação aluno x professor, para se poder ter uma maior precisão na definição de quantos professores serão necessários para se manter um curso de qualidade, em uma visão ampla de vários anos a frente. Dessa forma, pode-se predizer com antecedência, quantos professores serão necessários contratar, para cada curso e disciplinas relacionadas. Isso é necessário, pois muitas vezes o problema é identificado com o semestre em curso e, como a contratação de professores pelo instituto, leva meses para ser concretizada. | Média |
| 5 | Analisar também a relação em alunos que concluíram o curso e alunos diplomados. Uma vez que, temos constantemente, situações onde, o aluno conclui todas as disciplinas e não se diploma, por vários motivos: doença, desiste sem explicação, não faz o trabalho de conclusão do curso, começa a trabalhar, etc. Portanto, por esses motivos, o índice de alunos concluídos e alunos diplomados devem ser analisado, para se tentar identificar os verdadeiros motivos e, tentar encontrar uma solução para os mesmos. | Alta |
|  |  |  |

**7.4.3.2 Requisitos Não Funcionais**

Os requisitos não funcionais fornecem um guia e restrições para a arquitetura do sistema. Algumas dessas restrições, é devido os padrões de TI, outras restrições são impostas pela equipe de arquitetura de TI e, algumas outras são impostas pelos administradores de banco de dados (**DBA**s) e administradores do sistema operacional. Após reuniões com os arquitetos de TI, administradores do sistema operacional e usuários finais, foram definidos os seguintes requisitos não funcionais, como mostra a Tabela 72.

**Tabela 7.2 Requisitos Não Funcionais do Data Warehouse do IFRN**

|  |  |
| --- | --- |
| **Nº** | **REQUISITO** |
| 1 | Todos os usuários do data warehouse poderão acessar a aplicação front-end do data warehouse (relatórios e cubos OLAP), logando-se apenas uma vez no sistema. |
| 2 | A aplicação front-end do data warehouse não deve ser acessível fora do IFRN. |
| 3 | A aplicação front-end do data warehouse deve ser essencialmente web. |
| 4 | O acesso a aplicação front-end do data warehouse deve ser através do portal do IFRN. |
| 5 | Alguns usuários terão permissão de executarem consultas SQL diretamente no sistema. |
| 6 | Alguns dados sensíveis serão vistos somente por alguns usuários. |
| 7 | O tempo máximo de resposta para um relatório ou consulta é de 30 segundos. |
| 8 | O data warehouse deve estar disponível 24 horas ao dia, 7 dias da semana. |
| 9 | Para construir o projeto do data warehouse, será usado o Microsoft SQL Server, incluindo a ferramenta de ETL, reporting e OLAP. |
| 10 | Deve-se ser feito o backup do data warehouse e testado o restore pelo menos a cada 6 meses. |
| 11 | Estima-se um número de usuários entre 300 e 1000. Cerca de 10% desses usuários, acessarão o sistema mais frequentemente; o resto ocasionalmente. |

O principal propósito de fazer o estudo de viabilidade é poder identificar os possíveis risco que possam fazer com que o projeto falhe. Estes riscos geralmente estão associados com a indisponibilidade ou a acessibilidade dos dados. Como por exemplo, precisa-se analisar o índice de concluintes e diplomados, mais os dados, do ano ou semestre, ainda não estão disponíveis no sistema, pois as notas ainda não foram informadas, por motivos externos ao sistema, impossibilitando, dessa forma, a execução do processo de ETL.

A Tabela 7.3, lista alguns dos riscos de dados para o projeto data warehouse do IFRN.

|  |  |
| --- | --- |
| **Nº** | **RISCO** |
| 1 |  |

**7.4.4. Arquitetura**

Um sistema *Data Warehouse* tem duas arquiteturas principais: a arquitetura de fluxo de dados e a arquitetura do sistema. A arquitetura de fluxo de dados define como os dados do Data Warehouse são armazenados e como os dados fluem do sistema fonte para o *Data Warehouse*. A arquitetura de sistema define a configuração física de servidores, rede, software, *storage*, e clientes. Portanto, será definido a partir desse ponto as duas arquiteturas do projeto.

**7.4.5. Arquitetura de Fluxo de Dados**

Em um *Data Warehousing*, a arquitetura de fluxo de dados é a configuração dos armazéns de dados dentro do sistema Data Warehouse, bem como os dados fluem dos sistemas fonte para os armazéns de dados e desses para os usuários finais do sistema. Isto inclui, como esses dados são controlados, monitorado, bem como os mecanismos para assegurar a qualidade desses dados. No referencial teórico desse trabalho foi apresentado e discutido, em mais detalhes, quatro arquiteturas para fluxo de dados: ***single DDS***, **NDS** + **DDS**, **ODS** + **DDS**, e ***Data Warehouse*** federado.

Para este projeto escolhi trabalhar com a arquitetura NDS + DDS, por facilitar a implementação, uma vez que, tenho disponível para total acesso, a principal fonte de dados, que é o banco de dados do sistema Acadêmico, uma base de dados SQL Server, que já é um sistema normalizado.

Em uma arquitetura NDS + DDS, há os seguintes componentes: a fonte de dados, o armazém de dados normalizado (NDS), o armazém de dados dimensional (DDS)

**7.4.6. Modelagem do Data Warehouse**

Nesta seção, será desenvolvido o projeto de *Data Warehouse* em si, ou seja, a modelagem das bases de dados dimensional e multidimensional do sistema, onde serão definidos os fatos e suas dimensões correspondentes.

Come teremos vários fatos, para serem analisados, vamos trabalhar através de estudos de casos.

**7.4.6.1**. **Estudo de Caso 1 – Analisar a Evasão/Cancelamento e Conclusão**

O primeiro estudo de caso a ser trabalhado é para analisar a situação dos alunos em relação a evasão, o cancelamento, a jubilação e o total de concluídos por campus, curso, ano letivo e período letivo. Para isto, foram criadas as seguintes dimensões e tabela de fatos: dimCursos, dimInstituicao, dimTempo, dimEvaJubPorCampusCurso, dimCanceladoPorCampusCurso, dimConcluidosPorCampusCurso e a tabela de fatos FatoEvaCanCon, cuja estrutura, passarei a detalhar a seguir:

**Tabela 7.4 Dimensão Cursos**

|  |  |  |
| --- | --- | --- |
| **Coluna** | **Tipo de dados** | **Descrição** |
| COD\_CURSO | Int | Código do curso |
| COD\_TURNO | Int | Código do turno |
| COD\_INSTITUICAO | Int | Código da instituição |
| DESC\_CURSO | Varchar (255) | Descrição do curso |
| SIGLA\_CURSO | Varchar (15) | Sigla do curso |
| DATA\_INICIO | Smalldatetime | Data que o curso inicou |

**Tabela 7.5 Dimensão Instituição**

|  |  |  |
| --- | --- | --- |
| **Coluna** | **Tipo de dado** | **Descrição** |
| COD\_INSTITUICAO | Int | Código da instituição |
| DESC\_INSTITUICAO | Varchar(50) | Descrição por extenso da instituição |
| SIGLA\_INSTITUICAO | Varchar(20) | Sigla da instituição |
| ENDERECO | Varchar(100) | Endereço da instiuição |
| CIDADE | Varchar(100) | Cidade onde está localizada a instituição |
| BAIRRO | Varchar(100) | Bairro da cidade onde está localizada a instituição. |
| CEP | Varchar(10) | CEP do endereço onde está localizada a instituição. |

**Tabela 7.6 Dimensão Tempo**

|  |  |  |
| --- | --- | --- |
| **Coluna** | **Tipo de dado** | **Descrição** |
| Ano | Smallint | Ano letivo da ocorrência do fato. |
| Período | Tinyint | Período do ano letivo da ocorrência do fato. |

**Tabela 7.7 Dimensão Evadidos por campus e curso**

|  |  |  |
| --- | --- | --- |
| **Coluna** | **Tipo de dado** | **Descrição** |
| Campus | Varchar(50) | Sigla do campus onde está sendo totalizado os evadidos/jubilados. |
| Curso | Varcha(15) | Sigla do curso onde está sendo totalizado os evadidos/jubilados. |
| Ano Letivo | Smallint | Ano letivo onde está sendo totalizado os evadidos/jubilados. |
| Período letivo | Tinyint | Período letivo onde está sendo totalizado os evadidos/jubilados. |
| TotalEvadidos | Int | Total de alunos evadidos por campus, curso, ano letivo e período letivo. |

**Tabela 7.8 Dimensão Cancelados por campus e curso**

|  |  |  |
| --- | --- | --- |
| **Coluna** | **Tipo de dado** | **Descrição** |
| Campus | Varchar(50) | Sigla do campus onde está sendo totalizado os que cancelaram o curso. |
| Curso | Varcha(15) | Sigla do curso onde está sendo totalizado os alunos que cancelaram o curso. |
| Ano Letivo | Smallint | Ano letivo onde está sendo totalizado os que cancelaram o curso. |
| Período letivo | Tinyint | Período letivo onde está sendo totalizado os que cancelaram o curso. |
| TotalCancelados | Int | Total de alunos que cancelaram o curso, agrupados por campus, curso, ano letivo e período letivo. |

**Tabela 7.9 Dimensão Concluídos por campus e curso**

|  |  |  |
| --- | --- | --- |
| **Coluna** | **Tipo de dado** | **Descrição** |
| Campus | Varchar(50) | Sigla do campus onde está sendo totalizado os que concluiram o curso. |
| Curso | Varcha(15) | Sigla do curso onde está sendo totalizado os concluíram. |
| Ano Letivo | Smallint | Ano letivo onde está sendo totalizado os que concluíram o curso. |
| Período letivo | Tinyint | Período letivo onde está sendo totalizado os que concluíram o curso. |
| TotalConcluidos | Int | Total de alunos que concluíram o curso, agrupados por campus, curso, ano letivo e período letivo. |

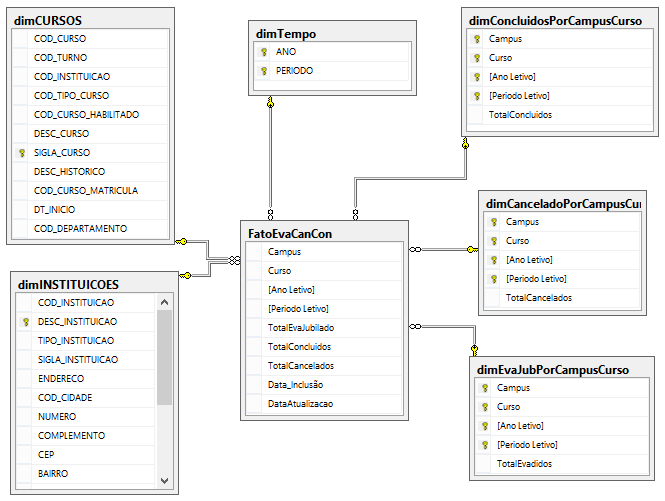
**Tabela 7.10. Tabela de medidas – Fato Evadidos Cancelados, Concluídos**

|  |  |  |
| --- | --- | --- |
| **Coluna** | **Tipo de dado** | **Descrição** |
| Campus | Varchar(50) | Campus onde ocorreu o fato. |
| Curso | Varchar(15) | Curso onde ocorreu o fato. |
| Ano letivo | Smallint | Ano letivo da ocorrência do fato. |
| Período letivo | Tinyint | Período letivo da ocorrência do fato. |
| TotalEvadidos | Int | Total de alunos evadidos por campus, curso, ano letivo e período letivo. |
| TotalCancelados | Int | Total de alunos que cancelaram matricula por campus, curso, ano letivo e período letivo. |
| TotalConcluidos | Int | Total de alunos que concluíram o curso por campus, curso, ano letivo e período letivo. |
| Data\_Inclusao | Date | Data que ocorreu a inserção dos dados na tabela de fatos. |
| Data\_Atualizacao | Date | Data que ocorreu a atualização da tabela de fatos. |

Esta coleção de tabelas de fato e dimensões é chamado de *data mart*. Um *data mart* é um grupo de tabelas de fatos e suas tabelas dimensões correspondentes, contendo as medidas e os eventos de negócio, categorizado por suas dimensões. A figura 7.2 mostra o modelo para este *data mart,* onde são combinados a tabela de fatos com suas tabelas de dimensões correspondentes.

Com este data mart, irei montar um cubo OLAP, para gerar relatórios administrativos, onde se possa analisar o total de alunos evadidos por campus, curso, ano letivo e período letivo, bem como, a quantidade de matrículas canceladas ou concluídas por campus, curso, ano letivo e períodos letivos. Este cubo também, servirá como entrada para estruturas de mineração de dados.

**Figura 7.2 Data mart para as medidas de Evasão, cancelado e concluídos**.

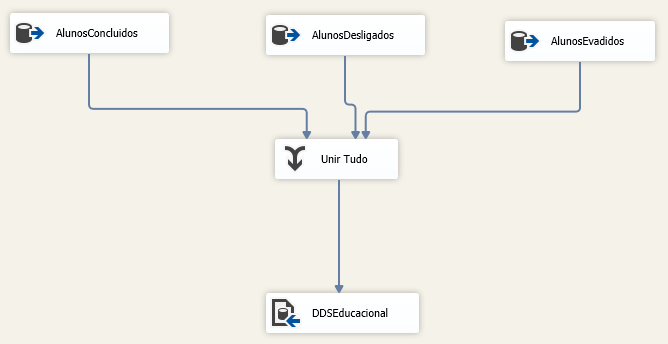


**Fonte**: autor

Antes de extrair os dados do sistema normalizado (fonte de dados - Sistema Acadêmico), criou-se as dimensões que conterão os totais dos alunos evadidos, cancelados e concluídos.

Após a criação das dimensões e da tabela de fatos, foi feita a extração e população das dimensões e da tabela de fato.

Após a análise dos dados do sistema Acadêmico, o sistema normalizado, percebeu-se que era necessário, juntar dados de diversas tabelas para podermos ter, uma maneira melhor de fazermos as agregações para totalizar as evasões, cancelamentos e conclusões. A figura 7.3.mostra o processo de extração dos dados do sistema normalizado, a união desses dados e a carga desses dados no sistema dimensional.



**Figura 7.3** Este pacote extrai dados do sistema acadêmico para o sistema dimensional.

**Fonte**: autor

O pacote da figura 7.3, tem a seguinte dinâmica: através de conexões SQL Server Nativa, são feitas a extração dos dados do sistema normalizado (Sistema Acadêmico), em seguida, faz-se a união desses dados, e finalmente, a carga no sistema dimensional.

O código **SQL** a seguir, faz a população da dimensão dimEvaJubPorCampusCurso:

INSERT INTO (Campus,

Curso,

[Ano Letivo],

[Periodo Letivo],

TotalEvadidos)

SELECT desc\_instituicao AS [Campus],

sigla\_curso AS [Curso],

ano\_let AS [Ano Letivo],

periodo\_Let AS [Periodo Letivo],

sum(QTD\_ALUNO) AS TotalEvadidos

FROM vJuntandoEvaDeslCon

WHERE Situacao\_Da\_Matricula = 'Evasão' OR

Situacao\_Da\_Matricula ='Jubilado'

GROUP BY DESC\_INSTITUICAO, Sigla\_Curso,ANO\_LET, PERIODO\_LET

O código **SQL** a seguir, faz a população da dimensão dimEvaJubPorCampusCurso:

INSERT INTO (Campus,

Curso,

[Ano Letivo],

[Periodo Letivo],

TotalCancelados)

SELECT desc\_instituicao AS [Campus],

sigla\_curso AS [Curso],

ano\_let AS [Ano Letivo],

periodo\_Let AS [Periodo Letivo],

sum(QTD\_ALUNO) AS TotalCancelados

FROM vJuntandoEvaDeslCon

WHERE Situacao\_Da\_Matricula = 'Cancelado'

GROUP BY DESC\_INSTITUICAO, Sigla\_Curso,ANO\_LET, PERIODO\_LET

O código **SQL** a seguir, faz a população da dimensão dimConcluidosPorCampusCurso:

INSERT INTO (Campus,

Curso,

[Ano Letivo],

[Periodo Letivo],

TotalConcluidos)

SELECT desc\_instituicao AS [Campus],

sigla\_curso AS [Curso],

ano\_let AS [Ano Letivo],

periodo\_Let AS [Periodo Letivo],

sum(QTD\_ALUNO) AS TotalConcluido

FROM vJuntandoEvaDeslCon

WHERE Situacao\_Da\_Matricula = 'Concluido' OR

Situacao\_Da\_Matricula = 'Formadodo'

GROUP BY DESC\_INSTITUICAO, Sigla\_Curso,ANO\_LET, PERIODO\_LET

A seguir temos a instrução SQL para popular a tabela de fatos:

**FatoEvaCanCon**:

INSERT INTO FatoEvaCanCon (Campus,Curso,

[Ano Letivo],

[Periodo Letivo],

TotalEvaJubilado,

TotalConcluidos,

TotalCancelados,

Data\_Inclusão,

DataAtualizacao)

SELECT coalesce(EJ.Campus,null),

coalesce(EJ.Curso,null),

coalesce(EJ.[Ano Letivo],null),

coalesce(EJ.[Periodo Letivo],null),

EJ.TotalEvadidos,

CO.TotalConcluidos,

CA.TotalCancelados,

getdate(),

getdate()

FROM dimEvaJubPorCampusCurso EJ inner Join dimConcluidosPorCampusCurso CO

ON EJ.campus = CO.Campus and EJ.Curso = CO.Curso AND

EJ.[Ano Letivo] = CO.[Ano Letivo] AND EJ.[Periodo Letivo] = CO.[Periodo letivo inner Join dimCanceladoPorCampusCurso CA

ON CO.campus = CO.Campus and EJ.Curso = CA.Curso AND

CO.[Ano Letivo] = CA.[Ano Letivo] AND

CO.[Periodo Letivo] = CA.[Periodo letivo] inner Join dimTempo T

ON CA.[Ano Letivo] = T.[ANO] AND

CA.[Periodo Letivo] = T.[PERIODO]

**7.4.7. Banco de Dados Multidimensional**

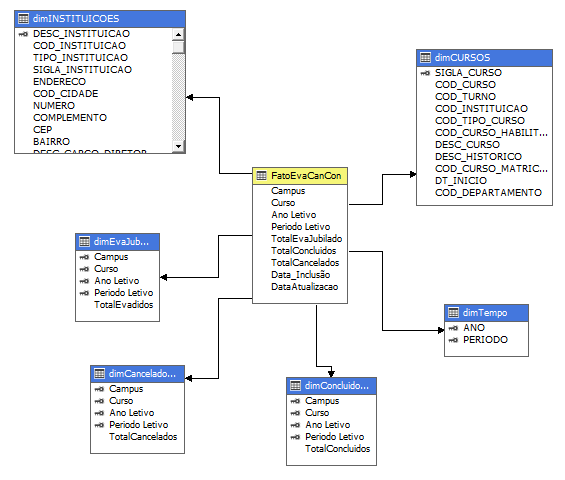
Nesta seção serão desenvolvidos os bancos de dados multidimensionais do projeto. Como já definido no referencial teórico. Um banco de dados multidimensional é uma forma de banco de dados onde os dados são armazenados em células e a localização de cada célula é definida por hierarquias chamadas de dimensões. Cada célula representa um evento de negócio, e os valores das dimensões mostram quando e onde este evento ocorreu.

A estrutura multidimensional, armazena as agregações geradas, bem como os dados da base de dados, em formato de matriz, em vez de tabelas relacionais. Os valores das agregações são pré-calculadas, sumarizadas, em função dos dados da base de dados.

Na seção anterior, criou-se um modelo dimensional, onde foram definidas as dimensões e a tabela de fatos, que relacionam as dimensões cursos, instituições, alunos evadidos, concluídos e cancelados por campus e curso e, a também a tabela de fatos que armazena os totais de alunos evadidos, concluídos e cancelados por campus e curso.

Como primeiro exemplo do desenvolvimento do banco multidimensional do projeto, tem-se a criação do banco multidimensional para o modelo dimensional definido na seção anterior, que servirá de base para os relatórios administrativos sobre os índices de evasão, cancelamento e concluídos por campus e curso. Este modelo multidimensional será utilizado também pelas estruturas de mineração de dados para análises de padrões de comportamentos, que possam ser úteis para o IFRN. A figura 7.4 mostra o cubo, gerado pelo Microsoft Analysis Services.

**Figura 7.4. Estrutura do Cubo para os fatos: Evadidos, Cancelados e Concluídos por Campos e Curso**.



**Fonte**: autor

A dimensão **dimTempo** define a hierarquia ano e período letivo onde ocorrem os eventos. Para este cubo, esta hierarquia possibilita, a análise dos dados relacionados, por ano e período letivos.

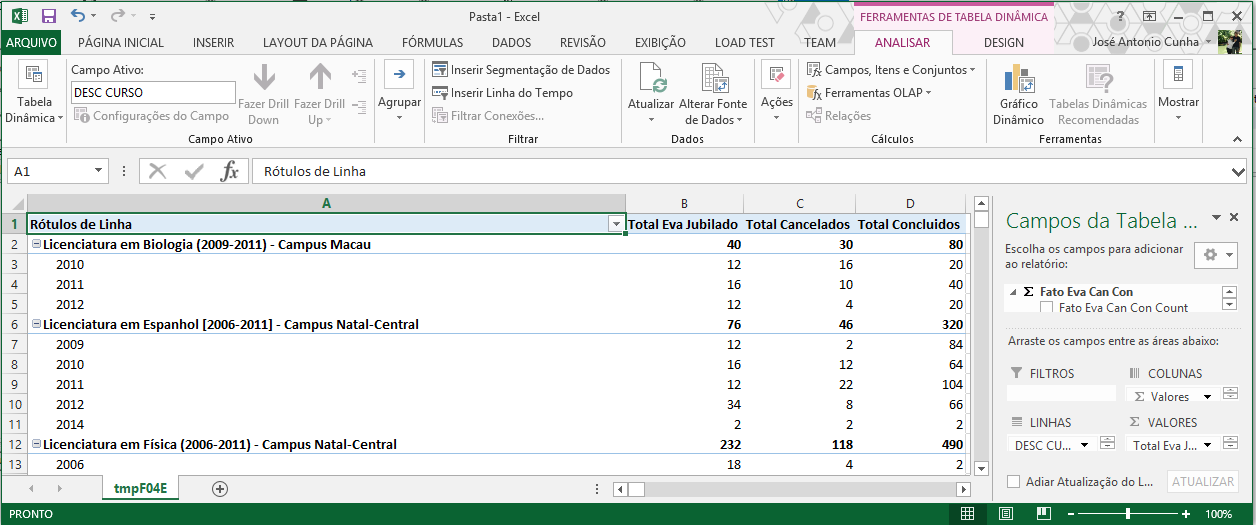
As outras dimensões servem para detalhar os fatos, por exemplo, a dimensão curso possibilita que seja identificado o nome descritivo do curso, bem como, a que departamento ele pertence.

**7.4.8. Alguns resultados para este estudo de caso em particular**

Com o cubo criado e processado, podemos manipulá-lo para extrair informações. Como por exemplo, podemos utilizar a planilha eletrônica ou um gerador de relatórios para manipular os dados do cubo.

A figura 7.5 mostra um exemplo do que se pode obter com a planilha Excel.

**Figura 7.5. Relatório exibindo o total alunos de evadidos, que cancelaram a matrícula e que concluíram o curso, relacionados por campus e ano letivo.**



**Fonte:** autor

Deve-se observar que este é um exemplo de um relatório puramente administrativo. Devo lembrar que, nessa fase, ainda não foi usado a mineração de dados. No entanto, pode-se perceber que, este tipo de relatório pode ser fundamental para os administradores do IFRN, terem ideia do que está acontecendo em seus respectivos campis e cursos. Devo lembrar que, este exemplo é apenas um tipo de informação que pode ser extraída desse cubo em particular.

Informações semelhantes, podem ser obtidas através de um gerador de relatório, como mostra a figura 7.6.

**7.3.6. Implantação**

Identificada as diversas fontes de dados, o próximo passo, foi realizar uma análise detalhada dos dados contidos nessas bases, para que fossem selecionados, aqueles dados, que realmente fossem de interesse, nos propósitos do projeto em questão. Feita esta análise, ficou comprovado que, a melhor forma de utilizar o *Data staging Area*, ou seja, a área de armazenamento para esses dados, seria na forma de arquivos de planilhas eletrônicas e arquivos simples (*flat file*). Isto foi constato, pelo fato de que, os dados do sistema acadêmico são facilmente extraídos, dessa base de dados, no formato de planilhas do Excel. O mesmo ocorre com as outras bases de dados, visto que a maioria delas já estão no formato de planilha Excel. Resumindo, esta etapa de extração de dados, envolve a leitura e a compreensão dos dados de origem e cópia dos dados necessários ao *Data Warehouse* na *stage* área para que sejam manipulados posteriormente.

7. **Etapa de anteprojeto**

**7.1.** **Levantamento das Necessidades do Sistema**

* 1. **Resultado do levantamento das necessidades**



* 1. **Arquitetura do Data Warehouse**

Um *Data Warehouse* é composto de duas arquiteturas principais: a arquitetura do fluxo de dados e a arquitetura de sistema. A arquitetura do fluxo de dados, define como os dados são organizados e armazenados no *data warehouse*, e como esses dados fluem para os usuários. A arquitetura física determina as configurações de servidores, redes, softwares, storages e clientes Rainardi, 2008).

**7.3.1** **Resultados obtidos no Planejamento**

* 1. Etapa de Definição
  2. Etapa de Execução

1. Conclusão