**Novo Paradigma em Automação Residencial**

**Projeto da Disciplina de BI**

Prof. Anderson Nascimento

prof.anderson@ica.ele.puc-rio.br

**Componentes do Projeto:**

Otávio Ciribelli Borges – otavio.ciribelli@gmail.com

|  |
| --- |
| **Histórico de Versões** |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Data** | **Versão** | **Descrição** | **Autor** | **Aprovado por** |
| 18/02/2023 | 1.0 | Proposição inicial | Otávio | Prof. Anderson |
| 02/03/2023 | 1.1 | Versão final | Otávio | Prof. Anderson |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

**Sumário**

Sumário

[1 Introdução 5](#_Toc128682125)

[2 Estudo de Caso 6](#_Toc128682126)

[3 Descrição do Modelo Transacional 8](#_Toc128682127)

[4 Proposta de Processo de BI 10](#_Toc128682128)

[5 Modelo Multidimensional 12](#_Toc128682129)

[6 Elaboração do Data Warehouse 13](#_Toc128682130)

[7 Projeto de ETL 14](#_Toc128682131)

[8 Dashboard 18](#_Toc128682132)

[9 Conclusão 21](#_Toc128682133)

[10 Arquivos 22](#_Toc128682134)

1. Introdução

Antes mesmo do lançamento do seu primeiro produto no mercado em 2018, a Ambinova, empresa brasileira que atua no segmento de tecnologia voltada à automação residencial, definiu em sua visão o firme propósito de gerar soluções habilitadoras para o uso racional dos recursos em um ambiente de negócios cada vez ávido por demonstrações de sustentabilidade.

Com uma linha de produtos já estabelecidos no mercado e uma marca reconhecida e respeitada pelo consumidor nacional, em 2023 a Ambinova deu início a um novo ciclo de pesquisas e desenvolvimento visando incorporar novos recursos ligados à eficiência energética às soluções existentes.

O Plano Estratégico da Ambinova no período de 2023 a 2027 prevê um crescimento de 17% em sua participação no mercado nacional, além do lançamento internacional de sua primeira linha de soluções de climatização de ambientes domésticos que visam reduzir o impacto ambiental em geração energética em até 30%.

Este documento tem por finalidade coletar, analisar e definir as principais necessidades do projeto do estudo de caso que propõe um Novo Paradigma em Automação Residencial. O documento procura demonstrar os principais problemas atuais e o foco investigativo desejado pelo cliente.

1. Estudo de Caso
   1. **Descrição do Estudo de Caso**

Com uma família de produtos já estabelecidos no mercado e uma retração das vendas no último semestre de 2022, a Ambinova abre uma nova frente de estudos de pesquisa e desenvolvimento que visam incorporar aspectos de Inteligência Artificial (IA) enquanto novos recursos ao seu portifólio atual.

O produto ‘carro chefe’ da empresa é o Autohome. Sucesso de vendas em 2018 e 2019 no mercado nacional, o Autohome é uma solução para monitoramento das condições de temperatura e umidade voltada para o ambiente doméstico. Fruto de um desenvolvimento pioneiro e com limitados investimentos, o Autohome combina a simplicidade do hardware com aplicação de sensores de mercado que fazem a aquisição das variáveis climáticas do ambiente com uma arquitetura de software igualmente simples, que estabelece rotinas para registro e apresentação dessas variáveis em forma paramétrica e/ ou gráfico de tendência.

Interface gráfica do usuário

Descrição gerada automaticamente

Figura 1 – Captura de tela da solução Autohome em sua versão original para mobile

Face à retração nas vendas do Autohome nos últimos meses, a Ambinova objetiva com esse estudo encontrar um novo paradigma no monitoramento das condições ambientais domésticas que sejam um diferencial competitivo para a empresa frente aos múltiplos concorrentes que surgiram nos últimos anos. O projeto, portanto, consiste na:

1. proposição de uma nova arquitetura de dados que possa combinar de forma integrada informações intrínsecas e extrínsecas ao ambiente doméstico;
2. criação de um Data Warehouse que permita aos cientistas de dados da empresa atuarem na fabricação de algoritmos para a nova versão do Autohome;
3. criação de um dashboard para visualização dos resultados dos estudos e que possam suportar a decisão de lançamento da nova versão do produto Autohome.

Confiante no papel da tecnologia enquanto vetor fundamental para a construção de soluções com menor impacto ambiental, este projeto é um passo concreto para a viabilização da meta do plano estratégico da Ambinova. Estudos preliminares indicaram que a solução Autohome, caso acrescida de algoritmos de IA, pode em seu primeiro ano de lançamento gerar um impacto de até 10% em redução do consumo energético de aparelhos de ar-condicionado.

Este projeto visa, portanto, fornecer insumos suficientes em termos de arquitetura e estruturação de dados para a definição da próxima linha de produtos que vai conduzir a Ambinova na desafiante jornada de seu planejamento estratégico.

1. Descrição do Modelo Transacional

Esta seção apresenta uma visão situacional da arquitetura de dados atual da solução Autohome da empresa Ambinova.

* 1. ***Fonte 1 – Diagrama de Classes***

Como parte da análise e entendimento do funcionamento do sistema transacional, foi coletado o diagrama de classes do sistema Autohome que é apresentado na Figura 2.

Texto preto sobre fundo branco

Descrição gerada automaticamente

Figura 2 – Diagrama de classes da solução Autohome

O modelo de dados transacional da solução Autohome consiste em 5 tabelas que combinam diferentes informações para realizar o monitoramento das condições ambientais de um determinado ambiente doméstico.

Uma descrição sucinta das tabelas é apresentada abaixo:

* Munícipios e estados: base de dados contendo o código IBGE, nome do município, capital, código UF, UF, estado, latitude, longitude, código SIAFI, DDD e fuso horário de todos os municípios brasileiros. Total de 5.570 registros. Disponível em <https://github.com/kelvins/Municipios-Brasileiros>
* Usuário: base de dados de usuários da aplicação Autohome contendo o nome, email e código IBGE do munícipio onde o usuário está cadastrado no sistema
* Ambiente: base de dados de aparelhos de ar-condicionado contendo o nome atribuído pelo usuário a cada ambiente a ser monitorado
* Monitoramento: Base de dados que concentra as informações para o monitoramento dos ambientes domésticos do Autohome. A base relaciona duas chaves estrangeiras que são os usuários e o ambiente, além de acrescer informações relativamente à temperatura, umidade e estado do ar-condicionado (se ligado ou desligado) para uma determinada data (timestamp) de aquisição.
  1. ***Fonte 2 – Base de dados de prototipação***

Com o propósito de enriquecimento da base de dados transacional, a empresa Ambinova disponibilizou uma base de dados adicional para fins de incorporação de informações já sintetizadas em um protótipo desenvolvido internamente.

A base de dados é um arquivo do tipo SQLite e encontra-se disponível no repositório GitHub no endereço <https://github.com/ciribelli/autohome>.

A Figura 3 mostra um exemplo de aquisições de temperatura e umidade realizadas em dois diferentes ambientes domésticos.

Interface gráfica do usuário, Aplicativo, Tabela

Descrição gerada automaticamente

Figura 3 – Banco de dados de prototipação do Autohome

* 1. ***Fonte 3 – Base de dados externos INMET***

Uma terceira fonte de dados foi indicada pela Ambinova para integração ao seu ambiente de análise. Os dados da terceira fonte são informações meteorológicas de municípios nacionais disponibilizadas pelo Instituto Nacional de Meteorologia (INMET) disponíveis no endereço <https://portal.inmet.gov.br/dadoshistoricos>.

Os dados desta base são informações históricas de precipitação, pressão atmosférica, temperatura, umidade e vento (direção e velocidade) para centenas de municípios brasileiros. A Figura 4 apresenta um exemplo de como são apresentadas essas informações para o caso da estação meteorológica localizada no Forte de Copacabana no Rio de Janeiro.

Interface gráfica do usuário, Aplicativo, Tabela, Excel

Descrição gerada automaticamente

Figura 4 – Planilha INMET para a estação meteorológica localizada no Forte de Copacabana no Rio de Janeiro com registros de 2021

1. Proposta de Processo de BI

O processo de BI proposto neste projeto é um projeto tradicional que se propõe a habilitar análises descritivas de dados históricos para a tomada de decisão no presente e atender os compromissos futuros do Plano Estratégico da Ambinova.

A Figura 5 apresenta o desenho esquemático do projeto de BI com ênfase para o caminho dos dados e a arquitetura da solução que permitirá os profissionais da Ambinova realizar suas atividades.

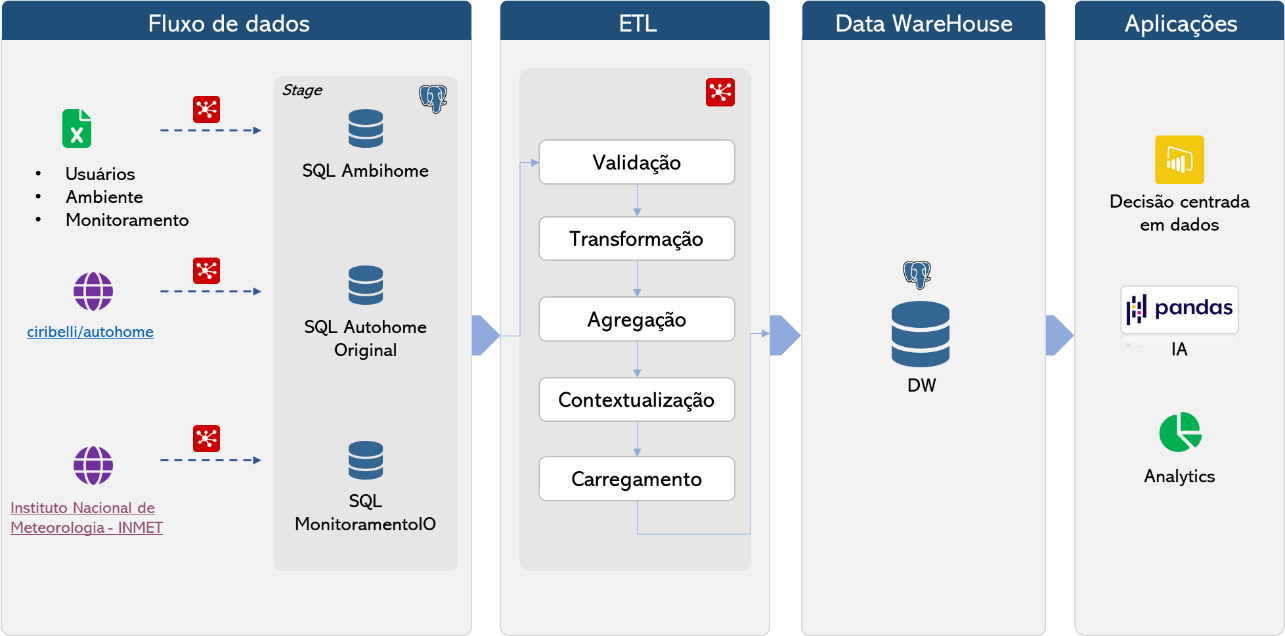


Figura 5 – Esquemático do projeto BI do Ambinova

Sobre os processos principais desse desenvolvimento, pode-se descrever:

* Fluxo de dados: etapa em que os dados transacionais da Ambinova são avaliados e preparados para receber as preparações necessárias para criação do Data WareHouse. Nesta etapa, outras bases de interesse da empresa também são organizadas e concentradas numa arquitetura interna denominada *stage*. Nesta etapa o software Pentaho Data Integration é utilizado para a movimentação e transformação dos dados enquanto o software c é o responsável pela guarda e manipulação de dados e tabelas.
* ETL: etapa em que são realizadas extrações e validações de formatos, transformações e agregações entre as diferentes bases (internas e externas). Por fim, são feitas operações de contextualização para a integração das bases e transporte dos dados (loading) para a base do Data WareHouse. O Pentaho Data Integration é o software principal utilizado na fase de ETL.
* Data WareHouse: etapa em que os dados são persistidos na nova base agora denominada DW. O desenho de arquitetura do DW é construído com vistas para endereçar o desafio da Ambinova.
* Aplicações: etapa em que os protagonistas pelo desenvolvimento das atividades são os próprios profissionais da Ambinova. Seja por meio de painéis BI ou criação de *datasets* para geração de algoritmos se valham de Inteligência Artificial que irão compor a nova versão do produto Autohome.

1. Modelo Multidimensional

Esta seção apresenta o modelo *snowflake* do estudo de caso desenvolvido para a empresa Ambinova. Na Figura T está apresentada uma captura do modelo *snowflake* desenvolvida no software SQL Power Architect 1.0.9 com destaque para a relação entre regiões (estados e municípios) que se relacionam de forma unívoca para então combinar com as outras três dimensões do projeto e a tabela fato proposta neste desenvolvimento.

Texto preto sobre fundo branco

Descrição gerada automaticamente

Figura 6 – Modelo multidimensional em configuração snowflake (projeto Ambinova 2023).

1. Elaboração do Data Warehouse

O Data Warehouse será a fonte integradora de informações da empresa, a tecnologia será utilizada com o intuito de servir de base para a camada de aplicação que será responsável por fornecer dados para a tomada de decisão na organização.

* 1. ***Definição do DW***
     1. Arquitetura

A arquitetura do projeto será Global e Integrada. Tal estratégia se deve em razão de uma boa adequação desta arquitetura ao modelo de negócio desenvolvido pela Ambinova que é do tipo Business to Consumer (B2C).

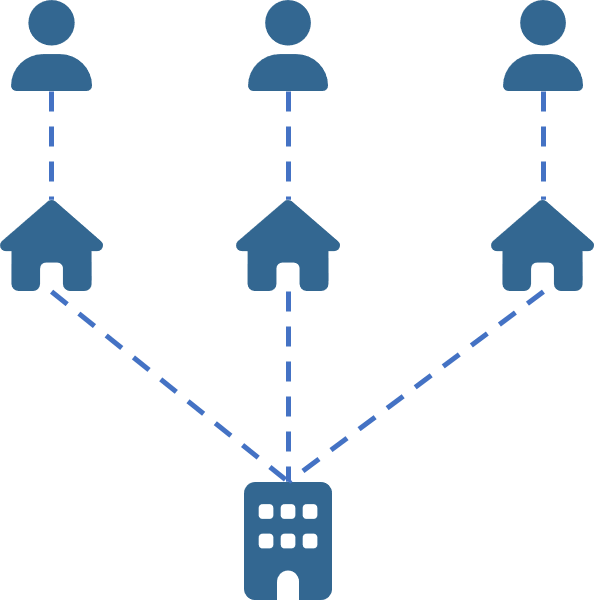


Figura 7 – Arquitetura simplificada do modelo de negócios Ambinova

* + 1. Abordagem de Construção

Como a arquitetura será Global e Integrada, não estão previstos entregáveis do tipo Data Marts neste projeto. No entanto, pode-se prever futuramente uma abordagem Bottom Up, caso a Ambinova queira explorar regionalmente novos pilotos de desenvolvimento.

* + 1. Arquitetura Física

Todo o projeto será criado numa estrutura de nuvem contratada pela empresa, ou seja, o DW ficará armazenado no servidor de um terceiro. Tal estratégia tem aderência com a estrutura atual da companhia e tem por objetivo reduzir custos com o trânsito de dados entre nuvem e servidores locais além de dispensar investimentos em links de maior conectividade entre a Ambinova e a nuvem contratada.

1. Projeto de ETL
   1. ***Descrição do Projeto de ETL***

Para atender o objetivo do projeto, foi criado o banco de dados denominado ‘dw\_ambinova’ que deverá desempenhar a função do Data WareHouse (DW). Suas tabelas então foram criadas a partir de rotinas SQL extraídas do software Power Architect. Na Tabela 1, são apresentadas as rotinas utilizadas para implementação deste DW.

Tabela 1 – Lista de arquivos SQL utilizados para geração do DW

|  |  |  |
| --- | --- | --- |
|  | Arquivo | Ação |
| Ícone  Descrição gerada automaticamente | script\_dim\_tempo\_resumido.sql | * Cria tabela dim\_data * Povoa a tabela com dados de 01/01/2018 à 31/12/2037 * Gera uma série com timestamps separados por uma hora |
| Ícone  Descrição gerada automaticamente | script\_cria\_dw\_ambinova\_snow.sql | * Cria tabela dim\_estados * Cria tabela dim\_municipios * Cria tabela dim\_ambiente * Cria tabela dim\_usuario * Cria tabela tabela ft\_monitoramento * Cria a dependência entre as tabelas |
| Ícone  Descrição gerada automaticamente | script\_dnqs\_snow.sql | * Inicializa a primeira linha das tabelas (exceção de dim\_data e ft\_monitoramento) com dados fictícios |
|  | script\_apaga\_dims.sql | * Apaga todas as tabelas do DW caso seja necessária uma nova implementação |

* + 1. Construção do *stage*

Parte essencial para atingimento do objetivo do projeto é realizar a adequada integração dos registros horários (temperatura e umidade) de cada ambiente para todos os usuários do banco transacional da Ambinova com os dados de registros externos disponíveis no site do INMET. Dada a complexidade da tarefa, esse processo foi dividido em duas ondas, sendo que a primeira consiste em transportar os dados do INMET, que são disponibilizados no formato 'xls' (Figura 8), para o *stage* do banco transacional da Ambinova. Importante destacar que todas as etapas no projeto de ETL foram realizados com o suporte do software PDI.

Interface gráfica do usuário, Texto, Aplicativo

Descrição gerada automaticamente

Figura 8 – Dados históricos horários para todas as capitais do Brasil para o ano de 2021 (fonte: site do INMET)

No quadro abaixo, é apresentado o processo de carregamento dos dados 'xls' para o banco transacional (SW), com ênfase para as etapas intermediárias de transformação dos dados:

|  |  |
| --- | --- |
| Nome do arquivo: | etl\_carrega\_meteo\_cidades\_lookup\_resumo.ktr |
| Descrição dos processos | Esse processo carrega todos os arquivos excel do INMET com registros horários diversos para todas as capitais:   1. Após identificação ordenação dos dados por municípios, faz-se uma junção com a adição de uma nova coluna com o código IBGE do município 2. Colunas recebem novos nomes e são retirados caracteres especiais 3. O campo hora, que é uma coluna separada nos dados do INMET, é concatenado com a data dos registros formando um string formado por 'data + hora' 4. Colunas desnecessárias num primeiro momento (velocidade e direção do vento, por exemplo) são eliminadas dos dados 5. A data é convertida do formato string para o formato timestamp 6. Dados são salvos no *stage* no SW na tabela meteo\_resumo |
| Uma imagem contendo Mapa  Descrição gerada automaticamente | | |

Neste ponto do processo, os dados externos do INMET já estão incorporados ao stage, porém, sem a devida contextualização que é um requisito do projeto. Uma segunda onda de integração dos dados acontecerá durante o carregamento dos dados no DW, quando registros de temperatura e umidade, tanto internos quanto externos, serão devidamente associados na base multidimensional.

* + 1. Carregamento dos dados no DW

Com o *stage* totalmente preparado após concluídas as etapas de limpeza, tratamento, agregação e validação dos dados, deve-se iniciar a carga propriamente dita do DW. Os quadros a seguir apresentam os processos de ETL para carregamento do DW:

|  |  |
| --- | --- |
| Nome do arquivo: | etl\_dw\_carrega\_usuario.ktr |
| Descrição dos processos | 1. Esse processo carrega os dados dos usuários no DW atribuindo uma chave do tipo surrogate key para cada usuário do SW. 2. Também são adicionadas colunas ligadas à versão e datas de início e fim da ETL. |
| Gráfico  Descrição gerada automaticamente | |

|  |  |
| --- | --- |
| Nome do arquivo: | etl\_dw\_carrega\_ambiente.ktr |
| Descrição dos processos | 1. Esse processo carrega os dados dos ambientes no DW atribuindo uma chave do tipo surrogate key para cada ambiente do SW. 2. Também são adicionadas colunas ligadas à versão e datas de início e fim da ETL. |
| Uma imagem contendo Gráfico  Descrição gerada automaticamente | |

|  |  |
| --- | --- |
| Nome do arquivo: | etl\_dw\_carrega\_municipios\_snow.ktr |
| Descrição dos processos | 1. Esse processo carrega os dados dos municípios no DW atribuindo uma chave do tipo surrogate key para cada município do SW. 2. Em razão do modelo multidimensional do tipo *snowflake*, uma etapa adicional de lookup é aplicada para também atribuir uma surrogate key para os estados. 3. Essa surrogate key dos estados também é atribuída enquanto uma chave estrangeira na tabela dim\_municipios. 4. Também são adicionadas colunas ligadas à versão e datas de início e fim da ETL para ambas as tabelas (estados e municípios) |
| Diagrama  Descrição gerada automaticamente | |

O quadro abaixo apresenta a transformação que conclui o projeto de ETL com a atualização da tabela fato ‘ft\_monitoramento’. Nesta mesma etapa, acontece a segunda onda de integração dos dados onde é feita a correspondência geográfica e temporal entre dados de temperatura e umidade registrados nos ambientes dos usuários e os dados de temperatura e umidade registrados em cada cidade e disponibilizados pelo INMET.

|  |  |
| --- | --- |
| Nome do arquivo: | etl\_dw\_carrega\_monitoramento\_snow.ktr |
| Descrição dos processos | Com os dados do INMET já atualizados no *stage* (SW), esse processo carrega esses dados e realiza as seguintes transformações:   1. Carrega a tabela monitoramento do SW original da Ambinova e recupera a informação do município de cada registro (primeiro lookup) 2. Converte o formato do timestamp para o padrão yyyy/MM/dd HH:mm:ss 3. Faz a agregação dos dados da tabela monitoramento (registros internos dos usuários) com os dados da tabela 'meteo\_resumo' (dados externos gerados na primeira onda) 4. Remove colunas desnecessárias, renomeia colunas de interesse e salva no DW a tabela fato 'ft\_monitoramento' |
|  | |

Com a gravação da tabela fato, dá-se por concluído o processo de ETL deste projeto da Ambinova. A partir de agora, produtos e aplicações poderão utilizar o acesso direto ao servidor do DW para construir visualizações do tipo BI que irão subsidiar decisões baseadas em dados *(decision driven*). Também poderão ser construídos *dataframes* para aplicação de técnicas de *data mining* e construção da nova versão do produto Autohome.

1. Dashboard
   1. ***Descrição da Elaboração***

Como etapa final do projeto Ambinova, foi desenvolvido um dashboard na plataforma Microsoft PowerBI para visualização dos dados do DW. O desenvolvimento do dashboard considerou os seguintes aspectos enquanto requisitos funcionais:

* Recurso de filtro/seleção de usuários e/ou ambientes
* Recurso de filtro/seleção do período a ser avaliado
* Apresentação de dados estatísticos principais para temperatura e umidade tanto para ambiente interno quanto para os dados externos (INMET)
* Visualização das séries temporais que permitam demonstrar a integração de dados internos e externos (INMET)
* Visualização do status de ar-condicionado ligado para os casos dos testes piloto utilizando recursos de Inteligência Artificial
* Exportação dos dados para utilização de técnicas de *Machine Learning* e/ou *Data Mining*.

Como requisitos não-funcionais, foram considerados:

* Desempenho e taxa de atualização de dados adequado para uma massa de dados de até um ano
* Identidade visual do dashboard aderente ao padrão de comunicação da Ambinova

Do ponto de vista de transformações no contexto do PowerBI, foi necessário converter os dados ligados ao status do ar-condicionado de natureza booleana para percentual. Essa etapa foi necessária para a representação destas informações em justaposição às séries temporais de temperatura e umidade.

A Figura 9 apresenta a tela principal do dashboard criado para o projeto Ambinova.

Uma imagem contendo Interface gráfica do usuário

Descrição gerada automaticamente

Figura 9 – Dashboard principal desenvolvido em PowerBI (projeto Ambinova 2023)

1. Conclusão

O projeto da Ambinova foi uma oportunidade excepcional para trilhar de ponta-a-ponta o caminho desde a concepção e criação de um banco de dados transacional até a conclusão do projeto com a geração do Data WareHouse e a visualização do dashboard em Power BI.

As dificuldades naturalmente encontradas ao longo do projeto foram felizmente superadas com o suporte da literatura e materiais disponibilizados ao longo da disciplina de BI. Alguns caminhos críticos do projeto, como a definição da topologia mais adequada para o modelo multidimensional (se *star* ou *snowflake*), provocaram a proposição de recursos automatizados para a reconstrução ágil das bases de dados. Técnicas para popular tabelas simulando bases de dados com temperaturas e umidades dos ambientes domésticos também foram desenvolvidas para testar o desempenho da solução com dados históricos extensos.

Quanto ao aspecto de agregação/integração dos dados internos dos usuários à base externa do INMET, pode-se confirmar por meio do dashboard desenvolvido uma plena aderência tanto de contexto como temporal desses dados. Cabe ressaltar o caráter crítico desta dimensão para o atingimento pleno dos objetivos do projeto.

Por fim, registra-se que o resultado em termos de qualidade e desempenho da solução final foram atingidos com total êxito. A escalabilidade e a versatilidade desta solução em abrangência a outros processos é algo muito valioso e que certamente será aplicado pelo time de projeto em outros desafios.

1. Arquivos

Os arquivos, rotinas, imagens e relatórios utilizados no projeto Ambinova, bem como os arquivos do INMET utilizados no exercício de integração, foram disponibilizados em repositório do GitHub no endereço <https://github.com/ciribelli/Ambinova>.

A Figura 10 apresenta uma captura da estrutura de arquivos do projeto Ambinova no contexto do GitHub.

Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente

Figura 10 – Arquivos do projeto Ambinova no GitHub

A estrutura de arquivos pode ser brevemente descrita por:

* Arch: pasta contendo arquivos do modelo multidimensional no formato SQL Power Architect
* Etl: pasta contendo arquivos para construção do DW no formato Pentaho Data Integration
* Img: pasta contendo arquivos do tipo png utilizados na construção do dashboard
* Sql\_scripts: pasta com duas subpastas com rotinas de
  + Criação do SW – contém arquivos do tipo SQL para criação e manipulação de tabelas do SW. Contem arquivo com rotina em python para população de tabelas com dados de temperatura e umidade para usuários
  + Criação do DW – contém arquivos do tipo SQL para criação e manipulação de tabelas do DW
* Xls\_cidades: pasta contendo os dados do INMET para o ano de 2021 para todas as capitais brasileiras
* ProjetoFinal\_OtavioCiribelli.docx: relatório final do projeto Ambinova