

MAA

Mestrado em Métodos Analíticos Avançados

Master Program in Advanced Analytics

INTEGRAÇÃO DE DADOS NO CONTEXTO DO SECTOR DE ENERGIA SOLAR FOTOVOLTAICO

Um caso de estudo EDP

Diogo Acabado

Relatório de Estágio apresentado como requisito parcial para
obtenção do grau de Mestre em Métodos Analíticos
Avançados



NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

INTEGRAÇÃO DE DADOS NO CONTEXTO DO SECTOR DE ENERGIA SOLAR
FOTOVOLTAICO - CASO DE ESTUDO EDP

por

Diogo Acabado

Relatório de Estágio apresentado como requisito parcial para a obtenção do grau de Mestre em
Métodos Analíticos Avançados

Orientador/Coorientador: Bruno Miguel Pinto Damásio

Novembro 2022

RESUMO

A sociedade moderna depende fortemente do uso de energia para todo o tipo de atividades, e o seu crescimento sustentável está assente no uso de formas de energia inesgotáveis e menos poluentes. De forma a transicionar para um uso predominante deste tipo de energia, foram definidas metas europeias de sustentabilidade, tendo-se a EDP posicionado como líder desta transição em Portugal. Uma das áreas de atividade dentro desta empresa que suporta esse posicionamento é a área de energia solar fotovoltaica, onde é feita a aquisição e gestão de sistemas fotovoltaicos para a produção de eletricidade a partir de energia solar. Estes sistemas precisam de ser geridos da forma mais eficiente e eficaz possível, e, para tal, foi proposto no contexto deste estágio implementar processos de automatização de integração de dados de inversores, que são o componente mais importante do sistema fotovoltaico. Estes dados foram obtidos através de uma API, que foi disponibilizada à empresa por um dos seus fornecedores de inversores, e a sua integração foi conseguida através da implementação de processos de extração, transformação, e inserção de dados, uma sequência comumente usada neste tipo de operação. O produto final deste estágio contribui para a implementação futura de métodos avançados de inteligência artificial, de forma a promover a inovação e o crescimento da área de energia solar da empresa.

PALAVRAS-CHAVE

EDP; Energia Solar Fotovoltaica; Integração de dados; Pipelines ETL; Automatização;

ABSTRACT

Modern society depends heavily on the use of energy for all kinds of activities, and its sustainable growth is based on the use of inexhaustible and less polluting forms of energy. In order to transition to a predominant use of this type of energy, European sustainability targets have been defined and EDP has positioned itself as a leader in this transition in Portugal. One of the areas of activity within this company that supports this positioning is the area of photovoltaic solar energy, where the acquisition and management of photovoltaic systems for the production of electricity from solar energy is carried out. These systems need to be managed as efficiently and effectively as possible, and to this end, it was proposed in the context of this internship to implement automation processes for integrating data from inverters, which are the most important component of the photovoltaic system. This data was obtained through an API, which was made available to the company by one of its inverter suppliers, and its integration was achieved by implementing processes of data extraction, transformation, and insertion, a sequence commonly used in this type of operation. The final product of this internship contributes to the future implementation of advanced artificial intelligence methods to promote innovation and growth in the company's solar energy area.

KEYWORDS

EDP; Solar Photovoltaic Energy; Data Integration; ETL Pipelines; Automation;

ÍNDICE

1. Introdução.....	1
1.1. A Energia	1
1.2. A EDP	2
1.2.1. Apresentação da empresa	2
1.2.2. Plano estratégico	2
1.2.3. Estrutura da empresa.....	3
1.3. Objetivos do estágio.....	4
1.4. Estrutura do relatório	5
2. Revisão da Literatura	7
2.1. O Sistema Fotovoltaico	7
2.1.1. Componentes de um sistema fotovoltaico	7
2.2. Tipos de sistemas FV	8
2.3. Integração de dados.....	9
2.4. Ferramentas	9
3. Metodologia.....	12
3.1. Enquadramento do estágio.....	12
3.2. Caso de estudo.....	13
3.2.1. Entendimento dos Dados.....	13
3.3. Implementação da <i>Pipeline ETL</i>	14
3.3.1. Criação de Conectores	15
3.3.2. Extração de Dados.....	15
3.3.3. Transformação de Dados	18
3.3.4. Inserção de Dados.....	18
3.3.5. Agendamento Automático.....	19
4. Conclusões	21
5. Referências.....	23

ÍNDICE DE FIGURAS

Figura 1: Organograma EDPC (Direção de destaque representada a azul).....	3
Figura 2: Organograma EDPC - Direção Serviços de Energia B2B.	4
Figura 3: Sistema FV ligado à rede, com um painel FV e um inversor. Fonte: (Messenger & Ventre, 2005)	8
Figura 4: Sistema FV autónomo, com um painel FV, um controlador de carga, uma bateria, e uma fonte de consumo. Fonte: (Messenger & Ventre, 2005)	9
Figura 5: Tabelas de dados de <i>Sunnyportal</i> , <i>Salesforce</i> e <i>MS Azure</i>	14
Figura 6: Estrutura de comunicação com a <i>API</i> da <i>SMA</i>	15
Figura 7: Método 1: Número da proposta.	16
Figura 8: Método 2: CPE.	17
Figura 9: Método 3: ID da Planta.	17
Figura 10: Inserção de um dispositivo em <i>MS Azure</i>	19

LISTA DE SIGLAS E ABREVIATURAS

B2B	Business to Business
CA	Corrente Alternada
CC	Corrente Contínua
CRM	Customer Relationship Manager
EDPC	EDP Comercial
FV	Fotovoltaico(s)
GA	Gestão de Ativos (Área de Negócio)
GEE	Gases do efeito estufa
O&M	Operações e Manutenção
RNC 2050	Roteiro da Neutralidade de Carbono 2050
SMA	SMA Solar Technology AG

1. INTRODUÇÃO

1.1. A ENERGIA

A energia é um recurso essencial para a maioria das atividades da sociedade moderna, e a sua existência em abundância está relacionada com o bem-estar e prosperidade do ser humano. Computadores, transportes, e iluminação são apenas alguns exemplos de como a energia se tornou uma necessidade para o funcionamento atual da sociedade. Assim sendo, é essencial para a humanidade desenvolver tecnologias que permitam aumentar o acesso a este recurso, assim como fazer a sua gestão adequada.

Consoante a capacidade de reposição da sua oferta, a energia pode ser classificada como não renovável ou renovável. Energia não renovável é definida como a energia obtida a partir de recursos energéticos cuja taxa de regeneração não é suficiente para manter o ritmo do seu uso, sendo estes considerados recursos esgotáveis. Por exemplo, o carvão, o petróleo, e o gás natural são fontes de energia não renovável. Por outro lado, a energia renovável é a energia que é obtida de fontes naturais que podem ser renovadas e regeneradas, tais como a energia solar, a energia hidráulica, e a energia eólica. Estas apresentam-se como a opção mais viável para o crescimento sustentável da sociedade, uma vez que a produção de energia não renovável é um dos principais responsáveis pelo aquecimento global, devido à emissão de gases do efeito estufa [GEE] que são libertados aquando do seu uso.

De modo a combater este problema, a UE desenvolveu o acordo de Paris, que tem como meta manter o aumento da temperatura global bem abaixo dos 2°C acima dos níveis pré-industriais, e limitar o aumento da temperatura a 1,5°C, reconhecendo que este limite irá reduzir significativamente os impactos das alterações climáticas (UNFCCC, 2018).

O Relatório Especial do IPCC afirma que, a fim de não exceder 1,5 °C de aumento da temperatura global, as emissões líquidas antropogénicas globais de CO₂ deverão diminuir até 2030 em cerca de 45% em relação aos níveis de 2010, com vista a atingir a neutralidade carbónica até 2050 (Rogelj, et al., 2018). Para alcançar este estado de neutralidade, que consiste em atingir um equilíbrio entre as emissões de GEE e o sequestro de carbono, serão necessárias reduções substanciais nas emissões destes gases, e/ou aumentos substanciais nos sumidouros nacionais de carbono, que têm de se materializar entre agora e 2050.

Com o objetivo de mapear a trajetória que permite alcançar a neutralidade carbónica até 2050 em Portugal, foi desenvolvido pelo governo o Roteiro da Neutralidade de Carbono 2050 [RNC 2050]. Este documento identifica os principais vetores de descarbonização em todos os sectores da economia, as opções de políticas e medidas a serem implementadas, e a trajetória de redução de emissões de GEE, para diferentes cenários de desenvolvimento socioeconómico. Todos os sectores devem contribuir para a redução das emissões de GEE, sendo o sistema energético o que mais contribui para este fim, particularmente no que diz respeito à produção e transporte de eletricidade.

Em resposta ao desenvolvimento do RNC 2050, a empresa Energias de Portugal, S.A. [EDP], como líder no sector energético em Portugal, posicionou-se como um dos principais contribuintes para o alcance dos objetivos definidos neste documento. A empresa tem como expectativa operar sem consumo de carvão até 2025, e ser 100% verde até 2030, antecipando em 20 anos metas previamente definidas para a descarbonização (EDP Annual Report, 2021).

1.2. A EDP

1.2.1. Apresentação da empresa

A EDP é uma empresa portuguesa do setor energético fundada em 1976, através da nacionalização e fusão da maioria das empresas existentes na altura no setor elétrico. A empresa está presente em toda a cadeia de valor da eletricidade, que consiste na produção, transmissão, distribuição e comercialização deste recurso, e na atividade de comercialização de gás. Trata-se do maior fornecedor de energia elétrica em Portugal e do terceiro maior na Península Ibérica, estando presente em 20 países e 4 continentes, onde conta com cerca de 12.000 colaboradores (EDP Annual Report, 2021).

Em termos da visão da empresa, a EDP ambiciona ser “Uma empresa global de energia, líder na transição energética com vista à criação de valor superior” (EDP Annual Report, 2021, p. 32), regendo-se por uma cultura de valores baseada na Sustentabilidade, na Inovação e na Humanização. Esta visa a melhoria da qualidade de vida das gerações atuais e futuras (“Sustentabilidade”), tem o intuito de criar valor nas várias áreas em que atua (“Inovação”), e pretende construir relações genuínas e de confiança com os colaboradores, clientes, parceiros e comunidades (“Humanização”) (EDP Annual Report, 2021).

Baseado na visão e nos valores da EDP, foram estabelecidos compromissos com o meio ambiente, colaboradores, clientes e acionistas. Estes compromissos são os seguintes (EDP Annual Report, 2021):

1. Sustentabilidade: Assumir as responsabilidades sociais e ambientais que resultam da atuação da empresa, contribuindo para o desenvolvimento das regiões onde esta está presente. Reduzir, de forma sustentável, as emissões específicas de gases com efeito de estufa da energia produzida pela empresa. Promover ativamente a eficiência energética.
2. Pessoas: Aliar uma conduta ética e de rigor profissional, ao entusiasmo e iniciativa, valorizando o trabalho em equipa. Promover o desenvolvimento das competências e o mérito, assim como o equilíbrio entre a vida pessoal e profissional.
3. Clientes: Colocar-se no lugar dos clientes sempre que houver uma tomada de decisão. Ouvir os clientes e responder de forma simples e transparente. Surpreender os clientes, antecipando as suas necessidades.
4. Resultados: Cumprir os compromissos assumidos perante os acionistas. Liderar através da capacidade de antecipação e execução. Exigir a excelência em todas as atividades da empresa.

1.2.2. Plano estratégico

Em 2021, a EDP apresentou o seu Plano Estratégico até 2025, que se baseia em 3 pilares estratégicos: crescimento acelerado e sustentável; organização orientada para o futuro; e retorno atrativo e excelência nos indicadores ESG (*Environment, Social, Governance*). Com base nestes pilares, a EDP propõe-se a alcançar, não exclusivamente, os seguintes objetivos (EDP Annual Report, 2021):

- Aumentar o crescimento nas energias renováveis, investindo cerca de 24 Bn. na transição energética (processo de substituição de energias não renováveis por energias renováveis) até 2025;
- Reforçar o foco na inovação e promover uma organização digitalmente habilitada, com um investimento de cerca de 2 Bn. até 2025;
- Operar sem consumo de carvão até 2025, e atingir a Neutralidade Carbónica em 2030.

1.2.3. Estrutura da empresa

A EDP atua em três áreas de negócio: *Renewables*, *Networks*, e *Client Solutions & Energy Management*. O segmento das *Renewables* corresponde à atividade de produção de eletricidade a partir de fontes renováveis, principalmente hídrica, eólica e solar; o segmento das *Networks*, por outro lado, corresponde às atividades de distribuição e transmissão de eletricidade; por último, o segmento das *Client Solutions & Energy Management* inclui as atividades de produção de eletricidade a partir de fontes não renováveis, fornecimento de eletricidade e gás, e a gestão de compras e vendas de energia nos mercados ibérico e brasileiro. Este último segmento inclui a empresa EDP Comercial - Comercialização de Energia, S.A. [EDPC], empresa onde decorreu o estágio relatado neste documento (EDP Annual Report, 2021).

Na EDPC existem atividades que exercem uma função de suporte à empresa, assim como gabinetes e direções responsáveis por funções corporativas e de negócio. Uma destas direções é a Direção Serviços de Energia B2B (Ver figura 1), que tem a Gestão de Ativos [GA] como área de atividade. É nesta área e dentro da equipa de Controlo e Otimização de Ativos que as tarefas descritas neste relatório se inserem. (Ver figura 2).

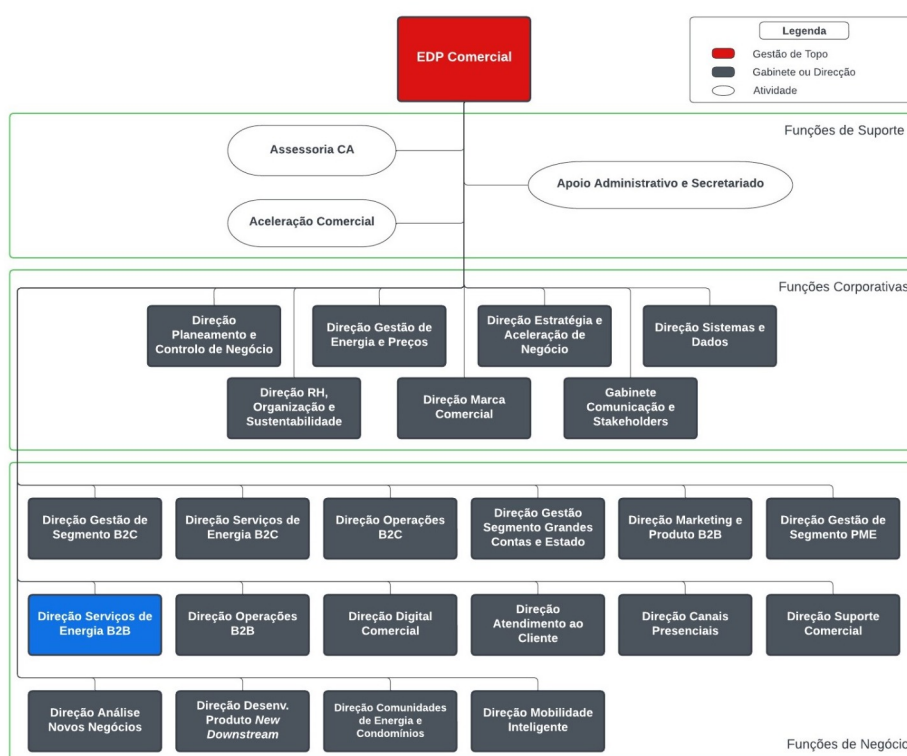


Figura 1: Organograma EDPC (Direção de destaque representada a azul).

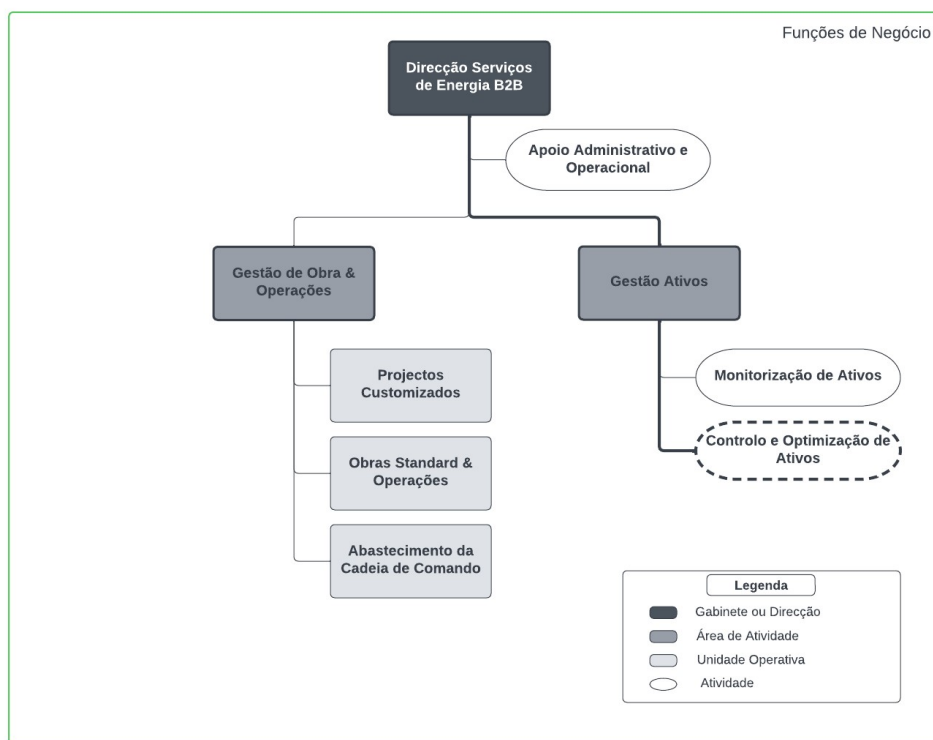


Figura 2: Organograma EDPC - Direção Serviços de Energia B2B.

1.3. OBJETIVOS DO ESTÁGIO

Em consequência do objetivo de aumentar o crescimento nas energias renováveis delineado no plano estratégico traçado pela empresa, a área de GA da EDPC tem vindo a crescer acentuadamente no número de ativos energéticos renováveis que são geridos no sector B2B nos últimos anos, nomeadamente nas áreas de energia solar e mobilidade elétrica. Relativamente à área de energia solar, houve um crescimento de 390 ativos em 2019, para 1.914 em 2022, representando um aumento relativo de 390% em três anos. Por outro lado, na área de mobilidade elétrica houve um crescimento de 37 ativos em 2019, para 1.424 em 2022 (aumento relativo de 3750%) (EDPC - Onboarding Gestão de Ativos, 2020). Este crescimento criou a necessidade na equipa da GA de fazer uma gestão o mais otimizada possível dos seus ativos.

De forma a satisfazer esta necessidade, e de acordo com a intenção de reforçar o foco na inovação e digitalização definido no plano estratégico da EDP, os analistas da equipa de Controlo e Otimização (também designada por equipa de *Analytics*) propuseram-se a desenvolver projetos de automatização e de data *analytics*, de modo a auxiliar os gestores de ativos da equipa de Monitorização (também designada por equipa de Gestores de Ativos) a fazer uma gestão mais eficiente e rigorosa dos seus ativos.

Assim sendo, o estágio realizado tem como função principal dar suporte aos gestores dos ativos de energia solar, através do desenvolvimento de um processo de integração de dados. Estes dados servem de base para integrações futuras, que permitirão fazer a avaliação do desempenho dos ativos,

assim como para fazer a deteção de anomalias nos mesmos, a fim de permitir aos gestores de ativos planejar intervenções de manutenção.

1.4. ESTRUTURA DO RELATÓRIO

Para além do presente capítulo, este relatório é constituído por quatro outros capítulos. O capítulo 2 (Enquadramento Teórico) introduz e discute criticamente os principais conceitos teóricos sobre sistemas fotovoltaicos, nomeadamente as suas componentes e os diferentes tipos. Segue-se a contextualização teórica do processo de integração de dados onde são referidas as etapas do processo ETL que é o método que será implementado no capítulo seguinte. Por último, são apresentadas as ferramentas utilizadas durante o estágio para implementar este método. O capítulo 3 (Metodologia) enquadra o estágio no âmbito do funcionamento da área de negócio e tendo em conta as necessidades das equipas e a forma como se relacionam. Segue-se uma descrição detalhada da metodologia ETL aplicada a um caso de estudo. Por último, o capítulo 4 (Conclusão) expõe as principais aprendizagens obtidas, impacto do trabalho desenvolvido e trabalho futuro a desenvolver.

2. REVISÃO DA LITERATURA

Este capítulo apresenta todos os conceitos teóricos relacionados com o funcionamento do ativo solar relevante no contexto do estágio realizado - o sistema (ou instalação) solar fotovoltaico(a) [FV].

2.1. O SISTEMA FOTOVOLTAICO

2.1.1. Componentes de um sistema fotovoltaico

Um sistema solar fotovoltaico tem como função principal gerar eletricidade a partir de radiação solar, sendo composto por vários componentes com funcionalidades específicas que podem ser agrupados nas seguintes categorias: dispositivos fotovoltaicos; componentes eletrônicos de potência; dispositivos de armazenamento de energia; e componentes elétricos e mecânicos. Todas as instalações fotovoltaicas têm os mesmos componentes básicos necessários para a produção de eletricidade, mas dependendo do tipo de instalação poderão ser necessários alguns componentes adicionais (Gorjian & Shukla, 2020).

2.1.1.1. Dispositivos fotovoltaicos

O elemento mais fundamental do sistema fotovoltaico é a célula fotovoltaica, onde ocorre a produção de eletricidade devido à propriedade semicondutora do material com que é feita. A fim de obter níveis de produção elétrica significativos, as células são agrupadas numa estrutura metálica denominada módulo (ou painel) fotovoltaico, sendo este o equipamento de geração de energia comercialmente disponibilizado para utilização. Por último, os módulos podem ser agrupados em *strings* (ou matriz), para instalações fotovoltaicas de maior dimensão (Gorjian & Shukla, 2020).

2.1.1.2. Componentes eletrônicos de potência

O tipo mais comum de dispositivos eletrônicos de potência num sistema fotovoltaico é o inversor, e este converte a corrente elétrica gerada pelos dispositivos fotovoltaicos de contínua (CC) para alternada (CA). Os inversores podem ser categorizados em micro-inversores, inversores de *string*, e inversores centrais, de acordo com o seu tamanho e a potência de saída da corrente elétrica (Gorjian & Shukla, 2020).

Entre os três tipos de inversores aqui mencionados, os micro-inversores são aqueles que têm uma classificação de potência de saída mais baixa, tendo em geral tamanhos mais pequenos. Os outros dois tipos (*string* e centrais) têm classificações de potência de saída mais elevadas, e estão disponíveis numa maior gama de tamanhos. Em termos da ligação ao sistema FV, os micro-inversores são conectados a um único módulo FV, enquanto os inversores *string* e central são ligados aos *strings* FV, com base nas classificações de intensidade de corrente e tensão do sistema FV em questão (Gorjian & Shukla, 2020).

2.1.1.3. Armazenamento de energia

Os dispositivos de armazenamento de energia mais utilizados nos sistemas FV são as baterias recarregáveis, cujo propósito é armazenar a energia gerada pelos dispositivos FV, podendo esta ser disponibilizada numa altura posterior. Devido à natureza dos sistemas FV, estes só geram eletricidade quando existe radiação captada pelos seus painéis FV, sendo necessário usar a energia armazenada nas baterias durante a noite, ou em períodos de baixa incidência solar (ex.: céu nublado). O uso de

baterias requer a incorporação de um controlador de carga no sistema, para que as baterias possam ser impedidas de atingir uma condição de sobrecarga ou de descarga excessiva (Messenger & Ventre, 2005).

2.1.1.4. Componentes elétricos e mecânicos

Por último, nos sistemas FV existem vários componentes elétricos e mecânicos. Os componentes elétricos incluem, entre outros elementos, cabos e caixas de junção, enquanto os componentes mecânicos incluem as estruturas de suporte utilizadas para a instalação.

2.2. TIPOS DE SISTEMAS FV

Um dos principais fatores que distingue os diferentes sistemas FV é a sua ligação à rede elétrica de serviço público, onde é feito o transporte e distribuição de eletricidade a nível nacional (E-Redes, 2021). Em função da sua relação com a rede, os sistemas FV são classificados em três tipos: sistemas FV ligados à rede; sistemas FV autónomos; e sistemas FV híbridos.

Os sistemas FV que produzem eletricidade para introduzir na rede elétrica são designados por sistemas FV ligados à rede. Os principais componentes de um sistema deste tipo são os módulos FV e o inversor, como ilustrado na figura 3.

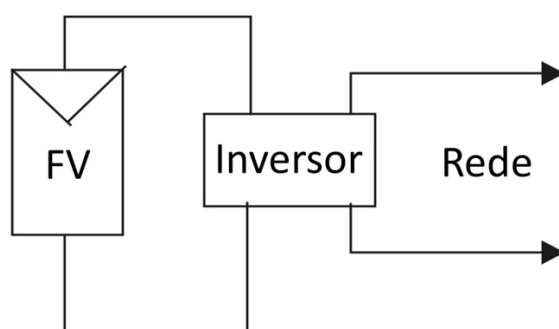


Figura 3: Sistema FV ligado à rede, com um painel FV e um inversor. Fonte: (Messenger & Ventre, 2005)

O segundo tipo de sistemas FV aqui referidos são aqueles que geram eletricidade para ser utilizada localmente, sendo estes designados por sistemas FV autónomos. Nestes sistemas, a maior parte da eletricidade que foi gerada é utilizada pelas fontes de consumo (ex.: eletrodomésticos), e qualquer excesso disponível será armazenado em baterias, não havendo injeção de energia na rede. A figura 4 que se segue apresenta de forma esquemática os principais componentes deste tipo de sistema FV, que incluem os módulos FV, a bateria, o controlador de carga, e uma fonte de consumo.

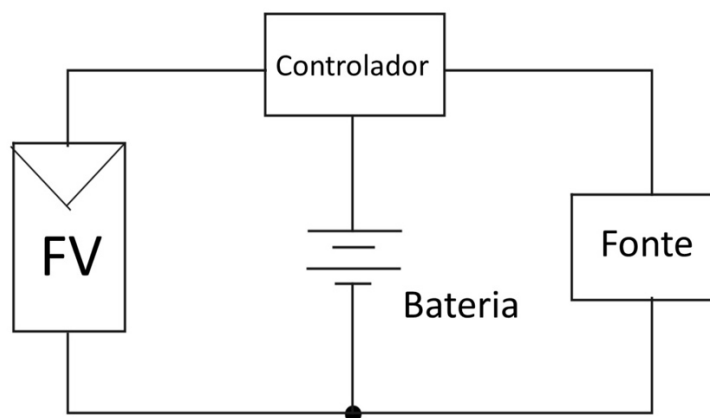


Figura 4: Sistema FV autônomo, com um painel FV, um controlador de carga, uma bateria, e uma fonte de consumo. Fonte: (Messenger & Ventre, 2005)

Por último, os sistemas FV híbridos representam a combinação de ambos os tipos referidos anteriormente. Na configuração híbrida, o sistema FV gera energia elétrica para ser consumida localmente ou armazenada em baterias, sendo que a energia produzida em excesso é injetada na rede. Estes sistemas usam todos os componentes dos sistemas FV ligados à rede e dos autônomos.

2.3. INTEGRAÇÃO DE DADOS

A integração de dados consiste na combinação de dados de diferentes fontes para que estes sejam acessíveis e utilizáveis num único sistema. A automatização da tarefa de integração de dados é feita com o recurso ao processo extração - transformação - inserção (*extract - transform - load*) [ETL]. Este é o processo no qual dados são extraídos de uma fonte, são configurados, e são armazenados numa base de dados (Sherman, 2015).

No primeiro passo do processo ETL é feita a extração de dados a partir de fontes de dados distintas, a fim de serem usados nas fases seguintes do processo. Cada fonte tem características diferentes no que toca aos protocolos de comunicação com a mesma, assim como ao tipo e formato dos dados que são extraídos (Ali El-Sappagh, Ahmed Hendawi, & El Bastawissy, 2011).

Na segunda etapa deste processo são implementados procedimentos de transformação dos dados recolhidos, que incluem tarefas de limpeza e conformação. A limpeza dos dados consiste em efetuar correções de erros e omissões nos dados recolhidos, de forma aumentar a sua consistência. Por outro lado, a conformação dos dados passa por resolver os conflitos de *labeling* entre fontes de dados potencialmente compatíveis para que possam ser utilizados em conjunto no repositório de destino (Kimball & Caserta, 2011).

Na terceira etapa do processo ETL é feita a inserção de dados no repositório de destino. Estes dados podem ser novos dados que precisam de ser inseridos no repositório de destino, ou podem ser atualizações dos dados existentes neste (Kimball & Caserta, 2011).

2.4. FERRAMENTAS

As tarefas relatadas neste documento foram realizadas com o apoio de diferentes ferramentas tecnológicas. A primeira ferramenta aqui apresentada, o *Python*, é a linguagem de programação usada

para desenvolver a automatização de todos os processos desenvolvidos pela GA, incluindo todos os programas (scripts) relacionados com a extração, processamento e armazenamento de dados. Esta linguagem é amplamente utilizada devido ao facto de ser bastante prática e legível, e é bastante popular como ferramenta de análise de dados.

Por outro lado, o armazenamento de todos os dados relativos a todos os processos dentro da GA é feito com recurso ao *Microsoft Azure SQL Database [MS Azure]*, que é um serviço que permite armazenar dados na *Microsoft Cloud*. A comunicação com bases de dados foi feita com recurso à linguagem de programação *SQL*, a linguagem padrão para este efeito.

Por último, a plataforma *Apache Airflow* foi usada para fazer a automatização de tarefas e de fluxos de trabalho, sendo esta baseada na linguagem de programação *Python*, e tendo sido usada para efetuar o agendamento periódico dos scripts realizados.

3. METODOLOGIA

Este capítulo tem como foco principal relatar todos os métodos e processos que foram concebidos e implementados ao longo do estágio. Para tal, é feito o enquadramento do estágio no âmbito do funcionamento das equipas de gestores de ativos e de *analytics*, tendo em conta as suas necessidades e a forma como se relacionam. De seguida, é apresentado um caso de estudo de uma integração de dados, sendo feita uma descrição detalhada dos dados recolhidos, assim como da implementação da *pipeline ETL* correspondente.

3.1. ENQUADRAMENTO DO ESTÁGIO

Existe uma vasta gama de ativos geridos na área de GA, sendo que para este estágio os únicos ativos que é relevante mencionar são os ativos de solar FV, nomeadamente os sistemas (instalações) FV ligados à rede. Como detalhado anteriormente [Cap. 2], os principais componentes de um sistema deste tipo são os módulos FV e o inversor, sendo este último o componente que recolhe os dados do sistema que serão apresentados neste capítulo.

As instalações FV geridas pela equipa de GA contêm inversores string de duas marcas diferentes, da SMA e da Huawei, e a monitorização e configuração destes inversores é feita a partir de portais disponibilizados por estas marcas, respetivamente o *Sunnyportal* e o *Fusion Solar*. Na área de GA, é a equipa de Gestores de Ativos que é responsável pela gestão e atualização dos dados que provêm destes portais.

Para além de gerir a informação proveniente dos portais, a equipa de Gestores de Ativos também é responsável por garantir a qualidade dos dados de ativos relativos a operações e manutenção [O&M] presentes na *Salesforce*, que é uma plataforma que oferece soluções de *customer relationship management* [CRM].

Por último, para além destas responsabilidades, os gestores de ativos têm também a função de analisar a rentabilidade dos ativos que gerem, identificar oportunidades de melhoria na cadeia de valor da empresa, e comunicá-las à macroestrutura da empresa. Para realizar estas tarefas, os gestores de ativos usam o *Power BI*, que é uma ferramenta de *reporting* da *Microsoft*.

Relativamente à integração dos dados usados pela equipa de Gestores de Ativos para efetuar as suas tarefas, esta é da responsabilidade da equipa de *Analytics*. Os dados provenientes desta integração são fornecidos aos relatórios realizados em *Power BI*, e, como tal, são processos de extrema importância no suporte que a equipa de *analytics* dá à equipa de Gestores de Ativos.

É neste contexto que surge a motivação para este estágio, que consiste na integração dos dados dos inversores da marca SMA. Esta integração foi possível devido à aquisição da *SMA Monitoring API* pela GA, um *software* que permite aceder a informação que está presente no portal *Sunnyportal*, contendo esta dados técnicos de sistemas FV e de inversores, dados referentes a registos de erro, e dados de leituras de energia e potência, entre outros. Relativamente aos dados disponibilizados por esta API, é importante referir que dados referentes a registos de erros, assim como dados de leituras de energia e potência foram integrados por outros membros da equipa de *analytics*, não sendo esta integração aqui relatada. Por último, é também importante referir que a integração dos dados de inversores da

Huawei já tinha sido implementada antes do início deste estágio, e, como tal, o foco deste relatório é apenas nos inversores da *SMA*.

3.2. CASO DE ESTUDO

3.2.1. Entendimento dos Dados

As tarefas efetuadas nesta etapa do caso de estudo usam como recurso dados de instalações FV e de inversores fornecidos por diversas fontes, de forma a permitir organizar e inserir na base de dados em *MS Azure* todos os dados técnicos dos inversores da marca *SMA*. Para entender os processos que foram efetuadas nesta etapa, é necessário conhecer a natureza dos dados que foram usados, e a forma como estão organizados.

Todos os dados que serão usados neste caso de estudo são referentes a dois ativos distintos: instalações e inversores.

Dados sobre uma instalação existem em três plataformas, no *Sunnyportal*, em *Salesforce*, e em *MS Azure*, estando esta informação em tabelas com nomes distintos. Em *Sunnyportal* a tabela que armazena dados de instalações denomina-se “Planta”, em *Salesforce* denomina-se “Ativo”, e em *MS Azure* denomina-se “Projeto”. Estas tabelas têm as seguintes características:

- A tabela “Planta” em *Sunnyportal* contém os campos ID, nome e descrição. Idealmente, o nome é um campo que contém também o número da proposta, ou o código de ponto de entrega [CPE] associado à instalação, que é o código que a identifica.
- A tabela “Ativo” em *Salesforce* contém um ID, um nome, uma descrição, um CPE, e um tipo. O campo descrição contém o link para a instalação correspondente no portal *Sunnyportal*, sendo que esta informação nem sempre está presente neste campo, em particular em ativos mais recentes. Relativamente ao CPE, este é um código que contém 20 dígitos (duas letras, dezasseis números, e mais duas letras), sendo que em Portugal o CPE começa por PT 00.
- Ainda relativamente à tabela “Ativo” em *Salesforce*, é importante referir que cada “Ativo” corresponde a uma “Proposta”, sendo possível associar estas tabelas a partir de uma terceira designada “Linha de Ativo”. A tabela “Proposta” contém um campo onde está registado o número da proposta, número este composto por oito algarismos, sendo os dois primeiros 00. É de realçar que nem sempre existe uma ligação entre as tabelas “Proposta” e “Ativo”.
- Por último, a tabela “Projeto” em *MS Azure* contém um ID, e o ID do “Ativo” correspondente em *Salesforce*.

Por outro lado, dados sobre um inversor existem em duas plataformas, no *Sunnyportal*, e em *MS Azure*, sendo que neste caso ambas as tabelas correspondentes se denominam “Dispositivo”. Estas tabelas têm as seguintes características:

- No *Sunnyportal*, a tabela “Dispositivo” contém um ID, um nome, um tipo, um modelo, a potência, um número de série, um vendedor, uma plataforma, e um estado.
- Em *MS Azure*, um “Dispositivo” contém um ID, um código, um ID do projeto, um ID do tipo de equipamento, um ID do equipamento, um ID da plataforma, um número de série, um vendedor, e um estado.

- Ainda relativamente à tabela “Dispositivo” em *MS Azure*, esta encontra-se ligada a várias outras tabelas que também contêm informação sobre os dispositivos. A tabela “Equipamento”, contém o modelo e a potência do inversor, assim como o ID do tipo de equipamento. As tabelas “Tipo de Equipamento” e “Plataforma” contêm um ID e um nome.

As tabelas aqui apresentados estão ilustrados na figura 5 que se segue, estando indicadas as fontes e repositórios que lhes deram origem. É importante mencionar que estão apenas a ser representadas as tabelas que permitem entender as tarefas que foram realizadas neste caso de estudo, e para cada uma destas tabelas estão apenas representados os campos considerados essenciais à compreensão dos dados no contexto do mesmo.

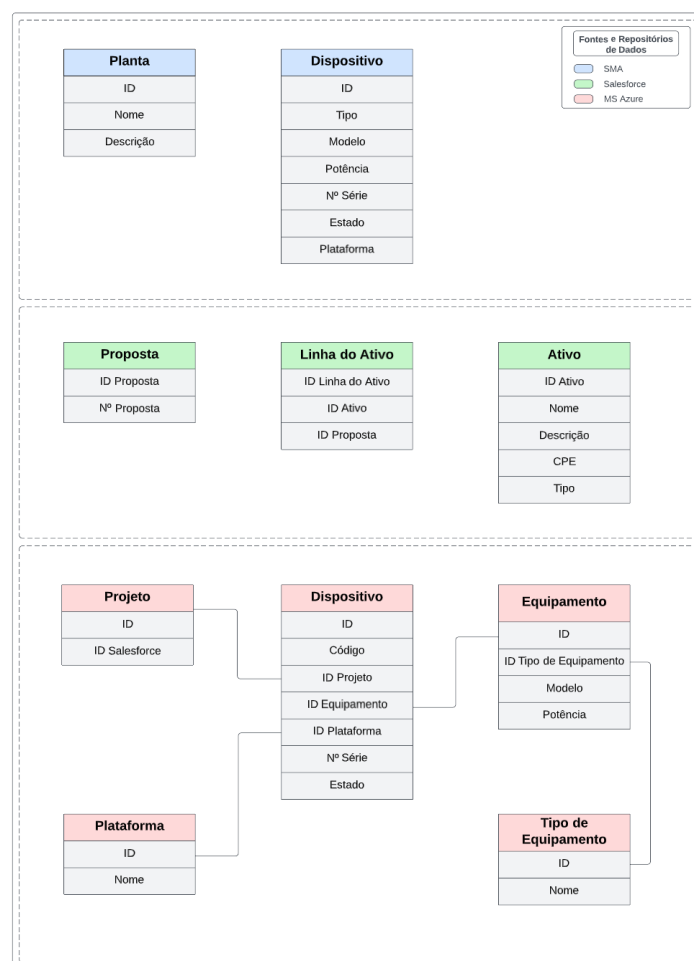


Figura 5: Tabelas de dados de *Sunnyportal*, *Salesforce* e *MS Azure*.

3.3. IMPLEMENTAÇÃO DA PIPELINE ETL

A *pipeline ETL* foi feita implementada na seguinte sequência de processos:

1. Criação de conectores às fontes e repositórios de dados
2. Extração de dados
3. Transformação de dados
4. Inserção de dados
5. Agendamento Automático

3.3.1. Criação de Conectores

A primeira tarefa da implementação da *pipeline ETL* consiste na automatização da comunicação com as fontes e repositórios de dados.

A comunicação com o portal *Sunnyportal* através da *API* da *SMA* foi feita através da estrutura apresentada na figura 6 que se segue.

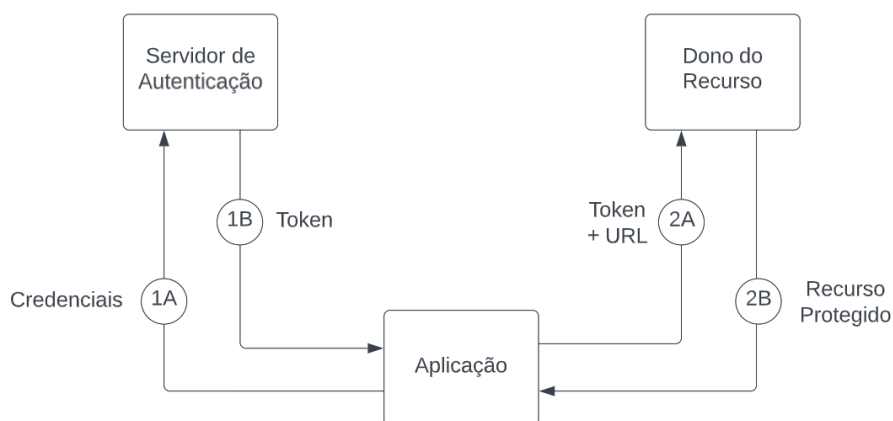


Figura 6: Estrutura de comunicação com a *API* da *SMA*.

Esta estrutura foi implementada de modo a obter um recurso protegido através da autenticação de uma aplicação, tendo este processo sido realizado em duas etapas:

- A primeira etapa consiste no fornecimento das credenciais de acesso (1A) (fornecidas pela *SMA* após a aquisição da *API* por parte da *GA*) ao servidor de autenticação, que devolve um *token* de acesso (1B).
- Na segunda etapa, o *token* obtido juntamente com o *URL* que caracteriza o recurso pretendido são fornecidos ao dono do recurso (2A), que fornece acesso ao recurso protegido requerido (2B).

A automatização deste processo foi feita através da criação de uma função em *Python* designada como conector no contexto deste relatório. Este conector permite a obtenção de dois recursos, o recurso “Planta” e o recurso “Dispositivo”. Para o recurso “Dispositivo”, o *URL* utilizado no processo (2A) inclui a planta para o qual o recurso é pedido.

Em relação à automatização da comunicação com o *Salesforce*, esta é feita através da criação de um conector com recurso à biblioteca *simple-salesforce*, que é uma biblioteca *Python* criada para o propósito de comunicar com esta plataforma. Da mesma forma, para automatizar a comunicação com a base de dados em *MS Azure* foi criado um conector com recurso à biblioteca *SQLAlchemy*.

3.3.2. Extração de Dados

Uma vez criados os conectores, passou-se à fase de extração de dados do processo *ETL*, onde foram extraídos dados de instalações e de inversores do portal *Sunnyportal* (através da *API* da *SMA*), da *Salesforce*, e da base de dados em *MS Azure*.

A primeira tarefa desta fase consistiu em extrair todos os dados do recurso “Planta” do portal *Sunnyportal*, assim como das tabelas “Ativo”, “Linha de Ativo” e “Proposta” de *Salesforce*, e da tabela “Projeto” de *MS Azure*. O conector desenvolvido para comunicar com a *API* da *SMA* permite a extração de todas as instalações do portal *Sunnyportal* num único pedido.

Tendo disponíveis todos os recursos “Planta”, foram feitos pedidos à *API* de todos os recursos “Dispositivo” associados às plantas que estivessem associadas a um “Projeto” em *MS Azure*. Esta associação consiste num “mapeamento” entre o ID da Planta e o ID do Projeto e este foi feito através da combinação de três métodos:

Método 1: Número da proposta - Primeiro, é extraído do campo “Nome” da tabela “Planta” o número da proposta, com recurso a uma função que extrai de um string os últimos 8 dígitos e verifica que são algarismos. Através do número da proposta obtido desta função, é possível fazer um *merge* da tabela “Planta” com a tabela “Proposta”. Depois, com o “ID” da tabela “Proposta” podemos fazer um outro *merge* com a tabela “Linha de ativo”, e de seguida fazer um outro *merge* com a tabela “Ativo” através do campo “ID Ativo” da tabela “Linha de Ativo”. Por último, a partir do “ID Salesforce” da tabela “Projeto” é possível fazer um *merge* à tabela “Ativo”. Este processo encontra-se representado na figura 7 que se segue para o exemplo do sistema FV Pena Aventura.

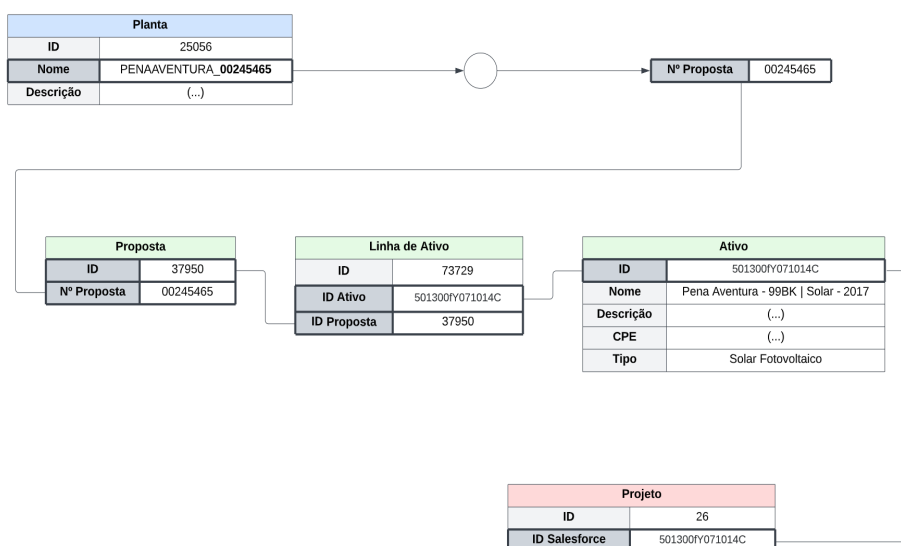


Figura 7: Método 1: Número da proposta.

Método 2: CPE - Este método inicia-se com a obtenção do CPE a partir do campo “Nome” da tabela “Planta”, com recurso a uma função que extrai de um string um padrão 20 dígitos, em que os quatro primeiros dígitos são PT00. Através do CPE obtido desta função, é possível fazer um *merge* da tabela “Planta” à tabela “Ativo”. Por outro lado, a partir do “ID Salesforce” da tabela “Projeto” é possível fazer um *merge* à tabela “Ativo”. Este processo encontra-se representado na figura 8 que se segue para o exemplo do sistema FV A.H.B.V Famalicão.

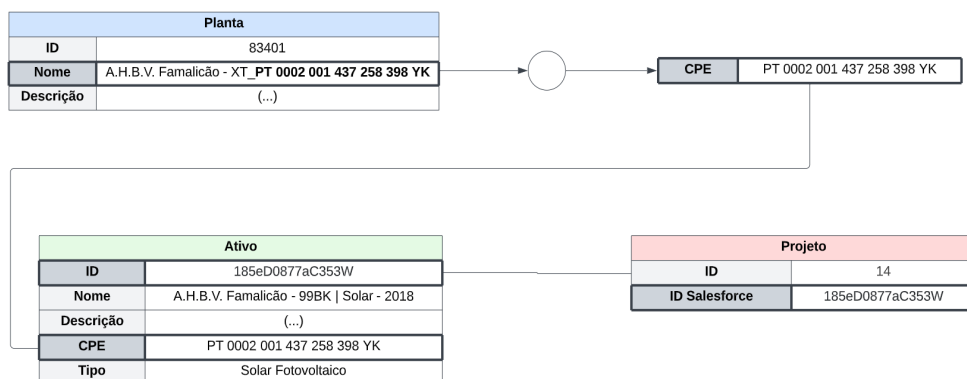


Figura 8: Método 2: CPE.

Método 3: ID da Planta - Este mapeamento inicia-se com a extração do ID da planta a partir do campo “Descrição” da tabela “Ativo”, com recurso a uma função que extrai de um string um padrão numérico. Através do ID da planta obtido é possível fazer um *merge* da tabela “Ativo” com a tabela “Planta” com este valor de ID. Por outro lado, a partir do “ID Salesforce” da tabela “Projeto” é possível fazer um *merge* à tabela “Ativo”. Este processo encontra-se representado na figura 9 que se segue para o exemplo do sistema FV Ivo Cutelarias.

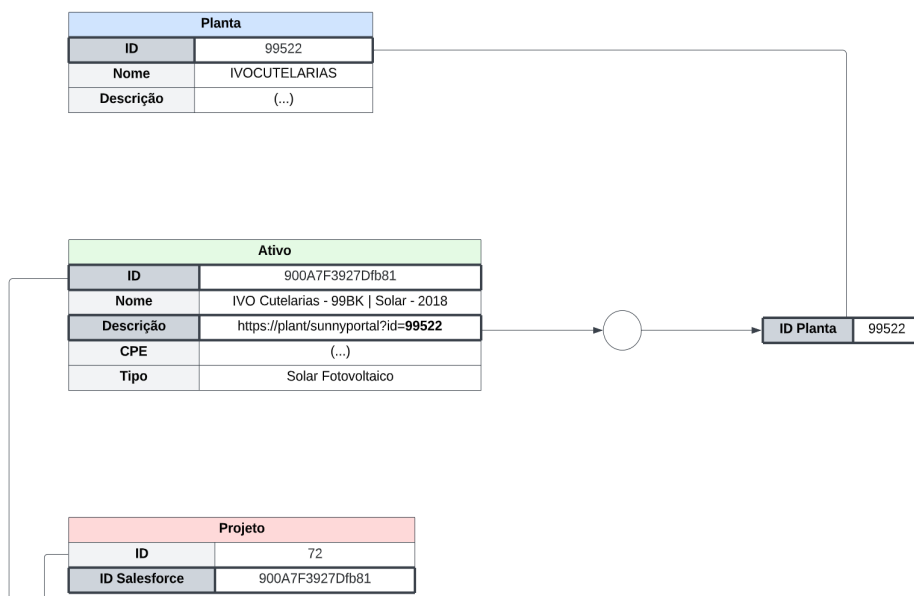


Figura 9: Método 3: ID da Planta.

Em termos de “mapeamento”, primeiro foram realizados os métodos do número da proposta e do CPE, consoante o conteúdo presente no campo “Nome”. Somente para as “Plantas” para o qual as funções implementadas não conseguiram extrair um número de proposta ou CPE do campo “Nome” é que se implementou o método do ID da Planta.

É de realçar que o método do ID da Planta é muito menos eficiente do que os outros. Isto deve-se ao facto deste método necessitar que seja feita a extração de todos os ativos que se encontram em *Salesforce*, em vez de se extraírem apenas os ativos necessários para fazer o merge com a tabela “Planta”. Como tal, foi instruído aos gestores de ativos que a configuração do campo “Nome” do recurso “Planta” no portal Sunnyportal seja feita da forma “nome_número-da-proposta” ou “nome_cpe”.

Note-se que de acordo com a arquitetura de dados de *MS Azure* representada na figura 5, a tabela “Projeto” está ligada à tabela “Dispositivo” através do campo ID do projeto. Tendo esta ligação em conta, segue-se que a inserção de dados na tabela “Dispositivo” vai ser realizada apenas para dispositivos associados a um projeto. Esta associação só é possível se a planta para a qual é feito o pedido do dispositivo através do conector da *SMA* estiver, por sua vez, associada a um projeto. Como tal, só é feito o pedido de dispositivos para os quais este critério se verifica.

Por último, foi feito um pedido pelas tabelas “Dispositivo” (filtrado para dispositivos da marca *SMA*), “Equipamento”, “Tipo de Equipamento”, e “Plataforma” à base de dados em *MS Azure*.

3.3.3. Transformação de Dados

A segunda etapa do processo *ETL* aqui aplicado, a transformação de dados, consistiu na limpeza e conformação dos dados recolhidos na primeira etapa, e esta foi realizada da seguinte forma:

- Converteu-se as unidades da potência dos dispositivos obtidos de KW para W, que são as unidades de potência na base de dados para os inversores da *Huawei*.
- Foram removidos os dispositivos que não tinham valores válidos no campo “Nº Série”, que é o campo que identifica o dispositivo.
- Conformou-se os tipos de dispositivos provenientes da *API* com os tipos de dispositivos presentes na base de dados em *MS Azure*. O tipo “Inversor Solar” proveniente da *API* foi transformado no tipo “Inversor String”.

3.3.4. Inserção de Dados

Finalmente, a terceira etapa do processo *ETL* aqui aplicado, a inserção de dados, consistiu apenas na inserção de novos dados, sendo que foi definido pela equipa de *Analytics* que a atualização de dados seria feita no futuro, quando necessário.

De modo a proceder à inserção de dados de dispositivos, são comparados os seus IDs com o campo “Código” da tabela “Dispositivo” extraída de *MS Azure*. Os dispositivos para os quais não existe uma *match* são inseridos, enquanto os restantes são ignorados. Este procedimento foi semelhante para inserções de dados de dispositivos nas tabelas “Plataforma”, “Tipo de Equipamento”, e “Equipamento”, sendo que a comparação foi feita com o campo “Nome” para as tabelas “Tipo de Equipamento” e “Plataforma” e com o campo “Modelo” para a tabela “Equipamento”.

A figura 10 representa o processo de inserção dos dados de um dispositivo nas diferentes tabelas em *MS Azure*.

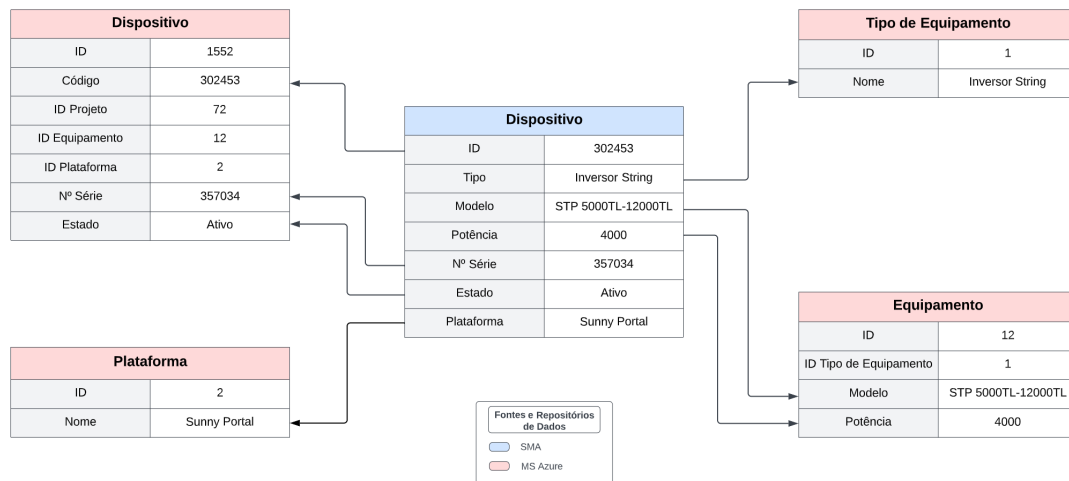


Figura 10: Inserção de um dispositivo em MS Azure.

3.3.5. Agendamento Automático

A *pipeline ETL* foi desenvolvida em dois programas em Python, um programa principal onde foram implementadas a sequência de processos que constituem a pipeline, e um programa secundário que contém todas as funções usadas pelo programa principal. Após o seu desenvolvimento, os programas foram inseridos no *Apache Airflow* de modo a correr o primeiro programa diariamente.

4. CONCLUSÕES

Durante o estágio foi possível aplicar uma variedade de conhecimentos adquiridos ao longo do ano curricular no mestrado de *Data Science* e *Advanced Analytics*. O desenvolvimento das tarefas em código permitiu a expansão do entendimento da programação em *Python*, nomeadamente no que toca ao uso de bibliotecas de análise de dados, ao uso de um modelo de programação orientado a objetos, e ao processo de extração de dados com recurso a uma *API*. Foi também adquirido um conhecimento profundo sobre o processo *ETL*, tendo sido desenvolvidas capacidades de planeamento, design e implementação de pipelines de dados. Por último, foi ganho um entendimento sobre o negócio da energia solar.

Em relação à importância do trabalho aqui desenvolvido para a empresa, é importante referir que a integração dos dados técnicos dos inversores permitiu ao resto da equipa integrar os restantes dados relativos a leituras de potência e energia, e registos de erro recolhidos pelos inversores da *SMA*. Estas integrações correspondem aproximadamente metade dos dados de inversores da EDP, algo que ajuda bastante os gestores de ativos de solar nas suas tarefas diárias.

Em termos de limitações do trabalho realizado, realçam-se a inexistência de um *data lake* na arquitetura de dados e a inexistência de ferramentas de *version control*. A existência de um *data lake* permitiria separar os processos ETL em dois scripts diferentes, sendo que o primeiro script faria a extração dos dados da *API* e inserção em bruto no *data lake*, e o segundo script transformaria os dados do *data lake* e faria a sua inserção na base de dados. Tendo em conta o grande volume de dados e que a comunicação com a *API* e com a base de dados é altamente suscetível a erros, esta separação seria desejável pois permitiria que a fase de extração de dados pudesse ser concretizada independente do sucesso da sua inserção. Em relação à existência de uma ferramenta de *version control*, esta permitiria a recuperação de versões anteriores de código, assim como a sua partilha pelos vários elementos da equipa, que teria facilitado, por exemplo, a coordenação entre programas que comunicam com a *API* da *SMA* através da partilha do conector desenvolvido para este fim.

Por último, este trabalho serve como base para desenvolvimentos futuros no âmbito da criação de soluções de inteligência artificial. Estes podem ser modelos de *Digital Twin*, onde são feitas versões digitais dos sistemas FV; modelos de manutenção preditiva, que permitem prever falhas nos inversores e noutros componentes do sistema; e uma melhor avaliação de desempenho dos ativos da empresa, através da comparação dos valores de energia expectáveis em modelos teóricos de produção, com os valores de energia efetivamente medidos. Estas melhorias irão permitir à EDP continuar a crescer na área de energia solar fotovoltaica, e como tal continuar a ser uma empresa líder no mercado mundial de energia.

5. REFERÊNCIAS

- Rogelj, J., Shindell, D., Jiang, K., Fifita, S., Forster, P., Ginzburg, V., . . . Mundaca, L. (2018). Mitigation Pathways Compatible with 1.5°C in the Context of Sustainable Development. Em *Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change* (pp. 93-174). Cambridge, UK and New York, NY, USA: Cambridge University Press.
- E-Redes. (2021). *Manual de Ligações à rede elétrica de serviço público - Guia técnico e logístico de boas práticas*. Obtido de <https://provedordocliente.e-redes.pt/Files/PDF/Manual-de-Ligacoes-a-Rede.pdf>
- Kimball, R., & Caserta, J. (2011). *The data warehouse etl toolkit: Practical techniques for extracting, cleaning, conforming, and delivering data*. Wiley.
- Ali El-Sappagh, S., Ahmed Hendawi, A., & El Bastawissy, A. (2011). A proposed model for data warehouse ETL processes. *Journal of King Saud University - Computer and Information Sciences*, 23, 91-105.
- Messenger, R., & Ventre, J. (2005). *Photovoltaic Systems Engineering*. Boca Raton London New York Washington, D.C.: CRC PRESS.
- Sherman, R. (2015). *Business Intelligence Guidebook From Data Integration to Analytics*. Waltham, MA,: Elsevier Inc.
- UNFCCC. (2018). The Paris Agreement. *Paris Climate Change Conference*.
- EDP Anual Report. (2021). Obtido em 2022, de <https://www.edp.com/en/investors/investor-information/results-reports#results---reports>
- Gorjian, S., & Shukla, A. (2020). *Photovoltaic Solar Energy Conversion Technologies, Applications and Environmental Impacts*. Academic Press.
- EDPC - Onboarding Gestão de Ativos. (2020).