

SERVIÇO NACIONAL DE APRENDIZAGEM COMERCIAL

SENAC

CURSO DE TECNOLOGIA EM BANCO DE DADOS

**PROJETO INTEGRADOR II: DESENVOLVIMENTO ESTRUTURADO DE
SISTEMAS**

Integrantes do grupo:

RAFAEL DE OLIVEIRA TEODORO

SAMARA DE PAIVA LACERDA

SABRINA MAI NAKAICHI PERDIZ

ANDREWS RIBAS DE AQUINO

RAIANA ROCHA OLIVEIRA

YAN SAVIO SANTOS DE MORAIS FONTES

IGOR VIANA CARVALHO

Professor Orientador: ROSEMEIRE DESIDERIO

EAD - ENSINO À DISTÂNCIA - 2025

Resumo

O presente projeto tem como objetivo desenvolver uma solução para análise e monitoramento da evolução da Covid-19 no Brasil. A proposta inclui a implementação de um processo de Extração, Transformação e Carga (ETL) de dados provenientes de fontes públicas confiáveis, permitindo a centralização e padronização das informações. O resultado será disponibilizado em forma de dashboards interativos, auxiliando gestores públicos e instituições de saúde na tomada de decisões estratégicas, como planejamento de recursos hospitalares e identificação de surtos regionais. Esta primeira etapa descreve a visão geral da solução, define as atividades de apoio decisório aos negócios e detalha o processo de ETL que será adotado.

Palavras-chave: ETL; Covid-19; tomada de decisão; BI.

Sumário

1. Introdução	5
2. Visão Geral da Solução Proposta	5
3. Atividades de Apoio Decisório aos Negócios	5
4. Definição do Processo ETL	6
5. Definição das Tecnologias.....	6
6. Detalhamento Técnico	8
7. Conclusão	11
Referências Bibliográficas.....	12

1. Introdução

A pandemia da Covid-19 trouxe desafios sem precedentes para a saúde pública e a economia global. Durante esse período, gestores precisaram de informações rápidas e precisas para planejar ações de contenção e atendimento hospitalar. No Brasil, os dados referentes à Covid-19 estão disponíveis em diferentes portais, mas muitas vezes de forma descentralizada e sem padronização, o que dificulta análises ágeis.

Este projeto visa integrar esses dados por meio de um processo estruturado de ETL, permitindo análises avançadas e a criação de dashboards que apoiem a tomada de decisão. O projeto está dividido em três etapas, sendo esta a Primeira Etapa, onde é definido o planejamento do processo e o detalhamento da solução.

2. Visão Geral da Solução Proposta

A solução proposta consiste em um sistema de Business Intelligence (BI) voltado à análise de casos de Covid-19 no Brasil.

Objetivos:

- Centralizar dados de casos confirmados, óbitos e vacinação em um único repositório.
- Garantir padronização e qualidade dos dados por meio de processos automatizados de transformação.
- Disponibilizar relatórios e dashboards dinâmicos para gestores da saúde pública.

Benefícios

esperados:

- Tomada de decisão baseada em dados.
- Agilidade no planejamento estratégico.
- Monitoramento em tempo real.
- Prevenção de surtos por meio de análises preditivas.

3. Atividades de Apoio Decisório aos Negócios

Área de Decisão	Atividade Apoiada	Benefício
Hospitais	Planejamento da capacidade de leitos hospitalares	Reducir risco de colapso do sistema
Governo	Alocação de recursos em regiões críticas	Melhor direcionamento de verbas
Gestores estaduais	Identificação de surtos locais	Intervenções rápidas e eficazes
População	Acesso a dados transparentes	Maior conscientização e confiança

4. Definição do Processo ETL

Fase	Descrição	Ferramentas / Formatos
Extração	Coleta de dados de Brasil.io (casos) e Ministério da Saúde (vacinação). Formatos: CSV / JSON. Automatização diária.	Python (requests / pandas), APIs, wget
Transformação	Limpeza, remoção de duplicidades, padronização de nomes e formatos de data, criação de métricas (média móvel, taxa diária).	Python (pandas), scripts ETL
Carga	Carga em Data Warehouse PostgreSQL; tabelas normalizadas; agendamento de atualizações diárias; integração com ferramentas de BI.	PostgreSQL, Airflow/cron, Power BI / Tableau

5. Definição das Tecnologias

Com base na análise de custo e integração, houve uma alteração estratégica:

Banco de Dados:

- Anterior: PostgreSQL.
- Atual: SQL Server, por oferecer melhor custo-benefício, integração com ambientes corporativos e suporte robusto.

ETL:

- Linguagem: Python (bibliotecas pandas, requests).
- Orquestração: Airflow ou agendamento via cron jobs.

Data Warehouse:

- Implementado em SQL Server, com tabelas normalizadas e índices otimizados.

Ferramentas de BI:

- Power BI (principal), com possibilidade de integração com Tableau.

Infraestrutura:

- Servidor local ou nuvem (Azure ou AWS, conforme orçamento).

Segurança:

- Controle de acesso baseado em perfis, criptografia em trânsito (SSL).

Implantação do ETL

O processo será dividido em três etapas:

Extração:

- Coleta diária via APIs e arquivos CSV/JSON.

Transformação:

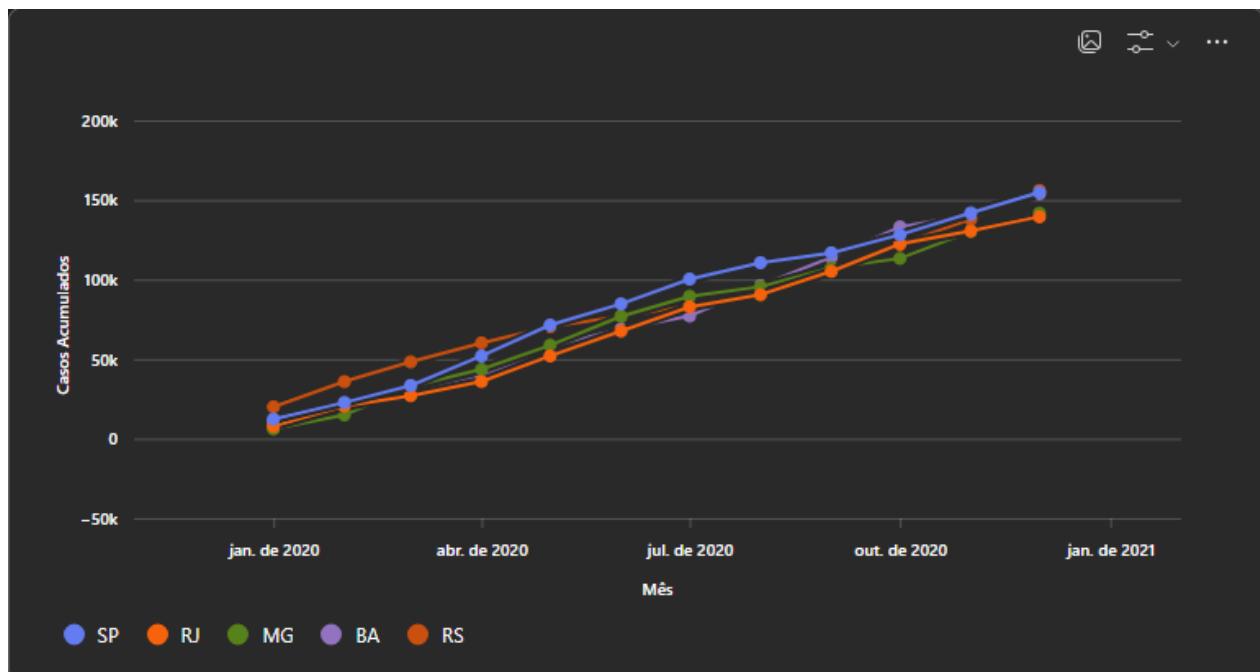
- Limpeza, padronização, cálculo de métricas (média móvel, taxa de crescimento).

Carga:

- Inserção no SQL Server, com agendamento automatizado.

Criação dos Dashboards

Os dashboards serão desenvolvidos no Power BI, com:



6. Detalhamento Técnico

Nesta seção, serão apresentados os aspectos técnicos e operacionais das tecnologias definidas para a implantação da solução, bem como os principais parâmetros necessários para configuração das ferramentas.

Arquitetura Técnica

A solução será composta por três camadas principais:

- Camada de Dados: SQL Server como banco de dados central.
- Camada de Processamento: Scripts ETL em Python para extração, transformação e carga.
- Camada de Visualização: Power BI para criação de dashboards interativos.

Configuração do Banco de Dados (SQL Server)

- Versão: SQL Server 2019 ou superior.
- Parâmetros principais:
 - Modo de autenticação: SQL Server Authentication.
 - Collation: SQL_Latin1_General_CI_AS (compatível com dados em português).
 - Tabelas: Estrutura normalizada com chaves primárias e índices para otimização.
 - Backup: Agendamento diário para garantir integridade.

Configuração do ETL

- Linguagem: Python 3.11.
- Bibliotecas: pandas, requests, sqlalchemy.
- Parâmetros:
 - Extração: APIs Brasil.io e Ministério da Saúde, formato CSV/JSON.
 - Transformação: Padronização de datas, remoção de duplicidades, cálculo de métricas (média móvel, taxa diária).
 - Carga: Conexão via sqlalchemy para inserir dados no SQL Server.

Configuração do Power BI

- Fonte de dados: Conexão direta com SQL Server.
- Parâmetros:
 - Atualização: Agendamento automático (refresh diário).
 - Modelagem: Relacionamentos entre tabelas (fato e dimensões).
 - Segurança: Publicação com controle de acesso por perfil.

Segurança e Governança

- Controle de acesso: Perfis de usuário no SQL Server e Power BI.
- Criptografia: SSL para tráfego entre ETL e banco.
- Logs: Registro de execução dos scripts ETL e acessos ao BI.

Infraestrutura

- Servidor: Windows Server ou Linux com suporte a SQL Server.
- Recursos mínimos:
 - CPU: 4 núcleos
 - RAM: 16 GB
 - Armazenamento: 200 GB SSD

7. Conclusão

Nesta primeira etapa, foram definidos os elementos essenciais para o desenvolvimento do projeto: a escolha de fontes de dados confiáveis, a definição das atividades de apoio decisório e o planejamento detalhado do processo ETL. Isso garante uma base sólida para a implementação técnica nas próximas etapas, visando a criação de uma ferramenta eficiente e útil para análise e tomada de decisão.

Referências Bibliográficas

BRASIL.IO. Dados Covid-19. Disponível em: https://brasil.io/dataset/covid19/caso_full/
MINISTÉRIO DA SAÚDE.Dados de vacinação. Disponível em: <https://www.gov.br/saude/pt-br>
KIMBALL,

KIMBALL, Ralph; ROSS, Margy. **The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling.** 3. ed. S. I: Wiley, 2013.