

Cenário de Validação da Modelagem de Arquitetura - IND07

Explicando o cenário de validação com a aplicação das ferramentas escolhidas

Para validar a arquitetura do sistema de análise de dados no ambiente acadêmico, a modelagem de arquitetura será submetida a um ambiente fictício que representa o Instituto de Informática da Universidade Federal de Goiás.

Cenário de Validação - IND07 - Taxa de Eficiência (TEF)

“Suponha que você seja o diretor acadêmico do Instituto de Informática da Universidade Federal de Goiás que está buscando maneiras de melhorar a eficiência e a tomada de decisões na instituição.

Decide utilizar o sistema de apoio à gestão acadêmica que inclui recursos avançados de BI para tal finalidade.

No início, você define os principais objetivos do sistema de BI como acompanhar o desempenho dos alunos, identificar tendências e padrões de matrícula nas disciplinas, bem como analisar a eficácia dos programas acadêmicos e monitorar a satisfação dos discentes, levando em conta o indicador de educação superior IND07 - Taxa de Eficiência (TEF).

Com o sistema de apoio à gestão acadêmica implementado, você pode acessar facilmente os dados relevantes (e.g. você pode visualizar um painel de controle que mostra o percentual de alunos que se formaram em cada curso até o prazo máximo para integralização, bem como concatenar essa informação junto com o percentual de adesão ao estágio e a média de notas dos discentes naquele curso).

Além disso, você pode usar recursos de análise avançada para identificar tendências (e.g. você pode descobrir que determinados cursos têm um percentual de adesão ao estágio menor e, com base nisso, tomar medidas para aumentar a adesão ou validar se existem fatores impeditivos de ensino que impeçam os discentes de aderir a prática do estágio nesses cursos).

O sistema de BI também permite que você faça previsões com base nos dados históricos (e.g. você pode prever a demanda por determinados cursos no próximo semestre e planejar adequadamente a alocação de recursos).

Além disso, você pode usar o sistema de gestão acadêmico para coletar feedbacks dos alunos e medir a satisfação geral em relação ao curso, disciplinas e metodologias aplicadas

no ensino. Com base nesses dados, é possível identificar áreas de melhoria e implementar ações corretivas.”

Levando em consideração o cenário de validação em relação ao indicador IND07 - Taxa de Eficiência (TEF), que diz respeito ao “Percentual de alunos formados em um determinado curso até o final do prazo de integralização”, aliado à utilização, após tratamento e integração, dos dados com fontes externas e internas, podemos identificar insights por meio dos dados tratados. Com isso, conseguimos responder às principais perguntas levantadas na estruturação do problema como:

- Qual o percentual de alunos que concluíram o curso dentro do prazo de integralização?
- Os alunos com estágio ativo tendem a ser mais aprovados ou reprovados nas disciplinas?
 - Qual a Taxa de Eficiência apresentada por eles?
 - Esses alunos, em comparação aos que não tem estágio ativo, tendem a ter uma taxa maior ou menor?
 - Os alunos que possuem estágio ativo estão se formando dentro do prazo máximo estipulado para integralização do curso?
- Em quais disciplinas os alunos que possuem estágio ativo tendem a ter maior aprovação e menor reprovação?

Em outro ponto de vista, pode-se pensar a respeito dos alunos e também dos coordenadores/professores de cada unidade acadêmica e um ponto que demanda muito esforço e muitas horas de análise é em relação ao que o aluno deve fazer no caso do mesmo estar com problemas com sua grade curricular.

Por exemplo, os alunos que por algum motivo estão fora do fluxo de matérias, muitas vezes ficam perdidos em relação ao que fazer e recorrem aos coordenadores para solucionar tal problema, mas essa solução é bem complexa. Essa situação demanda que os coordenadores tenham uma análise macro e micro da situação atual do aluno, além de analisar o Plano Pedagógico de Curso e vislumbrar os pré-requisitos, complexidade de matérias e vários outros fatores que possam auxiliar o aluno em cada período subsequente para que o mesmo possa ter sucesso ao final do trajeto, aumentando assim, a Taxa de Eficiência e diminuindo as chances de evasão daquele estudante.

Além de disponibilizar informações que possam auxiliar nesse desenvolvimento da grade curricular, seria interessante em etapas posteriores do projeto construir uma ferramenta avançada que consiga fazer esse trabalho de maneira automatizada com base em fluxos curriculares seguidos por outros estudantes com o mesmo perfil.

Complementando o ponto de vista da Taxa de Eficiência e interligando com a sugestão de grade curricular com o aluno, poderia-se utilizar as informações integradas a respeito do aluno para a construção de mais ferramentas de suporte. Nos ensinos básico e médio é bastante comum a utilização de ferramentas de auxílio de estudo e com a integração das informações e com um perfil completo do aluno criado, seria possível se construir ferramentas de estudo voltadas para o ensino superior de forma a auxiliar os alunos e consequentemente aumentando as taxas de eficiência. Essa ferramenta poderia

ser constituída por um gestor de horários, sugestões de assuntos e materiais a serem consultados.

Utilizando os dados extraídos, transformados e/ou normalizados, integrados e disponibilizados, podemos inferir que ao submetermos o cenário de validação na Modelagem de arquitetura proposta, conseguimos validar que esta seja capaz de fornecer uma visão ampla de como seria o funcionamento do sistema após ser implementado e que seria possível atender aos seguintes requisitos:

- Armazenar e processar grandes volumes de dados acadêmicos de forma segura, escalável e eficiente.
- Fornecer análises estatísticas precisas e confiáveis, identificando padrões, tendências e insights relevantes para a área acadêmica.
- Apresentar as informações de forma clara, compreensível e interativa, por meio da plataforma de visualização de dados.
- Ser intuitivo e fácil de usar para os usuários, facilitando a exploração e interpretação dos dados.
- Garantir a segurança dos dados, protegendo-os contra acessos não autorizados e garantindo medidas de privacidade adequadas.

Explicando como as ferramentas são aplicadas em cada da arquitetura

Camada de Extração de dados

Coleta de Dados do nas fontes de dados internas (Planilhas com dados de Estágio), externas (Sissa) e o Banco de dados operacional (SIGAA):

- Utilizar Python com bibliotecas como Beautiful Soup ou Pandas para extrair dados brutos da plataforma do SIGAA, onde extraímos informações sobre a taxa de conclusão anual, informações sobre os alunos e qualquer outra informação útil para a análise são disponibilizadas (i.e. de maneira automatizada utilizando Cronjobs ou manualmente a depender da consulta).
- Seria disponibilizado um modelo de dados criado com base nas necessidades informacionais do sistema (i.e. selecionar quais campos das fontes de dados são de interesse para realizar extração de dados) e seria disponibilizado a instituição que deseje utilizar o sistema uma API para que os dados fossem submetidos no formato correto.
- É possível também importar dados em formato CSV através da biblioteca Pandas da linguagem Python, neste caso, dados relacionados aos estágios dos alunos. Então seria necessário

desenvolver uma interface para que se possa enviar os dados em planilhas pré-formatadas. Outra saída possível, seria a utilização da ferramenta Metabase para o importação desses dados.

Camada de Integração de Dados

Transformação de Dados:

- Utilizar o Python e suas ferramentas para realizar transformações avançadas nos dados, como limpeza, normalização e agregações (e.g. calcular a taxa de conclusão anual com base nos dados brutos coletados).
- Padronização dos formatos é importante, pois convertemos os dados para um formato comum, como CSV ou JSON, para facilitar a integração.
 - Carregar os dados que você deseja padronizar em uma estrutura de dados adequada, como um Data Frame do pandas.
 - Selecionar as colunas a serem padronizadas: identificar as colunas do Data Frame que você deseja padronizar.
 - Criar um objeto Standard Scaler para realizar a padronização dos dados.
 - Ajustar e transforme os dados: ajuste o escalador aos dados selecionados e, em seguida, transforme os dados para a escala padronizada.
 - Atualizar o Data Frame com os dados padronizados: substitua as colunas originais do Data Frame pelos dados padronizados.
 - Salvar os dados padronizados no Operational Data Storage do ambiente de Staging.
- Após isso, carregar no ambiente de Staging (Oracle Database/ MySQL) os dados transformados e/ou processados.

Carregamento no Data Warehouse:

- Utilizar ferramentas ETL como Apache Airflow ou Apache NiFi para orquestrar o fluxo de trabalho ETL.
- Carregar os dados transformados e seus metadados no data warehouse, que pode ser um Oracle Database ou Microsoft SQL Server, otimizado para consultas analíticas.

Camada de Visualização de Dados

Ferramentas de BI:

- Configurar conexões entre o data warehouse e ferramentas de BI, através de uma api com camada de autenticação para garantir a segurança no acesso dos dados por meio do Tableau, Power BI, Data Dashboard orientado ao contexto do LenioLabs ou Metabase.
- As ferramentas orientadas a contexto geralmente fazem a parte de ETL, carregamento e montagem de gráficos e relatórios de forma dinâmica e interna. As chaves primárias e os campos de dados são de grande importância para que o sistema tenha um bom funcionamento. Portanto, é necessário que esses campos estejam padronizados e de fácil entendimento para o sistema. Outro ponto a se destacar, são os feedbacks do usuário. Caso o sistema não consiga relacionar os dados de forma correta e/ou completa, é de grande importância que o usuário forneça um feedback via chat e que faça um complemento do que ele está buscando, fornecendo maiores detalhes e possíveis informações que possam estar correlacionadas. É possível também criar categorias para as informações, ou seja, rotular e caracterizar os Data Lakes.
- Em vista das funcionalidades disponíveis nas ferramentas atuais, uma sugestão seria a criação de uma ferramenta de BI orientada ao contexto da Universidade, utilizando a biblioteca React do Javascript para desenvolver as telas, recomendando-se utilizar o framework NextJS para auxílio do desenvolvimento front-end, sendo assim possível visualizar os dados em formato de dashboard, relatório e painéis interativos.

Processo:

Conectar Ferramentas de BI ao Data Warehouse:

- Configurar conexões para que ferramentas de BI possam acessar dados armazenados no Oracle Database.
- Utilizar as ferramentas de BI para criar consultas e visualizações interativas.
- A utilização de ferramentas orientadas a contexto seria uma grande contribuição para a aplicação como um todo, utilizando um chat para que se possa buscar informações úteis e complementares a respeito de cenários específicos e de forma dinâmica.

Criação de Dashboards:

- Utilizar o Tableau, Power BI ou Data Dashboard LenioLabs para criar dashboards interativos que exibam a taxa de conclusão anual ao longo do tempo, desagregada por região, instituição de ensino, relação entre a taxa de conclusão e como o

trabalho/estágio impacta nesse indicador, ou qualquer outra dimensão relevante.

Análise Exploratória:

- Permitir aos usuários explorar os dados de forma interativa, identificando tendências, padrões e insights relacionados à taxa de conclusão anual.

Integração e Segurança

Segurança:

- Autenticação com tokenização por meio da API que disponibiliza os dados para as ferramentas de visualização de dados (BI).
- Implementar autenticação e autorização usando LDAP/AD (Active Directory) para controlar o acesso aos dados.
- Garantir a segurança de dados em trânsito usando SSL/TLS.

Integração:

- Utilizar Apache Kafka para streaming de dados em tempo real, se houver necessidade de atualizações frequentes nos dados.

Escalabilidade e Tolerância a Falhas

Considerações:

- Configurar o ambiente para escalar horizontalmente conforme necessário, especialmente durante períodos de pico de carga.
- Implementar replicação de dados e monitoramento contínuo para garantir tolerância a falhas.

Monitoramento e Manutenção

Ferramentas:

- Implementar Prometheus e Grafana para monitoramento de métricas e criação de dashboards.
- Utilizar ELK Stack (Logstash, Elasticsearch, Kibana) para análise de logs.

Processo:

- Monitorar o desempenho do sistema, identificar possíveis problemas e realizar ajustes conforme necessário.

- Manter documentação atualizada para futuras referências e treinamento.