

Parte 1 - Solução para gestão da qualidade de dados integrada

1. A solução que adotaria para garantir a padronização e visibilidade adequados em relação à qualidade desses dados;

Para padronização entre as empresas, acredito ser interessante adotarmos regras e melhores práticas comuns, como seguir o framework do DAMA-DMBOK. Além disso, criar um catálogo de dados para que possamos saber quais dados temos disponíveis e onde encontrá-los.

Considerando que cada empresa possui uma plataforma de cloud diferente, podemos utilizar o Databricks como uma plataforma centralizadora para acessar e gerenciar esses dados. O Databricks nos permitirá integrar e processar dados de diferentes fontes e utilizando diferentes linguagens de programação como Python, SQL e Pyspark. No quesito visibilidade, definiríamos indicadores claros para os dados que precisamos gerenciar. Geraríamos bases de dados que alimentariam relatórios. Esses relatórios seriam exibidos em dashboards dinâmicos no Power BI em periodicidades conforme a necessidade e criticidade do dado, com possibilidades de filtros e extrações das inconsistências.

Dessa forma, todas as empresas teriam acesso a esses dashboards e poderiam acompanhar suas performances de maneira transparente e eficiente.

2. Os processos de qualidade de dados que implementaria e como, assim como os papéis e responsabilidades envolvidos nestes processos;

Olhando para essa parte prática, entendo que poderíamos começar analisando os dados, focando em aspectos como completude, padronização, consistência, duplicidade e atualidade. Como estamos lidando com dados regulados, buscaríamos em resoluções, normas e padrões as regras do negócio e as aplicaríamos em nossas bases de dados através de análises utilizando linguagens de programação.

Identificaríamos a causa raiz das inconsistências para acionar a área de negócio responsável pelo dado, permitindo que verifiquem seus processos para possíveis melhorias, como travas sistêmicas ou tratamento direto na fonte do dado, evitando novos problemas. Caso a correção de algumas inconsistências fique sob nossa responsabilidade, já aplicaríamos as correções necessárias.

Por fim, criaríamos um monitoramento contínuo através de relatórios para acompanhar e reportar aos stakeholders a situação das inconsistências e as evoluções nas correções. Realizaríamos reuniões periódicas para discutir os problemas e as ações corretivas.

Papéis e Responsabilidades

Squad de Qualidade de Dados: Realizar análises de dados, identificando inconsistências, monitorando a performance da qualidade, gerando relatórios e comunicando as áreas responsáveis pelos dados. Garantir que todos os stakeholders estejam cientes dos problemas de qualidade de dados.

Áreas de Negócio: Receber os relatórios de inconsistências e tomar as ações corretivas necessárias. Colaborar com o analista de qualidade de dados para melhorar continuamente a qualidade dos dados.

Equipe de TI: Implementar as soluções técnicas necessárias para possibilitar as análises, monitoramento e geração de relatórios. Realizar melhorias sistêmicas para mitigar a entrada de dados incorretos.

Gestão e Liderança: Promover a cultura de qualidade de dados dentro da organização. Apoiar na remoção de possíveis obstáculos na implementação das análises e correções sistêmicas.

3. As tecnologias necessárias;

Plataformas de cloud: Como AWS, Azure, Google Cloud.

Databricks: Para centralizar o acesso e gerenciamento dos dados de diferentes plataformas de cloud e para processamento de dados.

Linguagens de Programação: Python, SQL, PySpark para desenvolver scripts de análise e ETL.

Jira: Para gerenciar tarefas e acompanhar o progresso das atividades de execução, tanto na parte de desenvolvimento de análises quanto nas de correção dos dados.

Confluence: Para documentar processos, procedimentos e resultados das atividades.

Power BI: Para criar as visualizações dos dashboards dinâmicos, gráficos, painéis e indicadores.

Microsoft Teams/ Outlook: Para comunicação e colaboração entre as equipes.

4. A estratégia de engajamento com as áreas envolvidas.

Para garantir o engajamento de todas as áreas envolvidas na melhoria da qualidade dos dados, acredito que seria interessante incluir um indicador de qualidade de dados nas metas globais da empresa. Esse indicador impactaria diretamente todas as áreas de negócio, incentivando-as a se comprometerem com a qualidade dos dados para atingirmos nossas metas.

Além disso, seria importante realizar workshops e treinamentos para orientar as equipes sobre a importância da qualidade dos dados e o impacto que ela traz nos processos e resultados financeiros da empresa. Esses workshops e treinamentos ajudariam a disseminar as regras utilizadas na qualidade dos dados e a ensinar como fazer os inputs corretos. Isso conscientizaria todos sobre a relevância de manter dados precisos e consistentes, garantindo que todos os colaboradores estejam alinhados com as melhores práticas de qualidade de dados.

Por fim, manteria uma comunicação contínua com as áreas, divulgando nossos relatórios de qualidade de dados, explicando os números contidos neles e coletando feedbacks para melhorar nossos processos de análise e comunicação. Isso ajudaria a manter todos informados sobre o estado atual da qualidade dos dados e garantiria que estamos sempre aprimorando nossas práticas para atender às necessidades do negócio.