# Catálogo de dados de Código Aberto

# 

Está cada vez mais comum ver empresas e organizações procurando profissionais com experiência em estrutura de dados para fazerem parte da revolução e mudança cultural da empresa para cultura DATA DRIVEN.

Organizações que já nasceram orientadas a dados, fazem uso de soluções de catálogo de dados. Os catálogos de dados ajudam a criar um ambiente único onde todos os dados da organização e seus contextos podem ser acessados. Isso reduz o tempo para obter insights e chegar rapidamente a decisões de negócios, informados por dados de qualidade.

O que antes era uma tarefa morosa e complexa, devido a necessidade de criar do zero de uma ferramenta de descoberta, classificação, catalogação e apresentação dos dados, hoje é um processo mais fácil de se realizar graças as organizações que abriram suas soluções como código aberto para que a comunidade fizesse uso e ajudasse a aprimorar... Não quero dizer com isso que não tenha um trabalho para a incorporação da solução no sistema e na cultura da empresa, apenas que agora não temos que codar do zero uma solução.

Existe bastante ferramentas open-source, mas destaco aqui apenas 2 soluções. Escolhi essas duas porque são as que já fiz uso; que mais gostei pela praticidade e facilidade de implantação; por ser altamente costumizaveis para necessidades específicas, e para que o caro leitor não tenha muitas opções, pois se forem como eu (que estou trabalhando pra melhorar isso...) quanto mais indicações tenho de opções, mais gasto tempo estudando essas diversas indicações, o que me demanda um tempo precioso que podia ser melhor usado pondo em prática.

# [Apache Atlas](https://atlas.apache.org/#/)

# 

**O Apache Atlas é uma ferramenta de gerenciamento de metadados e plataforma de governança de código aberto que foi incubada pela Hortonworks sob a égide da Data Governance Initiative**.

Principais recursos do Apache Atlas:

* **Classificação:** oferece a capacidade de classificar automaticamente para PII, dados confidenciais e outros. Os ativos de dados podem ser associados a várias classificações. As políticas também se propagam através da linhagem, garantindo assim que os dados derivados herdam a mesma classificação e controles de segurança.
* **Tipos:** ‘Tipo’ é uma definição de como determinados tipos de objetos de metadados são armazenados e acessados ​​no Atlas. Isso permite que os administradores de dados definam metadados técnicos e comerciais.
* **Pesquisa:** A IU intuitiva permite a exploração pré-definida e AD-HOC de tipos de dados por tipo, classificação, valor de atributo ou texto livre. Mantém um histórico de como uma fonte de dados ou dados explícitos foram construídos e como eles evoluíram ao longo do tempo.
* **Segurança:** Apache Atlas é uma ferramenta de governança de dados. Permite segurança granular para acesso a metadados, permitindo configurar controles de acesso a instâncias de entidade e configurar operações como adicionar/atualizar/remover classificações.

**Recursos do Apache Atlas:** Apache Atlas Overview, Apache Atlas Demo, Apache Altas GitHub Repository

# [Amundsen Lyft](https://www.amundsen.io/amundsen/)

# 

**Amundsen é uma plataforma de catálogo de dados de código aberto que foi originalmente construída pela equipe de engenharia da Lyft.**

Desfruta de uma comunidade coesa de contribuidores e usuários. É amplamente adotado por organizações que promovem iniciativas de democratização de dados, governança e serviços de metadados.

Principais recursos da Amundsen:

* **Fácil descoberta de dados:** Ajuda a encontrar dados em diversas fontes por meio de uma simples pesquisa de texto. Os resultados da pesquisa mostram até metadados em linha.
* **Metadados automatizados e com curadoria:** Descrição detalhada dos metadados e seu comportamento, que são selecionados manualmente e gerados automaticamente.
* **Compartilhamento:** Atualização das descrições dos ativos de dados, reduzindo assim o retrabalho de pesquisa em um ativo específico.
* **Aprendizado e compreensão de uso:** Verificação facilitada dos ativos de dados mais usados, possuídos ou marcados como favoritos com maior frequência.

**Recursos Amundsen:** Visão geral do Amundsen, Demonstração do Amundsen, Guia de configuração do Amundsen, Repositório Amundsen GitHub