

Governança de Dados

Com a mudança nos últimos tempos, o mundo vem acumulando mais e mais dados, a cada dia que se passa é necessária a administração abrangente de todos os dados, e para isso ser feito com eficiência utilizamos a governança de dados. De acordo com pesquisas, as ferramentas de governança de dados Apache Atlas, Collibra e Amundsen são as melhores e mais conhecidas neste momento para soluções de ferramentas de catálogo de dados para governança, gerenciamento e análise de dados em toda empresa de forma eficaz e eficiente.

Catálogo de dados significa coletar, organizar e controlar dados de operações, as ferramentas de catálogo de dados auxiliam na administração de dados, gerando economia de custos, mitigação de fraudes e riscos, eficiência e ganho de vantagem competitiva

Collibra

Uma empresa da qual se descreve como uma "empresa de inteligência de dados". Sua plataforma chamada DGC inclui catálogo de dados, glossário de negócios, gerenciamento de mordomia, gerenciamento de dados de referência e mais ferramentas de governança de dados, incorporando ferramentas de automação e colaboração que ajudam a melhorar a produtividade. Faz parte do Data Intelligence Cloud e se integra com outros produtos da Collibra. Alguns de seus clientes são JPMorgan Chase & CO, Credit Suisse, AXA que oferecem serviços financeiros.

Recursos:

- Flexível
- Oferece excelente suporte ao cliente
- Catálogo de Dados
- Glossário de negócios
- Dados de Referência
- Gerenciamento de Políticas
- Curadoria

O DGC da Collibra, lidera há anos o quadrante da Gartner em gerenciamento de metadados, por possuir recursos de Governança de Dados. Ele incorpora ferramentas de automação e colaboração que ajudam a melhorar a produtividade.

Apache Atlas

Apache Atlas é a solução de código aberto que continua tendo desenvolvimento estável ano a ano com contribuições de organizações como Hortonworks, Aetna, Merck e Target. Uma ferramenta completa para governança de dados e gerenciamento de metadados, permitindo que as empresas atendam seus requisitos de conformidade no Hadoop de forma eficaz permitindo a integração com todo ecossistema de dados corporativos. O foco inicial era o

ambiente Apache Hadoop, entretanto Apache Atlas consegue trabalhar em outras plataformas além da Hadoop.

Recursos:

- Classificação de Dados
- Pesquisa e linhagem
- Mecanismo de segurança e política
- Auditoria centralizada
- Troca de Metadados
- Modelagem de dados ágil
- Linhagem de dados

Amundsen

Ferramenta de descoberta de dados e metadados utilizada para melhorar a produtividade de analistas, cientistas e engenheiros para interagir com dados, indexando recursos de dados e alimentando uma pesquisa de estilo de classificação com base em padrões usuais. Amundsen é hospedado pela LF AI & Data Foundation. Sua lista de clientes é composta por Asana Grupo DHI, Instacart, Lyft, Remitly e outros mais.

Recursos:

- Pesquisa de dados através de palavras-chave
- Portal interface do usuário web
- Serviço de metadados
- Biblioteca de ingestão de dados
- Página de detalhes da tabela de dados
- Diretrizes do desenvolvedor para construção em ambiente local

Comparativo de Recursos

	Collibra	Apache Atlas	Amundsen
Classificação	Somente catálogo de dados	Somente catálogo de dados	Somente catálogo de dados
Otimizado para	Empreendimento	Empreendimento	Mercado intermediário
Principais casos de uso	Conformidade	Conformidade	Produtividade dos Analistas
Busca por palavra-chave	Sim	Sim	Sim
Qualidade de dados	Sim	Não	Não

	Collibra	Apache Atlas	Amundsen
Linhagem de Mesa	Sim	Sim	Não
Linguagem de coluna [WIP]	Não	Não	Não
Histórico de consultas	Não	Não	Não
Governança e Segurança	Sim	Não	Não
Indexação de ferramentas de BI	Não	Não	Sim
Integrações	Bearer Google Cloud Data Catalog IRI Voracity IRI FieldShield OpenLegacy	Azure Data Lake Beats Hive data Amazon Web Services Hackolade Spark	AWS Glue Amazon Redshift Apache Cassandra Apache Druid Apache Hive Google Cloud Platform

Conclusões

Ferramenta de dados Apache Atlas

Ferramenta gratuita onde o designer de código aberto do Apache Atlas tornou-se um componente fundamental da plataforma de dados auxiliando as organizações através do uso de metadados simplificando operações de análise de dados em toda organização para tomada de decisão e adicionando entidades aos metadados para facilitar a pesquisa de dados.

Com sua arquitetura pré-construída, pode-se criar processos de desenvolvimento para novos casos de uso, trazendo melhorias de análise de governança de dados onde os usuários podem definir novos tipos de metadados e compartilhar metadados facilmente com outras ferramentas por meio de metadados comuns.

Engenheiros de grandes empresas estão trabalhando juntos para ajudar a garantir que o Atlas seja construído para resolver problemas reais de governança de dados de uma área ampla em todos os setores, sendo assim mostra ser uma ferramenta que não se tornará obsoleta.

A interação com o Apache Ranger é primordial para a segurança da ferramenta Apache Atlas, permitindo a autorização e mascaramento de dados ao acessar os dados, através da classificação dinâmica dos mesmos.

Sendo assim, a conclusão que temos é que a Apache Atlas é uma ferramenta de governança de dados completa, da qual irá acompanhar o mercado de trabalho através de

suas necessidades, procurando se desenvolver de acordo com suas demandas, e oferecendo fácil acesso e gratuito ao usuário.

Referências

<https://www.collibra.com/us/en/platform/data-governance#customer-stories>

https://www.linkedin.com/pulse/melhores-ferramentas-e-software-de-governan%C3%A7a-dos-2021-asakura/?trk=pulse-article_more-articles_related-content-card&originalSubdomain=pt

<https://abracd.org/ferramentas-de-gerenciamento-de-dados-qual-escolher/>

<https://www.softwaretestinghelp.com/data-governance-tools/>

<https://www.amundsen.io/>

<https://www.notion.so/Catalog-of-Catalogs-4bcbee621de243b6a34deaebd28180d0>

<https://sourceforge.net/software/compare/Amundsen-vs-Apache-Atlas-vs-Collibra/>

<https://hevodata.com/learn/data-catalog-tools/>

<https://atlas.apache.org/#/>