

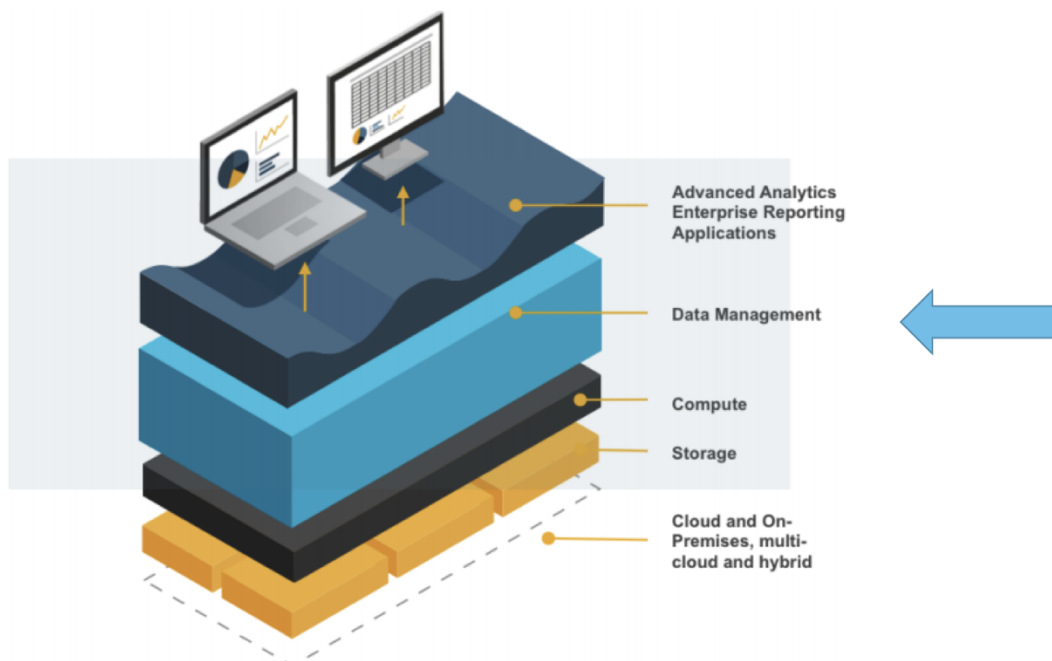


Data Science Academy

www.datascienceacademy.com.br

Data Lake – Design, Projeto e Integração

Gestão e Governança do Data Lake



Neste componente do Design de um Data Lake, as empresas precisam de ferramentas para ingerir e gerenciar seus dados através de várias camadas de armazenamento e processamento, mantendo o caminho claro dos dados em todo o seu ciclo de vida. Isso não apenas fornece uma maneira eficiente e rápida de obter insights, mas também permite que as empresas atendam aos seus requisitos regulamentares em relação à privacidade de dados, segurança e governança.

Os Data Lakes criados com uma estrutura integrada de gerenciamento de dados podem eliminar o complicado processo de preparação de dados do ETL que o Data Warehouse tradicional requer. Os dados são facilmente ingeridos no Data Lake, onde é gerido utilizando **tags de metadados** que ajudam a localizar e se conectar a informação quando os usuários de negócios precisam. Esta abordagem libera analistas para a importante tarefa de encontrar valor nos dados sem o envolvimento de TI em cada etapa do processo, conservando-se assim recursos. Hoje, todos os departamentos de TI estão sendo obrigados a fazer mais com menos. Em tais ambientes, os Data Lakes bem gerenciados ajudam as organizações a utilizar de forma mais eficiente todos os seus dados para obter insights de negócios e tomar boas decisões.

A governança de dados é extremamente importante e as organizações precisam de recursos mais avançados para assegurar que os usuários de negócios



e Cientistas de Dados podem rastrear a linhagem de dados e acesso a dados e tirar proveito de metadados comuns para utilizar totalmente os recursos de dados corporativos.

A chave para uma estratégia sólida de gerenciamento e governança de dados é ter a estrutura correta de gerenciamento de metadados. Com metadados precisos e descritivos, você pode definir políticas e padrões para gerenciar e usar dados. Por exemplo, você pode criar políticas que reforçam a capacidade dos usuários para adquirir dados de determinados lugares, que esses usuários, em seguida, possuem e são, portanto, responsáveis; quais usuários podem acessar os dados; como os dados podem ser usados e como são protegidos, incluindo como são armazenados, arquivados e armazenados em backup.

A estratégia de governança também deve especificar como os dados serão auditados para garantir que você está cumprindo com as regulamentações governamentais que se aplicam a indústria (às vezes em escala internacional, como o European Union's General Data Protection Regulation [GDPR]). Isso pode ser complicado de controlar, quando diversos conjuntos de dados são combinados e transformados. Tudo isso é possível se você implantar uma plataforma de dados robusta para gerenciamento que fornece os metadados técnicos, operacionais e comerciais necessários.