 [Governança de Dados em Data Lake](https://github.com/elainefabiola/datalake)

* [Como promover a Governança em um Data Lake](https://github.com/elainefabiola/datalake)

**Governança Corporativa**

A Governança é um tema da alta administração das empresas maduras, e na maioria das vezes é uma exigência dos acionistas, e a sua implantação tem como objetivo estratégico a obtenção de melhores resultados nos negócios, através da definição, medição e divulgação das de grandes metas da organização.

Podemos definir governança como um conjunto de processos e procedimentos para atingir os objetivos estratégicos de uma organização, ou também, como uma prática de planejar, controlar, medir e divulgar os resultados da estratégia das empresas.

Deve-se implantar a governança para atingir os objetivos estratégicos de uma organização, para isso é necessário definir e acompanhar a implantação de programas alinhados a estratégia para obtenção de melhores resultados.

Para que haja um forte engajamento das pessoas com a governança, é importante estabelecer os responsáveis para cada atividade, os fóruns para definição, estudo e de tomada de decisão.

Estrutura de Governança Corporativa

Interface gráfica do usuário, Texto, Aplicativo

Descrição gerada automaticamente

Fonte: <https://www.itau.com.br/relacoes-com-investidores/Home.aspx?linguagem=pt> – consulta em 28/10/2020

Algumas empresas, com a governança corporativa já estabelecida, e a partir dos conselheiros da administração ou seu Presidente e seus Vices Presidentes, defini as grandes metas de médio e longo prazo, este encontro ou reunião, pode ser definido como *steering comite*, ou comitê executivo, ou ainda comitê estratégico, e a partir desta definição, é possível submeter ao PMO para o detalhamento de um programa (conjunto de projetos ou de esforços ou de ações com objetivo de produzir um bem ou serviço em comum), também é nomeado um patrocinador (uma pessoa ou um grupo de pessoas que fornece recursos e suporte financeiro para o programa), da alta administração.

**O PMO, Project Management Office, ou em português, Escritório de Gerenciamento de Projetos.**

O PMO tem como objetivo desenvolver e fornecer **recursos intelectuais**, processos e procedimentos para definição do **escopo**, planejamento das atividades, **pessoas** e **aquisições** necessárias para execução, a qual fornece um estudo prévio de **prazo** e de entregáveis, critérios de **qualidade** e aceitação, também fornece um **orçamento** prévio para atingimento dos objetivos com **integração** de pessoas, áreas e ferramentas, identificação e monitoramento de **riscos** e a **comunicação** do andamento do programa , performance e resultados medidos por indicadores pré-definidos para as **partes** **envolvidas** nos diversos fóruns comitês e demais canais.

Ao receber uma demanda do comitê executivo, o responsável pelo PMO, seleciona uma pessoa e o apresenta como Gerente do Programa, esta pessoa tem como responsabilidade, entender a demanda e definir, planejar, executar, monitorar e divulgar os resultados do programa.

Para uma melhor contextualização do nosso artigo, vamos descrever o objetivo do programa como “adequação a legislação de proteção dos dados, identificar e corrigir possíveis vazamentos de informação ou acesso indevido, com fins alheios aos interesses da empresa, que possam vir a trazer algum prejuízo financeiro via indenizações ou a imagem da empresa”.

**As áreas de suporte intelectual e operacional, Pessoas, Tecnologia e Jurídico**

Conforme nosso artigo, a área de tecnologia juntamente com a área Jurídica tem papel fundamental no programa, e é a partir das necessidades apresentadas pelo jurídico que os estudos tecnológicos que são realizados, e algumas alternativas são discutidas, sendo necessário estabelecer um grupo multidisciplinar com pessoas capacitadas em suas áreas de conhecimento para administrar os dados da empresa, canais de entrada, transformação e armazenamento, este grupo pode recomendar um estudo de maturidade da empresa em relação aos dados, onde estão e como são protegidos. O grupo pode definir o uso conceitual do DataLake, e neste momento é submetido ao PMO a carta de viabilidade, recursos intelectuais aquisições necessárias, bem como plano de implantação e de investimentos.

**As áreas de monitoramento, Auditoria e Controles Internos**

Como vimos, nossa demanda é para atender a legislação, e como vamos tratar de riscos operacionais e de imagem da organização afim de evitar prejuízos financeiros, as áreas de auditoria e controles internos define como avaliar o programa, e fazem recomendações de políticas e regras de proteção e acesso aos dados, também desenvolvem um checklist de verificação para acompanhamento durante a implantação e após sua operacionalização.

**As áreas de negócios, logística e produção**

Atualmente temos fonte infinita para captura dos dados, que pode ser por um aplicativo com interação humana, IoT, aprendizado de máquina, sinais e transmissores de qualquer natureza, ás áreas de negócios, mais do que nunca estão entrelaçadas com a ás áreas de logística e produção, pois um pedido de manufatura, não necessariamente passa por uma pessoa, as vendas são realizadas por uma coke ou coffee machine, os pedidos são enviados automaticamente para logística que cuida da distribuição, que pode ser no modelo tradicional, usando carros e equipamentos com operador, ou até nos modelos atuais de drones, carros e equipamentos autônomos, responsáveis pelo abastecimentos dos estoques finais, que a partir de métricas de armazenamento, dispara um pedido de produção aos setores de manufatura, e não se trata apenas de um produto, pode ser um serviço ligado a vida, que faz o diagnóstico conforme parâmetros pré-estabelecidos, e envia para um sistema que a partir de indicadores faz análise e recomendações para evitar enfermidades e a até morte, maquinas ligadas a vida monitoram sinais vitais e a partir da dinâmica do aprendizado de máquina, verifica-se a suplementação vitamínica e/ou proteica necessária.

A origem, os canais de captura e o formato dos dados não tem tanta importância em relação ao armazenamento, pois devemos estar preparados para armazenar todo tipo de dado, mas para implementar uma estratégia de proteção e acesso aos dados, as áreas de negócios devem conhecer profundamente qual o tipo de informação que é importante para alavancar os negócios, bem como o tipo de acesso que cada pessoa ou grupo de pessoas devem ter em seus canais de captura e inserção de dados e manipulação de relatórios.

Um pedido de venda originado pela área de negócios, em seus infinitos canais, contém informações que devem ser protegidas, como dados pessoais do comprador, documentos, endereço forma de pagamento, e de suas preferências de compra, produtos, preços, quantidade e intervalos de compras, pois ele passa pelo setor produtivo, onde é feito o planejamento de produção com dados do produto e quantidade, e depois enviados ao cliente; ou, um cliente que abre uma conta bancária por um canal digital, informa seus dados pessoais e patrimoniais, necessita transitar informações financeiras de uma organização para outra, e recebe as suas correspondências físicas; ou, no envio de um pedido de vitamina D para uma farmácia, gerado por um equipamento ligado a vida, em comum, todos passam por um setor de logística e de distribuição, que pode ter muitas pessoas, através de suas empresas, parcerias e terceirizações e envolvidas com a informação, assim, embora tenhamos um ponto de encontro dos dados, eles estão replicados em outras estruturas de dados nos canais de parceria ou de prestação de serviços, sendo necessário uma política completa para a fragmentação da informação, de modo que cada acesso seja monitorado e controlado, uma definição clara de quem pode ver o que, de forma centralizada na origem da demanda, o demandado deve controlar e proteger a informação nos diversos canais, oferecendo acesso seguro e simplificado nas suas relações empresariais.

**Escritório de Investimentos**

Após viabilidade técnica e identificação das áreas envolvidas e suas responsabilidades, o programa é apresentado no escritório de investimentos pelo patrocinador, gerente do programa e os executivos das áreas envolvidas. O escritório de investimento tem uma cadeira permanente no comitê executivo, tem como responsabilidade estabelecer processos e procedimentos para análise de viabilidade financeira, ferramentas e técnicas para definição e controle de orçamento para liberação de recursos financeiros conforme entregas realizadas e o monitoramento do retorno do investimento.

**Governança Data Lake**

Agora que entendemos o que precisa ser feito, quem faz o que e quando e qual o investimento, é desenvolvida a governança para o Data Lake, um dos itens mais importante, e é o fator de sucesso do programa, que após definida, deve ser submetida para aprovação do comitê executivo, que aprovada com pessoas, fóruns e responsabilidades, inicia o trabalho.

Estrutura de governan**ça** de um Data Lake

* [Comitê Executivo](https://github.com/elainefabiola/datalake)

O comitê executivo deve ser formado por executivos das diversas áreas envolvidas com dados, desde a captação pelos diversos canais, pela transformação por infinitos sistemas, aplicativos e máquinas, até o armazenamento, proteção e disponibilização da informação. Mensalmente, delibera e aprova ações e recursos necessários para completude do programa.

Comitê Operacional

Formado por representantes executivos e especialistas de cada área, aprova o detalhamento do programa e os recursos intelectuais, realiza integração entre áreas, identifica riscos (ameaças e oportunidades), define critérios de qualidade para aceitação dos entregáveis, faz o monitoramento e controle e divulgação do avanço do programa, apresenta resultados e pontos de decisão ao Comitê Executivo.

Escritório de Gerenciamento de Dados

Com uma equipe multidisciplinar formada por especialistas de cada área envolvida no programa, atua fortemente em pesquisa e desenvolvimento das melhores práticas para definição dos tipos e formato de dados (metadados), ferramentas para inclusão e manipulação dos dados, formas e técnicas de armazenamento e de proteção da informação.

Cada empresa pode eleger sobre qual aspecto olhar para os dados, um banco por exemplo, deixará a responsabilidade do escritório de gerenciamento dos dados para área de riscos e de capital, onde a partir de modelos matemáticos defini o apetite de risco que a organização quer correr, pois para crescer sua carteira de negócios será necessário um aporte de capital dos acionistas, e este aporte pode ser calculado através da probabilidade de inadimplência da sua carteira, a exposição ao risco no momento da inadimplência e a qual a perda provável no momento da inadimplência, tudo isto é calculado para definir a perda esperada, o principal indicador na formação do preço (juros) que o banco vai cobrar nas suas linhas de crédito; se for uma indústria de transformação, provavelmente a área de produção fica com a responsabilidade, pois a partir dos dados recebidos pela área comercial, dados recebidos pelos fornecedores de matéria prima e a captura das informações das máquinas será possível planejar a produção de forma orquestrada, sem a perda de produção por falta de matéria prima ou produção acima da capacidade de vendas, pois nos dois pode-se gerar prejuízos a empresa; ou uma rede varejista, onde tem grande chance de ficar com a área de marketing, pois a partir dos dados recebidos pelos clientes nos seus diversos canais, é possível direcionar o crescimento da empresa, por produto, região, preço afinidades e até gênero. O importante é que as empresas percebam a necessidade de ter centralizado o gerenciamento dos dados, um lugar para discutir sobre o assunto.

O escritório de gerenciamento de dados pode definir um framework para governança dos dados, hoje existem alguns disponíveis, segundo o autor Barbieri, 2019, em seu livro Governança de Dados, editora Alta Books, os principais são:

* 5W2H, What (o quê), Why (o porquê), Where (onde), When (quando), Who (quem), How (como) e How much (quanto);

IMAGEM

* IBM, viabilizadores, disciplinas centrais e disciplinas de apoio;

IMAGEM

* EDM (Enterprise Data Management Council) e DCAM (Data Management Capability Assessment Model), Ontologia para indústria financeira, assessment e networking

IMAGEM

* Data Management Maturity (DMM) do CMMI Institute, estratégia de gestão dos dados, governança do dados, qualidade dos dados, operação dos dados, plataforma e arquitetura e processos e suporte.
* IMAGEM
* Dama DMBok V2, modelo de alinhamento estratégico, modelo de informações de Amsterdã, composição, evolução, ética, gerencia de mudanças, Big Data e ciências de dados, maturidade em dados
* IMAGEM
* [Política de acesso a dados](https://github.com/elainefabiola/datalake)
* [Captação, transformação e expurgo dos dados](https://github.com/elainefabiola/datalake)

 [Conclusão](https://github.com/elainefabiola/datalake)