

# O alto custo da codificação manual

Por que o gerenciamento de dados na nuvem inteligente,  
automatizado e moderno agora é mais importante do que nunca

### **Sobre a Informatica**

A transformação digital muda expectativas: melhor serviço, entrega mais rápida e com menos custos. As empresas devem mudar para continuar competitivas, e a solução está nos dados.

Como líder mundial em gerenciamento de dados corporativos na nuvem, estamos preparados para ajudar você a ocupar de maneira inteligente uma posição de liderança — em qualquer setor, categoria ou nicho. A Informatica oferece a perspectiva para que você se torne mais ágil, aproveite novas oportunidades de crescimento ou invente coisas novas. Estamos 100% focados em todos os tipos de dados para oferecer a versatilidade de que você precisa para prosperar.

Convidamos você a explorar tudo o que a Informatica tem a oferecer — e estimular o poder dos dados para impulsionar sua próxima revolução com inteligência.

## Índice

Apresentação .....	4
As escolhas que a TI enfrenta .....	4
Os desafios do gerenciamento de dados em um mundo multinuvem e híbrido .....	5
Por que algumas organizações ainda consideram a codificação manual .....	5
Quatro obstáculos da codificação manual para a TI.....	6
Três obstáculos da codificação manual para os negócios.....	7
Quatro elementos de gerenciamento de dados inteligente e automatizado de lakehouse na nuvem .....	8
A soluções de gerenciamento de dados de lakehouse em nuvem da Informatica.....	10
Adote uma abordagem abrangente para um gerenciamento de dados na nuvem moderno, automatizado e inteligente .....	10
Próximas etapas.....	11

"Para gerenciar toda esta complexidade dos dados será necessário um exército de pessoas com codificação manual e planilhas".

— Diretor de gerenciamento e governança de informações corporativas, organização de saúde

## Apresentação

Hoje em dia, as organizações estão investindo muito em novos processos, plataformas e ambientes baseados na nuvem para obter benefícios como escalabilidade, elasticidade, agilidade e eficiência em custos. Ao mesmo tempo, as organizações também compreendem que os dados são fundamentais para a transformação digital bem-sucedida e não podem ser vistos como uma ideia para o futuro.

À medida que essas tendências convergem, os departamentos de TI são incumbidos de ajudar a empresa a se preparar para a nuvem ou adotá-la imediatamente, enquanto modernizam as análises. As empresas estão modernizando ou criando armazéns de dados e data lakes na nuvem, ou algo a que eles começam a se referir como lakehouse. Em uma plataforma de dados na nuvem, você tem uma solução combinada para análise histórica e preditiva.

Entretanto, quando se trata de gerenciar os dados para acelerar a primeira relação tempo/custo e entregar um ROI com investimento em armazéns de dados, data lakes e lakehouses na nuvem, a primeira abordagem típica que os departamentos de TI tendem a assumir pode ter implicações significativas, como aumento nos custos, estouro de orçamento do projeto e complexidade de gerenciamento, eliminando qualquer benefício da modernização da análise na nuvem.

## As escolhas que a TI enfrenta

Considere o caso de uma firma de serviços financeiros que deseja consolidar e modernizar vários armazéns de dados e data lake on-premise em um lakehouse na nuvem como parte de uma iniciativa contínua da experiência do cliente. A primeira iniciativa da empresa é visar clientes que mostraram interesse nos produtos de diferentes categorias on-line, mas nunca fizeram uma compra. Entretanto, há sinais digitais (por exemplo, fluxo de cliques na Web, social, transações, etc.) que podem ajudar a firma a influenciar ou se engajar melhor com seus clientes para aumentar o wallet share.

A TI deve mover todos os dados de bancos de dados e de arquivos on-premise ou dados CRM e ERP para a nuvem e estar pronta para suportar várias análises avançadas e projetos de IA. O caminho mais rápido para o sucesso, e a armadilha em que muitas equipes de TI caem, é montar rapidamente o protótipo de uma solução de integração interna personalizada na nuvem para resolver o primeiro caso de uso comercial. Entretanto, as iniciativas de negócios consistem em vários projetos e fluxos de trabalho e, assim, não terminam com um projeto único e a solução não é escalada para atender aos desafios de gerenciamento de dados complexos de hoje em dia.

Novos tipos de dados (como sociais, arquivos de registro e de sensor de máquina, texto não estruturado, dados de terceiros e assim por diante) têm levado organizações a adotar novas tecnologias como estruturas de processamento de dados com código aberto, como o Apache Spark, para processar e armazenar dados em ambientes de nuvem, como o Amazon Web Services, Microsoft Azure ou Google Cloud Platform. E muitas organizações devem gerenciar uma paisagem complexa e multinuvel. É normal que empresas tenham dois ou mais ambientes de nuvem. A TI enfrenta a possibilidade real de não conseguir integrar sistemas on-premise e aplicativos na nuvem para aproveitar os dados de forma eficiente e eficaz.

"Se você é o responsável pela codificação manual e criação desses scripts para a ingestão direta dos dados, então é você quem nós queremos. A última coisa que desejo é que as equipes de desenvolvimento também estejam de prontidão para ajudá-los. Realmente preciso que o código possa ser reutilizado e mantido".

— Diretor de arquitetura de dados, fabricante

Muitos departamentos de TI precisam tratar de vários desafios novos para esses novos requisitos:

- Como a TI obtém dados em silos para os tomadores de decisão?
- Os gerentes de TI devem confiar em desenvolvedores internos para criar integração de dados personalizados?
- Os líderes de TI devem considerar as ferramentas de integração de dados?

Este white paper pretende oferecer orientação aos líderes de TI para implementar soluções de gerenciamento de dados inteligentes e automatizadas em vez de gastar tempo criando soluções personalizadas codificadas manualmente.

### Os desafios do gerenciamento de dados em um mundo multinuvel e híbrido

À medida que as organizações de TI começam a oferecer suporte às iniciativas de nuvem e de análise ou de IA, a tentação é carregar seus desenvolvedores técnicos com o projeto, desenvolvimento e implantação de uma solução. Entretanto, elas rapidamente se deparam com desafios de dados se decidirem trilhar o caminho da codificação manual. Em muitos casos, essas dificuldades são as mesmas que assombraram os armazéns de dados e data lakes on-premises:

- **Dados variados e em silos:** A maioria das organizações tem muitos tipos de dados residindo em vários sistemas e formatos de armazenamento diferentes, seja on-premises ou na nuvem. Normalmente, os dados estão distribuídos em armazéns de dados em silos, data lakes, aplicativos de nuvem ou recursos de terceiros. Entretanto, cada vez mais dados se originam de sistemas de transações e interações on-line, como arquivos de registro da web e de máquina e mídias sociais. Por exemplo, em um ambiente de varejo, os dados são distribuídos por muitos sistemas diferentes. Eles incluem sistemas de ponto de venda (POS) que contêm dados de transação na loja, dados de clientes em um sistema CRM e MDM, dados de fluxos de cliques na web e em mídias sociais armazenados em um data lake na nuvem e muito mais.
- **A falta de qualidade e de governança de dados:** Com dados variados e em silos, com frequência há graus variados de qualidade e governança de dados. As políticas (se existirem) raramente são aplicadas de maneira consistente. Os dados são despejados em data lakes, produzindo data swamps onde os dados são difíceis de pesquisar, entender, gerenciar, proteger e que restringe o acesso a poucos cientistas de dados. É ainda pior quando dados sujos se infiltram em um armazém de dados na nuvem do qual centenas de analistas comerciais e outros consumidores de dados dependem para tomada de decisão, análise preditiva e IA.
- **Há tecnologias demais que estão emergindo e mudando:** Com mais dados, há novos fornecedores, tecnologias e projetos de código aberto que modificam a paisagem da TI. Há tecnologias obsoletas, novas e em evolução para armazenamento, computação, bancos de dados, aplicativos, análises e, mais recentemente, IA e aprendizado de máquina. Os desenvolvedores podem se esforçar para acompanhar essa paisagem em evolução, dificultando a padronização ou a implementação de uma metodologia.

### Por que algumas organizações ainda consideram a codificação manual

Algumas organizações optam pela codificação manual, acreditando que é um caminho mais fácil do que obter ferramentas de integração de dados, que podem ter uma curva de aprendizado pequena. Além disso, os desenvolvedores podem sentir que as ferramentas de integração restringem sua criatividade para um determinado caso de uso. Na maioria das vezes, essas objeções são pouco ambiciosas quanto a uma solução inteligente e automatizada. Entretanto, a codificação manual pode ser adequada para provas de conceito (POC) rápidas com baixo custo de entrada.

## Quatro obstáculos da codificação manual para a TI

As integrações de dados codificadas manualmente podem atrair os departamentos de TI por parecer uma forma rápida e barata de criar pipelines de dados. Mas há desvantagens significativas a considerar.

### **A codificação manual é cara**

Com o tempo, fica caro implantar, operar e manter a codificação manual na produção. A codificação manual precisa ser ajustada e otimizada desde o desenvolvimento até a implantação. E, já que boa parte dos orçamentos de TI é direcionada a operações e manutenção, o preço da codificação manual aumenta com o tempo. Além disso, o preço de uma solução interna se eleva à medida que a complexidade aumenta na forma de mais origens, mais alvos, mais transformações de dados avançados ou simplesmente agendamento de eventos. Por volta da terceira ou da quarta iteração, a TI percebe que a folha de pagamento sofre um aumento drástico, modificando a economia da solução. E então já é tarde demais, um compromisso substancial foi feito. Ninguém quer explicar aos gerenciantes que os milhões gastos em codificação manual devem ser desperdiçados em favor do novo investimento em uma solução de gerenciamento de dados de um fornecedor, que vai além da ingestão e integração de dados básicas.

### **A codificação manual não resiste ao tempo**

Com tecnologias que se renovam e emergem com rapidez, os desenvolvedores têm que recriar e refazer o código sempre que há mudanças na tecnologia, upgrades ou até mudanças no mecanismo de processamento subjacente. Usávamos o Hadoop até recentemente. Hoje usamos o Spark. Amanhã haverá outra versão do Spark ou até mesmo uma tecnologia melhor. Assim que sua integração de dados codificada manualmente estiver funcionando, ela ficará obsoleta. E mais, uma solução codificada manualmente não atende à realidade multinuvel de hoje em dia. A maioria das organizações terminam com várias nuvens por departamentos ou vários tipos de projetos ou como resultado de uma fusão ou aquisição.

### **A codificação manual carece de automação**

A codificação manual não é escalada para organizações orientadas por dados e não acompanham o ritmo das demandas corporativas. Há solicitações demais por pipelines de integração de dados para as equipes de TI darem conta. A única forma de escalar a entrega de trabalhos de integração de dados de produção é por meio da automação. Isso exige o uso de IA e de aprendizado de máquina.

A codificação manual significa que o desenvolvimento, o teste e o gerenciamento do código de origem são, em sua maior parte, manuais. A documentação é trabalhosa e, normalmente, não acompanha o ritmo. Com a codificação manual, você deve usar ferramentas diferentes para depurar ambientes diferentes, programar, monitorar e esperar que possa se recuperar normalmente de trabalhos com falhas. Além disso, os processos de integração de dados codificados manualmente são limitados nas transformações envolvidas. Por exemplo, para tratar a qualidade de dados, o engenheiro de dados precisará codificar manualmente um processo para padronizar códigos ou limpar endereços de uma região específica.

### **A codificação manual carece de amplitude corporativa**

São necessários muitos anos para que os programadores tradicionais de codificação manual de integração de dados percebam como a qualidade e a governança de dados são importantes para garantir que a empresa tenha dados confiáveis. Hoje, isso é ainda mais importante com as empresas orientadas por dados e a ascensão da IA e do aprendizado de máquina. A codificação manual não consegue oferecer amplitude corporativa para integração de dados, qualidade de dados e gerenciamento de metadados.

## Três obstáculos da codificação manual para os negócios

As limitações da codificação manual não se restringem à TI. No fim das contas, a codificação manual afeta os resultados comerciais. Aqui estão as três principais áreas nas quais a codificação manual pode causar impacto comercial negativo:

### **Custos maiores**

Os custos da mão de obra da codificação manual e da elaboração de scripts podem ser significativos e os recursos especializados são limitados. Esses custos aumentam com o tempo. Apesar de projetos individuais de escopo limitado poderem ser entregues de maneira rápida e barata, essa economia é perdida em escala corporativa. Não há conceito de capacidade de reutilização na codificação manual — às vezes, os engenheiros de dados acabam criando scripts que já existem sem perceber. As capacidades não são repetíveis para esforços futuros. Por exemplo, as técnicas modernas de integração de dados incluem ingestão de dados com várias latências, transformações ETL/ELT complexas, streaming, limpeza e análise de dados. Com uma solução personalizada codificada manualmente, essas técnicas são difíceis de desenvolver e não são reutilizáveis com dados ou projetos mais recentes. Portanto, a empresa de TI precisará treinar de novo os desenvolvedores, o que implica grandes despesas e uso futuro limitado.

Os altos custos de desenvolvimento levam ao alto custo de manutenção. Se for necessária uma modificação no código, os engenheiros de dados precisam determinar o impacto das alterações, reescrever código, testar de novo e reimplantar o código e quaisquer artefatos impactados como os relatórios da BI. Se houver erro no processamento de dados, uma solução que usa codificação manual é difícil de depurar. Como resultado, os desenvolvedores estão presos a seus códigos em um ambiente em constante alteração.

### **Risco maior**

Upgrades frequentes na infraestrutura subjacente exigem uma coordenação complexa. Na codificação manual, você não desenvolve no ritmo da inovação. A qualquer momento em que houver uma alteração na paisagem dos dados (por exemplo, um upgrade da Oracle, uma mudança de um armazém de dados Teradata on-premises para o Amazon Redshift, ou uma mudança de SAP para Salesforce CRM), os desenvolvedores devem parar o que estão fazendo e manter seu código.

Essas dependências aumentam o risco geral quando recursos especializados são limitados. Dependendo das capacidades exigidas, as organizações podem ter um gasto significativo para encontrar e treinar o recurso certo para dar suporte a soluções internas. Em muitas regiões, pode ser difícil encontrar desenvolvedores de codificação manual altamente especializados e proficientes nas tecnologias mais recentes e a concorrência por desenvolvedores especializados é feroz.

Por exemplo, um desenvolvedor especializado em uma empresa farmacêutica global criou uma estrutura de ingestão para obter dados do data lake da empresa aproveitando o Sqoop (um componente de código aberto do Apache Hadoop). O processo era totalmente dependente de uma pessoa para manter e operar a estrutura de ingestão. O desenvolvedor se mudou e não deixou ninguém para manter ou operar a estrutura existente. Nenhum dado novo conseguiu entrar no data lake porque ninguém mais entendia ou conseguia manter o código. A empresa percebeu seu erro e mudou para uma solução de integração de dados corporativos.

"[Com a Informatica], conseguimos fazer atualizações oportunas do sistema disponíveis quando necessário sem codificação manual e conectar facilmente origens separadas em diferentes locais sem limitações relacionadas ao tipo ou ao formato dos dados".

— Diretor sênior de TI  
Varejista viajante

### **Tempo de desenvolvimento mais lento**

As organizações correm o risco de atrasos quando confiam em uma equipe limitada para desenvolver, implementar e manter a solução personalizada interna. Em alguns casos, essas dependências podem impactar significativamente o tempo de desenvolvimento e resultar na perda de oportunidades.

Uma grande empresa de energia tinha um ambiente Hadoop que não estava retornando o valor esperado. Foi constatado que, apesar de procurar por seis meses, o desenvolvedor chefe não conseguiu encontrar recursos especializados para criar os trabalhos Spark exigidos. A empresa decidiu investir em uma ferramenta de integração de dados corporativos para obter mais de seus investimentos e ajudar o desenvolvedor chefe com a lista de pendências de solicitações.

Uma falta de governança, qualidade e visibilidade de dados completos diminui a disponibilidade de dados confiáveis para a empresa para qualquer tipo de iniciativa de transformação digital. Sempre que houver uma falta de confiança ou governança, a empresa aciona os freios, reduzindo a velocidade de adoção do usuário e, assim, atrasando o valor comercial esperado.

### **Quatro elementos de gerenciamento de dados inteligente e automatizado de lakehouse na nuvem**

À medida que as organizações consolidam e modernizam seus armazéns de dados on-premises e data lakes na nuvem ou os criam na nuvem, é mais importante do que nunca evitar as armadilhas da codificação manual. Especialmente hoje, com o surgimento dos lakehouses que prometem o melhor dos armazéns de dados e data lakes com a agilidade e escalabilidade da nuvem, é essencial aproveitar a inteligência e automação orientadas a metadados para criar pipelines de dados eficientes.

Enquanto muitos departamentos de TI se concentram apenas na integração de dados, uma solução mais ampla é necessária para resolver as necessidades das empresas atuais em todo o ciclo de vida do gerenciamento de dados. Aqui estão os principais elementos.

#### **Integração de dados**

Uma solução existente de ótima qualidade é essencial para integração de dados inteligente e automatizada para gerenciar armazéns de dados, data lakes e lakehouses na nuvem. As capacidades a seguir permitem que você crie de forma rápida e eficiente pipelines de dados para alimentar seu armazém de dados, data lake e lakehouse na nuvem:

- Integração sem código com modelos e transformações futuras recomendadas por IA
- Ingestão em massa para arquivos, bancos de dados, dados modificados e streaming
- Otimização de empilhamento para bancos de dados, armazéns de dados na nuvem e lakehouses de PaaS
- Dimensionamento elástico sem servidor
- Processamento baseado em Spark na nuvem
- Conectividade nativa e ampla
- Processamento de fluxo
- Aumento de IA e aprendizado de máquina para lidar com desvio de esquema e análise de arquivos complexos
- Suporte para operações de dados e aprendizado de máquina (DataOps e MLOps)



### **Qualidade de dados**

No mundo de hoje dos lakehouses na nuvem, não é suficiente ter uma integração de dados de classe mundial. Você também precisa da melhor qualidade de dados. Capacidades de qualidade de dados inteligentes e automatizadas garantem que os dados sejam limpos, padronizados, confiáveis e consistentes por toda a empresa. O que procurar:

- Criação de perfil de dados que seja integrada à governança de dados
- Regras de qualidade de dados e geração automatizada de regras
- Dicionários de dados para gerenciar listas de valores
- Processos de limpeza, padronização, análise, verificação e deduplicação/consolidação
- Integração com sua solução de integração de dados
- Análise da qualidade de dados
- Processamento baseado em Spark na nuvem

### **Gerenciamento de metadados**

Uma base de metadados corporativos comuns proporciona visibilidade e linhagem inteligentes, automatizadas e completas por todo seu ambiente. A ampla conectividade de metadados por tipos e origens de dados distintos garante que você tenha visibilidade e possa acessar os dados bloqueados em aplicativos de transação heterogêneos, armazéns e sistemas de dados, aplicativos SaaS e sistemas legados proprietários. Uma base de metadados corporativos comuns proporciona, de forma inteligente e automatizada:

- Descoberta de dados
- Linhagem completa
- Identificação de ativos e compilação de dados
- Compreensão dos metadados técnicos, corporativos, operacionais e de uso
- Conectividade on-premises e na nuvem para bancos de dados (armazém de dados e data lakes), aplicativos, ETL, ferramentas BI e outros

### **Capacidades nativas da nuvem criadas em uma base de IA e aprendizado de máquina**

O quarto elemento é fundamental e é executado por baixo dos outros três. A integração de dados, a qualidade de dados e o gerenciamento de metadados devem ser criados com base em IA e em aprendizado de máquina para manipular o crescimento exponencial nos dados corporativos. Procure uma solução nativa na nuvem que seja multinuvm, orientada a API e baseada em microsserviços e que ofereça as capacidades a seguir:

- Automação baseada em IA/aprendizado de máquina, como próximas melhores recomendações de transformação, similaridade de pipeline de dados, alerta operacional e ajuste automático
- Containerização
- Arquitetura sem servidor
- Instalação e configuração mínimas
- Atualizações automáticas
- Preço baseado em uso
- Certificações confiáveis
- Alta disponibilidade e segurança avançada de pilha completa integradas (por exemplo: plataforma, rede, infraestrutura)

"O Intelligent Data Platform da Informatica nos permite aproveitar o melhor da tecnologia de código aberto para execução ideal e implantação mais rápida. Agora podemos gerenciar nossos dados de maneira sistemática, repetível e sustentável".

— Engenheiro de dados sênior,  
Empresa jurídica

## A soluções de gerenciamento de dados de lakehouse em nuvem da Informatica

O Cloud Lakehouse Data Management da Informatica® é a única solução de gerenciamento de dados de classe corporativa, nativa na nuvem, de ponta a ponta para lakehouses, bem como para armazéns de dados e data lakes na nuvem.

A solução é criada no Intelligent Cloud Services<sup>SM</sup> (IICS) da Informatica, o iPaaS corporativo líder mais avançado no setor. O Cloud Lakehouse Data Management da Informatica combina o melhor da integração de dados, qualidade de dados e gerenciamento de metadados em uma solução completamente automatizada nativa na nuvem com capacidades fundamentais de IA e aprendizado de máquina orientadas a metadados.

A solução trata dos vários desafios de gerenciamento de dados complexos que atingem as empresas que mudaram, ou estão mudando, seus dados para a nuvem. Com ela, você pode:

- Reduzir o TCO aumentando a eficiência em todo o ciclo de vida do desenvolvimento de software (criação, implantação, manutenção, operação).
- Aproveitar a conectividade pronta para centenas de origens de dados para tirar vantagem de integração gráfica e sem código e de transformações integradas complexas
- Eliminar os riscos de usar codificação manual e soluções pontuais limitadas para gerenciamento de dados, ocultando a complexidade e se adaptando às modificações de dados e da tecnologia
- Obter todos os benefícios que as soluções de nuvem como serviço oferecem: escala, agilidade, instalação e configuração mínimas, atualizações automáticas, alta disponibilidade e segurança avançada
- Obter um tempo de desenvolvimento mais rápido ao acelerar seus esforços para migrar os data lakes e armazéns de dados para a nuvem com automação com base em IA/aprendizado de máquina que elimina milhares de tarefas manuais
- Torne suas iniciativas de análises de dados à prova do tempo, desde o ecossistema de nuvem pública em constante alteração, e evite o bloqueio de fornecedores em um mundo multinuvm
- Estabeleça uma infraestrutura moderna de gerenciamento de dados com integração de dados, qualidade de dados e gerenciamento de metadados abrangentes

## Adote uma abordagem abrangente para um gerenciamento de dados na nuvem moderno, automatizado e inteligente

Muitas organizações precisam de dados para realmente entenderem, operarem e expandirem sua empresa, mas a complexidade dos dados as impede. As organizações de TI estão procurando uma solução de gerenciamento de dados automatizada e inteligente que feche a lacuna entre as implantações on-premises e na nuvem sem ter de reconstruir tudo do zero antes que possam obter os benefícios de uma implementação bem-sucedida.

Sem uma plataforma unificada e abrangente de dados, as organizações são forçadas a remendar soluções de pontos distintos que nunca foram projetados para funcionar juntos em primeiro lugar. A integração desses sistemas é demorada, cara, arriscada e difícil de mudar. Se uma solução pontual é alterada, você precisa refazer e testar novamente todas as integrações do sistema.

Adotar uma abordagem corporativa não exige uma implementação totalmente nova. Um dos benefícios do gerenciamento de dados inteligente e automatizado é que as organizações podem implantar metodologias, processos e tecnologias comuns de forma gradual, começando com um ou dois projetos.

Ao utilizar uma plataforma de gerenciamento de dados corporativos de alta produtividade, a TI pode acelerar projetos iniciais para entregar valor comercial imediato. À medida que a organização de TI executa projetos adicionais, ela pode aproveitar e reutilizar ativos existentes, diminuindo significativamente o custo e o tempo para entregar novas funcionalidades ao negócio e melhorar a consistência e o controle.

Com a solução de gerenciamento de dados do lakehouse na nuvem orientada a metadados líder no setor, você pode liberar o potencial completo de seu armazém de dados e data lake na nuvem em um ambiente híbrido multinuvem. Você ganha eficiência e economia, pode começar pequeno e aumentar com a melhor integração de dados, qualidade e governança de dados e gerenciamento de metadados, criados para a nuvem em uma plataforma de dados inteligentes com base em IA.

### Próximas etapas

Experimente o gerenciamento de dados de lakehouse na nuvem inteligente e automatizado da Informatica sem custo por 30 dias. Inscreva-se para fazer um [teste gratuito do Informatica Cloud Data Integration](#).

