Arquitetura MDM

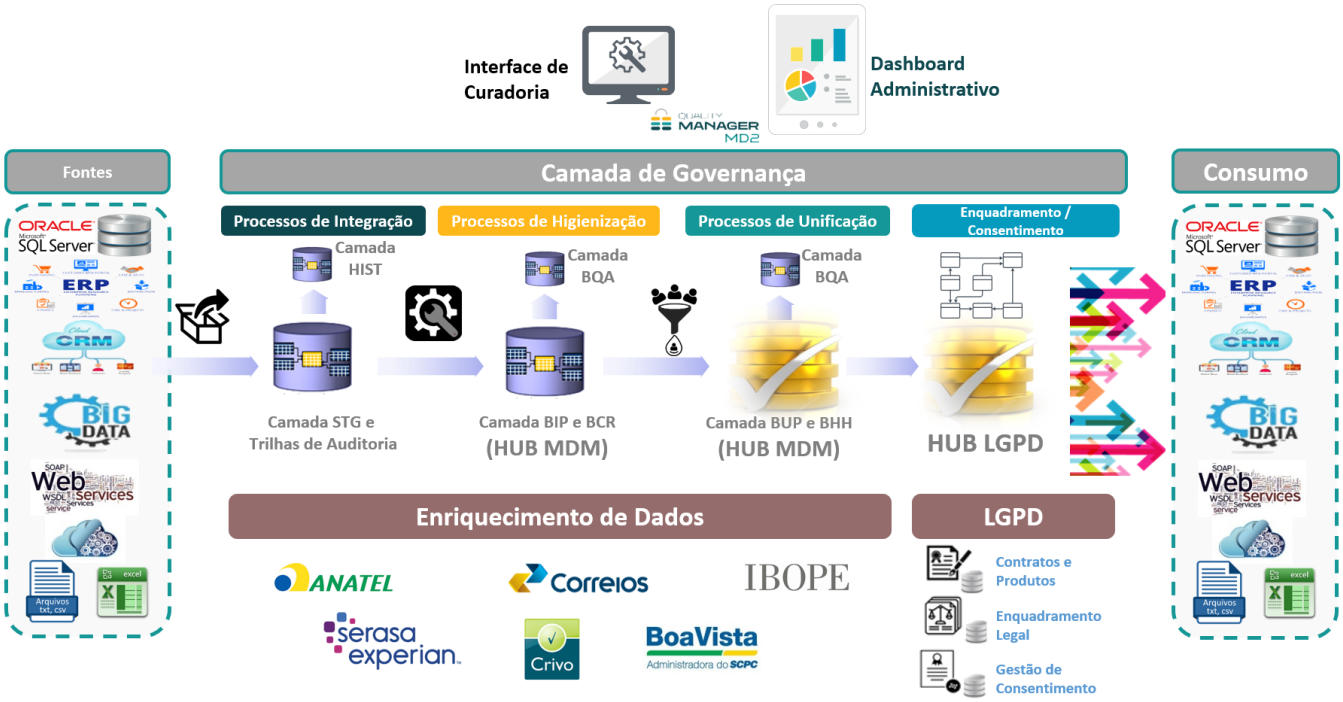
- Arquitetura
- Estilos de Implementação MDM
- Arquitetura Ambiente
  - 2.1. Camada Servidor
  - 2.2. Camada Cliente
  - 2.3. Engine de Processamento Paralelo Unificado
  - 2.4. Conectividade Ampla
  - 2.5. Metadados Unificados

Arquitetura

A solução de MDM da MD2 oferece uma plataforma centralizada de gestão de dados mestre em nível corporativo. A partir do monitoramento contínuo dos sistemas conectados a esta plataforma.

As informações inseridas ou alteradas são submetidas a uma bateria de regras de tratamento. As regras de validação, saneamento e enriquecimento de dados serão aplicadas a estas informações com intuito de garantir que o Golden Record (registro mestre que representa a visão 360º de cada pessoa) contenha a informação mais atual e assertiva possível, consolidada em nível corporativo a partir de todas estas fontes de dados.

Abaixo, um diagrama da arquitetura da solução MDM:



- Fontes: Views de cada assunto para cada sistema de origem (Pessoa, Endereço, Telefone, E-mail, Documentos, etc.);
- Integração: Monitoramento das alterações cadastrais nas fontes de origem, captura de dados, rastreabilidade;
- Higienização: Regras de validação, padronização das informações, enriquecimento de dados;
- Unificação: Resolução de identidade, regras de sobrevivência, composição do Golden Record;
- Enquadramento: controle na automação das hipóteses legais ao longo do ciclo de vida da informação e gestão de consentimento caso os respectivos fluxos de tratamento de dados estejam formalizados no Quality Manager;
- Consumo: serviços de publicação de dados mestre de referência, serviços de consumo, serviços de qualidade de dados;

- Governança: Controle de indicadores através de dashboards de gestão e curadoria de dados;

Uma vez implantada, a solução permite que futuros sistemas sejam adicionados de forma fácil e rápida, fornecendo uma arquitetura altamente flexível baseada em camadas que representam os diversos domínios funcionais. Os sistemas legados podem estar implantados em plataformas e tecnologias diferentes e o objetivo da camada de Ingestão é cuidar do trabalho da captura e integração de dados entre estes sistemas de armazenamento de dados e o MDM. Recomendamos a estratégia de integração por Views, por ser a menos intrusiva para os sistemas de origem e por trazer maior desacoplamento entre estas aplicações.

A carga inicial será aplicada na integração de todos dados com a base centralizada (HUB MDM), sendo estabelecidos a partir daí ciclos de cargas incrementais recorrentes de forma contínua para manter as informações sempre atualizadas e confiáveis, sendo gerenciados através de controle de carga próprio da solução.

Estilos de Implementação MDM

O sucesso da implantação de um sistema MDM é diretamente relacionado a correta escolha do método de implementação da solução, visando as diretrizes de governança dos dados, a gestão e o acesso aos dados resultantes da esteira de integração, qualificação e unificação do motor.

Segue abaixo um descritivo dos estilos de implementação MDM suportados pela solução e também como são providos na arquitetura do produto.

Centralizado/Transacional

Dados são criados, armazenados e processados pelo motor MDM, sendo após isso publicados para consumo por parte dos sistemas de origem. Esse estilo de implementação pressupõe a interferência do MDM na consistência dos dados na origem.

Consolidação

Os dados são extraídos de diversas fontes e são processados pelo Motor MDM, gerando dessa forma o Golden Record, sendo esse a visão única dos dados mestre, que é armazenado em um hub visando o consumo.

Coexistência

Mescla os métodos centralizado e consolidação, onde o Golden Record é gerado a partir de inserções e edições no próprio MDM ou nos sistemas de origem, sendo passíveis de consumo visando inclusive atualização dos sistemas de origem. É importante salientar que os dados são tratados e unificados pelo motor MDM.

Registro

Atribuição de identificação exclusiva aos registros visando criar uma estrutura de rastreabilidade onde é possível visualizar as associações entre os dados da origem que culminaram na formação da visão única. As alterações no MDM são realizadas a partir da manipulação nos sistemas de origem, não existindo uma retroalimentação.

A solução MD2 MDM provê funcionalidades e uma estrutura de modelo dos dados que permite a implantação dos estilos MDM descritos acima de forma separada ou híbrida.

A implementação centralizada/transacional prevê a criação de registros no MDM. Para esse intuito a solução fornece duas formas para cadastro e edição de registros golden no MDM.

API Inserção Hub Pessoa

API Atualização Hub Pessoa

Cadastro e Edição de Golden Records

Em implementações centralizada/transacional, consolidação e de coexistência é necessário prover uma forma de consumo dos dados, seja por aplicações terceiras ou pelos sistemas de origem. Nesse contexto a solução provê uma arquitetura com diversificados métodos de publicação de dados.

Publicação de Dados MDM

Para a implementação no modo registro a solução provê em seu modelo de dados tabelas de trilha inseridas nas camadas STG e BUP, sendo a primeira responsável por garantir a rastreabilidade do registro no sistema origem, e a segunda responsável por garantir a as associações e vigências entre os registros STG e os Golden Records gerados a partir dos mesmos.

Arquitetura Ambiente

A suíte IBM InfoSphere Information Server impulsiona a solução MD2 MDM com todo o ferramental necessário para realizar a gestão de metadados, integração as bases de dados e implementação do motor de ingestão, qualificação e unificação do dados mestre. Esse capítulo apresenta as camadas da arquitetura do ambiente IBM InfoSphere Information Server.

2.1. Camada Servidor

O Information Server é inteiramente construído sobre um conjunto de serviços compartilhados que centraliza as tarefas essenciais em toda a plataforma. O compartilhamento dos serviços permite que os mesmos sejam controlados em único local, independente de qual componente esteja sendo utilizado. A plataforma Information Server é baseada na arquitetura Cliente/ Servidor e inclui as camadas de Serviços, Engine, Repositório e Working Areas.

Camada Servidor

Inclui diversas camadas que interagem do lado Server da solução, compreendida por componentes distintos.

Camada Serviços

É suportada pelo IBM WebSphere Apllication Server. Pode ser desmembrada para que funcione em um servidor distinto dos demais componentes.

- **Common services** – são usados por toda a suíte Information Server para tarefas de segurança, administração de usuários, logs, metadados e execução;
- **Product-specific services** – fornece tarefas específicas para um determinado produto dentro da suíte Information Server;
- **Application Server**: Servidor de Aplicações WebSphere que realiza a autenticação dos usuários na suíte, além de hospedar WebServices desenvolvidos

Camada Repositório

O repositório compartilhado é um banco de dados usado para armazenar todos os objetos de todos os componentes da suíte Information Server.

Camada Engine

É o motor do QualityStage, que estabelece as conexões entre as fontes e alvos de dados, além de executar os processos desenvolvidos. Este serviço deve estar ativo para que as conexões de Pessoas QualityStage sejam aceitas.

Working Areas

São áreas de armazenamento temporário utilizada pelos componentes da suíte.

2.2. Camada Cliente

A suíte IBM Information Server fornece várias interfaces otimizadas para cada perfil de usuário dentro da organização.

Administrative clients - Permite gerenciar áreas de segurança, licenciamento, logs e scheduling. Tarefas administrativas tais como criação de usuários e grupos de usuários, definição de perfis de usuários e gerenciamento de sessões ativas são fornecidas pelo cliente Information Server Web console; Tarefas administrativas em projetos QualityStage são fornecidas pelo cliente QualityStage Administrator (interface desktop). Ele é usado para criar, deletar e alterar as propriedades de cada projeto. Privilégios de usuários e grupos também podem ser definidas pelo Quality Administrator;

User clients – Permite a criação, desenvolvimento, gerenciamento, execução, scheduling e monitoramento de Jobs. Quality Designer permite a criação, o gerenciamento e o desenvolvimento de jobs; Quality Director é responsável pela validação, execução, schedule e monitoramento de jobs;

2.3. Engine de Processamento Paralelo Unificado

O trabalho executado pelo Information Server é realizado dentro do engine de processamento paralelo. Este engine manipula necessidades de processamento de dados diversos assim como executa as análises requeridas pelos componentes InfoSphere Information Analyzer, higienização pelo InfoSphere QualityStage, e ainda transformações complexas pelo componente InfoSphere DataStage. Este engine é desenhado para:

- Paralelismo pipelining para tratar grandes volumes de dados e diminuir a janela de carga;
- Por adição de hardware escalar sem necessidade de alteração nos processos desenvolvidos.

Com o paralelismo pipelining é possível executar simultaneamente processos de transformação, limpeza e carga. Como uma linha de produção o Engine move linhas de processo a processo: ele inicia o processo de gravação enquanto o processo de leitura ainda está em execução.

Vantagens:

- Reduz a utilização de disco (áreas temporárias);
- Evita a ociosidade dos processadores;
- Suporta todas as configurações de hardware (Single Processor, SMP, MPP, Cluster, Grid);
- Aumenta a utilização de recursos pela simples adição de processadores ou nós de processamento sem a necessidade de alteração dos processos desenvolvidos.

2.4. Conectividade Ampla

O Information Server conecta fontes de dados seja estruturada, não-estruturada, no Mainframe, em plataforma distribuída ou em aplicações de mercado. Os metadados são compartilhados através dos módulos da suíte e objetos de conexão são reutilizáveis em outras funções. Conectores proveem em tempo de construção a importação dos metadados, navegação nos dados e acesso aos metadados operacionais.

Interfaces pré-construídas para aplicações fechadas, chamadas de Packs, proveem adaptadores para aplicações como SAP, Siebel, Oracle, entre outras, habilitando a integração com aplicações corporativas e associadas a ferramentas de reporting e sistemas analíticos.

2.5. Metadados Unificados

O Information Server foi construído em uma infraestrutura que habilita o compartilhamento do entendimento entre domínios técnicos e de negócio. Esta arquitetura reduz o tempo de desenvolvimento e provê registros persistentes que aumentam a confiança nas informações geradas. Todas as funções do Information Server compartilham o mesmo metamodelo, facilitando o trabalho de diferentes pessoas e seus respectivos papéis na empresa.

- **Metadados Dinâmicos** - Metadados dinâmicos incluem informação em tempo da construção dos processos de integração de dados.
- **Metadados Operacionais** - Metadados operacionais incluem informação de performance, monitoração, auditoria, log e data profiling de dados amostrais.

Destaca-se a existência do IBM InfoSphere Information Governance Catalog, uma ferramenta que fornece um ambiente colaborativo que auxilia os membros da organização a criar um catálogo central de terminologia específica da empresa, incluindo os relacionamentos com ativos. Tal catálogo é projetado para ajudar os usuários a entender a linguagem empresarial e o significado comercial dos ativos de dados, tais como bancos de dados, empregos, tabelas e colunas de bancos de dados e relatórios de inteligência empresarial.

Por intermédio do IBM InfoSphere Information Governance Catalog os usuários designados podem definir termos, categorias, políticas de governança da informação e regras de governança da informação, podendo obter informações sobre terminologia comercial comum, descrições de dados, propriedade de termos e metadados e como os termos se relacionam com os ativos de informação.