

Inteligência artificial para inovação conduzida por dados

Como as inovações baseadas em aprendizagem automática no
CLAIRE estão promovendo um grande salto em produtividade de
dados

Este documento contém informações confidenciais, exclusivas e segredos comerciais (“Informações confidenciais”) da Informatica que não podem ser copiados, distribuídos, duplicados ou de outra forma reproduzidos, de nenhuma maneira, sem o consentimento prévio por escrito da Informatica.

Apesar de todos os esforços terem sido realizados para assegurar que as informações deste documento sejam precisas e completas, alguns erros tipográficos ou imprecisões técnicas podem existir. A Informatica não aceita responsabilidade por nenhum tipo de perda resultante do uso das informações contidas neste documento. As informações contidas neste documento estão sujeitas a alterações sem aviso prévio.

A incorporação dos atributos de produto abordados neste material em qualquer versão ou atualização de qualquer produto de software da Informatica – bem como o momento de tal versão ou atualização – será definida a critério exclusivo da Informatica.

Protegido por uma ou mais das seguintes patentes norte-americanas: 6.032.158; 5.794.246; 6.014.670; 6.339.775; 6.044.374; 6.208.990; 6.208.990; 6.850.947; 6.895.471; ou pelas seguintes patentes pendentes norte-americanas: 09/644,280; 10/966,046; 10/727,700.

Edição publicada em maio de 2017

Índice

Apresentação	2
Tendências em gerenciamento de dados	3
O que isso significa para os líderes de TI	4
O que isso significa para os líderes de negócios	4
O que é aprendizagem automática?	5
Por que usar aprendizagem automática para gerenciamento de dados?	5
A base para a aprendizagem automática no gerenciamento de dados	5
Informatica CLAIRE: a “inteligência” da Intelligent Data Platform	6
CLAIRE em ação	7
Similaridade de dados inteligentes	7
Descoberta inteligente de domínio com tags	8
Descoberta de entidades inteligente	9
Recomendações de dados inteligentes	9
Descoberta de estruturas inteligente	10
Detecção inteligente de anomalias	11
Conclusão	12

Apresentação

A transformação digital é real e já está em progresso. Agora é a hora de inovar, ou sofrer as consequências. As organizações estão promovendo iniciativas transformadoras para melhorar o desempenho financeiro e a posição competitiva em seus setores. Exemplos dessas iniciativas incluem aprofundar o relacionamento com o cliente, otimizar operações, personalizar cuidados com a saúde e prevenir fraudes.

O fator essencial que impulsiona essas iniciativas é a capacidade de alimentá-las com dados confiáveis e oportunos. É muito simples: estratégias digitais bem-sucedidas são baseadas em dados. A competência que você desenvolve com o gerenciamento de dados determinará o grau de sucesso da estratégia digital. Em outras palavras, sua estratégia digital será apenas tão eficiente quanto os dados que a alimentam.

O mais provável, porém, é que continuar gerenciando dados da mesma maneira não seja suficiente. Os líderes de TI estão buscando maneiras de aumentar a produtividade do gerenciamento de dados para disponibilizar a todos dados melhores e mais rapidamente.

O mecanismo CLAIRE™ (Cloud-scale AI-powered Real-time Engine, mecanismo em tempo real habilitado por inteligência artificial em escala de nuvem) da Informatica, que usa inteligência artificial (AI) e técnicas de aprendizagem automática alimentada por dados e metadados corporativos, impulsionará significativamente a produtividade de todos os administradores e usuários de dados da organização.

Tendências em gerenciamento de dados

É hora de uma nova visão sobre dados e arquitetura de dados. Por décadas, o foco esteve em sistemas e processos de negócios. Embora esses fatores ainda sejam importantes, a habilidade de entregar dados melhores, mais rápidos e mais completos às iniciativas de negócio será o verdadeiro fator diferencial para a sua organização no mercado. Porém, o orçamento da maioria das organizações de TI aumenta em ritmo inferior às necessidades de dados. Por isso, você também precisa considerar fazer mais com os recursos que tem no momento.

O desafio de gerenciar dados corporativos nunca foi tão grande. Para aproveitar todos os benefícios dos dados, a organização de TI precisa ser capaz de gerenciar:

1. Mais dados:

- **Volume de dados:** São 15,3 zettabytes por ano em tráfego global de data center.
- **Complexidade e variedade de dados:** Há muitas novas fontes e tipos de dados dentro e fora da empresa.
- **Velocidade dos dados:** O surgimento da Internet das coisas (IoT), com 20 bilhões de dispositivos conectados, significa transmissão de dados sempre ativa.

2. **Mais usuários:** São 325 milhões de usuários de dados corporativos — um número que não para de crescer — e todos eles, de analistas de negócios e cientistas de dados a gestores de dados, querem acesso direto e imediato aos dados.

3. Mais padrões de integração:

- **Mudança para a nuvem:** Os pacotes de ERP estão se separando e indo para a nuvem.
- **Tecnologia de análise de dados:** O setor está adotando novas tecnologias, como Big Data, NoSQL e análise preditiva, para complementar o armazenamento de dados.
- **Experimentação:** Os usuários agora querem usar os dados para formar rapidamente uma hipótese, testá-la, avaliarem seu sucesso ou falha e iterarem rapidamente. O foco é na precisão sobre a velocidade, até que comprovem o mérito da hipótese.

O que isso significa para os líderes de TI

Todas essas tendências combinam-se para tornar o processo de gerenciar dados ainda mais complexo, justo quando as organizações começam a perceber que os dados são o combustível da transformação digital.

Essa é uma oportunidade ideal para oferecer liderança conduzida por dados para contribuir com o sucesso da sua organização. Como os líderes de TI atendem à necessidade de negócio de obter dados melhores mais rapidamente sem exigir um exército de desenvolvedores de alto custo, por exemplo?

Com orçamentos de TI que crescem pouco, quando crescem, há três maneiras principais de fazer isso:

- Aumentar a automação e a eficiência para tarefas e projetos de gerenciamento de dados
- Aumentar a capacitação para autoatendimento empresarial
- Aumentar a colaboração para promover alinhamento entre equipes de negócios e técnicas

O que isso significa para os líderes de negócios

Os líderes de negócios sentem que têm o poder de dirigir iniciativas inovadoras e fazer perguntas que nunca tiveram o alcance econômico para fazer. Porém, os resultados de suas iniciativas digitais serão apenas tão bons quanto os dados que utilizam.

Sua maior prioridade deve ser a definição de um plano para aproveitar os benefícios de todos os seus dados.

É importante desenvolver uma competência de gerenciamento de dados como base para todas as iniciativas digitais. Você precisa gerenciar dados como um ativo que pode ser descoberto e utilizado por qualquer usuário da organização. Os dados devem ser de uma qualidade adequada à sua finalidade: alta qualidade para decisões e interações importantes e boa qualidade para inovação e iteração rápidas. Quanto à tecnologia, a codificação manual ou um conjunto de ferramentas de gerenciamento de dados não integradas não é dimensionável para atender às necessidades da empresa.

O que é aprendizagem automática?

Aprendizagem automática é uma técnica pela qual os programas aprendem de maneira iterativa com os dados, em vez de serem estáticos. Os sistemas de aprendizagem automática são usados para criar um modelo baseado em entrada que possa ser usado para fazer previsões ou tomar decisões. Esses sistemas aprendem com os dados e podem ajustar-se de acordo para gerar resultados melhores. Quanto mais dados eles têm, mais rápido aprendem e mais precisos são os resultados.

Por que usar aprendizagem automática para gerenciamento de dados?

Para aumentar a velocidade da distribuição de dados nas iniciativas de negócios cruciais, é necessário aumentar a automação. É aqui que a aprendizagem automática entra em cena. Com visibilidade de metadados em toda a empresa e aprendizagem automática, as ferramentas de gerenciamento de dados podem ser “ensinadas” a fazer recomendações inteligentes e automatizar muitas tarefas de gerenciamento de dados. A aprendizagem automática não substitui os analistas de dados e outros usuários. Em vez disso, é o segredo para aumentar a produtividade e a eficiência da equipe de administração de dados da organização.

A aprendizagem automática pode ser utilizada para aprimorar tarefas entediadas ou impossíveis de serem feitas em escala humana. Alguns exemplos incluem:

1. Descoberta e identificação

- Regras de qualidade de dados e descoberta de entidades de negócios
- Pesquisa semântica, identificação de padrões e classificação de dados
- Detecção e notificação de anomalias

2. Operações preditivas

- Explosão para lidar com picos de dados
- Priorizar investigações de problema operacional
- Auto reparação para lidar com mudanças nos ambientes

3. Melhorar a próxima ação e recomendações

- Sugerir conjuntos de dados, transformações e regras
- Mapear automaticamente, limpar e padronizar da origem ao destino
- Auto integração a novas fontes de dados

A base para a aprendizagem automática no gerenciamento de dados

Uma aprendizagem automática eficaz requer grandes conjuntos de dados de treinamento. Em um contexto de gerenciamento de dados, a melhor fonte de dados é um catálogo de dados corporativo. A maioria das empresas tem centenas de bancos de dados, arquivos de dados, aplicativos e sistemas de análise. Ao reunir os metadados desses repositórios de dados, as empresas podem criar um catálogo amplamente preenchido. A combinação de aprendizagem automática e catálogo de dados com visibilidade de metadados em nível corporativo oferece a base para inteligência capaz de gerar impacto verdadeiramente significativo e positivo sobre a produtividade do gerenciamento de dados.

Nesta era da nuvem, é importante observar que essa abordagem também funciona para aplicativos SaaS. Os metadados podem ser coletados de aplicativos de SaaS, como Salesforce e Workday, e adicionados a um catálogo corporativo.

Informatica CLAIRE: a “inteligência” da Intelligent Data Platform

A abordagem da Informatica para promover a produtividade do gerenciamento de dados com aprendizagem automática é:

1. A Intelligent Data Platform (IDP): Desenvolvemos uma plataforma de gerenciamento de dados completa para máxima produtividade. Ao fornecer gerenciamento unificado de conectividade, metadados e operações, a plataforma unificada acelera o desenvolvimento e a implementação de novos projetos de gerenciamento de dados. A plataforma oferece um conjunto avançado e consistente de funcionalidades para gerenciar dados entre fontes locais, na nuvem e de Big Data. Chamamos essa plataforma de gerenciamento de dados unificada de "Intelligent Data Platform".

Essa plataforma é modular. Comece com qualquer ferramenta única e expanda ao seu ritmo.



Figura 1: A Intelligent Data Platform integra funcionalidades de gerenciamento de dados a conectividade compartilhada, visão operacional e inteligência de dados e metadados.

2. Metadados: A Informatica é líder reconhecida de longa data por seu gerenciamento de metadados técnicos e de negócios. A Informatica aumentou suas funcionalidades nessa área coletando um espectro mais amplo de metadados em toda a empresa, inclusive:
 - Metadados técnicos, como tabelas de banco de dados, informações de coluna e estatísticas de perfil de dados
 - Metadados de negócios, que capturam o contexto dos dados, seu significado, relevância e importância para vários processos e funções
 - Metadados operacionais, sobre sistemas e execução de processo – por exemplo: Quando foi a última atualização dos dados? Quando foi a última execução do processo de carregamento? Que dados foram mais acessados?
 - Metadados de uso, sobre a atividade do usuário, inclusive conjuntos de dados acessados, resultados de pesquisa que receberam cliques, classificações ou comentários fornecidos

Essa coleção mais ampla de metadados é fundamental para a aprendizagem automática. Ela fornece conjuntos de dados que são usados para “treinar” os algoritmos de aprendizado e possibilitar seu ajuste para produzir resultados melhores.

3. Inteligência: A Informatica está oferecendo uma combinação integrada de metadados e AI/aprendizagem automática com o CLAIRE.

Os metadados coletados pela Intelligent Data Platform fornecem uma ampla gama de informações avançadas que os algoritmos do CLAIRE podem usar para aprender sobre o panorama de dados de uma empresa. Esse conhecimento ajuda o CLAIRE a fazer recomendações inteligentes, automatizar o desenvolvimento e o monitoramento dos projetos de gerenciamento de dados e adaptar-se a mudanças dentro e fora da empresa. O CLAIRE é o que impulsiona a inteligência de todas as funcionalidades de gerenciamento de dados na Intelligent Data Platform.

CLAIRE em ação

O CLAIRE ajuda uma ampla variedade de usuários:

- Desenvolvedores de dados verão que muitas tarefas de implementação são parciais ou totalmente automatizadas
- Os analistas de dados terão mais facilidade em localizar e preparar os dados necessários
- Os usuários de negócio identificarão rapidamente os dados que devem ser submetidos a governança e controles de conformidade determinados
- Os cientistas de dados entenderão os dados com mais rapidez
- Gestores de dados terão mais facilidade em visualizar a qualidade dos dados
- Os profissionais de segurança de dados consideram mais simples detectar uso indevido de dados, proteger dados confidenciais e demonstrar que os controles adequados são mantidos
- Os administradores e operadores terão o poder de fazer a manutenção preditiva e a otimização de desempenho dos processos de gerenciamento de dados.

Estes são alguns exemplos de como a inteligência oferecida pelo CLAIRE está sendo utilizada hoje.

Similaridade de dados inteligentes

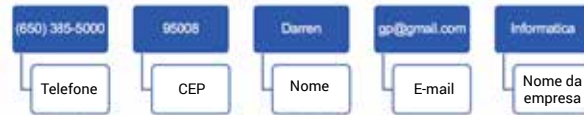
O CLAIRE usa técnicas de aprendizagem automática, como clustering, para detectar dados similares em milhares de bancos de dados e conjuntos de arquivo. A similaridade inteligente de dados é uma das principais funcionalidades utilizadas para diversos fins, inclusive identificar dados, detectar duplicatas, combinar campos de dados individuais em entidades de negócios, propagar tags entre conjuntos de dados e recomendar conjuntos de dados a usuários.

A similaridade dos dados calcula o quanto os dados em duas colunas são iguais. Uma abordagem de força bruta para tentar comparar todos os pares de colunas em um contexto empresarial (digamos, entre 100 milhões de colunas) seria computacionalmente proibitiva. Em vez disso, a similaridade de dados usa técnicas de aprendizagem automática para agrupar em cluster colunas similares e identificar correspondências prováveis.

O processo funciona em vários estágios. Primeiro, as colunas são agrupadas em clusters com base em suas características. Então, a sobreposição de dados é calculada para valores exclusivos em cada um desses clusters. Por fim, os pares mais promissores são escolhidos para calcular a similaridade de dados usando os coeficientes de Bray-Curtis e Jaccard.

Descoberta inteligente de domínio com tags

O CLAIRE é capaz de classificar campos de dados aplicando rótulos semânticos a cada coluna. Esses rótulos semânticos chamam-se domínios de dados.



Em geral, rótulos semânticos serão aplicados avaliando regras com base em expressões regulares, tabelas de referência ou outra lógica complexa de codificação manual. Definir e atualizar milhares dessas regras é uma tarefa entediante.

O CLAIRE utiliza o conceito de tags para simplificar drasticamente o processo de descobrir e rotular os campos de dados. Para as colunas que ainda não foram classificadas, o usuário precisa apenas fornecer uma tag simples (digamos, "Data de pagamento das reivindicações") indicando o conteúdo da coluna. O sistema aprende por associação e então propaga automaticamente essa tag a todas as colunas similares. A técnica de "reconhecimento facial" de dados é equivalente a marcar pessoas em uma foto do Facebook, com o efeito em de que a mesma pessoa é marcada em milhões de outras fotos.

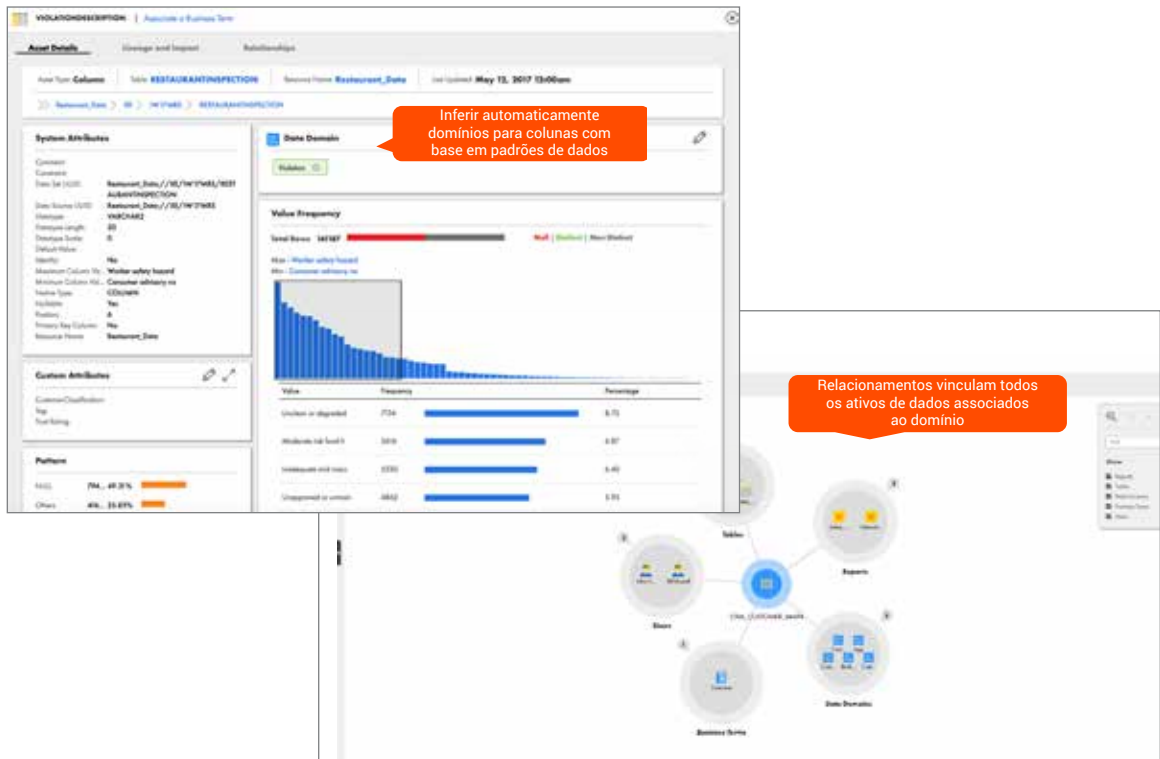


Figura 3: Classificação automática de dados.

Descoberta de entidades inteligente

Depois que os domínios para colunas foram identificados, o CLAIRE pode montar esses campos individuais em entidades de negócio de nível mais alto. O exemplo a seguir mostra como uma entidade chamada Purchase Order (Ordem de compra) é criada combinando campos identificados como Customer (Cliente) e Product (Produto). A descoberta de entidades aprende com base no modo como os usuários montaram campos de dados distintos em seus processos de integração ou análise de dados e aplica esse aprendizado para derivar entidades em todo o ambiente de dados corporativos.

Field0	Field1	Field2	Field3	Field4	Field5	Field6	Field7	Field8	Field9
4/5/2015	Estelle	Chambers	7332 Branch St.	Far Rockaway	NY	11651	70296	4-Porting SD Card SGB Class 6	308276.28
8/30/2016	Alfred	Sanchez	7578 Maiden St.	Potomac	MD	20854	71889	Halqae UTP CAT5e 400c cable Grango 6,5M Gint	301080
10/3/2015	Brandon	VaSci	33 N. Longfellow Lane	Atlantic City	NJ	8403	79038	Y-trunk tablet TA8344 8" GoTab (rev)	335500
12/21/2013	Lee	Morton	7 Sunbeam Dr.	Lippert Dorby	PA	19082	72526	Asus NB A7750 T952V (3.250/17.1"/4/500/W/16	97508

Entidades identificadas:

- Cliente**: Campos 2, 3, 4 (Nome, Sobrenome, Endereço)
- Endereço**: Campos 5, 6, 7, 8 (Rua, Cidade, Estado, CEP)
- Produto**: Campo 9 (ID do produto, Nome do produto)

Figura 4: Combinação de domínios de dados para detectar entidades de tabelas e arquivos.

Recomendações de dados inteligentes

O CLAIRE sugere aos analistas e cientistas de dados quais conjuntos de dados usar nos projetos. Ele observa os conjuntos de dados que os usuários selecionaram e sugere outros que sejam similares e mais bem classificados ou conjuntos de dados adicionais que podem complementar aqueles que estão sendo usados. Recomendações de dados inteligentes ajudam os usuários a evitar repetir o mesmo trabalho que muitos colegas já realizaram. As recomendações incluem:

1. Uma versão preparada dos mesmos dados (dados substituíveis)
2. Outra tabela contendo o mesmo tipo de registro (dados que podem ser unidos)
3. Uma tabela que possa ser acrescentada para aprimorar os dados com mais atributos (dados que podem ser acrescentados).

Recomendações de dados usam técnicas de filtragem com base em conteúdo para fornecer sugestões sobre conjuntos de dados adicionais. As características (termos) usadas para conjuntos de dados incluem informações de linhagem, classificação de usuário e similaridade de dados. Várias medidas de similaridade são usadas para pontuar a equivalência de diferentes conjuntos de dados. Essas pontuações então são usadas para recomendar conjuntos de dados com propriedades similares. Itens complementares são recomendados consultando o gráfico de metadados para localizar conjuntos de dados comumente usados juntos por diferentes usuários.

Descoberta de estruturas inteligente

O CLAIRE pode derivar a estrutura a partir de arquivos de registro e dispositivo desorganizados, tornando mais fácil entendê-los e trabalhar com eles. Ao usar a abordagem baseada em conteúdo para analisar arquivos, pode adaptar-se a mudanças frequentes de arquivo sem afetar o processamento do arquivo.

A descoberta de estruturas inteligente usa um algoritmo genético para automatizar o reconhecimento de padrões nos arquivos. Nessa abordagem, ela usa o conceito de “evolução” para melhorar os resultados. Cada solução candidata tem um conjunto de propriedades que podem ser alteradas e então testadas para determinar se fornecem uma solução mais adequada. Não requer contribuição do usuário para definir a estrutura do arquivo, nem é específica de um conjunto de formatos de arquivo do setor. As estruturas iniciais do arquivo são derivadas de análise básica baseada em delimitador. Essas estruturas então são pontuadas de acordo com diversos fatores, como cobertura da entrada e domínios derivados. As estruturas mais bem pontuadas entram em uma fase de “mutação”, na qual são feitas várias mudanças às estruturas, como combinar subestruturas, para tentar aumentar a pontuação. O processo termina quando é determinada a adequação correta de estrutura aos dados.

```
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir1
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir2
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir3
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir4
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir5
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir6
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir7
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir8
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir9
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir10
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir11
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir12
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir13
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir14
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir15
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir16
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir17
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir18
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir19
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir20
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir21
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir22
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir23
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir24
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir25
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir26
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir27
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir28
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir29
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir30
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir31
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir32
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir33
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir34
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir35
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir36
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir37
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir38
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir39
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir40
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir41
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir42
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir43
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir44
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir45
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir46
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir47
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir48
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir49
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir50
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir51
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir52
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir53
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir54
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir55
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir56
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir57
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir58
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir59
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir60
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir61
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir62
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir63
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir64
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir65
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir66
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir67
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir68
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir69
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir70
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir71
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir72
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir73
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir74
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir75
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir76
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir77
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir78
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir79
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir80
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir81
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir82
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir83
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir84
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir85
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir86
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir87
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir88
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir89
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir90
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir91
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir92
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir93
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir94
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir95
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir96
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir97
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir98
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir99
21/1/2014 9:21:14 am Found Directory as \\192.168.1.100\csm_data\dir100
```

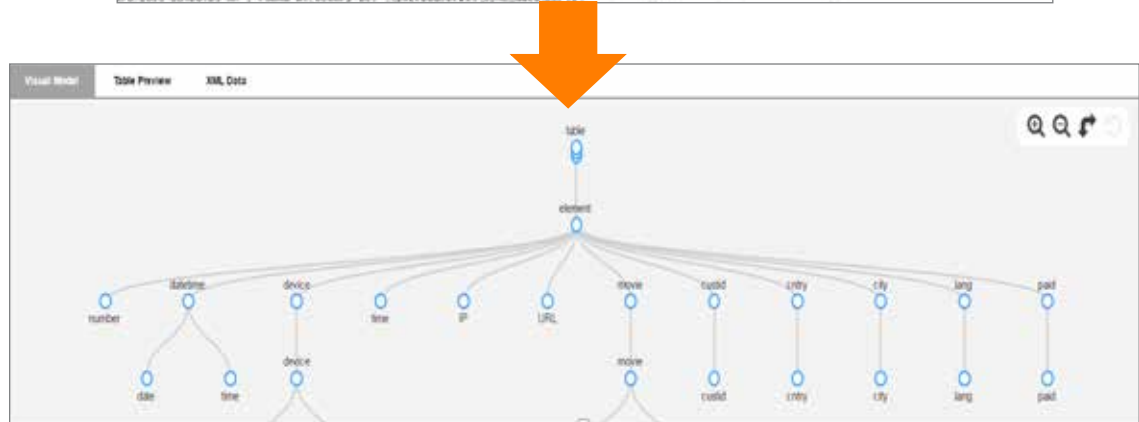


Figura 5: Localização inteligente de estruturas em arquivos de dados não estruturados

Detecção inteligente de anomalias

O CLAIRE usa abordagens de aprendizagem automática e estatísticas para detectar desvios e anomalias de dados. A funcionalidade UBA (análise de comportamento do usuário) detecta padrões de comportamento do usuário que podem ser arriscados e expor uma organização ao uso indevido de dados. A UBA é capaz de detectar personificação, sequestro de credenciais e ataques de aumento de nível de privilégios.

A UBA aplica aprendizagem automática não supervisionada a um modelo multidimensional de atividades do usuário, que inclui o número de repositórios de dados acessados pelo usuário, o número de solicitações feitas e o número de registros afetados em diferentes sistemas. A análise de componente principal é aplicada a esse modelo para redução de dimensionalidade. A técnica BIRCH é aplicada para clustering hierárquico não supervisionado para localizar usuários cujo comportamento foi diferente durante um dado período. Para validar o comportamento anômalo, métodos de detecção de desvio com base em distância e densidade são utilizados, e o teste estatístico de Grubb para desvios é realizado para confirmar que os objetos indicados pelos dois primeiros métodos são, de fato, desvios no sistema de cluster.

Estes são alguns exemplos das funcionalidades do CLAIRE a serem lançadas no futuro:

Auto integração: Integrar automaticamente dados recém-chegados aos processos de integração de dados. Identificar dados, localizar padrões de integração que processam dados similares, transformar automaticamente e mover dados com aprendizados de milhões de mapeamentos e ações do usuário existentes.

Assistência de desenvolvimento: Fornecer recomendações aos usuários e sugerir melhorias nas próximas ações durante o processo de desenvolvimento, inclusive:

- Autopreenchimento da transformação
- Recomendações de modelo
- Sugestões de tipo de mascaramento para dados confidenciais
- Sugestões de qualidade de dados para limpeza e padronização
- Otimizações automáticas de desempenho

Mapeamento automático: Detectar entidades de dados mestre em toda a empresa e mapeá-las automaticamente para o modelo de dados mestre aplicando as transformações obrigatórias e regras de qualidade

Auto reparação: Lidar tranquilamente com problemas de sistema externo, como pouca memória ou potência de computação. Por exemplo, incluir computação adicional (“explosão para a nuvem”) para lidar com picos de dados

Auto ajuste: Com base em informações históricas, volumes de dados atuais e recursos do sistema disponíveis preveem e ajustam os cronogramas ou calculam recursos para atender aos critérios de desempenho

Autoproteção: Detectar automaticamente dados confidenciais e mascarar-los antes que deixem uma região segura

Sobre a Informatica

A transformação digital está mudando nosso mundo.

Como líder no gerenciamento corporativo de dados em nuvem, estamos preparados para ajudá-lo a abrir caminho de maneira inteligente. Para fornecer a você a perspectiva para tornar-se mais ágil, aproveitar novas oportunidades de crescimento ou até mesmo inventar novos produtos. Convidamos você a explorar tudo que a Informatica tem a oferecer — e libertar o poder dos dados para gerar sua próxima revolução inteligente. Não apenas uma vez, mas sempre.

Conclusão

As atuais estratégias de negócio centradas em dados são baseadas em dados. Vencer requer o desenvolvimento de uma competência em gerenciamento de dados para aproveitar os benefícios dos dados.

Com todos os desafios que o gerenciamento de dados apresenta em circunstâncias normais, as abordagens tradicionais não podem ser dimensionadas para atender às exigências de hoje, sem falar das futuras. Uma maneira de aproveitar seus dados para promover a inovação é padronizar uma plataforma de gerenciamento de dados completa que aproveite os benefícios dos dados, metadados e aprendizagem automática/AI para aumentar a produtividade de todos os usuários da plataforma: técnicos, operacionais, de negócios e, especialmente, de autoatendimento de negócios.

[Entre em contato conosco](#) para saber mais sobre como usar o CLAIRE e a Intelligent Data Platform para aproveitar toda a capacidade dos seus dados.



Informatica

Sede mundial, 2100 Seaport Blvd, Redwood City, CA 94063, EUA Tel.: 650.385.5000 Fax: 650.385.5500

Ligação gratuita nos EUA: 1.800.653.3871 informatica.com [linkedin.com/company/informatica](https://www.linkedin.com/company/informatica) twitter.com/Informatica

© 2017 Informatica LLC. Todos os direitos reservados. Informatica, o logotipo da Informatica e CLAIRE™ são marcas comerciais ou marcas registradas da Informatica LLC nos Estados Unidos e em diferentes jurisdições por todo o mundo. Uma lista atualizada de marcas comerciais da Informatica está disponível na web em <https://www.informatica.com/trademarks.html>. Outros nomes de empresas e produtos podem ser nomes comerciais ou marcas comerciais de seus respectivos proprietários.

IN09_0517_3328