

Informatica Enterprise Data Catalog

Vantaggi

- Catalogazione e classificazione automatiche di tutti i tipi di dati nell'azienda grazie a un catalogo basato su intelligenza artificiale
- Sistema di metadati di record per l'azienda con un catalogo di cataloghi
- Estrazione automatica dei dati più dettagliati da un'ampia gamma di fonti dati, tra cui complessi sistemi aziendali
- Possibilità di trovare risorse dati attraverso una ricerca semantica avanzata simile a quella di Google
- Identificazione e comprensione delle risorse dati con una vista olistica che include lineage, viste delle relazioni e statistiche sulla profilazione dei dati e scorecard sulla qualità
- Identificazione di domini ed entità con intelligent curation
- Arricchimento delle risorse dati con annotazioni, valutazioni e recensioni governate e generate tramite crowdsourcing
- Associazione automatica dei termini di business glossary con risorse dati tecniche
- API aperte da integrare nel tuo ambiente e utilizzare per esporre ovunque metadati intelligenti

Libera il potenziale dei dati con un intelligent data catalog

I dati sono la linfa vitale della nostra economia e le aziende basate sui dati riescono a trasformare le loro risorse dati in fatturato e profitti. Il primo passaggio per qualsiasi iniziativa di digital transformation basata sui dati consiste nel gestire i dati come risorse aziendali, cioè eseguendone l'inventario, valutandone il valore e ottenendo il massimo dal loro utilizzo, proprio come con tutti gli altri importanti investimenti operativi e di capitale.

I dati sono di vario tipo e vengono distribuiti in diversi dipartimenti, applicazioni, data warehouse e data lake (on-premise o nel Cloud); di conseguenza è difficile sapere esattamente quali dati hai e dove risiedono. E con il proliferare delle fonti dati, il panorama si fa ancora più complesso.

Informatica® Enterprise Data Catalog è un data catalog basato su intelligenza artificiale che fornisce un discovery engine basato su machine learning per rilevare e catalogare le risorse dati nell'intera azienda (on-premise e nel Cloud) e i Big Data ovunque si trovino. Enterprise Data Catalog è alimentato dal motore CLAIRE®, che fornisce intelligence utilizzando i metadati per fornire consigli e suggerimenti, oltre ad automatizzare i task di gestione dei dati. Questo garantisce una maggiore produttività agli utenti IT, mentre gli utenti business possono collaborare appieno nella gestione e nell'utilizzo dei dati.

Informatica Enterprise Data Catalog offre a data analyst e utenti IT una potente ricerca semantica e facet dinamici per filtrare risultati della ricerca, lineage dei dati, statistiche di profilazione, scorecard sulla data quality, viste olistiche delle relazioni, consigli relativi alla somiglianza dei dati e un business glossary integrato.

Le funzionalità di collaborazione utilizzano competenze in materia e social curation, insieme alla potenza dell'intelligenza artificiale, per guidare la user experience e automatizzare la data curation. Ora puoi gestire in modo semplice ed efficiente le risorse dati di livello enterprise per ottenere il massimo del valore in tutta l'azienda. Gli utenti possono trovare rapidamente i dati e gestire in modo semplice il ciclo di vita di termini di business, definizioni, dati di riferimento e così via.

Informatica Enterprise Data Catalog è un data catalog basato su intelligenza artificiale che fornisce un discovery engine basato su machine learning per rilevare e catalogare le risorse dati nell'intera azienda (on-premise e nel Cloud) e i Big Data ovunque si trovino.

Funzionalità principali

API di metadati per l'integrazione nel tuo ambiente

Informatica Enterprise Data Catalog include API REST che permettono l'integrazione nel tuo ambiente e l'utilizzo dei contenuti del catalog dove preferisci. Le organizzazioni possono condividere qualsiasi tipo di intelligent metadata (applicazioni, report BI e dashboard) con gli utenti business. Gli utenti possono esportare e condividere determinati contenuti del catalog e metadati di enrichment associati.

Estensione Tableau di Enterprise Data Catalog per analytics self-service governati

L'estensione Tableau per Informatica Enterprise Data Catalog fornisce analytics agili e self-service con dati governati. Consente agli utenti di accedere a tutte le risorse di Informatica Enterprise Data Catalog dall'interfaccia utente nativa di Tableau. Gli utenti business che impiegano report Tableau possono utilizzare una barra di ricerca intelligente per trovare risorse dati certificate da data steward, inoltre possono accedere a contesto tecnico e di business, data quality, statistiche di profilazione e consigli di collegi. Gli utenti possono anche collaborare alle risorse dati con i colleghi attraverso recensioni, valutazioni, domande e risposte, rimanendo all'interno dell'interfaccia di Tableau.

Ricerca semantica con facet intelligenti

È possibile eseguire la ricerca e la discovery dei più importanti set di dati per l'analisi utilizzando la potente ricerca semantica con facet intelligenti. La ricerca avanzata delle keyword con corrispondenza token trova le risorse dati più rilevanti nel catalogo. La ricerca semantica viene applicata anche a domini di dati ricavati in modo da accertarsi di sottoporre a discovery tutte le risorse dati. I facet intelligenti basati sui risultati della ricerca consentono agli utenti di modificare la ricerca, limitandola ai set di dati di interesse.

Provisioning self-service dei dati

Dopo aver trovato il set di dati pertinente per la tua analisi, puoi spostarlo facilmente alla destinazione che preferisci con un semplice provisioning click-through da Informatica Enterprise Data Catalog. Puoi scegliere tra un'ampia gamma di fonti e destinazioni, tra cui Amazon Redshift, Azure Synapse Analytics (precedentemente noto come Azure SQL Data Warehouse), Google BigQuery, Snowflake e strumenti BI come Tableau. Questa funzionalità utilizza l'integrazione di Informatica Enterprise Data Catalog con Informatica Cloud Data Integration.

Lineage dei dati e analisi dell'impatto

Puoi tracciare in modo interattivo l'origine dei dati attraverso viste del lineage a qualsiasi livello, da vista business a livello di sistema che evidenziano gli endpoint fino ad arrivare a viste dettagliate che comprendono tutti i dettagli complessi delle viste intermedie. Una vista dettagliata del lineage espande il percorso per mostrare lineage dettagliate a livello di colonne e metriche. Gli utenti possono eseguire un'analisi dettagliata dell'impatto sulle risorse dati upstream e downstream.

Discovery olistica delle relazioni

Una vista olistica dei dati in un knowledge graph permette di cercare, eseguire la discovery e capire rapidamente i dati di livello enterprise e le relazioni significative tra i dati. È possibile eseguire la discovery automatica delle relazioni tecniche, di business, semantiche e basate sull'utilizzo dei set di dati. La vista olistica dei dati mostra gli insiemi di dati correlati, le tabelle, le viste, i domini dei dati, i report e gli utenti. Questo aiuta la discovery progressiva di altri set di dati di interesse.

Classificazioni automatizzate con identificazione intelligente di domini ed entità

Classifica e identifica automaticamente domini ed entità, come cliente, prodotto e ordine, in tutte le risorse dati strutturate e non strutturate a livello di campo, colonna e tabella. Si tratta di un passaggio cruciale nella possibilità per le aziende di catalogare, gestire e trarre valore dalle proprie risorse dati. I dati classificati consentono di effettuare ricerche, filtrare i risultati e trovare consigli nel business glossary in modo più efficiente. Informatica offre oltre 60 domini di dati pacchettizzati ad esempio email, numero di carta di credito, numero di previdenza sociale, paese, città, URL e nome dell'azienda. Gli utenti possono anche aggiungere i propri domini personalizzati. Le risorse dati possono essere classificate utilizzando regole dei dati (ovvero, colonne con dati che corrispondono a una specifica logica definita nella regola) o regole per i nomi delle colonne (colonne che corrispondono alla logica del nome delle colonne definita nella regola).

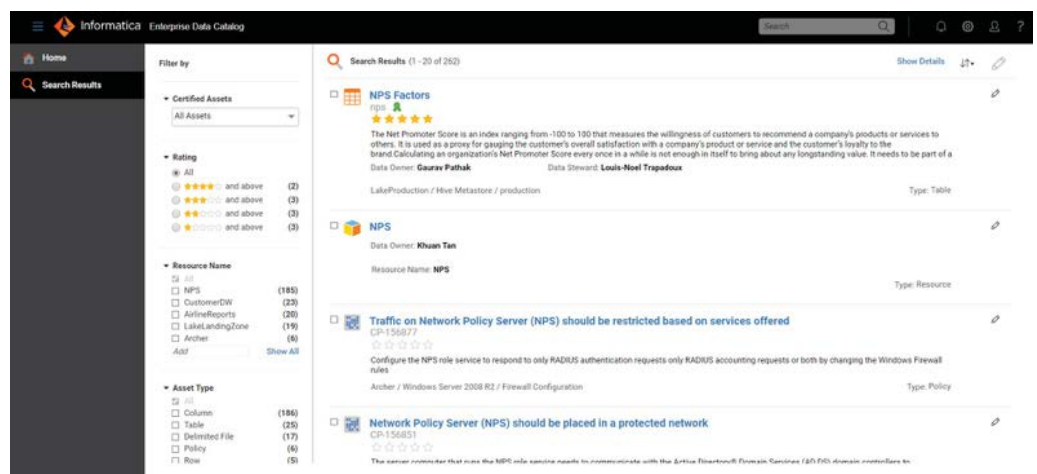


Figura 1: Rapida ricerca dei set di dati con ricerca semantica intelligente e facet dinamici. Vista di valutazioni e set di dati certificati.

Collaborazione e social curation

Informatica Enterprise Data Catalog consente a data analyst e data scientist di trovare facilmente i dati più rilevanti e affidabili per gli analytics impiegando la potenza combinata di intelligenza artificiale, esperienza umana e collaborazione. I set di dati possono essere certificati da data owner ed esperti in materia. I data consumer possono fornire valutazioni e recensioni per i set di dati, consentendo la social curation dei dati.

Gli utenti possono seguire i set di dati di loro interesse e ricevere notifiche in merito alle modifiche, mentre una piattaforma per domande e risposte consente a esperti in materia di rispondere alle domande più frequenti poste dagli utenti. Inoltre, gli utenti possono aggiungere annotazioni e attributi personalizzati ai set di dati, migliorando ulteriormente la collaborazione business-IT e i risultati di ricerca.

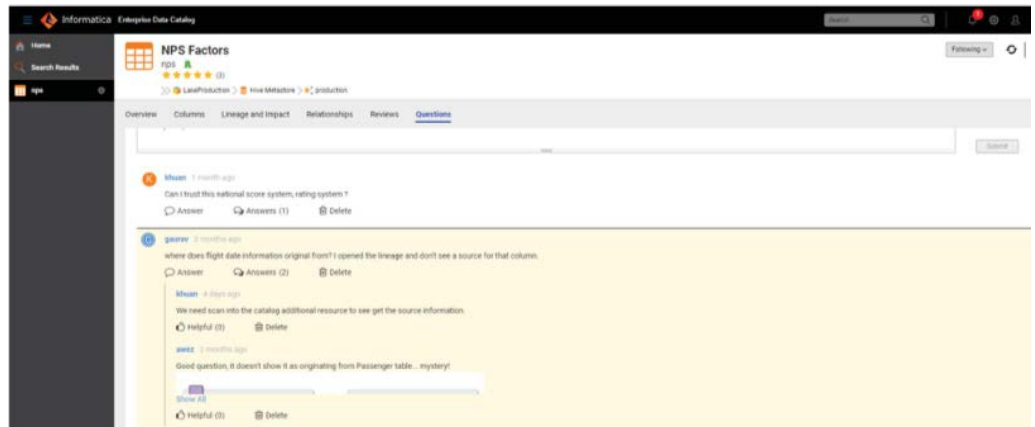


Figura 2: Consentire la collaborazione attraverso funzionalità di domande e risposte.

Data quality integrata

Puoi vedere statistiche di profilazione dei dati, regole di data quality, scorecard e gruppi di metriche insieme ai metadati tecnici per comprendere la qualità delle risorse dati per le analisi. Le statistiche di profilazione comprendono distribuzione di valore, modelli e tipi di dati e inferenza dei domini di dati.

Associazione automatica dei termini di business glossary

Informatica Enterprise Data Catalog permette di importare facilmente le risorse di business glossary come termini, policy e classificazioni da Informatica Axon™. Aggiungi ai dati un contesto di business dettagliato associando termini di business ai giusti metadati tecnici. Informatica Enterprise Data Catalog permette di associare automaticamente i termini di business con metadati tecnici, contribuendo a eliminare un noioso processo manuale per la data governance, permettendo agli steward di business e IT di gestire in modo collaborativo i metadati di business che comprendono l'automazione efficiente del workflow umano. Informatica Enterprise Data Catalog supporta anche l'importazione di risorse di business glossary di Informatica Business Glossary e altri strumenti di terze parti.

Identificazione intelligente della somiglianza tra dati

Algoritmi statistici e di machine-learning avanzati identificano dati e subset di dati simili. Questa potente funzionalità aiuta gli utenti a trovare i dati più rilevanti e affidabili di cui hanno bisogno. Ad esempio, un analista del settore telecomunicazioni interessato all'analisi dell'abbandono dei clienti può eseguire query di dati contenenti l'attività dei clienti pre-pagati per il trimestre corrente. Informatica Enterprise Data Catalog è in grado di consigliare una versione dei dati più pulita (dati sostituiti), dati contenenti l'attività del cliente per il trimestre precedente (dati unificabili) e una tabella con i dettagli del cliente per arricchire il set di dati (dati abbinabili).

Connettività universale dei metadati con scanner avanzati

Estrai metadati da qualsiasi tipo di fonte dati nell'azienda, ad esempio database, Data Warehouse, Cloud data lake, strumenti BI, cluster Hadoop, NoSQL e sistemi enterprise complessi tra cui sistemi legacy e mainframe, strumenti ETL multi-vendor, SQL dialect e varie applicazioni enterprise in ambienti ibridi e multi-Cloud.

Con gli scanner avanzati di Enterprise Data Catalog puoi ispezionare visivamente ogni script e procedura per comprenderne appieno la logica e il flusso di dati interno. Puoi ottenere un lineage di dati completo a livello di colonna, incluso un inventario completo di tutte le potenziali fonti di lineage ricche di dettagli. Gli scanner consentono di eseguire la scansione di codici statici e dinamici, nonché il parsing del linguaggio per ottenere un lineage di dati automatizzato.

Seguono alcuni esempi di fonti dati supportate per l'estrazione dei metadati:

- **Database/Data Warehouse:** Oracle, Microsoft SQL Server, Script SQL, Sybase ASE, IBM Netezza, Teradata, JDBC, SAP HANA, SAP BW, SAP BW/4HANA, stored procedure
- **Big Data:** Cloudera Navigator, Hive (Cloudera/Hortonworks/MapR/IBM BigInsights/EMR), HDFS, Hortonworks Atlas, Cassandra, MongoDB, Kafka
- **Mainframe:** DB2 z/OS, DB2 i5/OS, COBOL, JCL
- **BI e analytics:** SAP BusinessObjects, Tableau, Microsoft Power BI, Cognos, MicroStrategy, OBIEE, QlikView, Qlik Sense, Microsoft SSRS & SSAS, SAS
- **ETL:** Informatica PowerCenter, Informatica Data Engineering Integration, Informatica Intelligent Cloud Services, Informatica Data Integration Hub, Microsoft SSIS, Oracle Data Integrator, IBM InfoSphere DataStage, AWS Glue
- **Business glossary:** Informatica Axon Data Governance, Informatica Business Glossary
- **Modellazione:** Erwin Data Modeler, SAP PowerDesigner
- **Applicazioni di livello enterprise:** Salesforce, Oracle, Workday, Informatica MDM, SAP ECC
- **File system:** Microsoft SharePoint, Microsoft OneDrive, file system Windows/Linux
- **Formati file:** MS Excel, MS Word, MS PowerPoint, Adobe PDF, file flat, CSV, delimitati, XML, JSON, Avro, Parquet
- **Piattaforme Cloud:** AWS S3, AWS Redshift, Azure SQL DB, Azure SQL DW, Azure ADLS, Azure ADLS Gen 2, BLOB di Azure, Google Cloud Storage, Google BigQuery, Snowflake

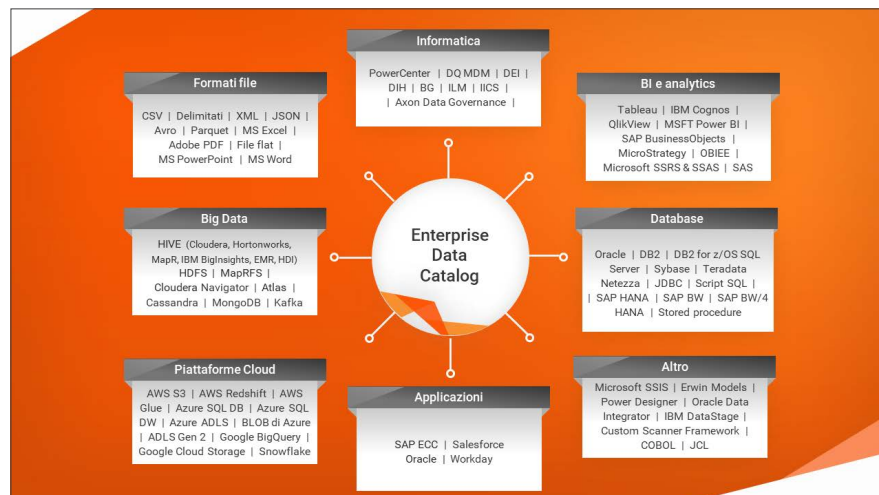


Figura 3: Informatica Enterprise Data Catalog supporta la connettività universale dei metadati.

Sicurezza a livello di risorsa

È possibile concedere autorizzazioni di lettura/scrittura per utenti e gruppi a livello di risorsa per consentire agli utenti di visualizzare o modificare gli attributi personalizzati, eseguire la domain curation e associare i termini del business glossary.

Implementazioni a livello enterprise

Informatica Enterprise Data Catalog è progettato per le implementazioni a livello di Big Data e offre la possibilità di scansionare decine di milioni di set di dati in centinaia di fonti dati. Supporta l'ingestion parallela dei metadati e l'indicizzazione distribuita ad alta velocità per caricare rapidamente i contenuti del catalog e fornire performance di ricerca senza paragoni e disponibilità elevata con tolleranza agli errori per implementazioni 24x7. Con la profilazione dei dati basata su Spark, puoi profilare enormi quantità di dati su scala, in modo da ottenere una maggiore comprensione dei dati enterprise.

Amministrazione unificata

Gestione e monitoraggio di risorse di catalogo, programmi di estrazione di metadati, profilazioni e molto altro da un'unica console di amministrazione unificata. Una dashboard di controllo dei processi offre widget per il monitoraggio delle attività e viste delle risorse. Avvisi via email assistono gli amministratori per la risoluzione proattiva dei problemi di catalogo.



Figura 4: Comprensione dei dati con viste olistiche delle relazioni tra dati.

Vantaggi

Catalogazione intelligente di tutti i tipi di dati nell'azienda

Informatica Enterprise Data Catalog rileva in modo intelligente molti tipi di dati e le loro relazioni all'interno dell'azienda. Gli scanner predefiniti raccolgono i metadati da database, Data Warehouse, Data Lake, archivi dati Cloud, applicazioni, strumenti BI, strumenti ETL, catalog di metadati di terze parti, NoSQL e altro ancora. Tutti i metadati sono indicizzati e catalogati in un database grafico altamente scalabile progettato per garantire aggiornamenti rapidi, ricerche intelligenti e query veloci.

Informazioni su Informatica

La Digital Transformation cambia le nostre aspettative: migliori servizi, consegne più rapide, il tutto a costi più contenuti. Le aziende devono trasformarsi per rimanere competitive e i dati sono la risposta per riuscirci.

Siamo leader mondiali nell'Enterprise Cloud Data Management e siamo pronti a supportarti in qualsiasi settore, categoria o nicchia di mercato. Informatica ti offre la possibilità di operare in un ambiente più agile, realizzare nuove opportunità di crescita o persino inventare cose nuove. Siamo focalizzati al 100% sui dati e questo ti offrirà la flessibilità necessaria per competere ed avere successo.

Ti invitiamo a scoprire tutto quello che Informatica ha da offrirti e a liberare "the power of data" per promuovere la tua prossima intelligent disruption.

Con la crescente creazione e la propagazione dei dati in tutta l'azienda, emergono inevitabilmente set di dati simili e duplicati. Informatica Enterprise Data Catalog utilizza algoritmi statistici e di machine learning avanzati per eseguire la discovery di dati e insiemi di dati simili, aiutando gli utenti a trovare i dati più affidabili e rilevanti di cui hanno bisogno.

Possibilità di trovare rapidamente risorse dati attraverso una ricerca semantica avanzata simile a quella di Google

Talvolta la ricerca dei dati necessari in centinaia di sistemi aziendali diversi può sembrare inutile. Solo attraverso una ricerca semantica avanzata basata su un'intelligence di metadati completa e su un'infrastruttura scalabile diventa possibile trovare dati rilevanti. Informatica Enterprise Data Catalog offre una ricerca semantica con facet intelligenti per affinare ulteriormente i risultati della ricerca. Dal momento che Informatica associa in modo univoco metadati di business, tecnici e operativi, gli utenti business possono cercare termini di business per trovare i dati necessari e successivamente esplorare viste olistiche delle relazioni per trovare le risorse dati correlate.

Identifica e comprendi le risorse dati con viste olistiche delle relazioni e lineage

Nel caso della gestione dei dati, il detto "non è possibile gestire ciò che non si può misurare" è vero. Per trarre il massimo valore dai dati, è necessario capire di quali dati disponi, da dove provengono, come sono cambiati e qual è il loro grado di affidabilità. Informatica Enterprise Data Catalog risponde a tutte queste e ad altre domande con un lineage end-to-end completo di riepilogo e dettagliato, statistiche di profilazione, scorecard sulla data quality e viste olistiche delle relazioni, offrendo un quadro chiaro dei dati.

Arricchimento delle risorse dati con contesto di business e annotazioni sia gestite che generate tramite crowdsourcing

Informatica Enterprise Data Catalog ottiene il massimo del riutilizzo e del valore dei dati classificando automaticamente le risorse dati di livello enterprise fino al livello di campo/colonna. Per aumentare ulteriormente il valore dei dati, Informatica Enterprise Data Catalog acquisisce il contesto relativo a chi utilizza i dati e a quale scopo, oltre a tag, annotazioni, valutazioni e recensioni generati tramite crowdsourcing. Questa "saggezza popolare" agevola l'arricchimento e la curation dei dati, rendendoli ancora più preziosi per tutta l'azienda. Informatica Enterprise Data Catalog si integra con Informatica Axon per importare facilmente da Axon le risorse di business glossary, come termini, definizioni e policy di business. Questi metadati di business vengono associati automaticamente ai metadati tecnici e operativi, consentendo a business analyst, data steward e altri utenti di trovare rapidamente le risorse dati, capirle e collaborare.

Ulteriori informazioni

Per ulteriori informazioni su Informatica Enterprise Data Catalog, visita <https://www.informatica.com/it/products/data-catalog/enterprise-data-catalog.html>.

