

IBM Spectrum Discover

エクサバイト規模のファイルやオブジェクトの格納をサポートする最新のメタデータ管理ソフトウェア

Highlights

- イベント通知およびポリシー・ベースのワークフロー
- ストレージ消費量に関するきめ細かなビュー
- ペタバイト規模のデータを効率的に高速検索
- 基幹業務のビジネス・データを素早く区別
- ポリシー・ベースのカスタム・タグ付け
- ソフトウェア開発者キット (SDK) による拡張可能性

デジタル・トランスフォーメーションにより、非構造化データが急速に増大しています。最近の Forrester 社アナリティクスの調査によると、1 ペタバイト以上の非構造化データを扱う企業数は、2016 年の 13% から 2017 年には 38% に増加しています。¹ つまり、複数のペタバイトの非構造化データを扱う企業の割合が、1 年で 3 倍近くに増加したことを意味します。従って、ストレージ管理者が、非構造化データの急速な増大への対応に苦労していることは驚きではありません。そして多くの場合、問題への対応として、単純により多くのストレージを用意するという状況に陥っています。

ストレージ容量についても課題ですが、保管データへの可視性が制限されていることもまた、大規模な非構造化データのストレージ管理者とユーザーの両者にとっては一段と大きな課題になります。多くの場合、ストレージ管理者は、システム・メタデータだけでは、ストレージの消費量とデータ品質をきめ細かに確認できないと感じています。こうした情報を把握する機能(ビュー機能)は、効率的なストレージ最適化に必要なものです。また、基本的なシステムレベルのメタデータだけでは、作業に必要なデータの検索にかなりの時間を費やしているデータ・サイエンティストやビジネス・アナリスト、ナレッジ・ワーカーにとっても不十分です。さらに、データの管理責任者も、社外秘扱いや機密扱いのデータを含むファイルおよびオブジェクト (レコード) の識別に苦労しています。

こうした非構造化データの課題を克服するために、大企業では、卓越したデータ可視性を提供するメタデータ管理ソリューションに注目しています。自らの非構造化データを明確に理解した組織は、ストレージ・システムの最適化、リスクの軽減、非構造化データの価値活用により、競争上の優位性と重要なデータ・インサイトを得ることができます。

非構造化データの経済性、ガバナンス、分析、ディスクバリーの改善

IBM Spectrum Discover は、最新のメタデータ管理ソフトウェアであり、エクサバイト規模の非構造化ストレージに対してデータ・インサイトを提供します。IBM Spectrum Discover は、IBM Cloud Object Storage System および IBM Spectrum Scale に簡単に接続して、数十億個ものファイルとオブジェクトのメタデータを迅速に取り込み、統合し、索引を作成します。これにより、これらのストレージ・ソース上に豊富なメタデータ層が形成されます。このメタデータにより、データ・サイエンティストやストレージ管理者、データ管理責任者は、膨大な量の非構造化データから効率的に管理、分類、そして知見を取得することができます。こうして得られた知見により、大規模分析の加速化、ストレージ経済性の向上、リスク軽減への支援が実現し、競争上の優位性が創出され、重要な研究が迅速に実施できます。

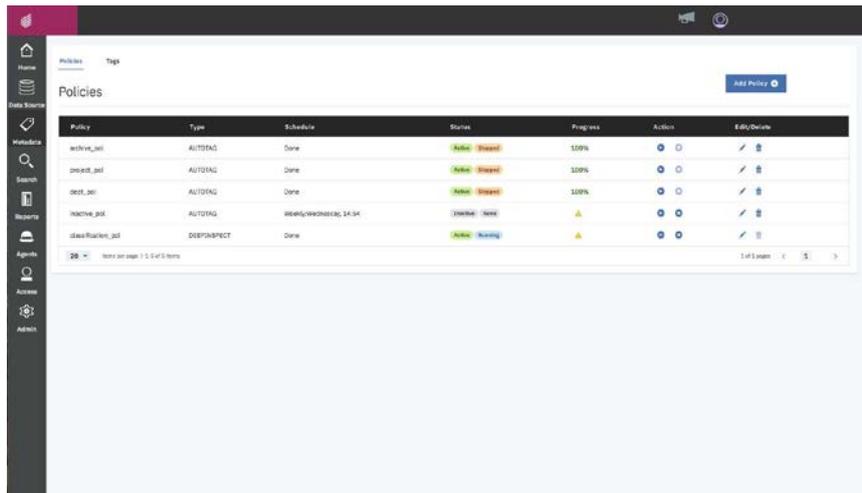
IBM Spectrum Discover には以下のような特長があります。

- イベント通知およびポリシー・ベースのワークフローにより、エクサバイト規模でメタデータの取り込みおよびメタデータの索引付けを自動化
- 幅広いシステムとカスタム・メタデータに基づきストレージ消費量をきめ細かに確認できるビュー。
- エクサバイト規模のデータを効率的に高速検索し、非常に関連性の高い検索結果を出して大規模分析を実現。
- 基幹業務のビジネス・データを、削除可能なデータや安価なコールド層のデータから素早く区別する機能。
- ポリシー・ベースのカスタム・タグ付けにより、組織はデータの分類化とカテゴリー化を行い、そのデータをビジネスのニーズに連動可能。
- ソフトウェア開発者キット (SDK) を利用してアクション・エージェントを構築。このアクション・エージェントは、ファイルのヘッダーとコンテンツからメタデータを抽出し、データ移動を自動化し、Apache Spark、Apache Tika、PyTorch、Caffe、TensorFlow などのオープン・ソース・ソフトウェアとの統合を提供するもので、データ識別を円滑にし、大規模なデータ処理を高速化。

ポリシー・ベースのメタデータ・タグ付けにより、データ分類を細分化

IBM Spectrum Discover は、ソースのストレージ・システムからシステム・メタデータを取り込み、検索結果からカスタム・メタデータを作成し、IBM Spectrum Discover のアクション・エージェント API を使用してファイルのヘッダーとコンテンツからキーワード・メタデータを抽出することを自動的に実行します。その結果、1つの一元管理ソリューションで管理される豊富なファイル/オブジェクト・メタデータからなるレイヤーが形成されます。ポリシー・ベースのメタデータ・タグ付け機能を以下に示します。

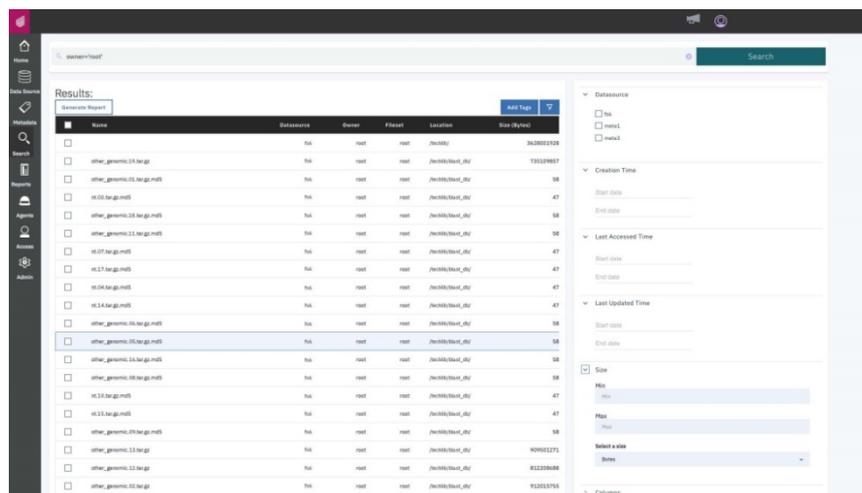
IBM Spectrum Discover では、ポリシーは、レコード・メタデータを強化するアクションの自動化に使用します。ユーザーはポリシーを任意のレコード・セットに適用でき、アクションを構成できます。例えば、ストレージ管理者は、部門と連携して古いデータをアーカイブすることができます。そのために、ソリューションのポリシー・エンジンを使って検索機能を利用し、部門(例:マーケティング部)が所有するレコードを検索し、例えば1年以上など、指定期間の間アクセスされていないレコードを検索します。次に、事前定義された「アーカイブ」タグをドロップダウン・リストから選択します。そのアーカイブ・タグが自動的に関連ファイルのグループに適用されます。ポリシーは、一回限りのイベントとして実行することも、定期的に行うようにスケジュールすることもできます。ポリシーで作成された新規タグはすべて索引付けされ、検索可能になります。



IBM Spectrum Discover provides an intuitive user interface for policy-based metadata tagging.

数十億個のメタデータ・タグを高速検索することにより、データ資産のディスカバリーを迅速化

IBM Spectrum Discover には、検索バーと詳細検索ペインの両方が用意されており、索引付けされたレコードのサブセットをユーザーが素早く検出するのに役立ちます。検索結果は、検索条件に関する情報が含まれているカラム型テーブルに表示されます。ユーザーが画面で確認できること、確認できないことは、役割ベースのアクセス制御によって決定されます。

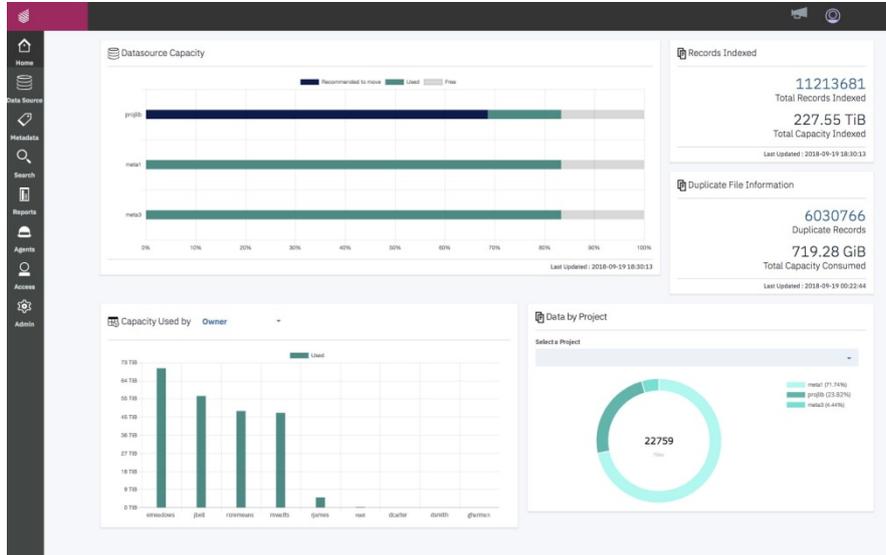


IBM Spectrum Discover search results are displayed in a columnar table that contains information correlated to search criteria.

SQL 構文に精通しているユーザーは、検索バーに検索文字列を入力できます。または、IBM Spectrum Discover が提供する、事前定義された選択ボックスを使ってレコードをフィルタリングする使いやすい検索ペインを利用できます。例えば、「File system」選択ボックスを使用すると、ユーザーは1つ以上のソース・ストレージ・システムを選択できます。「Time」セクターを使用すると、ユーザーは、最後にレコードがアクセスされた時刻に基づいて時間範囲を指定できます。「Size」セクターを使用すると、ユーザーは、ファイルの最小サイズまたは最大サイズ、あるいはその両サイズに基づいてレコードを特定できます。これらの検索機能やその他の検索機能を利用することで、ニーズに最も適した組み合わせの検索ボックスを使用できるようになります。

レコードを視覚化するためのダッシュボードとカスタマイズ可能なレポート作成

IBM Spectrum Discover ダッシュボードは、ユーザーの環境を一目で確認できます。ユーザーが画面で確認できること、確認できないことは、役割ベースのアクセス制御によって決定されます。このダッシュボードには、IBM Spectrum Discover クラスターで索引付けされたレコードの情報をグラフィカルに表示するウィジェットが含まれているため、ユーザーはデータ環境を可視化することができます。例えば、ダッシュボードでは、登録済みストレージ・システムの使用量と容量、潜在的な重複ファイルに関する情報、プロジェクトや部門がどれくらいの容量を使用しているかの内訳などを示すことができます。



IBM Spectrum Discover Dashboard includes widgets that show a user's environment at a glance.

追加のレコード詳細を必要とするユーザーの場合、IBM Spectrum Discover はカスタマイズ可能なレポートを提供します。要約レポートと詳細レポートの両方を生成することができます。要約レポートでは、次のような各種条件に基づいて、レコード数やレコード容量などの情報が集約され、グループ化されます。例えばその条件とは、オブジェクト・ポリシー、ファイル・システム、ユーザーなどです。詳細レポートでは、レポートのフィルター条件に一致する、システム内の各レコードの詳細情報が表示されます。

1 「Global Business Technographics Data And Analytics Survey, 2017」(2017年グローバル・ビジネス・テクノグラフィックス・データおよび分析に関するアンケート調査)、Forrester 社、2017年6月。

<https://www.forrester.com/Global+Business+Technographics+Data+And+Analytics+Survey+2017/-/E-sus3671#>

IBM Spectrum Discover の概要

単一ノード	
メモリー	228 GB
CPU	16 コア
ストレージ	
1次プール	3 TB のローカル SSD または 3 TB の SAN 接続 SSD/フラッシュ・ストレージ
バックアップ・プール	3 TB
ネットワーク	1 Gb イーサネットまたは 10 Gb イーサネット
マルチノード (3 ノード)	
メモリー	256 GB
CPU	24 コア
1次プール	15 TB のローカル SSD または 15 TB の SAN 接続 SSD/フラッシュ・ストレージ
バックアップ・プール	15 TB
ネットワーク	10 Gb イーサネット
ソフトウェア前提条件	
VMware ESXi 6.5	

IBM Spectrum Discover の機能

継続的なメタデータの取り込み	<ul style="list-style-type: none"> • 組み込みコネクタが、IBM クラウド・オブジェクト・ストレージおよび IBM Spectrum Scale との緊密な統合を提供 • イベント通知により、継続的なメタデータの取り込みを自動化 • メタデータの索引付けにより、データ照会を迅速化
系統的なメタデータ・キュレーション	<ul style="list-style-type: none"> • ポリシー主導のワークフローにより、カスタム・ラベル付けを自動化 • カスタム・データ・ラベルにより、大規模分析時のデータを正確に特定 • システム・データ・ラベルとカスタム・データ・ラベルをリンクする機能により、ストレージの最適化を加速
リアルタイムのデータ・インサイト	<ul style="list-style-type: none"> • 高速検索により、関連性の高いファイルとオブジェクトを秒単位で特定 • ドリルダウン機能を備えたチャート・エレメントを有するダッシュボードにより、ストレージ管理をシンプル化 • カスタマイズ可能なレポートにより、監査とコミュニケーションを円滑化
セキュアで拡張性のあるアーキテクチャー	<ul style="list-style-type: none"> • 役割ベースのアクセス制御により、承認されたデータ・アクセスのみを許可 • アクション・エージェントの API により、顧客が開発したソフトウェアやサードパーティ製のソフトウェア、あるいはその両方のソフトウェアとの統合をサポート • ポリシー・エンジン・フックにより、ワークフローの自動化を実現

Why IBM?

IBM は、データ・ストレージ製品の業界をリードするプロバイダーとして、ストレージの経済性、データ品質、データ・ガバナンス、大規模アナリティクスのデータ識別を改善するデータ管理ソリューションに投資しています。IBM Spectrum Discover は、IBM が提供する総合的なデータ管理の利点における重要な要素であり、IBM Cloud Object Storage System および IBM Spectrum Scale のお客様に対して、可視化と分類化をもたらす強力なメタデータ管理ソフトウェアを提供します。

詳細情報

IBM Spectrum Discover の詳細については、日本 IBM 営業担当者または IBM ビジネス・パートナーにお問い合わせいただくか、次の Web サイトをご覧ください。

<https://www.ibm.com/jp-ja/marketplace/spectrum-discover>



© Copyright IBM Corporation 2019.

IBM, the IBM logo, and ibm.com are trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the Web at <https://www.ibm.com/legal/us/en/copytrade.shtml>, and select third party trademarks that might be referenced in this document is available at https://www.ibm.com/legal/us/en/copytrade.shtml#section_4.

All statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.