

Module 1

Introduction

هنا بنشرح أهمية Protect data وإيه اللي ممكن يحصل لو مش محمية و ليه نحميها

The Informatica Data Privacy portfolio helps

1-organizations protect their data in a constantly changing environment.

حماية الداتا في بيئات دايمًا متغيره

2- Our data privacy solutions are designed to help you discover personal: صممت حلول للمساعدة في:

اتكشافات البيانات الشخصية

- sensitive data والبيانات الحساسه
- understand data movement فهم حركة البيانات
- link identities ID ربط ال
- analyze risk تحليل المخاطر
- and remediate problems with A1-driven automation. معالجة المشاكل باستخدام A1

1- Identify all **sensitive data** that exists in relational databases, cloud applications, and files systems.

2- Identify where it is in your environment. You might not know where sensitive data exists in all applications.

3- Identify the quantity of sensitive data that resides in your environment.

4- Identify how it moves in your organization. After you discover individual table columns that contain sensitive data

5- Identify if you have groupings of sensitive data as defined by the data security standards.

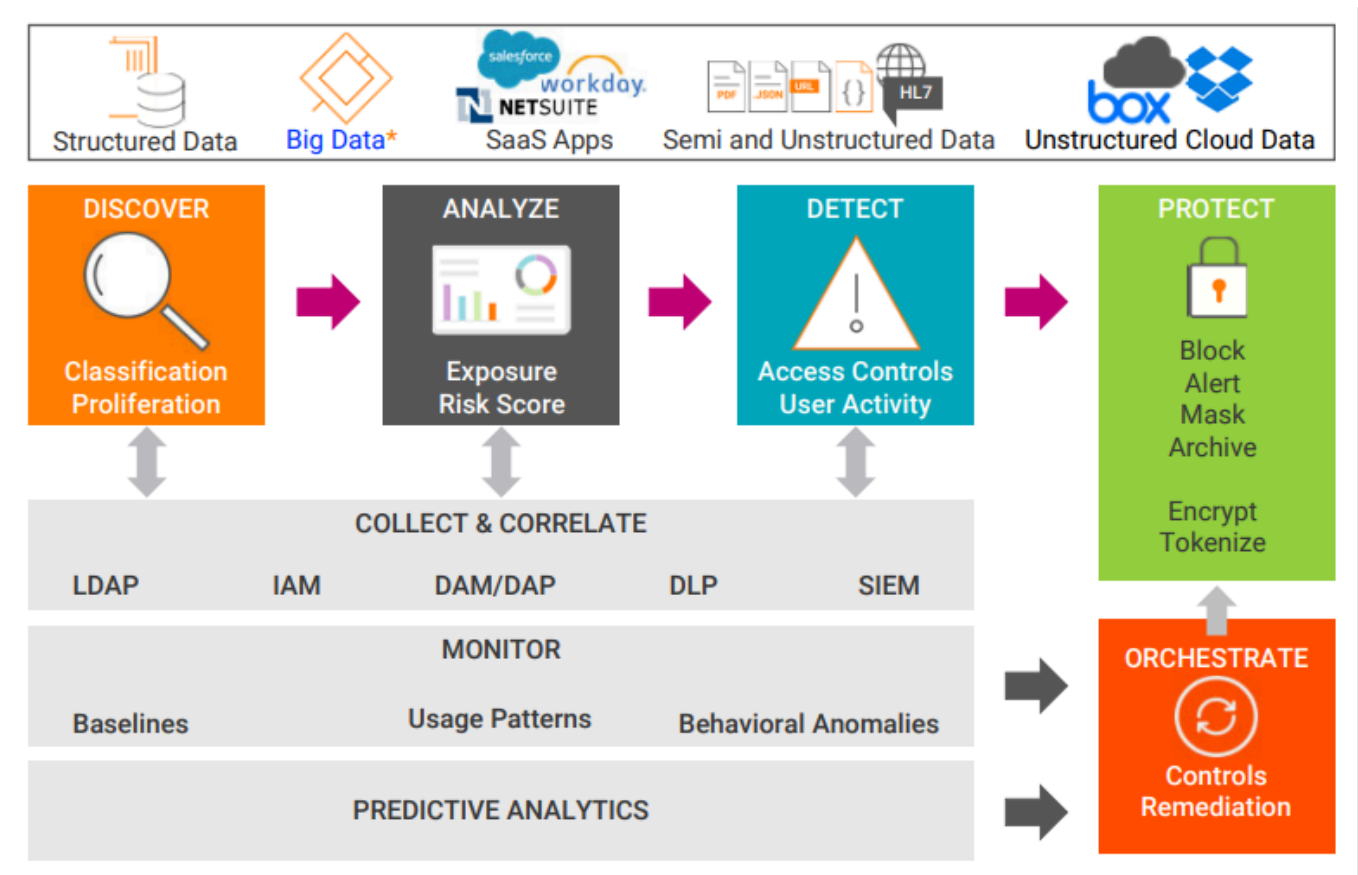
6-Identify if sensitive data is protected.

7- Identify how many and which users have access to sensitive data. You can determine the type of information the user accessed. You can also detect any unusual patterns of access such as increased activity on sensitive data during non-business hours.

8- Identify the security standards or policies governing it, if any.

9- Quantify the cost and the risk of sensitive data. The dashboard displays the risk cost and risk score of sensitive data. The risk cost is the total monetary loss that the organization might have to bear if sensitive data is lost, exposed to an unauthorized user, or becomes unavailable for use in ongoing operations.

Benefits of DPM



1- Discover, classify, and analyze sensitive data: Gain visibility across available data platforms and types to better understand risk exposure.

اكتشف البيانات الحساسة، وصنّفها، وحللها: اكتسب رؤية شاملة عبر منصات وأنواع البيانات المتاحة لفهم التعرض للمخاطر بشكل أفضل.

2-Analyze and track data risk: Perform continuous risk analysis of personal and sensitive data exposure.

حلل مخاطر البيانات وتتبعها: أجر تحليلًا مستمرًا لمخاطر التعرض للبيانات الشخصية والحساسة.

3- Map identities to data: Index, inventory, and look up data subjects and identities for transparency into access and use.

ربط الهويات بالبيانات: فهرسة، وجرد، والبحث عن أصحاب البيانات وهوياتهم لضمان الشفافية في الوصول والاستخدام.

4- Protect data, respond to access requests, and monitor activity for risk remediation:
Protect data from unauthorized access and manage data subject access requests with automated orchestration.

حماية البيانات، والاستجابة لطلبات الوصول، ومراقبة الأنشطة لمعالجة المخاطر:
حماية البيانات من الوصول غير المصرح به، وإدارة طلبات وصول أصحاب البيانات من خلال التنسيق الآلي.

-
- A single pane of glass to continuously Monitor sensitive data stores and their risks
منصة واحدة لمراقبة مخازن البيانات الحساسة ومخاطرها بشكل مستمر.
 - Enterprise-wide sensitive data risk analytics
تحليل مخاطر البيانات الحساسة على مستوى المؤسسة.
 - Classification & Discovery
التصنيف والاكتشاف.
 - Proliferation analysis
تحليل الانتشار.
 - User access to sensitive data
وصول المستخدم إلى البيانات الحساسة.
 - User activity on sensitive data
نشاط المستخدم على البيانات الحساسة.
 - Policy-based alerting
التنبيهات المستندة إلى السياسات.
 - Multi-factor risk scoring
تقييم المخاطر متعدد العوامل.
 - Identification of the highest risk areas
تحديد المناطق الأكثر عرضة للخطر.
 - User behavior analysis
تحليل سلوك المستخدم.
-

1-User Access and User Activity

Data Visibility — How many users and who has access to sensitive data

User access — Add users, groups, and access information imported from LDAP (Active Directory, IBM Tivoli)

Collect and correlate user activities against sensitive data Activity logs can be imported from the database, Hadoop audit logs, or DAM solutions in CEF, LEEF or JSON formats

The risk model and score incorporate user access and activity



2- Proliferation with Enhanced Performance

بمعني الانتشار مع تحسين الاداء

يدعم DPM في Informatica توسع البيانات الحساسة (Data Proliferation) عبر أنظمة مختلفة، مع تحسين الأداء والقدرة على التوسع باستخدام تقنيات Big Data.

المهام الأساسية

- **Discovery & Classification**

- تلقائيًا (PII, PCI, PHI) يكتشف البيانات الحساسة DPM

- يُصنف البيانات بناءً على قواعد وأنماط محددة مسبقًا.
- يدعم مصادر متنوعة مثل قواعد البيانات، ملفات، Salesforce، وHive.

دعم مصادر البيانات

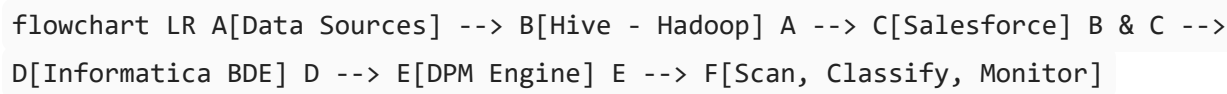
المصدر	الشرح
Hive	يدعم اكتشاف البيانات داخل مستودعات Hadoop مثل Cloudera و Hortonworks.
Salesforce	يمكن لـ DPM الوصول إلى البيانات السحابية وتحليل الحقول الحساسة في Salesforce.

Proliferation

تعني تتبع البيانات الحساسة أثناء انتقالها أو انتشارها عبر الأنظمة المختلفة:

- من قواعد بيانات إلى Hadoop.
- من Salesforce إلى تقارير الأعمال.
- من ملفات Excel إلى مستودعات البيانات.

التكامل مع Hadoop و BDE



- **Informatica Big Data Edition (BDE):** Hadoop على بيئة DPM يُستخدم لتشغيل عمليات.
- يدعم التكامل مع Hive لتحديد البيانات الحساسة ضمن جداول Hadoop.

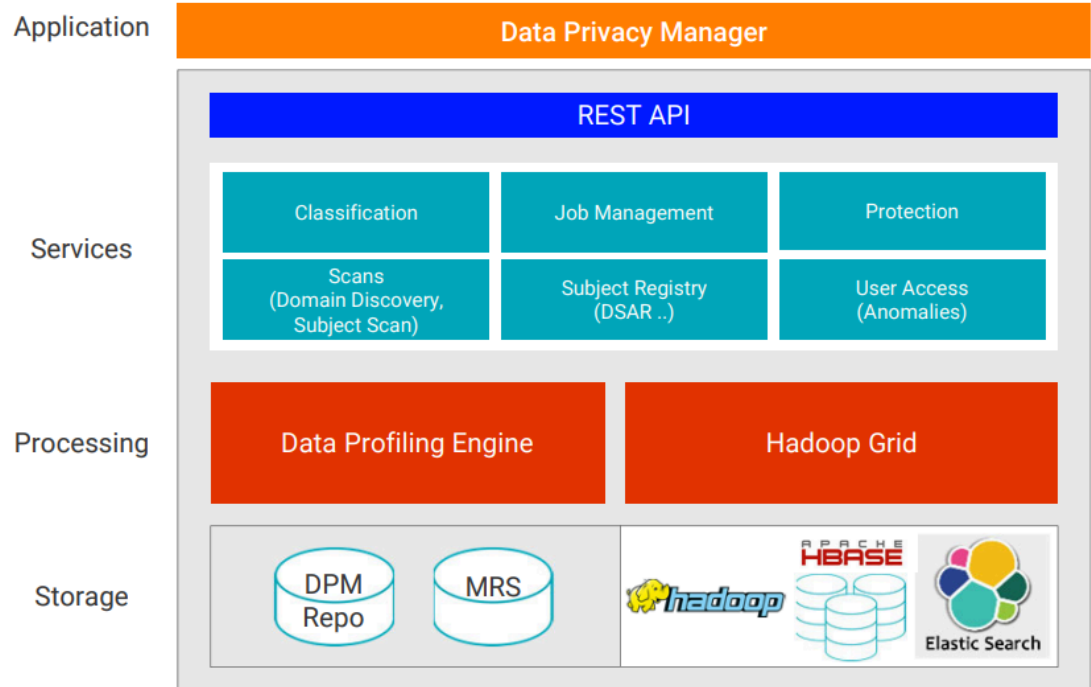
الأداء والتوسع

- **Scalability & Performance:**
 - يدعم أنظمة ضخمة بأداء عالي.
 - يستخدم تقنيات مثل **Grid computing** لتوزيع الحمل.
- **Grid Enabled:**
 - يسمح بتوزيع المهام على عدة عقد (Nodes) لتحسين السرعة والكفاءة.
 - يُفيد في البيانات الكبيرة حيث توجد بيانات ضخمة ومعقدة.
- **Leveraging Big Data Platforms:**
 - يستفيد DPM من منصات Big Data لزيادة السرعة والتغطية.
 - يعالج ملايين السجلات بكفاءة.

DPM Architecture

DPM Stack

DPM Application Stack



تتكون معمارية DPM من أربع طبقات أساسية:

1. Application Layer

- **Informatica Domain** يعمل داخل نطاق DPM.
- يستخدم **REST API** للتواصل بين واجهة المستخدم والـ Backend.
- واجهة الاستخدام تعتمد على المتصفح.

2. Services Layer

تشمل مجموعة من الخدمات المركزية:

الوظيفة	الخدمة
تصنيف البيانات الحساسة حسب السياسات	Classification

الخدمة	الوظيفة
Job Management	إدارة وتنفيذ ومراقبة المهام
Protection	تنفيذ سياسات حماية البيانات
Scans	تشغيل عمليات الفحص (Scanning)
Subject Registry	إدارة Object واستخدامها في DSAR
User Access & Anomalies	مراقبة سلوك المستخدمين واكتشاف الحالات الشاذة

3. Processing Layer

- محرك المعالجة الرئيسي هو **Data Profiling Engine** (من منصة Informatica وليس من DPM نفسه).
- ميزات مراقبة **User Access & Activity** تعتمد على:
- Spark Jobs.
- Hadoop Grid.
- هذه الميزات اختيارية، وليست جزءًا افتراضيًا من DPM.

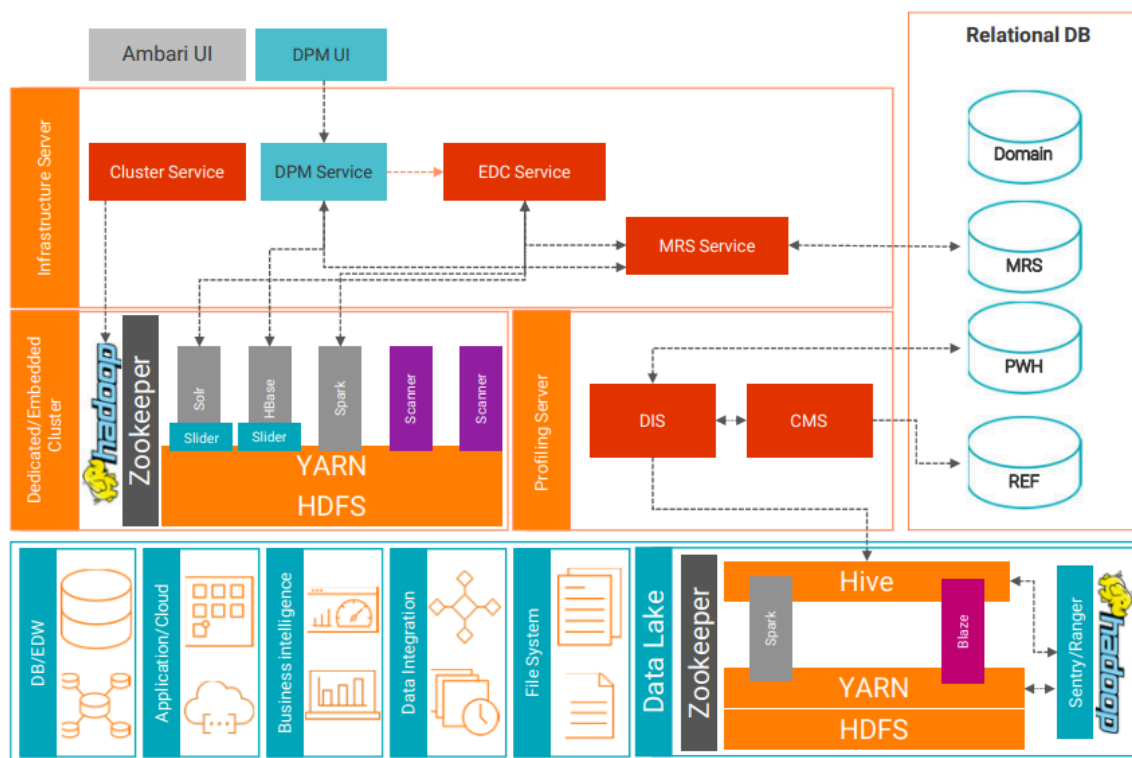
4. Storage Layer

النظام	الوظيفة
Relational DB	تخزين إعدادات (DPM (Datastores, Classifications...
MRS	تخزين بيانات الـ Data Domains
HBase + HDFS	تخزين كيانات Subject Registry
Elasticsearch	فهرسة واسترجاع تقارير نشاط المستخدم

كيانات او Object في مجال البنوك تكون : عميل - حساب - قرض - فرع

DPM Architecture – Informatica

DPM Architecture



مثل Informatica لا يعمل بشكل منفصل، بل يتكامل مع عدة خدمات أساسية من منصة DPM:

- EDC (Enterprise Data Catalog)
- DIS (Data Integration Service)
- CMS (Model Repository Service)
- HBase/HDFS
- Embedded Hadoop Cluster

المكونات الأساسية ودور كل منها:

المكون	الوظيفة
DPM IJI (Informatica Job Integration)	وحدة داخل DPM مسؤولة عن إرسال المهام وتشغيل الـ scan
DPM Service	المحرك الأساسي المسؤول عن التنسيق بين DPM وبقية الخدمات
EDC Service	يستخدم لاكتشاف البيانات وتصنيفها (بدون واجهة IJI هنا)
MRS (Model Repository Service)	يخزن تعريفات Data Domains، ونتائج التصنيف
HBase + HDFS	يستخدم لتخزين بيانات subject registry (مثل معلومات الأشخاص)
DIS (Data Integration Service)	يُنفذ عمليات الفحص والتحليل على البيانات
CMS (Content Management Service)	يستخدم لدعم عمليات الـ profiling (تحليل البيانات)

المكون	الوظيفة
Embedded Hadoop Cluster	بيئة تشغيل للمهام الكبيرة مثل Spark أو التخزين الموزع

(Workflow)

1. المستخدم يشغل مهمة من DPM UI.

2. DPM Service يتواصل مع DPM IJI.

3. DPM Service:

- يطلب من EDC Service تشغيل scan.
- يتصل بـ MRS للحصول على معلومات Data Domains.

4. EDC Service:

- يشغل DIS و CMS للقيام بعمليات profiling.
- يرسل النتائج إلى MRS.

5. DPM Subject Registry كـ HBase/HDFS يسجل الكيانات في DPM.

6. ElasticSearch يستخدم لفهرسة بيانات (user access monitoring غير مذكور هنا لكنه موجود غالبًا في جزء) الاستخدام.

`flowchart TD

UI[DPM UI]

IJI[DPM IJI]

DPM[DPM Service]

EDC[EDC Service]

MRS[Model Repository Service]

DIS[Data Integration Service]

CMS[Content Management Service]

HBase[HBase (Subject Registry)]

HDFS[HDFS Storage]

Hadoop[Embedded Hadoop Cluster]

UI --> IJI --> DPM

DPM --> EDC

DPM --> MRS

EDC --> DIS

EDC --> CMS

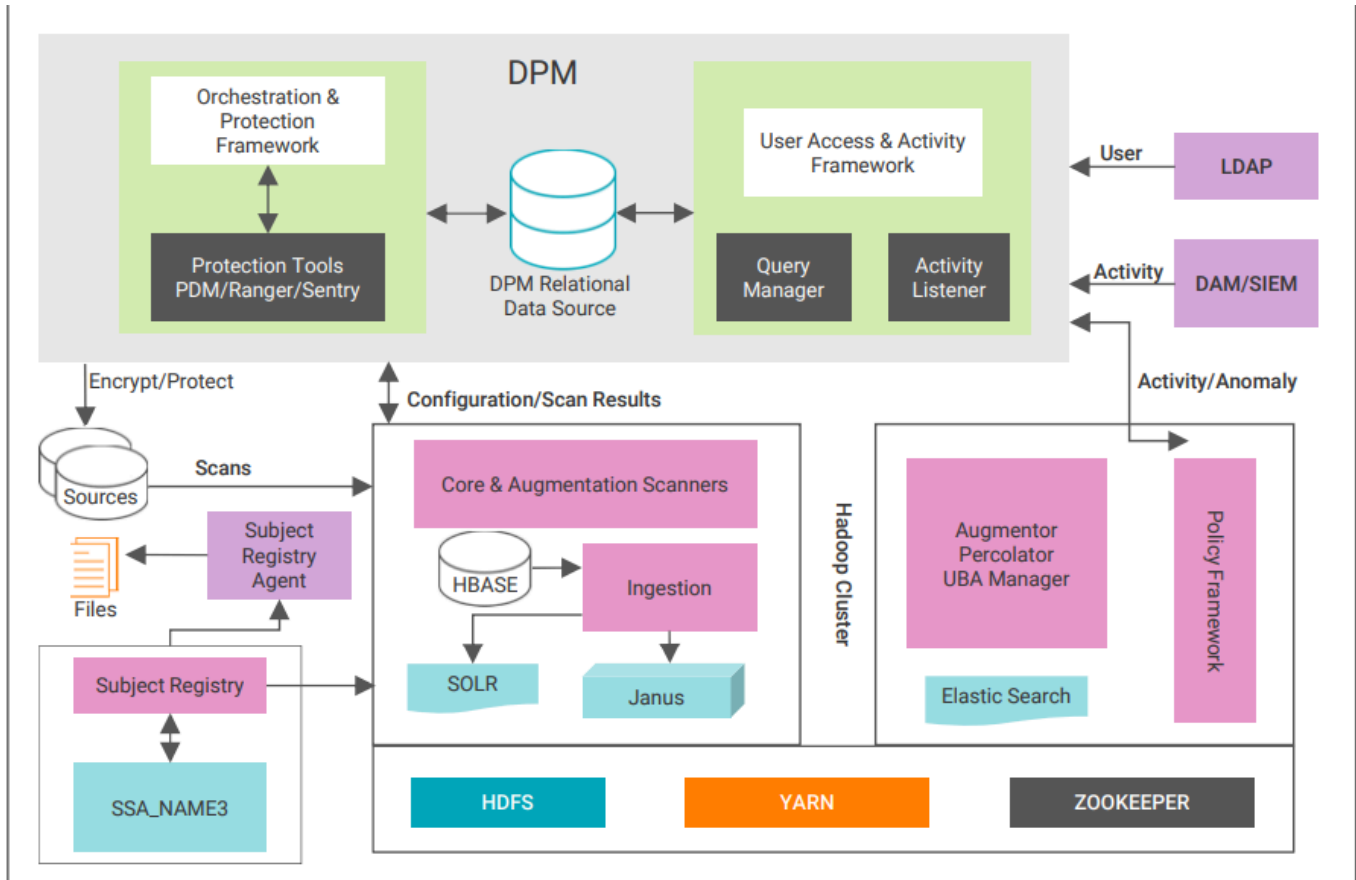
EDC --> MRS

DPM --> HBase
DPM --> HDFS
HBase --> Hadoop
HDFS --> Hadoop`

ملاحظات مهمة:

- مستخدمة EDC غير متاح في هذا السياق، لكن باقي خدمات EDC IJI.
- عنصر مركزي لأنه يحفظ كل التعاريف والنتائج MRS.
- HBase و HDFS تديران الـ Subject Data المرتبطة بالـ (DSAR) مثل معلومات الأشخاص المرتبطة بالـ.
- تُشكّل معًا خط المعالجة والتحليل للبيانات EDC + DIS + CMS.

DPM Services Architecture



مش أداة واحدة، ده نظام بيتكوّن من مكونات كتير بتشتغل مع بعض. خيلنا نقسمهم إلى 4 أجزاء DPM:

1. Orchestration & Protection (إدارة وتشغيل الحماية)

المعنى:

زوي (Jobs) هو المسؤول عن تنظيم وتنفيذ المهام DPM:

- استيراد تعريفات البيانات الحساسة (زوي الاسم، رقم البطاقة).
- تشفير أو إخفاء البيانات.

أدوات الحماية:

- PDM (Persistent Data Masking): يشفّر أو يخفي البيانات.
- Ranger/Sentry: لحماية البيانات داخل Hadoop.

2. User Activity Monitoring (مراقبة سلوك المستخدمين)

المعنى:

بمراقبة الناس التي يشتغلوا على البيانات، ويشوف هل في سلوك مريب.

الأدوات المستخدمة:

- DAM أو SIEM: أنظمة بتسجل النشاطات (زوي مين فتح الملف، ومتى).
- Hadoop Cluster: يخزن نشاط المستخدمين.
- ElasticSearch: يبحث بسرعة في النشاطات عشان نطلع تقارير.

3. Scanning & Metadata Collection (فحص البيانات وجمع معلومات عنها)

المعنى:

بيعمل Scan على قواعد البيانات ويجيب:

- أسماء الجداول والحقول.
- هل فيها بيانات حساسة ولا لأ.
- نوع البيانات (أرقام، نصوص، ...).

الأداة المستخدمة:

- بتعمل الفحص وتحليل البيانات Informatica أداة ذكية من EDC من Percolator Scanner

4. Subject Registry (تحديد وتسجيل الأشخاص داخل البيانات)

المعنى:

فيه أداة اسمها "Subject Registry Agent" دورها تلاقي الأشخاص جوا البيانات، حتى لو البيانات غير منظمة (زي ملفات PDF أو CSV).

محرك المطابقة:

- ببستخدم حاجة اسمها SSA Name3:
- بتقارن الأسماء وتقول لك "الاسم ده شبه ده بنسبة 95%".
- كويس جدًا لو الاسم مكتوب غلط أو فيه تشابه.

مثال عملي مبسط

تخيل شركة بتخزن بيانات عملاء في 3 أماكن:

- قاعدة بيانات SQL.
- ملفات Excel.
- Salesforce.

الشركة عايزة تعرف:

- مين العملاء؟
- فين بياناتهم؟
- هل البيانات دي حساسة؟
- مين فتحها أو عدلها؟

هيشغل كده DPM:

الخطوة	اللي بيحصل
1. Orchestration	يحدد المهام اللي هيعملها DPM



الخطوة	التي يحصل
2. Scan	لـكل مصدر بيانات يعمل Percolator Scan.
3. Classification	يعرف إن "رقم البطاقة" و"البريد الإلكتروني" بيانات حساسة DPM.
4. Protection	يطبق تشفير أو إخفاء على رقم البطاقة باستخدام PDM.
5. Subject Discovery	يلاقى كل العملاء من الملفات وقاعدة البيانات Agent.
6. SSA Name3	يقارن الأسماء ويقول إن "محمد أحمد" هو نفسه "م. أحمد".
7. Monitoring	يشوف مين فتح ملف فيه بيانات عميل من يومين باستخدام ElasticSearch.

UI

Header :

الوظائف الأساسية في الـ Header:

العنصر	الوظيفة
System Settings	ضبط إعدادات النظام (الاتصالات، المستخدمين، الأمان...)
Subject Registry	عرض الكيانات (Subjects) التي تم اكتشافها في البيانات
Tasks	عرض المهام الجارية أو المكتملة (زي Scans أو حماية)
Security Violations	عرض المخالفات في سياسات الأمان (زي وصول غير مصرح به)
Anomalies	عرض الأنشطة الشاذة للمستخدمين (زي تحميل بيانات كثيرة فجأة)

العنصر	الوظيفة
 Filter Workspace List	تصفية قائمة المساحات (Workspaces) حسب القسم أو النوع
 Online Help	فتح مركز المساعدة أو التوثيق الرسمي للـ DPM

قائمة الـ Menu في الـ Header:

- **Scan Summary:** ملخص سريع لنتائج عمليات الفحص.
- **Open Workspace:** معين بسرعة Workspace الانتقال إلى.
- **Change Password:** تغيير كلمة مرور المستخدم الحالي.
- **Log Out:** DPM تسجيل الخروج من.

Work Space

هي صفحة بتجمع كل الوظائف والبيانات المتعلقة بجانب معين من إدارة البيانات الحساسة.

Think of it like a dashboard for a specific function (e.g. Subject Discovery, Protection Policies, Scan Results...)

أمثلة على Workspaces:

محتواه	Workspace
كل ما يخص الفحص والتصنيف والـ metadata	Data Discovery
إدارة الكيانات واكتشاف الأشخاص	Subject Registry
إعداد قواعد الحماية والتشفير	Protection
عرض وتحليل مخالفات السياسات	Violations
تقارير طلبات الوصول للبيانات	DSARs

كيف تدخل إلى Workspace؟

- من **Main Menu** الجانبي.
- من أيقونات الـ **Header**.
- من مهام **(Tasks)** داخل Workspace آخر.