**DISSENY DE Seguretat del sistema informàtic**

**ÍNDEX**

[1. Sigles i definicions 3](#_Toc533346203)

[1.1 Sigles 3](#_Toc533346204)

[1.2 Definicions 4](#_Toc533346205)

[2. Introducció 6](#_Toc533346206)

[3. Descripció del plantejament metodològic de la seguretat 7](#_Toc533346207)

[3.1 Descripció general de la metodologia 7](#_Toc533346208)

[3.2 Pas 1: definició del context 8](#_Toc533346209)

[3.3 Pas 2: expressió dels requisits de seguretat 8](#_Toc533346210)

[3.4 Pas 3: avaluació dels riscos 8](#_Toc533346211)

[3.5 Pas 4: definició dels objectius de control 8](#_Toc533346212)

[3.6 Pas 5: definició dels controls 8](#_Toc533346213)

[3.7 Pas 6: revisió de la seguretat de la informació 8](#_Toc533346214)

[4. Anàlisi funcional 9](#_Toc533346215)

[4.1 Accés unificat a la consola 9](#_Toc533346216)

[4.2 Seguretat de les Big Data 10](#_Toc533346217)

[5. Arquitectura 11](#_Toc533346218)

[5.1 Arquitectura funcional 11](#_Toc533346219)

[5.1.1 Seguretat global 11](#_Toc533346220)

[5.1.2 Gestió d'identitats i d'accessos - IAM 12](#_Toc533346221)

[5.1.3 Seguretat del Big Data 13](#_Toc533346222)

[5.2 Arquitectura del programari 14](#_Toc533346223)

[5.2.1 Gestió d'identitats i d'accessos 14](#_Toc533346224)

[5.2.1.1 Descripció 14](#_Toc533346225)

[5.2.1.2 Components tecnològics 15](#_Toc533346226)

[5.2.1.3 Solucions alternatives 17](#_Toc533346227)

[5.2.1.3.1 OpenAM 17](#_Toc533346228)

[5.2.1.3.2 Solució "casolana" 17](#_Toc533346229)

[5.2.2 Seguretat del Big Data 17](#_Toc533346230)

[5.2.2.1 Descripció 17](#_Toc533346231)

[5.2.2.2 Components tecnològics 18](#_Toc533346232)

[5.2.2.3 Èmfasi en l'autenticació 19](#_Toc533346233)

[5.2.2.4 Èmfasi en la protecció de les dades 22](#_Toc533346234)

[5.2.2.5 Èmfasi en l'autorització 22](#_Toc533346235)

[5.2.2.6 Èmfasi en la gestió de dades 23](#_Toc533346236)

[5.2.2.7 Èmfasi en la seguretat de l’accès a les dades 24](#_Toc533346237)

# Sigles i definicions

## Sigles

A continuació, trobareu una llista ordenada alfabèticament de les sigles que s'utilitzen en aquest document.

| **Terme** | **Definició** |
| --- | --- |
| EBIOS | **Expression des Besoins et Identification des Objectifs de Sécurité (Expressió de les necessitats i identificació dels objectius de seguretat)** |
| SI | **Sistema Informàtic** |
| ISS | **Information System Security (Seguretat del Sistema Informàtic)** |
| RACI | **Responsible / Accountable / Consulted / Informed (Encarregat / Responsable / Consultat / Informat)** |
| IAM | **Identification and Access Management (Gestió d'identitats i d'accessos)** |

## Definicions

A continuació, trobareu una llista ordenada alfabèticament dels termes que s'utilitzen en aquest document.

| **Terme** | **Definició** |
| --- | --- |
| Disponibilitat | El fet de poder utilitzar o accedir a quelcom quan es demani a una entitat autoritzada. |
| Confidencialitat | Propietat d'una informació que no es vol divulgar o revelar a persones, entitats o processos no autoritzats. |
| Conseqüència | Resultat d'un esdeveniment que afecta els objectius. |
| Control | Mesura que modifica el risc. |
| Objectiu de control | Enunciat que descriu el que es vol aconseguir com a resultat de la implementació dels controls. |
| Integritat | Propietat de ser exacte i complet. |
| Risc | Efecte de la incertesa dels objectius. |
| Anàlisi del risc | Procés que pretén comprendre la naturalesa del risc i determinar-ne el nivell. |
| Amenaça | Causa potencial d'un incident no desitjat que pot ocasionar efectes negatius en un sistema o una organització. |
| Vulnerabilitat | La debilitat d'un actiu o d'un control, la qual pot ser explotada per una o més amenaces. |
| Identificació | En la seva versió més senzilla, la identificació consisteix en introduir una combinació d'un identificador + una contrasenya que permet conèixer la identitat de l'usuari. |
| Autenticació | L'autenticació és el procés que consisteix en verificar la identitat dels usuaris comparant-la amb el repositori d'usuaris. |
| Autorització | Un cop que l'usuari s'ha autenticat, l'autorització recupera la informació necessària des del repositori per tal de verificar que l'usuari tingui legitimitat per accedir a les aplicacions. |
| Auditoria | Procés que consisteix en revisar i examinar la identificació, l'autenticació, l'autorització i les activitats d'un servei o d'un usuari. El resultat de l'auditoria ha de satisfer la política de seguretat. |

# Introducció

Aquest document tracta sobre el disseny de la seguretat del sistema informàtic de la plataforma. L'abast d'aquest documents és el següent:

* Definició del plantejament metodològic de la seguretat que es farà servir per definir les especificacions de seguretat de la plataforma.
* Disseny de l'arquitectura que s'utilitzarà per a implementar la seguretat. Aquesta part es centrarà en els següents punts:
  + L'accés unificat al sistema,
  + La seguretat aplicada al repositori de Big Data i la gestió del cicle de vida de les dades.

# Descripció del plantejament metodològic de la seguretat

Aquest capitol descriu el plantejament que s'utilitzarà a l'hora de definir la seguretat de la plataforma. Aquest plantejament es basa en una metodologia d'anàlisi dels riscos que s'anomena EBIOS, un mètode estructurat i contrastat que compleix la norma ISO 27005.

La metodologia aplicada a l'hora de definir els objectius de seguretat serà una versió lleugera, però prou àmplia, de la metodologia EBIOS.

## Descripció general de la metodologia

Aquesta metodologia de seguretat té sis passos principals:

1. Definició del context (característiques, actors, limitacions, ...),
2. Expressió dels requisits de seguretat, amenaces i conseqüències potencials,
3. Avaluació dels riscos (amenaces, vulnerabilitat, origen de l'amenaça, probabilitat),
4. Definició dels objectius de control,
5. Definició dels controls,
6. Revisió de la seguretat de la informació.

El quadre següent és un diagrama d'aquesta metodologia:



## Pas 1: definició del context

L'objectiu d'aquest pas és recollir tota la informació que cal per dur a terme l'estudi de gestió de riscos. Per exemple:

* Actius essencials,
* Restriccions legals,
* Mètriques,
* Organització...

Durant aquest pas, es defineix clarament l'abast de l'estudi.

## Pas 2: expressió dels requisits de seguretat

L'objectiu d'aquest pas és definir els requisits pel que fa a la seguretat de la informació de cada actiu essencial tenint en compte les limitacions del sistema. En general, els requisits de seguretat de la informació s'expressen en termes de disponibilitat, confidencialitat i integritat. El valor de cada criteri de seguretat de la informació es determina mitjançant l'ús d'amenaces, que també es determinen en aquest pas.

## Pas 3: avaluació dels riscos

L'objectiu d'aquest pas és definir l'impacte dels escenaris d'incidents i determinar quins cal tractar. Aquest pas ve després de l'especificació tècnica del sistema. Els requisits de seguretat s'han definit en el pas 2 d'acord amb una visió funcional del sistema. En el pas 3, els actius essencials s'assignen a dispositius tècnics i les amenaces s'associen als escenaris d'incident, l'impacte dels quals està associat als dispositius tècnics. Al final d'aquest pas, s'identifiquen els riscos de les amenaces i dels escenaris d'incident avaluats utilitzant l'escala descriptiva definida al pas 1.

## Pas 4: definició dels objectius de control

L'objectiu d'aquest pas és definir la forma de tractar els riscos: modificar, contenir, evitar compartir. Es farà servir aquest anàlisi per especificar solucions que permetin que el sistema compleixi aquests objectius.

## Pas 5: definició dels controls

L'objectiu d'aquest pas és definir els controls, és a dir, especificar els objectius de control. La manera en què s’assoliran aquets objectius de control serà definida per cada usuari.

## Pas 6: revisió de la seguretat de la informació

Aquest pas va des del final de l'etapa d'especificació fins al final del projecte. Les solucions de seguretat implementades en l'arquitectura del sistema s'analitzen per comprovar que es compleixen els objectius de control. Al final del projecte, totes les proves verifiquen que els controls desplegats compleixin els objectius de control i que els riscos residuals siguin coneguts i acceptats pel responsable.

# Anàlisi funcional

L'anàlisi preliminar de la seguretat va fer palesa la necessitat d'abordar dues qüestions importants:

* L’accés unificat al sistema,
* La seguretat aplicada al repositori de Big Data i a la gestió del cicle de vida d'aquestes dades.

En conseqüència, aquest capítol està dedicat a descriure les necessitats funcionals dels dos punts anteriors.

## Accés unificat a la consola

Per tal que els usuaris es puguin connectar a les eines, el sistema plataforma hauria de proporcionar una solució que garanteixi un accés segur i la gestió d'identitats.

Els usuaris que necessitin accedir al sistema hauran de fer un pas preliminar de connexió. Sense aquest pas, no podran veure les dades ni accedir a les funcions. La identificació i l'autenticació dels usuaris es duen a terme amb el mòdul que s'encarrega de la gestió d'identitats i d'accessos.

La plataforma permet que hi accedeixin tots els usuaris coneguts i admesos pel sistema. Per accedir-hi, tanmateix, cal introduir un nom d'usuari i una contrasenya específics de cada usuari. Si es produeix un error en accedir, es notificarà a l'usuari la causa per la qual se li ha denegat l'accés. Si l'usuari inicia la sessió correctament, podrà accedir a les eines que li permetin els seus permisos.

Nota: l'ús d'un nom d'usuari i una contrasenya específics per a cada usuari és un element necessari si es vol garantir la traçabilitat de l'accés i poder fer auditories.

Cada eina del sistema s'encarrega de determinar els permisos dels usuaris, és a dir determinar si són capaços de veure dades, accedir a les funcions o realitzar accions. Aquests permisos es calculen en funció dels atributs de l'usuari:

* Perfils,
* Rols o missions,
* Perímetres geogràfics,
* Etc.

La figura següent representa el procés que permet calcular els permisos dels usuaris que accedeixen al sistema:



D'altra banda, en el plataforma, es prefereix un accés unificat dels usuaris a la consola. Això vol dir que la identitat dels usuaris es trasllada d'una aplicació a una altra, fent servir uns mecanismes d'inici de sessió únic proporcionats pel sistema de gestió d'identitats i accessos.

Per permetre el comportament descrit anteriorment, el mòdul de gestió d'identitats i accessos es basa en un directori central usuaris. Per tant, cal que la solució proporcioni funcions que permetin administrar els usuaris, els perfils, els rols, les missions, etc.

## Seguretat de les Big Data

Les Big Data ofereixen a l'organització una gran oportunitat per accedir a una gran varietat de dades amb la finalitat de realitzar anàlisis molt potents o facilitar la presa de decisions.

Tanmateix, aquesta tecnologia presenta nous reptes en termes de gestió de la seguretat. Per això, cal tenir en compte els àmbits següents:

* **Fonts de les dades.** El repositori de Big Data contindrà una gran varietat i quantitat de dades, estructurades o no, procedents de diferents sistemes d’informació. Aquestes dades poden contenir informació personal identificable, dades de targetes bancàries, dades protegides per la propietat intel·lectual, informació sobre serveis urbans, noms d'organitzacions, etc. Per tant, totes les fonts de dades i el repositori de Big Data han d'estar protegits per tal de satisfer les polítiques de seguretat i compliment.
* **Marcs de programació distribuïda de Big Data.** L'entorn d'execució d'un sistema de Big Data, que s'encarrega de fer els càlculs computacionals, també ha d'estar protegit per tal d'evitar que s'hi pugui accedir explotant els recursos del sistema.
* **Resultats analítics.** Són els resultats de totes les operacions de càlcul aplicades a les dades ingerides pel sistema. Proporcionen informació valuosa (i, de vegades, privada) per a determinats camps o negocis. Per això cal garantir la seguretat dels accessos si es vol evitar la filtració d'informació.

# Arquitectura

## Arquitectura funcional

## Seguretat global

Els protocols de seguretat utilitzats pel sistema depenen de “l’àrea”. S’identifiquen dues àres, una “Interna” (dins del sistema), amb eina de gestió de la dada de “baix nivell” i una “Externa”, amb API “d’alt nivell” que permet a terceres parts de l’aplicació d’accedir a les dades.

Per l’àrea externa, es necessita implementar un protocol SSO, permetent d’identificar els usuaris a través de multliples aplicacions.

Per l’àrea interna, es necessita implementar un protocal estàndard de seguretat compatible amb tots els components escollits.

Independentment de l’àrea, el repositori de l’usuari és l’Active Directory de la ciutat.

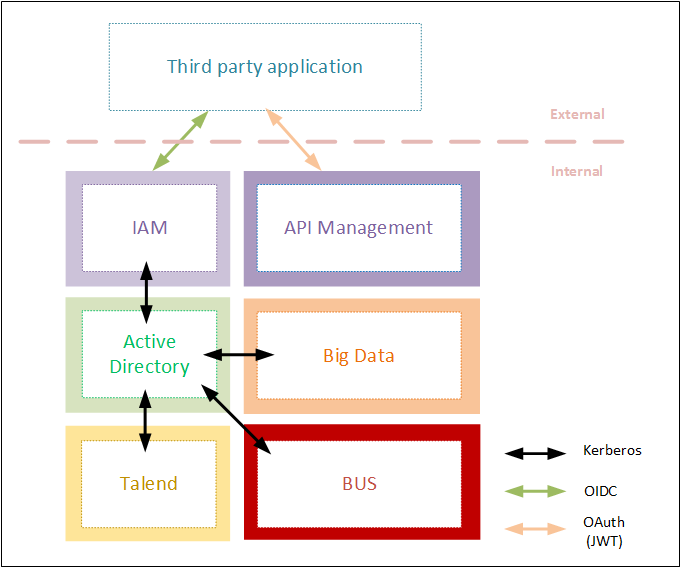
OpenId Connect (OIDC) és un protocol SSO lleuger que s’utilitza majoritàriament per assegurar l’accès a les multiples aplicacions. Associat amb el flux OAuth JWT, aquest protocol permet d’identificar l’usuari a través de multiples aplicacions i autoritzar-lo a accedir a les dades a través de la solució API Management.

Es tria per l’aspecte de la seguretat externa i s’implementa per la solució IAM, la cual està lligada a l’Active Directory.

Kerberos és un potent protocol d’autentificació basat en claus criptogràfiques. Es àmpliament utilitzat i compatible amb totes les aplicacions utilitzadas en el sistema. S’implementa directament per l’Active Directory (depenent de la seva versió) o bé a través d’un component agregat “add-in” (MIT KDC).

Es tria per l’aspecte de la seguretat interna.

La figura següent mostra les interaccions de la seguretat global entre els components disponibles:



## Gestió d'identitats i d'accessos - IAM

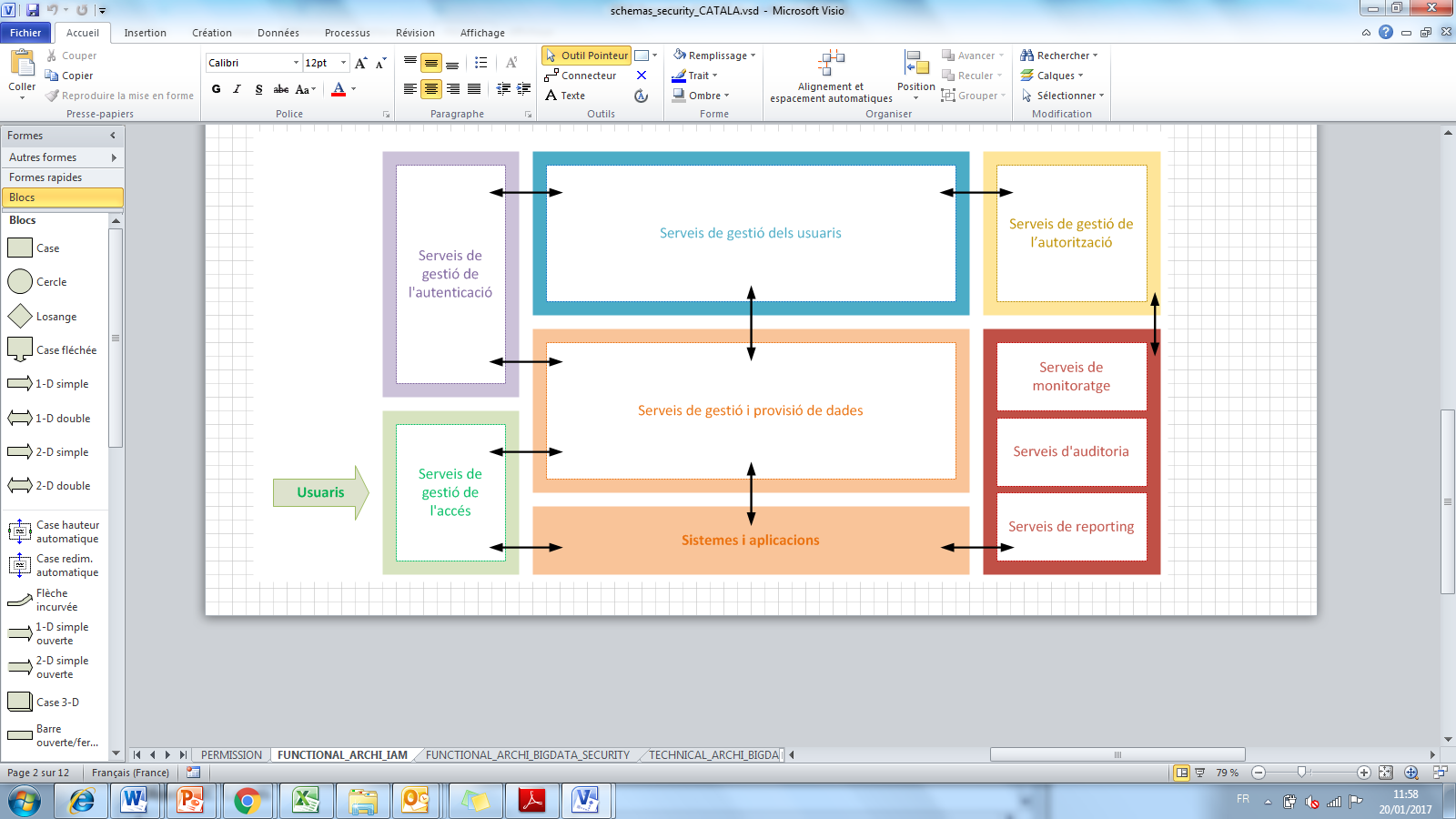
Les sigles IAM (Identity and Access Management) fan referència al conjunt de processos que fa servir un sistema a l'hora de gestionar la identificació, l'autenticació i l'autorització de les persones que accedeixen a un servei o a unes dades. També inclou un conjunt de funcions pensades per a garantir la seguretat, la traçabilitat i la auditabilitat dels sistemes.

El procés IAM es divideix en tres etapes:

* **Identificació i autenticació.** En la seva versió més senzilla, la identificació consisteix en introduir una combinació d'un identificador més una contrasenya que permet conèixer la identitat de l'usuari. L'autenticació és el procés que consisteix en verificar la identitat dels usuaris comparant-la amb el repositori d'usuaris.
* **Autorització.** Un cop que l'usuari s'ha autenticat, l'autorització recupera la informació necessària des del repositori per tal de verificar que l'usuari tingui legitimitat per accedir a les dades, processos o sistemes interns de la plataforma.
* **Auditoria.** Procés que consisteix en revisar i examinar la identificació, l'autenticació, l'autorització i les activitats d'un servei o d'un usuari. El resultat de l'auditoria ha de satisfer la política de seguretat.

La gestió d'identitats i accessos també proporciona la possibilitat de treballar amb inicis de sessió únics, és a dir, proporciona el procés d'identificació i autenticació per a totes les eines del plataforma. Cada eina s'encarrega del seu propi procés d'autorització.

La figura següent presenta l'arquitectura funcional de la solució IAM:



Descripció dels serveis:

* **Serveis de gestió dels usuaris.** Aquest component és l'encarregat de la governança i la gestió efectives dels cicles de vida de les identitats.
* **Serveis de gestió de l'autenticació.** Aquest component és l'encarregat de verificar la identitat d'un usuari per a una organització.
* **Serveis de gestió de l'autorització.** Aquest component és l'encarregat del procés que permet determinar els drets d'accés que decideixen a quins recursos d'una entitat (usuari o servei) es pot accedir d'acord amb les polítiques de l'organització.
* **Serveis de gestió de l'accés.** Aquest component és l'encarregat d'aplicar les polítiques de control d'accés en resposta a una sol·licitud d'una entitat.
* **Serveis de gestió i provisió de dades.** Aquest component s'encarrega de sincronitzar les dades d'autorització amb els recursos.
* **Serveis de monitoratge, auditoria i reporting.** Aquest mòdul s'encarrega de comprovar que es compleixi la política de seguretat de l'organització.

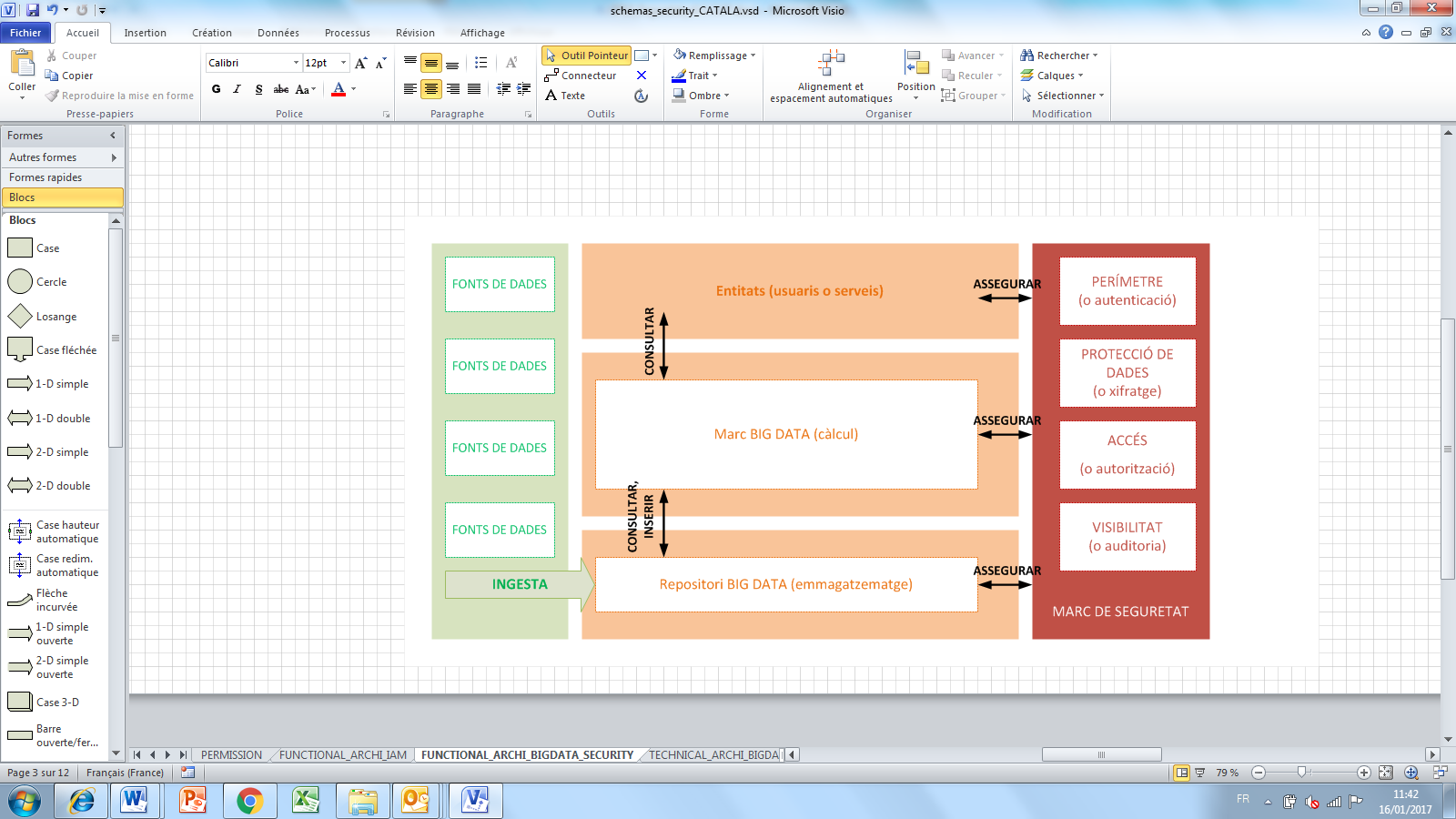
## Seguretat del Big Data

La seguretat del Big Data es basa en diversos components funcionals que permeten garantir la seguretat de les seves dades i, especialment, de les de caire privat.

Per tal de complir els requisits de seguretat que imposen les normatives i controls de governança del sector, cal que el marc de seguretat tingui en compte el següent:

* **Autenticació (seguretat perimetral).** L'autenticació s'encarrega de controlar l'accés a les dades i els seus serveis per part de les entitats (usuaris o serveis).
* **Xifrat (protecció de dades).** El xifrat s'encarrega de protegir les dades contra els accessos no autoritzats.
* **Autorització (drets o accés).** L'autorització s'encarrega de determinar el que les entitats poden fer amb les dades i els serveis.
* **Auditoria (transparència o visibilitat).** L'auditoria consisteix en controlar l'ús de les dades (quan, on, com).

La figura següent presenta l'arquitectura funcional de la solució de seguretat del Big Data:



Descripció dels serveis:

* **Fonts de les dades.** Les fonts de dades representen el sistemes que generen dades. Aquestes dades poden estar estructurades o no, i poden provenir de sistemes diferents (Sistemes d’Informació interns de l'organització, externs, etc.) i fer referència a diversos àmbits (financer, industrial, laboral, etc.).
* **Repositori de Big Data.** Aquest component és el lloc (sistema d'arxius) on s'emmagatzemen les dades procedents de les fonts de dades un cop ingerides pel sistema.
* **Marc del Big Data.** Aquest marc proporciona un conjunt d'eines que ofereixen funcions de càlcul (anàlisi, algoritmes predictius, etc.) o eines per fer consultes.
* **Entitats (usuaris o serveis).** Les entitats són els usuaris i serveis que necessiten accedir a les dades o resultats analítics que hi ha al repositori.
* **Marc de seguretat.** Conjunt d'eines que permeten garantir la seguretat de la solució de big data d'acord amb els conceptes que hem explicat abans (seguretat perimetral, protecció de dades, drets o accés, transparència o visibilitat). La seguretat abasta tot el cicle de vida de les dades i tots els components del Big Data.

## Arquitectura del programari

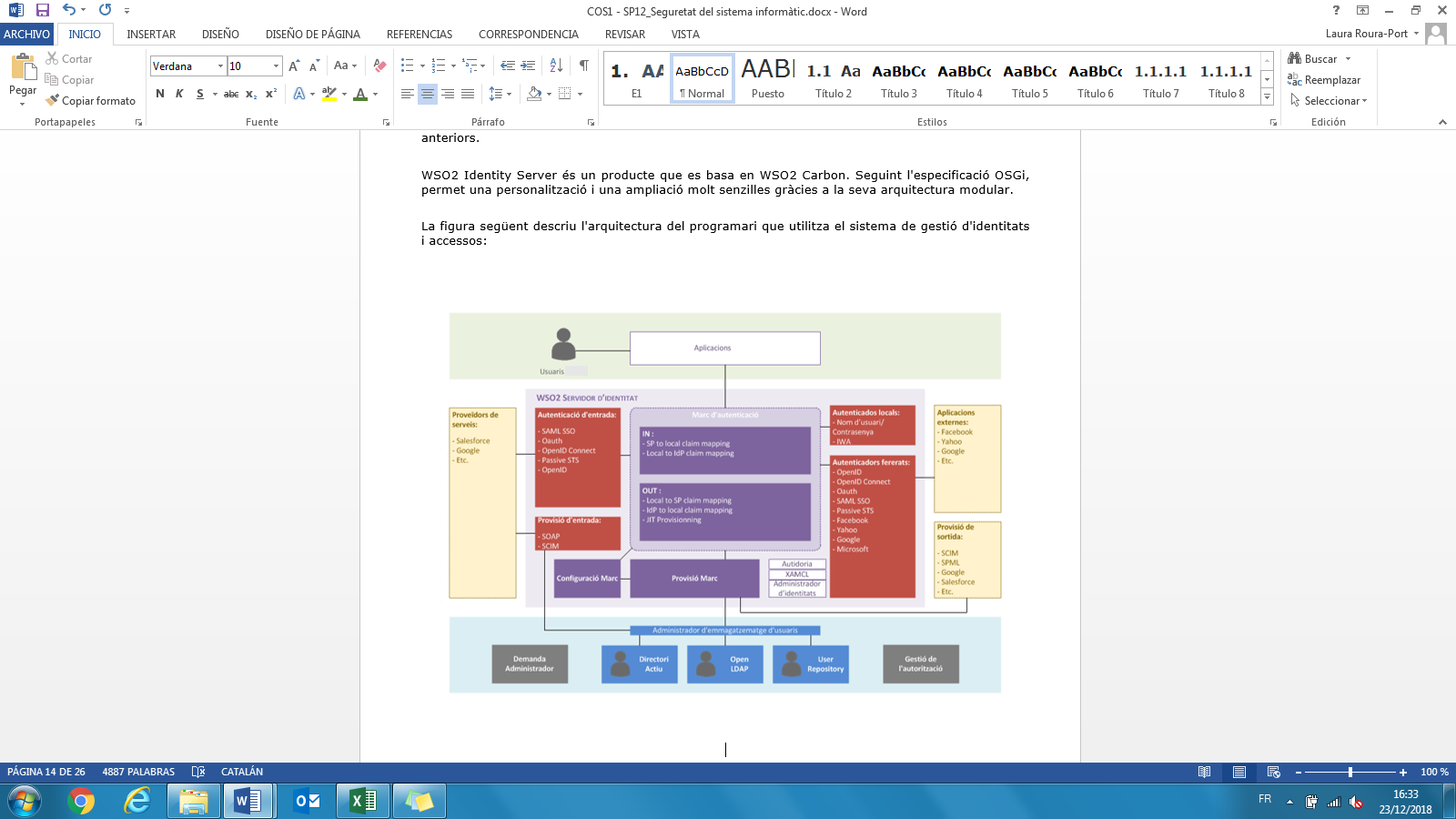
## Gestió d'identitats i d'accessos

## Descripció

La solució que s’ha triat per proporcionar la gestió d'identitats i accessos és WSO2 amb el seu producte WSO2 Identity Server, que ofereix totes les funcions que es descriuen en els capítols anteriors.

WSO2 Identity Server és un producte que es basa en WSO2 Carbon. Seguint l'especificació OSGi, permet una personalització i una ampliació molt senzilles gràcies a la seva arquitectura modular.

La figura següent descriu l'arquitectura del programari que utilitza el sistema de gestió d'identitats i accessos:



## Components tecnològics

Els components de WSO2 Identity Server són els següents:

| **Component** | **Descripció** |
| --- | --- |
| Service providers | **Els service providers (proveïdors de serveis) no formen part de WSO2 Identity Server**, sinó que representen entitats amb les quals WSO2 es pot connectar. El proveïdor de servei proporciona autenticació i autorització externes. Google i Salesforce són exemples de proveïdors de serveis. |
| Inboud authentication | El component Inboud authentication (autenticació d'entrada) rep la sol·licitud d'autenticació del proveïdor de serveis. Depenent de la configuració i del proveïdor de serveis, pot rebre **sol·licituds de SSO en SAML, OAuth/OpenID connect, Passive STS o OpenID**. A continuació, la sol·licitud s'envia al component Authentication framework (marc d'autenticació). |
| Authentication framework | Aquest component és **l'encarregat d'encaminar la sol·licitud d'autenticació**, en funció de la seva configuració, cap als autenticadors locals i/o federats. Un cop realitzada la sol·licitud d'autenticació, s'envia la resposta al component Inboud authentication. |
| Local authenticators | Els Local authenticators (autenticadors locals) **són processos d'autenticació disponibles dins d'Identity Server**. Aquests autenticadors comproven la identitat mitjançant el nom d'usuari i la contrasenya o mitjançant el sistema d'autenticació de Windows (IWA). |
| Federated authenticators | Els Federated authenticators (autenticadors federats) **són processos d'autenticació que no estan disponibles dins d**'**Identity Server**. Identity Server és compatible amb diversos proveïdors d'identitat:   * OpenID, * OpenID Connect, * OAuth, * SAML SSO, * Passive STS, * Facebook, * Yahoo, * Google, * Microsoft (Hotmail, MSN, Live). |
| External applications | **Les External applications (aplicacions externes) no formen part de WSO2 Identity Server**. Aquest component proporciona funcions d'autenticació. |
| Provisioning framework | **El Provisioning framework (marc de provisió) s'encarrega de tota la feina de provisió que realitza Identity Server**. Aquest marc s'integra amb el component User Store Manager i també rep sol·licituds de provisió del marc d'autenticació. |
| Authorization manager | **Authorization manager (administrador d'autoritzacions) és un component avançat d'auditoria i gestió de drets**. |
| IdP and SP configurations | **Aquests components són els encarregats d'establir les bases de totes les accions** que tenen lloc dins del marc d'autenticació i del marc de provisió. |
| Inbound provisioning | **El component Inbound provisioning (provisió d'entrada) rep la sol·licitud de provisió del proveïdor de serveis**. Depenent de la configuració i del proveïdor de serveis, fa servir el mètode **SOAP o SCIM** per a intercanviar informació, i després enviar la sol·licitud al User Store Manager. |
| Outbound provisioning | **El component Outbound provisioning (provisió de sortida) pot enviar sol·licituds de provisió a aplicacions** compatibles amb els següents connectors:   * SCIM, * SPML, * Google, * Salesforce. |
| User Store Manager | **WSO2 Identity Server té el component User Store Manager (administrador del repositori d'usuaris)**, que es basa en LDAP (ApacheDS), LDAP externs, Microsoft Active Directory o qualsevol base de dades JDBC. |
| Claim manager | Aquest component **s'encarrega de gestionar la informació d'identitat**  relacionada amb els usuaris. **Propaga la informació de la identitat** empaquetant-la en tokens (com ara SAML). |
| XACML | XACML (eXtensible Access Control Markup Language) **és un llenguatge basat en XML per al control de l'accés** que ha estat estandarditzat pel Comitè Tècnic del consorci OASIS. XACML ofereix:   * Un mètode d'autorització refinada, * Un llenguatge de polítiques de control, * Un llenguatge de sol·licitud/resposta. |
| Auditoria | Aquest component permet, entre d'altres, auditoria operativa, monitoratge i gestió dels KPI, i estadístiques de rendiment mitjançant la seva **consola d'administració**. |
| Identity manager | **L'Identity manager (administrador d'identitats) és la consola d'administració que permet gestionar les polítiques d'identitat**. |

## Solucions alternatives

## OpenAM

Hi ha diversos proveïdors de solucions de gestió d'identitats i accés. Un exemple podria ser ForgeRock, que proporciona una solució de codi totalment obert anomenada OpenAM.

OpenAM està escrita en Java i inclou els dominis següents:

* Autenticació,
* Autorització,
* Inici de sessió únic (Single-Sign-On o SSO),
* Identitat federada.

Les seves principals funcions són:

* Autenticació,
* Avaluació adaptativa del risc,
* Autorització
* Federació,
* Inici de sessió únic,
* Inici de sessió mitjançant xarxes socials,
* Autoservei bàsic,
* Privadesa i consentiment,
* Administració de sessions d'alt rendiment,
* Autenticació mòbil,
* Funcions d'autenticació multifactorial.

## Solució "casolana"

Per tal de proporcionar les funcions d'IAM, també és possible integrar un component de codi obert per a construir la solució. Entre aquests components, es troben els següents:

* Directori d'Ldap – ApacheDS,
* Eines de provisió – Apache Directory,
* Inici de sessió únic – CAS (Sistema d'Autenticació Central).

## Seguretat del Big Data

## Descripció

Per tal de garantir la seguretat de la solució de Big Data (repositori, marc, processos), els components i productes que hem triat provenen de l'ecosistema Hadoop.

La figura següent descriu l'estructura del programari d'una solució típica de Big Data. En el quadre vermell, hi ha els components dedicats a garantir la seguretat de la solució de Big Data:



## Components tecnològics

Els components de seguretat de la solució de Big Data són:

| **Component** | **Descripció** |
| --- | --- |
| Active Directory / Kerberos | Protocol d'autenticació, que es podria desplegar en Active Directory. Kerberos serveix per verificar la identitat dels usuaris o hosts.  Kerberos proporciona seguretat de cadenes i és capaç d'inutilitzar els paquets d'autenticació que puguin ser interceptats per un atacant. |
| Apache Sentry | Apache Sentry és un component que permet una autorització de nivell refinat basada en rols per accedir a dades i metadades emmagatzemades en un clúster Hadoop.  Apache Sentry està en relació amb l’Active Directory, i per tant, la gestió de les autoritzacions dissenyada es basa en grups definits dins del mateix l’Active Directory. En el desenvolupament de la solució de Big Data, es defineixen grups d’autoritzacions dins d’Apache Sentry i són associats a diferents rols (accessos i rols sobre les diferents aplicacions i sobre les dades). A continuació, aquets grups d’autoritzacions són associats als grups de l’Active Directory. L’autorització dels usuaris a poder accedir o no a les diferents funcions i dades es fa doncs per la seva afectació als grups de l’Active Directory corresponents. |
| RecordService | RecordService actua com una capa de seguretat troncal i proporciona una ruta d'accés unificada a les dades per a tots els components de l'ecosistema Hadoop. S'utilitza per a aplicar seguretat de nivell refinat. |
| Cloudera Navigator Key Trustee Server | Key Trustee és un component que emmagatzema i gestiona les claus criptogràfiques i els artefactes de seguretat, i permet emmagatzemar per separat les claus criptogràfiques i les dades xifrades. |
| Cloudera Navigator  Encrypt | Cloudera Navigator Encrypt permet xifrar i protegir la seguretat de les dades en repòs de manera transparent. |
| Cloudera Navigator | Cloudera Navigator és una solució de governança de dades per a Hadoop que ofereix possibilitats com ara el descobriment de dades, l'optimització contínua, auditoria, llinatge, gestió de metadades i aplicació de polítiques. |

## Èmfasi en l'autenticació

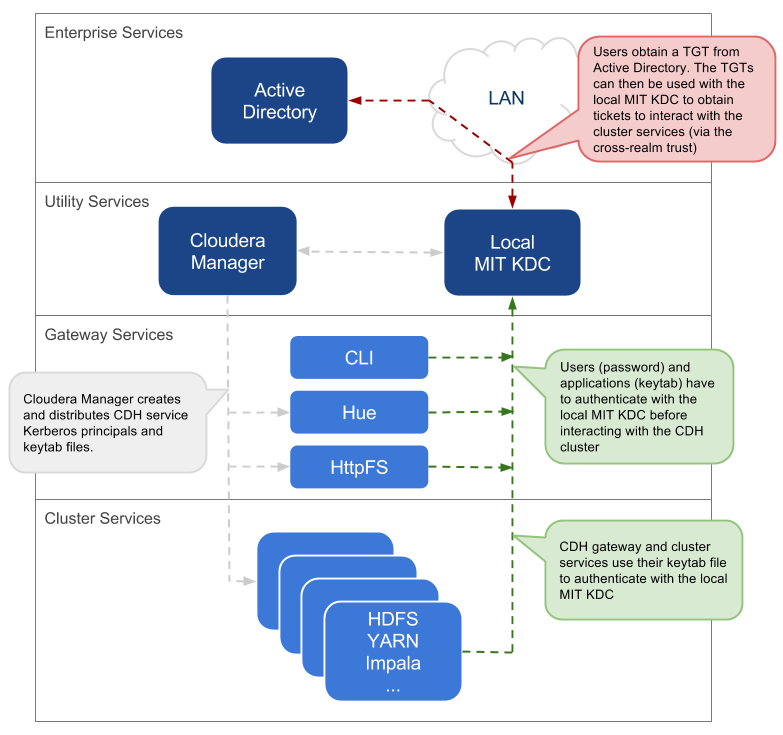
Un sistema informàtic necessita mecanismes d'autenticació per demostrar que un usuari o servei és qui diu que és. L'autenticació en l'ecosistema Hadoop es du a terme mitjançant Kerberos, un protocol d'autenticació d'empresa.

Existeixen diversos patrons de desplegament de Kerberos en un sistema informàtic:

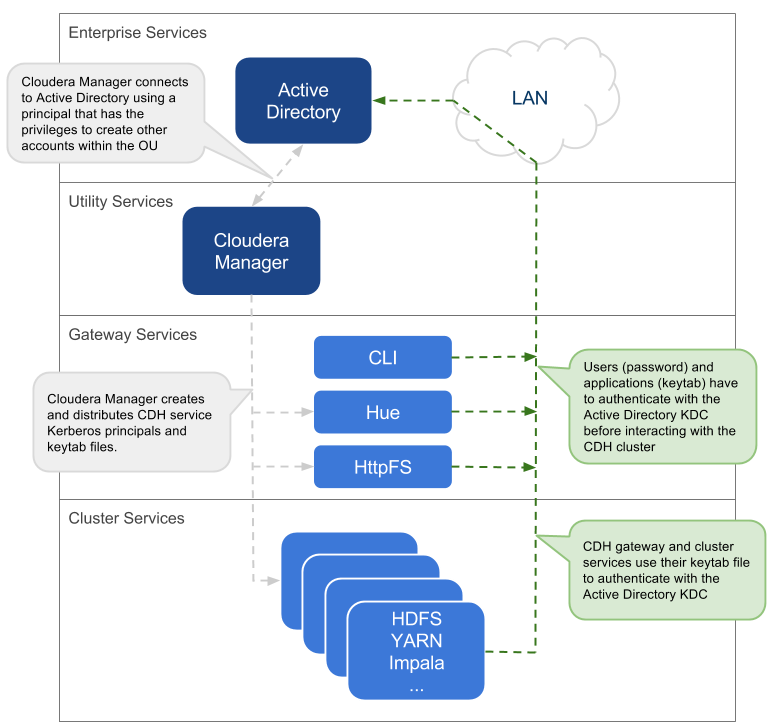
* KDC MIT local. Consisteix en desplegar un sistema d'autenticació directament en un clúster Hadoop.
* KDC MIT local amb integració a Active Directory. Consisteix en desplegar un sistema d'autenticació en un clúster Hadoop. No obstant això, l'autenticació dels usuaris o serveis es delega a l'Active Directory central.
* Directe a Active Directory. L'Active Directory central s'utilitza com a KDC, sense cap KDC desplegat en el clúster Hadoop.

Depenent de les capacitats d'Active Directory que s'hagin desplegat per al cas concret de sistema, l'escenari de desplegament de Kerberos serà "KDC MIT local amb integració a Active Directory" o "directe a Active Directory".

La figura següent representa aquests dos escenaris de desplegament:



KDC MIT local amb integració a l’Active Directory



Directe a l’Active Directory

## Èmfasi en la protecció de les dades

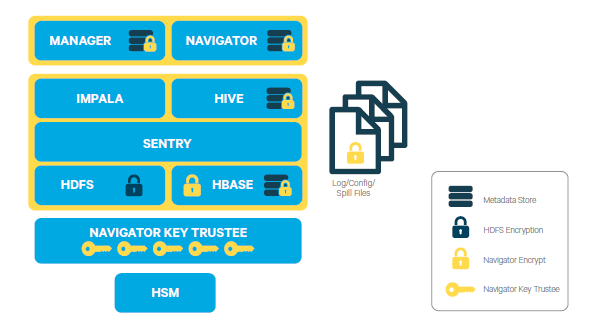
El principal objectiu de la protecció de dades és garantir que només puguin veure, utilitzar i introduir dades els usuaris o serveis autoritzats. La protecció de dades s'aplica a:

* Les dades en respòs. Consisteix en protegir les dades mentre es troben en un disc.
* Les dades en trànsit. Consisteix en protegir les dades mentre es mouen d'un procés o un sistema a un altre.

La protecció de les dades en trànsit es realitza amb **certificats SSL/TLS**  desplegats en el clúster Hadoop.

La protecció de les dades en repòs es realitza ambla combinació del **sistema de xifrat transparent HDFS de Hadoop**, **Navigator Encrypt** iel **Navigator Key Trustee** de Cloudera.

A la figura següent es descriu un desplegament típic de la protecció de dades per al sistema Hadoop.

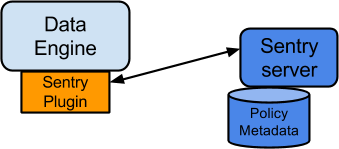


## Èmfasi en l'autorització

L'autorització fa referència a qui o què accedeix o controla un recurs o servei determinat. Els mecanismes d'autorització que es fan servir en Hadoop s'acostumen a manifestar de tres formes:

* Permisos tipus POSIX sobre els arxius i directoris,
* Llistes de control d'accés (ACL) per a administrar serveis i recursos,
* Control de l'accés basat en el rol (RBAC) per a alguns serveis que tenen controls avançats d'accés a les dades.

Apache Sentry permet gestionar l'RBAC. La figura següent presenta l'arquitectura de Sentry.



Apache Sentry consta de tres components:

* **Sentry Server**. El servidor sentinella gestiona les metadades d'autorització i proporciona les interfícies que permeten recuperar i modificar les metadades.
* **Data Engine**. El motor de dades fa referència a l'aplicació que ha d'autoritzar l'accés a les dades o metadades.
* **Sentry Plugin**. S'executa en el motor de dades. Proporciona les interfícies que permeten modificar les metadades d'autoritzacions emmagatzemades en el Sentry Server, i inclou el motor de polítiques d'autorització.

Sentry utilitza el sistema d'autenticació subjacent Kerberos per a identificar l'usuari. Sentry és capaç de gestionar el següent:

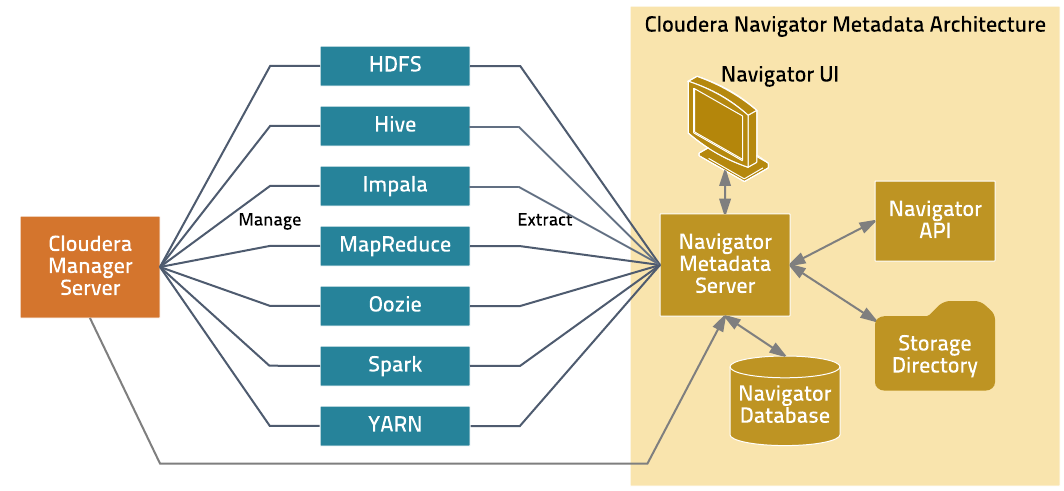
* Usuari. Individual.
* Grup. Un conjunt d'usuaris.
* Privilegis. Una instrucció o regla que permeti l'accés a un objecte.
* Rol. Un conjunt de privilegis.

## Èmfasi en la gestió de dades

La solució escollida per gestionar el cicle de vida de les dades a Hadoop és Cloudera Navigator. Cloudera Navigator proporciona funcions d'auditoria, gestió de metadades i llinatge.

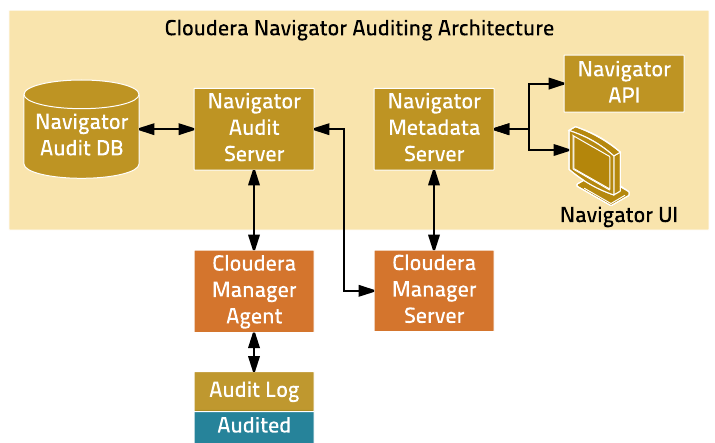
Per proporcionar aquestes característiques, l'arquitectura de Cloudera Navigator es divideix en 2 parts:

* Arquitectura de les metadades de Cloudera Navigator, que proporciona funcions de descobriment de dades i de llinatge de dades.



Descripció dels seus components:

* + Servidor de metadades: s'encarrega dels extractes de metadades, gestiona les polítiques d'extracció, indexa i emmagatzema les metadades, gestiona autoritzacions, auditoria, informes, etc.
  + Base de dades del Navigator: emmagatzema les polítiques, autorització dels usuaris, metadades d'informe d'auditoria i dades analítiques.
  + Directori d'emmagatzematge: emmagatzema l'estat d'extracció i metadades extretes.
  + UI i API del Navigator: proporciona la interfície d'usuari i la interfície de programació d'aplicacions per als usuaris/serveis.
* Arquitectura d'Auditoria de Cloudera Navigator, que proporciona les funcions d'auditoria de dades i d'accés.



Descripció dels seus components:

* + Agent d'administració de Cloudera: supervisa els arxius del registre d'auditories i envia els successos al servidor d'auditoria del Navigator.
  + Servidor d'auditoria del Navigator: Seguiment i agrupació de successos, emmagatzemar successos a la base de dades d'auditoria.
  + Base de dades d'auditoria del Navigator: emmagatzema els successos extrets.

## Èmfasi en la seguretat de l’accès a les dades

Les aplicacions de tercers, introduïdes al capitol 5.1.1, permeten als usuaris d’accedir a les dades sintesi.

Per tractar la seguretat, es necessita gestionar diversos aspectes, tals com:

L’identitat de l’usuari,

L’obtenció del seu perfil i grup d’autorització,

L’autorització dels seus accessos.

Dos protocols de seguretat son utilitzats:

El protocol OIDC per identificar l’usuari i obtenir la seva autorització,

El protocol OAuth per autoritzar els accessos a les dades a través de la solució API Management.

El protocol OIDC permet a les aplicacions de tercers de delegar l’autenticació a la solució IAM.

Per cumplir l’autenticació, l’usuari és redirigit cap a una Web form proporcionada per la solució IAM. El protocol entra les seves credencials i les envia a l’IAM (les credencials no són mai enviades a les aplicacions de tercers). L’IAM valida l’identitat de l’usuari i després envia una petició “callback” a l’aplicació. La petició “callback” conté un codi personal d’identitat (firmat per l’IAM), que demostra l’identitat de l’usuari i un token d’accès que permet a l’aplicació de demanar més detalls sobre l’usuari cap a l’IAM. El perfil i grups d’autorització de l’usuari queden integrats en el token com a “Claims”.

Un cop l’aplicació rep el token, el valida verificant la seva firma. Si la firma és vàlida, l’usuari ja és autentificat.

Els drets “fine grains” de l’usuari es gestionen amb la mateixa aplicació en funció del perfil d’usuari i les seves autoritzacions.

Per accedir a l’API, que proporciona les dades, l’aplicació inicia una seqüència OAuth, basada amb el flux OAuth JWT.

Aquest envia el token ID de l’usuari (rebut de l’IAM) cap a la solució API Management. L’API Management valida el token verificant la seva firma i retorna un token d’accès cap a l’aplicació si aquest és vàlid.

A continuació, l’aplicació utilitza el token d’accès per accedir a l’API.

La figura següent mostra les interaccions entre cada part:

