Indice

Capitolo 1

Legge di Darcy e mezzi porosi

1.1 Descrizione dello scenario

Capitolo 2

Implementazione

Un sistema di Identity and Access Management, d'ora in poi definito IAM, identifica le componenti tecnologiche e procedurali atte a supportare i processi di Identificazione, Autenticazione e Autorizzazione degli utenti che accedono al proprio sistema informativo. Un sistema IAM quindi coadiuva la gestione delle identità digitali di una persona fisica, dal processo di creazione e organizzazione, fino all'eliminazione dell'identità digitale. Ogni singola persona, dipendente di una azienda o utente esterno di un servizio, è oggetto di un processo di assegnazione di molte identità digitali che possono riguardare l'account di posta o l'accesso alla intranet aziendale, in maniera coerente con il suo specifico ruolo aziendale. Nell'Identity Management ciascuna di queste identità digitali va attivata (Provisioning) e disattivata (deProvisioning) in maniera coerente alle politiche aziendali e in modo automatico per evitare errori o ritardi. Ad esempio nel caso di licenziamento di un dipendente l'intero processo di disattivazione dei suoi account deve essere eseguito nel modo più rapido ed efficace possibile. Processi ripetitivi possono essere facilmente eliminati riducendo notevolmente i costi operativi di gestione delle identità digitali grazie all'automazione di processi e servizi self service, ad esempio il processo di rigenerazione di una password smarrita.

Il motivo principale per cui risulta essere conveniente un sistema di questo tipo, soprattutto per realtà di grandi dimensioni, è rendere più gestibile e controllabile l'infrastruttara tecnologica. Al crescere della quantità di utenti da gestire, dei diversi sistemi e risorse a cui questi accedono e al numero di diversi gruppi di amministratori che regolano l'accesso ai sistemi la gestione dell'infrastruttura tecnologica tende a diventare complicata. La complessità del sistema comporta, schematicamente, i seguenti problemi:

- *Difficoltà per l'utente:* Accedere a molti sistemi comporta ricordarsi le diverse credenziali di accesso e le diverse peculiarità per sottomettere le stesse ai sistemi (interfaccia web, login via terminale, login integrata con il sistema operativo, ...)

- *Difficoltà di gestione:* Consentire agli utenti l'accesso alle risorse in maniera da garantire sempre il minimo privilegio e revocare i privilegi di accesso quando non più necessario, il tutto in maniera facilmente monitorabile dalle funzioni aziendali preposte (processo di audit)

- *Operazioni di supporto:* Monitoring degli accessi, operazioni di mantainance (rimozione utenti, cambio password, ...)

- *Sviluppo di meccanismi omogenei di autenticazione:* L'accesso a ambienti disomogenei comporta la scrittura da parte degli sviluppatori di meccanismi di autenticazione propri per ogni applicazione

A questo si sono aggiunte la preoccupazione per la tutela della Privacy e i recenti scandali finanziari (Enron, Parmalat, Cirio ...) che hanno convinto i legislatori di tutti i paesi a considerare in modo sempre più incisivo la tutela dei dati fino a emendare leggi sempre più severe atte a garantire la disponibilità, integrità e Riservatezza delle informazioni e dei dati trattati.

Figura 2.1: Evoluzione dell'architettura di una rete tipica

I sistemi di IAM nascono quindi dall'esigenza di migliorare la gestione dell'infrastruttura tecnologica di un'azienda, permettendo di rafforzare il rispetto delle policy di accesso ai sistemi e fornendo tutti i servizi a supporto necessari, quali:

- Servizi a supporto della creazione, revoca, propagazione e sincronizzazione delle "Identità Digitali"

- Servizi di controllo accessi basati sul ruolo (RBAC)

- Servizi per armonizzare la gestione delle identità con il modello organizzativo aziendale

- Servizi per gestire Workflow e processi di approvazione.

2.1 Tipologie di IAM

MANCA

2.2 Componenti logiche

In un sistema IAM un concetto fondamentale è *l'identà digitale* di ogni utente del sistema. L'identità digitale di un utente non è altro che una rappresentazione del suo profilo che comprende tutte le informazioni ed attributi che lo caratterizzano (matricola, unità organizzativa, ruolo, mansione...), le sue credenziali di accesso, i servizi cui l'utente deve essere abilitato e i diritti della persona su ogni servizio.

Gli strumenti coinvolti in un sistema di gestione delle Identità Digitali possono essere così suddivisi:

1. Provisioning
2. Work Flow
3. User Self Service e password management
4. Delegated Administration
5. Change Reconciliation
6. Audit and Reporting
7. Single Sign On

In figura 2.2 è possibile vedere come questi moduli interoperino tra loro.

2.2.1 Provisioning

Il servizio di Provisioning è l'insieme di servizi di gestione della vita di un utente all'interno del sistema. La componente di Provisioning in un sistema di Identity and Access Management deve:

- Consentire la creazione, la manutenzione e la revoca (de-provisioning) automatica degli utenti;
- Prevedere l'utilizzo di un servizio centralizzato per la definizione dei diritti di accesso;
- Supportare la comunicazione asincrona con i sistemi target;
- Consentire di definire e di monitorare le dipendenze di provisioning

Figura 2.2: Schema architett

2.2.2 Workflow

Il servizio di Workflow supporta i processi di autorizzazione in un sistema di Identity and Access Management. Un servizio di Workflow deve poter:

- Rendere automatico i processi di approvazione e provisioning;
- Garantire processi di approvazione basati su Gruppi e gestire l'escalation dei ruoli definiti;
- Fornire strumenti di controllo, anche grafico, per lo stato delle richieste;
- Gestire processi di approvazione sia paralleli che sequenziali.

2.2.3 User Self Service e password management

Gli strumenti di User Self Service consentono agli utenti di gestire in modo automatico e personalmente le proprie credenziali in modo che questi possano, in autonomia aggiornare delle informazioni relative al proprio profilo di accesso, Modificare, sincronizzazione e resettare le proprie password, formalizzare le richieste di abilitazioni e nuovi accessi a risorse non contemplate nel proprio profilo utente e altro ancora. Gli strumenti di Password Management forniscono inoltre la possibilità di gestire in modo centralizzato tutte le tematiche relative alle password delle risorse aziendali. Grazie ad essi si possono rafforzare le policy aziendali relative alla durata massima della password, lunghezza minima, tentativi ammissibili prima di bloccare di bloccare l'account etc....

2.2.4 Amministrazione delle deleghe

L'amministrazione di una rete di grandi dimensioni può risultare onerosa per un solo soggetto: è quindi utile delegare ad altri utenti non privilegiati alcuni compiti. I servizi di "Amministrazione delle Deleghe" consentono di impostare le regole che stabiliscono quali operazioni e con che regole possano essere delegate. Con questi strumenti un Data Base Administrator può per esempio abilitare la completa gestione di uno specifico data base ad una classe di utenti, e revocare il permesso una volta che il compito sia terminato in maniera controllata, e potenzialmente usando un'interfaccia web di controllo.

2.2.5 Change reconciliation

La gestione centralizzata di tutte le regole che gestiscono le identità digitali consentono la diffusione semplice e immediata di cambiamenti. In generale servizi di "Change Reconciliation" consentono findamentalmente di standardizzare la diffusione/erogazione di modifiche legate all'identità di un

determinato insieme di utenti a tutte le risorse dello stesso insieme o verso altri insiemi e di gestire queste modifiche secondo delle regole ben definite.

2.2.6 Audit and reporting

Le funzioni di logging e reportistica di un sistema di Identity and Access management risultano indispensabili ai fini di auditing e risultano il nucleo per tutto ciò che riguarda i benefici dimostrabili dell'adozione del sistema stesso. Il sistema devo poter consentire di:

1. Censire tutte le operazioni e gli eventi (logging dettagliato);
2. Realizzare reportistica personalizzata con i log;
3. Realizzare reportistica che possa essere protetta, segmentata e filtrata per contenuti;
4. Estendere le funzionalità di reportistica a valle di nuove richieste dell'azienda.

2.2.7 Single Sign On

Il Single sign on (SSO, traducibile come autenticazione unica o identificazione unica) è un sistema specializzato che permette la connessione degli utenti alle applicazioni aziendali con un'unica autenticazione, offrendo una fruibilità ininterrotta per quanto riguarda tutti i tipi di applicazioni aziendali. Segue la descrizione tratta da Wikipedia [**?**]

Architettura

Vi sono tre approcci per la creazione di un sistema di SSO: l'approccio centralizzato, l'approccio federativo e l'approccio cooperativo.

Approccio centralizzato

Il principio è di disporre di un database globale e centralizzato di tutti gli utenti e di centralizzare allo stesso modo la politica della sicurezza. Questo approccio è destinato principalmente ai servizi dipendenti tutti dalla stessa entità, per esempio all'interno di una azienda.

Approccio federativo

Con questo approccio differenti gestori (federati tra loro) gestiscono dati di uno stesso utente. L'accesso ad uno dei sistemi federati permette automaticamente l'accesso a tutti gli altri sistemi.

Un viaggiatore potrebbe essere sia passeggero di un aereo che ospite di un albergo. Se la compagnia aerea e l'albergo usassero un approccio federativo

avrebbero un accordo reciproco sull'autenticazione dell'utente. Il viaggiatore potrebbe ad esempio autenticarsi per prenotare il volo e essere autorizzato, in forza di quella sola autenticazione, ad effettuare la prenotazione della camera d'albergo.

Questo approccio è stato sviluppato per rispondere ad un bisogno di gestione decentralizzata degli utenti: ogni gestore federato mantiene il controllo della propria politica di sicurezza.

Approccio cooperativo

L'approccio cooperativo parte dal principio che ciascun utente dipenda, per ciascun servizio, da uno solo dei gestori cooperanti. In questo modo se si cerca, ad esempio, di accedere alla rete locale, l'autenticazione viene effettuata dal gestore che ha in carico l'utente per l'accesso alla rete.

Come per l'approccio federativo, in questa maniera ciascun gestore gestisce in modo indipendente la propria politica di sicurezza. L'approccio cooperativo risponde ai bisogni di strutture istituzionali nelle quali gli utenti sono dipendenti da una entità, come ad esempio in università, laboratori di ricerca, amministrazioni, etc...

2.3 Soluzioni presenti sul mercato

Quando si parla di gestione delle identità è difficile prescindere dal mercato per determinare le diverse tipologie di soluzioni disponibili. Essendo un argomento molto attuale (e remunerativo) le aziende che propongono soluzioni di livello cosiddetto enterprise offrono delle soluzioni complete per la gestione dell'identità, spesso integrabili facilmente con altri prodotti presenti nei propri cataloghi. In figura2.3 è mostrato l'insieme dei potenziali moduli che compongono una soluzione IAM, con le specifiche tecnologie impiegabili nelle singole componenti.

I produttori di maggior rilievo che attualmente propongono soluzioni complete di gestione delle identità comprendono Sun[1], ComputerAssociates[2], Novell[3], Microsoft[4], Entrust[5] e Oracle[6].

I componenti fondamentali di un sistema di IAM sono comuni a tutte le solzioni sopraccitate, e anche alla soluzione oggetto del lavoro di tesi che, come vedremo nei prossimi capitoli, è una soluzione che mutua le proprie componenti da diversi prodotti commerciali.

[1] http://www.sun.com
[2] http://www.ca.com
[3] http://www.novell.com
[4] http://thesource.ofallevil.com
[5] http://www.entrust.com/
[6] http://www.oracle.com

Figura 2.3: Tecnologie impiegate i

Capitolo 3

Classi Base

3.1 Descrizione dello scenario

Capitolo 4

Problema numerico

4.1 Descrizione dello scenario

Capitolo 5

Classi Fratture

5.1 Descrizione dello scenario

Capitolo 6

Il caso Unicredit - titolo provvisorio

La realtà in analisi è un grosso gruppo finanziario internazionale. L'ambiente progettato e realizzato durante il mio stage aziendale elabora i dati di circa 90000 dipendenti dell'azienda, ed è predisposto per gestirne altrettanti aggiuntivi, alla luce di una grossa fusione avvenuta di recente. Il sistema di gestione delle identità realizzato è subentrato gradualmente all'architettura precedentemente esistente e, di conseguenza, la sua progettazione ha dovuto considerare la possibilità di inserimento graduale nel ciclo di vita di tutto il resto dell'architettura aziendale.

Subentrare ad una realtà già esistente è stato un notevole vantaggio perchè ha permesso di analizzare quanto già creato e migliorare i difetti più evidenti mantenendone i punti di forza.

6.1 Ambiente AS-IS

In figura6.1 è illustrata la soluzione completa attualmente funzionante. La maggior parte delle componenti di IAM presenti è implementata con prodotti di Computer Associates [1].

Il sistema si occupa della gestione del personale di quasi tutto il gruppo bancario. Alcune banche e assicurazioni controllate non sono state ancora inserite, benchè la loro integrazione sia prevista entro la fine del 2009. Un'immagine rappresentativa delle prestazioni ottenute è rappresentata in figura 6.2

La gestione delle utenze avviene a partire da quanto presente nei data base degli uffici del personale (HR). L'ufficio HR inserisce per ogni dipendente un profilo contenente tutte le informazioni necessarie in base alle policy della singola azienda in un data base. Normalmente i DB utilizzati sono ospi-

[1] CA - http://www.ca.com

Figura 6.1

tati su mainframe e principalmente sono di tipo DB2 [**?**], anche se alcune fonti usano SQL server 2005 [**?**] e perfino flat file. Le basi di dati vengono denominate *fonti autoritative*, per evidenziarne il ruolo chiave che questi dati rivestono ai fini della definizione del profilo utente. La soluzione implementata comporta il minor impatto possibile sulla struttura già presente in tutte le realtà che compongono il gruppo finanziario, nella totalità dei casi ci si limita a leggere i dati forniti senza apportare alcuna modifica alla sorgente. I dati letti dalle sorgenti vengono elaborati dal nucleo del sistema di IAM, componente che è stato reingenierizzato durante lo stage di laurea, e vengono generate delle istruzioni di modifica del profilo utente in relazione alla situazione consolidata dell'esecuzione precedente. Queste istruzioni comportano delle sequenze di comandi sui profili degli utenti, che consistono in operazioni di cancellazione, modifica o aggiunta di attributi. Le istruzioni e il meccanismo di modifica del profilo verrà trattato approfonditamente nel seguito della relazione. Il sistema target di tutte le azioni di modifica è un directory LDAP [**?**] che contiene un sottoalbero per ogni utente del sistema, la modifica di un attributo comporta una serie di operazioni che vengono portate a termine sui server in produzione, configurando l'immagine di ogni utente come richiesto dal sistema di IAM. La scelta di avere un server LDAP come interfaccia tra lo IAM e i sistemi target è molto utile perchè permette di rendere indipendenti i due ambienti. Alla luce di questa decisione implementativa è risultato facile sostituire il blocco dell'ETL custom precedentemente in essere con una soluzione diversa senza rompere l'equilibrio degli altri componenti dell'architettura.

Il nucleo del framework di IAM è composto da tre macro-componenti che svolgono, ciascuna, una parte delle elaborazioni necessarie per portare a termine in modo efficace ed efficiente l'interno processo di provisioning delle identità e di gestione degli accessi:

- **ETL**: la componente di ETL (Extract, Transform, Load) si occupa dell'estrazione, trasformazione e caricamento dei dati in un sistema di sintesi. I dati vengono estratti da sorgenti quali database transazionali (DB2, SQL Server), comuni file di testo (presenti in locale o prelevati da sistemi remoti) o da altri sistemi informatici che forniscono le informazioni raccolte dai sistemi di HR (Human Resource). I dati prelevati subiscono quindi un processo di trasformazione, che consente ad esempio di selezionare solo quelli che sono di interesse per il sistema, normalizzare i dati, derivare nuovi parametri, cercare correlazioni tra dati recuperati da differenti tabelle, etc. Tale trasformazione ha lo scopo di rendere omogenei dati provenienti da sorgenti diverse e di fare in modo che siano più aderenti alla logica di business del sistema per cui viene sviluppato il processo di provisioning. L'ultima operazione svolta dalla componente ETL è quella di determinare le variazioni pre-

senti sui profili normalizzati rispetto alla situazione presente in Admin e consolidata in precedenza.

- **Transaction**: la componente di Transaction ha il compito di prelevare le informazioni generate dal modulo ETL e di procedere con l'aggiornamento dei profili utente presenti in Admin applicando le trasformazioni indicate, così da allineare il profilo utente su Admin con la situazione fornita dai sistemi di HR.

- **Admin**: eTrust Admin è una soluzione robusta e scalabile per il provisioning di utenti tra molteplici ed eterogenei sistemi (Active Directory, LDAP, RACF, etc.). Questo meccanismo consente, attraverso una corretta configurazione di policy, la creazione, la modifica o la rimozione degli utenti e dei relative oggetti sui diversi sistemi di un organizzazione. Il vantaggio offerto da questa soluzione è che astrae i sistemi target, presentando le principali caratteristiche in formato LDAP.

Mentre l'ultima componente, eTrust Admin, è un prodotto commerciale di Computer Associates installato e configurato in modo appropriato, le prime due sono frutto di attività di analisi, progettazione e implementazione svolte appositamente per l'ambiente presente in azienda.

La parte di cui mi sono occupato principalmente riguarda il modulo di ETL perché è la parte che più rappresentava un problema in termini di prestazioni e scalabilità di tutta l'architettura esistente. Quella che segue è una disamina dettagliata del processo eseguito da ETL e Transaction, ovvero dalle due componenti custom dell'infrastruttura IAM in questione.

6.1.1 Soluzione precedente

Il modulo ETL è stato implementato con del codice custom, realizzato da un gruppo di programmatori Java e progettato per lavorare strettamente con eTrust admin attraverso l'interfacciamento LDAP. Lo stato del sistema target era mantenuto mediante l'ausilio di un database di appoggio di tipo INGRES [**?**], in cui era memorizzato un record per ogni utente e un hash che identificasse lo stato di ogni riga. Lo pseudocodice è presentato nel listato 1, e descrive le operazioni di elaborazione effettuate per ogni fonte autoritativa.

La trasformazione così implementata presenta diversi punti deboli e inefficienze, che si è cercato di risolvere con la nuova implementazione:

1. La soluzione non è scalabile

2. Il codice è difficilmente riutilizzabile, vista la diversa natura delle sorgenti

3. Non c'è una correlazione forte tra quanto presente nel directory di admin e l'immagine consolidata dell'utente

```
Connessione alla fonte autoritativa;
while esistono utenti non elaborati do
    Estrai tutte le informazioni presenti;
    Normalizza il profilo rispetto al consolidato ;
    Applica le policy e ricava i privilegi ;
    Calcola Hash ;
    if Hash != Hash consolidato then
        ModificaProfiloLDAP(ProfiloUtente);
        if Modifica_riuscita then
            ConsolidaProfilo ;
        end
    end
end
```

Algorithm 1: Soluzione di ETL precedente

4. Il codice è difficilmente manutenibile perchè la curva di apprendimento è troppo ripida

5. Non è possibile dividere il calcolo delle modifiche e l'esecuzioni delle operazioni in fasi distinte

6. Eventuali errori di acquisizione dei dati possono portare a modifiche massive e indesiderate dei profili utente

7. É prevista solo un'esecuzione di tipo batch notturna, quando i dati non vengono modificati

Il raddoppio della base di utenza è stato il motivo principale che ha spinto il passaggio dal meccanismo sopra descritto a uno più flessibile, che garantisse degli standard di scalabilità e manutenibilità in linea con il resto dell'architettura di IAM. Un altro punto debole significativo è la resistenza a errori nei dati in ingresso. Potrebbe potenzialmente capitare (ed è capitato) che delle mancanze di dati nelle sorgenti scatenassero processi di rimozione massiva indesiderata di profili utente dai sistemi target, comportando fastidiosi disservizi.

6.2 Nuova implementazione del modulo ETL

Come introdotto poco sopra, il compito della componente di ETL è quello di prelevare i dati provenienti da HR sotto forme e da sorgenti diverse per poi procedere con la loro normalizzazione, il completamento degli stessi con informazioni prelevate da tabelle aggiuntive, la rilevazione delle modifiche sui profili utente e la generazione dell'input necessario alla Transaction

per allineare, su Admin, i profili degli utenti rilevati come modificati. Tutte queste operazioni devono essere fatte rispettando dei vincoli di tempo e in modo parallelo, riducendo al minimo (possibilmenete eliminando) gli svantaggi dell'architettura precedentemente in essere.

Lo strumento scelto come base di tutto l'ETL è SQL Server Integration Services [**?**], che consiste in un modulo integrato in Visual Studio per elaborare dati provenienti da data base. Il prodotto in questione risulta essere molto comodo perchè permette agilmente di:

- Trattare sorgenti eterogenee (file, DB2, fogli excel, SQL server) in modo omogeneo
- Lavorare in modo visuale sui dati
- Utilizzare il linguaggio SQL (Transaction-SQL) quando necessario
- Compiere elaborazioni complesse usando dei linguaggi di programmazione (Vb.net, VBS)

A parte la propaganda commerciale che, si sa, spesso lascia il tempo che trova, SSIS si è dimostrato essere uno strumento robusto e affidabile, che ha davvero reso facile l'implementazione della logica applicativa. Lavorare sui dati ingresso sempre sotto forma di tuple e con gli strumenti tipici del data mining ha permesso inoltre di avere delle performance molto buone in termini di tempo di elaborazione e soprattutto ha permesso di mantenere il codice dell'applicazione spesso immediato da capire.

In figura6.3 si può vedere lo schema logico di lavoro della nuova implementazione del modulo ETL.

6.2.1 Concorrenza nell'accesso ai dati

6.2.2 Scalabilità

Figura 6.3: Flusso logico ETL