

Tesi di laurea

Candidato: Giampiero Bono

Relatore: Paolo Prinetto

Titolo: Piattaforma cloud per servizi LIS: progetto e implementazione in Windows Azure

Il lavoro di questa tesi è stato svolto al Microsoft Innovation Center (MIC) di Torino all'interno dell'Istituto Superiore Mario Boella (ISMB) e consiste nella progettazione e nella realizzazione di una piattaforma per servizi LIS basata su tecnologia cloud-computing, nello specifico Microsoft Windows Azure.

In particolare l'idea è di elaborare e sovrapporre in modo automatico normali filmati con attori virtuali (avatar) al fine di generare e distribuire su diverse piattaforme (es. smartphone) contenuti a supporto di persone sordo-mute. L'obiettivo è quindi quello di realizzare la parte di elaborazione finale e distribuzione agli utenti, di un sistema più ampio in grado di tradurre automaticamente tracce audio di filmati nel linguaggio dei segni (LIS – Language In Sign), mediante un attore virtuale.

L'avatar in questione è realizzato con un'accurata grafica 3D, utile a riprodurre fedelmente tutte le parti del corpo necessarie per un'esecuzione fluida di tutti i segni del linguaggio. La tecnica nasce all'interno del progetto Automatic Translation into sign LanguageS (ATLAS), che ha l'obiettivo di realizzare un sistema avanzato di generazione automatica di questi attori virtuali, analizzando semanticamente i contenuti video di un filmato e prelevando da un database i segni necessari per ottenere una traduzione fedele.

La piattaforma, oggetto di questa tesi, è stata progettata secondo criteri di sviluppo di tipo enterprise, ricercando la massima flessibilità, scalabilità e robustezza. Il pattern architetturale scelto è stato quindi un enterprise 3-tier su cloud computing PaaS così organizzato:

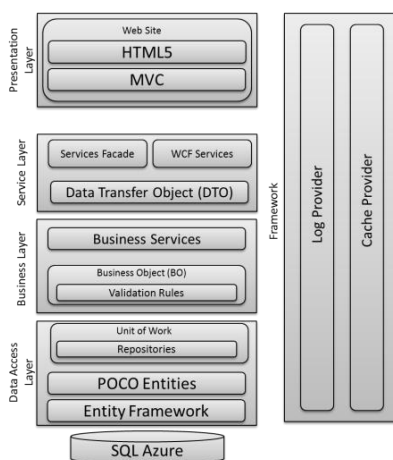


Figura 1 - Pattern architetturale

- **Data Access Layer (DAL):** accesso ai dati memorizzati nel database. Per questa tesi si è scelto di utilizzare *Entity Framework* così da avere un mapping diretto tra oggetti relazionali e classi.
- **Business Logic Layer (BLL):** svolge le funzioni principali, validando e operando sui dati provenienti dal DAL.
- **Presentation Layer (PL):** presenta all'utente i risultati provenienti dal BLL.

La piattaforma si compone, secondo la terminologia propria di Windows Azure, di un **web role** ed un **worker role**, ovvero di un componente che ospita un web server e uno che si occupa di elaborazione in background. Il web role consente all'utente la configurazione di parametri utili al processing dei video; il worker role, processa, in maniera automatica, coppie di video: il filmato originale (video master) e la corrispondente traduzione in LIS (video avatar).

Il vantaggio più grande dalla scelta di una soluzione cloud è la possibilità di realizzare una piattaforma sicura, affidabile e flessibile senza la necessità di affrontare i costi di acquisto e gestione di strumenti hardware e software e di manutenzione dell'infrastruttura. Il cloud offre potenza di calcolo completamente adattabile alle necessità in base al carico di lavoro, evitando rischi e costi dovuti a sovrastima/sottostima dei requisiti dell'applicazione.

Tenendo conto dell'aspetto economico si è deciso di non tenere sempre attiva l'istanza del worker role, attivandola solo nel momento in cui è presente del lavoro da svolgere. Per questo sono stati gestiti l'accensione e lo spegnimento delle istanze di calcolo, ovvero la creazione e l'eliminazione del deployment in cloud, mediante l'uso delle **Service Management API**. Attraverso queste API è possibile accedere da codice a tutte le funzioni presenti sull'Azure Management Portal, da cui si amministrano le diverse istanze e i deployment. Per utilizzare queste API è necessario creare richieste http opportune, contenenti specifici header e body, oltre a una URI dedicata composta da diversi parametri come la subscription-id, il nome DNS associato al servizio ed altre parti che dipendono dal tipo di richiesta da effettuare. A ognuna delle richieste è obbligatorio allegare un certificato X.509 per autenticare e rendere sicura la connessione. Il certificato deve essere caricato in precedenza sull'account Azure, per far sì che la richiesta sia presa in carico dal servizio che altrimenti la rifiuterebbe in quanto proveniente da un dispositivo non riconosciuto.

La procedura implementata si compone di diverse fasi. Il web role si collega a un web service esterno, da cui riceve un dataset contenente informazioni su diversi video, tra i quali cerca delle coppie (master + avatar) che rispettino le specifiche date (il tutto utilizzando una query LINQ sul dataset). Successivamente controlla che i video selezionati non siano già stati elaborati in precedenza, andando a verificare che l'identificativo del video non sia presente nel database SQL Azure. Nel caso ci siano video non ancora processati, il web role crea un messaggio, con tutte le informazioni utili per permettere al worker role di processarli, che viene trasmesso in modo asincrono attraverso una Azure Queue (componente dello Azure Storage), utilizzata dai due ruoli per comunicare. Il web role attiva quindi l'istanza del worker role, creando un nuovo deployment. Una volta avviato il worker role verifica la presenza di messaggi all'interno della Azure Queue e, in caso affermativo, estrae e processa il messaggio dal quale ottiene tutte le informazioni necessarie per poter iniziare l'elaborazione, al termine della quale si ottengono 5 diversi video: quattro con



Figura 2 - Effetto chroma key

l'avatar posizionato ai diversi angoli del video master e l'ultimo con l'avatar a schermo intero. A questo punto il worker role si autosospinge e successivamente il web role, non appena si accorge del completo spegnimento, procede all'eliminazione del deployment precedentemente creato. Una volta pronti, i video vengono salvati sullo Azure Storage a disposizione per la fruizione.

Per l'elaborazione dei video ci sono stati utilizzati due tool di processamento audio/video: **FFMpeg** e **ImageMagick**. Il primo serve per effettuare la conversione dei video nei formati dall'utente attraverso il web role. Il secondo serve per eseguire l'elaborazione dei frame per sovrapporre l'avatar al video originale. L'avatar 3D, che effettua la traduzione, è infatti realizzato con uno sfondo verde (RGB(0, 127, 0)) che necessita di essere eliminato prima di effettuare la sovrapposizione con il video master. La tecnica in questione, detta di **chroma key** e facilmente realizzabile con software di video-editing professionali, ha presentato una delle difficoltà superate nella realizzazione di questa tesi, in quanto è stato complicato trovare e configurare un software che fosse eseguibile da linea di comando. La soluzione scelta consiste nell'estrarre, in formato PNG, tutti i frame del video avatar per poi elaborarli singolarmente per eliminarne lo sfondo, e infine sovrapporli ad ogni frame del video master. La rimozione dello sfondo rappresenta il collo di

bottiglia per le performance di tutta la catena di processamento. Per velocizzarla è stato scelto di sottocampionare i frame estratti dal video avatar, riducendo così il numero di immagini da elaborare pur mantenendo una qualità accettabile in riproduzione. Il sottocampionamento porta però una riduzione del framerate del video avatar richiedendo quindi un intervento per risincronizzare i video da sovrapporre. La soluzione adottata è quindi di duplicare i frame del video avatar dopo il processamento riportandone il totale al valore originario e ottenendo un video finale perfettamente fluido.

Per rendere disponibili ai dispositivi mobili le informazioni riguardanti i video pronti per lo streaming, è stato realizzato un web service WCF che genera un RSS feed. Per effettuare i test si è utilizzato un dispositivo Windows Phone 7 con installata un'applicazione che contatta il servizio, scarica l'RSS, lo parsifica e popola una lista con i video disponibili. Dopo aver selezionato un video l'utente può decidere di modificare la posizione dell'avatar senza dover interrompere la riproduzione del video.

La scelta di utilizzare Windows Azure per questa tesi non è stata casuale, bensì guidata da una ricerca effettuata tra i diversi provider cloud: Microsoft, Amazon e Google. La scelta di escludere Amazon e Google è dipesa dal tipo di servizio da loro offerto. Amazon, infatti, permette di noleggiare macchine virtuali che richiedono una completa fase d'installazione e configurazione prima di essere operative, non necessario per lo scopo della tesi. Google, nonostante offrisse un servizio simile a quello Microsoft, è stato scartato per via della poca flessibilità dell'ambiente di deployment offerto e per il fatto che non permetteva di utilizzare C# come linguaggio di programmazione.

Per il futuro si potrebbe pensare di effettuare un porting completo sul cloud dell'intera piattaforma del progetto ATLAS. Potrebbe essere la scelta ideale visti i numerosi vantaggi che verrebbero introdotti. Infatti, potendo facilmente aumentare la potenza di calcolo, si potrebbe rendere ancora più "automatica" la traduzione in LIS dei contenuti, realizzando un traduttore speech-to-text per convertire il parlato in testo, così da essere passato al traduttore text-to-signs che effettuerà la traduzione generando l'attore 3D. Questo permetterebbe una più veloce creazione dei contenuti, facendone aumentare la quantità disponibile. L'aumento dei contenuti, con l'aggiunta dei sottoprodotti intermedi delle elaborazioni, richiederebbe uno storage molto più grande e soprattutto affidabile ed anche in questo caso il cloud risulterebbe essere la soluzione migliore.