

Lezione 8 modulo 1

In questo modulo e nei due moduli successivi parleremo di integrazione di moduli applicativi. Innanzitutto introdurremo, diremo, che cosa vuol dire integrare, nell'ottica di avere una delle tecnologie di supporto a questa fase di integrazione e poi discuteremo alcuni aspetti che sono rilevanti per questo tipo di integrazione; in particolare parleremo di cosa vuol dire integrare dei dati, poi parleremo di tecnologie di interazione per scambiarsi messaggi fra moduli applicativi, declinato in varie modalità, comprese le modalità relative all'utilizzo di servizi. Quali sono le problematiche che noi dobbiamo affrontare? Innanzitutto, appunto, vogliamo capire perché vogliamo integrare; premessa: quando parliamo di quasi integrazione non intendiamo integrare tutto all'interno di una singola applicazione; abbiamo già visto l'esempio nel caso della tecnologia basata sulle API che questo è il tentativo che viene fatto; il tipo di sistemi: noi abbiamo un'architettura, per cui, sostanzialmente, noi andiamo a costruire un sistema che fornisce delle funzionalità basato su delle funzionalità che sono di integrazione tecnologica dei dati, quindi avremo un DBMS che sarà alla base dell'integrazione dei dati e quindi, concettualmente, tutti i dati che vengono utilizzati dalle funzionalità, il sistema ERP, sono integrati, ovviamente; poi il DBMS a sua volta utilizzerà una base tecnologica all'interno del sistema operativo; questo dal punto di vista concettuale, ovviamente, poi dal punto di vista tecnologico può essere distribuito e questa è una rappresentazione, diciamo, di tipo logico di questo sistema; lo scopo è proprio quello, come abbiamo visto, di fornire funzionalità che sono funzionalità integrate facendo vedere una vista unica dei dati, quindi abbiamo una integrazione dei dati all'interno del sistema. Come abbiamo però visto, abbiamo visto altre tecnologie che sono state sviluppate a livello applicativo per fornire funzionalità avanzate per problematiche specifiche: per esempio nel caso del CRM, noi abbiamo visto che vogliamo dare delle funzionalità che sono tipiche della gestione del cliente e della gestione di campagne di contatto con i clienti di servizi forniti al cliente per risolvere i problemi e così via; come per il CRM abbiamo anche discusso più in dettaglio alcuni aspetti tecnologici che riguardano i datawarehouse e abbiamo anche citato altri tipi di sistemi che si sono sviluppati di tipo specialistico, per esempio Supply Chain Management e altri che abbiamo visto nelle lezioni precedenti. Quindi, cosa succede? Che noi ci ritroveremo comunque all'interno del nostro insieme di tecnologie disponibili a livello applicativo, come una serie di moduli; una serie di moduli che saranno appunto un modulo che sarà dedicato alla gestione integrata dei dati dell'azienda, ma sarà affiancato anche in alcuni casi da un modulo CRM che fornisce funzionalità avanzate per gestire i clienti e così via, datawarehouse e altri a seconda di quelle che sono state le scelte dell'azienda. Il problema è che questi sistemi scambiano dati fra di loro perché non sono i dati, non sono separati, ma sono dati che fanno parte di un unico insieme di servizi che l'azienda dà ai propri clienti e che sono già al proprio interno. Quindi, quando parleremo, ad esempio, di clienti è chiaro che i clienti li verremo a gestire all'interno del CRM ma li avremo anche rappresentati all'interno di ERP perché avremo informazioni, ad esempio, su tutte le transazioni di tipo, ad esempio, contabile che sono state fatte con i clienti; anche detto warehouse, proprio per definizione, raccoglierà dati che provengono da altri sistemi fornendo delle funzionalità avanzate per la loro analisi. Un altro punto che dobbiamo sottolineare, quindi, diciamo qua: quello che abbiamo visto è che abbiamo un problema aperto: il fatto che questi diversi sistemi hanno i loro dati, però dovranno scambiarsi delle informazioni e queste informazioni in parte si sovrappongono. Un altro problema che noi andremo studiare a mano a mano nel corso e abbiamo già discusso all'interno del framework bot e che noi avremo sia delle funzionalità all'interno dell'organizzazione e tutto quello che abbiamo visto finora è all'interno dell'organizzazione e funzionalità che invece sono fra organizzazioni, quindi inter-organizzazione. È chiaro che in un'ottica che abbiamo già introdotta a servizi, alcune delle funzionalità che l'azienda utilizza, potrebbero essere fornite come servizi da altre aziende, quindi, quando stiamo parlando di gestione dei clienti, questo modulo CRM potrebbe essere un modulo interno all'azienda oppure potrebbe essere fornito da altre organizzazioni e quindi essere un caso di partecipante a un processo di business, ad esempio, che si presenta come un intermediario, oppure come un fornitore o un cliente. Quindi, quando parleremo delle tematiche diverse che si presentano, integrazione e possibili soluzioni tecnologiche, vogliamo coprire un po'



un vasto spettro che abbiamo di temi di integrazione. Il nostro obiettivo non è quello di prendere tutto e creare un unico sistema, appunto questo è stato un tentativo fatto da alcune tecnologie di tipo ERP, ma il nostro sistema è fare colloquiare questi sistemi, cioè come e ERP e CRM potranno scambiarsi dati, che cosa vuol dire e lì sono, soprattutto, le tecnologie di supporto per questo. Cominciamo a trattare un aspetto che è fondamentale in tutti i progetti di integrazione e in particolare l'aspetto di integrazione dei dati. Cominciamo a vedere quali possono essere i problemi da risolvere e quali possono essere le situazioni, discutendole secondo una classificazione che viene proposta in letteratura discutendo prima il formato e i dati e poi la semantica dei dati. Vediamo cosa intendiamo con questi termini, facendo qualche esempio e discutendo, poi, successivamente, una casistica più ampia. Allora, innanzitutto, quando parliamo di formato dei dati vuol dire avere rappresentato, all'interno del nostro sistema, i nostri dati secondo quelle che saranno le caratteristiche della rappresentazione di quel dato, con una certa struttura e con una certa presentazione. Cosa vogliono dire questi tre termini che vengono definiti in letteratura? Supponiamo di voler rappresentare un indirizzo, ad esempio un indirizzo, quello del Politecnico di Milano, piazza Leonardo da Vinci 32, 20133 cap, Milano, Italia. È chiaro che qui abbiamo alcuni elementi: abbiamo il nome di una via, in realtà qua una piazza, numero civico, poi il cap, abbiamo la città e il Paese; questo può essere rappresentato all'interno di un sistema informativo in tanti modi diversi e vediamo qua, ad esempio: un'unica stringa, oppure, strutturato in campi, in cui mettiamo prima l'indirizzo, inteso come piazza Leonardo da Vinci, ad esempio qua poniamo di aver scelto di aver messo insieme via e numero civico, poi il cap, poi la città e poi il Paese. Questa è una delle tante possibili strutture; ovviamente avrei potuto anche decidere di avere prima la via, poi il numero civico, poi il cap... Potrei anche decidere di non mettere la città, perché il numero di cap identifica univocamente anche la città e la stessa cosa vale per il Paese. Quindi, la prima cosa che vogliamo decidere è che struttura dare, dal punto di vista concettuale: quindi, quali sono gli elementi che vogliamo trattare. Quando poi parleremo di struttura che è il punto successivo, invece parleremo anche dell'ordine di questi elementi. Quindi, quali sono i vantaggi e gli svantaggi di avere l'indirizzo rappresentato in un unico campo o in tanti campi diversi? Ovviamente poi questi campi potranno essere rappresentati all'interno di un database, in una struttura tabellare, ad esempio in un sistema relazionale, oppure - e questo sarà molto importante per il colloquio fra sistemi - essere la base per la struttura di messaggi, quindi vorrò scambiarmi indirizzi, ad esempio dei miei clienti e voglio farlo in modo da avere indirizzi che sono comprensibili all'altro sistema, quindi noi vogliamo un'automazione di questo scambio di messaggi. Allora, quando parliamo dell'indirizzo, ovviamente noi stiamo parlando dell'indirizzo Italiano, è chiaro che alcune regole, composizioni di indirizzo, possono variare a seconda della località geografica: ad esempio quando parliamo di piazza, tipicamente in Italia abbiamo vie e piazze, ma in realtà poi abbiamo altri termini che utilizziamo: indirizzo locale ad esempio calle, campo eccetera, che potrebbero rientrare all'interno di un indirizzo italiano. Abbiamo questa struttura in Italia, ma in altri posti abbiamo altri modi di rappresentare gli indirizzi, altri modi di rappresentare quello che per noi è il cap, quindi un identificativo di zona, altri modi di rappresentare la località che possono variare; ad esempio, è tipico in una località degli Stati Uniti indicare sia la località sia lo stato, che poi il Paese, USA. Quindi, abbiamo tanti modi di rappresentare indirizzi; è chiaro, che se io voglio essere generale e comprendere qualunque tipo di indirizzi, il formato stringa è quello più comodo, perché non avrò mai il problema di aver preso una decisione; ad esempio, cosa succede se non ho nel mio formato degli indirizzi una struttura in via e numero civico? Cosa succede se non ho un cap, oppure se il cap ha strutture completamente diverse, eccetera? Quindi, la rappresentazione in un'unica stringa è sicuramente più comoda, ma è anche molto difficile poi da utilizzare in uno scambio di informazioni fra sistemi, perché il sistema che riceverà questi dati dovrà riconoscere, da questa rappresentazione in un'unica stringa, le informazioni che interessano: ad esempio sapere che un indirizzo è in Italia oppure che un indirizzo è relativo a Milano; quindi dovrà interpretare la stringa a ricevere questi dati. Quindi, spesso si va a cercare un giusto bilanciamento, che è una rappresentazione che può essere, chiamiamola 'comoda', è una rappresentazione che ha più elementi, identificando gli elementi. Ovviamente il bilanciamento deve essere fatto in modo che questa rappresentazione non fissi una struttura che non vada bene in alcune situazioni, quindi è una



rappresentazione sufficientemente flessibile per poter rappresentare tutti i possibili indirizzi in un contesto, ad esempio internazionale. Ad esempio, nella struttura, noi possiamo avere un ordine diverso degli elementi e questo ordine può essere un elemento posto in un'altra posizione, ma anche può essere una struttura gerarchica; per esempio il nostro stesso indirizzo potrebbe essere rappresentato in questo modo: prendo Italia, poi considero come un'unica informazione la località e quindi Milano e il suo cap vengono viste gerarchicamente, quindi, come un unico campo strutturato a sua volta; la stessa cosa posso farla per l'indirizzo: posso dividere quello che è la via, quindi qua abbiamo la piazza Leonardo da Vinci e il numero civico; quindi, posso avere, in questo caso, come struttura in realtà, tre diversi campi: il Paese, la località e l'indirizzo; a loro volta alcuni di questi strutturati. Ovviamente, a partire da questa discussione possiamo immaginarci tante situazioni che si possono presentare, anche con alternative che possono avere caratteristiche locali. Ultimo aspetto la presentazione: quando discutiamo in un aspetto, è un aspetto che è critico nell'interazione fra sistemi, che è la possibilità di rappresentare la stessa informazione con stringhe di lunghezza variabile. Questo è rappresentato qua graficamente, ma supponiamo di avere un indirizzo, qua, che mi consenta di scrivere indirizzi della lunghezza, ad esempio, di 50 caratteri; mentre invece supponiamo di avere qui l'indirizzo con un massimo di 30 caratteri. Ovviamente, per il nostro esempio questo non fa molta differenza, purché io abbia definito dove mettere gli spazi; quindi, nella presentazione abbiamo anche la formattazione, diciamo, di questa stringa, ma, sostanzialmente, abbiamo lo stesso contenuto. È chiaro, però, che quello che può succedere è che alcuni indirizzi siano molto lunghi: ad esempio, presentando le caratteristiche dell'indirizzo di questa parte di indirizzo, indicando anche quello che potrebbe essere il nome di un quartiere, che a sua volta può essere un nome lungo e così via... E potrebbero esserci indirizzi che superano i 30 caratteri. Cosa succede nello scambio d'informazioni fra sistemi? Se io passo da un indirizzo da 50 caratteri a 30, quello che succederà è che taglierò queste informazioni, quindi avrò una stringa diversa rispetto all'inizio; se io passo da un indirizzo 30 caratteri a 50, ovviamente potrò aggiungere degli spazi bianchi, ma il problema è se l'indirizzo da 30 era sufficiente a rappresentare tutto quello che si vuole avere rappresentato all'indirizzo che ha più caratteri. Quindi, noi vediamo qui, - ovviamente ci sono anche tanti altri problemi, relativi alla codifica di carattere, relativa appunto aspetti strutturali, eccetera che qui abbiamo solo accennati - cosa vediamo qua? Vediamo qui un problema che è comune nei sistemi informativi: il fatto che i dati possano essere rappresentati in tanti modi diversi e con strutture diverse e presentati in modo diverso; e questo è un problema che nell'interazione fra sistemi può essere molto significativo. Un esempio: le operazioni di pulizia, cleaning dei dati; il termine che viene usato tecnicamente all'interno dei sistemi informativi, che riguardano, contengono anagrafi di clienti, sono operazioni che possono essere molto più pesanti, nel senso che dovremo decidere, per ogni problema che si può presentare, nell'avere indirizzi di tipo diverso per lo stesso cliente: se sono un problema di rappresentazione, come convertirli, se è una struttura diversa, come passare da una struttura all'altra; se abbiamo problemi di presentazione, come adottare un sistema di presentazione che consente di rappresentare correttamente tutti i nostri dati. Alla fine, quello che vogliamo avere, se un cliente ha un solo indirizzo, è questo indirizzo, descritto in modo univoco all'interno del nostro sistema. Ovviamente, un altro aspetto che si presenta, è il fatto che ciascun sistema, sottosistema - abbiamo parlato anche di interazioni fra diverse organizzazioni - gestisce i propri dati e quindi quello che poi vedremo, avremo come soluzione tecnologica, dovremo fare un'operazione che sarà soprattutto di adattamento; quindi parleremo di adattatori. Non possiamo pensare di cambiare il formato dei dati all'interno di ogni sistema per adattarlo a un formato unico, ma dobbiamo fare in modo che i nostri sistemi possano parlarsi, senza introdurre errori, quindi ad esempio, un errore che si potrebbe creare, in questo caso, è avere due indirizzi per lo stesso cliente, quando in realtà i due indirizzi sono logicamente, concettualmente uguali. Ovviamente, problemi di presentazione poi si presentano non solo per le stringhe, ma anche per tanti altri problemi; ad esempio potete provare a pensare alcuni problemi di presentazione; un altro tipico sono le date: ci sono tanti formati possibili per le date e, ad esempio se io scrivo 6/1/2018, a seconda del Paese in cui mi trovo, questo potrà essere il 6 gennaio oppure il primo di giugno. Ovviamente questo è un problema che anche deve essere risolto comprendendo qual è il formato di presentazione che viene utilizzato e, eventualmente,



facendo le conversioni per scambiarsi dati quando si utilizzano formati diversi e sistemi diversi. Il problema del formato dei dati non è l'unico problema che si può presentare, dicevamo. Un altro problema è il significato dei dati, si parla di semantica dei dati. Ad esempio, prendiamo qui un esempio: troviamo quattro campi, P10, 10, 100, offerta; questa potrebbe essere la presentazione di un ordine, in cui si ordina il prodotto 10, quantità 10, il prezzo 100 e lo stato di questo ordine può essere lo stato dell'offerta di un prodotto. Allora, quale può essere il problema che abbiamo in questo caso, di comprensione? Abbiamo detto questi sono strutturati, possiamo ovviamente utilizzare le intestazioni che vengono date, per esempio una forma tabellare per indicare il significato dei vari campi, quindi ad un identificatore il prodotto, quantità, prezzo, note come indicato nella figura. Però il problema che si presenta di significato, è ad esempio nel prezzo: nel prezzo, quando noi scriviamo 100, ovviamente, sottointendiamo la valuta che stiamo utilizzando; sarò in Italia, supponiamo di avere un prezzo in euro, questo prezzo però nel caso di un ordine risolto a livello internazionale potrebbe avere quotazioni in valute diverse: non solo l'euro, ma anche dollari, sterline così via. Quindi, il significato di questo dato è comprensibile solo quando abbiamo effettivamente associato questo prezzo anche alla valuta di cui si parla. Ma questo non è l'unico problema che si presenta nell'interazione fra sistemi; nel momento in cui noi abbiamo definito una valuta specifica per questo ordine, il problema nell'interazione da risolvere è anche di quando e come verranno effettuate le conversioni fra valute; ad esempio, se di cambio sono variabili, per la maggior parte delle valute, questo vorrà dire che un prezzo 100 euro varrà prezzo diverso in dollari a seconda del giorno in cui avviene la quotazione. Quindi, nella integrazione di queste informazioni, sarà importante definire non solo qual è il significato di questo dato, ma anche quali potranno essere le modalità per scambiare questa informazione tra il mio sistema che sta parlando in euro e un altro sistema che sta parlando in dollari, in cui tutte le quotazioni vengono memorizzate in dollari. Dovranno essere definite alcune regole per convertire in modo corretto quella che è l'informazione che viene scambiata. Abbiamo visto quindi, in questo modulo, alcune tematiche che riguardano lo scambio di informazioni fra sistemi quando noi vogliamo guardare gli aspetti di integrazione dei dati. Come abbiamo visto, il nostro problema non è integrare i dati nel senso di metterli tutti all'interno di uno stesso sistema, ma fare in modo che le informazioni che vengono memorizzate in sistemi diversi possano essere scambiate mantenendo quello che è il loro significato originario e questo si declina in varie tematiche, che devono essere affrontate da un sistema che supporta lo scambio di informazioni fra sistemi. Dobbiamo definire vari aspetti di rappresentazione di dati, di struttura, di presentazione e dobbiamo anche definire la semantica dei dati e, insieme alla semantica, anche tutte quelle regole che consentono la modifica di un dato, che viene rappresentato secondo un sistema - abbiamo parlato delle valute ma questo si rappresenta anche in altri casi, ad esempio quando parliamo di dimensioni, di pesi, in generale quando parliamo di unità di misura - dobbiamo definire delle regole per consentire che il colloquio fra i vari sistemi, lo scambio d'informazioni avvenga in modo corretto.

