

# Master Data Management e migrazione dei dati

Questo documento contiene informazioni riservate, proprietarie e coperte da segreto commerciale ("Informazioni confidenziali") di Informatica Corporation che non possono essere copiate, distribuite, duplicate o altrimenti riprodotte in alcun modo senza previa autorizzazione scritta di Informatica.

Nonostante siano state adottate tutte le precauzioni necessarie per garantire la precisione e la completezza delle informazioni contenute nel presente documento, non si può escludere la presenza di errori tipografici o imprecisioni tecniche. Informatica declina qualsiasi responsabilità in relazione a perdite di qualsiasi natura derivanti dall'utilizzo delle informazioni contenute in questo documento. Le informazioni contenute in questo documento sono soggette a modifica senza preavviso.

L'integrazione delle caratteristiche di prodotto descritte nel presente materiale in qualsiasi versione o upgrade dei prodotti software Informatica (oltre ai tempi di rilascio di tali versioni o upgrade) è interamente a discrezione di Informatica.

Protetto da uno o più dei seguenti brevetti USA: 6,032,158; 5,794,246; 6,014,670; 6,339,775; 6,044,374; 6,208,990; 6,208,990; 6,850,947; 6,895,471; o dai seguenti brevetti USA in attesa di registrazione: 09/644,280; 10/966,046; 10/727,700.

Edizione pubblicata a novembre 2014

## Indice

|  |          |
|--|----------|
| <b>Perché l'MDM è così importante per la migrazione dei dati. . . . .</b>            | <b>2</b> |
| <b>Problema n.1: diversità di esigenze . . . . .</b>                                 | <b>2</b> |
| <b>Lancio tempestivo: il passaggio intermedio . . . . .</b>                          | <b>3</b> |
| <b>Caso studio: milioni e milioni di prodotti. . . . .</b>                           | <b>4</b> |
| <b>Problema n.2: la qualità è importante . . . . .</b>                               | <b>4</b> |
| <b>Caso studio: riunire i sistemi della sede principale e quelli locali. . . . .</b> | <b>5</b> |
| <b>La migrazione dei dati come percorso per un MDM di alto livello . . . . .</b>     | <b>5</b> |

In questo documento viene evidenziata la crescente importanza del Master Data Management nei progetti di migrazione dei dati. Si approfondisce la sua utilità e le migliori opzioni di implementazione con alcuni casi di studio significativi.

## Perché l'MDM è così importante per la migrazione dei dati

Ogni nuovo sistema richiede dati per stimolare l'avviamento di un'attività. La maggior parte dei nuovi sistemi necessita di moltissimi dati. Oggi, migriamo i dati da numerose fonti legacy verso nuovi sistemi, e arricchiamo nuovi sistemi da fonti esterne come i file degli indirizzi degli uffici postali. Il kit di tool per la migrazione dei dati di Informatica implementa una suite di tecnologie e best practice per risolvere le numerose problematiche collegate agli attuali scenari di migrazione dei dati.

Innanzitutto, esaminiamo lo scenario. Anche se il sistema target può offrire molte soluzioni, è molto probabile che le fonti dati siano soluzioni deludenti, poiché ciascuna è stata progettata tenendo in considerazione un processo di business con una visione globale diversa. Sebbene il sistema target richieda coerenza, l'ambiente legacy spesso stenta a garantirla nelle strutture e nei contenuti. Si vengono a creare due problemi; in entrambi i casi, la soluzione è l'utilizzo di tecniche di MDM.

### Problema n. 1: diversità di esigenze

Esempio: stiamo installando una nuova applicazione per pianificare la produzione. Promette di migliorare la gestione, partendo dal livello più basso del business, e di aumentare l'efficacia creando un collegamento tra le applicazioni per la gestione delle risorse umane e la contabilità. Tuttavia, ci rendiamo conto che tutti i dipartimenti e i sistemi su cui si basano tali dipartimenti adottano un approccio diverso sulle stesse identiche cose. Agli occhi di un contabile, il mondo è fatto di centri di costo, centri di profitti, ammortamento, capital asset e asset operativi. Un production engineer, invece, descriverebbe lo stesso spazio fisico parlando di processi automatizzati, semi-automatizzati e manuali, workflow, programmi di manutenzione e quote di produzione. Il dipartimento HR vede, nella stessa scena, solo staff interno, staff esterno, requisiti di formazione, livelli di competenze, contratti e problemi di salute.

Tutti osservano la stessa realtà, ma la vedono in modo diverso.

Di conseguenza, non bisogna sorprendersi se quando scelgono e disegnano i loro sistemi, li modellano in modo diverso. Le esigenze sono tutte ugualmente importanti, ma il problema è che non sono comuni a tutti. E così, quando dobbiamo eseguire la migrazione dei dati ci rendiamo davvero conto che non è possibile accomunare le esigenze, perché sono troppo diverse. La soluzione è adottare un approccio universale,

perché non possiamo scartare nessun punto di vista. Non esiste una verità assoluta, così come nessun aspetto è sbagliato a priori. D'altro canto, anche se il punto di vista del dipartimento di produzione fosse il più giusto (considerato che nel nostro caso è quello che ci spinge a cambiare), non potremmo costringere gli altri data store ad adattarsi a questo modello, senza riprogettarlo su una scala tanto ampia quanto la migrazione.

In ogni caso, è più probabile convincersi che ogni data store legacy sia appropriato perché elabora un suo dominio. Di conseguenza, il dipartimento di produzione ha ragione dal suo punto di vista, così come il dipartimento HR e via dicendo. Pertanto, abbiamo bisogno di un modello che possa soddisfare tutti i punti di vista. È interessante notare che questa difficoltà è la stessa che si riscontra nel progetto per produrre il sistema target. Perché non limitarsi ad attendere che il target sia pronto ed eseguire un'analisi dei punti deboli?

## Lancio tempestivo: il passaggio intermedio

Se il primo problema è quello di ricostruire un panorama di dati frammentato, facciamo velocemente un passo indietro e riflettiamo sulle tempistiche richieste da una normale migrazione.

Partendo dal presupposto che la progettazione e l'installazione del nuovo sistema richiedono un anno, significa che il sistema target sarà disponibile solo dopo almeno otto mesi e, oltretutto, inizialmente sarà in forma provvisoria. Inoltre, questo richiede una metodologia di progetto standard con attività di start-up, analisi dello stato corrente, configurazione di un nuovo sistema e riprogettazione dei processi di business. In base alla nostra esperienza, il target sarà disponibile solo dopo dieci mesi, e anche allora saranno richieste modifiche sostanziali.

Potremmo, quindi, avere solo due mesi per eseguire la gap analysis, occuparci di estrazioni, trasformazioni e caricamenti di dati, scrivere e provare i load script e così via. Purtroppo il tempo non è sufficiente per eseguire un'analisi approfondita delle differenze strutturali tra i data store legacy, tanto meno per produrre i dati composti di cui abbiamo bisogno nei file di load.

Un approccio per risolvere il problema legato al tempo sarebbe la creazione di un modello intermedio, o prototipo di migrazione. Si registrano le differenze tra i data store legacy e il prototipo e si inizia l'attività di bonifica e preparazione dei dati. Quando il target sarà finalmente pronto, esamineremo le differenze con il prototipo. Conoscendo già le differenze tra i data store legacy e il prototipo, possiamo isolare le differenze, semplificando la nostra attività nei momenti di maggior pressione e attività più intensa.

Il trucco è realizzare un modello di migrazione il più realistico possibile, analizzare le differenze tra la migrazione legacy e il modello proposto e progettare trasformazioni, arricchimento dei dati, miglioramenti e così via. Infine, quando il target prende forma, possiamo perfezionare le trasformazioni, affrontando i problemi meno evidenti che possono presentarsi alla fine. In questo modo, se il nostro prototipo si avvicina al target solo per l'80% (e l'esperienza ci insegna che non basta), avremo comunque progettato l'80% della logica di trasformazione prima delle impegnative attività delle ultime settimane.

Ma perché una soluzione di Master Data Management dovrebbe aiutarci in questo senso?

Secondo la definizione di Wikipedia:

**"L'obiettivo dell'MDM è fornire i processi per raccogliere, aggregare, abbinare, consolidare, garantire la qualità, portare avanti e distribuire i dati attraverso un'organizzazione, per assicurare coerenza e controllo negli interventi di manutenzione e nell'uso delle informazioni all'interno delle applicazioni"**

Dalla prospettiva dei dati, i tre punti di vista analizzati nel nostro esempio (contabile, production engineer e dipartimento HR) verranno dimostrati nei valori codificati nei data store di pertinenza. Siamo, quindi, molto interessati a rendere disponibile una procedura per "raccogliere, aggregare, abbinare, consolidare e garantire la qualità". Ai fini del nostro progetto, non siamo direttamente preoccupati di "portare avanti e distribuire i dati attraverso un'organizzazione" per i motivi a cui abbiamo accennato. Tuttavia, garantire "coerenza e controllo negli interventi di manutenzione e nell'uso delle informazioni all'interno delle applicazioni" è uno degli obiettivi principali del nostro progetto di migrazione. In altre parole, adottiamo un approccio MDM nei confronti della migrazione di dati, anche se probabilmente non chiuderemo il cerchio attorno ai sistemi che utilizzano tali dati, perché il nostro lavoro è sostituire i sistemi, non migliorarli.

## Caso studio: milioni e milioni di prodotti

Ora esaminiamo un esempio reale per capire meglio.

La migrazione dei dati che vogliamo esaminare ha avuto luogo in un'importante società di telecomunicazioni. Tanto per farci un'idea, parliamo di un'azienda con decine di milioni di clienti e centinaia di milioni di installazioni. Anche il numero dei prodotti è elevato: fino a un centinaio di migliaia di singoli prodotti. E, a rendere il tutto ancora più interessante, il settore delle telecomunicazioni sforna prodotti su prodotti, quindi abbiamo dovuto tenere in conto anche la loro sempre crescente diffusione.

Quello delle società di telecomunicazioni è un ambito difficile anche a causa della quantità di sistemi legacy in uso. Ordini, progettazione, fornitura e fatturazione sono compiti che implicano diversi livelli di attività fisiche e logiche, quindi richiedono la costruzione di soluzioni nuove per ogni fase. Di conseguenza, in genere ci occupiamo di circa 30 sistemi legacy per installazione su un potenziale gruppo di oltre 400.

Certamente ognuno di questi sistemi aveva una sua personale vista in base alla posizione nel flusso del processo e tale vista era logica, fisica o finanziaria.

In questo documento, prenderemo in esame esclusivamente la struttura dei prodotti e cercheremo di capire perché una soluzione basata sull'MDM si è rivelata necessaria.

La presenza di centinaia di migliaia di prodotti, alcuni dei quali formati da altri prodotti, implicava una quantità molto elevata di regole. Dove un'applicazione legacy richiedeva l'installazione di un singolo telefono, un altro sistema richiedeva un ricevitore, un segnale di chiamata, chiamate in uscita, esclusione di chiamate in uscita, possibilità di richiamare, segreteria telefonica e così via. In poche parole, per un telefono di casa servivano molti prodotti, senza parlare delle complesse reti di dati e traffico vocale. Sembrava che ogni sistema nel processo di provisioning avesse una propria concezione del telefono. Avevamo bisogno di coerenza e dovevamo iniziare a garantirla il prima possibile. Non potevamo aspettare che il target fosse pronto.

La soluzione era creare un Master Data Management hub per prodotti e build di prodotti. I requisiti del settore, in rapida evoluzione, sono stati soddisfatti attraverso un aggiornamento settimanale, che ha richiesto la progettazione e l'aggiunta di nuovi prodotti. L'hub ci ha consentito di verificare la presenza di prodotti corrispondenti nei data store legacy e di continuare a notare le differenze tra le rappresentazioni nei diversi punti del processo di distribuzione. Man mano che si verificava una fase della migrazione, apprendevamo le differenze tra le strutture del sistema sorgente e del sistema target e potevamo ricodificare di conseguenza processi di mapping e trasformazioni.

## Problema n.2: la qualità è importante

Ora esaminiamo il secondo tipo di problema: il valore dei dati o dei contenuti. Così come i data store legacy possono avere più strutture per lo stesso oggetto di business, anche noi possiamo avere più valori. Osserviamo alcuni esempi frequenti.

Uno dei problemi più fastidiosi per tutti i dipartimenti marketing sono i duplicati negli elenchi di clienti raccolti da numerosi store di dati che compongono il panorama. Alcuni sono facili da identificare ed eliminare, altri si scontrano con il temuto problema di omonimi o sinonimi. Mario Rossi sarà la stessa persona indicata altrove come M Rossi o M G Rossi? Quando eseguiamo una migrazione, tutti noi riscontriamo gli stessi problemi, che con ogni probabilità si verificano nello store di dati dove possono essere presenti più copie della stessa persona.

Spesso questi problemi sono causati da una combinazione di problemi strutturali. Ad esempio, in ambienti business-to-business (B2B), spesso si fa fatica a inquadrare il cliente. Il cliente è la persona giuridica principale (cioè l'ente da citare in giudizio in caso di necessità)? Oppure è l'ente commerciale con cui siamo in affari?

È un magazzino o un negozio? Ancora una volta, il punto di vista cambia a seconda del ruolo in azienda. Il team di responsabili della logistica avrà bisogno di sapere dove trovare le scorte, mentre il team che si occupa della fatturazione vorrà conoscere anche l'indirizzo dove inviare la fattura. I team di vendita potrebbero operare a seconda dell'area geografica, ma con la collaborazione di manager che vantano relazioni strategiche e che si occupano dei clienti più grandi. Quindi, a livello legale, potrebbe esistere una sola persona fisica, ma centinaia di sedi per l'ordine, la consegna e la fatturazione.

Ancora una volta, potrebbe essere impossibile correggere queste anomalie nei sistemi sorgente, soprattutto se i sistemi sorgente non sono sbagliati (è naturale che ci siano duplicati causati da errori nei processi che hanno creato due volte lo stesso utente e sono errati anche in termini di data store legacy e possono essere corretti al loro interno).

Da dove possiamo iniziare allora? Quale aiuto potremmo ricevere da una soluzione MDM? È evidente che l'MDM è la soluzione perfetta per gestire le entità principali. Qui ci occuperemo di quasi tutte le funzioni di un MDM completo, cercando allo stesso tempo di garantire "coerenza e controllo negli interventi di manutenzione e nell'uso delle informazioni all'interno delle applicazioni", almeno per quanto riguarda i duplicati che non dovrebbero essere presenti nel nostro set di dati legacy.

## Caso studio: riunire i sistemi della sede principale e quelli locali

Un altro esempio tratto da un'esperienza reale ci aiuterà a capire meglio il problema. Una società bancaria di medie dimensioni doveva effettuare una migrazione. Le sue sedi erano distribuite in varie aree geografiche e, al loro interno, c'erano diversi tipi di sistemi aziendali, a loro volta collegati alla sede principale con una serie di processi in batch. Ci si presentavano entrambi i tipi di problemi: c'erano varianti locali nelle interpretazioni strutturali e duplicati sia tra un ramo e l'altro sia al loro interno. Grazie all'aiuto della tecnologia MDM, abbiamo potuto raggiungere un'unica versione della verità, ricollegandola ai relativi sistemi sorgente. L'obiettivo è riunire i dati collegati in un'unica struttura, eliminando i duplicati tra una filiale e l'altra e al loro interno in tempo reale. In questo caso, l'MDM non si è limitato a questo, ma ha anche offerto un'unica vista della verità al termine della migrazione.

## La migrazione dei dati come percorso per un MDM di alto livello

La migrazione dei dati è un efficace presupposto per beneficiare dei vantaggi dell'MDM. Durante la preparazione dei dati per la migrazione, dovremo eseguire la deduplica e risolvere problemi semantici legati a differenze strutturali. In passato, spesso abbiamo creato applicazioni inefficaci che non hanno fornito un'unica versione della verità lungo il percorso del progetto. Se utilizziamo una soluzione MDM per un utilizzo industriale, portare avanti il progetto per realizzare un MDM e integrarlo perfettamente nell'architettura è ancora più complicato. Abbiamo già mostrato il valore della tecnologia MDM, e i problemi semantici dell'adozione dell'MDM.

## Informazioni su Informatica

Informatica Corporation (Nasdaq:INFA) è il principale fornitore indipendente di software per l'integrazione dei dati. Aziende in tutto il mondo si affidano a Informatica per valorizzare il potenziale racchiuso nelle informazioni - information potential - ottenendo un evidente vantaggio competitivo. Informatica Vibe, la prima e unica virtual data machine (VDM) integrabile, offre l'innovativa ed esclusiva funzionalità "Map Once. Deploy Anywhere" di Informatica Platform. Oltre 5,500 aziende a livello mondiale utilizzano Informatica per sfruttare appieno il loro patrimonio informativo, da dispositivi verso mobile, social e Big Data, siano essi on premise, nel Cloud o sui social network. Per ulteriori informazioni, visitare [www.informatica/it](http://www.informatica/it).



Piazza della Repubblica 14/16 - 20124 Milano, Italia. Tel: +39 02 37 05 80 00 Via Luca Gaurico 9/11 - 00143 Roma, Italia.  
Tel: +39 06 54 83 21 34. [informatica.com/it](http://informatica.com/it) [linkedin.com/company/informatica](https://www.linkedin.com/company/informatica) [twitter.com/InformaticaCorp](https://twitter.com/InformaticaCorp)

© 2014 Informatica Corporation. Tutti i diritti riservati. Informatica® e Put potential to work™ sono marchi o marchi registrati di Informatica Corporation negli Stati Uniti e in giurisdizioni a livello mondiale. Tutti gli altri nomi di aziende e di prodotti possono essere nomi commerciali o marchi.