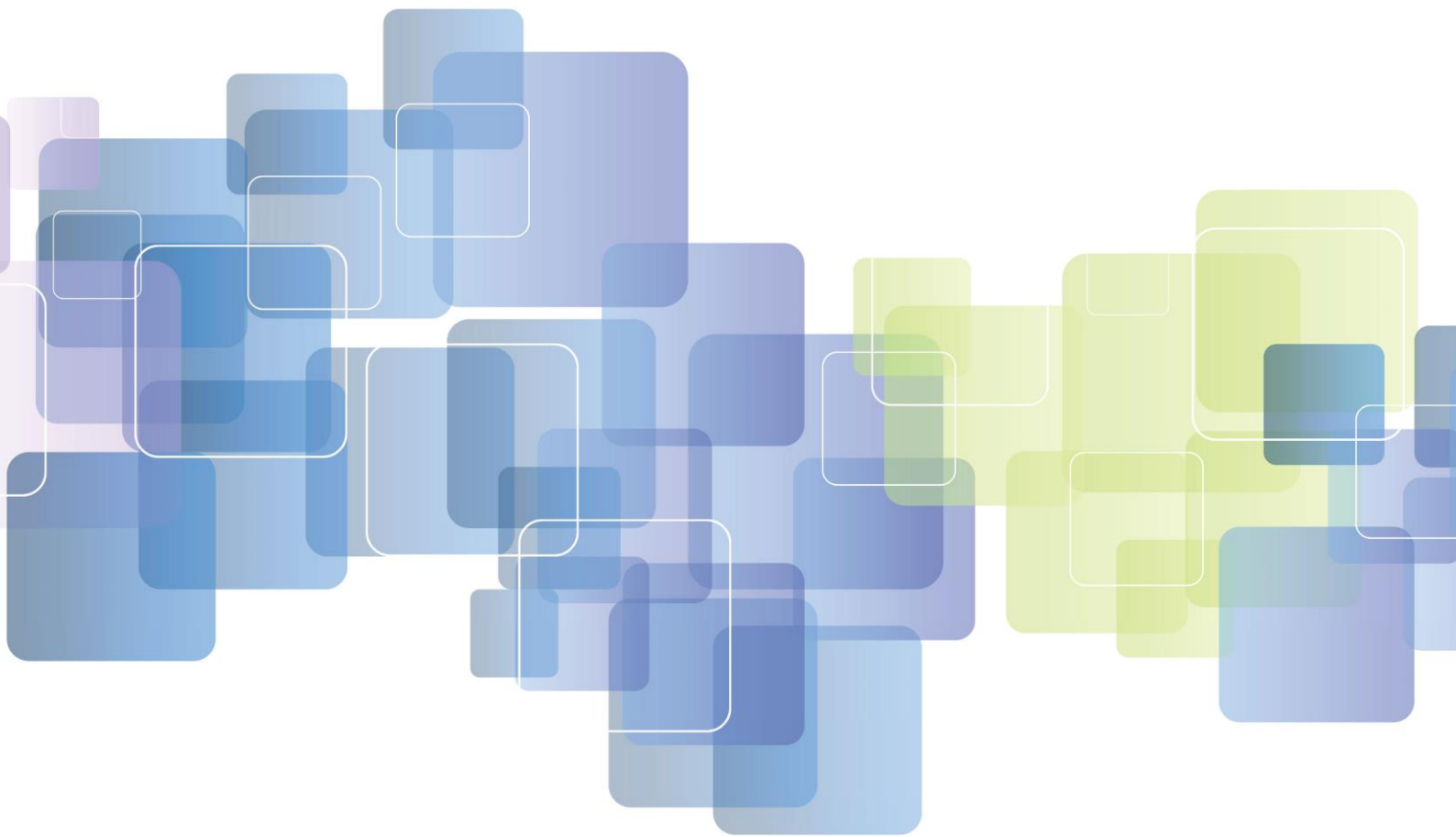


DBA, una razza in via di estinzione?

Renato Comes



Una tragedia vera in pochi atti!

Chi è un DBA ? Un Data Base Administrator, una figura professionale del mondo informatico nata negli anni 70 quando apparvero sul mercato i primi DBMS, Data Base Management System, il software per gestire le Basi di Dati, evoluzione concettuale, logica e fisica dei sistemi generalizzati di memorizzazione dei dati (i file). Per chi non lo ricordasse, il modello concettuale di una Base Dati è la rappresentazione delle Classi o Entità di Business (p.es. Ordini, Certificati, Cartelle Cliniche ...) e delle relazioni intercorrenti fra loro a prescindere dal modello di DBMS. Il modello Logico è la trasposizione del modello concettuale nel modello di DBMS prescelto (Relazionale, Object Relational, Reticolare etc.). Il modello fisico è la successiva trasposizione che tiene conto della tecnologia specifica prescelta. I tre modelli sono legati fra loro con una gerarchia che vede il modello concettuale al più alto livello seguito da quello logico che poggia su quello fisico. La descrizione dei modelli e dei loro contenuti è memorizzata un meta modello erede del tradizionale dizionario (o catalogo) dei dati. Il Data Base quindi non è quindi solo un insieme di tabelle (file) con un insieme di campi: nel tempo ha acquisito una serie di significati semantici e logici di cui molti hanno oggi perso la percezione, come se la capacità di mettere in relazione due entità e di esprimerne nel meta modello i relativi vincoli semantici e d'integrità sia la cosa più naturale del mondo. Era naturale che accadesse questa cosa: oggi si danno un mucchio di cose per scontate così come si è portati a sottovalutare la specificità del mestiere di Data Base Administrator, un ruolo che esprime in sé due professionalità: la prima è di Data Analyst, cioè capire e tradurre la realtà in un modello concettuale dei dati e poi seguire passo passo l'evoluzione di questo modello fino al fisico ed occuparsi, seconda professione, dell'ottimizzazione fisica del modello dei dati a seconda del DBMS utilizzato.

Atto Primo, i bei tempi andati

Negli anni 70 e fino agli anni 90, la figura del DBA è stata sempre ben delineata: una professionalità che prestava il proprio supporto professionali prima ai gruppi di sviluppo nelle fasi di analisi e poi nella soluzione dei problemi di ottimizzazione delle varie basi dati. In questo contesto professionale il DBA è quindi un "consulente" ed è di staff allo sviluppo.

La gestione dei dati e dei metadati è quindi sotto controllo, non si muove foglia senza che il DBA non abbia partecipato e validato un modello dati.

La duplicazione dei dati, la cattiva gestione dei metadati, la non documentazione degli modelli concettuali, logico e fisico sono quindi scongiurati.

Il DBA è quindi il Grande Sacerdote che veglia sulla risorsa dati aziendale e quindi ha titolo per interfacciarsi con eventuali Auditor e controllori della qualità dei dati (privacy e correttezza in primis).

Lo stipendio di un DBA, in azienda è una risorsa pregiata, è superiore a quello di un Capo Progetto.

Atto secondo, il declino

Negli anni 80 e specialmente negli anni 90 avvengono alcuni accadimenti che porteranno al declino professionale del DBA ed ad un suo ridimensionamento operativo, fino alla chiusura dell'ufficio DBA:

Evento	Effetto
I sistemi informativi passano da una struttura centralizzata (mainframe centrica) ad una distribuita in modo sempre più caotico	Più sistemi, sparsi, quindi i DBA perdono la completa conoscenza dei dati. Come possono reggere poche persone su centinaia di database ? Non reggono ed infatti in qualche azienda vengono concentrati solo sulla parte mainframe (per quei pochi che l'hanno ancora), dove resiste il modello centralizzato.
Sempre più package con DBMS e schemi "nascosti" solo da interfacciare e non integrare	I package sono documentati e le loro basi dati non necessitano di manutenzione diretta del metamodello ma di conoscenza delle interfacce, cosa prettamente applicativa in quanto l'enfasi si sposta dal modello dati al processo (cosa fare rispetto al dove)
La proliferazione di DBMS di ogni tipo con la disgregazione del modello Data Base Integrato	La proliferazione dei tipi di DBMS e la loro crescente complessità comporta la specializzazione della componente tuning (cioè la parte fisica) rispetto a quella logica e concettuale. I DBA diventano dei sistemisti dei DBMS e quindi vanno nelle Operation.
Proliferazione di progetti di taglia minore e specifici per i vari utenti che mettono alla frusta l'ambiente di sviluppo	Le nuove architetture web facilitano la produzione di applicazione di media e piccola taglia che richiedono gruppi di sviluppo sempre più piccoli e specializzati: questi vengono quotidianamente messi alla frusta dagli Utenti che hanno il budget e sono sempre più capricciosi sui requisiti. I DBA, da sempre pochi ed abituati a lavorare con metodo, diventano un impiccio per chi

	non ha tempo di pensarci su troppo.
Il Data Warehouse	<p>Data Warehouse e loro proliferazione significano duplicazione dei dati e moltiplicazione delle Basi di Dati senza contare il disegno della parte ETL di competenza funzionale.</p> <p>I DBA, usi ai dati strutturati e normalizzati, non servono molto, neanche a disegnare le procedure ETL, evidentemente. Magari servono coloro che sono esperti nelle problematiche di tuning, ovvero a risolvere i problemi in produzione ahi loro, invece che a prevenirli, come avveniva in passato</p>
Nuove metodologie di sviluppo (RAD, RUP, Agili) che privilegiano la formalizzazione del ciclo di vita dei requisiti e, in cascata la formalizzazione in schemi concettuali di processi e anche di dati	<p>Queste metodologie esaltano il ruolo degli analisti di processo e di dominio ed il dialogo con l'utente nella gestione dei requisiti (specie i requisiti di business) ed il disegno del database è una delle attività che segue a ruota, un di cui. I vari diagrammi delle attività e delle sequenze operative fino a quelli di stato e di collaborazione sono al centro dello sforzo progettuale. In questo contesto, è molto più importante un analista che un DBA avulso dai requisiti</p>
La terziarizzazione dello sviluppo e/o della gestione	<p>L'Application Management (AM) comporta l'outsourcing dello sviluppo e della gestione applicativa e quindi la fase di analisi e disegno della base dati sono affidate a persone diverse e di aziende diverse, altro che DBA ! Se si pensa poi al cloud computing in cui anche fisicamente le applicazioni sono in spazi diversi lo sfilacciamento è totale e quindi un DBA aziendale è inutile.</p>

Il primo decennio del nuovo secolo vede quindi consumarsi il dramma di un'informatica che a parole è sempre più sensibile ai requisiti di business degli utenti e dall'altra,

complice la crisi economica e la gestione dei budget in mano agli utenti, taglia molte funzioni di controllo e di qualità, fra cui i DBA.

L'ufficio DBA sparisce e i DBA vanno ad occuparsi di ottimizzazione degli strumenti DBMS, ovvero solo la seconda parte del loro mestiere: le attività di Data Analysis sono assorbite dai Business Analyst e Analyst, dagli Application Designer e dagli esperti di architetture, spesso senza confini certi.

Il risultato è il caos attuale, sia sul parco applicativo, sia sui dati. Un caos che cresce al crescere delle dimensioni delle aziende, del numero dei poli di sviluppo e del numero dei fornitori di AM e dei package acquisiti. E non c'è SOA che tiene, anzi. C'è qualcuno che ne è seriamente preoccupato? Pochi.

Atto terzo, prendere atto della situazione

Nel panorama appena delineato, i gruppi di sviluppo (interni ed esterni) lavorano a testa bassa per star dietro ai requisiti funzionali degli utenti, i momenti di raccordo sono davvero pochi e riguardano i requisiti non funzionali, ovvero quelli non legati alle funzionalità applicative. Gli enti IT più avveduti cercano di mantenere un presidio proprio su questi requisiti trasversali così ben definiti dalla normativa ISO 9126: l'usabilità, l'accessibilità, la chiarezza architeturale e la sicurezza, tanto per esemplificare.

La conoscenza semantica dei dati è sparsa nei vari progetti così come il controllo dei metadati, l'integrazione delle basi di dati e la non ridondanza sono demandati ai singoli progetti e, dato che tutti lavorano a testa bassa sul proprio progetto, non c'è molta governance.

Possiamo pensare che esistano principalmente due modelli organizzativi per quelli che riguarda la gestione del parco dati, per chi se ne preoccupa ancora:

- un modello centralizzato, spesso valido solo per chi ha ancora dati sul mainframe con uno o due DBA che li presidiano e svolgono ancora l'antico ruolo di consulente dei progetti in essere, progetti spesso di sola manutenzione poiché le applicazioni web e data warehouse sono oramai altrove
- un modello che nella migliore delle ipotesi potrebbe essere federato, ove i vari capi progetto costituiscono un direttorio virtuale per promuovere il riutilizzo della conoscenza sui metadati. Il direttorio non è un organismo nato per queste cose (di solito ci riunisce per fare il punto della situazione sui progetti), ma da cosa nasce cosa e spesso nel confronto nascono idee e suggerimenti che vengono poi sviluppati a livello di tecnici

Il rischio del caos e quindi, come lo chiamo io, il rischio di logorio da dati moderni (poco controllati ed integrati) è veramente alto ed è un'altra delle contraddizioni di una IT "moderna" che vorrebbe riutilizzare, pubblicare, dare supporto al business e magari fare carriera in azienda come motore d'innovazione.

Quali sono i rischi di una non gestione del patrimonio dati ?

- Duplicazione fisica (possibile) senza una coerenza semantica (male) e oltretutto non documentata (molto male)
- Ambiguità semantica come conseguenza del passo precedente
- Dati che dicono la stessa cosa ma provenienti da fonti diverse
- Propagazione incontrollata di errori ed imprecisioni nei dati
- Nessuna possibilità di fare auditing in ripetitivo e serio (non regressione)
- Dati con pedigree non controllabile, utenti che potrebbero non fidarsi e quindi decisioni di business a rischio

Naturalmente i sistemi informativi e i processi di business sono talmente complessi ed inefficienti che l'impatto del caos dei dati è un po' mitigato ma ciò non toglie il senso di precarietà regnante fra gli utenti più esperti. La soluzione sarà forse l'outsourcing ? La storia ha dimostrato che l'allungamento della catena delle responsabilità può magari ridurre i costi (nel breve e medio) ma crea altri problemi, primo fra tutti la perenne turbolenza dei rapporti professionali basata a questo punto su rigide procedure da rispettare e su SLA improbabili e non più sull'abnegazione dei singoli, cosa che risolve spesso le situazioni più incresciose dell'organizzazione aziendale.

Il caos dei dati, il logorio dei dati moderni, è quindi una realtà: che cosa si può fare ?

Atto quarto, happy end?

Al crescere del caos dei dati si è cercato di porre riparo definendo una nuova pratica: il Master Data Management (MDM), ovvero, prendendo atto che non si riesce a tenere sempre tutto sotto controllo, almeno si applichi la legge di Pareto sui Dati più importanti. MDM è quindi un processo (regole, tempi e risorse) che regoli il ciclo di vita dei Master Data, i dati fondanti di un'azienda, il basamento sul quale almeno le applicazioni strategiche fondano la propria elaborazione. Va da sé che solo una parte dei dati sono dei Master Data; se volessimo fare un esempio, potremmo dire che le classi degli oggetti (entità) su cui fonda il business di un'azienda sono dei Master Data - Prodotti, clienti, fornitori, ordini etc. Gran parte delle anagrafiche sono Master Data. I Master Data hanno requisiti speciali di qualità intrinseca (anche tracciamento, auditing etc.) poiché fondamenta del sistema informativo: non ci possono essere ambiguità di sorta o duplicazioni non gestite, altrimenti si mette a repentaglio la veridicità delle elaborazioni e l'affidabilità del servizio.

I Master Data devono avere un pedigree (metadati) ben curato con un tracciamento delle versioni, la catena delle derivazioni, le procedure di controllo, le responsabilità gestionali ben definite, ownership e stewardship etc.

Un progetto di Master Data Management deve partire dalla costruzione di Glossario business oriented che descriva puntualmente i dati con linguaggio aziendale e gli utenti finali devono sottoscrivere queste definizioni come primo passo verso la vera integrazione semantica. Dal Glossario si deriverà il Dizionario Dati che è ad uso e consumo dei gruppi di sviluppo. A parità di dato, le sua definizione nel Glossario devono essere coerente con quella del Dizionario Dati.

Chi può fare questo lavoro ? La risposta è intuibile, solo qualcuno che è al di sopra delle parti, dei progetti ed è misurato sulla qualità dei dati e dei metadati, un DBA. C'è però una grossa differenza rispetto al passato: il DBA non si mette più di traverso per validare modelli dati oramai troppi e sparsi in azienda, ma discute e gestisce l'evoluzione dei Master Data in accordo con i vari progetti che li utilizzano e riporta a fattor comune e pubblica le modifiche in modo che tutti siano al corrente dell'evoluzione dei metadati.

Un nucleo di DBA, con opportuni strumenti di design e di documentazione è in grado di integrare le fasi di raccolta dei requisiti con l'analisi e con la gestione e la standardizzazione dell'utilizzo dei Master Data reali. La gestione degli altri dati "locali" all'interno dei progetti è bene lasciarla agli analisti di progetto, tanto non si avrebbero più le risorse sufficienti.

Questa soluzione prende atto della situazione oggettiva e vale sia in un modello centralizzato, sia in un modello federato. In quest'ultimo, il nucleo dei DBA è quanto mai utile nel tirare le fila dei progetti così come può rappresentare la vera interfaccia di auditor che vorrebbero verificare la qualità dei dati critici, normalmente Master Data.

Il percorso virtuoso per recuperare la qualità passa dunque per un progetto di MDM da affidare ai gloriosi DBA; che abbiano trovato una nuova vita professionale decorosa ? Fine della tragedia (speriamo) e giù il sipario.



Tecnet Dati s.r.l.
C.so Svizzera 185 -
10149 - Torino (TO), Italia
Tel.: +39 011 7718090 Fax.: +39 011 7718092
P.I. 05793500017 C.F. 09205650154
www.tecnetdati.com

