Gli strumenti di ETL

PDF created with pdfFactory trial version www.pdffactory.com

Gli strumenti di ETL

Il ruolo degli strumenti di ETL, Extraction, Transformation e Loading, è quello di alimentare la sorgente dati singola, dettagliata, esauriente e di **alta qualità** che possa a sua volta alimentare il DW.

IN BREVE SI PARLA DI RICONCILIAZIONE DEI DATI.

Durante il processo di alimentazione del DW, la riconciliazione avviene in più occasioni:

1. quando il DW viene popolato per la prima volta,
2. periodicamente quando il DW viene aggiornato.

PDF created with pdfFactory trial version www.pdffactory.com

Gli strumenti di ETL

- La riconciliazione dei dati consta,in sostanza, di quattro fasi:

1. estrazione
2. pulitura
3. trasformazione
4. caricamento

PDF created with pdfFactory trial version www.pdffactory.com

Gli strumenti di ETL: **l' estrazione**

- Si suddivide in estrazione statica e estrazione incrementale.

1. STATICA: avviene quando il DW deve essere popolato per la prima volta e consiste concettualmente in una fotografia dei dati operazionali.
2. INCREMENTALE: viene usata per l'aggiornamento periodico del DW e contiene solo i cambiamenti avvenuti nelle sorgenti dall'ultima estrazione.

PDF created with pdfFactory trial version www.pdffactory.com

Gli strumenti di ETL: **la pulitura**

- La pulitura dei dati risulta di estrema importanza per garantire **la qualità dei dati** che popoleranno il DW.
- Infatti, normalmente le sorgenti risultano essere “sporchi”, e presentano diversi tipi di errori, tra cui:

1. Dati duplicati
2. Inconsistenza tra valori logicamente associati
3. Dati mancanti
4. Uso non previsto di un campo
5. Valori non ammissibili o errati
6. Valori inconsistenti dovuti a errori di battitura
7. Ecc.

PDF created with pdfFactory trial version www.pdffactory.com

Gli strumenti di ETL: **la trasformazione**

- La fase di trasformazione consiste nella conversione dei dati dal formato operazionale sorgente a quello del DW.
- Tra le funzionalità di trasformazione ricordiamo:

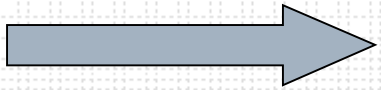
1. La conversione e normalizzazione, che operano sia a livello di formato di memorizzazione sia a livello di unità di misura al fine di uniformare i dati
2. Il matching che stabilisce corrispondenze tra campi equivalenti in sorgenti diverse
3. La selezione che riduce il numero di campi e di record rispetto alle sorgenti.

PDF created with pdfFactory trial version www.pdffactory.com

Esempio di pulitura e trasformazione di un dato anagrafico

Carlo Bianchi
P.Zza Grande 12
50126 Bologna (I)

Normalizzazione



Nome: Carlo

Cognome:Bianchi

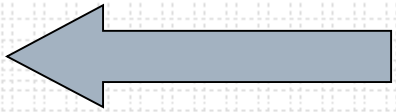
Indirizzo: P.zza Grande 12

CAP:50126

Città: Bologna

Nazione: I

Standardizzazione



Nome: Carlo

Cognome:Bianchi

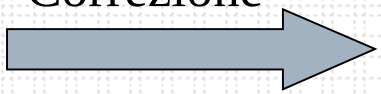
Indirizzo:Piazza Grande 12

CAP:50126

Città: Bologna

Nazione: Italia

Correzione



Nome: Carlo

Cognome:Bianchi

Indirizzo:Piazza Grande 12

CAP:40126

Città: Bologna

Nazione: Italia

PDF created with pdfFactory trial version www.pdffactory.com

Gli strumenti di ETL: **il caricamento**

- Il caricamento dei dati nel DW può avvenire secondo due modalità:

1. ***Refresh***: i dati vengono riscritti integralmente, sostituendo quelli precedenti, in genere durante il primo popolamento del DW.
2. ***Update***: solo i cambiamenti vengono aggiunti al DW senza distruggere o alterare quelli esistenti.

PDF created with pdfFactory trial version www.pdffactory.com