

Comarch Data Warehouse Manager

10.5

EINFÜHRUNG IN DAS DATA-WAREHOUSE-MANAGEMENT

Die Speicherung von Geschäftsdaten aus verschiedenen dezentralen Systemen und deren Zusammenführung an einem Ort sind häufig komplizierte Prozesse, die jedoch notwendig sind, wenn ein Unternehmen zuverlässige Analysen durchführen und wissensbasierte Entscheidungen treffen möchte. Unternehmen, die im Rahmen ihrer Geschäftstätigkeit große Datenmengen erzeugen, nutzen daher häufig zentrale Datenbanken, in denen sie die Daten aus sämtlichen Quellen vereinen können. Eine solche Datenbank, die auch als Data Warehouse bezeichnet wird, wird in regelmäßigen Zeitabständen mit Daten aus allen laufenden Systemen, Dateien und Transformationsprozessen beladen. Die Benutzer des Data Warehouse, bei denen es sich zumeist um den Unternehmensvorstand und andere Führungskräfte handelt, greifen auf diese Daten mittels unterschiedlicher Reporting- und Analysetools zu. Je nachdem, wie groß der Datenbestand ist und welchen Bereich die Daten betreffen, lassen sich bei der Durchführung von Analysen Trends, Abhängigkeiten oder Muster erkennen. Comarch Data Warehouse Manager (DWM) ist ein entsprechendes modernes Tool, mit dem sich ein Data Warehouse erstellen und verwalten lässt.

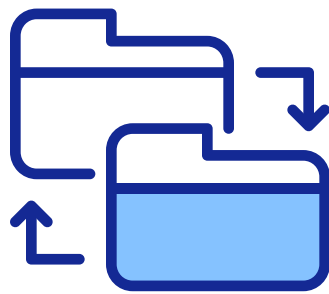
TRANSFORMATION UND LADEN VON DATEN (ETL-PROZESS)

Hinter dem ETL-Prozess (engl.: Extract Transform Load) verbirgt sich die Erfassung von Daten aus unterschiedlichen Quellsystemen sowie ihre Verarbeitung, Bereinigung und Zusammenführung; abschließend werden die Daten in die finale Datenbank geladen. Da die Mehrheit der Unternehmen von heute datenbasiert agiert, ist ein schneller Datenfluss dabei von signifikanter Bedeutung. Tatsächlich sind moderne Organisationen auf Echtzeitdaten angewiesen und möchten über Reporting-Anwendungen kontinuierlich auf aktuelle Informationen zugreifen können. Durch den Einsatz spezieller Lösungen, wie der Anwendung Comarch Data Warehouse Manager, wird es möglich, die Datenverwaltung deutlich zu vereinfachen und den Betrieb eines Data Warehouse zu optimieren.



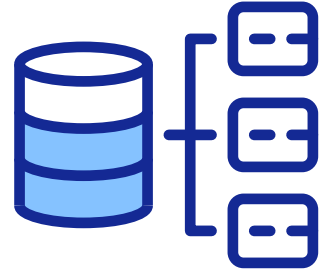
EXTRAKTION

Extraktion der Daten
aus Quellsystemen



TRANSFORMATION

Umwandlung und
Integration in das Zielmodell



LADEN

Laden der Daten in
das Data Warehouse

DIE ANWENDUNG COMARCH DATA WAREHOUSE MANAGER (DWM)

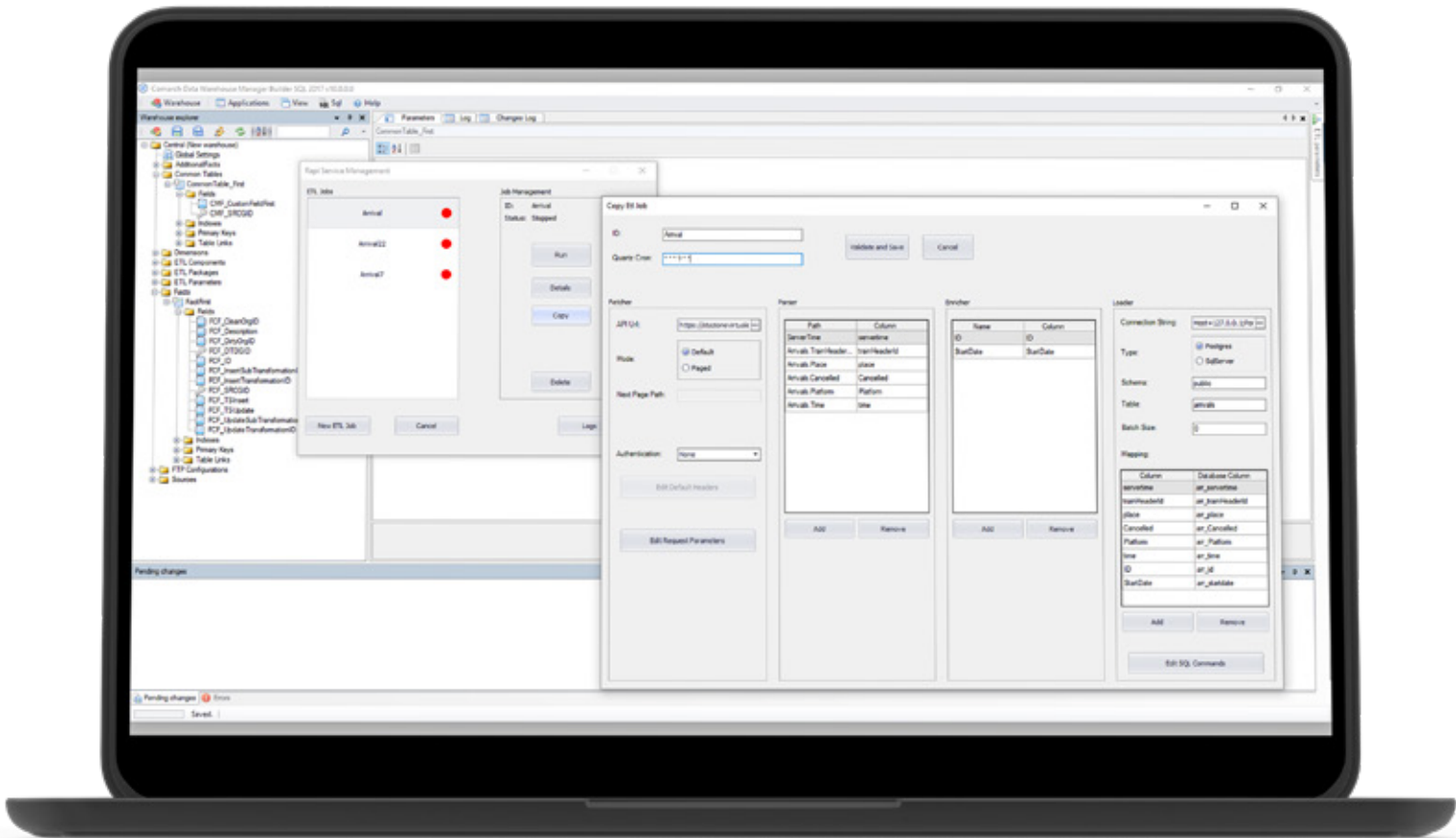
Die Anwendung Comarch Data Warehouse Manager (DWM) ist Bestandteil der Lösung Comarch Business Intelligence (BI). Ihre Aufgabe besteht darin, Daten mittels ETL-Prozesses an einem Ort zu sammeln – dem Data Warehouse. Hierdurch ermöglicht DWM die Verwaltung und die optimale Nutzung der Daten. Zu den weiteren Aufgaben von Comarch DWM gehören die Ausführung und Überwachung der Datenübertragung in das Data Warehouse sowie die Vereinfachung der Prozesse rund um die Verwaltung, Sortierung und Beschreibung der betreffenden Daten. Je nach Spezifikation des umzusetzenden Projekts, der Anzahl unterschiedlicher Datenquellen und den Anforderungen des Kundenunternehmens kommen unterschiedliche Module zum Einsatz, wie zum Beispiel das Modul Rapi und das Modul Kapi.

Comarch DWM ermöglicht:

- Verwaltung der Data-Warehouse-Struktur
- Steuerung der Lade- und Transformationsprozesse
- Protokollieren von Ereignissen während des Lade- und Transformationsprozesses sowie Bereitstellung von Informationen über den Prozessverlauf
- Kontrolle und Verarbeitung zusätzlicher Mappings, Gruppierungen und Aufteilungen sowie Pflege von Wörterbüchern in zusätzlichen Datenquellen

GRUNDLEGENDE FUNKTIONEN

- Mit Comarch DWM lässt sich ein Data Warehouse strukturieren, es können Dimensionen erzeugt sowie Daten und Fakten geordnet und gespeichert werden. Die Anwendung ermöglicht zudem die automatische Erzeugung der ETL-Pakete, über die die einzelnen Data-Warehouse-Tabellen mit Daten versorgt werden.
- Sämtliche Daten bezüglich des ETL-Prozesses und der Data-Warehouse-Struktur werden in Metatabellen gespeichert. Dadurch lassen sich Berichte erstellen, die die Datenbankstruktur bzw. den Ablauf des ETL-Prozesses wiedergeben.
- Das System ermöglicht die Verwaltung von Datenquellen, wobei die Angaben zur Datenherkunft in einer Zieltabelle gespeichert werden. Dadurch können mühelos Berichte zu den einzelnen Datenquellen erstellt werden.
- Der Zugriff auf die Anwendung wird den Benutzern in unterschiedlichen Berechtigungsstufen gewährt. Bei Änderungen ist es möglich, die ETL-Prozessdefinition an die neuen Anforderungen anzupassen sowie zusätzliche, in der Datenbank auszuführende Vorgänge zu definieren.
- Zu den einzelnen Prozessen werden Logdateien in einer eigenen Datenbank gespeichert, wodurch jederzeit Zugriff auf Informationen zum Verlauf von ETL-Prozessen, zu eventuellen Problemen, fehlerhaften Quelldaten oder zur Erstellung des Data-Warehouse besteht. Diese Informationen können in eine Datei exportiert und anschließend in eine andere Instanz von Comarch DWM importiert werden.



Hauptvorteile von Comarch DWM:

- Automatische Erstellung eines Data Warehouse
- Überwachung und Automatisierung der Ladeprozesse
- Paralleles Befüllen mehrerer Tabellen
- Laden von Daten mit unterschiedlichem Detaillierungsgrad
- Unterstützung von sich langsam verändernden Dimensionen (engl.: Slowly Changing Dimensions, SCD)
- Zusammenführung unterschiedlicher Datenquellen wie bspw. Flat Files, XLS-Dateien, MS-SQL-Datenbanken, Oracle-Datenbanken, OLE DB
- Datenübernahme aus Postgres-Datenbanken und JSON-Dateien
- Einblick in den Verlauf des ETL-Prozesses in Echtzeit

WEITERE ENTWICKLUNG

Um den aktuellen Trends im Business-Intelligence-Bereich gerecht zu werden, wird die Anwendung Comarch DWM zu einem verteilten ETL-System weiterentwickelt, das sowohl mit kommerziellen als auch mit Open-Source-Datenquellen zusammenarbeiten kann. Durch die Einführung von Mechanismen, die bspw. die Zusammenarbeit mit Kafka unterstützen, können Daten in Echtzeit extrahiert und im Data Warehouse weiterverarbeitet werden.

Die weitere Entwicklung der Anwendung erfolgt nach den Prinzipien der Cloudfähigkeit. Als Netzwerkanwendung ist Comarch DWM einfacher zu warten und ermöglicht eine schnelle Reaktion auf Änderungen und Implementierungsanforderungen. Gleichzeitig kann die Rechenleistung, die zur Ausführung der einzelnen ETL-Aufgaben benötigt wird, auf mehrere unterschiedliche Umgebungen verteilt werden.



Weitere Informationen zu Comarch BI finden Sie unter:
www.comarch.de/produkte/business-intelligence

NEUESTE FUNKTIONEN VON COMARCH DWM

In der neuesten Version wurde die Leistungsfähigkeit von Comarch DWM verbessert. Zudem wurden folgende Funktionen implementiert:

- Fakten- und Dimensionstabellen können mit Daten aus JSON-Dateien beladen und die Dateistruktur kann auf ihre Richtigkeit geprüft werden.
- Änderungen im Data Warehouse können zusammen mit Kommentaren zu jeder einzelnen Objektänderung gespeichert werden.
- In einem neu hinzugefügten Protokollbereich können alle im Data Warehouse vorgenommenen Änderungen nach ausgewählten Kategorien eingesehen und sortiert werden.
- Eine separate Anwendung – Migrator – wurde ergänzt, mit der Metadatenbanken migriert und Migrationsskripts erstellt werden können.
- In Tabellen wurde die Gruppierung nach Fakten und Dimensionen ermöglicht.
- Ein neuer, universeller Mechanismus zur Aktualisierung von Datenstrukturen im Data Warehouse wurde implementiert.
- Es wurden Zusatzmodule entwickelt, die mit REST-API und Kafka-Queues zusammenarbeiten. Dadurch konnten folgende Funktionen bereitgestellt werden:
 - Aufbau von Verbindungen zu externen REST-API- und Kafka-Services (betrieben auf Linux- und Windowssystemen)
 - Testen der Verbindungen und Speichern von deren Einstellungen
 - Einsehen und Ändern der Verbindungsdaten eines Service sowie Entfernen eines Service aus der Liste
 - Verbinden mit den ausgewählten, gespeicherten Services
 - Überblick über die ETL-Jobs für einen verbundenen Service
 - Verwalten der ETL-Jobs eines verbundenen Service, d. h. Starten, Anhalten, Einsehen, Bearbeiten, Entfernen, Kopieren, Neuerstellen
 - Einsehen der Logdateien eines Service

