



Wetterdaten punktgenau

MeteoSchweiz setzt beim Data Warehouse auf Software von Informatica

HERAUSFORDERUNG

Ziel war der Aufbau einer zentralen Drehscheibe für alle meteorologischen und klimatologischen Daten zur Sicherstellung einer hohen Daten- und damit Dienstleistungsqualität nach internationalen Standards.

BENEFIT

Mit dem jetzigen Data Warehouse ist MeteoSchweiz nun für die vielfältigen Aufgabenstellungen eines Wetterdienstes bestens gerüstet: Die von MeteoSchweiz-Stationen sowie Partnerwetterdiensten zugeliferten Messwerte können jetzt erfasst, geprüft und validiert werden. Das Data Warehouse enthält sowohl kurzfristige, sogenannte „near-realtime“, als auch langfristige Messwerte, die bis ins Jahr 1755 zurückreichen. Das Data Warehouse von MeteoSchweiz ist mithin die Grundlage für eine zeitgemäße, fachbereichsübergreifende Datenbasis für Wetterforschung, auf die eine Vielzahl von Anwendungen aufsetzen kann.

INFORMATICA-LÖSUNG

- PowerCenter

INDUSTRIEZWEIG

- Behörde

„Unser Ziel war es, eine metadatenbasierende Infrastruktur aufzubauen und die Staging-Prozesse möglichst weitgehend zu automatisieren. Hauptargumente für Informatica waren Raschheit bei der Implementierung, Flexibilität sowie die Möglichkeit, eigene Libraries einzubinden.“

— Christian Häberli, Projektleiter von MeteoSchweiz

MeteoSchweiz ist das Bundesamt für Meteorologie und Klimatologie unserer eidgenössischen Nachbarn und hat als solches eine Vielzahl von Aufgaben wahrzunehmen. Hierzu zählen unter anderem Wetterprognosen, Klimaanalysen und Verarbeitung von meteorologischen Daten. Die Präzision der Auswertungen hängt dabei entscheidend von der Qualität der Daten ab – und die stammen aus unterschiedlichen und verteilten Quellen. Datenintegration ist also bei der MeteoSchweiz eine strategische Aufgabe – hier spielt Software von Informatica eine entscheidende Rolle.

Die Herausforderung

Beim Aufbau der geforderten zentralen Drehscheibe für alle meteorologischen und klimatologischen Daten haben sich die Verantwortlichen schnell für eine Lösung mit einem Data Warehouse (DWH) entschieden. Das DWH-System von MeteoSchweiz sollte die zugeliferten Messwerte von den MeteoSchweiz eigenen Messstationen und die Daten der Partnerwetterdienste erfassen, prüfen und archivieren. Die neue Lösung musste kurzfristige (near-realtime) sowie auch langfristige Messwerte (bis zum Jahr 1755) enthalten. Und schließlich ist das DWH die Grundlage für eine moderne fachbereichsübergreifende Datenbasis, auf die sowohl fachspezifische Anwendungen über genormte Schnittstellen zugreifen, aber auch Off-the-Shelf-Reporting-Werkzeuge.

In einem solchen Konzept spielt daher die Plattform zur Datenintegration eine entscheidende Rolle. Denn nur mit einem leistungsfähigen ETL-Tool (extract, transform, loading) können die riesigen Datenmengen nahtlos aus den verschiedenen Quellen integriert werden. Zudem ist Performance gefragt: Das System muss den Zugriff auf Milliarden Datensätze erlauben, muss alle gängigen Hardund Softwaresysteme im Griff haben, und zwar mit einer Systemoffenheit, die auch zukünftige Applikationen erlaubt. Offene Skalierbarkeit und freie Konfigurationsmöglichkeiten ohne Schnittstellenproblem sind Grundvoraussetzung für eine detaillierte Sicht auf alle Wetterdaten.

„Für uns war schnell klar, dass Informatica PowerCenter die zentrale Plattform für die Datenintegration sein sollte“, so Christian Häberli, Projektleiter von MeteoSchweiz, zur Entscheidung für Informatica.

„Hauptargumente für Informatica waren Raschheit bei der Implementierung, Flexibilität sowie die Möglichkeit, eigene Libraries einzubinden.“

Data Warehouse im Einsatz

Entsprechend dieser Anforderung nahm MeteoSchweiz gemeinsam mit Siemens Business Services den Aufbau eines Data Warehouse-Systems über mehrere Standorte hinweg in Angriff. Das Data Warehouse sollte eine Infrastruktur zur Verfügung stellen, mit der meteorologische und klimatologische Grunddaten und die dazu gehörigen Kontextinformationen gesammelt, gespeichert, aufbereitet, bearbeitet, bereitgestellt und archiviert werden können. Die Daten sind dann die Grundlage für Wetterprognosen und Warnungen und dienen der Erstellung von Flugwetterplänen und Klimaanalysen.

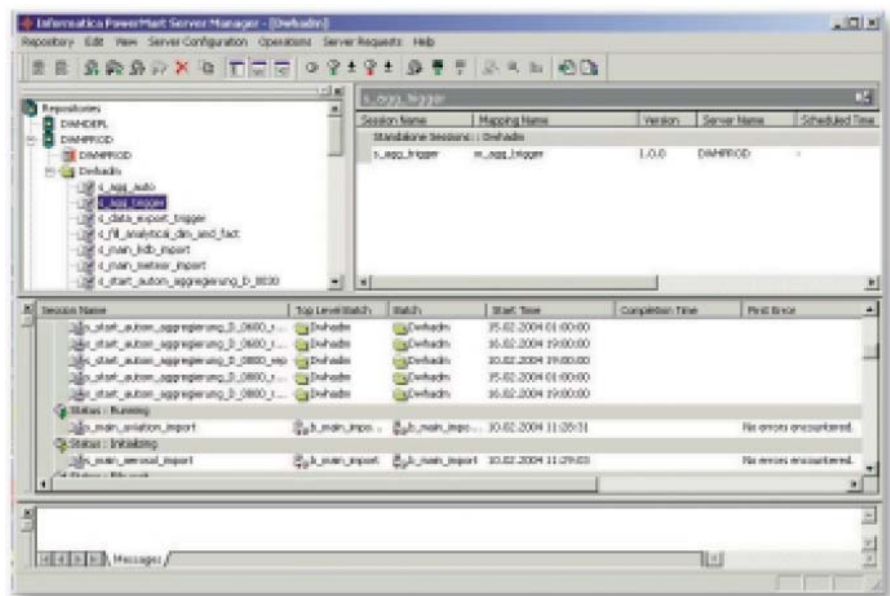
Die meteorologischen Daten und Klimainformationen fließen aber auch in verschiedene Projekte wissenschaftlicher, akademischer und privater Forschungsprojekte ein. Zusätzlich sind die Erhebungen Basis des Schweizer Klimaatlasses. Und nicht zuletzt können alle Privatpersonen, z. B. über das Internet, von den Dienstleistungen rund um das Wettergeschehen direkt profitieren. Grundlage für die breite Datenbank ist ein Data Warehouse-System, für das die folgenden Grundprinzipien gelten:

- Das System (siehe Abbildung 1) ist in vier verschiedene getrennte Ebenen unterteilt: Die Quellebene, die Ebene zur Datenaufbereitung, die Datenspeicher- und Nutzungsebene sowie die Applikationsebene. In diesen Ebenen werden die meteorologischen Daten generisch aufbereitet. Dieser Prozess dient der kontinuierlichen Qualitätsverbesserung der Mess- und Beobachtungsdaten.

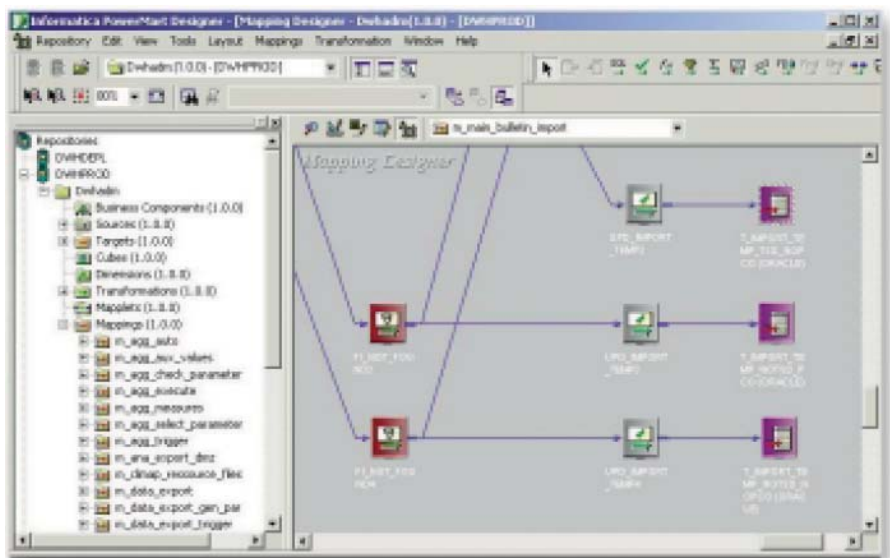
- Ein „Metadatenrepository“, das aus allen vier Ebenen zugänglich ist, enthält die Kontextdatenbank, die die klimatologischen Metadaten (zum Beispiel Stationsinformationen) speichert. Hierdurch wird sichergestellt, dass in allen Schichten dieselben Kontextdaten verwendet werden.
- Werkzeuge für die Datenaufbereitung sorgen dafür, dass die Qualität der meteorologischen Mess- und Beobachtungsdaten überprüft und verbessert werden kann.

Der Datenaufbereitungsprozess

Betrachtet man die Funktionsweise des Data Warehouse, bei der die Daten im Lauf der Prozesskette verschiedentlich transformiert werden nun etwas genauer, so wird deutlich, welche Bedeutung der Datenintegration auf den fünf Daten-Levels zukommt (siehe Abbildung 2). Daten aus dem untersten Level werden vom Datenerhebungssystem erzeugt. Dies kann eine automatische Wetterstation am Boden, ein Satellit oder auch eine Radiosonde sein. Falls notwendig, werden die von den Sensoren gelieferten Daten



PowerCenter Screenshots bei MeteoSwiss



PowerCenter Screenshots bei MeteoSwiss

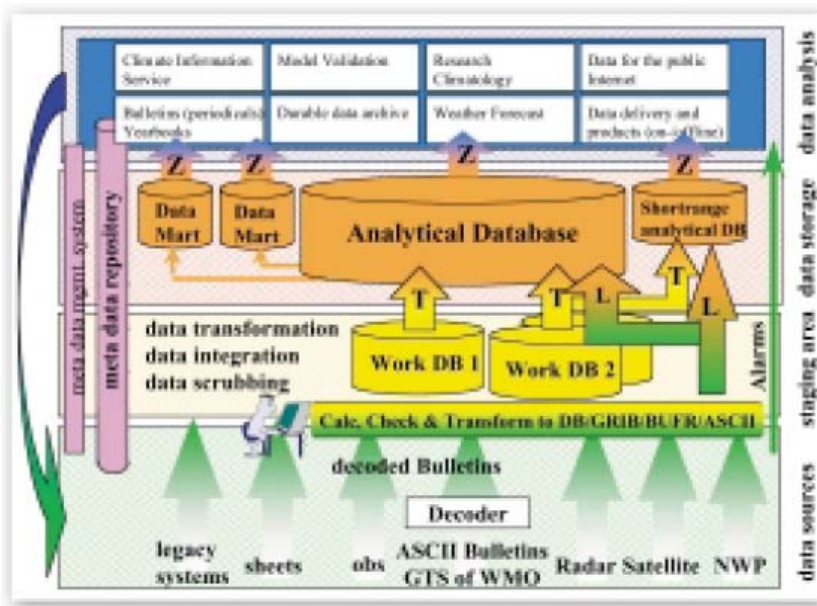


Abbildung 1

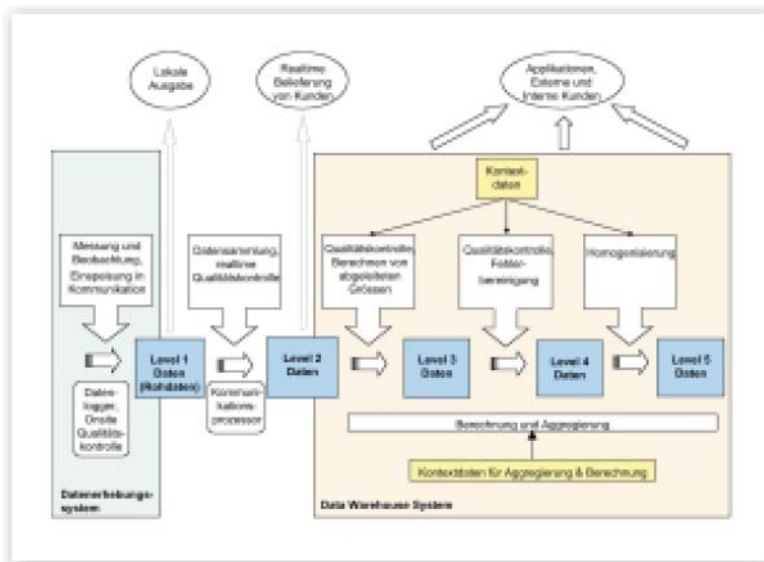


Abbildung 2

(Spannung, Zählung etc.) in physikalische Einheiten umgewandelt, die in die meteorologische Anwendung passen.

Beim Schritt vom ersten in den zweiten, also vom Quell- in den Datenaufbereitungslevel werden Daten gesammelt: Die an verschiedenen Messorten gewonnenen Daten werden hier von einem Kommunikationsprozessor zentralisiert. Einige Daten werden dabei im zentralen Sammelsystem anhand von Definitionen und Vorgaben vereinheitlicht und umgerechnet, so dass diese verglichen werden können. Hierbei erfolgt oft auch schon eine erste

systemspezifische Plausibilitätskontrolle, die zum Ziel hat, offensichtliche „Ausreißer“ zu kennzeichnen.

Die Rohdaten werden dann in einem sogenannten „raw data repository“ (Rohdaten) gespeichert. Dieser Datenpool bildet die Auswertungsbasis für alle Abfragen, die bei der Transformation von Level 1 zu Level 2 auftreten – sowohl für neue Abfragen aber auch bei wiederholten Vorgängen.

„Unser Ziel war es, eine metadatenbasierende Infrastruktur aufzubauen und die Staging-Prozesse möglichst weitgehend zu automatisieren“, so Ch. Häberli,

MeteoSchweiz. Bei diesen Prozessen spielt PowerCenter von Informatica die entscheidende Rolle. Die Software integriert die Daten auf Anfrage sowohl in Echtzeit, durch die Batch-Verfahren oder auch bei Veränderungen von Datensätzen in den verschiedenen Anwendungen, Datenbanken oder Systemen. Dabei kann es sich, wie bei der Infrastruktur von MeteoSchweiz um unterschiedliche Quellen mit verschiedenen Formaten handeln.

Im Schritt von der Datenquellschicht in die Staging Area, d.h. am Übergang von Level 2 zu Level 3, werden die verschiedenen Quelldaten zunächst identischen Plausibilitätstests unterzogen. Aus den hieraus abgeleiteten, plausiblen Wetterwerten wird ein einheitliches Set abgeleiteter Größen berechnet. Die Daten werden im Anschluss in den Arbeitsbereich des Data-Warehouse Systems geladen, wo sie für die weitere Verarbeitung bereitstehen. Außerdem können auf der Basis dieser Daten erste Berechnung von Aggregationen, zum Beispiel Tageswerten, durchgeführt werden. Im Schritt von Level 3 nach Level 4 werden die Daten dann im eigentlichen Sinne bereinigt.

Diese Bereinigung erfolgt in zwei Schritten, in einem automatischen und in einem interaktiven. Im automatischen Schritt werden die Daten weiter auf Plausibilität hin überprüft, um Tests auf zeitliche und später auch räumliche Konsistenz durchzuführen. Ziel bei diesen Plausibilitätstests ist es, offensichtlich falsche Werte automatisch zu korrigieren und kleine Datenlücken nach bestimmten Regeln aufzubereiten. Im interaktiven Teil werden die nicht automatisch korrigierbaren Daten manuell modifiziert. Bei diesem Vorgang werden die Daten historisiert – sie bleiben damit für nachträgliche Untersuchungen erhalten.

Aus diesen Bereinigungsprozessen und Plausibilitätsprüfungen gehen schließlich auf Level 5 homogenisierte Daten hervor. Basierend auf Referenzdatenreihen und Angaben aus der Stationsgeschichte werden mit Hilfe statistischer Methoden Brüche in den Zeitreihen festgestellt und ausgeglichen. Ziel ist dabei, Zeitreihen zu erhalten, die lediglich die Variationen umfassen, die tatsächlich durch Wetter und Klima hervorgerufen werden und somit andere Einflussquellen auszuschließen.

Unternehmensprofil MeteoSchweiz

Die MeteoSchweiz, die an fünf Standorten rund 250 Mitarbeiter beschäftigt, erfüllt meteorologische Aufgaben als Dienstleister für die Bevölkerung, die Wirtschaft und öffentliche Institutionen.

Dabei ist es eine der Hauptaufgaben der MeteoSchweiz, meteorologische Daten zu erfassen, zu archivieren, auszutauschen, auszuwerten sowie klimatologische Informationen bereitzustellen. Allerdings kennt das Wetter keine Grenzen. Und deshalb arbeitet MeteoSchweiz auch eng mit anderen Wetterdiensten und internationalen meteorologischen Organisationen zusammen, die ihrerseits Daten bereitstellen.

Das Ziel der Sicherstellung einer hohen Daten- und damit Dienstleistungsqualität nach internationalen Standards kann allerdings nur dann erreicht werden, wenn die relevanten meteorologischen Daten in einem zentralen Auswertungssystem kontinuierlich gepflegt werden.

UNTERNEHMENSPROFIL INFORMATICA

Informatica ist ein führender Anbieter von Software für Datenintegration. Mit Hilfe der Produkte von Informatica können Unternehmen system-, prozess- und anwenderübergreifend auf ihre Daten zugreifen, sie integrieren, migrieren und konsolidieren. Dadurch wird Komplexität verringert, Konsistenz gewährleistet und die Organisation wird leistungsfähiger. Weltweit nutzen mehr als 2.200 Unternehmen die Softwarelösungen von Informatica und decken damit den gesamten Bedarf an unternehmensweiter Datenintegration ab.

KURZÜBERSICHT

Gründungsjahr: 1993

Firmensitz: Redwood City, Kalifornien/USA

Mitarbeiter: 800+

Kunden: 2.200+

NASDAQ: INFA

PRESIDENT UND CEO

Sohaib Abbasi

INFORMATICA
The Data Integration Company™

Informatica GmbH Lyoner Strasse 15 60528 Frankfurt am Main
Tel: +49 (0)69 92 88 09-0 Fax: +49 (0)69 92 88 09-500 www.informatica.com/de

Informatica Niederlassungen weltweit: Australien · Belgien · China · Deutschland · Frankreich · Großbritannien · Irland · Japan · Kanada · Korea · Niederlande · Schweiz · Singapur · USA

© 2009 Informatica Corporation. Alle Rechte vorbehalten. Informatica, das Informatica Logo, Informatica PowerCenter, Informatica Data Quality, Informatica Data Explorer und PowerExchange sind Warenzeichen oder eingetragene Warenzeichen von Informatica Corporation in den Vereinigten Staaten und im rechtlichen Sinne weltweit. Alle anderen Firmen- und Produktnamen sind Markennamen oder Warenzeichen der entsprechenden Unternehmen.

(05/05/2009)