



Informatica Built Data Warehouses Provide Critical Information On Nation's Water Quality

“We chose Informatica as our data integration solution because of its maturity, wide range of features, ease of use, and industrial strength, integrated architecture.”

—Harry House, Data Warehouse Practice Leader, USGS

FAST FACTS

CUSTOMER

The USGS has served the United States for more than 125 years through exploration and scientific study.

BENEFITS

- Cost savings and rapid results
- Improved decision-making capabilities
- Improved performance
- Ease of use and quick implementation

INFORMATICA SOLUTION

Informatica PowerCenter®

NUTS AND BOLTS

- Data Integration: Informatica PowerCenter
- Sources: Flat files, Oracle, SQL Server, IMS
- Target: Oracle
- Platform: NT

The United States Geological Survey (USGS), with a budget of close to \$1 billion a year, serves the nation by providing reliable scientific information to describe and understand the earth; minimize loss of life and property from natural disasters; manage water, biological, energy, and mineral resources; and enhance and protect the quality of life.

The Challenge

The United States Geological Survey's (USGS) National Water Quality Assessment Program (NAWQA), established in 1991, collects and interprets water quality data on the major groundwater and surface water systems of the nation. The USGS assesses conditions of water quality, causes of those conditions and trends so that water managers and scientists have a sound, scientific basis on which to act.

In looking to build the largest water quality dataset collected, the challenges confronted by the USGS team largely concerned data integration. The USGS has four major legacy data source systems, each requiring a different method for manipulating and accessing data because of unique data formats and reference systems. Since water quality has many different dimensions that help determine its status and future trends, scientists need an integrated and detailed view of all relevant information in order to effectively assess it.

Previously, researchers would use a variety of integration methods for water quality analysis and pull subsets of data into Microsoft Access or Microsoft Excel. However, it often took up to a day or more to compile data for analysis operating in this environment and sometimes the system would simply crash under the weight of moving several million records.

The Solution

To build a consolidated and integrated water quality data warehouse, the USGS data warehouse team selected the Informatica data integration platform because of its strong performance capabilities, integrated architecture, and ease of use. And now the USGS is powering several national-level data warehouse systems, including the NAWQA program and several solutions for regional cooperators.

The NAWQA has enabled scientists to quantify water quality problems and explain probable causes, providing water quality managers and scientists with a reliable and sound mechanism to make decisions. In fact, the NAWQA contains over six million water and biological results and has revolutionized the way USGS researchers find and verify water quality trends and hot spots by empowering researchers to test hypotheses that were not possible previously.

And to continue effectively analyzing water quality and stay current with trends, the NAWQA data warehouse requires periodic refreshes from multiple data sources, including flat files, Oracle databases, Microsoft Access systems, and an SQL Server transactional system. Data is integrated across a dozen different local, regional, and national agencies into an Oracle staging area. From there the data is transformed with multiple concurrent sources and with extensive use of key lookups, joined, and reformatted into a denormalized star schema for end user access.

The Results

Cost savings and rapid results

By visually defining mappings and transformations through the Informatica Designer GUI interface instead of writing custom code, the USGS has been able to save on development costs. In the project's first year, the USGS saved at least 30 days of development time and resources. And now, using Informatica data integration platform, the productivity of each developer is fully maximized.

Quicker, better decision-making capabilities

By having a consolidated view of water quality data, USGS scientists can select from \$70 million of chemical, biological and physical water quality data collected over the past 10 years in about 50 basins. For example, now through a consolidated view, scientists can analyze how different regions compare based on a particular pollutant.

Improved system performance

By using the Informatica in-memory server side caching to leverage systems resources, the USGS has been able to improve system performance by over 30 percent for complex data loads over what a custom solution offered.

Ease of use and quick implementation

Demonstrating the ease of use of the Informatica data integration platform, the USGS was able to begin developing mappings with minimal training. In fact, the USGS was up and running within a week of licensing the Informatica solution.

INFORMATICA®

The Data Integration Company™ Worldwide Headquarters, 100 Cardinal Way, Redwood City, CA 94063, USA
phone: 650.385.5000 fax: 650.385.5500 toll-free in the US: 1.800.653.3871 www.informatica.com

Informatica Offices Around The Globe: Australia · Belgium · Canada · China · France · Germany · Ireland · Japan · Korea · the Netherlands · Singapore · Switzerland · United Kingdom · USA

© 2008 Informatica Corporation. All rights reserved. Printed in the U.S.A. Informatica, the Informatica logo, and The Data Integration Company are trademarks or registered trademarks of Informatica Corporation in the United States and in jurisdictions throughout the world. All other company and product names may be trade names or trademarks of their respective owners.

First Published: 2004

6393 (09/16/2008)