



Bioinformatics Mining Database

Customer: A leading biotechnology company running a large biological data factory discovering information, both for external clients and internal use.

Business Problem

The customer employs a variety of experimental techniques, including high throughput mass spectrometry, all of which generate vast volumes of data. Because of this, the experimental and data processing techniques developed by the customer are at the core of their business.

The data from these different sources could be linked together in a very powerful and informative manner. However, this proved to be a time consuming and complex task due to the different data sources involved in the various processes. These included information stored on a variety of databases, including Informix, an Oracle database containing sample information, a SQL Server database at the heart of a data processing system and further data stored on a UNIX file system.

The customer envisaged a single data source that would collect and amalgamate the data from these various sources so the data can be mined as one collective whole. In addition to internal data mining this needed to be used as a source of subsets of data for external client data deliveries.

Tessella Solution

After an analysis of the different data sources involved, Microsoft SQL Server Data Transformation Service (DTS) packages were identified as being able to offer the flexibility to take data from all of the different sources, perform any required reprocessing and import the data into a single SQL Server database – the Bioinformatics Mining Database.

An automated loading system was developed which loads data incrementally, as it becomes ready in the source systems. The system comprises of a Visual Basic control program that calls a series of DTS packages which each perform the data loading for a different section of the Bioinformatics Mining Database. This allowed both greater control and loads of smaller granularity.

A database subsetting mechanism was also defined for the creation of external client data deliveries using stored procedures and database links, controlled by a Visual Basic user interface.



Results and Benefits

The customer now has a continually updating data resource, which allows almost immediate access to new experimental and processed data with a rich set of data mining tools.

Using a previous system, external client data deliveries took 2-3 weeks to build. Now data deliveries can be created at a rate of 1 – 2 per week.

Tessella plc 26 The Quadrant, Abingdon Science Park, Abingdon, Oxfordshire OX14 3YS, UK
T: +44 (0)1235 555511 | F: +44 (0)1235 553301 | E: info@tessella.com

Tessella Inc 233 Needham Street, Suite 300, Newton, MA 02464, USA
T: 1 617 454 1220 | F: 1 617 454 1001 | E: info@tessella.com

Tessella – successfully delivering IT and consulting services to world leaders in R&D, science and engineering.

For decades, Tessella has been successfully delivering IT and consulting services to world leaders in R&D, science, and engineering. Through the application of scientific methods and rigorous quality procedures, we enable clients in life sciences, energy, the public sector, and consumer industries to achieve a wide range of objectives, including, forecasting floods, developing fusion power, enhancing military sensor capability, increasing drug discovery and development efficiency, and reducing risk to health and the environment in the extraction and production of oil and gas. With offices in Europe and North America, global companies rely on Tessella for business critical assignments.

Copyright © Tessella plc 2009, all trademarks acknowledged.

