

## Mass Spectrometry Analysis Pipeline

### Customer

A leading biotechnology company which runs a large biological data factory using high throughput mass spectrometry to produce information for both external and internal clients, with the aim of identifying potential drug targets.

### Business Problem

The customer identifies proteins that are contained in their clients' samples by using an innovative, custom-designed analysis pipeline. This uses high throughput mass spectrometry followed by automated data analysis to identify proteins contained in each sample.

The customer initially used proprietary analysis software integrated into an in-house analysis framework that adequately performed this task for small volumes of data. They decided to replace this with a bespoke system that would allow them to produce algorithms tailored to their business needs, and to have a data pipeline that is more adaptable to future changes in laboratory practices and that would cope with greater data volumes.

### Tessella Solution

Tessella provided database experts to join the team of specialist scientific programmers and mass spectrometrists that the customer had put together to deliver the new system. A prototype database was quickly constructed using SQL Server. This was integrated with a range of other databases that were in operation (Oracle and Informix) with the help of SQL Server DTS (Data Transformation Services) packages.

Part of the spectral identification process involves making comparisons with known

sequences contained in public and proprietary databases such as the Human Genome project. This amounts to several hundred Gigabytes of data, so careful design decisions had to be made in order to optimize the loading and searching of this data.

A sophisticated queuing system was built to interact with the database. This uses PERL scripts running on a bank of PCs allowing parallel processing of the specialist analysis code. This ensured that the system was scalable and was able to cope with the expected large data volumes.

### Results and Benefits

The prototype quickly proved its worth, and proved so popular that it was soon adapted to provide data deliveries to customers.

Tessella have subsequently carried out a project to create a fully automated, high reliability, high availability production pipeline based on the prototype and designed to cope with the estimated data volumes of several Terabytes a year as the customer's business grows.

Copyright © Tessella Support Services plc

V1.R2.M0

#### HEAD OFFICE

Tessella Support Services plc

3 Vineyard Chambers • Abingdon • Oxfordshire • OX14 3PX • England.

Telephone: +44 (0)1235 555511 • Facsimile +44 (0)1235 553301 • E-mail: info@tessella.com

Registered in England No. 1466429. VAT Registration No. GB 229 0489 52.