Big Data Netezza Migration

Client Overview
The customer is a leading provider of information, services and technology for the healthcare industry. The customer has a presence in over 100 countries worldwide with over 55 years of business experience, providing Analytics and Strategic Consulting services for the Pharma and Healthcare industries.

Scope of Work:
The customer needed to migrate and consolidate its existing data, stored in an Oracle DB and in a mainframe system, to a robust Netezza platform. They also had a need to convert the Ab Initio ETL process to Netezza code.

Achieving these goals required:
- Understanding the current application landscape
- Data modeling for the target schemas in the Netezza platform
- Redesigning ETL routines to leverage the abilities of the Netezza platform for data transformation and load processes using Netezza scripts
- Data load and validation

Business Challenges:
- The client's mainframe-based application holds data on drug sales and distribution of prescription drugs, which the client offers to its customers. These data include manufacturers' direct sales information from more than 100 pharmaceutical companies, as well as indirect sales data from over 500 other sources
- The client needed a more flexible solution to handle the growth and variation in their data arising out of an expanding business
- The legacy mainframe application was difficult to maintain and use
- Data volume was more than 2 TB, and increasing at the rate of approximately 10% every year.
- Several performance issues in data loads led to lower frequency of data refreshes and limitations in generating reports from the legacy application
- The infrastructure and license costs required to operate and maintain the different technologies in the environment were increasing significantly

Technical Solution:
- A dimensional model was created on the Netezza platform.
- The flat files used by existing Ab Initio and mainframe (SAS/COBOL) scripts were utilized as source data files to avoid redesign of the data extraction process.
- Netezza SQL scripts were created to incorporate the logic in the current Ab Initio ETL and mainframe SAS/Cobol scripts for the data transformation and load process.
- Data was loaded to staging tables using Netezza NZLOAD / NZSQL, and then Netezza capabilities were used to execute the transformation jobs in parallel, allowing for migration of data to the Netezza database

Benefits:
- Using Netezza improved the performance of the ETL jobs by up to 30-fold compared to the earlier mainframes-based platform
- The Netezza platform-based solution provided extensibility for the addition of data for new countries and markets with few modifications. Extensibility of the application was demonstrated by loading the data for another country post-migration of the data for the US business unit
- The Netezza solution facilitated extensive and easier reporting. Data extracted from multiple sources using the applications were collated together to produce quality reports for end users
- Manual preparation of data for producing Excel sheets and PowerPoint slides was minimized, thereby making reporting less error prone
- Few other benefits included real-time data load capability, Massive Parallel Processing capabilities, lower maintenance costs, ease of scalability, and optimized in-database analytics

Technology Environment:

<table>
<thead>
<tr>
<th>Hardware</th>
<th>Mainframe</th>
</tr>
</thead>
<tbody>
<tr>
<td>Software</td>
<td>Netezza</td>
</tr>
<tr>
<td></td>
<td>Oracle</td>
</tr>
<tr>
<td></td>
<td>SQL Server</td>
</tr>
<tr>
<td></td>
<td>Ab Initio</td>
</tr>
<tr>
<td>Language Environment</td>
<td>Unix Shell Scripts</td>
</tr>
<tr>
<td></td>
<td>SQL</td>
</tr>
<tr>
<td></td>
<td>PL/SQL</td>
</tr>
<tr>
<td></td>
<td>SAS</td>
</tr>
<tr>
<td></td>
<td>COBOL</td>
</tr>
</tbody>
</table>
Solution Overview:
Architecture Framework
The architecture framework below highlights the application landscape before and after the migration to a Netezza platform.

Solution Flow:
Flat files used to load Legacy Applications ➔ Datawarehouse ➔ Various Reporting Tools

Flat files used to load Legacy Applications:
- Oracle
- Mainframe
- SQL Server

Datawarehouse:
- Netezza Performance Server
- Load & Transform
  - Reusable scripts for Loading & Staging data.
  - Mainframe/Ab initio logic incorporated into Netezza scripts for transformation

Various Reporting Tools:
- Ad hoc querying & analysis
- Reports

Notes:
-----------
Data flow prior to migration ➔ Data flow after migration

About Hexaware
Hexaware is a leading global provider of IT & BPO and consulting services. The company has achieved a leadership position in Business intelligence solutions across a wide range of domains such as Banking, Financial Services, Insurance, Travel, Transportation, Logistics, etc. Hexaware focuses on delivering business results and leveraging technology solutions by specialising in Business Intelligence & Analytics, Enterprise Solutions, Quality Assurance and Testing Services. Founded in 1990, Hexaware has a well-established global delivery model armed with proven proprietary tools and methodologies, skilled human capital and SEI CMMI-Level5 certification. Hexaware has a significant presence in the European market with local offices in the UK, Germany & the Netherlands. Our global footprint spans across 5 continents, 20 countries and 31 locations.

For all Hexaware office locations, visit: http://www.hexaware.com/worldwide.htm