Harnessing the Power of Advanced Analytics

How an appliance approach simplifies the use of advanced analytics

Introduction

The Netezza TwinFin i-Class™ advanced analytics appliance pushes the limits of analytics by fusing our ground breaking data warehouse appliance with high performance, scalable analytics that can process massive data to solve complex problems orders of magnitude faster than typical solutions. Netezza i-Class advanced analytics is an open and flexible advanced analytics platform that enables the development and deployment of analytics to drive game changing results.

Thomas Davenport in his book Competing on Analytics describes the requirements necessary for companies to move to the next frontier of intelligent applications using advanced analytics. Top on his list of requirements is the need for embedded analytics in an appliance. By embedding analytics into an appliance and making them an integral part of the appliance simplifies the deployment of analytics while fully capitalizing on the processing throughput of the underlying architecture.

With TwinFin i-Class, parallelized analytics for data preparation, data mining, predictive modeling and optimization can exploit the Netezza asymmetric massively parallel processing™ (AMPP™) architecture to achieve high throughput of advanced analytics on huge data. TwinFin i-Class can easily be extended to create your own powerful, advanced analytics and embed them into the appliance. Analytic applications, visualization tools and business intelligence tools can harness the parallelized advanced analytics via a variety of programming methods such as SQL, Java, MapReduce, Python, R, C, C++ and Fortran to deliver powerful, insightful analytics. The comprehensive advanced analytics environment makes it easy to have it your way – using your preferred tools for ad-hoc analysis, prototyping and production deployment of advanced analytics.
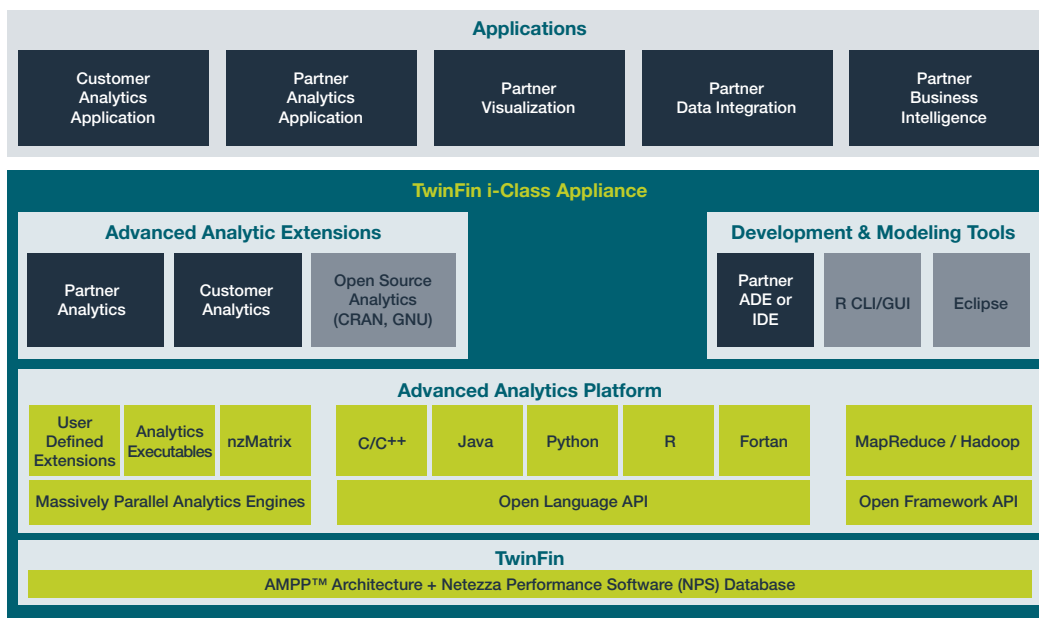
Figure 1 - Netezza TwinFin i-Class is a robust open platform for advanced analytics.

What does Netezza i-Class advanced analytics do for me?

Netezza i-Class advanced analytics exploits the convergence of huge data and parallel processing to enable high performance, scalable analytics applications. The Asymmetric Massively Parallel Processing ™ (AMPP™) architecture of the Netezza appliance enables analytics to capitalize on data and task parallelism to deliver blisteringly fast insightful results. By tightly integrating the analytics with our unique AMPP architecture, Netezza has pushed the technology envelope to deliver faster performance with the ability to process data that was impossible to process previously. This scaling for both data volume and computational complexity allows statisticians, modelers and developers to use the right tool at the right time to discover high value opportunities.

Today it is commonplace to copy data from a data warehouse in order to process it for analytics which results in longer development cycle times and higher software maintenance costs. This includes programs to extract, transform and load data between the data warehouse and other platforms. TwinFin i-Class simplifies the process by moving the analytics processing next to the data inside the data warehouse eliminating or significantly reducing these non-value added steps. By simplifying the environment, analysts can focus on higher value added activities such as modeling, simulation and optimization that can drive significant value for their organization.

TwinFin i-Class incorporates a rich set of advanced analytics from Netezza, our partners and the open source community. Application developers and modelers can use the right combination of advanced analytics for their unique situation to design and develop a comprehensive solution to their unique challenges. Companies can leverage the high-performance, parallelized analytics as a starting point to create their own "secret sauce" – or competitive edge – by creating their own high performance analytics or by leveraging the analytics to more quickly develop analytic applications.

For example, in CRM where there are a huge number of transactions and a high degree of variability among customers, customer segmentation is often used to identify high profit customers and predictive modeling can be used to determine how to convert low profit customers into high profit customers. Modeling and optimization can be combined to determine the best design, messaging and promotion to offer to customers to increase response rates. Data mining and predictive modeling can be used to identify and sustain loyal customers as well as approaches to increase customer loyalty with the appropriate incentives.

How do we use Netezza's i-Class advanced analytics?

Netezza i-Class is a platform for:

1. Building and deploying advanced analytic applications
2. Leveraging parallel analytics via visualization or business intelligence tools
3. Performing ad-hoc analysis especially on huge data or computational intensive problems

Visualization and business intelligence tools leverage the analytics in Netezza i-Class advanced analytics via SQL for targeted inquiries. However, application developers and modelers require more flexibility to develop applications or ad-hoc analysis including:

1. Data preparation and transformation
2. Model creation and publishing
3. Model scoring
4. Algorithm development

Let's take a closer look at each of these scenarios.

Data Preparation and Transformation

An analyst performing data prep and transformation activities can use i-Class tools and utilities to aid in the process. Such tools include:

1. Descriptive statistics
2. Data cleansing
3. Data transformation
4. Feature selection and dimension reduction
5. Statistical testing and utilities

When data prep and transformation are necessary, Netezza speeds up the process with high performance packages while scaling petabytes of user data that can be stored on the appliance.

Creating and Publishing a Model

Creating a model is typically an iterative process. The data used in building the model may be prepared and transformed by the modeler or by an analyst who prepares the data for the modeler. The modeler may start off performing an ad-hoc analysis by request of some executive. Once the results are reviewed the modeler may create a refinement that is more of a prototype for eventual production deployment. The modeler oftentimes develops multiple models during this process and evaluates and compares the performance of each of the model before selecting a model for final deployment. At each stage in the iterative development cycle, the modeler may use different development tools and analytics depending on the evolving requirements and performance targets.

i-Class offers flexibility at each stage of this process. The modeler can create their model using one of several supported development environments such as Eclipse, R GUI or directly using:
- Netezza's plug-in for Eclipse
- Netezza's plug-in for the R GUI
- Partner's integrated development environment (IDE)
- Partner's analytic development environment (ADE) to create an analytic workflow
- Any of the Netezza supported programming languages – SQL, Java, MapReduce, Python, R, Fortran, C or C++

The model can use any combination of advanced analytics including:
- Open Source Analytics – R CRAN packages and GNU Scientific library which are easily installed via i-Class
- Partner Analytics – Industry proven analytics that can be leveraged in application development

Once a model has been used to predict future outcomes, it can be used to automate optimal decisions by combining optimization along with the predictive modeling. By embedding optimization, the application can solve tough, complicated business problems such as:
- Minimizing shipping costs at the same time as minimizing inventory
- Minimizing capital investment at the same time as minimizing the risk of an entire portfolio
- Maximizing revenue at the same time as maximizing customer loyalty
- Minimizing marketing budgets at the same time as maximizing lead generation

This is done while taking into account diverse real world constraints including budget, staffing, customer preferences and behaviors, company preferences and many others.

Model Scoring

Once a model is built, the model can be applied to the voluminous data typically associated with a predictive model. This process is referred to as scoring. Examples of scoring include applying the model to millions of records to:

- Detect fraud
- Calculate credit scores
- Re-price customer accounts
- Predict future customer purchases
- Decide who should be sent various marketing promotions
- Recommend if credit lines should be increased/decreased
- Conclude if delinquent accounts should be sold off
- Determine product warranty strategies based on field support calls
- Determine the appropriate customer promotions to increase sales

Algorithm Development

The algorithm developer has many development tool choices including integrated development environments (IDE) for a GUI based approach to development or programming languages. Netezza provides a plug-in for Eclipse, an open source IDE. Additionally, Netezza partners provide IDEs which can speed up the algorithm development cycle.   Netezza supports the R GUI for development of algorithms based on the R language. Netezza supports several programming languages for algorithm development including Java, MapReduce, Python, R, Fortran, C, and C++ as well as the User Defined Extensions. User Defined Extensions include::

- UDFs – user defined functions process a row
- UDAs – user defined aggregates process a group of records
- UDTFs – user defined table functions return any size table
- Stored Procedures – are used to group functions, aggregates and other logic into procedures

After an algorithm is developed and tested can be registered to on the appliance which makes the algorithm accessible as any other in-database analytic via the supported programming languages.

Summary

The TwinFin i-Class is a powerful and extendable advanced analytics platform that simplifies the development, deployment and use of advanced analytics while delivering unmatched performance and scalability.   Netezza, partner, open source and your own advanced analytics can be combined and exploited via SQL, Java, MapReduce, Python, R, Fortran, C or C++ to deliver game changing results for your business by harnessing the high performance parallel architecture of the Netezza appliance. Ⓝ

Netezza Corporation   26 Forest Street. Marlborough, MA 01752   +1 508 382 8200 TEL   +1 508 382 8300 FAX   www.netezza.com

About Netezza

Netezza (NYSE: NZ) is the global leader in data warehouse and analytic appliances that dramatically simplify high-performance analytics across an extended enterprise. Netezza's technology enables organizations to process enormous amounts of captured data at exceptional speed, providing a significant competitive and operational advantage in today's data-intensive industries including digital media, energy, financial services, government, health and life sciences, retail and telecommunications. Netezza is headquartered in Marlborough, Massachusetts and has offices in North America, Europe and the Asia Pacific region. **For more information about Netezza, please visit www.netezza.com.**

NET7679_HARNESSING_WP_02_10