PolyAnalyst 6
Technical capabilities and system requirements

PolyAnalyst Features

Client/Server architecture

Server architecture enables the implementation of PolyAnalyst 6 as an enterprise level analytical system. It facilitates the collaboration between data analysts working on the same projects and sharing various related resources such as analysis scenarios, dictionaries, taxonomies, and multi-dimensional matrices. Server architecture helps enhance the performance of the system by performing calculations on the most powerful machines, reducing data transfer over the network, scheduling execution of tasks at a given time, and generating custom reports and condition based alerts for different groups of business users. It provides the centralized management and audit of the list of system users and their actions.

Security

Recognizing that data is one of the most valuable and sensitive assets of a modern organization, PolyAnalyst provides solid mechanisms to ensure data security. Communications between client and server are performed in a fully encrypted manner with a new encryption key generated by the server for every communication session. PolyAnalyst supports secure user login based on user rights and passwords and keeps track of individual and group rights and sequences of actions carried out by the users. In addition, full compliance with the requirements of the HIPAA legislation facilitates PolyAnalyst implementations at healthcare and insurance organizations.

Scalability and Performance

PolyAnalyst provides industrial level scalability: it can handle huge amounts of data within reasonable time intervals. This scalability is ensured through a combination of several factors. PolyAnalyst utilizes hard disk instead or RAM for holding all data and supporting meta-information. Special scalable implementation of analytical algorithms enables the system to process large volumes of data. PolyAnalyst user interface allows the user to develop complex data analysis scenarios without loading data in the system, thus saving analyst's time. The availability of PolyAnalyst as the first 64 bit analytical system and its server farm implementation help dramatically increase the performance of the system and allow numerous users to perform data analysis and report viewing simultaneously.

Reusable analysis scenarios

A typical data analysis project involves a sequence of steps of data loading, preparation, analysis and reporting. Frequently, the same sequence of analytical steps has to be carried out repetitively on new batches of data. PolyAnalyst is a self documenting system that provides intuitive visual tools for developing and editing reusable multi-step data analysis scenarios. The system is easy to learn and fun to use. Data analysis and report generation scenarios can be scheduled for re-execution at any given time. This ensures that business users have timely access to up-to-date reports built on most recent data.

Data loading and integration

Whatever your data sources are, PolyAnalyst provides means for loading and integrating these data. PolyAnalyst can load data from disparate data sources including all popular database, statistical, and spreadsheet systems. In addition, it can load collections of documents in html, doc, pdf and txt formats, as well as load data from an internet source. PolyAnalyst offers visual on-the-fly integration and merging of data coming from disparate sources to create data marts for further analysis. It supports incremental data appending and referencing data sets in previously created PolyAnalyst projects.

Data manipulation

Every data analysis project starts with data normalization, manipulation, and exploratory analysis. In fact, the dominant part of the analyst's time is spent on data manipulations preceding the application of machine learning algorithms. PolyAnalyst provides the analyst with a vast set of powerful manipulation tools for data cleansing, aggregation and derivation of new attributes. Virtually any data transformation task can be solved through sequential application of PolyAnalyst data manipulation tools.

Analytical algorithms

The main quality the user expects in a knowledge discovery system is its ability to use mathematical algorithms to learn from historical data and predict outcomes of future situations. PolyAnalyst provides a broad selection of analytical algorithms for clustering and categorization of data, solving prediction, link analysis and affinity grouping tasks, learning patterns and discovering anomalies in data. Ranging from neural networks and decision trees, to Bayesian Networks and Support Vector Machines, to CHAID and logistic regression, and to Case-based Reasoning and Evolutionary Programming, PolyAnalyst scalable algorithms enable the user to solve the analytical task at hand.

Text Analysis

The data might contain attributes holding free form text, like in incident reports, claims notes, or in survey responses. Or the project might require the analysis of huge collections of documents in various formats, possibly harvested from the Internet in real time. Whatever the task, PolyAnalyst offers a collection of text analysis algorithms that enable the data analyst to solve it. Based on Megaputer linguistic platform and the incorporation of various semantic dictionaries, PolyAnalyst represents a powerful natural language processing tool. PolyAnalyst provides tools for handling both analyst driven analysis, such as taxonomy based categorization and Text OLAP, and data driven analysis, such as intelligent spell checking, keyword, and entity extraction, clustering, and taxonomy creation.

Multi-dimensional analysis

The development of multi-dimensional cubes based on the analyzed data allows the user to answer a variety of business questions by slicing data across various dimensions. This technology is widely known as OLAP (for on-line analytical processing). PolyAnalyst offers a robust OLAP engine as one of its data analysis algorithm. In addition, PolyAnalyst enhances standard OLAP by adding dimensions defined on fields containing free form text and offering a unique multi-dimensional OLAP interface.

Interactive visualization

PolyAnalyst provides the data analyst with an immediate feedback on the results of their analysis. It offers interactive and visual user experience whenever possible. In fact, exploratory analysis represents a very important stage of the complete data analysis cycle. The user of PolyAnalyst commands a variety of tools for interactive visualization of the data and the results of the analysis: one and two dimensional histograms, pie charts, line charts, scatter plots, snake charts, link charts and trends graphs help the user make sense of the data.

Reporting

PolyAnalyst enables than data analyst to create custom reports delivering key results of the analysis to business users across the organization in a clean, consistent and easy to comprehend format. Interactive reports include a mixture of graphs, tables, numbers, text and links to other PolyAnalyst objects. Reports can be scheduled for re-execution at a given time to provide business users with results based on the analysis of the most up-to-date data. Static snapshots of reports can be exported to HTML and RTF format.

Technical Capabilities

1. **Data loading and integration**
    a. Loading from numerous data formats
    b. On-the-fly data integration on a set of keys
    c. Data merging
    d. Referencing data sets in other PolyAnalyst projects
    e. Exporting results of the analysis to external RDBMS
2. **Data cleansing, manipulation and exploratory analysis**
    a. String mapping
    b. Data aggregation
    c. Data transformations

       d. Sampling and partitioning
       e. Derivation of new attributes
       f. OLAP
       g. Statistics
       h. Link analysis

3. Predictive modeling and machine learning
       a. Classification
           i. Decision Tree
           ii. Bayesian Network
           iii. R-Forest
           iv. CHAID
           v. Case-based Reasoning
           vi. Neural Network
           vii. Support Vector Machine
       b. Clustering
           i. CHAID
           ii. Localization of Anomalies
       c. Prediction
           i. Neural Network
           ii. Evolutionary Programming (Symbolic Knowledge Acquisition)
           iii. Logistic Regression
           iv. Linear Regression
       d. Anomaly detection
           i. MediCop
           ii. Data Audit
       e. Affinity grouping
           i. Market Basket Analysis
           ii. Visual Link Analysis

4. Text Mining
       a. Text cleansing and normalization
           i. Deduplication
           ii. Built-in, imported and user defined dictionaries of phrases, stop-words, synonyms and other semantic relations
           iii. Intelligent spell checker
       b. User driven analysis
           i. Information retrieval with search queries of any complexity
           ii. Query expansion with synonyms and hypernyms/hyponyms
           iii. Taxonomy-based categorization
           iv. Finding similar reports
           v. Multi-dimensional reporting from text data (Text OLAP)
       c. Data driven analysis
           i. Text clustering (binary and hierarchical)
           ii. Taxonomy generation from the analysis of raw data
           iii. Keyword extraction
           iv. Visual cluster analysis
           v. Classification model training based on pre-categorized examples

5. Graphical presentation of results with interactive drill-down feature
       a. Histograms (2D and 3D)
       b. Scatter plots
       c. Line charts
       d. Thermal charts
       e. Rule graphs
       f. Link charts
       g. Snake charts
       h. Trends graphs

6. Reporting
       a. Custom reporting templates
       b. Reports summarizing up-to-date results
       c. Alerts
       d. Interactive drill-downs for business users

e.   Publishing reports to popular document formats

7. **Analysis scenarios**
    a.   Visual development of reusable analysis scenarios
    b.   Drag-and-drop development of reports
    c.   Group nodes feature for managing large scenarios

8. **Scheduling**
    a.   Automated execution of analytical scripts at given times
    b.   Automated publishing of reports

9. **Scalability**
    a.   Client/Server implementation
    b.   Client – Server communications over TCP/IP protocol
    c.   Utilization of hard disk instead of RAM
    d.   Scalable implementation of algorithms
    e.   64 bit implementation available
    f.   PolyAnalyst Server farm architecture available
    g.   Analytic scenario development prior to actual data loading

10. **Usability**
    a.   Interactive drag-and-drop experience throughout the system
    b.   Visual development of reusable data analysis scenarios
    c.   Tight integration of analytical and reporting applications
    d.   Visual creation of nice looking interactive reports for business users
    e.   Publishing reports to popular document formats for better collaboration
    f.   Group nodes
    g.   Subject areas

11. **Security**
    a.   User name and password based authentication
    b.   Support for LDAP and MS Windows based authentication
    c.   Fully encrypted client-server communications
    d.   User activities logging
    e.   Compliance with HIPAA regulations

System Requirements

PolyAnalyst Server

Minimum Server Requirements:
1.4 GHz Intel or AMD processor
1 GB RAM
10 GB of free hard drive space
Windows 2000 or XP operating system
Internet Explorer 6.0 or higher

Recommended Server Platform
2 GHz or higher Intel or AMD processor
3 GB or more RAM
100 GB or more of free hard drive space
Windows 2003 or XP operating system
Internet Explorer 6.0 or higher.

PolyAnalyst Clients

Minimum Client Requirements
1 GHz or higher Intel or AMD processor
128 MB or more RAM
50 MB or more of free hard drive space
64 MB video card
Windows 2000 or XP operating system
Internet Explorer 6.0 or higher

Recommended Client Platform
1 GHz or higher Intel or AMD processor
512 MB or more RAM
100 MB or more of free hard drive space
64 MB video card
Windows 2000 or XP operating system
Internet Explorer 6.0 or higher

Megaputer
Your Knowledge Partner™

Tel: **+1.812.330.0110**, *email:* **info@megaputer.com**