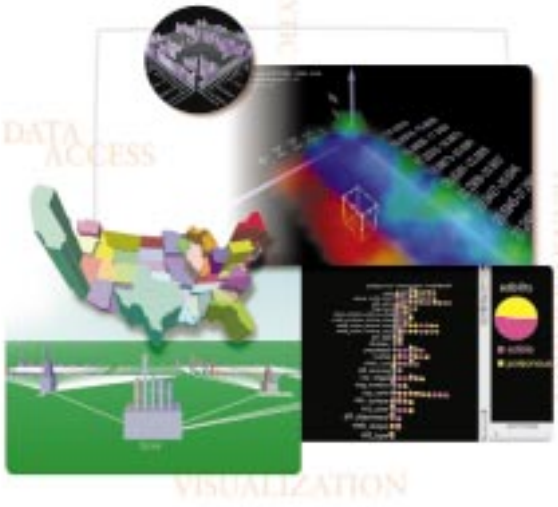
MineSet™ 2.6



www.sgi.com/go/mineset/

Maximize the Value of Your Data with Advanced Data Mining and Visualization Technology

Buried within your vast store of transaction logs and legacy data, subtle correlations and elusive relationships hold the secrets to enhance your business. To reveal the hidden value in your data warehouse, Silicon Graphics developed MineSet, an integrated suite of software tools for data mining and data visualization. Powered by industry-leading Silicon Graphics® work-stations and servers, MineSet revolutionizes the way you use data mining for strategic business analysis.

Intuitive Visual Data Mining Tools

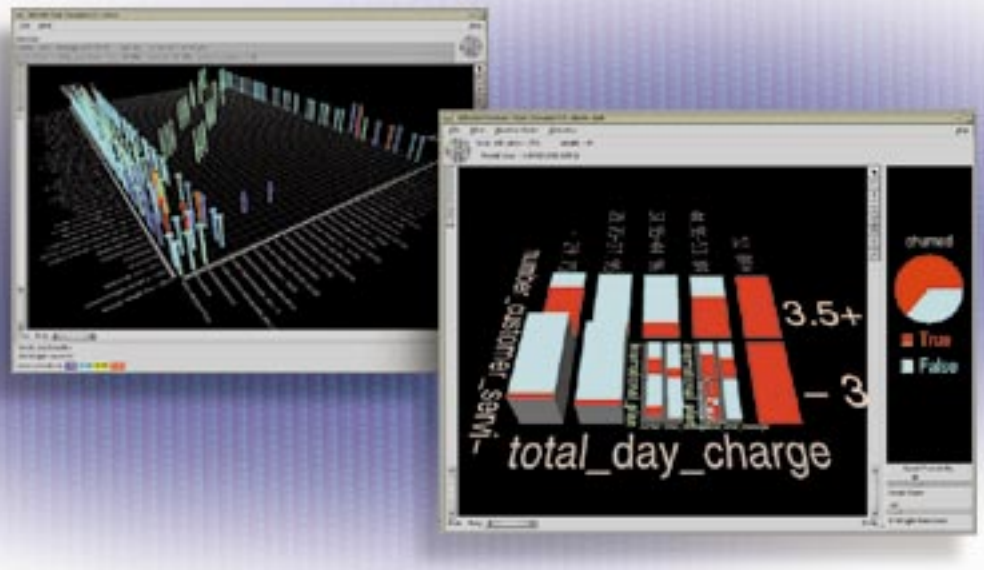
MineSet enables interactive exploration of data through an advanced suite of visual tools for faster discovery of meaningful trends and relationships. The *Splat Visualizer* and the *Scatter Visualizer* represent complex data in up to eight dimensions. The *Map Visualizer* displays data with geographical relationships by using a map metaphor. Animation and view synchronization techniques are used to reveal patterns over critical dimensions such as time. The *Tree Visualizer* depicts data with hierarchical relationships utilizing a fly-through technique set in a 3D landscape. The *Statistics Visualizer* presents a visual summary of basic statistical information. Advanced drill-through techniques give you fast access to the original records that created entities within your visualization for additional analysis. The *Record Viewer* allows quick access to the original data with column sorting and HTML output capabilities.

Powerful Analytical Data Mining Algorithms with Matching Visualizers

Below the elegant interfaces of the visual tools, state-of-the-art analytic data mining algorithms build comprehensive models. MineSet includes multiple classifier inducers, including a *Decision Tree*, an *Option Tree*, *Evidence*, and *Decision Tables*. Boosting allows MineSet to further improve the accuracy of these classifiers. *Decision Tree* and *Option Tree* classification models are viewed through the *Tree Visualizer*. An associated tool, the *Evidence Visualizer*, displays the structure and properties of the *Evidence Classifier* and supports instantaneous what-if analysis. The *Decision Table Visualizer* displays multidimensional data on important columns automatically chosen by the matching inducer. The visual-OLAP paradigm supports multilevel drill-through capability. The *Regression Tree* inducer supports regression, and the resulting model is visualized using the *Tree Visualizer*. The *Association Rules Generator* analyzes data to discover product affinities and relationships between data entities. This feature can extract both one-to-one and multiway associations. The resulting one-to-one rules are depicted by the *Scatter Visualizer* in an easy-to-understand graphical format. Multiway rules are summarized in a text report. The *Column Importance* feature enables automatic or user-guided selection of important columns for use with the MineSet visualizers. The k-means based *Clustering* algorithm is used for segmentation analysis and is visualized using the *Cluster Visualizer*.

Intuitive User Interface and Innovative Enterprise Solutions

MineSet is designed to fit the needs of users at several levels of technical ability. For technical users, MineSet offers unlimited potential with a complete suite of data mining tools, database integration, and scalable performance. To accommodate business users, data mining results and visualizations are easily deployed across corporate networks through point-and-click access. Analytical models developed in MineSet can be applied to any data set with the touch of a button. This easy scoring capability requires no programming or running of complicated batch programs. The MineSet Application Interface allows developers to embed MineSet tools within their customized solutions.



New MineSet 2.6 Features

- Internationalization (support for international characters and date formats)
- 64-bit analytics and support for very large memory
- Java™ Record Viewer with column sorting and HTML output capabilities
- Best-of-breed third-party analytics (optional)
- Multiway association rules with record weighting
- Ability to generate Web-launchable visualizations directly from the visualizers

Integrating MineSet into Your Computing Environment

With seamless support for Oracle®, Informix®, and Sybase® databases, MineSet makes it easy to add new capabilities to your data warehouse. Queries to your relational database are supported through graphical or SQL commands. An open architecture allows MineSet to coexist with other data mining and visual tools, including SAS® software. Hot links help MineSet users share discoveries and results through Web pages, launching appropriate MineSet visualization tools for further analysis.

Using X servers supporting OpenGL®, such as Hummingbird Communications Exceed® and Exceed 3D®, your existing PCs and workstations from other vendors interact with the full power of MineSet.

Industry-Leading Scalability and Large Memory Support

MineSet is optimized for the ultimate scalable performance on the Origin™ server. In addition to parallelization, MineSet offers very large memory support (over 2GB) and data handling capability using the 64-bit implementation.

Data Access and Transformation Features

- Support for access to Oracle, Informix, and Sybase running on all major UNIX® platforms, including IBM®, Sun™, HP®, and Digital
- Support for access to flat files (ASCII or binary)
- Data transformation history and graphical editing facility
- SAS file import/export utility
- Data transformation support for:
  - Automatic or user-specified binning of data
  - Data aggregations with indexed arrays using average, minimum, maximum, sum, and count
  - Data distribution (transpose)
  - New column creation using expressions
  - Column removal
  - Data type conversion
  - Sampling
- Save and restore session management
- Client/server architecture
- Statistical tool for determining minimum, maximum, means, median, standard deviation, histograms, and quartiles

Analytical Data Mining Features

Classification
- Decision Tree
- Option Tree
- Evidence (simple Bayes)
- Decision Table

Clustering
- K-means
- Iterative k-means
- Autoclass (optional)

Regression
- Regression trees

Associations
- One-to-one
- Multiway

Column Importance
- Automatic attribute selection

Advanced Analytical Options

User-specified parameters for control of classifier induction, including:

- Boosting
- Node-splitting criteria for decision tree induction
- Pruning factors for decision tree induction
- Laplace correction for evidence induction
- Automatic feature selection for evidence induction
- Data holdout percentage for automatic accuracy estimation (holdout and cross-validation)

Advanced features for classification, including:

- Loss matrices
- Lift curves
- Binning options
- Scoring
- Record weights
- Learning curves
- Backfitting

Visual Data Mining Features

- User-defined mapping of data to visualization components
- Visual and numeric normalization of data across visual components such as height, color, and size
- Discrete or continuous mapping of data to colors
- Web launching support
- Visual drill-up, drill-down, and drill-through to source data
- Animation of dependent data by up to two user-defined independent variables for trend analysis
- VCR-style animation playback
- Visual filtering and querying of data
- Automatic visual scaling adjustment
- Synchronous animation of multiple visualizations
- Visual level-of-detail controls
- 3D fly-through

Other Platform Support

All visualizations are based on OpenGL and can be viewed and controlled from platforms running an OpenGL enabled X server.

MineSet includes trial versions of Hummingbird Communications Exceed and Exceed 3D software, which run on PCs with Windows 95® and Windows NT® (at least 4MB of VRAM required).

Product Requirements

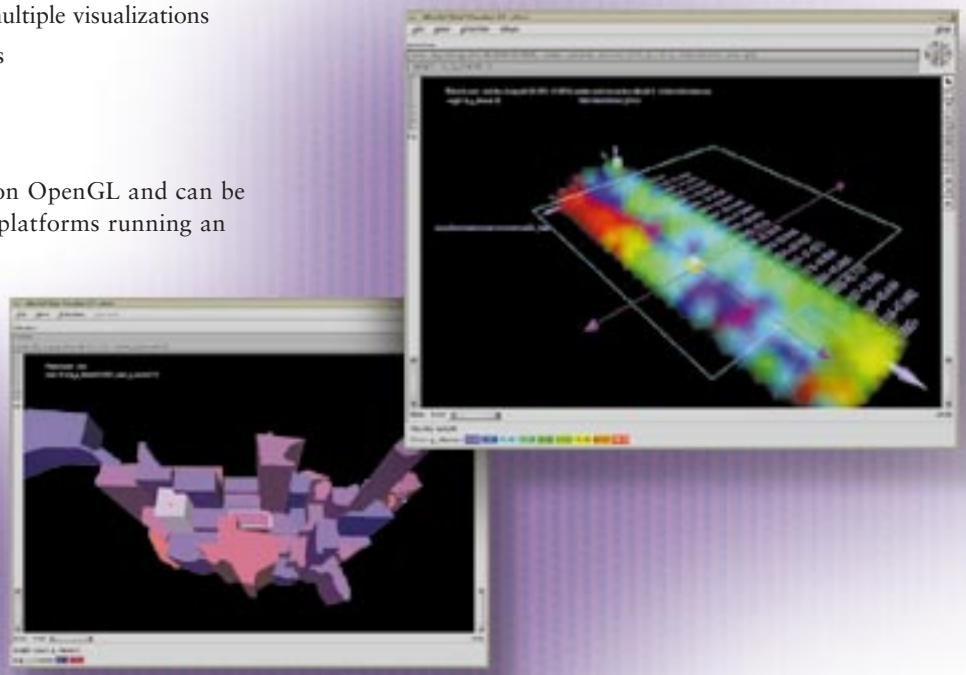
MineSet server and client are supported on all Silicon Graphics platforms running IRIX® 6.2 and above. IRIX 6.4 or higher is required for 64-bit support and parallelization. Memory requirements for the server vary with the size of the data. Clients must have at least 64MB of memory; 96MB is recommended.

MineSet Light

To acquaint new customers with the unique capabilities and power of MineSet, Silicon Graphics is offering MineSet Light, an entry-level version of MineSet at a fraction of the cost. It provides all the standard MineSet features, but the MineSet server automatically samples the data to 5,000 records. An upgrade to full MineSet is provided through a purchase of the MineSet server component.

Optional Analytics

Also available, at additional cost, is Ultimode ACPro, an automatic segmentation and clustering tool that works within MineSet. Ultimode Systems also provides consulting services and turn-key solutions using MineSet and ACPro. Visit the Ultimode Systems Web site at *www.ultimode.com*.

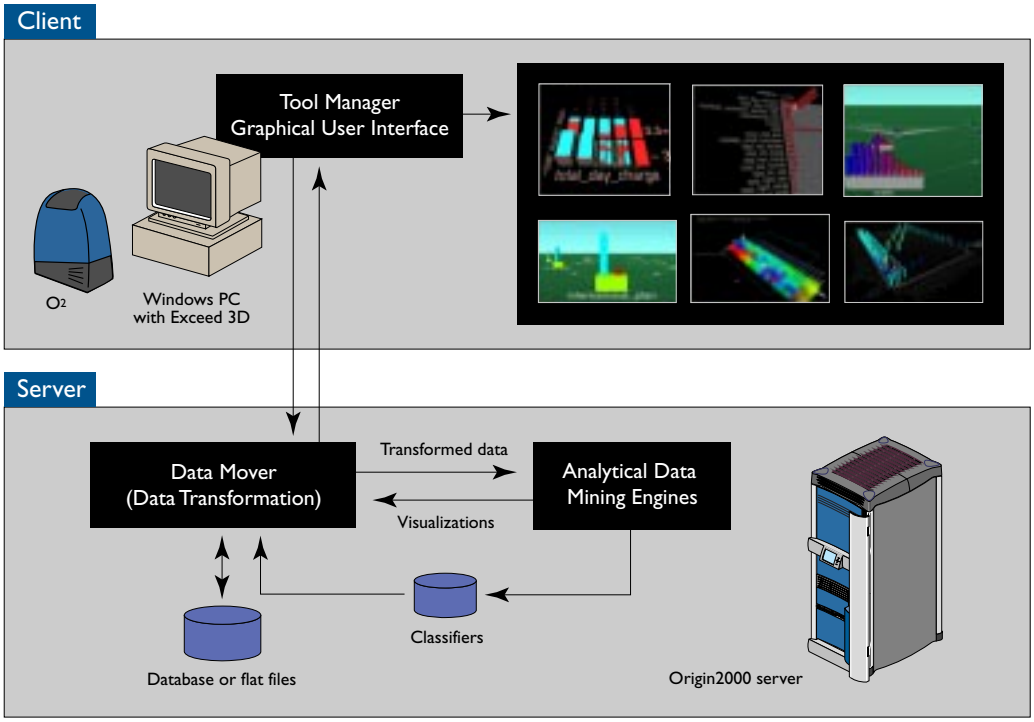


MineSet

Sample Configurations

	O2™	OCTANE™	ORIGIN200™	ORIGIN2000™
				
Sample Configurations	1 user entry level client/server	1 user heavy graphics client/server	1-4 users heavy analytics server only	2 or more users heavy analytics, big data server only
Model	1 CPU 200 MHz R5000®	1 CPU 225 MHz R10000®	2 CPU 250 MHz R10000®	8 CPU 250 MHz R10000®
Memory	128MB	512MB	1GB	4GB
System Disk	4GB	9GB	9GB	9GB
Additional Disk	*as needed*	*as needed*	*as needed*	*as needed*
Maximum Memory	1GB	2GB	4GB	256GB
Maximum CPUs	1	2	4	128

MineSet Client/Server Architecture



SiliconGraphics
Computer Systems

Corporate Office
2011 N. Shoreline Boulevard
Mountain View, CA 94043
(650) 960-1980
www.sgi.com

U.S. 1(800) 800-7441
Europe (44) 118-925.75.00
Asia Pacific (81) 3-54.88.18.11
Latin America 1(650) 933.46.37

Canada 1(905) 625-4747
Australia/New Zealand (61) 2.9879.95.00
SAARC/India (91) 11.621.13.55
Sub-Saharan Africa (27) 11.884.41.47

www.sgi.com/go/mineset/

312 (9/98)