

# STATISTICA

## Data Warehouse

- ✓ The solid foundation for your entire, enterprise-wide business intelligence system
- ✓ The core of the high-performance intelligence delivery infrastructure, designed to meet even the most demanding data integration and knowledge extraction needs
- ✓ The ultimate system for intelligent management of unlimited amounts of data, distributed across unlimited numbers of locations worldwide



## Contents

- [Overview](#)
- [The Crucial Features and Benefits](#)
- [Architecture and Connectivity](#)
- [Web-Enablement](#)
- [Advanced Security and Authentication](#)
- [Document Control](#)
- [Advanced Analytics](#)
- [Programmability and Customizability](#)

## Overview

The *STATISTICA Data Warehouse* system is a complete, powerful, scalable, and customizable intelligent data warehouse solution, which also optionally offers the most complete analytic functionality available on the market, fully integrated into the system.

*STATISTICA Data Warehouse* consists of a suite of powerful, flexible component applications, including:

- *STATISTICA Data Warehouse Server Database*
- *STATISTICA Data Warehouse Query* (featuring *WebSTATISTICA Query*)

- *STATISTICA Data Warehouse Analyzer* (featuring *WebSTATISTICA Data Miner*, *WebSTATISTICA Text Miner*, *WebSTATISTICA QC Miner*, or the complete set of *WebSTATISTICA Analytics*)
- *STATISTICA Data Warehouse Reporter* (featuring [WebSTATISTICA Knowledge Portal](#) and/or [WebSTATISTICA Interactive Knowledge Portal](#))
- *STATISTICA Data Warehouse OLAP* (featuring [STATISTICA OLAP](#))
- *STATISTICA Data Warehouse Document Repository* (featuring *WebSTATISTICA Document Management System*)
- *STATISTICA Data Warehouse Scheduler*
- *STATISTICA Data Warehouse Real Time Monitor and Reporter* (featuring *WebSEWSS*, or *WebSEDAS*)

If you are new to data warehousing, StatSoft consultants will guide you step by step through the entire process of designing the optimal data warehouse architecture, from a comprehensive review of your information storage and extraction/analysis needs, to the final training of your employees and support of your daily operations.

## The Crucial Features and Benefits

The crucial features and benefits of *STATISTICA Data Warehouse* solutions include among many others:

- Complete data warehousing application tailored to your business
- Platform independent architecture for seamless integration with your existing infrastructure
- Facilities to integrate data from a wide variety of sources
- Virtually unlimited scalability
- Options to update/synchronize data from multiple sources via automatic schedulers or on demand
- Completely Web-enabled system architecture to provide ultimate enterprise functionality for all company locations around the world (e.g., access via Web browsers from any location)
- Advanced security model and authentication of users
- Complete document management options to optimize management of documents of any types and satisfy regulatory requirements (e.g., FDA 21 CFR Part 11, ISO 9000)
- Comprehensive OLAP functionality for online data exploration and reporting
- Advanced analytic components to clean/verify data and to integrate automated data mining, artificial intelligence, and real-time process monitoring
- Options to automatically run and post on [Knowledge Portals](#) (or broadcast) highly customized reports, including interactive (i.e., drillable, sliceable, and user-customizable) reports and results of advanced analytics
- Backup and archiving options
- Programmable, customizable, and expandable to adapt to specific mission profiles (open architecture, exposed to extensions using the most industry standard languages, such as VB, C++, Java, HTML)

- Built on robust, well tested, highly scalable, cutting-edge technology to leverage your investment (including highly optimized in-place database processing (IDP) technology, and true multithreading, distributed/parallel processing and support for pooling CPU resources of multiple servers to deliver supercomputer-like performance)

*STATISTICA Data Warehouse* is a complete intelligent data storage and information delivery/distribution solution enabling you to customize the flow of information through your organization, to provide all authorized members of your organization with flexible, secure, and rapid access to critical information and intelligent reporting.

The system is virtually platform independent and will fit into any existing database architecture and hardware environment. It will efficiently combine information from multiple database formats and sources (from manual data entry forms to large batteries of automatic data collection devices). The system can be further enhanced through integration with other fully compatible components of the *STATISTICA* line of applications and solutions; to name just a few:

- [\*STATISTICA Data Miner\*](#) for advanced data mining and artificial intelligence (e.g., neural networks) based solutions to provide decision support through cutting-edge methods for knowledge extraction and prediction
- [\*Quality Control Miner\*](#) and [\*Enterprise Wide SPC System \(SEWSS\)\*](#) for tight integration with quality control, process control, and yield management activities
- [\*STATISTICA Text Miner\*](#) for automatic processing of unstructured information in documents, databases, or Web directories (Web-crawling of URL's)
- [\*STATISTICA Knowledge Portal\*](#) for presenting summary reports, charts, and action items to end users (management, sales force, engineers, etc.) through secure access portals via the Web; to deliver key intelligence and decision support to stakeholders world wide (e.g., you access the [\*STATISTICA Knowledge Portal\*](#) via standard Web browsers from anywhere in the world)

## Architecture and Connectivity

- The *STATISTICA Data Warehouse* connects to any platform, database, or data source, and will scale to businesses and applications of any size.

The *STATISTICA Data Warehouse* is built on a database and database schema customized for your particular business. The solution can be installed either inclusive of a high performance database engine (SQL Server) or as a (virtual) database schema compatible with most industry standard databases; therefore, it will seamlessly integrate into existing database systems.

Because *STATISTICA Data Warehouse* does not depend on one particular database vendor or hardware platform, it is itself entirely platform-independent. The main *Data Warehouse* software will connect to any database format, and hence can efficiently combine and pool information from multiple sources.

The *STATISTICA Data Warehouse* application software will run on servers with multiple processors, or banks of multiple-processor servers for super-computer like performance. The

system will scale effortlessly and economically to even huge data sizes and analysis (intelligence) problems.

## Web-Enablement

- The *STATISTICA Data Warehouse* extracts information from sources anywhere in the world, and delivers intelligence anywhere in the world.

The *Web* component of the system is built on the proven *WebSTATISTICA* technology that is used by organizations world-wide to provide secure access via standard Web browsers: Unlike other Web-based solutions, the *STATISTICA Data Warehouse* does not require any additional components to be installed on the (thin) client machines. Hence, the system can be utilized by (authorized and authenticated) users world wide, from hotel rooms via dial-up modems, from home, or from office and production facilities located at the most remote places on earth (e.g., via satellite Web links).

## Advanced Security and Authentication

- The *STATISTICA Data Warehouse* implements a detailed and sophisticated security system to ensure that your proprietary knowledge and intelligence is safe from unauthorized access.

The *STATISTICA Data Warehouse* system will likely become the most important repository of business intelligence and decision support resources in your organization. Therefore, the security of the system is a crucial priority, so that those valuable resources are shielded from unauthorized access.

*STATISTICA Data Warehouse* implements the highest level of security by establishing groups of users with different levels of authority (regarding the information that is accessible, and the operations that can be performed), requiring regularly updated passwords, etc. Special methods are also in place to detect and guard against systematic electronic intrusions ("hacking").

## Document Control

- The *STATISTICA Data Warehouse* enables full document management, compliant with government and industry standards.

[\*STATISTICA Document Management System\*](#) can be seamlessly integrated into your *STATISTICA Data Warehouse* application to optimize the flow of information within your organization and thus increase your productivity. This system can also be configured to comply with all (corporate) documentation management policies or regulatory requirements for document security, audit trails, and electronic signatures/authentication (e.g., as, for example, stipulated by FDA CFR Title 21 Part 11: *Electronic Records; Electronic Signatures*; or ISO 9001 4.5: *Document and data control*).

## Advanced Analytics

- The *STATISTICA Data Warehouse* can incorporate the most advanced data analysis and knowledge extraction methods available; you can go far beyond *OLAP* to simplify and extract knowledge about even the most complex - and inaccessible to other applications - patterns in the data.

Because the *STATISTICA Data Warehouse* is built from the same high performance components as the entire *STATISTICA* line of analytic solutions software, those analytic solutions can easily and seamlessly be integrated into your *Data Warehouse*. *STATISTICA* offers the most comprehensive set of tools for data mining, text mining, data analysis, graphics and visualization, quality and process control (incl. Six Sigma), etc. on the market. These resources and technologies can be connected to the data sources in the *STATISTICA Data Warehouse*, to leverage the most advanced technologies and algorithms available for analyzing and extracting key intelligence from all sources. For example, you can apply hundreds of neural networks architectures, highest performance tree classifiers (e.g., stochastic gradient boosting trees), flexible root cause analyses, control charting methods, powerful business forecasting methods, or sophisticated analytic graphics methods to convert raw data in the *Data Warehouse* into useful and actionable intelligence with clear implications for decisions affecting your business.

## Programmability and Customizability

- The *STATISTICA Data Warehouse* is an open-architecture system that will not lock you into a relation with a single vendor or solution; you can respond quickly to new business demands and requirements that need to be incorporated into the *Data Warehouse*.

As all applications and solutions in the *STATISTICA* family of products, the *STATISTICA Data Warehouse* is fully programmable and customizable, using industry standard programming tools such as Visual Basic, C++, Java, or HTML. This feature is of key importance when your business depends on your ability to quickly adapt to new information and business realities: Because you can customize the system without being forced to rely on the programmers of a single vendor or knowledge of idiosyncratic scripting conventions (required by many competing solutions), you have the freedom to develop your proprietary extensions to the data warehouse, and to add not only your own reports but also custom analytic and data transformation/cleaning procedures, using widely available resources and industry standard tools (e.g., VB, C++, Java, or HTML tools and programmers). Of course, StatSoft can always offer to you a full complement of consulting, system integration, and programming services delivered by an experienced staff, if you choose to work with us.