

FROM ERP TO DATA MINING ON THE WEB

Șerban GHENEA¹

Funcționalitățile specifice ERP (Enterprise Resource Planning) oferite de platforma ISIS.NET sunt completate de capacități avansate de data mining, obținute în baza integrării în sistem a serviciilor de business intelligence din Microsoft SQL Server. Folosind un simplu calculator conectat la Internet, utilizatorii pot actualiza și întreține baza de data operațională, depozitul de date și modelul multidimensional de date, administrând întregul ciclu de viață al datelor, de la informația primară până la rezultatele analizelor business intelligence.

The ERP (Enterprise Resource Planning) specific functionalities provided by the ISIS.NET web platform gain business intelligence capabilities following the integration of analysis and data mining services based on Microsoft SQL Server. From the operational database to the data warehouse to the multidimensional data model that supports business intelligence processing, the user is guided to conduct the conversion of raw data to a powerful decision-making foundation.

Keywords: ERP, SOA, web, data mining, business intelligence

1. Introduction

The first worldwide four vendors of enterprise applications (in 2008 market shares, based on the IDC report [1]) are SAP, Oracle, Siemens and Microsoft. Their solutions are of course prepared for the current major transformation undergone by the software industry: vendors offering infrastructureless applications, IT management, and platforms “as a service”, using their or a partner’s cloud services, will continue to acquire share from vendors that only offer on-premise software licenses. This phenomena will spread to a wider range of capabilities, formerly delivered as on-premise solutions. Software Oriented Architecture (SOA) is an evolving software and systems architectural style that takes advantage of distributed, networked, and modularized capabilities that can be more discretely and flexibly applied to specific requirements. Such capabilities are paving the way for new and innovative solutions to emerge from the IT industry and are also forming a foundation for Software-as-a-Service (SaaS) and cloud-based services, business process management solutions, and the next generation of Web-based systems.

The paper describes the capabilities of a web-based platform aimed to establish a unique solution that provides ERP and business intelligence SaaS.

¹ Eng., Omnis GROUP, Bucharest, Romania, e-mail: serban@omnis.ro

Users may access, over the web, solutions for the management of the following enterprise activities: financial-accounting, procurement, sales and invoicing, fixed assets, human resources, documents workflow, legal reporting to authorities, on top of standardized catalogs of partners, activities and products. ERP operational data can further be analysed using Microsoft SQL Server data mining algorithms and services, implemented under the same concept of web utilities.

The solution is able to draw businesses away from managing their own IT infrastructures and enables them to subscribe to information services, under a concept similar to classic utilities consumption (water, electricity, gas, cable TV), saving time and money, reducing administrative overhead and answering to the lack of IT specialists.

2. ERP and Business Intelligence solutions

Prominent solutions such as Microsoft Dynamics ERP, SAP ERP or Oracle's E-Business Suite carry with them a functional and architectural complexity that make the configuration and implementation process, even in a cloud installation, a very difficult task. If the user further needs to integrate analytics/data mining with business processes, he will have to choose from one of the best tools available in the field: Rapid Miner of Rapid-I GmbH (<http://rapid-i.com/content/view/36/210/>), SAS Enterprise Miner [2], IBM SPSS Modeler [3], MATLAB [4], Weka [5] or Microsoft Excel (<http://www.kdnuggets.com/polls/2010/data-mining-analytics-tools.html>). These tools offer, in general, complete integration with back-end solutions and provide complete functionalities for data warehousing, reporting and analysis tools, models for the best practices, applications for business analysis and administrative resources. The data mining solutions are scalable and adaptable to the changing requirements, including tools and interfaces for extending their content or for integration with third-parties' analysis and reporting tools. They get unified access across heterogeneous data sources, and combine data from different data sources to answer business questions. Most of them are based on standards such as XML, XML for Analysis (XML/A), OLE DB for OLAP, Common Warehouse Metadata Interchange (CWM), business application programming interfaces - Java 2 Platform Enterprise Edition (J2EE) and JDBC, or specific APIs, such as the ABAP programming language (for SAP). In terms of ad hoc reporting and analysis, these solutions provide interactive and intuitive analysis with organized toolbars, charts, context relevant menus, drag-and-drop features, drill, slice and format information based on the user's needs, interrogate data and identify trends based on existing data-mining algorithms.

Nevertheless, there are huge problems to surpass, in terms of identifying the simplest and most efficient integrated solution that would allow users to

generate daily invoices, run payroll and accounting reports, track cash-flows and, optimistically, dig into historical data to identify trends and to make predictions.

3. ISIS Solution Architecture

The work is based on the ISIS.NET solution developed by Omnis Group Bucharest over the last 12 years (first described by the author in the volume “*Enterprise Information Managerial System*”, [6]).

The ISIS.NET solution represents the first and only ERP platform developed in Romania, by local specialists, allowing access to its functionalities from a web browser, to any authorized user connected to the Internet. The solution establishes a central platform that can be used by large numbers of clients to run daily operations, to store and analyze large amounts of historical data.

ISIS.NET is structured on three layers and uses .NET Framework 3.5:

- Presentation services, built using ASP.NET web forms and server controls, XML, Java scripting and web service behaviors;
- Business services, that include managed components developed in C# and unmanaged components written in C++, using ADO.NET for data access;
- Data services, in SQL Server 2008 stored procedures, user defined functions, session context information and XML capabilities.

Fig. 1 shows the configuration of a portal based on the ISIS.NET platform and Microsoft's .NET technology, hosted in a data center. The platform is supported by Microsoft SQL Server 2005/2008 and Windows Web Server 2003/2008 clusters, in a failover architecture. Users can register their companies in the portal and operate the solution in exchange of a monthly tax. There are two catalog databases in the portal: the PARTICIPANTS database, that stores an index of registered companies, and the CONFIGURATIONS database, storing the profiles of the companies, created after registration. For each registered company, three databases are created in the portal: one ISIS_DB database, for the management of ERP specific daily operations, one ISIS_DW data warehouse, periodically loaded with data from ISIS_DB, and one ISIS_AN *Analysis Services* database, used for data mining processes. The solution uses a service-oriented architecture, implementing functionalities that can further be used by other services or applications : catalogs of articles and partners, personnel files, document lists, etc. A registered portal may open a web page that provides access to a wide range of services, according to the selected subscription: from restricted pieces of functionality, such as Human Resources data processing, to a full set of services (from complete ERP functionalities to data warehousing and business intelligence tools).

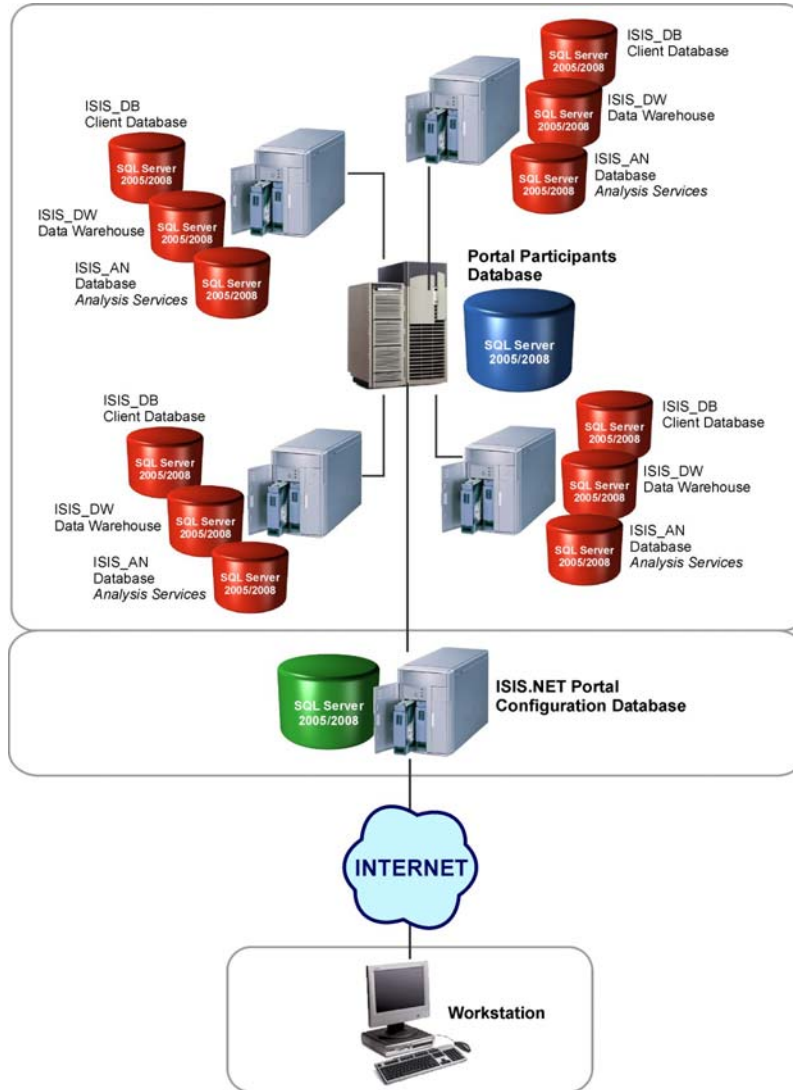


Fig. 1. The ISIS.NET Portal platform

4. Available services and user presentation

The web interface allows access to the entire set of available services. The platform core is represented by the ERP engine, that sustains all operations, on top of the ISIS_DB client database. There are separate solution sections in the interface for each service category, as shown in Fig. 2.

Data sets from ISIS_DB are periodically loaded in the ISIS_DW data warehouse. ISIS_DW can be generated automatically or, at user's request, may be configured to host only domain or department specific data. The data warehouse is used only for OLAP reporting, leaving ISIS_DB available for heavy transactional activities.

The OLAP and data mining services work directly with the ISIS_DW and ISIS_AN database, while all the other services make use of the business layer built around ERP functionalities, as described in [7].

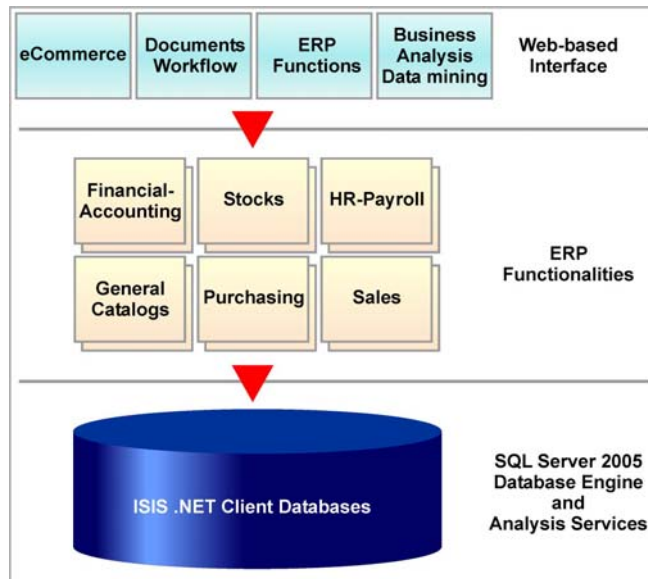


Fig. 2. ISIS.NET solution configuration and available services

Typically, presentation of the results of the data mining process is based on Excel-oriented solutions. Microsoft Excel is the most widely deployed business intelligence tool in the world for the self-service analysis style. SQL Server 2008 data mining add-in for Excel 2007/2010 and the PowerPivot add-in for Excel 2010 build on top of the Excel functionality and add additional features to empower the desktop user for managed self-service business intelligence. In order to present the results on the web, the Excel reports can be published to SharePoint Server and viewed using Excel Services in Microsoft SharePoint Server 2010. Excel Services is a Microsoft SharePoint technology that allows users to publish whole or partial Excel workbooks to SharePoint Server.

The ISIS.NET business intelligence (BI) solution is implemented as a set of web services, but does not use the Microsoft SharePoint technology. ISIS.NET

BI uses custom web controls for implementing cube functionalities and charts. In this manner, the footprint of ISIS.NET BI is low and the response times are reduced, as the solution does not employ SharePoint Server complexity.

5. Business Intelligence and data mining results

Besides a BI solution dashboard, that displays general results at organizational level, the system provides the capabilities to analyze sets of data by applying data mining algorithms. In order to display in web pages the data mining results, a library of data mining web controls was developed. These controls allow to visualize the output of data mining models on any computer that runs Internet Explorer. The library provides a simplified version of the data mining viewers integrated into Microsoft Analysis Services. The data mining web controls function in the web server's memory space. The web server must run Internet Information Services and ASP.NET 2.0 or later versions, and may also host the Microsoft Analysis Services. The controls are implemented in C#, using the client components *Adomd.net* from the namespace *Microsoft.AnalysisServices.AdomdClient*.

The data mining web controls library includes the following controls:

- *DMClusterViewer* – for the clustering model;
- *DMDecisionTreeView* – for the decision tree model;
- *DMNaiveBayesViewer* – for the Naïve Bayes model;
- *DMNeuralNetworkViewer* – for the neural network model.

The class hierarchy implementing the library is represented in Fig. 3.

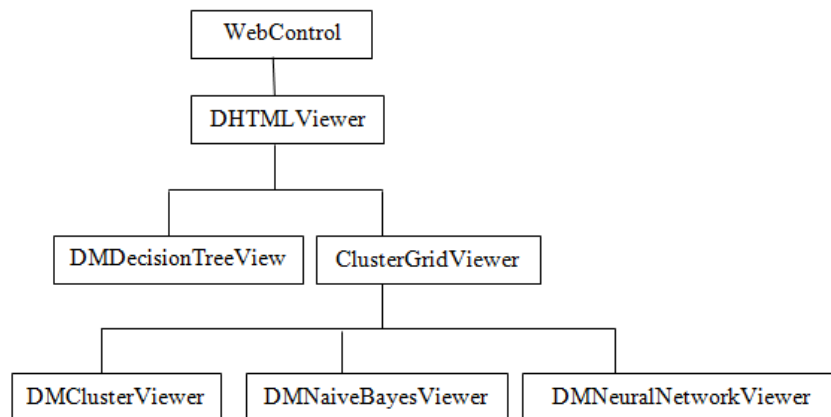


Fig. 3. Class hierarchy for the implementation of data mining web controls library

The data mining web controls are implemented in the *Microsoft.AnalysisServices.DataMiningHtmlViewers.dll* library, that must be copied into the \bin directory of the web site. The following operations have to be performed:

- In the .ASPX page in which the control is included, the reference to the control is registered:

```
<%@ Register
Assembly="Microsoft.AnalysisServices.DataMiningHtmlViewers"
Namespace="Microsoft.AnalysisServices.DataMiningHtmlViewers"
TagPrefix="cc1" %>
```

- The control is added and its properties are set:

```
<cc1:DMNaiveBayesViewer ID="DMNaiveBayesViewer1"
runat="server" ViewType="tree" BorderColor="black"
OverflowBehavior="scroll" BorderStyle="Solid" BorderWidth="1px"
Height="700px" Width="1300px" ForeColor="AntiqueWhite"
EnableTheming="true" ShowToolbar="true" />
```

- In the .ASPX.CS page, on the *Page_Load* event, other settings are established:

```
DMNaiveBayesViewer1.Visible = true;
DMNaiveBayesViewer1.Server = ConfigurationManager.AppSettings
    ["Server_Isis"].ToString();
DMNaiveBayesViewer1.Database = ConfigurationManager.AppSettings
    ["Database_Isis"].ToString();
DMNaiveBayesViewer1.Model = ConfigurationManager.AppSettings
    ["DataMiningModel_Isis"].ToString();
```

- In the *web.config* configuration file, the *Data Mining Content* properties are set:

```
<configuration>
...
<appSettings>
  <add key="Database_Isis" value="IsisWarehouseAnalysis"/>
  <add key="Server_Isis" value="WIN2008_VIRT\WIN2008SQL"/>
  <add key="DataMiningModel_Isis" value="Association_Sales_DT"/>
</appSettings>
```

- A reference to the *AdomdClient* library is added, in order to be able to access specific *ADOMD.net* functionalities:

```
<system.web>
  <compilation debug="true">
    <assemblies>
      <add assembly="Microsoft.AnalysisServices.AdomdClient,
        Version=10.0.0.0, Culture=neutral,
        PublicKeyToken=89845DCD8080CC91"/>
      <add assembly="Microsoft.AnalysisServices,
        Version=10.0.0.0, Culture=neutral,
        PublicKeyToken=89845DCD8080CC91"/>
```

- The user identity information is specified – otherwise, the page will run under the MACHINE\ASP.NET identity, that lacks sufficient permissions to instantiate and call controls:

```
<system.web>
...
  <identity impersonate="true"/>
```

As an example, Fig. 4 shows the results displayed by the *DMNaiveBayesViewer* control, following the application of the Naïve Bayes algorithm (described in [8]) on a data mining model hosted in the ISIS_AN database. The Naïve Bayes algorithm is provided by Microsoft SQL Server - a detailed description of the platform's data mining capabilities is available in [9]. In our example, a hotel manager was interested in identifying the criteria that describe employees that are inclined to benefit from an unacceptable large number of sick leave days. The problem can be reduced to counting significant events, identified as present in a specific situation. These events may be considered as hypothesis of a given assumption: a lot of sick leave days. The algorithm is based on the Bayes' theorem, with strong (naïve) independence assumptions: the presence of a particular feature of a class (an attribute of the employee record) is unrelated to the presence (or absence) of any other feature. In our example, the system analyzes relations between the attributes of the employee records: years in service, age, marital status, number of children, income, occupation, health insurance house. Fig. 4 displays the criteria describing employees who scored a sick leave coefficient greater than a critical value (value 1 = acceptable, value 2 = over the critical value).

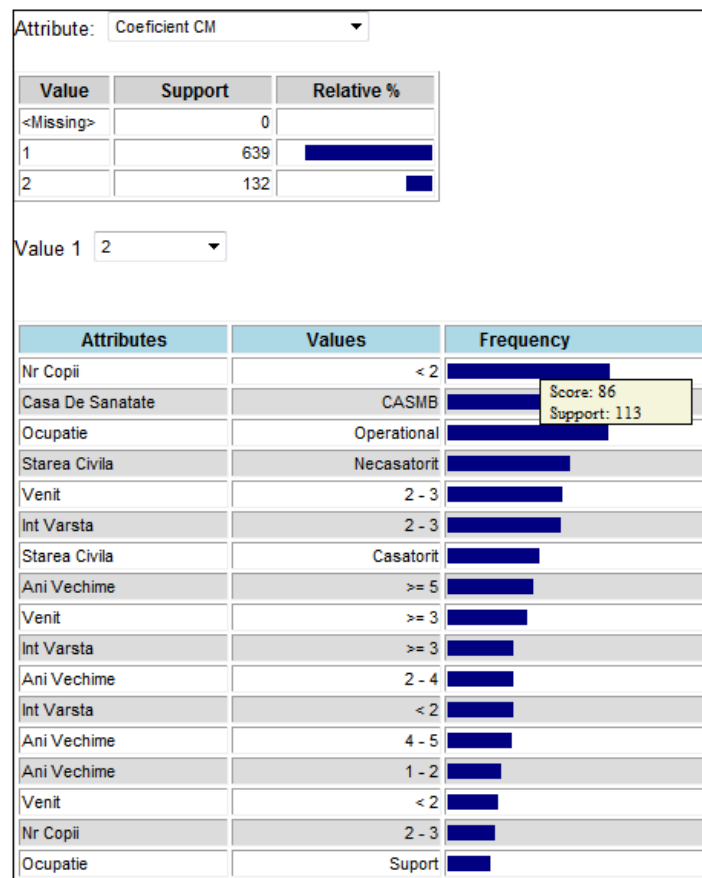


Fig. 4. Sick leaves evaluation based on Naïve Bayes algorithm

6. Conclusions

The ISIS.NET system applies the Microsoft .NET technology to the creation of an original and innovative solution for data management and analysis. The system may be installed on a laptop, for individual use, or can be configured as a SaaS platform, in a data center or in the cloud, to support a large number of companies. As of today, the business intelligence functionalities are implemented as an extension of the ISIS.NET platform. The platform's administrator is required to perform the following important operations, using the regular SQL Server interface:

- Define the integration packages that describe loading of data from ISIS_DB into ISIS_DW, and running the packages;
- Define the mining models in ISIS_AN, on users' request.

The ISIS.NET BI solution uses custom web controls for presenting the data mining analysis results - one control for each type of data mining algorithm - using the new web charting functionality introduced in Visual Studio 2008/2010. The originality of ISIS.NET BI consists in the way it implements the front-end functionality for presenting the business and data mining analysis results on web pages. The plans are to separate this ISIS.NET BI extension and to build a set of tools and web services able to process data received in a standard format (XML, CSV, etc.). The next system version will allow users themselves to perform the above listed actions, directly interacting with web viewer controls. The BI modules will also be able to process and analyze data extracted from ISIS_DW warehouses or from external systems.

REFERENCES

- [1] *A. Pang*, Market Analysis - Worldwide Enterprise Applications 2009 – 2013 Forecast Update and 2008 Vendor Shares, IDC Report, International Data Corporation, 2009
- [2] *SAS Publishing*, Data Mining Using SAS Enterprise Miner: A Case Study Approach, 2 edition, 2009, ISBN-13: 978-1590471906
- [3] *R. Heck, S. Thomas, L. Tabata*, Multilevel and Longitudinal Modeling with IBM SPSS, Routledge Academic, 2010, ISBN-13: 978-1848728639
- [4] *S. Attaway*, Matlab, Second Edition: A Practical Introduction to Programming and Problem Solving, 2 edition, Butterworth-Heinemann, 2011, ISBN-13: 978-0123850812
- [5] *I. Witten, E. Frank, M. Hall*, Data Mining: Practical Machine Learning Tools and Techniques, 3 edition, Morgan-Kaufmann, 2011, ISBN-13: 978-0123748560
- [6] *Șerban Ghenea*, "Sistemul ISIS", in Sistemul Informațional Managerial al Organizației, pp. 453-465, Ed. Economică, București, 2001
- [7] *Ș. Ghenea, C. Oprean*, "Web-based Data Mining Services – A Solution Proposal", in Proceedings of ICSOFT 2007 – 2nd International Conference on Software and Data Technologies, pp. 203-208, Barcelona, 2007
- [8] *T. Bayes*, "An Essay towards solving a Problem in the Doctrine of Chances", in Philosophical Transactions of the Royal Society of London, London, 1763
- [9] *B. Crivat, J. MacLennan, Z. Tang*, Data Mining with Microsoft SQL Server 2008, Wiley Publishing, 2009, ISBN -10: 0470277742