The Municipality of Venice
Data Warehousing Project

SAS European Users Group International
Florence, 29th May – 1st June 2001
Index

- Introduction
- The Municipality of Venice Data Warehousing Plan
- Projects accomplished
- The evolution of the plan
Introduction

Venice Municipality

- 275,368 inhabitants (31st December 2000)

- City Center: 66,386
- Estuary: 32,451
- Mainland: 176,531

- Districts: 12
The Municipality of Venice

- Approximately 3,000 employees
- 1,500 PC in the network of the 2,000 managed
- 75 council buildings cabled out of 150, amounting at 70% of the users
- 80 network servers
- More than 1,300 employees linked-up (electronic mail)
Venis S.p.A.

- Venis is the enterprise of the Telecom Italia Information Technology Group operating in Venice on behalf of the local authorities.

- This company helps the management and the development of the Venice Council IT system.

- The Municipality of Venice is the main relative shareholder of Venis S.p.A.
The relationship between Venis and the Venice council is controlled by the Agreement that has assigned Venis the role of System Integrator.

Migration from Mainframe to Web in 2 years (1999-2001)

Managerial system completely renovated with net computing and Web interfacing.

Demography, Elections, Human Resources Juridical ed Economic, Civil Status, Accounting, Taxes, Technological Services, Public Works, Patrimony, Housing, Call Centers.....
Architecture

- Managerial Services on UNIX and NT servers with Oracle and SQL Server databases

- A SAS Data Warehouse resident on an IBM Netfinity 5000 Server with NT operating system

- Web Servers on last generation computers working on the Linux operating system
Objective

- To complete the Data Warehousing structure in the SAS environment

- To provide access for institutional offices (Intranet)

- To provide access to citizens (Internet)
The Data Warehousing Project for the Venice Council

- Quick and easy access to data specific to various sectors of the Council administration

- Need to keep low link-up costs

- The possibility to interact with data in terms of classification of reports, and data navigation
Data Warehousing Project for the Venice Council

- The valorisation of managerial system data, providing useful information for the administrators

- The classification of a standard instrument, fit for the integration of the various managerial realities

- A Data Warehouse as the exclusive container common to all the information submitted in a coherent and uniform way
The stages of production

- Identify the information to be distributed
- Identify the managerial structures from which the data is extracted
- Loading of the managerial structure into the Data Warehouse
- The preparation of DW data following the requirements expressed during analysis
- Definition of front-end applications for the subject
Web publication reports

- Phase during which the presentation of information is released from Data Warehouse implementation
- The choice of internet/intranet technology solutions
- Definition in agreement with the Council of standard format applications to be published
- Possible modalities of publication:
Publishing

The Web static publishing
The processed information, can be automatically translated onto HTML format and positioned on a web server.

Dynamic Web Page publishing
the possibility to implement a system of dynamic analysis with the possibility of selecting the parameters that allow the user to customise the interrogations following his own requirements
Multidimensional Web Analysis

- The possibility to navigate through data and observe the latter following predefined classification criteria
Projects accomplished or in the phase of realisation

- The Statistical and Demographic Structure
- The visualization of historical data from past elections
The Statistical and Demographic Structure

- Creation of a Data Warehouse composed of statistical data specific to the population resident in the borough of Venice
- The visualization of information in a Web format
- The data coming from a managerial environment is manipulated to adapt suitably for the subsequent stage of visualization
- To have access to reserved information
- Weekly up-grading of DW data
The Statistical and Demographic Structure

- The categorization of data following geographic variables:
  - Zone (mainland, island e town centers)
  - District
  - Locality
  - Street
The Statistical and Demographic Structure

- Categorization of data following individual variables:
  - Age
  - Age Segment
  - Sex
  - Civil Status
The Statistical and Demographic Structure

- For family units:
  - The geographic distribution of families
  - Number of units within families

- Definition of classified levels of navigation:
  - Zone – District – Street
  - Zone – District – Locality
  - Zone – District – Census Area
The visualization of historical data from past elections

- The necessity to gain access to information specific to the historical data of past elections reported to the Municipality of Venice

- The possibility to gain access to the Internet

- The project has been achieved (www.comune.Venezia.it/elezioni2001/) for data specific to 1996 General Election, and the Local Government election in 2000
The visualization of historical data from past elections

- The implementation of data relative to the past elections and referendum of particular local interest
- Processing of data coming from managerial databases
The Visualization of historical data from past elections

- The possibility to view data according to different levels:
  - Borough
  - Zone
  - District
  - Chamber of Deputies
  - Chamber of Senators
  - Electoral Sector
Other subjects of the Municipality of Venice DW

- The Web visualization of supplier data
- The production of managerial reports for the Bookkeeping and Human Resource Departments of the Venice Municipality
The Web visualization of Supplier Data

- The necessity for a supplier to access through the Internet to the data relative to his own invoices and/or mandate for payments

- Security is guaranteed by providing access only to authorized subjects, possessing a valid account.
Production of managerial reports for Bookkeeping and Human Resources

- The necessity of council access to the managerial system for extraction and subsequent processing of the information aimed towards the production of specific reports (spread-sheet, printouts, files).

- To gain access to the managerial database for the extraction and successive processing in function of the specific requirements of the Council Officials.
Production of managerial reports for Bookkeeping and Human Resources

- Visualisation via the Intranet by means of a menu customized on the basis of the User profile, and the specific functions specific to the production of reports.

- The possibility to parameterise data extraction from the DW for the subsequent production of customized reports.
VENIS Ltd

The Municipality of Venice
Data Warehousing
Introduction

The present document illustrates the guidelines that define the accomplishment of the Data Warehousing plan proposed by the Municipality of Venice. The aim of this project is towards the creation of a technological infrastructure based on a SAS solution, able to provide the solutions necessary to enable administration offices engaged (via the Intranet) and, local citizens (in the Internet) to gain dynamic access to the information available by means of standard interfaces and applications.

1 Venis Ltd

Venis Ltd., the Venice I. and T. company is a part of Telecom Italia Group, established in May 1989, and having amongst others the Venice Council as a partner, fulfils the major part of its activities for this municipality, which has assigned Venis Ltd with the production, development and technical management of the Local Council Information Technology System.

Venis Ltd. is the technological partner of the Venice Municipality for Information Technology, and in this role is engaged towards the development and renovation of the towns IT structure. Apart from this Venis Ltd. Collaborates with the Venice Municipality by planning and implementing of IT solutions that enable the local administration to work at its best while, contemporarily conveying the Council as a reliable interlocutor for all the local citizens.

In these last two years, in concomitance with the problem relevant to the Year 2000 of providing the Local Council Administration, with an ever more innovative and efficient instrument, also in terms of flexibility, Venis Ltd. has foreseen toward the migration of the principle IT applications from a mainframe ambient, with an MVS/ESA operating system, towards a local environment based on a number of UNIX servers run on Oracle databases.

Other applications, specifically developed for the Venice Municipality, are added to these databases, in the Windows NT System and, in some cases, with SQLServer databases.

Within its role as an integrator of the various systems, Venis Ltd. has foreseen to create the actual departmental processes as being compatible with each other and capable, of providing the practical qualities in any case necessary towards the perfect running of town council administration.
2 Concepts of the Data Warehousing

The process towards the implementation of Data Warehousing, is that which creates an architecture aimed to the distribution of information. For a state-owned enterprise this implies on one hand, the potential to supply IT employees with a means of sharing technologically advanced information, and therefore improve the quality and the quantity, of work carried out. On the other hand, this process provides a method of easily and efficiently reaching the citizen/customer, ever more the focal point of the divulgation of data.

The technological principle and structural characteristics of Data Warehousing are as follows:

- A Data Warehouse is typically constructed around subjects; these subjects are of particular importance, as for example taxes or, the composition of the population, that are defined by the level and the nature of the information requested by the user;
- A Data Warehouse is logically and physically distinct from the system that runs the managerial processes, and does not interfere with the latter;
- A Data Warehouse is coherent and integrated; all the information is integrated to define the subject, therefore the eventual discrepancies in data deriving from the managerial system are corrected before being loaded;
- A Data Warehouse is not volatile and supplies a unique “fountain of truth”. This means that the moment in which two people request the same information, they will obtain the same answer, and the data will not be modified during the phase of analysis;
- Data Warehousing provides a historical prospective of the various subjects. The data is usually archived in a concise form for a few years, and this grants a possibility of gaining access towards trends, for statistic and forecasting analysis.

The different realities in the private world as in the public one, have been archived in huge moles of data for many years, but the partial exploitation of their potential has always represented an unresolved problem. The extraction of information generally implies, the creation of a specific programme for extracting data from the managerial system, and the execution of the subsequently complicated steps required to attain to the information. A similar request repeated at different intervals called for in any case the re-extraction of data and the execution of the
successive processing steps. The growing vorticose of recent years, and the demand for information, has created an unsustainable burden of work for the structure assigned to supply information.

As a result, the concept of Data Warehousing was born, that is, from a structure of data designed to reply in a rapid and efficient manner to several of the users requests with added value, described by the following characteristics: standardization, rapid response and flexibility.

Another strong feature of Data Warehousing is that it simplifies ad hoc, queries or, the exploration of data not previously anticipated. The users, depending on the level of autonomy allowed to them by the system administrator, have the independence to analyse in different ways, their own data rather that being limited to a predefined ensemble of analysis.

3 The Data Warehousing Project and the Venice Municipality

An aspect of the present day needs specific to the Venice Municipality, is that of providing the conditions for internal staff (of the Council and/or Venis Ltd.) to access to specialised information of the various sectors, in a simple and quick way. Another aspect is that distributing to the public by means of the Internet, a flow of information otherwise difficult to reach or of time-consuming retrieval. The aim is to make external users proficient at visualizing, by means of a browser, a series of predefined reports and statistics, or to allow adept employees to navigate through the different levels of information following the various hierarchy of prospective.

It is likewise a shared opinion of the prerequisite to define a standardized instrument fit, to be positioned as the subject integrator, of the various managerial realities at currently present in the IT framework of the Venice Municipality, but distributed on separate systems.

Such needs cannot be fulfilled by approaching the various subjects individually, this would only result in eventually having to repeat the same, or similar, extraction processes on the same databases, and run the not irrelevant risk of inserting into the various phases of preparation, different criteria of selection, validation and/or aggregation, hence obtaining results that are not coherent.

An alternative, the only one fit to deal globally with this problem, would be to plan the creation and growth of a Data Warehouse destined to be the sole commune container of all the information that has been prepared, in a coherent and uniform way specific to the various subjects involved.
A generalized approach used for the implementation of every subject can be described as follows:

- The identify of the key information to be presented;
- The identify of the managerial structures from which to extract the necessary data;
- The loading of the managerial structure into the Data Warehouse;
- The preparation of Data Warehouse information, depending on analysis requirements;
- The definition of the output/outputs of analysis, (client server o Web) according to the subject.

The technical particulars of the stages described above will looked at in detail in the following chapter.

The activities accomplished until now are yet in a stage of execution. The first subjects identified are the following:

- A Statistic Service: For this kind of service an initial Web analysis application based on data of the population at present resident in the municipality. Such data is visible in terms of residents and family units, grouped together in the basis of various classifications. For example: area, sex and age group.

- A Polling Service: An analysis application of the historical results relevant to the Venice municipality due to be released, is to provide also in view of the electorate vote of March 13th information relative towards recent trends.

- An Accounting (Expenses) Service: This service anticipates the execution of a function which will allow local citizens in their role of suppliers of goods or services, to have access to information concerning the payment of their own notariale expenses.

- A Revenue Accounting Service: This organization will provide the realization of a monitoring system specific to the collection of council tax, eventually to be administered by the office concerned.

- A Managerial and Administration Service: This service proposes a phase of integration between the current managerial applications concerning town council accountancy (bookkeeping) and the monitoring systems of PDO targets.
4 The Technical Stages of the Project.

As previously mentioned, for every subject it is possible to define a generalised approach, which goes through various standard stages.

4.1 The access to managerial data.
The initial phase towards Data Warehouse implementation is that relative to the extraction of managerial data. This can be achieved owing to the functional characteristics of SAS applications, which provides easy access to the various structures involved, as if these were effectively SAS structures. The characteristics of the SAS engine, makes the process of data extraction easy and rapid. It is also possible to exploit all other eventually foreseen intermediary structures, (for example Oracle) previously created for analysing purposes.

4.2 The maintenance and management of a Data Warehouse
Through SAS/Warehouse Administrator - the specific SAS product for the maintenance and management of a Data Warehouse - it is possible for the end-user to control and administer the Data Warehouse in terms of framework and workflow.

In this way the overall architecture is centrally controlled, thus attention can be concentrated on analysis and planning, considering that as a result of these factors the technical area results less complicated.

In the specific context concerning Data Warehousing, for the Venice Council, SAS/Warehouse Administrator will control from an end-user terminal assigned to the administration department, the entire loading and processing course to be carried out on a selected NT server.

4.3 The Web Publishing of Statistical and Dynamic Reports
An important phase, if not the most important is that concerning the submitance of data, as Data Warehousing implementation does not exclude before hand, any form of front-end.

A proposed solution for the case in question is that based on Internet/Intranet technology. The advantages of which can be considered in the terms of easy access by the part of the outside end-user (the Internet), a reduction of link-up costs, in that every user to gain access must have at his disposal a public browser, and facility towards the creation of applications which consequently means maintaining costs (no client/server application needed).
Therefore it should be emphasised how the proposed solution can guarantee a reduction of development and management costs in terms of hardware and software, in alternative to a Client Software Solution. Also the fact that the Data Warehouse is composed of data that is to be physically found on an assigned framework, can guarantee a universal quality performance for managerial systems.

In a field such as that of Web publications on behalf of the Venice municipality, it is extremely important that the information is submitted not only in a formally correct manner, but that it is organised following a standard, simple, uniform and engaging format. For this reason it is important for Venice Ltd. and the Council, to define in a uniformed way a standard layout of the applications destined for publication.

The primary level of the submittance of the information produced in the Data Warehouse, is that of the so-called statical web publications. The processed information, from data analysis can be automatically translated onto HTML format and positioned on an Internet or Intranet web server.

This solution is without a doubt satisfactory for standard distribution, but in an environment that requires in-depth study, will result as being too restrictive. The moment in which Intranet/Internet users are allowed to build their own customized reports, it will prove necessary to implement a dynamic assessment system. This method of analysis can be achieved by planning from the selection page, a series of parameters that will allow users to parameterize the interrogation according to his requirements. To accomplish this and, rapidly receive constructed (i.e. dynamic) feedback.
Server SAS

Posizionamento
batch
Pagine html

web server

1: richiesta

2: invio

web browser

2: richiesta
elaborazione

Server SAS

3: elaborazione

web server

1: richiesta
report

4: invio pagina

web browser

figure 7.1

figure 7.2
4.4 Multidimensional Analysis by Web

The most advanced level of analysis by web, of data warehousing information, is that in which the user is able to conduct an in-depth study of the aspects related to the data. This implies, for example the possibility to carry out an analysis starting from the aggregated data specific to a town, and to descend towards the detailed analysis of a district, followed by the street and the street number, a particular indicator. For example in the research of a detail relative to a particular anomaly that has been found.

This level evidently presupposes an adept interpretation of data analysis, and is therefore considered to be used by an users.

5 The Primary subjects of the Venice Municipality DW

The subjects so far identified, and for which, has already begun the activity of Data Warehousing implementation, are those already described in the previous chapters. To accomplish the total evolution of the process, the most convenient choice is to observe the complexity of the numerous organizations now existing, that compose the organization of the Venice Council as represented in figure. 7.1.
figura 7.3
5.1 The Statistical and Demographic Structure

This project in phase of execution, foresees the creation of a Data Warehouse composed of statistical and demographic data relative to the population who reside within the city of Venice, and the subsequent visualisation of these in a Web environment.

These functions based on data, source of the managerial organization, must necessarily include a preliminary stage of manipulation, in order to adapt suitably to the successive stages of visualization.

The statistical and demographic structure is developed in a Web environment, maintaining the possibility to be only partially shown, in order to avoid violating the citizen's rights of privacy. The remaining, not authentically statistical data can in any case be read, always by means of IT administration staff Web technology by the establishment workforce. These employees for institutional reasons, have the necessity to retrieve this data.

It should be emphasized how the choice of the method previously described, is determined by the necessity to maintain development application costs. Especially if paragoned to the vast basis of users.

The Data Warehouse of particular interest in the demographic and statistical area foresees the revision roughly on a weekly basis. This does not mean however that for specific needs, the updating phase cannot be scheduled more or less frequently, until reaching a daily refresh of the DW.

For that which concerns the resident population, the data is submitted by variables of classification that are either geographical:

Zone (Mainland, Island and Town Centers)
District
Locality'
Census Areas
Streets

Or directly linked to individuals

Age
Age Segment
Sex
Civil Status

Or else directed towards family units
The geographic collocation of the number of families
Large families (families composed of 1 to 6 individuals)

For the typology of analysis that necessitates, it is also foreseen the possibility to define the hierarchy levels through which the user is allowed to navigate directly by means of a drop-down menu:

Zone - District - Street
Zone - District - Locality
Zone - District - Areas out of Bounds

The possibility of effectuating multiple selections of classified variables is also anticipated
Analogical criteria of classification are used for data analysis specific to city population trends: births, deaths immigration, emigration and domestic swings, i.e. movement, specific to the changing of residence within the town itself.

5.2 The visualization of historical data from past elections.

This project has been achieved, and the relative data specific to the 1996 General Election, and the Local Government election in 2000, is at present available on the Internet.

A successive implementation is foreseen with data deriving from other elections including, the Referendum, which is of particular interest.

The difficulty in disponing of an instrument that facilitates access to past election results, has always been a principal setback for the City council Electoral Office.

More and more frequent are the requests from: political parties, public officials, universities, institutions or research agencies to the individual citizens, who demand not only the data specific to their own requirements, but also the processing of this data form in every possible perspective.

As a result it is therefore evident how it has been necessary to provide a final answer to satisfy these requests. To ensure that the retrieval of such information is as rapid and complete as possible, and that the instrument used to gain access to the data is user-friendly.

It should also be considered that in this case, as in other realities, the Municipality of Venice and the election results, originate from a managerial environment. The characteristics of which are orientated towards the aspect of acquisition towards this establishment. For the same reason the processing procedures were finalized towards this managerial organization. For that which concerns the representation
in a Web environment, the problem of visualization in this particular case required the preliminary processing of the data; therefore a phase of implementation of the Venice Municipality Data Warehouse was obligatory. Precisely, a Data mart containing data and election results.

Facing every election, and within the latter for typology are processed, (local-government elections, proportional board etc.) the data originating from the managerial system is processed and various charts containing the data itself organized and grouped together for constituency, that in any case represents the minimum grouping that guarantees every other type of superior grouping.

For that which concerns the visualization of the data, it can be a hypothesis towards the creation of a series of functions that allow the representation of the information at various classified levels:

- Town Council
- Zone
- District
- Electoral Sector

Or else:

- Town Council
- Zone
- Constituency
- District
- Electoral Sector

Or

- Town Council
- Zone
It should be remembered that these typologies of aggregation/representation of data, are independent from the specific type of election, and as such it is possible for example that the results of the local government areas are visualised by constituency of the Chamber of Deputies.

Finally it is to be considered that access to such functions take place by means of a link to the site of the Municipality of Venice.

6 Other subjects of the Municipality of Venice DW
These realities currently in the planning phase together with the Venice Council concern in particular the possibility of reaching invoice data by Web, and the execution of managerial reports for domestic use by the IT administration employees.

6.1 The Web visualization of Supplier Data
The necessity to visualise the data relative to the suppliers of the Venice City Council, arises in the moment that a citizen in the ambit of the transparency process, that must characterize the Public Administration, requests information concerning his rapport with the Public Institutions.

In particular the competent organ of the Venice City Council, has the necessity to allow a supplier of goods to this institution, to verify in real time the state of his rapport with this Institution, in terms of mandate for the payment of invoices.

It is evident that these functions find their natural development in the Internet environment, which provides the possibility for anybody previously authorised, to gain direct Web access to such information. Consequently providing the Council with considerable advantages in terms of lower expenditure.

The proposal of Venis Ltd. consists in the development of a project, appropriate towards this objective, which backed by SAS technology, is able to gain access to this information.

This project which ascribes to SAS Technology for the development of Data warehousing applications, can be articulated into two distinct phases:
- Daily data extraction from managerial databases into Dw architecture and the utilization of the Data Warehouse, which is physically resident on a processor assigned as a support towards the memorization of data, previously attained from the managerial database and consecutive processes. This phase could configure at daily intervals, and would free the managerial servers from possible requests deriving from the consequent maintenance of the performing managerial systems.

- The possibility for every authorized user to gain access through the Internet to the information and only that of his own competence. In this way the user can verify his situation, with the only obligation of possessing the disponibility of a link-up to the Internet.

Data access security is guaranteed by providing access only to authorized subjects, who posses a valid account.

6.2 Production of managerial reports for Bookkeeping and Human Resources

As previously specified, the phase of integration of the new departmental organizations, and the continuous requests from the offices involved, has bought about the demand for this function, by council officials, in charge of functions aimed towards the production of specific reports (spread-sheet, printouts, files).

This implies a primary phase of back-office with the implementation, of the Municipality Data Warehouse concerning information coming from the two sectors in question.

In particular we plan to use the managerial database, with a frequency to be accorded with the final user. To the processing of the information in function of those that are the specific needs that gradually emerge, and finally towards the memorization of such information stored within the Data Warehouse of the institution.

The successive front-end phase foresees the visualization of data according to user's needs. These functions may be explicate as has been done up to this point, directly from Venis Ltd. employees facing the specific demand of the council officials, or else by means of the visualization of data by Web, on an Internet framework, with the opportune criteria of security to avoid undesired intrusions.

This last hypothesis results without a doubt the most desirable, as it puts the Municipality in the position of being totally self sufficient, while contemporarily frees up Venis Ltd. personnel of continuous and repetitive requests.
It is envisaged the creation of opportune customized menus in base of the profile of the user, characterized conjointly with those in charge, in the aim of being able to guarantee that eventual perceptible data is visualised only by authorised personnel.

By choosing the item of interest and eventually specifying the opportune selection variable, one can gain access to the Data Warehouse, and a Web page can be visualised showing the data of interest.

It should be considered that over a certain period of time this type of project requires successive implementation that can be finalised only on the basis of the necessities that gradually emerge from council officials.