INDIAN LANGUAGE LOCALIZATION OF ORACLE - Server & Tools

Shashank Shekhar
Senior Development Manager,
Oracle India Private Limited, Bangalore, India

Contents

Introduction

With the changing rules of business and enhanced role of Internet in potentially, boosting the top line of the business, it is of paramount importance that businesses revamp their storefront to cater to global customers. At the very least, visitors to their online store, should be able to see the content in their own language, with the content dynamically transforming itself to cater to his own territory definitions, example currency sign, date format etc. Oracle products have always been designed with this very underlying principle, so that when one uses Oracle one doesn't have to worry about how to make their business global. Globalization of software ingrains localization. And we fructified this idea by enabling Oracle products to handle Indian Territory and Language definitions, so that your customers in India could use their own native language while doing business with you. Since, we have used Unicode as base for storing all your data, it will truly be global.

While urban India is predominantly English speaking, the vast majority of Indian population who live in district towns and countryside are not. Indian constitution provides for 14 specified languages for Government work, English being one of them. There are 10 languages, each used by minimum 2.5% of population while Hindi is used by 40% of the population of the country.

There had been initiatives in each state government's to implement the state's native language for official communications. In the Hindi belt of North India, majority of states insists on Hindi or their respective state language to be used for transactions with the government. Property registrations, land records are predominantly made in these languages in rural India .So are the Tax notices, Government gazettes etc. Ability to provide utilities bills by government held institution in the native language is growing in demand. Education has always been delivered and best understood in one's mother tongue. This is more important when offering adult education at the district and village levels.

The goal of this project was to enable Oracle Server and Tools for multiple Indian languages. We aim that user be able to not only enter the data in Hindi or Tamil or any other Indian language and English, but explore the data using various other database operations as supported by oracle and its tools. A primary design goal of this product was to enable Oracle Server, Tools and Application to handle database operations, software development and deployment using Indian languages. It should enable developers to store data and write software that can be deployed easily across a variety of world markets. These markets use very different languages, written in a variety of writing systems. To be acceptable in these markets, you must be able to present text in a way that conforms to the rules of their writing systems.

The Indian language support is available in Oracle9i server and Oracle 9i application server.

Why Localize?

The benefits of localization are

- The target audience for your application increases
- Develop once use anywhere
- Easier Web deployment

Target Users

Government - Central and State governments for e-Governance applications, Government departments, PSU's, Educational and Research Institutions etc.

Web Sites - Portals wanting to develop Indian language channels, Newspapers, magazines hosting online editions, Govt/ PSU's/ educational/ Institutional web sites etc.

Endless Possibilities

Some of the applications for which Oracle's solution could be used are:

e-governance applications
Customer Care & Billing
Census records

Land Records
Generic Records
Kiosks
Web Portal and Web Sites
Public Information Systems
Email
Chat
Online forms
Information archives
Wireless messaging
Data warehousing and data mining
Document publishing and printing
Web search engines
Business Applications

ORACLE SERVER

Oracle offers a unique capability that is unparalleled and solves a long-standing problem of true data processing capability in Indian languages. The solution is Unicode compliant and can be used with Oracle 9i RDBMS for data entry, indexing, searching, sorting and generating reports. The data is universally portable and can be displayed using any Unicode compliant fonts and application software. Data and text produced using Oracle's solution can be converted to formats such as XML, HTML and WML and hence can be re-purposed for use in any environment. Since it uses Java as the backbone, applications developed using Oracle's technology can be deployed on any hardware/OS platform and the data produced by such applications will be universally exchangeable and portable.

Oracle server 8.1.5 was changed to handle Hindi and Tamil characters. Oracle Server's National Language Support (NLS) architecture allows you to store, process, and retrieve data in native languages. It ensures that database utilities and error messages, sort order, date, time, monetary, numeric, and calendar conventions automatically adapt to the native language and locale.

Oracle RDBMS is implemented using three-tier architecture. The language-dependent operations are controlled by a number of parameters and environment variables on the client, the middle tier and the server. On the server, each session started on behalf of a client may run in the same or different locale, and have the same or different language requirements specified.

A database itself also has a set of session-independent NLS parameters specified at its creation time. Two of them are the database and the national character set. They specify the character set used to store data in the database. Other parameters, like language and territory, are used in the evaluation of CHECK constraints.

In the event that the client and server specify different character sets, Oracle9i will handle character set conversion of strings automatically.

As far as NLS (National Language Support) architecture is concerned, all applications, even those running on the same physical machine as the Oracle instance, are considered clients. For example, SQL*Plus started by the Unix user which owns Oracle software, from the Oracle Home in which RDBMS software is installed, and connecting to the database through an adapter by specifying the ORACLE_SID, is still considered a client and its behavior is ruled by client-side NLS parameters.

When a client application is started, it initializes its client NLS environment from environment settings. All NLS operations performed locally are executed using these settings.

When the application connects to a database, a session is created on the server. The new session initializes its NLS environment from NLS instance parameters specified in the initialization parameter file. These settings can be subsequently changed by an ALTER SESSION statement. The statement changes the session NLS environment only. It does not change the local client NLS environment. The session NLS settings are used in the processing of SQL and PL/SQL statements executed on the server.

LANGUAGE SUPPORT

Oracle9i allows users to store, process, and retrieve data in Hindi, Tamil, Kannada, Malayalam, Gujarati, Oriya, Punjabi, Telugu, Assamese and Marathi. Through the Unicode (UTF8) character set.

Additional support is available for a subset of the native languages, for which Oracle9i knows, for example, how to display dates using translated month names or how to sort text data according to cultural conventions.

TERRITORY SUPPORT

Oracle9i supports different cultural conventions, which are specific to a given geographical location. Local time, date, numeric and monetary conventions are handled. From Oracle 8.1.5 onwards, Oracle supports “India” territory.

DATE AND TIME FORMATS

The world's various conventions for hour, day, month, and year can be handled in Indian language formats.

MONETARY AND NUMERIC FORMATS

Currency, credit, and debit symbols can be represented in Indian language format. Radix symbols and thousands separators can be defined by locales (Hindi_India.UTF8 in this case).

CALENDARS

Gregorian, Japanese Imperial, ROC Official, Thai Buddha, Persian, English Hijrah, and Arabic Hijrah are supported. The one used by the Indian locale is Gregorian.

LINGUISTIC SUPPORT

Oracle9i provides linguistic definitions for culturally accurate sorting and case conversion of Hindi, Tamil, Kannada, Malayalam, Gujarati, Oriya, Ppunjabi, Telugu, Assamese and Marathi languages.

CHARACTER SET SUPPORT

Oracle supports a large number of single-byte, multi-byte, and fixed-width encoding schemes, which are based on national, international, and vendor-specific standards. The character set for Indian languages is UTF8.
You can store Indian language specific data in the database in Unicode format. Oracle server has been changed to handle Indian language characters like month names, day names, AM/PM/AD/BC string. You can use these features to store data in any of these languages.

Why was the SERVER Localized?

DATA MIGRATION

If you want to use these Indian language localization suite you need to have the database with UTF8 character set .If you haven't customized it to the UTF8 character set until now and you have existing tables, you can convert your existing tables to the UTF8 character set, dynamically.

Oracle Tools

The features that are provided by the localization of the oracle tools are

- A set of pluggable java components, which could be used to develop your applications in Indian languages using Oracle tools.
- With the help of these tool you could store your data in the database in the language of your choice
- The data stored will be unicode data so you can perform all the operations you would normally do in english data like sorting, indexing, querying etc.
- The Keyboard layout and rendering are as per the Inscript's format.
- A Java Input Method Engine for Indian language is also provided that could be used with JDK 1.3.x to input data in any of Java applications.

· Unicode fonts pertaining to all the supported languages could be procured from our font vendors

ABOUT PJC

Oracle tools support Java client side user extensions through the use of standard JavaBeans and specialised Java components called Pluggable Java Components (PJCs). Using PJCs and JavaBeans it is possible for application developers to insert custom code to run inside the Forms applet and to tailor or customize the default operation of the default Forms Java components.

ABOUT JIME

Input methods are software components that let the user enter text in ways other than simple typing on a keyboard. The input method framework in the Java 2 platform enables the collaboration between text editing components and input methods in entering text. Its service provider interface (SPI) for input method engines provides interfaces that enable the development of input methods in the Java programming language that can be used with any Java runtime environment; it also supports native input methods of the host platform. The Java Input Method Engine is developed for Indian language.

Conclusion

Oracle offers a completely portable, maintainable and scalable solution that caters to any business requirement for Indian market. The primary design goal of this project was to enable Oracle Server, Tools and Application to handle software development and deployment using Indian languages. It enables developers to write software that can be deployed easily across a variety of world markets. This suite of tools will aid in storing, retrieving and processing Indian Language Data and also provide a means to input data in their native language in a way that conforms to the rules of their languages. The database is able to store multi-lingual data, which is universally exchangeable and portable.