TECHNICAL OVERVIEW

single view eXchange

Single View Exchange

Single View eXchange (*SVX*) is a solution framework that allows you to fully integrate disparate customer information from your operational systems into a single cohesive customer identifier ready for use throughout the enterprise.

Using precise matching processes and algorithms, *SVX* successfully provides the basis for customer-orientated business systems, enabling better business decisions and driving customer-centric processes. *SVX* presents the customer as a total enterprise valued customer – not a specific divisional number.

SVX Structure

SVX is built with the key criteria of speed, efficiency and flexibility at the forefront of the product development processes. *SVX* is a software framework. We say a framework because it is highly customisable and operates in conjunction with existing processes at your site.

SVX is a collection of UNIX and Perl language constructs, together with customised consulting and implementation services that define processes for data matching and data handling. Operated together they provide the basis of *SVX* processing.

The operation of *SVX* is straightforward and has the ability to be run on demand or at any time interval that is suitable.

It is important to understand that *SVX* does not actually perform the extraction of data – it expects records for processing to be available in a formatted flat file for use. This approach was taken as the prime function of *SVX* is to match and consolidate multiple customer records – representing the same physical entity – to a single logical entity – and not to extract data.

SVX provides the user with the ability to fully customize how they wish to treat records during the processing and how intense they want the record comparisons to be. Users have almost an unlimited ability to compare multiple records to ensure a high degree of matching is done. Also, they have the ability to apply certain business rules defined within *SVX*, which are parsed at processing time. These functions are performed via the *configuration file*, which we will discuss later.

SVX – How is it Executed?

SVX is executed by a series of scripts that take the input files, process them according to the content of the configuration file, and output to a pre-defined output file. The execution of the scripts may occur at anytime that is suitable. Additionally, the scheduler may be used to automate this process.

To gain a basic understanding of the elegance of the *SVX* solution, we will discuss the main component of the solution – the Framework, which governs most of the engine and logic behaviour. In addition we discuss the Process flow depicting overall movement of data through the solution. Together, they form the integral part of the solution.

SVX Framework Architecture

Relevant to the business is the ability for the framework to operate in a business sensitive time frame and perform important matching processes at the best possible speed while incorporating the *SVX* defined business rules.

The *SVX* framework has been designed to dynamically handle these requirements as well as the power to:

Adapt to Changing Business Needs - The way in which *SVX* operates can be customized to suit the organisational requirements by using its configuration files for both the selection criteria as well as setting confidence levels for the resulting output. Further flexibility is available through the process itself, which enables the user to optionally review and adjust the results.

Tolerate changes over time in data sources - Data sources may change according to business requirements and/or business growth. The *SVX* architecture and implementation is flexible enough to handle ongoing changes of data sources.

Also, it may take any data source from spreadsheets, external commercially available data, to any operational system data that may be extracted to a flat file. It may then be matched with existing data.

Because of the changeable character of data sources, *SVX* is kept separated from the 'TL' part of the ETL process. Extraction, Transformation and Loading processes must provide a consistent outcome according to a constant definition defined during the implementation.

The ETL process and structures are determined by the data structures and model. The *SVX* definition, however, may change to reflect business changes.

Data Matching - *SVX* may be used for the process of matching customer, data records together to identify commonalities or exceptions. This given ability to perform exception and commonality reporting lends itself to be very powerful in quickly ascertaining data matches.

It is also extremely useful when you have the need to refresh publicly available data like credit rating, census, demographics, psychographics and other external data into your operational systems.

Multiple Deployment Options - *SVX* has been written in Perl so that it can be run on many popular operating systems including OpenVMS, UNIX, Microsoft Windows NT and Microsoft Win2000

High Speed Processing - *SVX* has been designed to be a fast and compact business solution. *SVX* can be configured to perform parallel processing of tasks (where supported by the hardware). The parallel processing is controlled by the Scheduler, which monitors the overall process and progress of each task.

Delta Awareness - *SVX* can be configured to be delta aware so that it only processes new data and/or changes to existing data rather than reprocessing the entire extracted file. In addition to minimizing error, this also greatly assists in the organisation's management and review of the output.

The 'delta awareness' is handled by referring to the History File, which operates like a master-file.

Inputs/Outputs Readily Imported/Exported - The use of ASCII text files for inputs and outputs ensure that *SVX* can be readily integrated with external systems. The use of flat files and internally managed hash tables for the actual processing also ensures that *SVX* is RDBMS independent.

SVX Key Components

The *SVX* framework consists of the following components:

- Configuration File
- Data Transformation Engine
- Data Matching Engine
- Scheduler

Configuration File

The Configuration File establishes the rules under which the other components will operate. It is set up in consultation with the user so that it reflects the organisational priorities and draws upon the on-site expertise. It contains many attributes that define how *SVX* works. These attributes are stored in defined segments that are described below.

Domains

SVX "domains" are groupings of "related" fields.
For example: all the telephone numbers that a person has – home phone, work phone, mobile etc are grouped into one domain.

Information within a domain will be compared and analysed to identify "likely matches" between customer records.

Domains are derived from the SVX_Order column. Those fields assigned a number within the same range of 100 will be grouped.

For example, assigning the last name, middle name and first name the values of 101,120,130 respectively will result in these fields belonging to the same domain, Domain 1, and assigning the various telephone numbers the values of 510, 520 and 530 will result in these belonging to a separate domain, Domain 5.

Note that the SVX_Order column also has certain values (i.e. those beginning with "9") that are reserved.

The value of being able to combine fields to make one domain is that concise matching can be achieved while eliminating extraneous matches.

For example, the typical use for this feature is the combining of Last_Name, Middle_Name and First_Name into one "name" domain. Matching on Last_Name alone would produce large numbers or erroneous matches for those customers that share the same Last_name (e.g., "Smith" or "Purcell") but who are not the same individuals.

Ordering and Weighting

The ability to order and thereby weight the various domains is one of the key aspects of configuring *SVX* to suit the customer's requirements.

Rank Order

The importance/relevance of fields within the *SVX* analysis is implied via the rank order of the SVX_Order column.

Those fields that are best likely to assist in identifying similar customers are defined first. Typically this would be the customer's last name.

SVX Weightings

SVX automatically assigns an internal "weighting" to each domain progressing from the lowest ranked to the highest ranked. Thus the fields deemed most important or relevant will attract the highest weighting and matches for these fields will be considered more significant than for matches of lower weighting.

Weightings are assigned using a Base 2 scale with the lowest being 1 and the highest being 128 when 8 domains are used, for example.

Parameters

Parameters may be defined against a source field to instruct *SVX* as follows:

- To include or exclude the record for analysis
- How to transform the data for analysis
- That a particular field is a primary or foreign key.

Setting Parameters for Conditional Processing

Before *SVX* begins to analyse the data, parameters for the inclusion or rejection of records can be set.
For example, if National Customers are indicated by a Customer type of 'N' and these are not required for analysis as there is a high level of confidence with National Customer data then all data where Customer Type = 'N, can be excluded from the processing.

Transformation

In order to maximize its efficiency, one of the layers in the *SVX* Framework undertakes the transformation of the data to be analysed. The nature of these transformations is also set in the Configuration File.

Data Transformation Engine

The Data Transformation Engine prepares the extracted data for analysis. It does this by first further extracting the records that match the conditional processing parameters set in the Configuration File and by then applying the data transformations necessary for analysis.

Conditional Processing

By being able to conditionally limit the data to be analysed *SVX* ensures that only the relevant data is considered. In doing this it can reduce processing time and increase the usefulness of the output.

Data Transformation

Data transformation is necessary to firstly "standardize" data formats to enable meaningful comparisons to be made and secondly, to make modifications to the data, which in effect "standardize" the data entries, so that matches are not incorrectly rejected. *For example, all double letters are reduced to single letters so that 'Twilley' becomes 'Twiley'. This reduces the chance that the record will incorrectly fail to be matched with another record that was actually entered as 'Twiley' through a typographical or data entry error.*

These transformations include:

- Modifying date fields
- Modifying telephone numbers
- Modifying name fields
- Assigning Soundex values to fields

Data Matching Engine

The Data Matching Engine compares the modified data in each domain, assigns scores based on the Configuration File and then groups the results based on threshold settings.

Scoring

SVX operates by analysing each domain and comparing rows in pairs and checking whether there are matches. An *SVX* score is maintained for every successful match per domain. The score assigned for each domain-match corresponds to relevant *SVX* weighting for the domain as specified in the Configuration File.

For example, if internally generated SVX row_ids 1 and 5 share a common date of birth, the weighting for this domain is 8 and a score of 8 would be added to the matched pair's running total.
No match results in a *SVX* score of 0.

Scoring is additive.
For example, should the same row_id 1 and row_id 5 match in each of the specified 6 domains specified in the Configuration File, then the maximum SVX Score for rows 1 and 5 would be 63 (i.e. 1+2+4+8+16+32).

Thresholds.
One of the key features of *SVX* is the ability to set thresholds. Scoring thresholds are set to define which matched customers are considered accurate enough. Without setting a threshold, many tens of thousands of "rubbish" matches will be returned.

For example, separate individuals whose only shared attribute may be the same Last Name or DOB or Workplace Phone Number.

The tuning of suitable thresholds involves multiple runs and analysis of *SVX* results until the best balance between accuracy and completeness is achieved. This process is typically performed by the consulting *SVX* expert when customising *SVX* for a particular installation.

Grouping

Once all scoring has been completed, *SVX* builds *SVX* groups from the *SVX* scored pairs.

For example, the following scenario involves 3 distinct rows that are distributed between two pairs:

- *Pair 1: (with internally generated SVX row_ids of 1 and 5) are assigned an SVX Score of 63.*

- *Pair 2: (with internally generated SVX row_ids of 5 and 9) are assigned an SVX Score of 59.*

SVX will define an SVX group with the following 3 members (and SVX Scores):

- *Row1: 63*

- *Row5: 63*

- *Row9: 59*

By grouping the records, *SVX* is ordering them so that each record is with similar records.

Ultimately, each group will be assigned an *SVX*_Id that will tie the disparate records together.

SVX Scheduler

The Scheduler oversees the running of the components and processes within the *SVX* Framework. By monitoring available system resources and task progress, it enables multiple processes to be run in parallel thereby reducing processing time.

SVX Process Flow

The *SVX* process flow is a structured multi entry process that allows processing of a given set(s) of records through a transactional engine, where customer defined logic is applied to produce the end results.

Processing and data flow is the same – irrespective of whether the record set is a complete set of data or delta changes.

The *SVX* process flow is shown below. Traditionally implemented as a post ETL process, the process flow shows the ability of having multiple entry points to satisfy business and technical requirements.
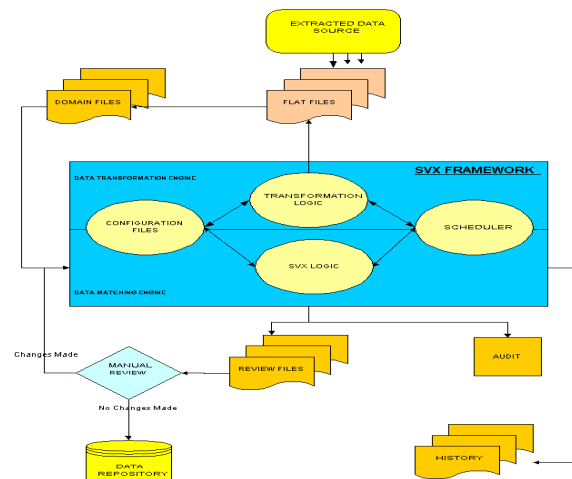
SVX Data Handling

The extraction of data for analysis and the uploading of the analysed data are handled outside the *SVX* framework. Data is handled within *SVX* by a 7-layer process. Extraction, the Optional Manual Review and Upload of the file output are external to the *SVX* framework as seen below.

The entire Data Handling process may also be automated for "lights out" processing.
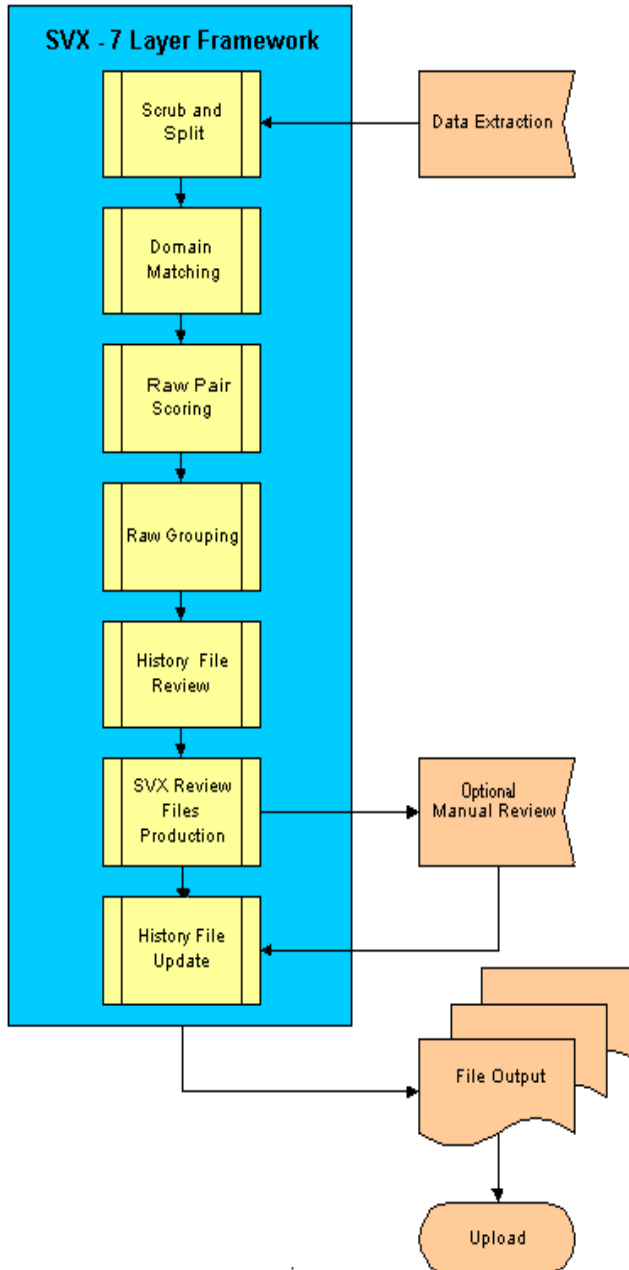
Business Rules

The *SVX* framework is intended to analyse information to identify a single individual or organisation within a company's potentially multiple data sources.

What is especially important are the possible changes of business rules during the lifetime of the Data Source and the resultant effect it may have on the data being analysed by *SVX*.

The application of business rules guiding how this data is interpreted is crucial to the conditional processing logic that needs to be customized for each customer.

Business Rule Types

Business rules are divided into two groups: standard and advanced.

Standard Business Rules define the process of customer data transformation. For Standard Business Rules to apply, customer data must exist and its accuracy has to be high.

If however, customer data is not available or data quality is questionable then Advanced Business Rules may be applied. Instead of using explicit values, customer identification is based on behavioural analyses and data element matching. This is the area where default assumptions are made relative to the data interpretations.

Standard Business Rules

Standard business rules form a set of transformation definitions for typical customer's attributes, which can be found in most information systems maintaining customer data. These are included in the *SVX* Data Transformation engine.

Advanced Business Rules

Advanced Business Rules may be applied independently from Standard Business Rules or if data quality of primary and/or secondary attributes is questionable.

Additionally, they may support Standard Business Rules to achieve the best match. They may be implemented by conditional logic within the *SVX* engines. An example of an advanced business rule could be a specific organisational rule governing treatment of dignitaries or government officials.

Thresholds.

Thresholds can be set across a number of parameters. The nature and reliability of the data will dictate which of, and how these parameters should be set. These should be set up in consultation with your *SVX* consultant and will be modified after multiple evaluative runs. The tuneable threshold parameters are:

- The *SVX* score, if any, from a Name match
- The number of matching domains that make up the total *SVX* score for a match
- The total *SVX* score for a match.

Operating Requirements

One of the key aspects of *SVX* is the ability to operate on a wide range of computing platforms and operating systems. These include:

- Commercially available UNIX operating systems supporting i86, PA-Risc and Sparc processors.
- Microsoft™ Windows 2000
- Microsoft™ NT

Operating Features

SVX processing is performed as a combination of UNIX scripts and Perl language constructs.

SVX has a feature that allows files created as part of the processing of records to be physically limited to a given size. *SVX* will create as many files as needed based on this size setting to complete the processing run. Setting a given size removes the problems that arise with the 2GB file size limitation found on most 32-bit commercial UNIX operating systems.

The Open Source programming language Perl is included in most UNIX distributions. Alternatively, it is readily available for download from the Internet.

SVX occupies approximately 15MB of disk space on the server.

SVX processing does not specifically perform the extract of data from the source systems. This is expected to be ready as a flat file input.

Conversely, *SVX* relies on the available loading tools the user has to load data into the target systems.

Summary

Today, businesses operate in a very complex and rapidly changing world where the *rate* of change is not only the fastest it has ever been, but is also increasing. Companies of all sizes are facing a hyper-competitive environment where they are forced to provide higher service at lower cost. At the same time, the general expectation of what is *good* customer service has gone up dramatically.

Applying a customer-centric approach to all aspects of a company's operations is a foreign concept for many organizations – but one that is gaining rapid approval and acceptance as more and more companies successfully implement such strategies and notice a difference in their bottom lines.

Finally, advances in technology have made it possible and quite economical for companies to invest in any one of a broad array of systems and tools to implement customer focused strategies via a whole host of software applications for automation, data gathering, data analysis, personalization, cost efficiency, and productivity improvement. Such technological advances have made it relatively easier and cheaper to provide personalization and customization in the online environment.

SVX provides this basic environment and creates a foundation for understanding your most prized business asset – your customers.

About Languix

Since 1990, Languix Pty Ltd has been providing information technology consulting services and solutions to companies addressing their most pressing data challenges.

Languix Pty Ltd

ACN 003 915 517

Level 2
366 Abercrombie Street
Darlington NSW 2008
Australia

www.lgx.com.au/svx

Tel: +61 (02) 9318 1021
Fax: +61 (02) 9318 2504