

## Course 596: Scripting for Data Manipulation (4 days)

### Course Description...

Data comes in from many sources and in many, many formats. The first objective of this course is to present the best available tools and techniques for selecting and re-formatting data. The course also shows how to use Perl's database independent methods for extracting, modifying, and inserting data.

### Prerequisites...

To get the maximum benefit from this course the student must have a working knowledge of Unix and some exposure to the different kinds of data formats available.

## Course Outline...

### Unix Tools

- Simple extraction with cut
- Simple joining with join
- Translations and character replacements with tr
- Specialized tools for data manipulation  
xargs, tee, and mastering the pipeline

### Mastering Regular Expressions

- Basic regular expressions
- How to read regular expressions
- How to write a regular expressions
- Extended Regular Expressions
- The grep family of tools

### Using sed

- How sed processes a file
- sed's command structure
- Regular Expressions and sed
- Simple sed commands
- Sed commands for substitution and replacement
- Moving lines with sed
- Writing sed scripts



## **Programming in awk**

- Language design
- Predefined variables and line processing
- Data manipulation with function calls
- Regular expressions and data selection
- Looping and field processing
- Writing awk scripts

## **Shell Scripting**

- Flow of control
- Shell scripts are co-ordinators
- I/O processing
- Field processing

## **Basic Perl**

- Scalars, lists, and arrays
- Program structure
- Line oriented I/O

## **Regular Expressions in Perl**

- The basic expression set
- Anchors and assertions
- Minimal match
- Command to make it fast

## **Perl Subroutines and Modules**

- How to write subroutines
- How to use predefined subroutines
- How to use OO modules

## **Database Access with Perl**

- Setup for implementation independent access
- Processing data from the database
- Methods of inserting into the database
- How to handle implementation uniqueness

## **Perl Data Manipulation Methods**

- Core techniques
- Some must have modules

*Please contact your ROI representative to discuss course tailoring!!!*