

CONTENTS

<i>Preface</i>	<i>xiii</i>
Chapter 1 Working with grep	1
What is the grep Command?	2
Meta Characters and the grep Command	3
Escaping Meta Characters with the grep Command	3
Useful Options for the grep Command	4
Character Classes and the grep Command	8
Working with the -c Option in grep	9
Matching a Range of Lines	10
Using Backreferences in the grep Command	12
Finding Empty Lines in Datasets	14
Using Keys to Search Datasets	14
The Backslash Character and the grep Command	15
Multiple Matches in the grep Command	15
The grep Command and the xargs Command	16
Searching Zip Files for a String	18
Checking for a Unique Key Value	18
Redirecting Error Messages	19
The egrep Command and fgrep Command	19
Displaying “Pure” Words in a Dataset with egrep	20
The fgrep Command	21
Delete Rows with Missing Values	22
A Simple Use Case	24
Summary	26
Chapter 2 Working with sed	27
What is the sed Command?	27
The sed Execution Cycle	28

- Matching String Patterns Using sed.....28
- Substituting String Patterns Using sed.....29
 - Replacing Vowels from a String or a File.....31
 - Deleting Multiple Digits and Letters from a String.....31
- Search and Replace with sed.....32
- Regular Expressions with sed.....34
- Datasets with Multiple Delimiters.....35
- Useful Switches in sed.....36
- Working with Datasets.....37
 - Printing Lines.....37
 - Character Classes and sed.....38
 - Removing Control Characters.....39
- Counting Words in a Dataset.....40
- Back References in sed.....40
- One-Line sed Commands.....41
- Populate Missing Values with the sed Command.....47
- A Dataset with 1,000,000 Rows.....48
 - Numeric Comparisons.....48
 - Counting Adjacent Digits.....49
 - Average Support Rate.....50
- Summary.....52

Chapter 3 Working with awk.....53

- The awk Command.....54
 - Launching awk Commands and File Extensions.....54
 - Built-In Variables that Control awk.....55
 - How Does the awk Command Work?.....56
 - Three Important Code Blocks in awk Commands.....56
 - Different Ways to Print Text.....57
- Working with the -F Switch in awk.....59
 - Splitting Strings into Characters.....60
 - The PROCINFO Array.....62
 - Ignore Case in awk.....62
- Working with OFS, ORS, and Linefeed versus “\n”.....63
 - Linefeed versus “\n”.....64
- Basic awk Examples with Text Files.....65
 - Display the Contents of a File.....65
 - Omit the Header Line of a File.....66
 - Display a Range of Lines in a Text File With NR.....66
 - Display a Range of Lines Based on Strings.....66
 - Insert Blank Lines.....67
 - Remove Blank Lines.....67
 - Specifying the Number of Fields.....67
 - Changing the Field Separator FS.....68
 - Exclude Fields.....70
 - Switch Adjacent Fields.....71
 - Display Fields in Reverse Order.....72

Count Non-Empty and Empty Rows	73
Detecting Transfer Protocols	73
Detecting Number Formats	74
Working with Record Lengths	75
Aligning Text with the printf() Statement	76
Working with Loops in awk	77
A for Loop in awk	77
Exponents in a for Loop	77
A for Loop with a break Statement	78
Working with while Loops in awk	78
A do-while Loop in awk	79
Conditional Logic and Control Statements	80
The break, continue, and next Keywords in awk	80
The exit Keyword	83
Conditionally Displaying Fields	83
Logical Operators in awk	85
Logical Operators and Partial Matches	85
Checking for Leap Years	87
Formatting Output	89
Floating Point Output	92
Inserting a Thousands Separator	94
Scientific Notation	95
Rounding and Truncating Numbers	97
Rounding Numbers	98
Truncating Numbers	98
Numeric Functions in awk	99
Convert Base 10 to Binary	101
Built-In String Functions in awk	102
Convert Strings to Lowercase in awk	103
Convert Strings to Uppercase in awk	103
Convert Strings to Mixed Case in awk	103
Counting Lines that Match a Character	104
Working with the match() Function	104
Characters and Hexadecimal Numbers in awk	105
Non-Printable Characters	105
Hexadecimal Numbers	105
Summary	106

Chapter 4 Working with awk, Part 2..... 107

Working with Arrays in awk	107
Initializing and Printing the Contents of an Array	108
Initializing and Deleting the Contents of an Array	108
Adding an Array of Numbers	109
Find the Maximum and Minimum of an Array of Numbers	110
Concatenating an Array of Strings	111
Generating Arrays from Input Strings	111
The split() Function with Linefeed Strings	112

Using the split() Function with the case Statement	112
The patsplit() Function.....	113
Multi-Dimensional Arrays	113
Higher Dimensionality Arrays.....	116
Calculating Invoice Totals (1)	117
Calculating Invoice Totals (2).....	118
Calculating Invoice Averages.....	120
Counting Fields in Text Files	120
Displaying the Number of Fields in Text Files.....	121
Deleting Alternate Lines in Datasets	122
Print a Range of Strings in Datasets.....	123
Print a Range of Dates in Datasets.....	124
Merging Lines in Datasets.....	126
Printing File Contents as a Single Line	127
Joining Groups of Lines in a Text File	127
Joining Alternate Lines in a Text File	128
Reversing the Lines in a File.....	129
Working with Duplicate Lines in a File	130
Display Duplicate Rows.....	130
Remove Duplicate Rows	132
Uniformity of Data Values	133
Count Duplicate Fields in a File	134
Calculating Invoice Totals.....	135
Printing Lines Using Conditional Logic.....	136
Duplicate and Unique Rows in Text Files	139
Splitting Filenames with awk.....	140
One-Line awk Commands	141
Useful Short awk Scripts.....	141
Printing the Words in a Text String in awk	143
Count Occurrences of a String in Specific Rows.....	143
Well-Known Sorting Algorithms.....	144
The Bubble Sort.....	145
Find Anagrams in a List of Words.....	147
Sort an Array in O(n) Complexity.....	148
Find the Median of an Array of Numbers	149
Linear Search	150
Binary Search (Iterative).....	151
Binary Search Walkthrough.....	152
Code for a Binary Search (Iterative)	152
Binary Search (Recursive)	153
Summary.....	155

Chapter 5 Working with awk, Part 3..... 157

Bit Operations in awk.....	157
Logical AND	158
Logical OR	158
Logical XOR.....	158

Logical NOT.....	159
Logical Left Shift	159
Logical Right Shift	160
Reverse a String	160
Find Palindromes.....	161
Check for Unique Characters.....	161
Check for Balanced Parentheses	163
Printing a String in a Fixed Number of Columns.....	164
Printing a Dataset in a Fixed Number of Columns	165
Aligning Columns in Datasets	166
Aligning Columns and Multiple Rows in Datasets	167
Displaying a Subset of Columns in a Text File	168
Subsets of Columns Aligned Rows in Datasets.....	170
Longest/Shortest Words in Datasets	172
Counting Word Frequency in Datasets	173
Displaying Only “Pure” Words in a Dataset	175
Delete Rows with Missing Values.....	176
Working with Multiline Records in awk.....	179
CSV Files with Multirow Records.....	180
Processing Multiple Files with awk.....	183
Inner Join of Two Files in awk.....	186
Logical Operators in awk	187
A Simple Use Case	188
Another Use Case	189
Date Conversion	191
A Dataset with 1,000,000 Rows	193
Numeric Comparisons	193
Counting Adjacent Digits	194
Average Support Rate	194
What is Recursion?	195
Arithmetic Series.....	196
Calculating Arithmetic Series (Iterative)	196
Calculating Arithmetic Series (Recursive)	197
Calculating Partial Arithmetic Series	198
Geometric Series.....	199
Calculating a Geometric Series (Iterative).....	200
Calculating Arithmetic Series (Recursive)	201
Factorial Values	202
Fibonacci Numbers.....	202
Euclid’s Algorithm (GCD)	203
Lowest Common Multiple of Two Positive Integers	204
Composite versus Prime Numbers.....	205
Factors of Positive Integers	207
Goldbach’s Conjecture.....	208
Environment Variables in awk.....	209
Summary.....	210

Chapter 6 Regular Expressions.....	211
What are Regular Expressions?	212
Metacharacters.....	212
Character Sets	213
Working with “^” and “\”	214
Character Classes	215
POSIX Character Classes	215
Generic Character Classes.....	216
Regular Expressions in awk	217
Matching Strings with the Dot “.” Metacharacter.....	217
Matching Strings with the “^” Metacharacter.....	218
Matching Strings with the “\$” Metacharacter	219
Matching Strings with the “ ” Metacharacter.....	219
Matching with ?, +, or * Metacharacters	220
Matching Subgroups with ?, +, or * Metacharacters.....	221
Matching with Character Classes.....	221
Working with Postfix Arithmetic Operators.....	224
Working with the sub() Function in awk.....	225
Examples Using the sub() Function.....	225
Working with the gsub() Function	226
Removing Digits with the gsub() Function.....	226
Removing Characters with the gsub() Function.....	227
Removing Consecutive Characters with the gsub() Function	228
Removing Complements of Strings with the gsub() Function.....	228
Removing Metacharacters with the gsub() Function	228
Removing Spaces in Text Fields.....	229
Splitting Strings with Regular Expressions	229
Dynamic Regular Expressions.....	230
Regular Expressions Involving Text Files	231
Counting Comments with Regular Expressions	232
Combining grep with awk.....	233
Combining sed with awk.....	233
Removing HTML Tags	234
The gensub() Function and Captured Groups	236
Summary.....	237
 Index	 239