# Table of Contents

Acknowledgements..............................................................................................................................................12

Chapter 1. Introduction........................................................................................................................................13

1.1. Purpose and Structure of this Book ........................................................................................................13

1.2. Data and Workflows for this Book ...........................................................................................................14

1.3. Memory Usage in KNIME Analytics Platform ........................................................................................16

Chapter 2. Database Operations ...................................................................................................................19

2.1. Database Nodes .......................................................................................................................................19

2.2. Connect to a Database: DB Connector Nodes .......................................................................................20

   DB Connector ........................................................................................................................................21

   Register your own JDBC driver .............................................................................................................22

   Workflow Credentials ..........................................................................................................................24

   SQLite Connector ...............................................................................................................................25

2.3. Select the Table to work on: the DB Table Selector Node ................................................................26

   DB Table Selector ...............................................................................................................................26

2.4. In-Database Processing ........................................................................................................................28

   DB Row Filter .....................................................................................................................................29

   DB Column Filter ..............................................................................................................................31

   DB Query ...........................................................................................................................................32

2.5. Utility Nodes for Databases ...................................................................................................................33

   DB SQL Executor ...............................................................................................................................34

   DB Query Injector .............................................................................................................................34

   DB Query Extractor ............................................................................................................................35
<table>
<thead>
<tr>
<th>2.6.</th>
<th>Reading Data resulting from a SQL Query</th>
<th>35</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DB Reader</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>Table Creator</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td>Parameterized DB Query Reader</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>DB Query Reader</td>
<td>39</td>
</tr>
<tr>
<td>2.7.</td>
<td>Writing Data resulting from a SQL Query</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>DB Writer</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>DB Connection Table Writer</td>
<td>41</td>
</tr>
<tr>
<td>2.8.</td>
<td>Database UPDATE and DELETE Commands</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td>DB Delete</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>DB Update</td>
<td>43</td>
</tr>
<tr>
<td>2.10.</td>
<td>Big Data Platforms and MongoDB</td>
<td>46</td>
</tr>
<tr>
<td>2.11.</td>
<td>Exercises</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>Exercise 1</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>Exercise 2</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td>Exercise 3</td>
<td>51</td>
</tr>
<tr>
<td>Chapter 3. Accessing Information from the Web</td>
<td>53</td>
<td></td>
</tr>
<tr>
<td>3.1.</td>
<td>Accessing Google Sheets</td>
<td>53</td>
</tr>
<tr>
<td></td>
<td>Google Authentication</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td>Google Sheets Connection</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>Google Sheets Reader</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td>Google Sheets Appender</td>
<td>58</td>
</tr>
<tr>
<td></td>
<td>Google Sheets Updater</td>
<td>59</td>
</tr>
</tbody>
</table>
5.6. Configurations, Widgets, Components, and KNIME WebPortal ................................................................. 110
5.7. Transform a Flow Variable into a Data Value ................................................................................................. 113
  Variable To TableRow ........................................................................................................................................ 113
5.8. Modifying Flow Variable Values ..................................................................................................................... 114
5.9. Configuration Nodes and Widget Nodes .......................................................................................................... 117
  Value Selection Widget ...................................................................................................................................... 118
  Local File Browser Configuration ...................................................................................................................... 119
5.10. Composite View in Components .................................................................................................................... 121
  Interactive Range Slider Filter Widget .............................................................................................................. 124
5.11. Exercises ....................................................................................................................................................... 126
  Exercise 1 .......................................................................................................................................................... 126
  Exercise 2 .......................................................................................................................................................... 127
  Exercise 3 .......................................................................................................................................................... 129
  Exercise 4 .......................................................................................................................................................... 131
Chapter 6. Loops .................................................................................................................................................... 133
6.1. What is a Loop ................................................................................................................................................ 133
6.2. Loop with a pre-defined number of iterations ................................................................................................. 135
  Data Generator ................................................................................................................................................ 136
  Counting Loop Start ......................................................................................................................................... 138
  Loop End .......................................................................................................................................................... 138
6.3. Dedicated Commands for Loop Execution ...................................................................................................... 141
6.4. Appending Columns to the Output Data Table ............................................................................................... 143
  Loop End (Column Append) ............................................................................................................................. 144
6.5. Loop on a List of Columns .......................................................................................................................... 146
    Column List Loop Start ............................................................................................................................... 147
6.6. Loop on a List of Values .............................................................................................................................. 150
    TableRow To Variable Loop Start ................................................................................................................ 150
    Cache ............................................................................................................................................................ 152
    Loop End (2 ports) ....................................................................................................................................... 153
6.7. Loop on Data Groups and Data Chunks ........................................................................................................ 154
    Group Loop Start .......................................................................................................................................... 155
    Chunk Loop Start ......................................................................................................................................... 156
    Breakpoint .................................................................................................................................................... 158
6.8. Keep Looping till a Condition is verified ..................................................................................................... 158
    Generic Loop Start ....................................................................................................................................... 158
    Variable Condition Loop End .................................................................................................................... 159
6.9. Recursive Loop .......................................................................................................................................... 160
    Recursive Loop Start .................................................................................................................................. 161
    Recursive Loop End .................................................................................................................................... 162
6.10. Exercises .................................................................................................................................................. 163
    Exercise 1 ..................................................................................................................................................... 163
    Exercise 2 ..................................................................................................................................................... 165
    Exercise 3 ..................................................................................................................................................... 166
    Exercise 4 ..................................................................................................................................................... 168
Chapter 7. Switches .......................................................................................................................................... 171
7.1. Introduction to Switches .............................................................................................................................. 171
7.2. The “IF Switch” - “END IF” switch block ................................................................................................................. 172
    IF Switch ........................................................................................................................................................................... 173
    End IF .................................................................................................................................................................................. 174
    Auto-Binner .......................................................................................................................................................................... 176

7.3. The “Java IF (Table)” node ............................................................................................................................................ 177
    Java IF (Table) ....................................................................................................................................................................... 178

7.4. The CASE Switch Block .................................................................................................................................................. 179
    CASE Switch Data (Start) .................................................................................................................................................. 180
    CASE Switch Data (End) ................................................................................................................................................... 181

7.5. Transforming an Empty Data Table Result into an Inactive Branch ............................................................................. 182
    Empty Table Switch .............................................................................................................................................................. 183

7.6. Exercises ............................................................................................................................................................................. 184
    Exercise 1 .............................................................................................................................................................................. 184
    Exercise 2 .............................................................................................................................................................................. 186

Chapter 8. Advanced Reporting ........................................................................................................................................ 189

8.1. Introduction ....................................................................................................................................................................... 189

8.2. Report Parameters from global Flow Variables............................................................................................................ 191
    Concatenate (Optional in) ................................................................................................................................................... 192

8.3. Customize the “PARAMETER SELECTION PAGE” web page ..................................................................................... 194

8.4. The Expression Builder....................................................................................................................................................... 197

8.5. Dynamic Text....................................................................................................................................................................... 200

8.6. BIRT and JavaScript Functions ........................................................................................................................................ 202

8.7. Import Images from the underlying Workflow .................................................................................................................. 204
Acknowledgements

We would like to thank a number of people for their help and encouragement in writing this book.

In particular, we would like to thank Bernd Wiswedel for answering our endless questions about calling external REST services from inside a workflow, Iris Adae for explaining the most advanced features of the Date&Time nodes, and Andisa Dewi for significantly contributing to the update of the Database Operations chapter.

Special thanks go to Peter Ohl for reviewing the book contents and making sure that they comply with KNIME intended usage and to Heather Fyson for reviewing the book’s English written style.

Finally, we would like to thank the whole KNIME Team for their support in publishing and advertising this book.
Chapter 1. Introduction

1.1. Purpose and Structure of this Book

KNIME Analytics Platform is a powerful tool for data analytics and data visualization. It provides a complete environment for data analysis which is fairly simple and intuitive to use. This, coupled with the fact that KNIME Analytics Platform is open source, has led a large number of professionals to use it. In addition, third-party software vendors develop KNIME extensions in order to integrate their tools into it. KNIME nodes are now available that reach beyond customer relationship management and business intelligence, extending into the field of finance, life sciences, biotechnology, pharmaceutical, and chemical industries. Thus, the archetypal KNIME user is no longer necessarily a data mining expert, although his/her goal is still the same: to understand data and to extract useful information.

This book was written with the intention of building upon the reader’s first experience with KNIME software. It expands on the topics that were covered in the first KNIME user guide (“KNIME Beginner’s Luck” [1]) and introduces more advanced functionalities. In the first guide [1], we described the basic principles of KNIME Analytics Platform and showed how to use it. We demonstrated how to build a basic workflow to manipulate, visualize, and model data, and how to build reports. Here, we complete these descriptions by introducing the reader to more advanced concepts. A summary of the chapters provides you with a short overview of the contents to follow.

Chapter 2 describes the nodes needed to connect to a database, import data, build an appropriate SQL query to select a subset of the data or for some required processing, and finally to write data back into the database. Accessing a database, importing data, and building SQL queries are the basic operations necessary for any, even very simple, data warehousing strategy.

Of course, the largest source of data is nowadays the Internet. Chapter 3 is dedicated to alternative ways of getting data besides files and databases, i.e. web data sources. Chapter 3 starts with the connectors to Google Sheets and continues with access to REST services. Those are definitely powerful tools to search for data elsewhere.

Chapter 4 introduces the Date&Time object and the nodes to turn a String column into a Date&Time column, to format it, to extract a time difference, and in general to perform date and time based operations. The Date&Time object provides the basis for working with time series. The last section of chapter 4 briefly describes a few nodes dedicated to time series analysis.

A very important concept for the KNIME workflows is the concept of “flow variables”. Flow variables enable external parameters to be introduced into a workflow to control its execution. Chapter 5 describes what a flow variable is, how to create it, and how to edit it inside the workflow, if needed.
Most data operations in KNIME Analytics Platform are executed on a data matrix, named data table. This means that an operation is executed on all data rows. This is a big advantage in terms of speed and programming compactness. However, from time to time, a workflow also needs to run its rows, one after the other, through an operation. That is, sometimes it needs a real loop. Chapter 5 introduces a few nodes that implement loops: from a simple “for” cycle to more complex loops, such as looping on a list of values or feeding the current iteration results into the next iteration.

Chapter 7 illustrates the use of logical switches to change the workflow path upon compliance with some predefined condition.

Chapter 8 is an extension of chapter 6 in “KNIME Beginner’s Luck” [1]: it describes a number of advanced features of KNIME reporting tool. First of all, it explains how to introduce parameters into a report and how flow variables and report parameters are connected. Later on, in the chapter, a few more reporting functions are discussed which can be used to create a more dynamic report.

In this introductory chapter, we list the data and the example workflows that have been built for this book and note the KNIME Extensions required to run some of the example workflows.

1.2. Data and Workflows for this Book

In the course of this book we will put together a few workflows to show how KNIME Analytics Platform works. In each chapter we will build one or more workflows and we will expect the reader to build a few more in the exercises.

When you purchased this book, in the purchase email containing the link to the pdf file, there should also have been a link to the Download Zone file. The Download Zone file is a .knar file and contains the data and workflows used and implemented in this book.

- Download the Download Zone knar file onto your machine
- Import it into KNIME Explorer:
  - Select “File” -> “Import KNIME Workflow ...
  - In the “Import Workflow” window, select option “Select File” and navigate to the Download Zone knar file.
1.1. Workflows and data used in this book, as imported from the Download Zone.knar file

At the end of the import operation, in your “KNIME Explorer” panel you should find a folder named “Advanced Luck” and containing Chapter2, Chapter3, Chapter4, etc ... subfolders, each one with workflows and exercises to be implemented in the corresponding chapters of this book. In addition, under the main folder “Advanced Luck”, you should find a KCBdata subfolder containing all necessary data.

The data used for the exercises and for the demonstrative workflows of this book were either generated by the author or downloaded from the UCI Machine Learning Repository [2], a public data repository (http://archive.ics.uci.edu/ml/datasets). If the data set belongs to the UCI Repository, a full link is provided here to download it. Data generated by the author, that is not public data, are located only in the KCBdata folder.

Data sets from the UCI Machine Learning Repository [2]:

- Automobile: http://archive.ics.uci.edu/ml/datasets/Automobile
- Slump_test: http://archive.ics.uci.edu/ml/datasets/Concrete+Slump+Test
This book is not meant as an exhaustive reference for KNIME Analytics Platform, although many useful workflows and aspects of it are demonstrated through worked examples. This text is intended to give you the confidence to use the advanced functions in KNIME Analytics Platform to manage and mine your own data.

1.3. Memory Usage in KNIME Analytics Platform

Sometimes some workflows require exceptional memory usage. The amount of memory available to KNIME Analytics Platform is stored in the knime.ini file. The knime.ini file is located in the directory in which KNIME Analytics Platform has been installed, together with the knime.exe file. The knime.ini file contains a number of required settings.

-Xmx<size> is the setting that defines the maximum heap size available to run the workflows. You can define this value by editing the knime.ini file or at installation time. During installation, in fact, you are required to insert the maximum amount of memory available to KNIME Analytics Platform. After that, if you run into memory problems, you probably need to manually increase the heap space (-Xmx option) directly in the knime.ini file to a size compatible with the memory you have on your machine.
Now, let’s suppose that you are having a memory problem while executing a workflow. You have already changed the -Xmx value in the knime.ini file to a higher value, but that was not enough. You are still running into memory problems when your workflow is executed. Even though this kind of problem occurs very rarely in KNIME, it can still happen. In this case, you need to know whether the problem is due to your workflow or to some other program running on your machine at the same time.

There is an easy way to monitor how much heap space is being used by a workflow and if this reaches the maximum limit assigned by the –Xmx option. In the Top Menu in the KNIME workbench:

- Click “File”
- Select “Preferences”
- “Preferences” window opens
- In “Preferences” window
  - Select “General”
  - In the frame on the right named “General”, enable the option “Show heap status”
  - Click “OK”
  - Now, in the lower right corner you can see a small number showing the heap status.
1.4. The "Preferences" window with the "Show heap status" option

1.5. The bottom right corner shows the heap status

To run the example workflows and the exercises provided in this book, you will need to install the following the whole “KNIME & Extensions” group.

In order to install a KNIME Extension:

- In the top menu of the KNIME workbench, select “File” -> “Install KNIME Extensions …”
- In the “Install” window:
  - Open the group containing your extension, like for example “KNIME & Extensions” group
  - If you do not know where your extension package is located, just run a search by inserting a few related keywords in the top textbox
  - Select your extension
  - Click “Next” and follow installation instructions