# Table of Contents

Acknowledgements ............................................................................................................. 11

Chapter 1. Introduction ....................................................................................................... 12

1.1. Purpose and Structure of this Book ........................................................................... 12

1.2. Data and Workflows for this Book ........................................................................... 13

1.3. Memory Usage in KNIME Analytics Platform ........................................................... 14

Chapter 2. Database Operations ......................................................................................... 18

2.1. Database Nodes: Modular vs. All-in-One Approach .................................................. 18

2.2. Connect to a Database: Database Connector Nodes .................................................. 19

   (Generic) Database Connector .................................................................................. 20

   Workflow Credentials .............................................................................................. 21

   Upload a new JDBC Driver ..................................................................................... 22

   (Dedicated) SQLite Connector .................................................................................. 23

2.3. Select the Table to work on: the Database Table Selector Node .............................. 24

   Database Table Selector .......................................................................................... 24

2.4. Connect to a Database AND Select the Table: the Database Table Connector Node ... 25

   Database Table Connector ....................................................................................... 26

2.5. In-Database Processing .............................................................................................. 27

   Database Row Filter ................................................................................................. 27

   Database Column Filter .......................................................................................... 29

   Database Query ....................................................................................................... 30

   Database SQL Executor .......................................................................................... 31

   SQL Inject ................................................................................................................ 32
Chapter 3. Accessing Information on the Web

3.1. Accessing Google Sheets

2.6. Looping on Database Data

2.7. Read and Write Data resulting from a SQL Query

2.8. All in one Node: Database Connection, Table Selection, SQL Query, and Exporting Data

2.9. Database UPDATE and DELETE Commands

2.10. Big Data Platforms and MongoDB

2.11. Exercises
Chapter 4. Date&Time Manipulation

4.1. The Date&Time Type

4.2. How to produce a Date&Time Column

String to Date&Time

Date&Time to String

Create Date&Time Range
Chapter 5. Flow Variables

4.3. Refine Date&Time Values ........................................................................................................78
    Modify Time ............................................................................................................................78
    Date&Time Shift .....................................................................................................................79

4.4. Row Filtering based on Date&Time Criteria ........................................................................80
    Date&Time-based Row Filter ..................................................................................................81
    Extract Date&Time Fields ......................................................................................................82
    Date&Time Difference ............................................................................................................84

4.5. Moving Average and Aggregation .........................................................................................85
    Moving Average ....................................................................................................................87
    Moving Aggregation ................................................................................................................88

4.6. Time Series Analysis .............................................................................................................90
    Lag Column ............................................................................................................................91

4.7. Exercises .............................................................................................................................93
    Exercise 1 ...............................................................................................................................93
    Exercise 2 ...............................................................................................................................94

Chapter 5. Flow Variables ...........................................................................................................96

5.1. What is a Flow Variable? ........................................................................................................96

5.2. Creating a Flow Variable for all Nodes in the Workflow .......................................................97

5.3. Flow Variable Values as Node Settings ................................................................................99
    The “Flow Variable” Button ..................................................................................................99
    The “Flow Variables” Tab in the Configuration Window .......................................................101

5.4. Creating a Flow Variable from within a Workflow ...............................................................102
    Transform a Data Value into a Flow Variable........................................................................102

5.11. Exercise 4

5.10. Composite View in Wrapped Meta-nodes

5.9. Other Quickform Nodes

5.8. Modifying Flow Variable Values

5.7. Transform a Flow Variable into a Data Value

5.6. Quickforms, Wrapped Meta-nodes, and KNIME WebPortal

5.5. Inject a Flow Variable through the Flow Variable Ports

5.4. Workflow Variable Injection into the Workflow

5.3. Merge Variables

5.2. Transform a Configuration Setting into a Flow Variable

5.1. Quickforms to Create Flow Variables

5.0. Integer Input

5.0. What is a Loop

TableRow To Variable

Transform a Configuration Setting into a Flow Variable

Quickforms to Create Flow Variables

Integer Input

5.5. Inject a Flow Variable through the Flow Variable Ports

Workflow Variable Injection into the Workflow

Merge Variables

5.6. Quickforms, Wrapped Meta-nodes, and KNIME WebPortal

5.7. Transform a Flow Variable into a Data Value

Variable To TableRow

5.8. Modifying Flow Variable Values

5.9. Other Quickform Nodes

Value Selection

File Upload

5.10. Composite View in Wrapped Meta-nodes

Range Slider Filter Definition

5.11. Exercises

Exercise 1

Exercise 2

Exercise 3

Exercise 4

Chapter 6. Loops

6.1. What is a Loop
6.2. Loop with a pre-defined number of iterations .......................... 134
   Data Generator ........................................................................ 135
   Counting Loop Start .............................................................. 137
   Loop End ............................................................................. 137
6.3. Dedicated Commands for Loop Execution .................................. 140
6.4. Appending Columns to the Output Data Table ........................ 142
   Loop End (Column Append) .................................................. 143
6.5. Loop on a List of Columns ..................................................... 145
   Column List Loop Start .......................................................... 146
6.6. Loop on a List of Values ....................................................... 149
   TableRow To Variable Loop Start ........................................ 149
   Cache ................................................................................ 151
   Loop End (2 ports) ............................................................... 152
6.7. Loop on Data Groups and Data Chunks .................................. 153
   Group Loop Start ................................................................. 154
   Chunk Loop Start ............................................................... 155
   Breakpoint ........................................................................ 157
6.8. Keep Looping till a Condition is verified .................................. 157
   Generic Loop Start .............................................................. 157
   Variable Condition Loop End ................................................ 158
6.9. Recursive Loop ................................................................... 159
   Recursive Loop Start .......................................................... 160
   Recursive Loop End ............................................................ 161
7.1. Introduction to Switches .............................................................................................................. 170
7.2. The “IF Switch”- “END IF” switch block ..................................................................................... 171
    IF Switch ........................................................................................................................................ 172
    End IF ........................................................................................................................................ 173
    Auto-Binner ................................................................................................................................. 175
7.3. The “Java IF (Table)” node .......................................................................................................... 176
    Java IF (Table) ........................................................................................................................... 177
7.4. The CASE Switch Block ............................................................................................................... 178
    CASE Switch Data (Start) ............................................................................................................ 179
    CASE Switch Data (End) .............................................................................................................. 180
7.5. Transforming an Empty Data Table Result into an Inactive Branch ............................................. 181
    Empty Table Switch ....................................................................................................................... 182
7.6. Exercises ..................................................................................................................................... 183
    Exercise 1 .................................................................................................................................... 183
    Exercise 2 .................................................................................................................................... 183
Chapter 8. Advanced Reporting ........................................................................................................ 188
8.1. Introduction .................................................................................................................................. 188
8.2. Report Parameters from global Flow Variables ................................................................. 190
     Concatenate (Optional in)........................................................................................................ 191
8.3. Customize the “PARAMETER SELECTION PAGE” web page .................................................. 193
8.4. The Expression Builder ........................................................................................................... 196
8.5. Dynamic Text........................................................................................................................... 198
8.6. BIRT and JavaScript Functions ............................................................................................... 201
8.7. Import Images from the underlying Workflow ......................................................................... 202
     Read Images .............................................................................................................................. 204
8.8. Exercises .................................................................................................................................. 207
     Exercise 1 .................................................................................................................................. 207
     Exercise 2 .................................................................................................................................. 209
     Exercise 3 .................................................................................................................................. 211
References ....................................................................................................................................... 214
Node and Topic Index ..................................................................................................................... 215
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, and Iris Adae for explaining the most advanced features of some of the Date&Time nodes.

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

The 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 the 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 the KNIME Analytics Platform. KNIME nodes are now available that reach beyond customer relationship management and business intelligence, extending into the field of finance, the 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. 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 KNIME user 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 some 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 4 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, continues with access to REST services, and concludes with a web crawling example workflow. 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 shows 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 the 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 the 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 the 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 the KNIME Explorer:
  - Select “File” -> “Import KNIME Workflow ...
  - In the “Import Workflow” window, select option “Archive File” and select 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 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 the KNIME software is stored in the knime.ini file. The knime.ini file is located in the directory in which KNIME has been installed, together with the knime.exe file. The knime.ini file contains a number of settings required by the KNIME software.
1.2. The “knime.ini” file

-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 the 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
  o Select “General”
  o In the frame on the right named “General”, enable the option “Show heap status”
  o Click “OK”
  o Now, in the lower right corner you can see a small number showing the heap status.
To run the example workflows and the exercises provided in this book, you will need to install the following KNIME Extensions:

- The whole “KNIME & Extensions” group
- “Palladian for KNIME” under “KNIME Community Contributions – Other”

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