

Table of Contents

Preface

Part 1: Introduction

1

Getting Started with Alteryx

Understanding the Alteryx platform

The software that makes the Alteryx platform 4

Using the Alteryx platform in a business scenario 5

How Alteryx benefits data engineers 6

Using Alteryx Designer

Why is Alteryx Designer suitable for data engineering? 8

Building a workflow in Designer 9

What can the InDB tools do? 9

Best practices for Designer workflows 13

2

Data Engineering with Alteryx

What is a data engineer? 22

Using Alteryx products as a data engineer 23

Leveraging Alteryx Server and Alteryx Connect

How can you use Alteryx Server to orchestrate a data pipeline? 15

How does Connect help with discoverability? 17

Using this book in your data engineering work

How does the Alteryx platform come together for data engineering? 17

Examples where Alteryx is used for data engineering 18

Summary

19

Creating with Designer for data engineers 26

Automating with Server for data engineers 29

Connecting end users with Alteryx Connect	31	Applying DataOps as an Alteryx data engineer	32
		Summary	33

3

DataOps and Its Benefits

The benefits the DataOps framework brings to your organization	36	Applying DataOps to Alteryx	42
Faster cycle times	36	Supporting the People pillar with Alteryx	42
Faster access to actionable insights	37	Using Alteryx to deliver data pipelines	42
Improved robustness of data processes	37	Building confidence with Alteryx	43
Provides an overview of the entire data flow	37	Using Alteryx software with DataOps	44
Strong security and conformance	37	Alteryx Designer	45
		Alteryx Server	45
		Alteryx Connect	46
Understanding DataOps principles	38	General steps for deploying DataOps in your environment	46
The People pillar	39	Summary	48
The Delivery pillar	40		
The Confidence pillar	41		

Part 2: Functional Steps in DataOps

4

Sourcing the Data

Technical requirements	52	Leveraging external data sources from authenticated APIs	65
Accessing internal data sources	52	Connecting to an API with URL parameters	66
Data source types	52	Connecting to an API in call headers	67
Using the Alteryx Input Data tool	54	Initial cleansing of datasets	68
Integrating public data sources with Download tool use	61	A simple cleansing process	68
Identifying whether the data structure has changed	63	A consolidated cleansing process	72
Creating our first validation test	64		

Constructing a data pipeline in Alteryx Designer	74	Calculating the first set of statistical values	77
Configuring the annotations and names	76	Saving the processed dataset	77
Adding initial documentation in the workflow	76	Summary	78
5			
Data Processing and Transformations			
Technical requirements	80	Investigating distributions in your dataset	100
The data cleansing process	80	Correcting missing values in your dataset	104
Selecting columns	81	Transforming our data pipeline	107
Filtering to relevant rows	84	Transforming the downloaded data	108
Generating features and modifying columns with formulas	89	Profiling our dataset	111
Summarizing the dataset	94	Summary	114
Profiling data with summary and statistical aggregations	95		
Investigating the variation and size range of your dataset	96		
6			
Destination Management			
Technical requirements	118	Integrating data pipelines across environments	130
Writing to destinations	118	Using a secrets file or environment variable	131
Writing to files	118	Creating a DSN in the Windows ODBC manager	136
Managing database connections	125	Getting the most from a connection	137
Using standard connections	125	Publishing the external data to a Snowflake destination	138
How to load data faster with bulk loaders	128	Installing the drivers	139
Leveraging a database's custom tools	129	Creating the Snowflake ODBC DSN	140
Accessing more data sources with custom connections	129	Summary	144

7

Extracting Value

Technical requirements	146	Making our datasets available for other BI tools	170
Exploratory data analysis in Alteryx and surfacing the datasets for BI tools	147	Using Alteryx to deliver standard reports	171
Identifying missing values and summarizing fields	147	Creating a formatted table	172
Understanding your value distribution	157	Adding visualizations to the report	173
Finding relationships between fields	163	Adding styling to the report	175
Identifying any outliers in the dataset	168	Outputting the report for consumption	180
Understanding the difference between the Interactive Chart tool and the Insights tool	170	Summary	183

8

Beginning Advanced Analytics

Technical requirements	186	Summarizing spatial information	193
Implementing spatial analytics with Alteryx	187	Beginning the ML process in Alteryx	195
Creating a spatial point	187	Using the Intelligence Suite	195
Geocoding addresses to make spatial points	189	Building workflows with R-based predictive tools	198
Generating trade areas for analysis	191	Creating a custom Python or R script in a workflow	200
Combining data streams with spatial information	192	Summary	202

Part 3: Governance of DataOps

9

Testing Workflows and Outputs

Technical requirements	205	Monitoring workflows with the Message tool	206
Workflow tests and messages	206		

Monitoring workflows with the Test tool	208	Confirming the country of our place search	216
Using the Community CREW test macros	210	Centralizing the monitoring outputs with Insights	218
Validating data outputs	211	Building a control chart monitoring system	219
Automating the result monitoring actions	211	Using insights on your Alteryx server	224
Running tests on the output dataset	215	Summary	225

10

Monitoring DataOps and Managing Changes

Technical requirements	228	Exporting the MongoDB database for custom analysis	240
Using the Alteryx Server monitoring workflow	228	The MongoDB schema	240
Accessing and installing the server monitoring workflow	228	Modifying the Server Monitoring workflow	241
Reading the PDF report	230	Using Git and GitHub Actions for continuous integration	247
Using the data output	234	Saving workflow changes with Git	248
Creating an insight dashboard for workflow monitoring	235	Verifying the XML workflow	253
Creating a monitoring dashboard	235	Applying standards with GitHub Actions	257
Exporting the data output for external reports	239	Summary	259

11

Securing and Managing Access

Technical requirements	262	Managing collections	268
Organizing content on Alteryx Server	262	Creating collections	269
My Workspace	264	Securing the data environment	276
Collections	264	Alteryx Server architecture	276
Districts	264	Summary	280

12

Making Data Easy to Use and Discoverable with Alteryx

Technical requirements	282	Using the prebuilt workflow apps	296
What is Alteryx Connect, and how does it help DataOps?	282	Creating a custom-built data source with the Connect APIs	298
What is Connect?	283	Data nexus	302
Areas of the Connect interface	284	Syncing the Connect data dictionary with other data catalogs	303
Using Connect for DataOps	285	Using the Connect API methods	303
Publishing the data lineage to Alteryx Connect	291	Tableau Data Dictionary API example	305
Loading metadata directly from Connect	291	Summary	307

13

Conclusion

The Alteryx data engineer	309	Governance of DataOps with Alteryx	320
The functional steps in DataOps	313	Testing workflows and outputs	320
Sourcing the data	314	Monitoring DataOps and managing changes	322
Data processing and transformations	315	Securing and managing access	323
Destination management	315	Making data easy to use and discoverable with Alteryx	325
Extracting value from data	317	Our Alteryx data pipeline	325
Beginning advanced analytics	318	Final summary	327

Index

Other Books You May Enjoy

Technical requirements	205	Monitoring workflows with the Message tool	206
Workflow tests and messages	206		