Data Analytics in Internal Audit

Applying data analytics in the Internal Audit Function of ING using KNIME

Tjasse Biewenga
KNIME Data Talk, May 2021

Open for Innovation ®
KNIME

do your thing

Data Analytics – KNIME - Internal Audit - ING

- ING is a global financial institution with a strong European base, offering retail and wholesale banking services to customers in over 40 countries. Bank's Headquarters are in Amsterdam (The Netherlands)

- ING has more than 55,000 employees. We have more than 39 million customers, with 13.9 million considered primary customers.

- Top 10 of European banks

- "ING has set the ambition to become a **data-driven organization** with **analytics as a business model**"

Data Analytics – KNIME - Internal Audit - ING

Our Audit World

- 400+ staff, 35 teams, 20+ locations

- Global / local, expert teams

- Number of Auditable Entities: 2.781
- Average cycle: 2,6 yrs
- Duration audit: 8 - 10 weeks

- Number of Audits 2021: 450+

In >60% of audits analytics will be applied

Data Analytics – KNIME - Internal Audit - ING

Audits 2021

- Many different topics

- Many audits per topic in different locations

Audit Analytics

- Different tests for different topics

- Same tests with different data sources

Our Analytics Approach

- Data Analytics indispensable part of every audit

- All staff proficient in applying Analytics in their audit

- Expert role for Advanced Analytics and Support

Data Analytics – KNIME - Internal Audit - ING

People

- Auditors with DA skills (perform DA with least dependency)

Audit Process

- Large variety of audit topics and data sources
- Relatively short timelines
- Automation of repetitive tasks

Tool Requirements

- Friendly interface => intuitive UI
- Automation made easy => no scripting

- Re-usability => sharing and collaboration
- Flexibility and integration =>
                          file types, databases, APIs
- Scalable => no limits for data volume

- Low license costs => expected >350 users

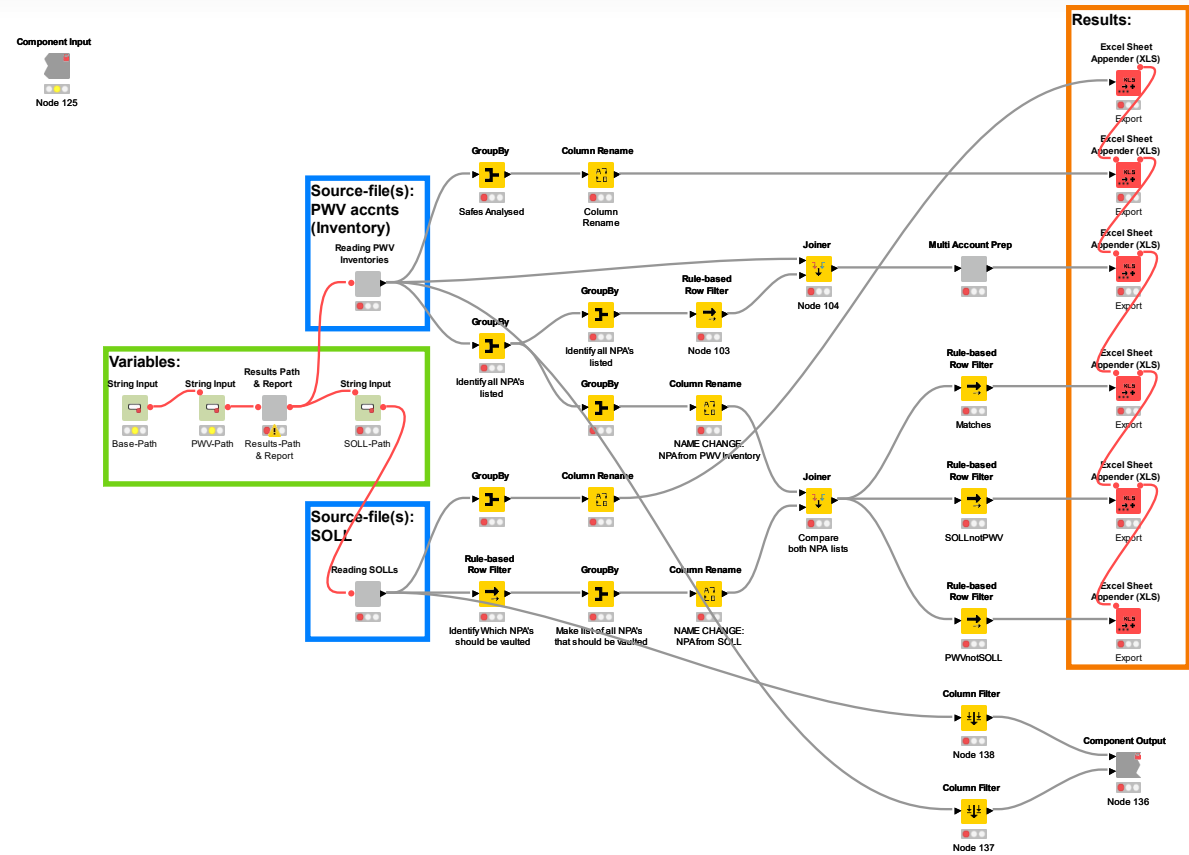# Data Analytics – KNIME - Internal Audit - ING
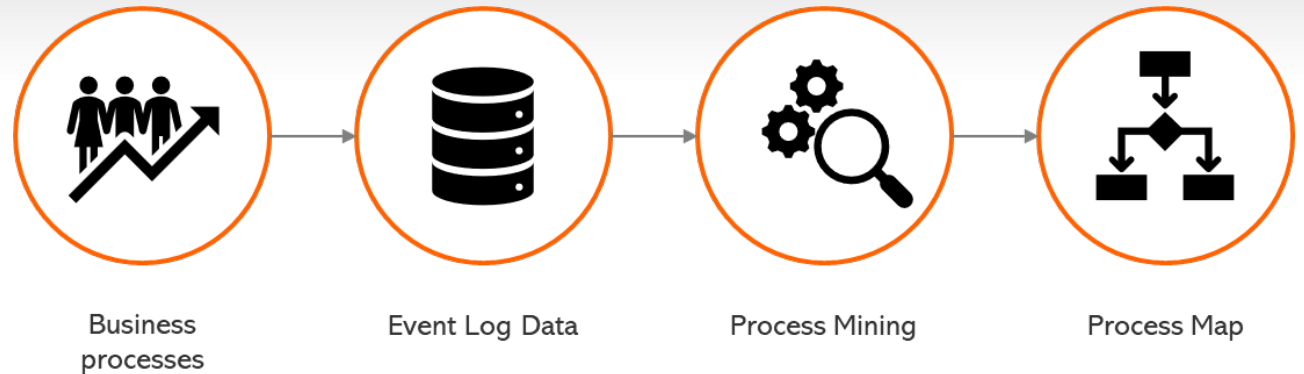
KNIME Analytics Platform

KNIME Server

Use Case

Data Analytics – KNIME - Internal Audit - ING

Use Case Process Mining for KNIME Server

Business processes → Event Log Data → Process Mining → Process Map

Analyzing Event Logs:

- Transaction id's ("Cases")
- Process steps ("Events")
- Attributes to that events (e.g. timestamps, resources)

Use cases primarily:

- Process Discovery

Tools:

- CAS Process Mining node in KNIME Server

ING CAS KNIME WebPortal

Process steps

Upload Eventlog — Validation — Assign Columns — Time Analysis — Transition Analysis — Trace Analysis — Download Results

Download file Full Path Analysis
Download

Back

Cancel   Next

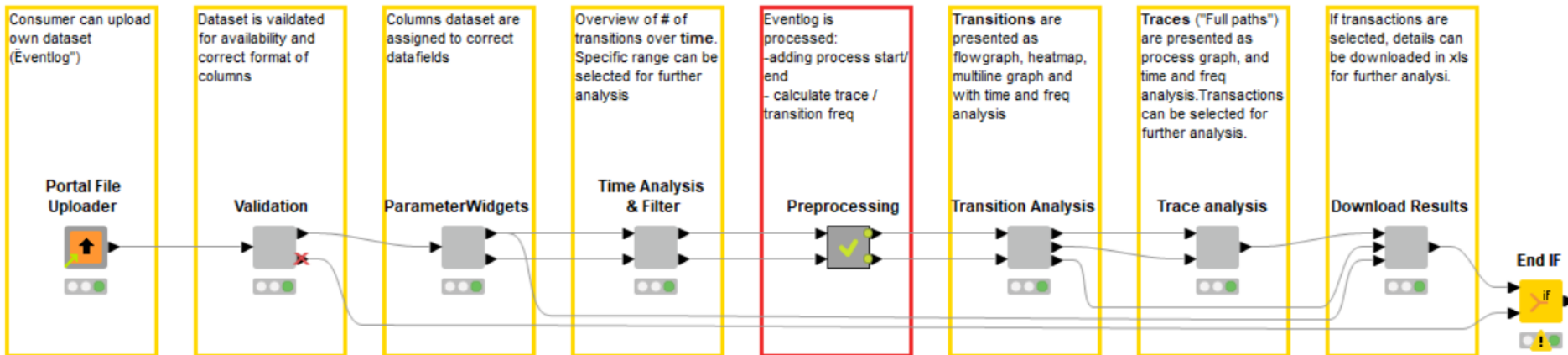# Data Analytics – KNIME - Internal Audit - ING

Process Mining

The purpose of the node is to present information regarding the process flow of a process dataset.

The main output of the node is an interactive view containing a dashboard with differenet kind of information and interaction.

The input columns should be indicated by the user, thus the proper configuration is used.

Data are prepared to present clear transition from one status to another, together with transition duration. Initiating and Ending statuses are also hardcoded, and included in the process.
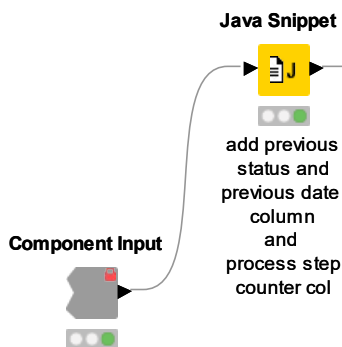

Consumer can upload own dataset ("Ëventlog")	Dataset is vaildated for availability and correct format of columns	Columns dataset are assigned to correct datafields	Overview of # of transitions over **time**. Specific range can be selected for further analysis	Eventlog is processed: -adding process start/ end - calculate trace / transition freq	**Transitions** are presented as flowgraph, heatmap, multiline graph and with time and freq analysis	**Traces** ("Full paths") are presented as process graph, and time and freq analysis.Transactions can be selected for further analysis.	If transactions are selected, details can be downloaded in xls for further analysi.
Portal File Uploader	**Validation**	**ParameterWidgets**	**Time Analysis & Filter**	**Preprocessing**	**Transition Analysis**	**Trace analysis**	**Download Results**

End IF

Technical details

For process graph Generic JavaScript View node is used with JS Mermaid module (see: https://mermaid-js.github.io/mermaid/#/).

Use Case Process Mining for KNIME Server

Java Snippet

add previous
status and
previous date
column
and
process step
counter col

Component Input

Dialog - 3:1429:0:1420:0:1400 - Generic JavaScript View (javascript view)

File

JavaScript View | Image Generation | Templates | Flow Variables

Maximum number of rows: 10.000   [Create Template...]

Flow Variables
- mermaid-js
- knime.workspace

Dependencies
- ☐ D3 - Version 4.2.6
- ☐ Plotly.js - Version 1.47.4
- ☑ jQuery - Version 3.1.1
- ☑ jQuery UI - Version 1.12.1

CSS

```css
 1  /* Flowchart variables */
 2  /* Sequence Diagram variables *
 3  /* Gantt chart variables */
 4
 5
 6  .mermaid .label {
 7      color: #333;
 8  }
 9  .node rect,
10  .node circle,
11  .node ellipse,
12  .node polygon {
13  /* colors the node */
14      fill: #DEDEDE;
15      stroke: #CCCCFF;
16      stroke-width: 1px;
17  }
18  .edgePath .path {
19      stroke: #24AEA6;
20  }
21  .edgeLabel {
22      background-color: #e8e8e8;
23      color:#ff780a;
```

JavaScript

```javascript
48          return dictnod;
49  }
50
51  function filterData (range, columnIndex) {
52      let min = range[0];
53      let max = range[1];
54      var idStart = '';
55      var idEnd = '';
56      var s = 'graph TD\n'; // used to build the code for the process
57      var ids = {}; // used to store all the different processes ids and their unique identificator th
58      var rec = knimeDataTable.getRows();
59      var counter = 0; // used for assigning unique identificators to nodes
60      var ddd = groupitby(min,max);
61  //    for (var i=0; i< knimeDataTable.getNumRows(); i++){
62  //        var nodeFrom = rec[i].data[0];
63  //        var nodeTo = rec[i].data[1];
64  //        var occur = rec[i].data[2];
65
66      for(var key in ddd){
67          var nodeFrom = ddd[key].nodeFrom;
68          var nodeTo = ddd[key].nodeTo;
69          var occur = ddd[key].occr;
70          var normfreq = ddd[key].normfreq;
```

Output Flow Variables

Replace	Flow Variable	Output Type	Script Field	Default Value

[Add] [Remove]

[OK] [Apply] [Cancel]

Data Analytics – KNIME - Internal Audit - ING

The End

Thank you!