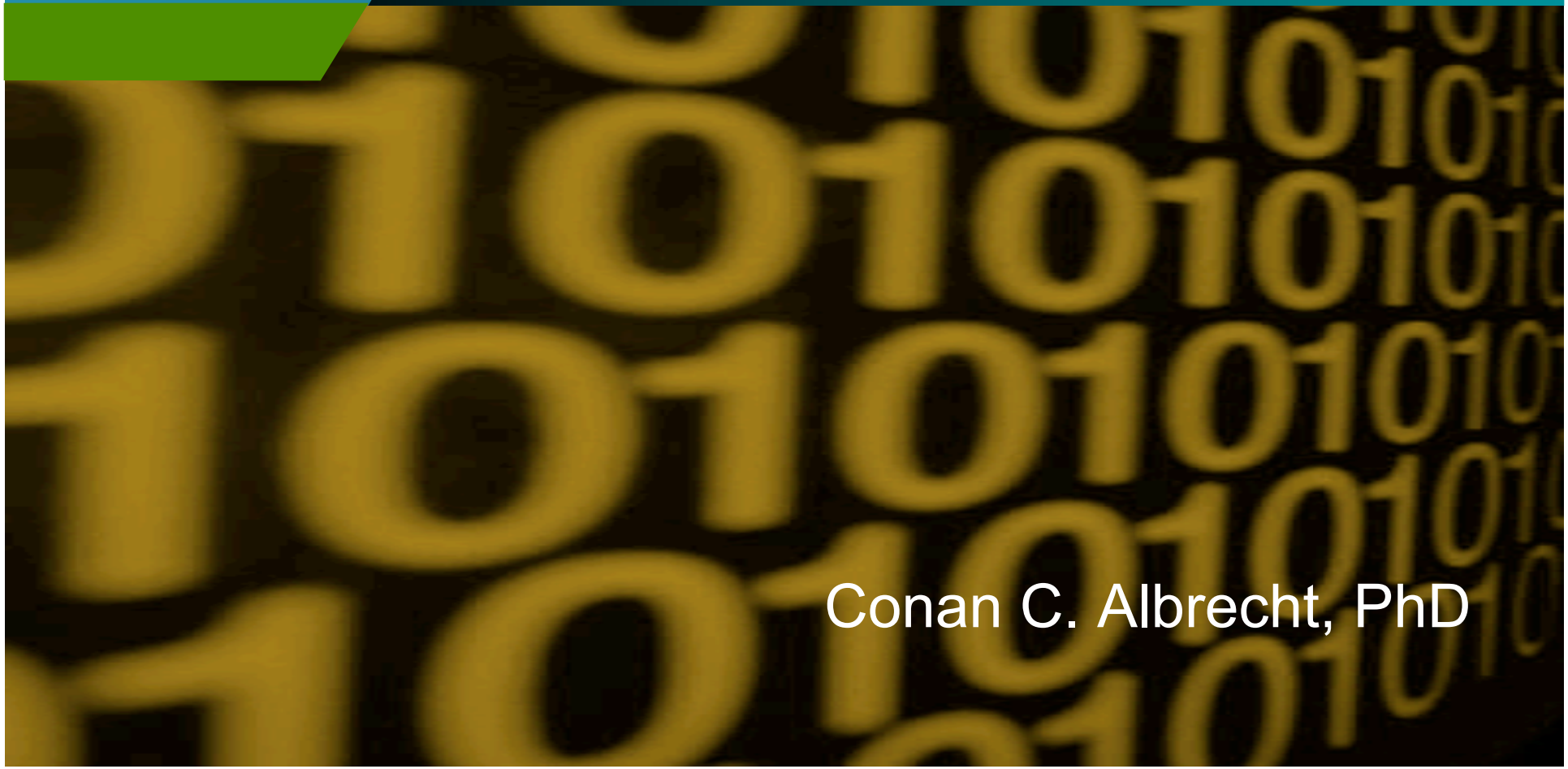
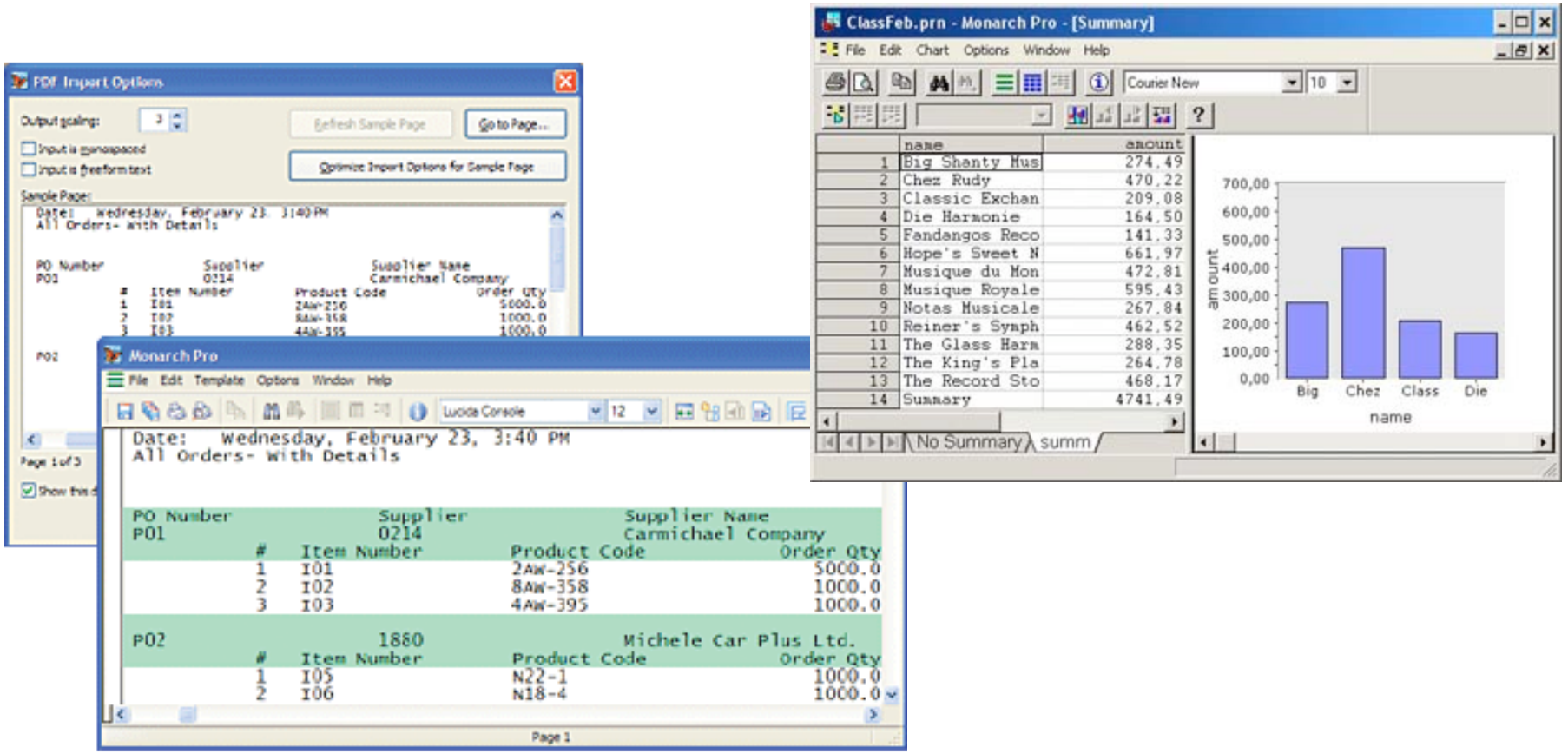
Popular Analysis Tools



Conan C. Albrecht, PhD

Monarch

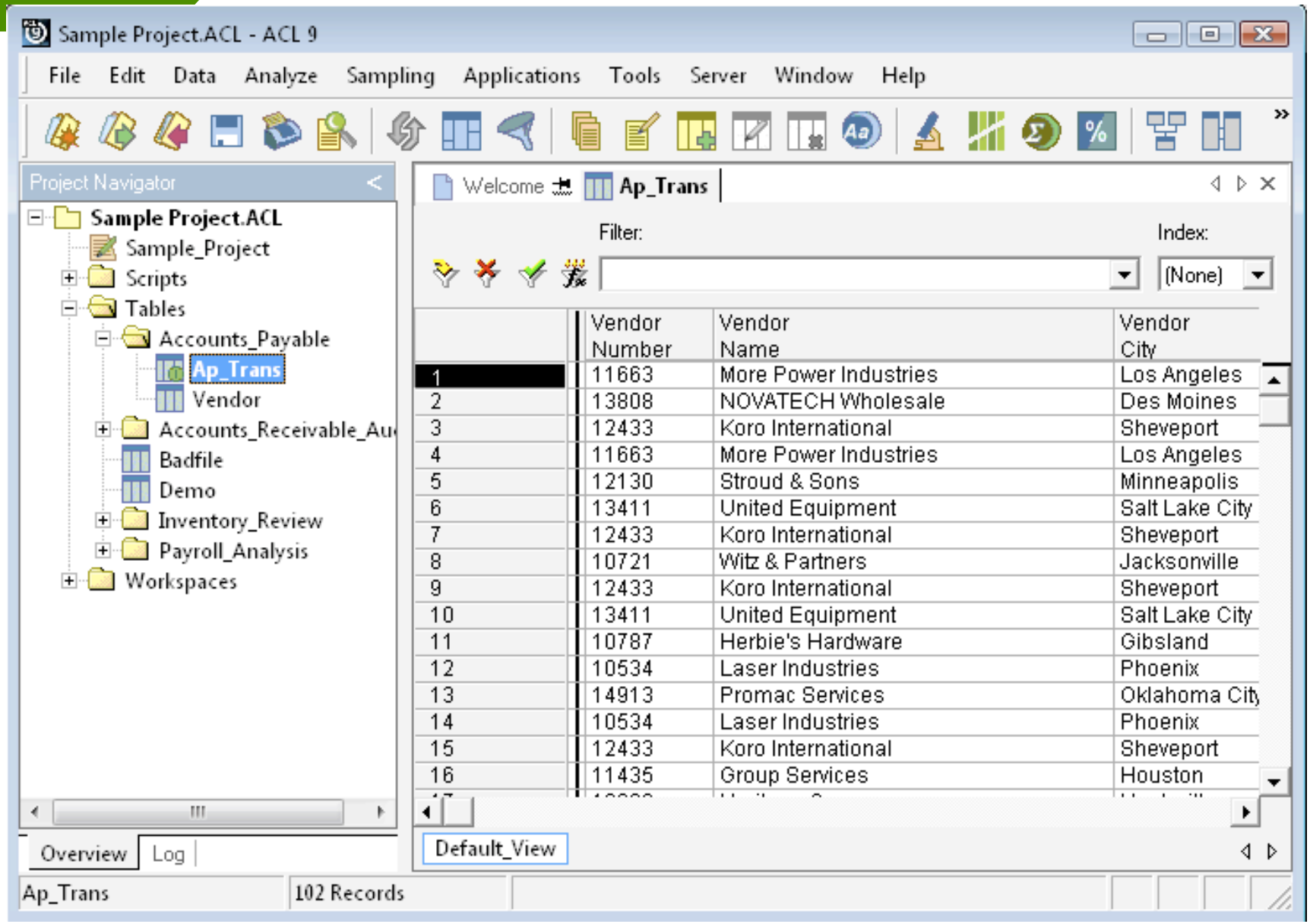

PDF Import Options
Output scaling:
Refresh Sample Page
Go to Page...
Input is monospaced
Input is freeform text
Optimize Import Options for Sample Page
Sample Page:
Monarch Pro
File Edit Template Options Window Help
Lucida Console
12
Date: Wednesday, February 23, 3:40 PM
All Orders- With Details
PO Number Supplier Supplier Name
P01 0214 Carmichael Company
Item Number Product Code Order Qty
1 I01 2AW-256 5000.0
2 I02 8AW-358 1000.0
3 I03 4AW-395 1000.0
P02 1880 Michele Car Plus Ltd.
Item Number Product Code Order Qty
1 I05 N22-1 1000.0
2 I06 N18-4 1000.0
Page 1
ClassFeb.prn - Monarch Pro - [Summary]
File Edit Chart Options Window Help
Courier New
10
name amount
1 Big Shanty Hus 274.49
2 Chez Rudy 470.22
3 Classic Exchan 209.08
4 Die Harmonie 164.50
5 Fandangos Reco 141.33
6 Hope's Sweet N 661.97
7 Musique du Mon 472.81
8 Musique Royale 595.43
9 Notas Musicale 267.84
10 Reiner's Symph 462.52
11 The Glass Harm 288.35
12 The King's Pla 264.78
13 The Record Sto 468.17
14 Summary 4741.49
No Summary
summ
700,00
600,00
500,00
400,00
300,00
200,00
100,00
0,00
amount
Big
Chez
Class
Die
name

Monarch

- Pros
  - Most popular import utility from “printed” reports
  - Imports PDF, text, HTML, DBF, MDB, XLS, etc.
  - ODBC compliant
  - Exports to most applications
  - Handles master/detail reports
- Cons
  - Can also be done with free regular expressions

~$500 range

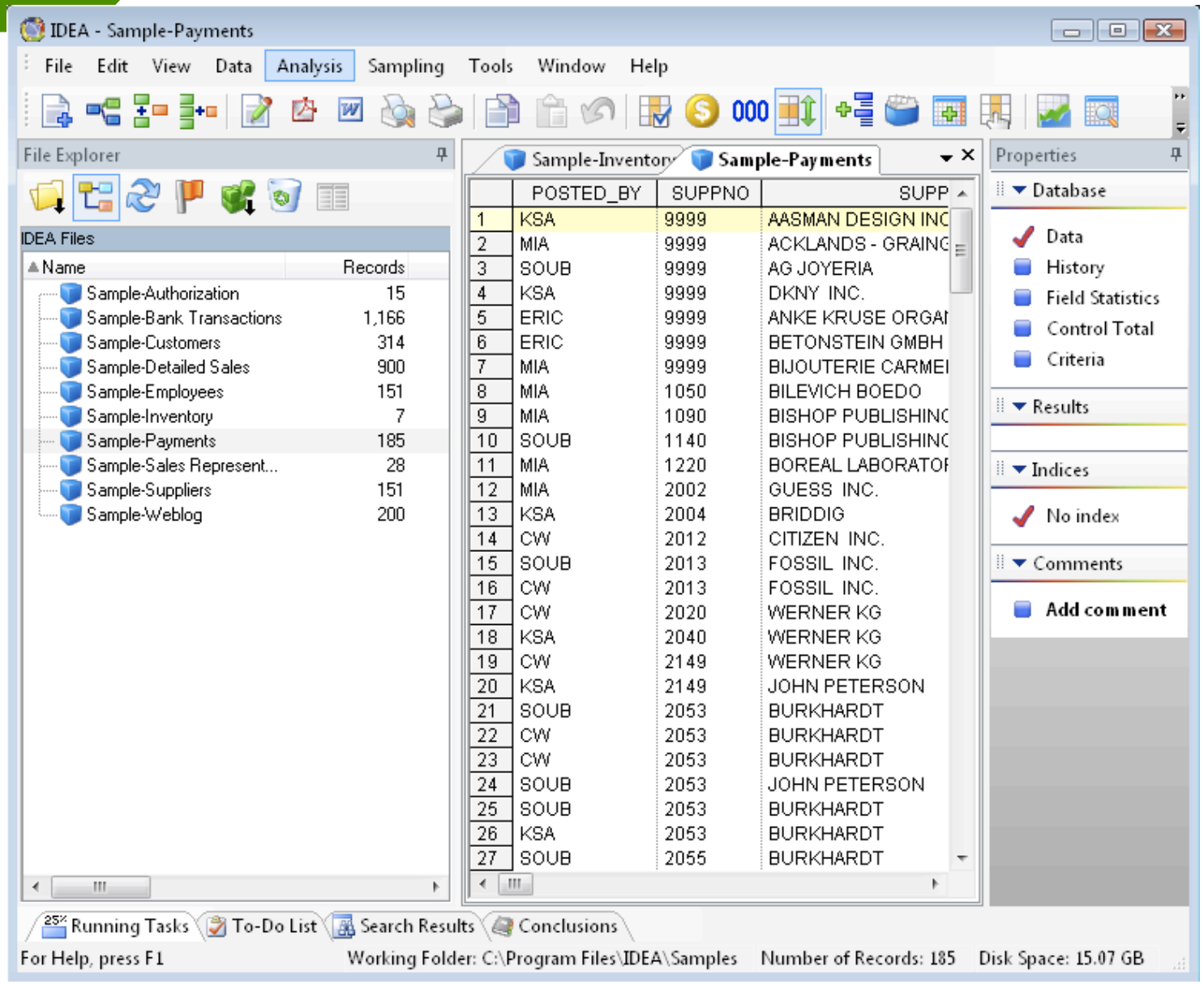
ACL


Sample Project.ACL - ACL 9
File Edit Data Analyze Sampling Applications Tools Server Window Help
Project Navigator
Sample Project.ACL
Sample_Project
Scripts
Tables
Accounts_Payable
Ap_Trans
Vendor
Accounts_Receivable_Au
Badfile
Demo
Inventory_Review
Payroll_Analysis
Workspaces
Overview
Log
Welcome
Ap_Trans
Filter:
Index:
(None)
Vendor Number
Vendor Name
Vendor City
1 11663 More Power Industries Los Angeles
2 13808 NOVATECH Wholesale Des Moines
3 12433 Koro International Sheveport
4 11663 More Power Industries Los Angeles
5 12130 Stroud & Sons Minneapolis
6 13411 United Equipment Salt Lake City
7 12433 Koro International Sheveport
8 10721 Witz & Partners Jacksonville
9 12433 Koro International Sheveport
10 13411 United Equipment Salt Lake City
11 10787 Herbie's Hardware Gibsland
12 10534 Laser Industries Phoenix
13 14913 Promac Services Oklahoma City
14 10534 Laser Industries Phoenix
15 12433 Koro International Sheveport
16 11435 Group Services Houston
Default_View
Ap_Trans
102 Records

ACL

- Pros
  - Most popular software package
  - ODBC compliant
  - Handles huge databases
  - ACLScript
    - Powerful scripting language
- Cons
  - Primarily written for audit work, not specifically for fraud work
    - Although a fraud module exists

IDEA


IDEA - Sample-Payments
File Edit View Data Analysis Sampling Tools Window Help
File Explorer
IDEA Files
Name Records
Sample-Authorization 15
Sample-Bank Transactions 1,166
Sample-Customers 314
Sample-Detailed Sales 900
Sample-Employees 151
Sample-Inventory 7
Sample-Payments 185
Sample-Sales Represent... 28
Sample-Suppliers 151
Sample-Weblog 200
Sample-Inventory
Sample-Payments
POSTED_BY SUPPNO SUPP
1 KSA 9999 AASMAN DESIGN INC
2 MIA 9999 ACKLANDS - GRAINC
3 SOUB 9999 AG JOYERIA
4 KSA 9999 DKNY INC.
5 ERIC 9999 ANKE KRUSE ORGAI
6 ERIC 9999 BETONSTEIN GMBH
7 MIA 9999 BIJOUTERIE CARMEI
8 MIA 1050 BILEVICH BOEDO
9 MIA 1090 BISHOP PUBLISHINC
10 SOUB 1140 BISHOP PUBLISHINC
11 MIA 1220 BOREAL LABORATOI
12 MIA 2002 GUESS INC.
13 KSA 2004 BRIDDIG
14 CW 2012 CITIZEN INC.
15 SOUB 2013 FOSSIL INC.
16 CW 2013 FOSSIL INC.
17 CW 2020 WERNER KG
18 KSA 2040 WERNER KG
19 CW 2149 WERNER KG
20 KSA 2149 JOHN PETERSON
21 SOUB 2053 BURKHARDT
22 CW 2053 BURKHARDT
23 CW 2053 BURKHARDT
24 SOUB 2053 JOHN PETERSON
25 SOUB 2053 BURKHARDT
26 KSA 2053 BURKHARDT
27 SOUB 2055 BURKHARDT
Properties
Database
Data
History
Field Statistics
Control Total
Criteria
Results
Indices
No index
Comments
Add comment
Running Tasks
To-Do List
Search Results
Conclusions
For Help, press F1
Working Folder: C:\Program Files\IDEA\Samples
Number of Records: 185
Disk Space: 15.07 GB

IDEA

- Pros
  - Popular with auditors
  - ODBC compliant
  - Handles huge databases
  - Scripting support
- Cons
  - Primarily written for audit work, not specifically for fraud work

Excel + ActiveData

- Pros
  - Works within Excel
    - Easy to learn
  - Brings many of IDEA and ACL's functions to Excel
  - Cheaper than IDEA or ACL

- Cons
  - Works within Excel
    - 65K record limit

~$500 range

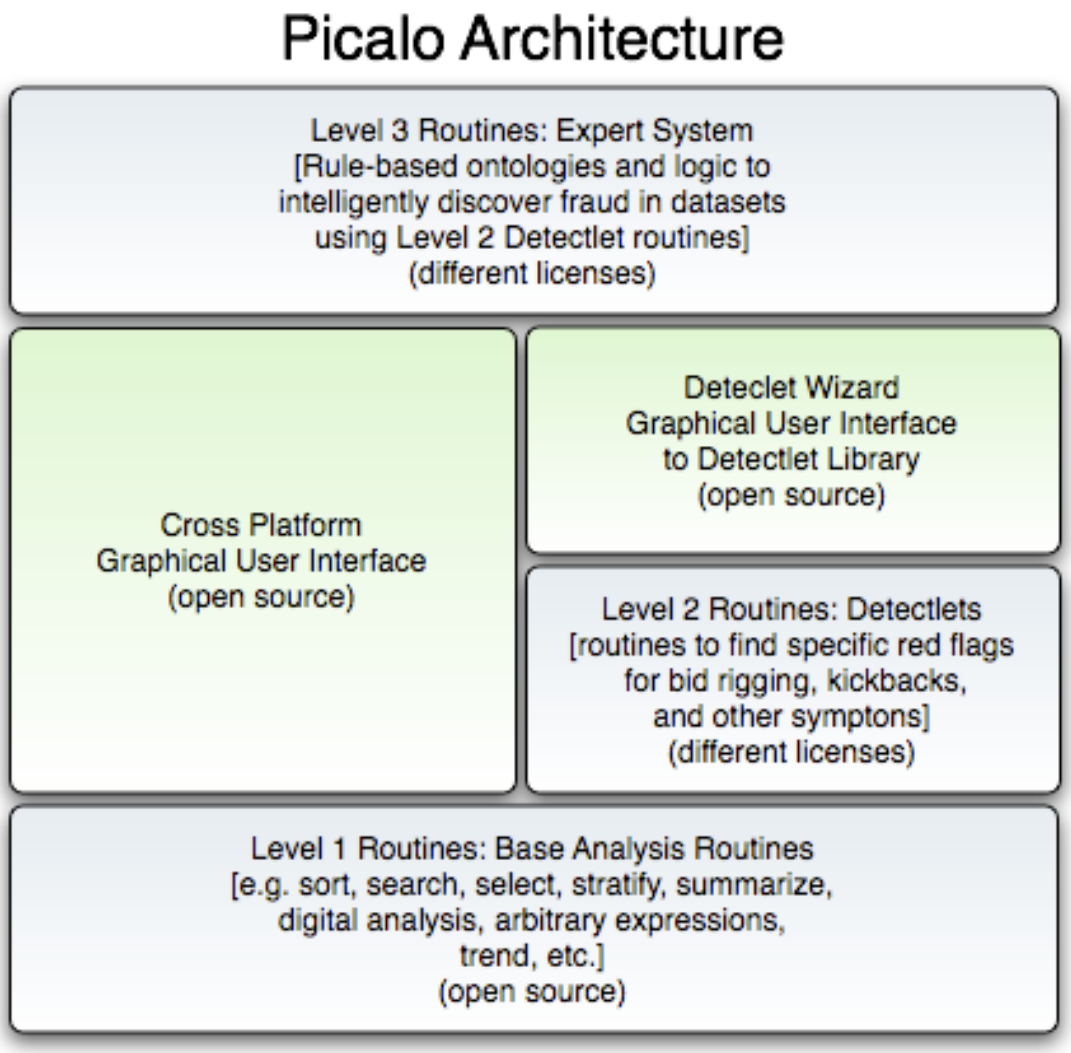
Microsoft Access

- Pros
  - Easy database
  - Real database
    - SQL, multiple tables
  - Visual queries
  - Excellent ODBC front end
  - Reporting, Forms
- Cons
  - 250K record limit on internal database
    - Put SQL Server behind it, and you get around this
  - General application - not specific to data analysis

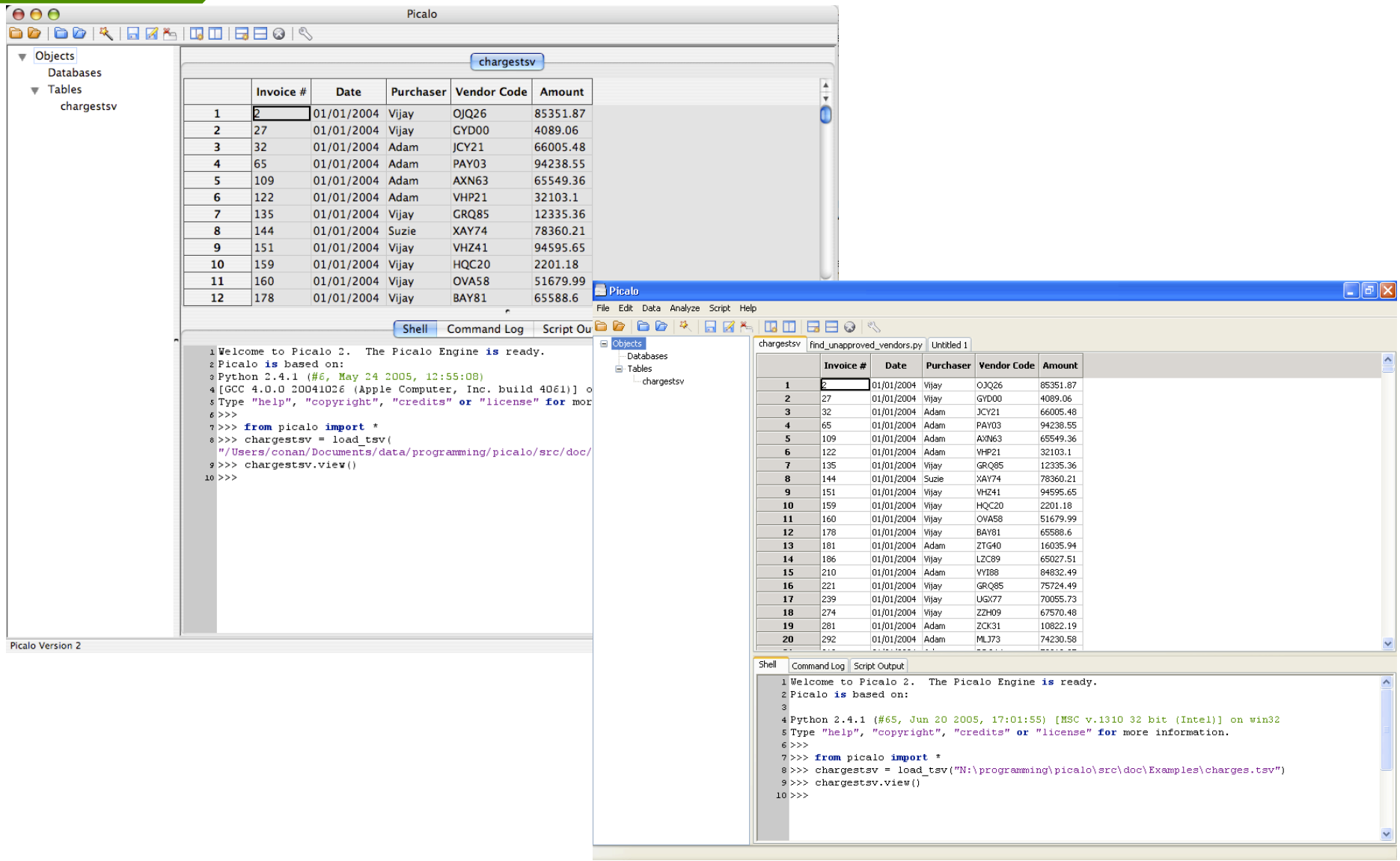
Picalo

- Data analysis toolkit with emphasis on fraud detection
- Community-supported
- Detectlets
  - Bridge the gap between non-technical analysts and technical analysts
  - Working toward an AI-based, automatic fraud detection program
- Get it free from http://www.picalo.org/

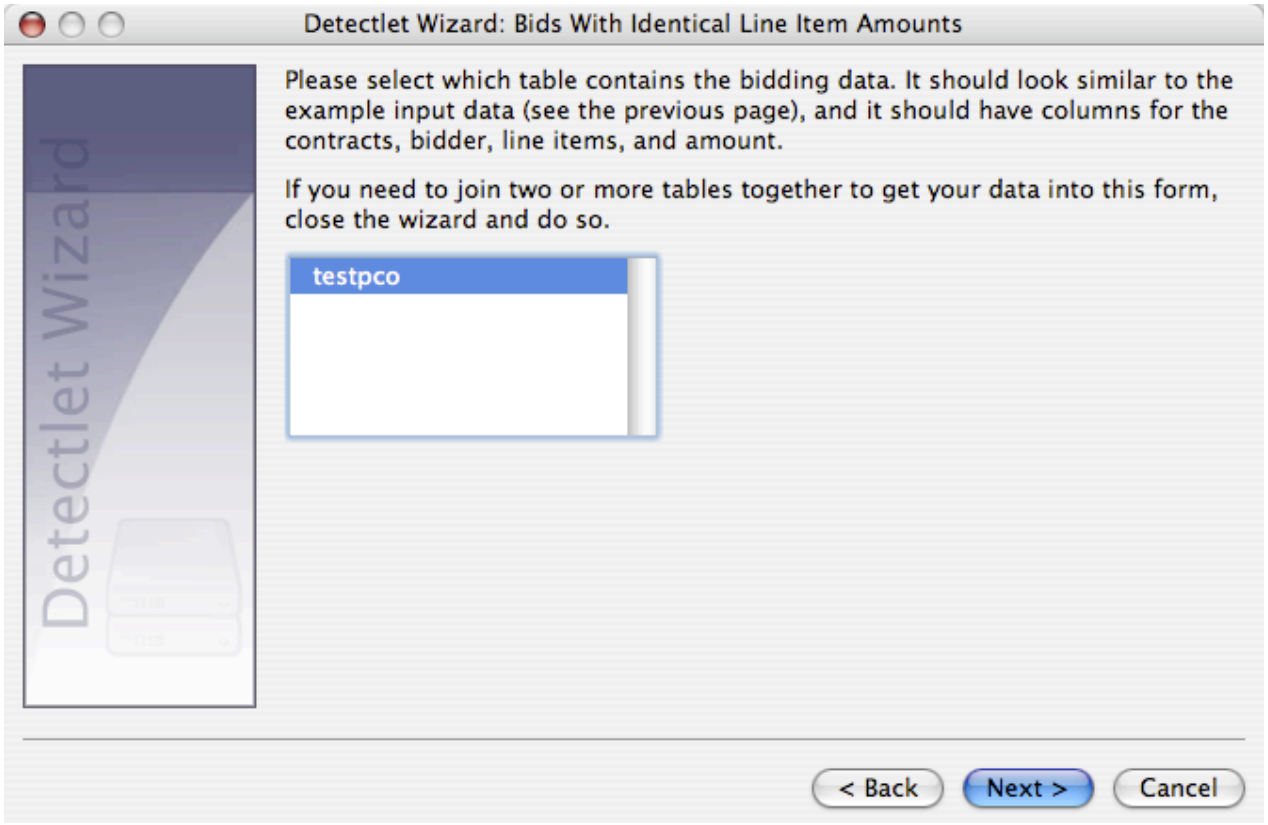
Picalo


Picalo Architecture
Level 3 Routines: Expert System
[Rule-based ontologies and logic to
intelligently discover fraud in datasets
using Level 2 Detectlet routines]
(different licenses)
Cross Platform
Graphical User Interface
(open source)
Deteclet Wizard
Graphical User Interface
to Detectlet Library
(open source)
Level 2 Routines: Detectlets
[routines to find specific red flags
for bid rigging, kickbacks,
and other symptons]
(different licenses)
Level 1 Routines: Base Analysis Routines
[e.g. sort, search, select, stratify, summarize,
digital analysis, arbitrary expressions,
trend, etc.]
(open source)

Picalo


Picalo
Objects
Databases
Tables
chargestsv
chargestsv
Invoice # Date Purchaser Vendor Code Amount
1 2 01/01/2004 Vijay OJQ26 85351.87
2 27 01/01/2004 Vijay GYD00 4089.06
3 32 01/01/2004 Adam JCY21 66005.48
4 65 01/01/2004 Adam PAY03 94238.55
5 109 01/01/2004 Adam AXN63 65549.36
6 122 01/01/2004 Adam VHP21 32103.1
7 135 01/01/2004 Vijay GRQ85 12335.36
8 144 01/01/2004 Suzie XAY74 78360.21
9 151 01/01/2004 Vijay VHZ41 94595.65
10 159 01/01/2004 Vijay HQC20 2201.18
11 160 01/01/2004 Vijay OVA58 51679.99
12 178 01/01/2004 Vijay BAY81 65588.6
Shell
Command Log
Script Ou
1 Welcome to Picalo 2. The Picalo Engine is ready.
2 Picalo is based on:
3 Python 2.4.1 (#6, May 24 2005, 12:55:08)
4 [GCC 4.0.0 20041026 (Apple Computer, Inc. build 4061)] o
5 Type "help", "copyright", "credits" or "license" for mor
6 >>>
7 >>> from picalo import *
8 >>> chargestsv = load_tsv(
"/Users/conan/Documents/data/programming/picalo/src/doc/
9 >>> chargestsv.view()
10 >>>
Picalo Version 2
Picalo
File Edit Data Analyze Script Help
Objects
Databases
Tables
chargestsv
chargestsv
find_unapproved_vendors.py
Untitled 1
Invoice # Date Purchaser Vendor Code Amount
1 2 01/01/2004 Vijay OJQ26 85351.87
2 27 01/01/2004 Vijay GYD00 4089.06
3 32 01/01/2004 Adam JCY21 66005.48
4 65 01/01/2004 Adam PAY03 94238.55
5 109 01/01/2004 Adam AXN63 65549.36
6 122 01/01/2004 Adam VHP21 32103.1
7 135 01/01/2004 Vijay GRQ85 12335.36
8 144 01/01/2004 Suzie XAY74 78360.21
9 151 01/01/2004 Vijay VHZ41 94595.65
10 159 01/01/2004 Vijay HQC20 2201.18
11 160 01/01/2004 Vijay OVA58 51679.99
12 178 01/01/2004 Vijay BAY81 65588.6
13 181 01/01/2004 Adam ZTG40 16035.94
14 186 01/01/2004 Vijay LZC89 65027.51
15 210 01/01/2004 Adam VYI88 84832.49
16 221 01/01/2004 Vijay GRQ85 75724.49
17 239 01/01/2004 Vijay UGX77 70055.73
18 274 01/01/2004 Vijay ZZH09 67570.48
19 281 01/01/2004 Adam ZCK31 10822.19
20 292 01/01/2004 Adam MLJ73 74230.58
Shell
Command Log
Script Output
1 Welcome to Picalo 2. The Picalo Engine is ready.
2 Picalo is based on:
3
4 Python 2.4.1 (#65, Jun 20 2005, 17:01:55) [MSC v.1310 32 bit (Intel)] on win32
5 Type "help", "copyright", "credits" or "license" for more information.
6 >>>
7 >>> from picalo import *
8 >>> chargestsv = load_tsv("N:\programming\picalo\src\doc\Examples\charges.tsv")
9 >>> chargestsv.view()
10 >>>

Picalo


Detectlet Wizard: Bids With Identical Line Item Amounts
Detectlet Wizard
Please select which table contains the bidding data. It should look similar to the example input data (see the previous page), and it should have columns for the contracts, bidder, line items, and amount.
If you need to join two or more tables together to get your data into this form, close the wizard and do so.
testpco
< Back
Next >
Cancel

Picalo

- Pros
  - Scripting-based
    - Powerful
  - Based in the free Python language
    - Language has a lot of features and history
  - Free
  - ODBC compliant
- Cons
  - Scripting-based
  - No support

Specialized Tools

- SAS and SPSS
  - Statistical packages with fraud modules
  - Useful in specialized areas
- Scripting languages
  - Perl, VB.net, VBA, Ruby, Python
  - More difficult to learn, but useful for any type of automation
- Full-text search tools
  - WinGREP, Lucine, Sphinx
  - Good for email analysis, full text search
- Database packages
  - MySQL, PostgreSQL, SQL Server, etc.
  - Very useful when combined with a front end like ACL or IDEA