


Data Mining and Governmental Accounting Applications

Dennis E. Peachey
 Governor's Office of the Budget
 Commonwealth of Pennsylvania
 dpeachey@state.pa.us



Data Mining is...

- The process of analyzing a specific data set with the intent to summarize it or distill it to reveal useful data
 - Decision-making – Are our internal controls effective?
 - Monitoring – Is depreciation posting properly?
 - Forensic accounting – Look for evidence of fraud
 - Pre-auditing – Find problems before the auditors do
- Approach Continuum:
 - Ad hoc analysis → Develop review methodology → Continuous Monitoring

2

Growth of Data Stores

- ERP produces massive amounts of data
- Larger data files measured in terabytes
- Pennsylvania's application of SAP:
 - Several different modules:
 - FM - Budgetary accounting
 - FI - Financial accounting
 - CO - Cost accounting
- 26,768 tables
- Several tables currently have more than 215 million records

3

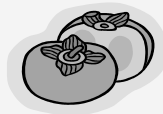
Data Mining Tools - Requirements

- Importing – Ability to extract a user-defined set of data
- Analyzing – Examine the data set for patterns, matches, anomalies
- Parsing – Dividing pieces of existing data into smaller parts
- Matching – Finding identical pieces of data in separate data sets
- Summarizing – Grouping data on one or more characteristics
- Aging – Summarizing data based on a chronological sequence
- Stratifying – Splitting data into distinct layers
- Verifying – Recalculate depreciation, etc
- Reporting – Create useful reports
- Documenting – Maintain a valid audit trail

4

Available Tools

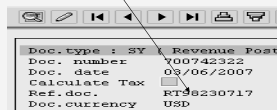
- Microsoft Excel
- Microsoft Access
- ACL
- Others
 - Clementine
 - Data Cruncher
 - Scenario



5

Data Mining With Excel

- Excel limited to approximately 65,000 lines
- Use VLOOKUP function to match and reconcile
- Debits and credits post on different documents in the Cash in Transit account
- Documents match on Ref Doc field



- Need a process to match debit documents with credit documents in the account

6

Using Access to Reconcile Accounts

Access is much more powerful but becomes very slow when table size surpasses 500,000 records

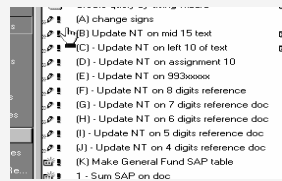
- Example – Using Access queries to reconcile differences between two data sets – BFM does this on a monthly basis
 - Two tables
 - SAP cash postings
 - Treasury cash postings – Treasury does not use SAP so document numbers are not consistent with those in SAP



10

Using Access to Reconcile Accounts

- Treasury documents all post with a positive amount – we use a query to change expenditures to negative amounts
- A series of queries is then used to parse out like document numbers, summarize on document numbers and eliminate records that have like document numbers and equal transaction totals



11

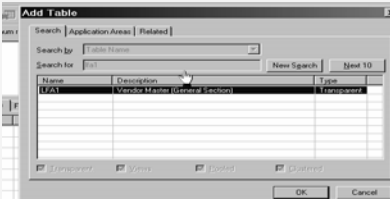
ACL with Direct Link to SAP

- Provides a powerful tool to analyze large amounts of data
 - Import directly from SAP
 - Maintains audit trail
 - Quickly analyze large data tables
 - Allows scripting of commands for commonly repeated tasks

12

Data Mining with ACL

- Cleansing Vendor File of dormant records
 - Query table LFA1



13

Vendor Analysis with ACL



- Allows user to define fields in query (109 fields available in this table)
- Vendor Table Query Results (239,111 vendor records)

14

Vendor Analysis with ACL

- Extract Funds Commitment documents from table FMIOI

Code	Name	Description	SQL Data	ACL Data	Length
1	Name				
2	LFA1#	Account number of vendor or creditor	Char	ASCII	1
3	LFA1#	County key	Char	ASCII	1
4	Name 1	Name 1	Char	ASCII	1
5	Name 2	Name 2	Char	ASCII	1
6	Name 3	Name 3	Char	ASCII	1
7	Name 4	Name 4	Char	ASCII	1
8	Name 5	Name 5	Char	ASCII	1
9	Name 6	Name 6	Char	ASCII	1

Code	Name	Description	SQL Data	ACL Data	Length
1	Name				
2	LFA1#	Account number of vendor or creditor	Char	ASCII	1
3	LFA1#	County key	Char	ASCII	1
4	Name 1	Name 1	Char	ASCII	1
5	Name 2	Name 2	Char	ASCII	1
6	Name 3	Name 3	Char	ASCII	1
7	Name 4	Name 4	Char	ASCII	1
8	Name 5	Name 5	Char	ASCII	1
9	Name 6	Name 6	Char	ASCII	1

15

Vendor Analysis with ACL

- Final Step – Find all Vendor records that contain no activity -Relate the two tables from the above steps and remove those vendor master records

16

ACL Audit Trail

- Each action taken by the user in a given project is recorded:

17

Data File Types

- “Flat file” – contains no formatting
- When importing the user must make decisions regarding field length and name

```

1100079274 44000020000000000000 00110000000000000000 LRFPRN
50439058272 40243024000000000000 35-1619INSERTING SYSTEM RETERCFRFO
90407074555 48000020000000000000 00700000000000000000 CEFRFO
90408014324 33808002000000000000 00185000000000000000 CEFRFO
90408014323 33808002000000000000 00180000000000000000 CEFRFO
20599074554 48000020000000000000 00201000000000000000 CEFRFO
20599028616 53110002000000000000 1011800000000000000000 CEFRFO
10821074556 49100020000000000000 00000000000000000000 CEFRFO
10821074557 49100020000000000000 01600000000000000000 CEFRFO
10409028649 78988002000000000000 11477000000000000000 CEFRFO
10111058078 48000020000000000000 00000000000000000000 CEFRFO
11114074558 25003002000000000000 04603000000000000000 CEFRFO
40204028640 13081400200000000000 110-01194000000000000000 CEFRFO
    
```

18
