Data Analytics and Data Access

Michael Kano

Audit Analytics Manager, eBay

Professional Techniques – T33

ISACA®
Trust in, and value from, information systems
San Francisco Chapter

Agenda

- Importance of Data Access
- The Current Data Environment
- Role of Audit Leadership
- Role of Data Analysis Specialist
- Documentation

THE IMPORTANCE OF DATA ACCESS

ISACA®
Trust in, and value from, information systems
San Francisco Chapter

Importance of Data Access

Foundation for success and growth

- Independence
- Confidence in data
- Expansion of scope
- Increased productivity
- Greater timeliness

THE CURRENT DATA ENVIRONMENT

ISACA®
Trust in, and value from, information systems
San Francisco Chapter

Complexity

- Diversity of applications
- Vast number of platforms
- Mergers and acquisitions
- Overlaps and gaps
- Massive data volumes

Opacity

- Lack of comprehensive/centralized oversight
- Inconsistent documentation standards
- Strong security requirements
- Lack of portals for data access

Culture

- Not accustomed to audit access
- Not prepared for audit access
- Unaware of power of desktop analytical apps
- Accountability for data security

Summary

- Obstacles are technical, cultural and historical
- Often unexpected or minimized

THE ROLE OF AUDIT LEADERSHIP

ISACA®

Trust in, and value from, information systems

San Francisco Chapter

To-Do List for the CAE

- Embed universal data access in audit charter
- Active, visible support for process owner
- Budget for software, training, servers and security compliance
- Create managerial DA specialist role
- Documented expectations, standards, and timelines for specialist

Audit Charter: Universal Data Access

"The Internal Audit team is authorized by the Audit Committee to:

- Have *unrestricted access* to all premises, corporate records, information and personnel;
- Require any officer of the company to supply such information and explanations as may be needed; …"

Active and Visible Support

- Introduce DA specialist to key data owners
- Reinforce DA specialist's message when challenged
- Acknowledge contributions of data owners

Budget

- Basic software and support
- Software add-ons
- Training
- Dedicated server(s)
- Connectivity
- Physical security
- Logical security

Create DA Specialist Role

- Primary focus on data analysis
- Clear and unique accountability
- Employee preferable
- Knowledgeable
- Experienced

Documented Deliverables

- Priorities
  - Derived from audit plan/risk assessment
  - Review with audit team
- Standards
  - Documentation
  - Repository
- Maintenance
  - Regular review

Summary

- Establish these principles up front
- Leadership needs you to tell them what you need
- Hold leadership accountable©

THE ROLE OF THE DATA ANALYSIS SPECIALIST

ISACA®

Trust in, and value from, information systems

San Francisco Chapter

DA Specialist Role

- Relationships with data owners
- Data access procedures
- Documentation
- Train the IA team
- Develop analytics

Managing Your Data Owner Relationships

- Know the person/role
- Know the data
- Articulate your objectives
- Be prepared to address challenges

Know the Person/Role

- Length of time in current position
- Career history
- Previous audit encounters
- Major achievements
- Accountabilities
- People on his/her team

Know the Data

- Production/warehouse
- Database
- Key tables and fields
- Unusual data types
- PCI/PII/Confidential/Restricted data

Articulate Your Objectives

- Access to the data
- READ and SELECT access
- Sufficient spooler capacity
- Metadata
- Added to distribution list

Your First Contact

- E-mail allows recipient time to consider
- Identify yourself and your mission
- Identify what data you need
- Explain how you will access/extract data
- Identify supporting features of your DA tool
- Request a meeting

Your First Meeting

- Recap your e-mail
- Address challenges
- Demonstrate DA tool access method
- Present name of technical contact at tool HQ
- Explain audit data retention policy
- Ask for name of key contact

Typical Challenges and Your Response (1)

The Challenges	Your Responses
"We're concerned about the integrity of our data."	"Our application is read-only, and we cannot write to the source database."
"We're worried that the data may be vulnerable in your server."	"Here are the security precautions we will be following."
"Why don't you just ask us for the reports you need?."	"We want to minimize the impact on your department's resources."

Typical Challenges and Your Response (2)

The Challenges	Your Responses
"This is going to add to network traffic."	"Let us know the best days/times for extracting the data."
"This is going to add to network traffic."	"How about monitoring our use and placing it in a low priority?"
"This is going to add to network traffic."	"We'll be filtering the data and pulling only the fields we need."

Summary

- Non-technical component is critical to your success
- Focus on building relationships
- Do your research on data and people
- Be prepared to address challenges

DOCUMENTATION

What To Include

- Corporate data matrix
- Technical documentation
- User documentation
- Data source guide

Corporate Data Matrix

- Basic information by BU and process
- Working document basis for data source documents
- Good document for IA team to learn data geography

Corporate Data Matrix Example

	Item	Org 1	Org 2	Org 3
Business Area	**Accounts Payable**			
	System Name	Oracle Financials	Teradata	SAP
	Resource(s)	Diana Delphi	Vishal Gupta	Hans Datameister
	Database	OrcFin1	TD_A100	SAP_AP_057
	Data Dictionary	www.corpdata.com/OracleFin.pdf	None	None
	Access Method	ODBC-Arbutus/ACL	Teradata SQL Assistant	Direct Link-ACL
	Analytic Application	Arbutus/ACL	Arbutus/ACL	Arbutus/ACL
	Key Tables	AP_AE_Lines_All AP_AE_Headers_All	dim_payables dim_vendors	BSEG, BKPF, LFA1
	Notes	Headers contain dates, lines do not.	Use SQL Joins with filters; extract to .txt for Arbutus/ACL import.	Column titles are truncated when exported from ACL; use alternate names.

DA Application Documentation

- User documentation
  - User guide
  - Sample projects/data
  - Best practices guide
- Technical documentation
  - Specifications
  - Requirements

Data Source Guide

- Expands on basic information in matrix
- Step-by-step data access guide
- Key table layouts
- Create in Word with change date embedded
- Save as PDF for library

Data Source Guide: Key Information

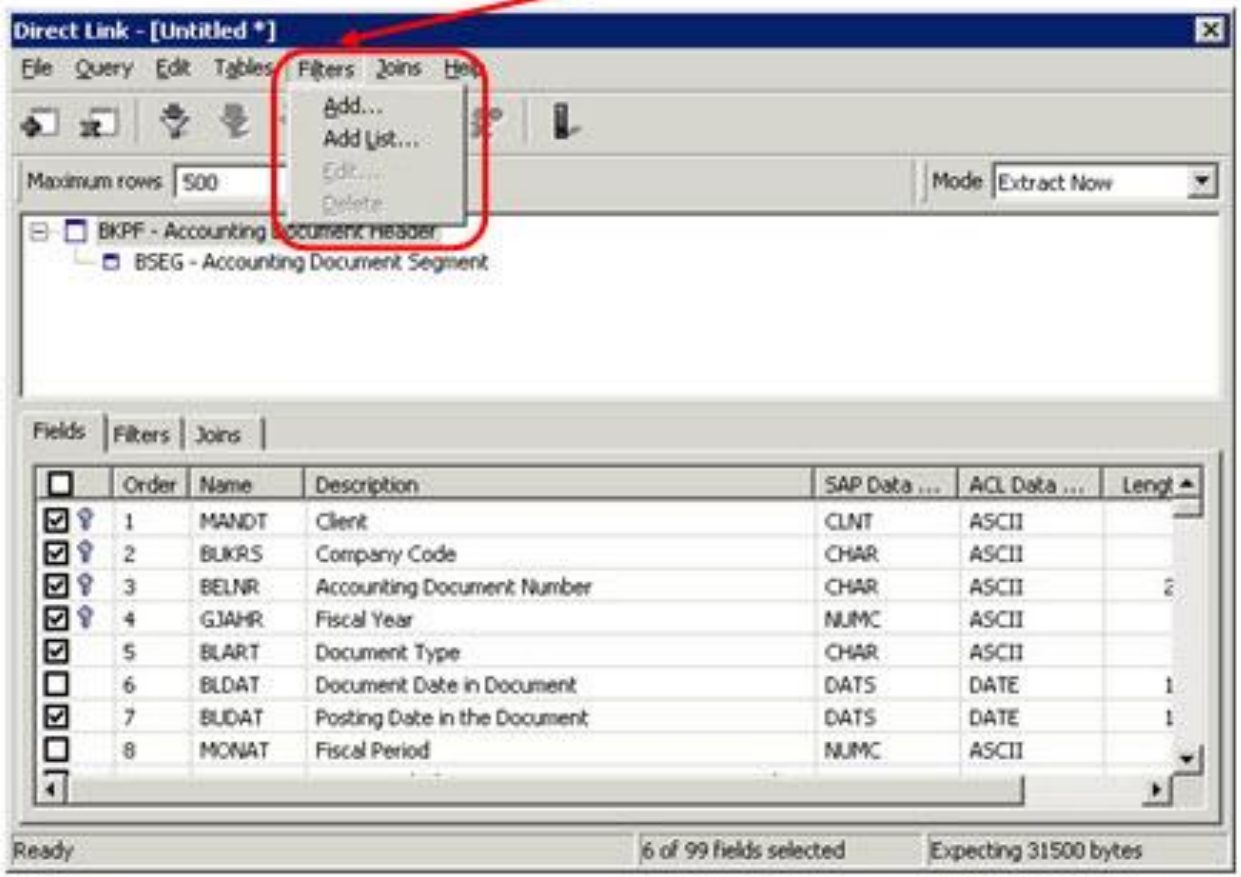
- System name
- Organization
- Key databases
- Data content
- Owner
- Server
- Network access
- Information resources
- Training
- User guide
- Data dictionary
- Key tables and fields
- Table relationships
- Data access methods

Step-by-Step Data Access

- Screen shot for every step
- Use red arrows, boxes to direct user
- Detailed text explaining process step

Step-by-Step Data Access

To add filters to the query, go to Filters >> Add:



Data Import Scripts

- Arbutus/ACL/SQL syntax with comments
- Preserves best practices for data imports

Import Scripts

SCRIPTED IMPORTS: ACL Import Syntax for Delimited Files

Line	Syntax	Explanation
1	IMPORT DELIMITED TO *<destination table name>* "*<destination file name (.fil)>*" FROM "*<source file name and address>*" 0 SEPARATOR "*<field separator/delimiter>*" QUALIFIER "*<text qualifier>*" CONSECUTIVE STARTLINE *<line number>* KEEPTITLE FIELD "*<field name>*" *<field type>* AT *<starting position>* DEC *<number of decimals>* WID *<field length>* PIC "*<format>*" AS ""	SEPARATOR: Recommended field delimiter is pipe "|" STARTLINE: The line on which the data begins. If field/column names are imported, this parameter is usually 2, as the first line is reserved for the column names. KEEPTITLE: Use the original database's field names for the ACL field names. <field type>: Usually C, N, or D (character, numeric, or date) PIC "<format>" is important for date fields. The format should reflect the source database's date format, e.g., "MM/DD/YYYY"

Key Table Layouts

- Arbutus: Use DISPLAY PRIM TO <output file> on command line; writes to file
- ACL: Use DISPLAY on command line
  - Copy + Paste from ACL Command Log to Word/Excel
- Teradata SQL Assistant: Tools >> List Columns
- Add comments where needed

Arbutus Example

Command Log | Detail_Data_TL

Smart S

	Name	Type	CNLD	Start	Length	Decimals	Modifiers
1	Detail_Data_1	DELIMITED		0	9707	0	C:\Users\mkano\Desktop\Folders\KRI Project\Data Analysis\Detail_Data_1.fil
2	TASK_ID	ASCII	C	1	13	0	NATIVE [D]
3	problem_manager	ASCII	C	14	8	0	NATIVE [D]
4	STATUS_LABEL	ASCII	C	22	14	0	NATIVE [D]
5	PRIORITY_LABEL	ASCII	C	36	7	0	NATIVE [D]
6	TITLE	ASCII	C	43	250	0	NATIVE [D]
7	DIRECT_CAUSE_TYPE	ASCII	C	293	8	0	NATIVE [D]
8	DIRECT_CAUSE_SUBTYPE	ASCII	C	301	14	0	NATIVE [D]
9	CUSTOMER_FEATURE	ASCII	C	315	18	0	NATIVE [D]
10	SUBFEATURE	ASCII	C	333	37	0	NATIVE [D]
11	DATABASE_VIP	ASCII	C	370	14	0	NATIVE [D]
12	POOL_NAME	ASCII	C	384	55	0	NATIVE [D]
13	RC_CFG	ASCII	C	439	60	0	NATIVE [D]
14	root_cause_category	ASCII	C	499	45	0	NATIVE [D]
15	jirabugid	ASCII	C	544	17	0	NATIVE [D]
16	TIME_SERVICE_IMPAIRED	DATETIME	D	561	19	0	PICTURE "yyyy-mm-dd hh:mm:ss" NATIVE [D]
17	YearMonth_Impaired	ASCII	C	561	7	0	
18	detection_date_time	DATETIME	D	580	19	0	PICTURE "yyyy-mm-dd hh:mm:ss" NATIVE [D]
19	diagnosis_date_time	DATETIME	D	599	19	0	PICTURE "yyyy-mm-dd hh:mm:ss" NATIVE [D]
20	TIME_SERVICE_RESTORED	DATETIME	D	618	19	0	PICTURE "yyyy-mm-dd hh:mm:ss" NATIVE [D]
21	DETAILED_DESCRIPTION	ASCII	C	637	2485	0	NATIVE [D]
22	PROBLEM_STATMENT	ASCII	C	3122	254	0	NATIVE [D]
23	IMPACT_ASSESMENT	ASCII	C	3376	2000	0	NATIVE [D]
24	ACTIONS_TO_RESTORE	ASCII	C	5376	2000	0	NATIVE [D]
25	ROOT_CAUSE	ASCII	C	7376	254	0	NATIVE [D]
26	REPAIR_ITEMS	ASCII	C	7630	2000	0	NATIVE [D]
27	CODE_BUG	PRINT	N	9630	1	0	DECIMAL '.' NATIVE [D]
28	F26	ASCII	C	9631	1	0	NATIVE [D]
29	c_Diff_Impaired_Restored	COMPUTED	N	0	0	0	
30				0	0	0	TIME_SERVICE_RESTORED - TIME_SERVICE_IMPAIRED
31	c_Root_Cause_Category	COMPUTED	C	0	0	0	
32				0	0	0	"TBD" IF root_cause_category = "TBD"
33				0	0	0	root_cause_category

<< End of File >>

Teradata SQL Assistant Example

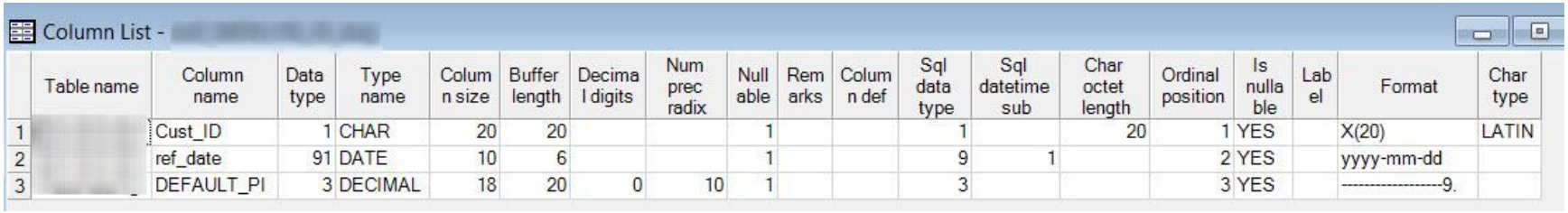

Column List -

	Table name	Column name	Data type	Type name	Column size	Buffer length	Decimal digits	Num prec radix	Nullable	Remarks	Column def	Sql data type	Sql datetime sub	Char octet length	Ordinal position	Is nullable	Label	Format	Char type
1		Cust_ID	1	CHAR	20	20			1			1		20	1	YES		X(20)	LATIN
2		ref_date	91	DATE	10	6			1			9	1		2	YES		yyyy-mm-dd	
3		DEFAULT_PI	3	DECIMAL	18	20	0	10	1			3			3	YES		----------9.	

Summary

- Good documentation is invisible but valuable
- Save time in the future
- Supports audit findings if challenged
- Supports newbie orientation

CLOSING REMARKS

Final Notes

- Maintain relationships with data owners and resources
- Get on distribution lists
- Update documentation regularly
- Proactively seek out data sources for future audits
- Remember: You're investing in the future!

QUESTIONS?

Michael Kano

Audit Analytics Manager

eBay, Inc.

mkano@ebay.com

(480) 862-8347