
Building Audit Trail Data Warehouse from DataFax Journal Files

Hanming Tu
Premier Research Group plc
30 South 17th Street
Philadelphia, PA 19103
Email: Hanming.Tu@premier-research.com
Tel: 215-282-5526; Fax: 215-282-5528
<http://www.premier-research.com/>



Content

- About The Golden Mines
- Built-in DataFax Reports
- DFjfv - Journal File Viewer
- Building Audit Trail Data Warehouse
- Conclusion



Golden Mines - Journal File & Fax Log

- **Journal File**
 - ✓ DataFax server records when images arrive in the system and who did what at what time in journal file.
 - ✓ Separate journal files are kept for each DataFax study, and within a study, a new journal file is created for each month.
 - ✓ The naming scheme for journal files is YYMM.jnl where YY is the last two digits of the year (e.g. 92) and MM is the two digit month of the year.
- **Fax Log**
 - ✓ Every fax received by DFmaster.rpcd is recorded in fax log. The most recently received fax is listed at the end of the file.
 - ✓ Each received fax represents one record in the file. Each entry records the 5 pieces of information: full pathname of fax file, total number of pages in fax, date and time fax was transmitted, sender identification of transmitting fax machine, and archive duration.



Golden Mines - Journal File Records

Journal files contain the following records:

- ✓ a new record arrives in the study database from the incoming daemon
- ✓ an existing record is modified in the validation tool
- ✓ the validation level of an existing record is modified in the validation tool
- ✓ the status of an existing record is modified in the validation tool
- ✓ an existing record is deleted in the validation tool
- ✓ a QC note is added, modified, resolved, or deleted in the validation tool
- ✓ a record is imported into the study database using DFimport.rpc



Golden Mines - Journal File Columns

Field #	Column	Description
1	Date Stamp	A date stamp, in <i>YYMMDD</i> format, identifying when the data record was written to the database
2	Time Stamp	a time stamp, in <i>HHMMSS</i> format, identifying when the data record was written to the database. Hours are reported in 24 hour notation
3	User Name	the username of the person who wrote the record to the database. This is the login name of the user who is modifying the records using DFvalidate or DFimport.rpc .
4	Record Type	Possible values are: d for data record, and q for QC record.
5-11	Record Keys	fields 5 through 11 contain the first 7 fields from the data record.
12+	Record Data Fields	the remaining data fields (8 to the end of the record) follow.



DataFax Built-in Reports - DF_ATcrfs

- DF_ATcrfs tracks keys associated with a CRF page over time
- It can search all the Journal files for all studies to find records matched with the CRF page id
- Journal files will be skipped if access is denied due to OS system permission restriction.
- It reports key history including: date, time, JF DBN, recorded study, pid, visit, plate, level, status, and user.



DataFax Built-in Reports - DF_ATfaxes

- DF_ATfaxes traces validation history of each page in selected faxes
- It reads fax_log and journal files if journal files are specified.
- It reports
 - ✓ Fax arrival: date, time, sender id, page #
 - ✓ Data Record Keys: id, visit/seq, plate
 - ✓ Journal: status, validation level, date, time,



DataFax Built-in Reports - DF_ATmods

- **DF_ATmods** traces database modifications over a time period
- It reads through journal files and could generate thousands pages if the search is not limited.
- It reports the modifications (additions, changes, and deletions) to a study database and QC notes over a specified period of time.



DataFax Journal File Viewer - Why Do We need it

- The built-in reports are helpful in tracking and reviewing individual faxes and changes but could not provide detailed data records
- The DFjfv is a web-based interface allowing users to retrieve journal file records by study, DataFax user, record type, record keys (PID, plate, visit, image id, status, and validation level), time range, and any field and its value.



DataFax Journal File Viewer - Web-based Interface

DataFax Journal File Viewer

Options:	<input type="checkbox"/> Real time			<input type="checkbox"/> Verbose mode			<input checked="" type="checkbox"/> Format output		
Select studies:	001 - Mallin 1092-96-01								
Select users:	abartley - Angela Bartley								
Select Records:	Data Type:	a - QC and Data			Pat IDs:				
	Seq (0-7):		Lvl (0-7):		Plates:		Visits:		
	Flt Name:		Flt Value:		Img IDs:				
	Time Range:	Start:	(YYYYMMDD.hhmmss)			End:	20050201.160750 (YYYYMMDD.hhmmss)		
Select Fields:	(only selected columns will be displayed)								
Actions:	<input type="button" value="Reset"/> <input type="button" value="Submit"/> <input type="button" value="Help"/> <input type="button" value="Refresh"/>								

Task 3 - DataFax Journal File Viewer

- **Purpose:** This web interface provides you a gateway to query journal files in a study. You can select records by user ids, data type, record status, validation levels, patient ids, visit numbers, plate numbers, image ids, and start and end times. You can specify fields (columns) by using field numbers. This task has the following task types:



DataFax Journal File Viewer - Search Result

- Study 253
- Patient 9002
- Time
between
2001/8/1 and
2002/6/1

Search Results

- 253 - TRAINING
 - 200108: [\[Page 1\]](#)
 - 200109: [\[Page 1\]](#) [\[Page 2\]](#) [\[Page 3\]](#)
 - 200110: [\[Page 3\]](#)
 - 200111: [\[Page 3\]](#)
 - 200112: [\[Page 3\]](#)
 - 200201: [\[Page 3\]](#) [\[Page 4\]](#)
 - 200203: [\[Page 4\]](#)
 - 200204: [\[Page 4\]](#)
 - 200205: [\[Page 4\]](#) [\[Page 5\]](#)
 - 200206: [\[Page 5\]](#)

Total selected records: 403 out of 16930

© Copyright Premier Research Group plc
30 South 17th Street, Philadelphia, PA 19103



DataFax Journal File Viewer - A Sample Record

```
# File name: /dlb/datafax/src_data/syswide/queries/t3311464/p0001.txt
( 0, 4): 010824
( 1, 5): 120021
( 2, 6): dyannett
( 3, 7): d
DFSTATUS( 4, 8): 2
DFVALID( 5, 9): 2
DFRASTER( 6,10): 0134/0571001
DFSTUDY( 7,11): 253
DFPLATE( 8,12): 1
DFSEQ( 9,13): 0
PATID(10,14): 9002
PTINIT(11,15): CWK
VDATE(12,16): 05/28/01
CONSDT(13,17): 06/28/01
INC_1(14,18): 2
INC_2(15,19): 1
INC_3(16,20): 1
INC_4(17,21): 1
INC_5(18,22): 1
INC_6(19,23): 1
INC_7(20,24): 0
INC_8(21,25): 0
INC_9(22,26): 1
INC_10(23,27): 1
INC_11(24,28): 1
INC_12(25,29): 1
INC_13(26,30): 1
INC_14(27,31): 1
INC_C(28,32):
DFSCREEN(29,33): 2
DFCREATE(30,34): 01/08/24 12:00:21
DFMODIFY(31,35): 01/08/24 12:00:21
(32,36):
```



Audit Trail Data Warehouse - Why Do We Need It

- The built-in reports and the DFjfv only provide the tools for examining the individual records in journal file or in fax log but could not provide high level reports from the rich data contained in the journal and fax log files.
- With an audit trail data warehouse, we can conduct online analytical processing (OLAP) and can convert data into knowledge to provide some business intelligence (BI)



Audit Trail Data Warehouse - How To Build One

- Developed an ETL application to extract, transform, and load journal and fax log files from DataFax into Oracle database.
- Build fact tables: The ETL application only extracts 18 common fields from the journal files and builds the columns consistently for all studies. It is easy to process fax_log. [As we know each plate in DataFax has different number of fields, and they may have different type definition for the same fields in different studies.]
- Add tools: Install and configure Oracle infrastructure, portal and business intelligence components. [optional]
- Define dimensions: define the summary and drill down paths. [optional]



Audit Fact Tables - Journal Columns

```
SQL> desc S000DF_ATJNL
Name                               Null?    Type
-----
DFSTUDY                            NOT NULL NUMBER(3)
CENTERNO                           NOT NULL NUMBER(5)
RECORD_TYPE                         NOT NULL CHAR(1)
DFSTATUS                            NOT NULL NUMBER(1)
DFVALID                             NOT NULL NUMBER(1)
DFRASTER                            NOT NULL VARCHAR2(12)
DFPLATE                             NOT NULL NUMBER(3)
DFSEQ                               NOT NULL NUMBER(5)
DFPATID                             NOT NULL NUMBER(10)
JNL_USER                            VARCHAR2(12)
JNL_DATE                            DATE
CRT_USER                            VARCHAR2(12)
CRT_DATE                            DATE
MOD_USER                            VARCHAR2(12)
MOD_DATE                            DATE
RES_USER                            VARCHAR2(12)
RES_DATE                            DATE
VISIT_DATE                          DATE
```



Audit Fact Tables - Fax Log Columns

```
SQL> desc S000_ATFAXLOG
Name                               Null?    Type
-----
FAX_FILE_NAME                       NOT NULL VARCHAR2(150)
PAGES                                NUMBER(6)
FAX_SENT_TIME                       DATE
FAX_IDENTIFIER                       NOT NULL VARCHAR2(100)
ARCHIVE_DURATION                     NUMBER(3)
DF_SERVER                            VARCHAR2(25)
JOB_ID                               NOT NULL NUMBER(9)
```



Audit Data Warehouse - Reporting Facts

- Drill-down to the details
 - ✓ How many re-faxes for this CRF page id?
 - ✓ List all the changes for this CRF image
 - ✓ Show me who has worked on this CRF at what time
 - ✓ How many QC notes posted to this CRF?



Audit Data Warehouse - OLAP Reports

- Summary Reports
 - ✓ How many faxes that we have received each month in the last 5 years?
 - ✓ How many CRFs that we have processed each week last year?
 - ✓ Who are the best performer each year in validating CRF forms?
 - ✓ Which site has the shortest average time to send in CRFs since each visit?



Audit Data Warehouse - Business Intelligence

- From Data to Knowledge: linking the audit data warehouse with other data
 - ✓ CRF design and validation performance: how big impact is a busy CRF to the performance of data validation and ICR?
 - ✓ How effective are site visits in reducing re-faxes and resolving QC notes?
 - ✓ Financial performance: How big is the profit margin for this type of study?



Conclusion

- Journal and fax log files are the golden mines in DataFax system.
- Converting journal and log files into fact tables, we can build audit trail data warehouse.
- Through mining the data in the warehouse, we can gain useful knowledge about the trial and performance of employees and centers.
- With OLAP tool, we can drill down to detailed history of each data record.

