

Supplement 2:

Legacy Technical Environment

Table of Contents

1	LEGACY TAX DISCOVERY DATA WAREHOUSE TECHNICAL ENVIRONMENT AND TOOLS.....	3
1.1	GENERAL SCOPE	3
1.2	TEST RDMS HARDWARE/SOFTWARE.....	3
1.3	PRODUCTION RDMS HARDWARE/SOFTWARE.....	3
1.4	BAR	4
1.5	MAINFRAME CONNECTION	4
1.6	SUPPORT ENVIRONMENT	5
2	TAXATION VOLUMETRICS.....	7
2.1	GENERAL	8
2.2	EXTRACT, TRANSFORM AND LOAD STATICS	8
2.3	CONSUMERS	10

1 Legacy Tax Discovery Data Warehouse Technical Environment and Tools

1.1 General Scope

This section provides the hardware and software being utilized in the ODT legacy DWH environment. The Department plans to retire this environment and its infrastructure by September 2021. This environment fully meets ODT's technical and business requirements.

Offerors must be able to extract, transform and load data from this environment into their proposed Tax Discovery Data Warehouse solution.

- Ohio Taxation Test on-premise environment
- Ohio Taxation Production on-premise environment
- Ohio Taxation Backup and Recovery on-premise (BAR)
- FICON Connect to Mainframe

1.2 Test RDMS Hardware/Software

1.2.2 Hardware

- TPA - Current Test Tax Compliance Database utilize Teradata 670c 1 node
- 4 TCore
 - 256GB Memory
 - 23 AMPs
 - 2 PE/Node
 - 1x670 Controller, 46xDrives (46x600GB HDD 10k), 46xDPN

1.2.3 Software

- Linux SLES 11 SP3
- Teradata 16.20 Enterprise WH Edition
- Teradata 16.20 Temporal Support for EWE
- Teradata Columnar Support for EWE
- Teradata 16.20 – Database
- TTU – 16.20

1.3 Production RDMS Hardware/Software

1.3.1 Hardware

TPA - Current Production Tax Compliance Database utilize Teradata 6700c 1+1HSN (Scalability to 1,296 nodes)

- 7 TCore
- 256GB Memory
- 36 AMPs
- 2 PE/Node
- 5x6700-Arrays, 144xDrives (144x300GB-HDD, 10k), 6xGHS-HDD, 144xDPN

1.3.2 Software

- Linux SLES 11 SP3
- Teradata 16.20 Enterprise WH Edition
- Teradata 16.20 Temporal Support for EWE
- Teradata Columnar Support for EWE
- Teradata 16.20 – Database
- TTU – 16.20

1.4 BAR

1.4.1 Hardware

- Master/Media Server on Dell PowerEdge R730
- Scalar Key Manager Appliance Pair - SKM – Version 2.1.1
- Quantum - Scalar i80
- Quantum- i80 Tape Drive HP LTO-5 8Gb

1.4.2 Software

Master/Media Server - Teradata Managed Storage Server (TMSS)

- SuSE Linux Enterprise Server 11 – SP3
- Teradata Data Stream Architecture
- NetBackup Enterprise Server
- NetBackup Library Based Tape Drive
- NetBackup Virtual Tape Option
- NetBackup Vault Base

1.4.3 Disaster Recovery

Full database level backups are taken of each database/user at least weekly via Teradata Data Stream Architecture and directed to the Teradata VTL. On Sundays', the entire VTL is copied to physical tapes using Netbackup and sent off site to Fireproof in a locked box. The tapes remain off site for three weeks then they are returned, expired, wiped clean and inventoried. In case of a disaster, Taxation plans to use the tapes to recovery the data warehouse on a system provided by Teradata within two weeks of the disaster.

1.5 Mainframe Connection

1.5.1 Hardware

- FICON Connect to Mainframe
 - FICON16s (rated up to 16Gb, running at 4Gb)
 - PROD Channel C1 – devices 420-42F
 - Test Channel B1 – devices 400-40F

1.5.2 Software

- Teradata 16.2 The Directory Program
- Teradata Tools and Utilities for z/OS package
 - Basic Teradata Query Utility (BTEQ) 16.20
 - Teradata FastExport 16.20
 - Teradata FastLoad 16.20

- Teradata MultiLoad 16.20
- Teradata TPump 16.20

1.6 Support Environment

1.6.1.1 Four Microsoft Window Servers managed by DAS

- **Test ETL server**
 - Window Server 2016 Standard 64-bit operating system
 - 8 CPU at 2400MHz
 - 24 GB RAM
 - Two drives, C: 125GB, D: 1T storage
- **Test Cognos Server**
 - Window Server 2012R2 Standard 64-bit operating system
 - 8 CPU at 2400MHz
 - 18 GB RAM
 - Three drives C: 99GB, D: 99GB, L: 10GB
- **Production ETL Server**
 - Window Server 2016 Standard 64-bit operating system
 - 8 CPU at 2400MHz
 - 24 GB RAM
 - Two drives, C: 99GB, D: 1T storage
- **Production Cognos Server**
 - Window Server 2012R2 Standard 64-bit operating system
 - 8 CPU at 2400MHz
 - 18 GB RAM
 - Three drives C: 99GB, D: 99GB, L: 10GB
- **Mainframe on DAS managed hardware**
 - Taxation is running on OIT's z13 2964.508.IBM (Model: N30,dialed down to a 508) , soft capped at 50MSU (approx. 375-430 mips).
 - sharing up to 4 GP's
 - 12GB central storage allocated to PRDA lpar
 - Currently NO crypto engines allocated to our Lpars.
 - Shared use of 3 ziiip engines
 - Shared use of 2 OSC / ICC OSA cards for Console connections
 - Dedicated use of 3 OSA Express cards for network connectivity to lpars

1.6.1.2 Three UNIX Linux SLES Servers on Taxation's Teradata Racks

- TD viewpoint - Teradata's strategic and innovative SOV (single operational view) for Teradata DB, Aster, and HDP Hadoop systems management and monitoring that enables Teradata's Unified Data Architecture (UDA).
- TD management workstation server for Test
- TD management workstation server for Production

1.6.2 Software

- **Test ETL server**

- ActivePerl 5.26.3
- Atanasoft 19.3
- IBM Db2 Runtime Client 11
- IBM Workload Scheduler Client
- IBM SPSS Modeler 18.2
- IBM SPSS OEM Connect and Connect XE for ODBC 8.0
- Intelligent Search Technology
 - CorrectAddress
 - MerlinMerge SpeedPro 3.1
- Oracle Client 12c
- Teradata Tools and Utilities

Version	Display Name
16.20.00.02	Teradata Azure Access Module nt-x8664 16.20.0.2
16.20.00.03	Teradata FastExport 16.20.0.3
16.20.00.03	Teradata FastLoad 16.20.0.3
16.20.00.03	Teradata TPump 16.20.0.3
16.20.00.04	Teradata MultiLoad 16.20.0.4
16.20.00.05	Teradata Kafka Access Module nt-x8664 16.20.0.5
16.20.00.06	Shared ICU Libraries for Teradata 16.20.0.6
16.20.00.06	Shared ICU Libraries for Teradata nt-x8664 16.20.0.6
16.20.00.07	Teradata BTEQ 16.20.0.7
16.20.00.07	Teradata Data Connector 16.20.0.7
16.20.00.07	Teradata Data Connector nt-x8664 16.20.0.7
16.20.00.07	Teradata S3 Access Module nt-x8664 16.20.0.7
16.20.00.10	Teradata SQL Assistant 16.20.0.10
16.20.00.102	ODBC Driver for Teradata 16.20
16.20.00.102	ODBC Driver for Teradata nt-x8664 16.20
16.20.00.16	Teradata CLiv2 16.20.0.16
16.20.00.16	Teradata CLiv2 nt-x8664 16.20.0.16
16.20.00.17	Teradata Parallel Transporter Base 16.20.0.17
16.20.00.17	Teradata Parallel Transporter Base nt-x8664 16.20.0.17
16.20.00.17	Teradata Parallel Transporter Stream 16.20.0.17
16.20.00.17	Teradata Parallel Transporter Stream nt-x8664 16.20.0.17
16.20.12.00	Teradata Studio nt-x8664 16.20.12
16.20.29.00	Teradata Tools and Utilities - Base 16.20.29
16.20.33.00	ODBC Driver for Teradata - Suite 16.20.33
16.20.8.0	.NET Data Provider for Teradata 16.20.8

- **Test Cognos server**

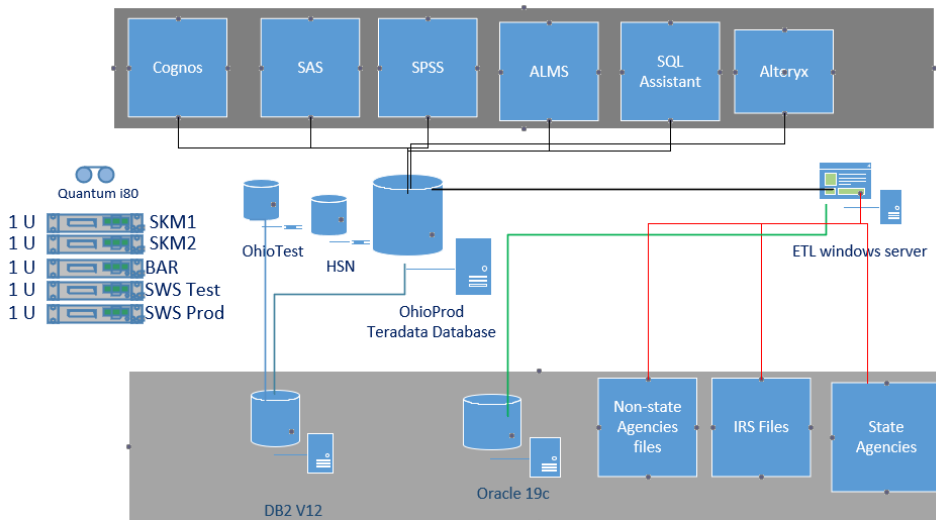
- IBM Cognos Analytics V11.0.8
- IBM Cognos Framework
- IBM DB2 Runtime Client V11.1
- IBM Workload Scheduler Client
- Microsoft SQL Server 2012
- Teradata Tools and Utilities 16.20 (see complete list above)

- **Production ETL Server**

- ActivePerl 5.26.3
- Atanasoft 19.3
- IBM Db2 Runtime Client 11
- IBM Workload Scheduler Client
- Intelligent Search Technology
 - CorrectAddress - USPS Address standardization
 - MerlinMerge SpeedPro 3.1 - Soft Matching tool.
- Oracle Client 12c
- Teradata Tools and Utilities (see complete list above)

- Veritas NetBackup 8.1 Administration Console
- **Production Cognos server**
 - IBM Cognos Analytics V11.0.8
 - IBM DB2 Runtime Client V11.1
 - IBM Workload Scheduler Client
 - Microsoft SQL Server 2012
 - Teradata Tools and Utilities 16.20 (see complete list above)
- **Production Cognos server**
 - IBM Cognos Analytics V11.0.8
 - IBM Cognos Framework
 - IBM DB2 Runtime Client V11.1
 - IBM Workload Scheduler Client
 - Microsoft SQL Server 2012
 - Teradata Tools and Utilities 16.20 (see complete list above)

1.6.3 Connections



- In Blue – FICON Connection from the Mainframe to both Teradata database Test and Production
- In Green – Oracle 12c client using SQL plus to extract data to a file on the windows ETL server
- In Red – secure FTP
- In Black – ODBC Driver for Teradata & Teradata Data Connector

2 Taxation Volumetrics

ODT's Tax Discovery Production Data Warehouse is over 6T and has 11 databases over 100G compressed. ODT's Quality Assurance (QA) environment is the same size as production. There are only the two environments of the warehouse, Production and QA (a.k.a. Test and non-production)

2.1 General

2.1.1 Production Data Warehouse Databases over 100G in size

Database	Storage	Tables
TTCS_IRS_DB	1.67T	58
TTCS_OHIO_DB	991G	222
PIT_STG_DB	466G	81
ODT_STG_DB	342G	98
ODT_UTIL_DB	269G	40
TTCS_RESTRICT_DB	200G	8
ODT_REPORT_DB	165G	16
TTCS_REPORT_DB	150G	16
TTCS_MISC_DB	105G	4

2.1.2 State Tax Types

Yearly Personal Income tax forms received, persist 10 years in the WH

- Data Source – Db2 z/OS V12
- Twice weekly complete ETL refresh

Current Sale, Use and Streamline Sales Tax

- Data Source – Oracle 12c on aix
- weekly complete ETL refresh

Current PTE

- Data Source – Oracle 12c on aix
- weekly complete ETL refresh

Current CAT

- Data Source – Oracle 12c on aix
- weekly complete ETL refresh

Current MUNI

- Data Source – Oracle 12c on aix
- weekly complete ETL refresh

2.1.3 IRS Tax Type

IRS data is provided in fixed record length files, CSV files, XML and access database annually, semi-annually and monthly. ODT provides the IRS 'Tickler' files throughout the year.

2.2 Extract, Transform and Load Statics

Offerors must be able to reuse, translate or recreate all ODT's existing ETL processes from this environment into their proposed Tax Discovery Data Warehouse solution.

Offeror is directly responsible for maintaining performance, data integrity and documentation of any new or modified ETL processes,

2.2.1 Number of ETLs by Type

	on windows	mainframe
Teradata Fast Load		30
Teradata Parallel Transport (TPT)	134	21
Teradata bteq	254	116
Teradata Fast Export		3
Oracle SQL plus	80	0

BTEQ, stands for Basic Teradata Query - is a general purpose, command driven utility that enables user to interact with one or more Teradata Database Systems. ODT uses it as a SQL/DDDL scripting tool for ETL processes.

TPT, stands for Teradata Parallel Transporter - is a single utility which uses other Teradata Utility depending on the operation the user specifies. It will invoke

- FASTLOAD – simple fast load operator
- MULTILOAD – complex load operator
- FASTEXPORT – export operator
- TPUMP – stream operator

2.2.2 Run times by type

- Ohio Tax Investigative System (OTIS) – daily full replacement
 - 12 TPT scripts utilizing TPump from Db2 z/OS streaming into Teradata

<u>Records streamed</u>	<u>elapsed time</u>
▪ 3,693,677	1 min
▪ 6,933,397	1 min 4 s
▪ 13,132,084	1 min 7s
▪ 12,676,309	1 min 42 s
▪ 5,143,546	7s
▪ 197,441	1s
▪ 24,885,516	6 min 30s
▪ 26,284	16s
▪ 2,392,790	22 min
▪ 6,942,760	28 min (8 table join)
▪ 5,255,490	1 min 8s
▪ 6,788,328	3 min
 - 1 bteq script (insert/update/deletes and call of a stored procedure)
 - 18,476,117 17 min
- Personal Income Tax and Assessments – twice weekly refresh – all scripts/jobs executed on the mainframe:
 - 548,374,134 records involved
 - 25 Teradata Staging tables are loaded using TD FASTLOAD in 59 minutes elapsed time, from previously unloaded Db2 tables to flat files.
 - 25 BTEQs select from Teradata staging tables transform and load into 25 target Teradata Tables. In 59 minutes, elapsed time.

2.3 Consumers

2.3.1 Cognos (number of reports and timings)

- 225 active users
- 96 Production reports utilizing a Cognos Framework model

2.3.2 SPSS predictive models

IBM's SPSS (Statistical Package for the Social Sciences) is used to build predictive models for Fraud, compliance, audit and other Taxation data. Majority of the models are Ad-hoc and are to gain a better understanding of the data or situation.

ODT has a couple 'CRISP-DM' like processes that re-analyze once every 6 months. The main process is for an audit lead generation application known to ODT as ALMS. It consists of a total of 56 streams and its longest running stream is the CAT's predictive script. It consumes 3 years of filing history for CAT, Sales, IRS, PTE, SUT and W2 wages and aggregates millions of records down to 191K and 97 fields. It takes about 10 minutes.