Enterprise Data Management (EDM) Solution Overview

Sources

Data entities from disparate systems
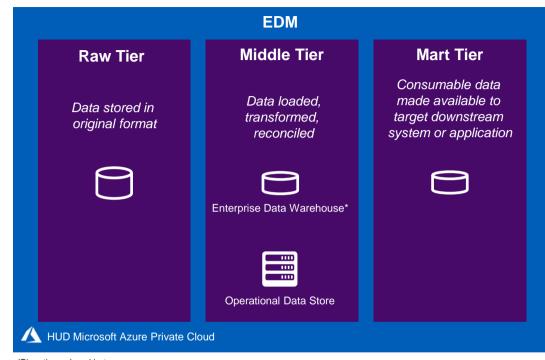
Unisys Mainframe Sources

IBM Mainframe Sources

Sybase ASE Sources

Oracle Database Sources

File-based Sources (SFTP server)

EDM

Raw Tier

Data stored in original format

Middle Tier

Data loaded, transformed, reconciled

Enterprise Data Warehouse*

Operational Data Store

Mart Tier

Consumable data made available to target downstream system or application

HUD Microsoft Azure Private Cloud

Service

External service consumer (system or application)

Applications

MicroStrategy

Power BI

SAS

MS Access

SFTP server

*Discretionary by subject

EDM Technology, Tools, Processing

Sources

Data entities from disparate systems

Unisys Mainframe Sources

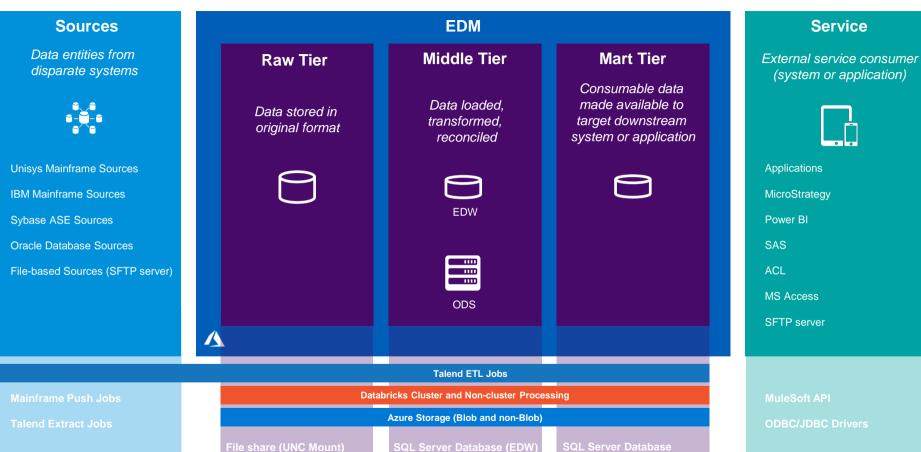IBM Mainframe Sources

Sybase ASE Sources

Oracle Database Sources

File-based Sources (SFTP server)

EDM

Raw Tier

Data stored in original format

Middle Tier

Data loaded, transformed, reconciled

EDW

ODS

Mart Tier

Consumable data made available to target downstream system or application

Service

External service consumer (system or application)

Applications
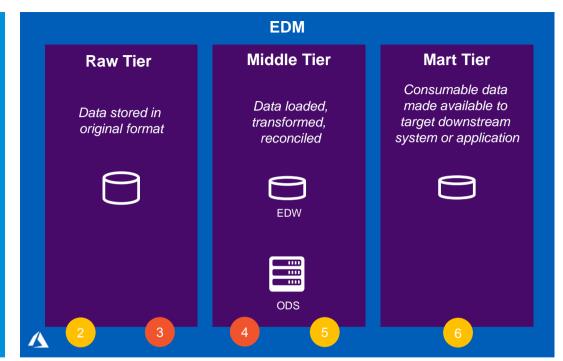
MicroStrategy

Power BI

SAS

ACL

MS Access

SFTP server

Talend ETL Jobs

Mainframe Push Jobs	**Databricks Cluster and Non-cluster Processing**		MuleSoft API
	Azure Storage (Blob and non-Blob)		
Talend Extract Jobs			ODBC/JDBC Drivers

File share (UNC Mount)	**SQL Server Database (EDW)**	**SQL Server Database**
	File based Objects (ODS)	

EDM ETL Jobs

Sources	EDM			Service

Sources

Data entities from disparate systems

Unisys Mainframe Sources

IBM Mainframe Sources

Sybase ASE Sources

Oracle Database Sources

File-based Sources (SFTP server)

1

EDM

Raw Tier

Data stored in original format

2　**3**

Middle Tier

Data loaded, transformed, reconciled

EDW

ODS

4　**5**

Mart Tier

Consumable data made available to target downstream system or application

6

Service

External service consumer (system or application)

Applications

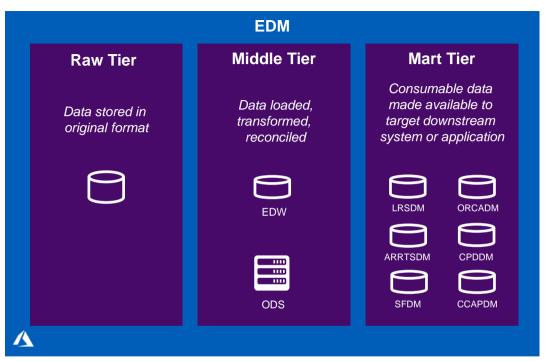MicroStrategy
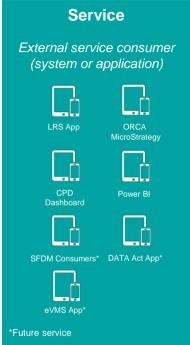
Power BI

SAS

ACL

MS Access

SFTP server

Job Types

1. Push (mainframe), Extract (relational), or Transfer (SFTP server) Job to push or pull data from source
2. Validate Job to validate data received vs. data expected (control file)
3. Intake Job to perform change data capture and limit ingest to updates, inserts, no changes (Databricks Cluster processing, Azure Blob Storage)
4. LoadFile Job to build table from different sources (Databricks Cluster processing, Azure Blob Storage)
5. Publish Job to deliver LoadFile to UNC mount
6. LoadDB Job to bulk load LoadFiles to target data mart

EDM Sources, Marts, Service (dev and prod)

Sources

*Data entities from disparate systems**

Single Family
- *ARRTS/F51A*
- *CHUMS/F17*
- *CLAIMS/A43C*
- *DCAMS-GD/F71A*
- *FASS/P092*
- *FHA-SL/P013*
- *HERMIT/P271*
- *LAMS/P292*
- *LEAP/P278*
- *SAMS/A80S*
- *SFDMS/F42D*
- *SFIS/A43*
- *SFPCS-U/A80R*
- *TIIS/F72*
- *TOTAL*

Enrichment/Quality
- *GSC/A15*

CPD
- *DRGR/C08A*
- *GMP/C07A*
- *IDIS/C04*

Ginnie Mae
- *RFS/P269*
- *GFAS/P237*

Treasury
- *ARC*

Multi-Family
- *DAP/F24A*
- *CSMS/P085*
- *iREMS/F24*
- *MDDR/P057*
- *TRACS/F87*

PIH
- *IMS-PIC/P113*
- *PASS-R/P109*
- *REACS/F89A*
- *VMS/P224*

*Other CCAP sources: CMS.gov, Moodys.com, GinnieMae.gov, Dynaxis (LSS), HUD SharePoint

EDM

Raw Tier

Data stored in original format

Middle Tier

Data loaded, transformed, reconciled

EDW

ODS

Mart Tier

Consumable data made available to target downstream system or application

LRSDM

ORCADM

ARRTSDM

CPDDM

SFDM

CCAPDM

Service

External service consumer (system or application)

LRS App

ORCA MicroStrategy

CPD Dashboard

Power BI

SFDM Consumers*

DATA Act App*

eVMS App*

*Future service

LRS Dataset Workflow

Sources

Data entities from disparate systems

Single Family
- *CHUMS/F17*
- *CLAIMS/A43C*
- *HERMIT/P271*
- *LEAP/P278*
- *SFDMS/F42D*
- *SFIS/A43*

EDM

Raw Tier

Data stored in original format

Middle Tier

Data loaded, transformed, reconciled

EDW

ODS

Mart Tier

Consumable data made available to target downstream system or application

LRSDM

Service

External service consumer (system or application)

LRS Application

Data ingested in varying frequencies (daily, weekly, monthly, etc.)

ETL processing based on file size and target business need

Data loaded to target data mart based on consumption need

Data delivered to downstream app via service call

ORCA Dataset Workflow

Sources

Data entities from disparate systems

Treasury
- *ARC*

EDM

Raw Tier

Data stored in original format

Middle Tier

Data loaded, transformed, reconciled

EDW

ODS

Mart Tier

Consumable data made available to target downstream system or application

ORCADM

Service

External service consumer (system or application)

ORCA MicroStrategy

Data ingested in varying frequencies (daily, weekly, monthly, etc.)
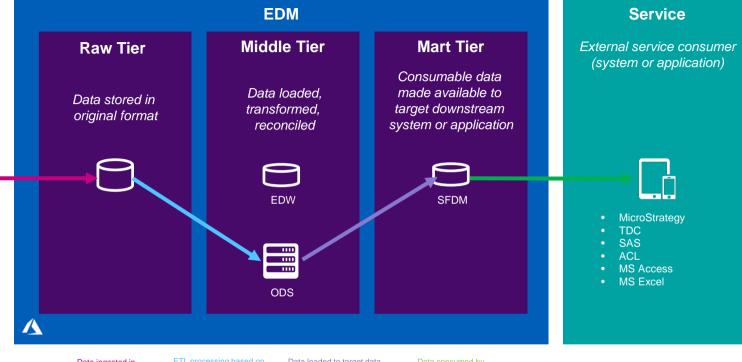
ETL processing based on file size and target business need

Data loaded to target data mart based on consumption need

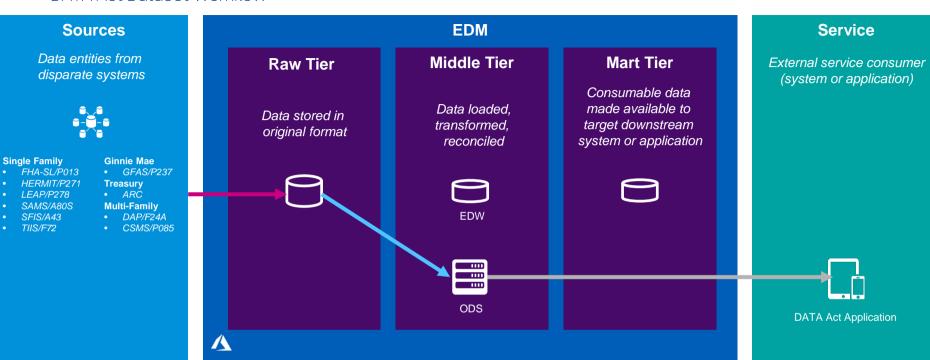Data consumed by downstream system, application, or user

SFDM Dataset Workflow

Sources

Data entities from disparate systems

Single Family
- *CHUMS/F17*
- *CLAIMS/A43C*
- *DCAMS-GD/F71A*
- *FHA-SL/P013*
- *HERMIT/P271*
- *LEAP/P278*
- *SAMS/A80S*
- *SFDMS/F42D*
- *SFIS/A43*
- *SFPCS-U/A80R*
- *TIIS/F72*
- *TOTAL*

Ginnie Mae
- *RFS/P269*

Enrichment/ Quality
- *GSC/A15*

EDM

Raw Tier
Data stored in original format

Middle Tier
Data loaded, transformed, reconciled

EDW

ODS

Mart Tier
Consumable data made available to target downstream system or application

SFDM

Service

External service consumer (system or application)

- MicroStrategy
- TDC
- SAS
- ACL
- MS Access
- MS Excel

Data ingested in varying frequencies (daily, weekly, monthly, etc.)

ETL processing based on file size and target business need

Data loaded to target data mart based on consumption need

Data consumed by downstream system, application, or user

DATA Act Dataset Workflow

Sources

Data entities from disparate systems

Single Family
- *FHA-SL/P013*
- *HERMIT/P271*
- *LEAP/P278*
- *SAMS/A80S*
- *SFIS/A43*
- *TIIS/F72*

Ginnie Mae
- *GFAS/P237*

Treasury
- *ARC*

Multi-Family
- *DAP/F24A*
- *CSMS/P085*

EDM

Raw Tier

Data stored in original format

Middle Tier

Data loaded, transformed, reconciled

EDW

ODS

Mart Tier

Consumable data made available to target downstream system or application

Service

External service consumer (system or application)

DATA Act Application

Data ingested in varying frequencies (daily, weekly, monthly, etc.)

ETL processing based on file size and target business need

Data delivered to downstream app via service call

Onboarding Data into EDM (raw tier only)

Process*

1. Identify and confirm consumption requirements (needed to develop data mart)
2. Identify data entities from one or more source systems
3. Confirm data entity info (database type, # of tables, # of records, primary/unique keys, capture method, data dictionary)
4. Schedule meeting with source system POCs to discuss requirements, request(s)
5. Obtain email approval from source system POCs to request interface with EDM/P302
6. Document Server Access Request (SAR) – include IPs, hostname, environment, directory, raw tables (unaltered source data, typically not views)
7. Submit DIAMS for mainframe developers (C-ID) to access mainframes (mainframe sources only)
8. Obtain email approval from Networking (Jimmy Bresnahan) to open firewall between source system and EDM/P302
9. Submit Service Desk Ticket to IT – DATA CENTER – SECURITY to open firewall (attach Networking approval)
10. Confirm firewall has been opened between on-prem and Azure resources
11. Submit Service Desk Ticket to IT – DATA CENTER – DATABASE or IT – DATA CENTER – MAINFRAME (include X-ID, attach SAR, attach approval) to authorize EDM/P302 access to source system
12. Validate access and privileges to source data entities
13. Begin capture development (push, extract, or transfer jobs, validate, intake jobs) per source data identified in step 3

Excludes PIV, VM, Talend License, Access to EDM Resources