Industry Workbench/Common Data Model

CDM driven Analytics Vision at Microsoft

Payal Tiwana | Principal Group PM | Azure Global
Data as a strategic asset
Today’s data realities

- **Volume**
  - What *data* do I have?

- **Variety**
  - Is it *trustworthy*?

- **Velocity**
  - Can people access the *data* needed to make the right decisions?
  - How can I enable faster *business insights*?
  - What’s my *compliance exposure*?
Analytics & AI is the #1 investment for business leaders, however they struggle to maximize ROI

80% report struggling to become mature users of data*

55% report data silos and data management difficulties as roadblocks*

* Harvard Business Review (2019), Understanding why analytics strategies fall short for some, but not for others
Cloud analytics enables business transformation

- **Simplicity**: Spend less time tuning on-prem appliances & more time understanding the new realities of your operations and customers using all your data.
- **Fast insights**: Adapt to a changing status quo with the ability to go from hypothesis to validation in minutes using ALL your data.
- **Reduce cost**: Quickly ramp-up data and analytics tools, save on costs and reduce CAPEX.
- **Security**: Secure a business-critical environment with enterprise-class features such as row and column security.
Digital transformation
Data & Intelligence

Engage customers

Transform products

Optimize operations

Empower people
Businesses are forced to maintain two critical, yet independent analytics systems.

Data science
- Big data
- Experimentation
- Fast exploration
- Semi-structured

Data lake

OR

Business analytics
- Relational data
- Proven security & privacy
- Dependable performance
- Structured

Data warehouse
Azure meets these challenges with a single service to provide limitless analytics.

Data science
- Data lake

Business analytics
- Data warehouse

Features:
- Ease of use
- Fast exploration
- Quick to start
- Proven security
- Airtight privacy
- Dependable performance
Data is stored within internal and external silos

Data intelligence is difficult to realize without a harmonized data estate especially if data scientists and business analysts use different tools and data.
Our Vision: Enable Data Harmonization using a Common Data Model

The key lies in building a unified data platform using a common data model that describes and governs that data and is accessible by all personas.
Our Vision: Enabling App Interoperability

Moving from **bespoke Apps** to **plug n play Apps** that can be more universally deployed quickly and cheaply requires a common data schema that the apps understand.
Comprehensive Common Data Models

Provide information blueprints for your business, create a plug and play ISV ecosystem and enable data consortiums across Industries

5,000 data entities and 30,000 data attributes (all with clear business names & definitions)

Common Data Models

22-30 Business Areas per industry model

Data produced by Customers & their ecosystem partners

Microsoft First Party Apps (e.g., D365)

Data from data providers/aggregators

Cross Industry Data Pools & Consortiums

Power Query / Power Platform

AI / ML

IOT / Digital Twins

Azure Data Share

Synapse Analytical Compute

Retail

Consumer Goods

Freight & Logistics

Agriculture

Sustainability & Carbon Mgmt.

Energy & Commodity Trading

Oil & Gas

Utilities

Automotive

Mfg.

Pharma & Clinical Trials

Genomics

Financial Services & Insurance

Healthcare

Media & Entertainment

Telecom

Purview

Data produced by Customers & their ecosystem partners

Microsoft First Party Apps (e.g., D365)

Data from data providers/aggregators

© Copyright Microsoft Corporation. All rights reserved.
Azure Synapse Analytics

Limitless analytics service with unmatched time to insight

Unified experience
- Azure Synapse Studio
  - Integration
  - Management
  - Monitoring
  - Security

Analytics runtimes
- SQL
- Spark
- Azure Data Lake Storage
- Azure Synapse Link
  - Available now
  - Future support

Data types
- On-prem data
- Cloud data
- SaaS data
- Streaming data

Integration with
- SQL Server
- Azure Machine Learning
- Power BI
Azure Synapse Analytics

- **Synapse SQL**: Query and analyze data T-SQL using both provisioned and serverless models.
- **Apache Spark for Synapse**: Apache Spark in Synapse for quick creation of notebooks with your choice of language.
- **Synapse Pipelines**: Build end-to-end workflows for your data movement and data processing scenarios.
- **Synapse Studio**: Execute all data tasks with a simple UI and unified environment.
Azure Synapse Analytics + Common Data Model (CDM)

**Synapse SQL**
Query and analyze data
T-SQL using both provisioned and serverless models

**Apache Spark for Synapse**
Apache Spark in Synapse for quick creation of notebooks with your choice of language

**Synapse Pipelines**
Build end-to-end workflows for your data movement and data processing scenarios
+ Automated Transformations

**Synapse Studio**
Execute all data tasks with a simple UI and unified environment
+ Low-code/no-code database design

**Common Data Models**
Comprehensive data models providing information blueprints for customers to describe their data estate for analytics
Democratize Data and unlock innovation

- **Acceleration to the cloud**
  Increase the speed with which data estates can be moved to the cloud and modernized.

- **Risk reduction**
  Use proven data models that are 85-90% complete "off-the-shelf" and do not require a large, lengthy and risky effort.

- **Common language for the organization**
  Enable clear communication among stakeholders and power business glossaries.

- **Better data integration**
  Avoid information "silos" by building a unified data lake.

- **Data Pipeline Automation**
  Enable automation of data pipelines in a low code, no code manner *(future release)*.

- **ISV ecosystem**
  Power an ecosystem of ISV solutions that interoperate with Microsoft's data models.
Customer case study
one of Europe's largest grocery retailers

Business scenario
The company wanted to bring together product and ingredient data sets together with customer data sets to determine nutritional choices customers were making in order to be able to recommend healthier choices

Benefits of using Common Data Models
• Process automation
• Providing data service: department after department
• Master Data Management
• DataMart in minutes
• Code generation versus Code typing

Business outcomes
• 80% reduction in time to set up end-to-end data pipeline
• Autogenerate data warehouse in 2 weeks instead of 6 months
• Hours to setup data pipelines instead of weeks
• $$$ savings due to use of ready-made industry data models and data automation tools
Demo
Reference Architecture

Source Data → Copy ETL → Raw layer → Output data to parquet → Stage layer

- Synapse Database: provides rich metadata of CDM
- Native Spark or SQL pools

- Integrated Synapse + IDW

- Power Apps
- Power BI
- Azure Data Share
- Azure Machine Learning

Data Lake (ADLS)

Purview natively understands Synapse
# Purview with CDM

## Purview Features

- **Automated data discovery, lineage identification, and data classification across on-premises, multicloud, and SaaS sources**

## CDM Enhancements

- **Provide semantics traits and meaning to optimize data discovery and classification**
- **Provide verbose relationship descriptors through model definitions**
- **Provide rich business descriptions for entities and attributes for inclusion into business glossary**
- **Provide model lineage for derived analytical data models**

Semantic search enables data discovery using business or technical terms. Insight into the location and movement of sensitive data across your hybrid data landscape.