

1

About me



- Solution architect at Inetum–Realdolmen
- Working for Toyota Motor Europe as lead CRM architect
- Blog: <https://jopx.blogspot.com>
- Twitter: @jopxtwits
- <https://github.com/jorisp>

2

2

Agenda

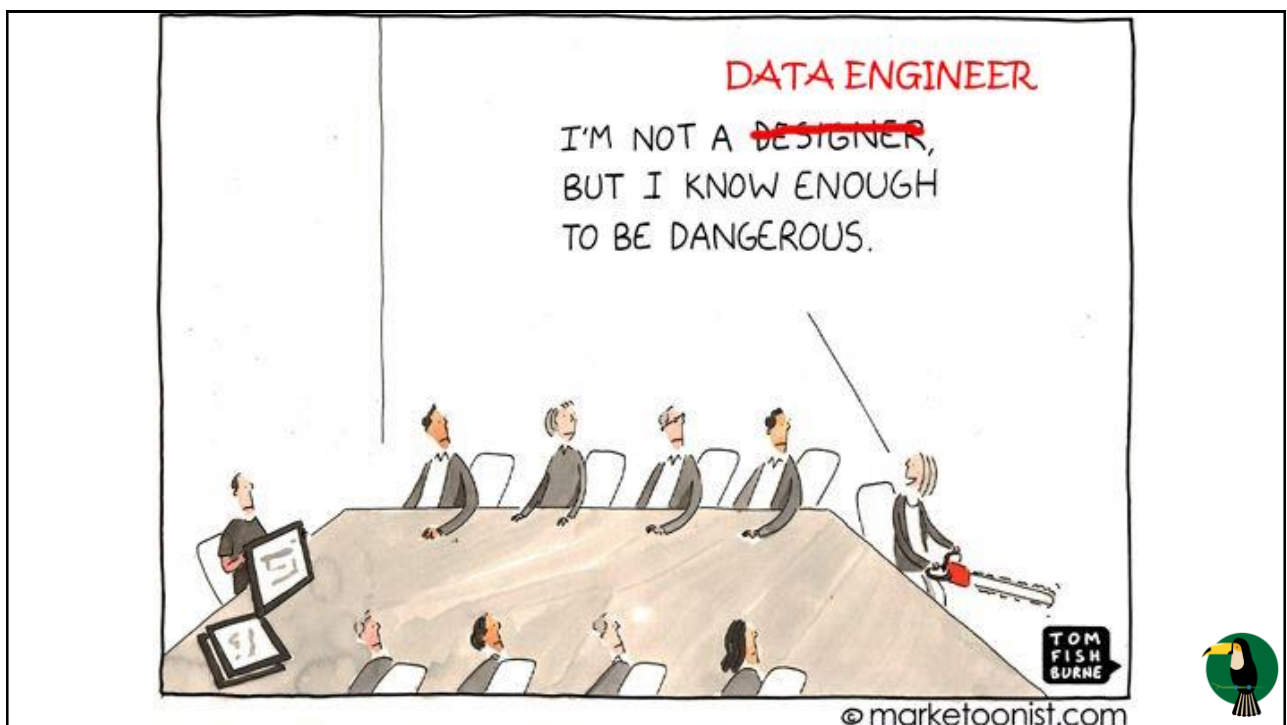
- Introduction to Azure Synapse Link for Dataverse
- Setting up Azure Synapse Link for Dataverse
- Consuming data from Azure Synapse Link for Dataverse
- Lessons learned at Toyota Motor Europe

Objectives

- Kickstart your deployment of Azure Synapse Link for Dataverse
- Provide practical tips & tricks and gotchas when deploying Azure Synapse Link for Dataverse
- No deep dive on Azure Synapse



4



5

Introduction to Azure Synapse Link for Dataverse



6

Why Azure Synapse Link for Dataverse



Dataverse



Azure Synapse Analytics



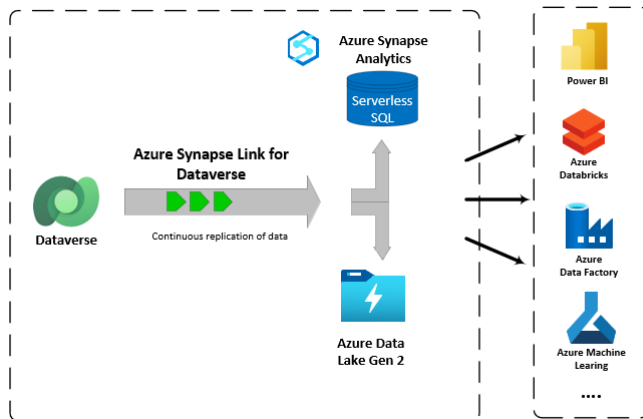
Accelerate to insights

- A single end-to-end way to work with data in Microsoft Dataverse, make it available for analytics and visualization, integrated with external datasets at any scale (supporting large volumes) as well as running AI and machine learning.



7

What is Synapse Link for Dataverse?

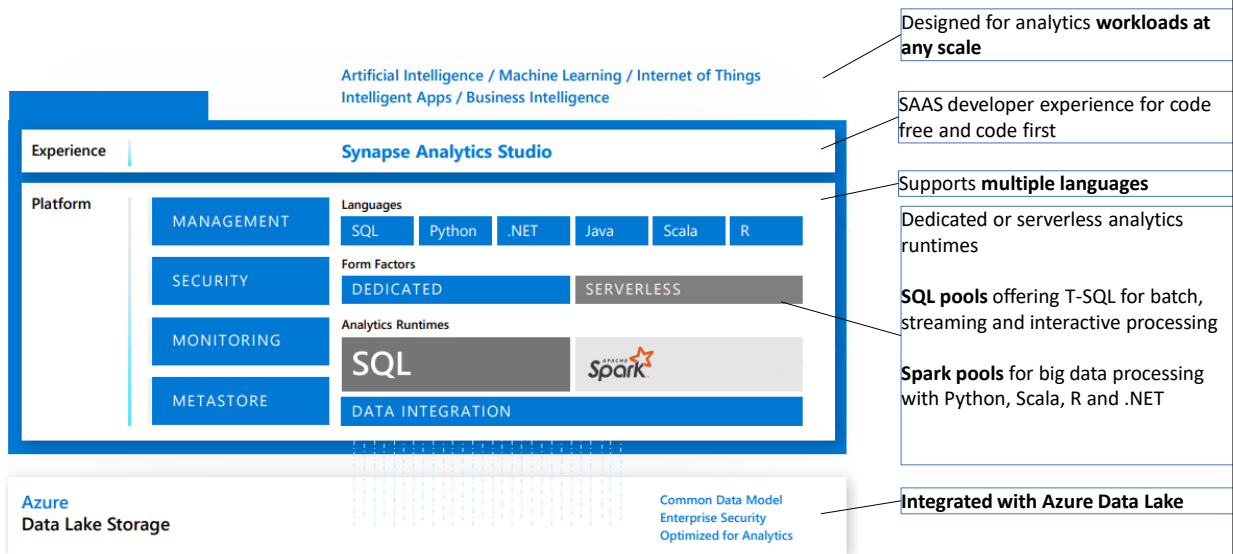


- Replication engine to ADLS and Synapse
 - Export to Data Lake (GA 02/2020)
 - Export to Synapse (GA 11/2021)
 - Near real time replica of selected tables/entities



8

What is Azure Synapse?



9

Azure Synapse

Serverless vs dedicated SQL pools

Serverless SQL pools

- Used for data preparation or ad-hoc queries against unstructured data
- Provides an always available SQL endpoint for unplanned workloads
- Enables interactive querying
- Pay per use - 4,48€/TB of data processed.

Dedicated SQL pools

- Used for Data Warehouse operations
- Provides predictable performance
- Reserves processing power for data stored in SQL tables
- Cost known upfront - smallest DW100C (100 DWU) – monthly cost +/- 1.000 EUR



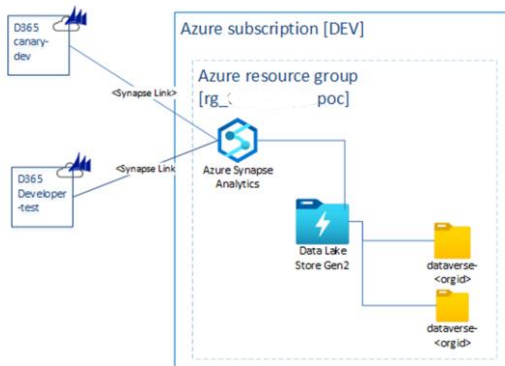
10

Azure Synapse
Link for
Dataverse setup



11

Azure Synapse Link for Dataverse setup



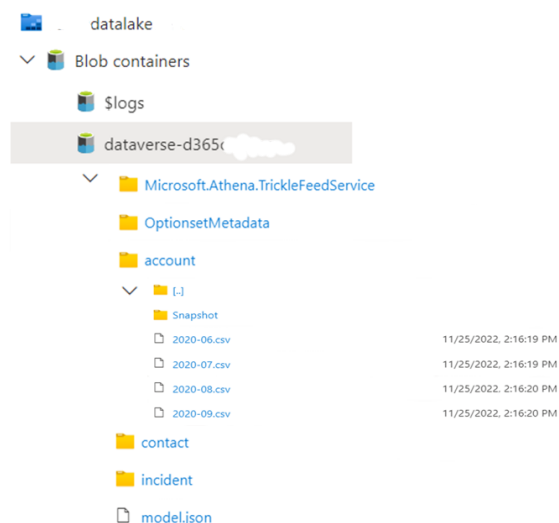
- Setup from Maker Portal
 - Requires Azure Data Lake Storage Gen2 and Azure Synapse upfront setup
- Prerequisites <https://bit.ly/3Mlunzk>
 - Same region
 - Owner role on Azure Blob Storage account
 - Entities/tables enabled for change tracking



12

Azure Synapse Link for Dataverse setup

Additional considerations

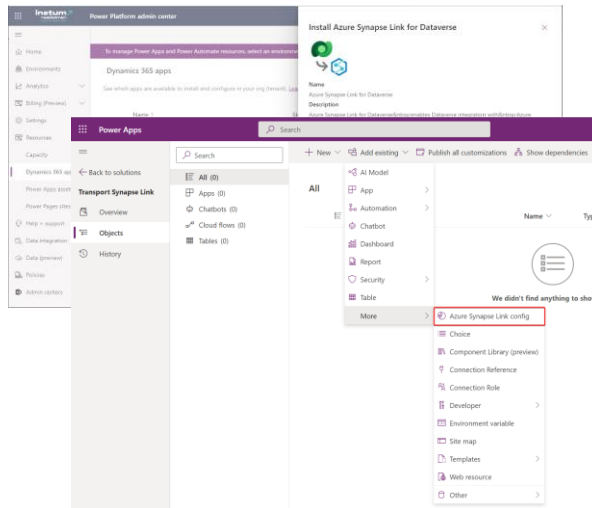


- Data Lake Storage based on CDM model
- Partition mode – monthly or yearly?
 - Monthly default mode, recommended for large volumes
- In-place updates vs append only writes
 - In place update – default setting.
- Unable to change after initial setup
- Reference: <https://bit.ly/3VVjzuW>



13

Azure Synapse Link for Dataverse setup



Possible to transport the Azure Synapse Link for Dataverse config across environments

<https://bit.ly/3I4OEgy>

- Install Microsoft solution from AppSource
- Add Synapse Link Config to your own solution

Limitations:

- One time only – no update scenario
- Not possible to use different ADLS & Synapse setup in source and target

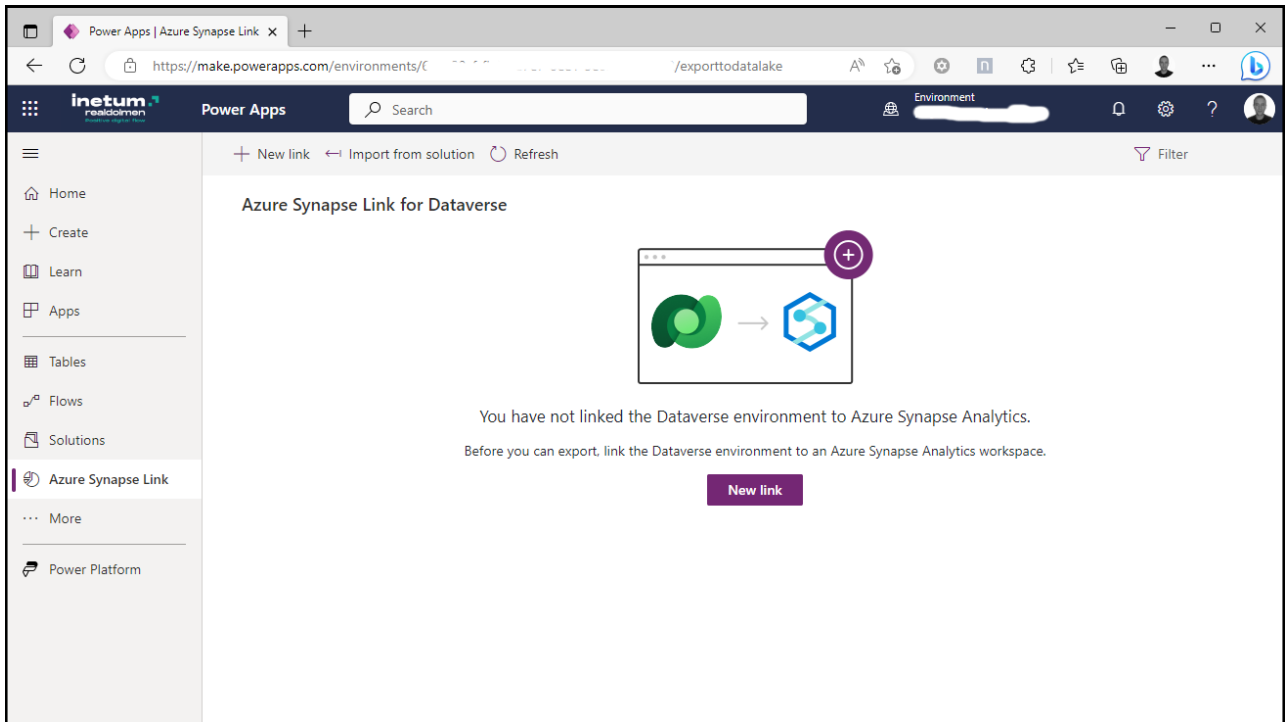


14

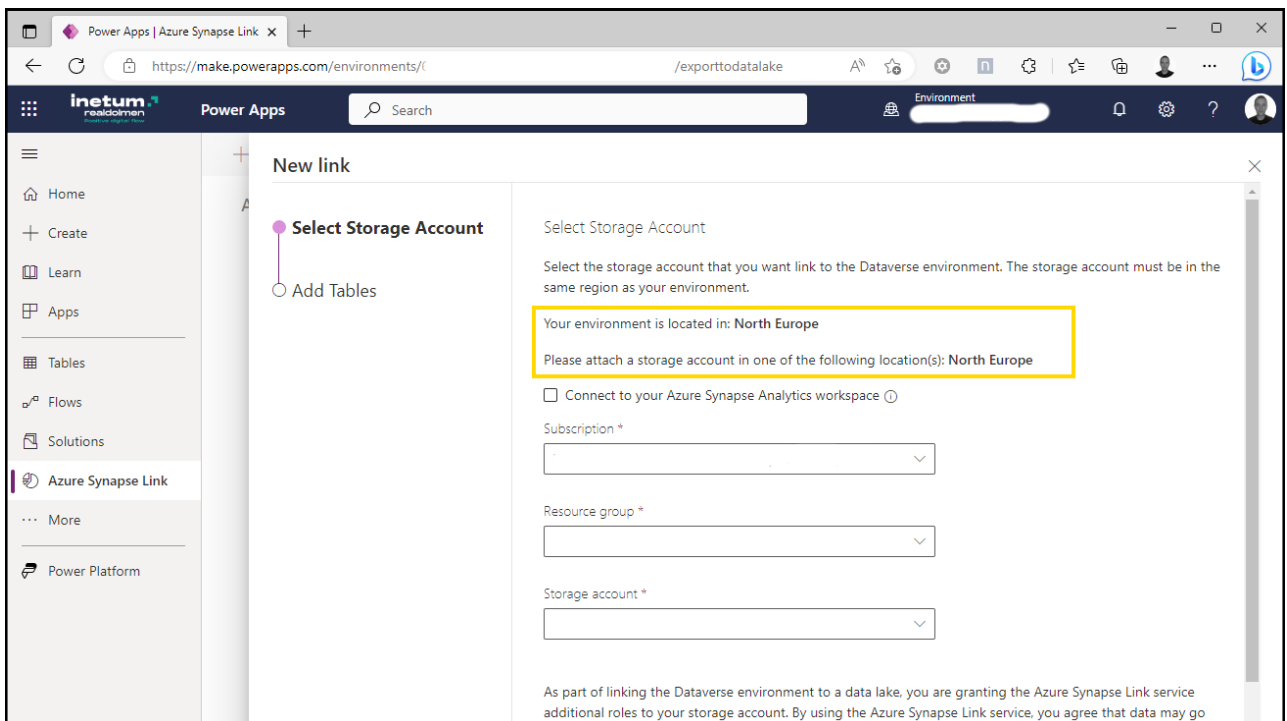
Azure Synapse
Link for
Dataverse setup
- walkthrough



15



16



17

Microsoft Azure

Search resources, services, and docs (G+/)

Home > Subscriptions > | Resource groups > | Marketplace > Storage account >

Create a storage account

Basics **Advanced** Networking Data protection Encryption Tags Review

Permitted scope for copy operations (preview)

Hierarchical Namespace

Hierarchical namespace, complemented by Data Lake Storage Gen2 endpoint, enables file and directory semantics, accelerates big data analytics workloads, and enables access control lists (ACLs) [Learn more](#)

Enable hierarchical namespace ☒

Access protocols

Blob and Data Lake Gen2 endpoints are provisioned by default [Learn more](#)

Enable SFTP ☐

Enable network file system v3 ☐

18

Home > All resources > Create a resource > Marketplace >

Azure Synapse Analytics

Microsoft

Add to Favorites

Microsoft | Azure Service

★ 4.1 (54 ratings)

Plan: Azure Synapse Analytics [Create](#)

Overview Plans Usage Information + Support Ratings + Reviews

Azure Synapse is a limitless analytics service that brings together data integration, enterprise data warehousing. It gives you the freedom to query data on your terms, using either serverless or dedicated resources-at scale. A experience to ingest, explore, prepare, manage, and serve data for immediate BI and machine learning needs.

Key service capabilities include:

- Unified analytics platform
- Serverless and dedicated options
- Enterprise data warehouse
- Data lake exploration
- Code-free hybrid data integration
- Deeply integrated Apache Spark and SQL engines
- Cloud-native HTAP
- Choice of language (T-SQL, Python, Scala, SparkSQL, R, .NET)
- Integrated AI and BI

Home > Marketplace > Azure Synapse Analytics >

Create Synapse workspace

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all of your resources.

Subscription *

Resource group * [Create new](#)

Managed resource group

Workspace details

Name your workspace, select a location, and choose a primary Data Lake Storage Gen2 file system to serve as the default location for logs and job output.

Workspace name *

Region *

Select Data Lake Storage Gen2 * ☒ From subscription ☐ Manually via URL

Account name * [Create new](#)

File system name * [Create new](#)

i We will automatically grant the workspace identity data access to the specified Data Lake Storage Gen2 account, using the [Storage Blob Data Contributor](#) role. To enable other users to use this storage account after you create your workspace, perform these tasks:

- Assign other users to the **Contributor** role on workspace
- Assign other users the appropriate **Synapse RBAC roles** using Synapse Studio
- Assign yourself and other users to the **Storage Blob Data Contributor** role on the storage account

[Learn more](#)

[Review + create](#) < Previous Next: Security >





19

[Home](#) > [All resources](#) > [Create a resource](#) > [Azure Synapse Analytics](#) >

Create Synapse workspace


Provide credentials that can be used for administrator access to the workspace's SQL pools. If you don't provide a password, one will be automatically generated. You can change the password later.


SQL Server admin login * 

SQL Password  ✓

Confirm password ✓


System assigned managed identity permission


Select to grant the workspace network access to the Data Lake Storage Gen2 account using the workspace system identity. [Learn more](#) 

☐ Allow network access to Data Lake Storage Gen2 account. 

 The selected Data Lake Storage Gen2 account does not restrict network access using any network access rules, or you selected a storage account manually via URL under Basics tab. [Learn more](#) 

Workspace encryption

 Double encryption configuration cannot be changed after opting into using a customer-managed key at the time of workspace creation.

Choose to encrypt all data at rest in the workspace with a key managed by you (customer-managed key). This will provide double encryption with encryption at the infrastructure layer that uses platform-managed keys. [Learn more](#) 

Double encryption using a customer-managed key ☐ Enable ☒ Disable

[Review + create](#)

[< Previous](#)

[Next: Networking >](#)



20


[Home](#) > [All resources](#) > [rg_rdcdatastorage](#) > [Create a resource](#) > [Azure Synapse Analytics](#) >

Create Synapse workspace


* Basics * Security **Networking** Tags [Review + create](#)

Configure networking options for your workspace.


Managed virtual network

Choose whether to set up a dedicated Azure Synapse-managed virtual network for your workspace. [Learn more](#) 

Managed virtual network  ☐ Enable ☒ Disable

 To control public network access to your Synapse workspace, you must enable managed virtual network.


Firewall rules

 Azure Synapse Studio and other client tools will only connect to the workspace endpoints if this setting is selected. Connections from specific IP addresses or all Azure services can be allowed or disallowed after the workspace is provisioned.

Allow connections from all IP addresses to your workspace's endpoints. You can restrict these permissions to just Azure datacenter IP addresses and/or specific IP address ranges after creating the workspace.

☒ Allow connections from all IP addresses

Encrypted connections

Azure Synapse Analytics applies TLS1.2 encryption protocol for all network connections. [Learn more](#) 

Minimum TLS version

[Review + create](#)

[< Previous](#)

[Next: Tags >](#)



21

New link

Select Storage Account

Select the storage account that you want link to the Dataverse environment. The storage account must be in the same region as your environment.

Your environment is located in: **North Europe**

Please attach a storage account in one of the following location(s): **North Europe**

☒ Connect to your Azure Synapse Analytics workspace

Subscription *

Resource group *

Workspace name *

Storage account *

As part of linking the Dataverse environment to a data lake, you are granting the Azure Synapse Link service additional roles to your storage account. By using the Azure Synapse Link service, you agree that data may go outside of Power Apps' compliance boundary. For more information, see the Privacy Notice in the technical documentation for the service.

Back **Next** **Cancel**

22

Manage tables for

Manage tables
2 of 251 selected

When you add new table(s), all changes for existing tables are temporarily paused. We will resume writing changes for existing table(s) after we complete exporting data for new table(s) to the Azure data lake.

Advanced

☒ Show advanced configuration settings

Search:

Table	Name	Append only	Partition
<input checked="" type="radio"/> Account	account	<input type="checkbox"/>	Month
<input type="radio"/> Accountleads	accountleads	<input type="checkbox"/>	Month
<input type="radio"/>		<input type="checkbox"/>	Month

Back **Save** **Cancel**

23


Refresh Manage tables Unlink Go to Azure data lake Go to Azure Synapse Analytics workspace All Filter

All changes for existing tables are temporarily paused when we are in the process of exporting data for new table(s). We will resume writing changes for existing table(s) after we complete exporting data for new table(s) to the Azure data lake.

Azure Synapse Link for Dataverse >

Tables Details Discover hub

Table ↑	Name	Sync status	Last synchronized on	Count	Append only	Partition
Account	account	Active	01/20/2022 10:51:56 AM	3	No	Month
Contact	contact	Initial sync in progress	01/20/2022 11:35:05 AM	-	Yes	Year



24

inetum Power Apps Search

Manage tables Unlink Tables Details Go to Azure data lake Go to Azure Synapse Analytics workspace

Azure Synapse Link for Dataverse

Links


Manage tables
Unlink
Tables
Details
Go to Azure data lake
Go to Azure Synapse Analytics workspace

+ New link Import from solution Refresh

Azure Synapse Link for Dataverse

Links

data



25

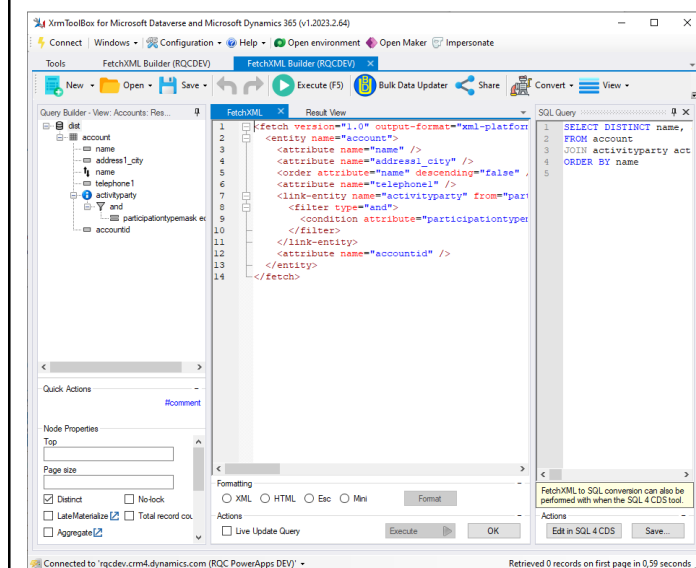
Consuming data
from Azure
Synapse Link for
Dataverse



26

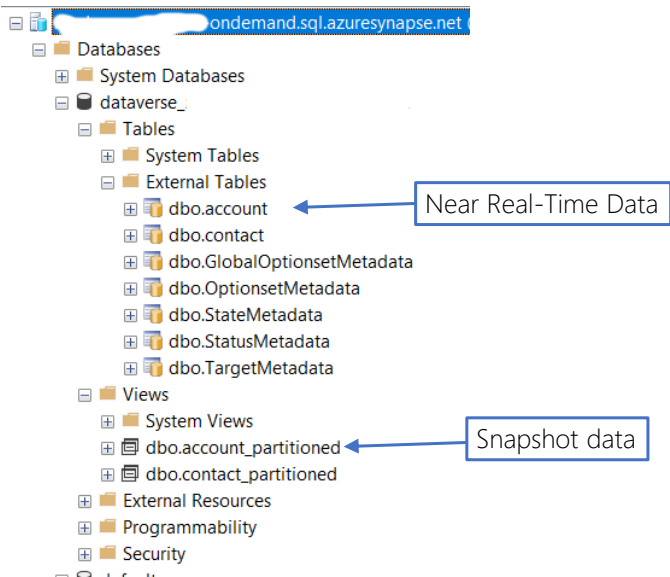
Synapse Studio

- Synapse Studio
<https://web.azuresynapse.net/>
- Use XrmToolBox FetchXML Builder
<https://bit.ly/3ldSy07> and SQL4CDS
to help building queries



27

Near real-time data and read-only snapshots

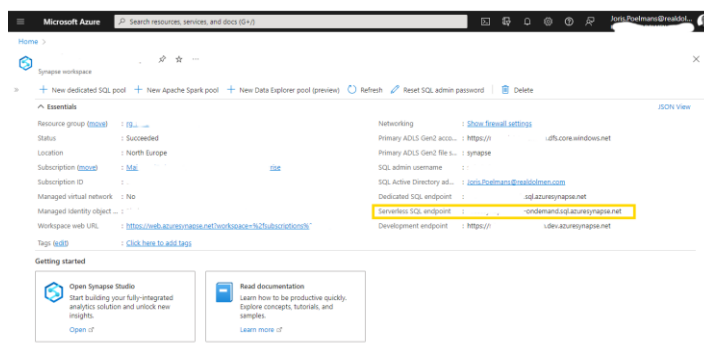


- Two versions of table data in Synapse
 - Near-real time
 - Snapshot - hourly refreshed
- SQL Error - Cannot bulk load because the file ">xxxxx.dfs.core.windows.net/.../2022-08.csv" could not be opened



28

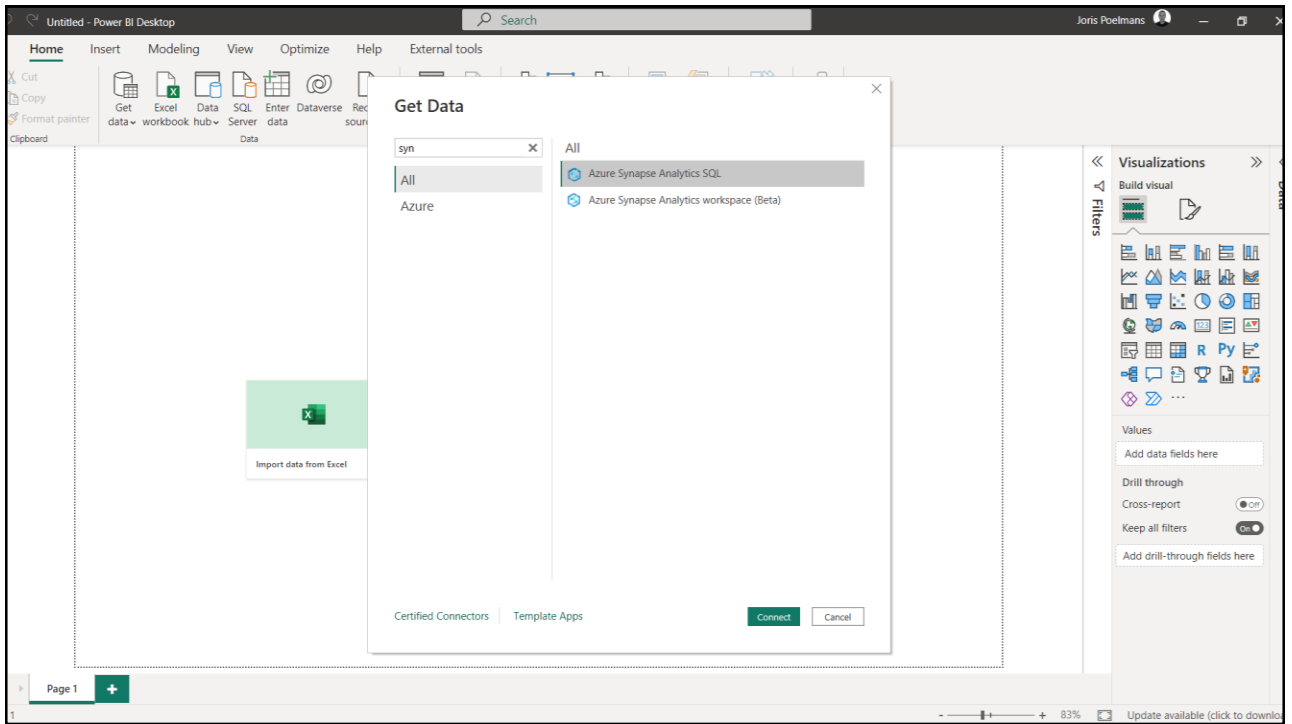
Use SQL Serverless endpoint



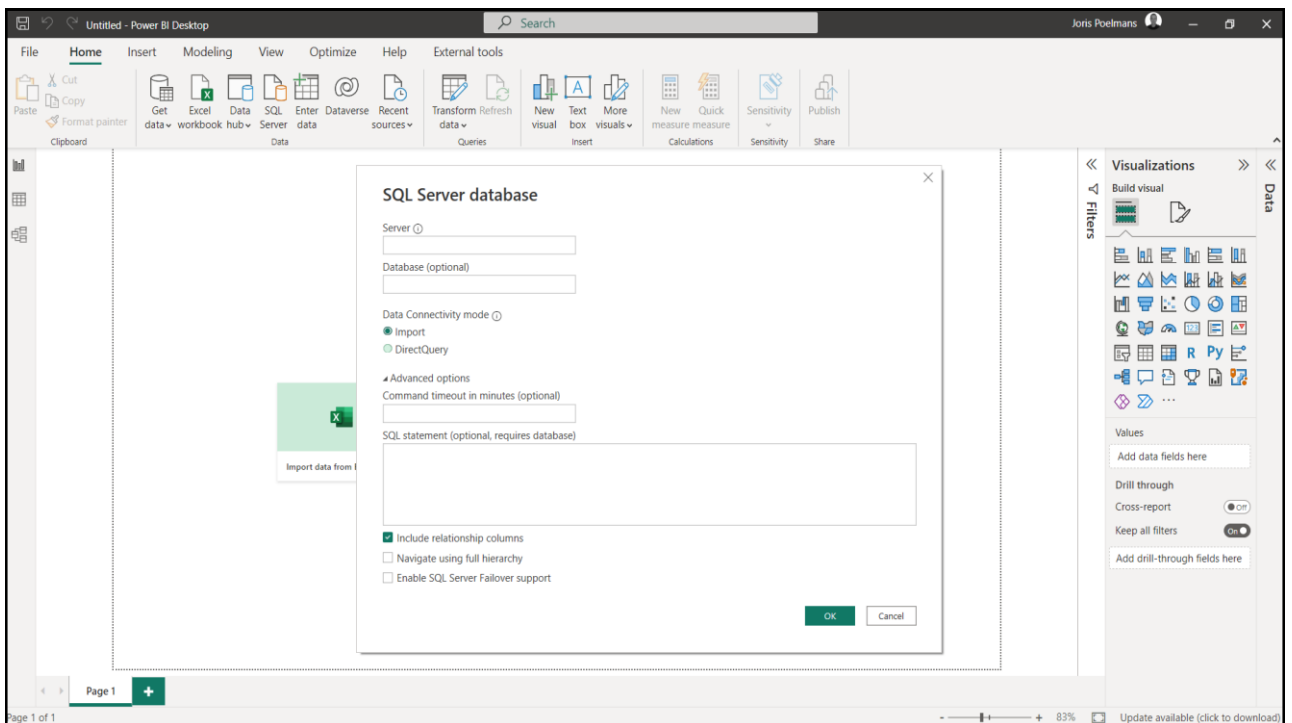
- Use SQL Serverless endpoint
 - SQL Server Management Studio
 - Power BI
- Security gotcha
 - Pass through LakeDB authentication into ADLS



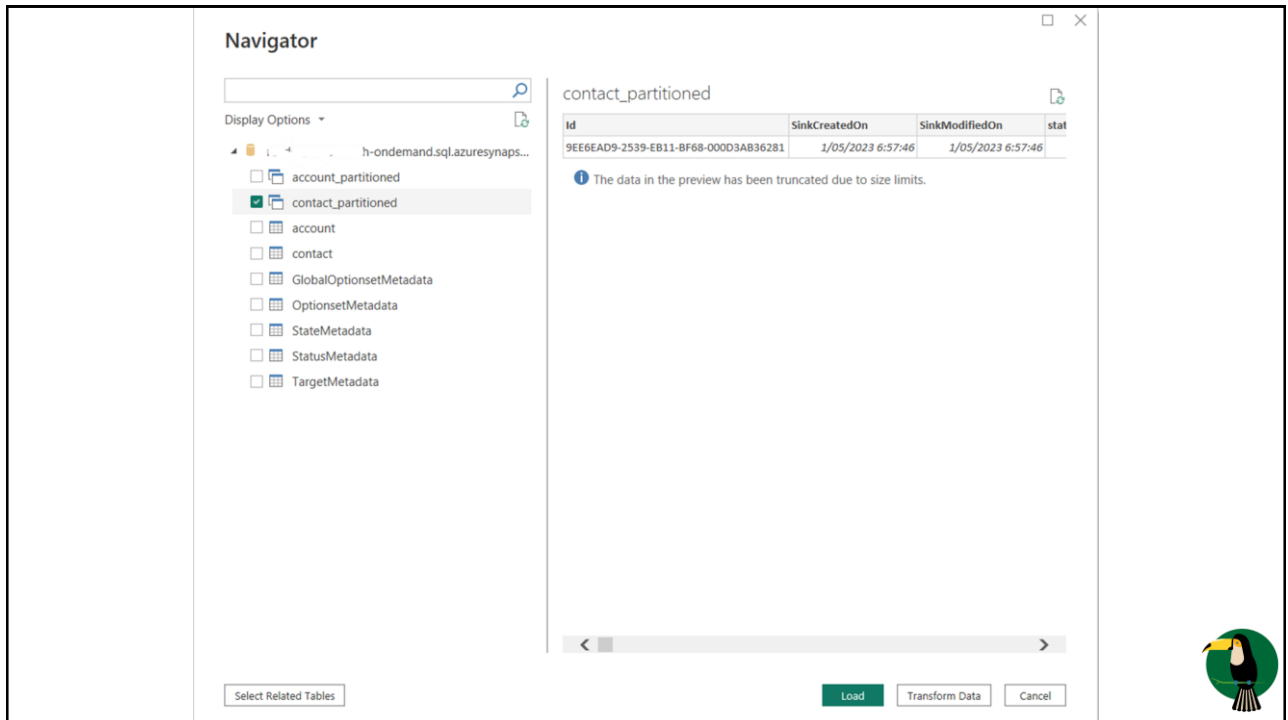
29



30



31



Navigator

Display Options ▾

h-ondemand.sqlazuresynaps...

- ☐ account_partitioned
- ☒ contact_partitioned
- ☐ account
- ☐ contact
- ☐ GlobalOptionsetMetadata
- ☐ OptionsetMetadata
- ☐ StateMetadata
- ☐ StatusMetadata
- ☐ TargetMetadata

Select Related Tables

contact_partitioned

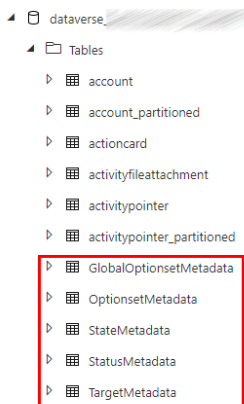
id	SinkCreatedOn	SinkModifiedOn	stat
9EE6EAD9-2539-EB11-BF68-000D3AB36281	1/05/2023 6:57:46	1/05/2023 6:57:46	

The data in the preview has been truncated due to size limits.

Load Transform Data Cancel

32

Working with choice (optionset) values



dataverse

Tables

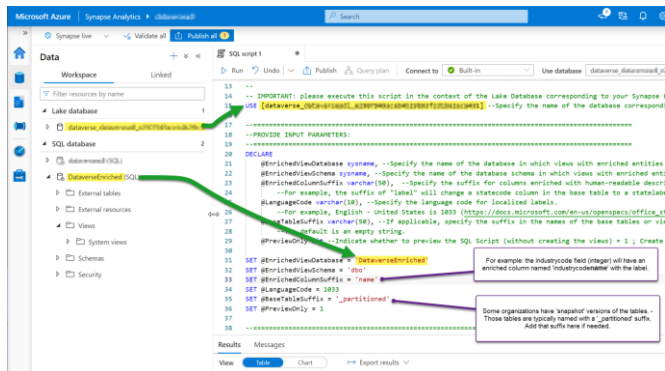
- account
- account_partitioned
- actioncard
- activityfileattachment
- activitypointer
- activitypointer_partitioned
- GlobalOptionsetMetadata**
- OptionsetMetadata
- StateMetadata
- StatusMetadata
- TargetMetadata

- Number of Synapse tables available to get localized text values of choice columns
- SQL Join with OptionSetMetadata to retrieve text labels
- Docs <https://bit.ly/3pNy8EL>

```
SELECT [LocalizedLabel] as [<COLUMN_NAME>] FROM [<DATABASE_NAME>].[dbo].[<TABLE_NAME>_partitioned] LEFT JOIN
[<DATABASE_NAME>].[dbo].[OptionsetMetadata] ON ([<DATABASE_NAME>].[dbo].[OptionsetMetadata].[Option] =
<DATABASE_NAME>].[dbo].[<TABLE_NAME>_partitioned].[<COLUMN_NAME>] AND [<DATABASE_NAME>].[dbo].[OptionsetMetadata].[OptionSetName] =
<COLUMN_NAME>)
```

33

Create views in Azure Synapse



- Lake DB is read-only
- Create views in separate Serverless DB
- Create views for choices – see GitHub repo <https://bit.ly/3o5HEmm>
- Join with other data – see CETAS feature in Azure Synapse



34

CETAS (Create External Table as Select)

- Use CETAS to create external table pointing to CSV, Parquet or Delta Lake
- Export results of a SQL select statement to Hadoop, Azure Blob Storage/ADLS – materialize the results of SELECT statements

```
CREATE EXTERNAL TABLE [db_name].[schema_name].[table_name]
WITH (
    LOCATION = '<yourstorage>',
    DATA_SOURCE = external_data_source_name
    FILE_FORMAT = external_file_format_name
)
AS <select_statement>
```

```
CREATE VIEW [dbo].[account_partitioned] AS (
    SELECT * FROM OPENROWSET (
        BULK (
            N'https://***.dfs.core.windows.net/dataverse-***/account/Snapshot/2020-12_1673414094/*.*csv',
            N'https://***.dfs.core.windows.net/dataverse-***/account/Snapshot/2021-01_1670595293/*.*csv' ),
        FORMAT = 'csv',
        FIELDTERMINATOR = ',',
        FIELDQUOTE = ''
    )
    WITH
        ([id] varchar(36) COLLATE Latin1_General_100_BIN2_UTF8, [SinkCreatedOn] datetime2, [SinkModifiedOn]
        [accountratingcode] bigint, [address_shippingmethodcode] bigint, [businessstypecode] bigint, [shippi
        [address2_shippingmethodcode] bigint, [territorycode] bigint, [accountclassificationcode] bigint, [p
        [paymenttermscode] bigint, [address1 addressstypecode] bigint, [ownershipcode] bigint, [industrycode]
        [address1 freighttermscode] bigint, [customertvcode] bigint, [accountcategoriescode] bigint, [market
```



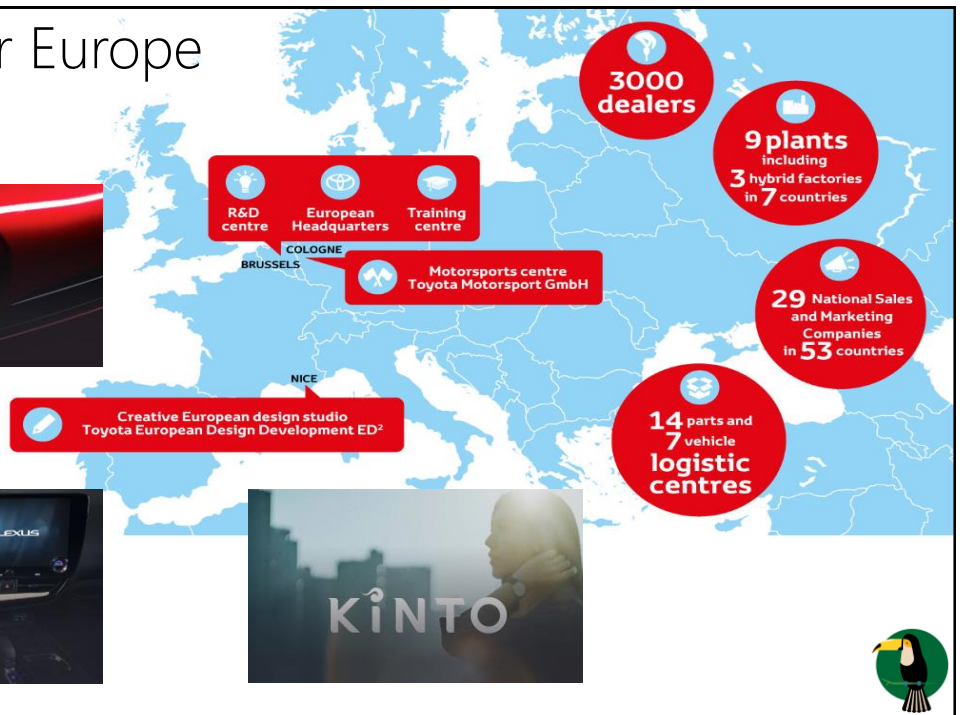
35

Lessons learned at Toyota Motor Europe

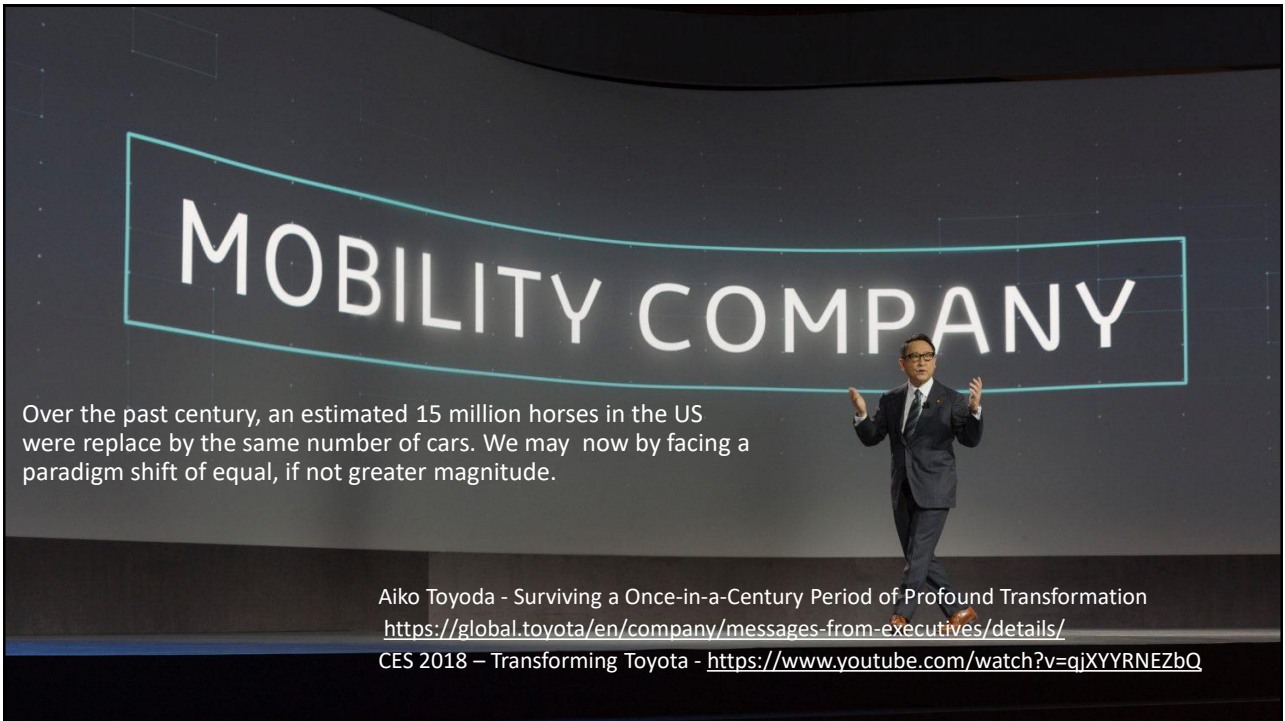


37

Toyota Motor Europe



38




MOBILITY COMPANY

Over the past century, an estimated 15 million horses in the US were replaced by the same number of cars. We may now be facing a paradigm shift of equal, if not greater magnitude.


Aiko Toyoda - Surviving a Once-in-a-Century Period of Profound Transformation
<https://global.toyota/en/company/messages-from-executives/details/>
 CES 2018 – Transforming Toyota - <https://www.youtube.com/watch?v=qjXYRNEZbQ>

39



Consolidate customer touch points throughout the customer lifecycle into a single customer-centric view

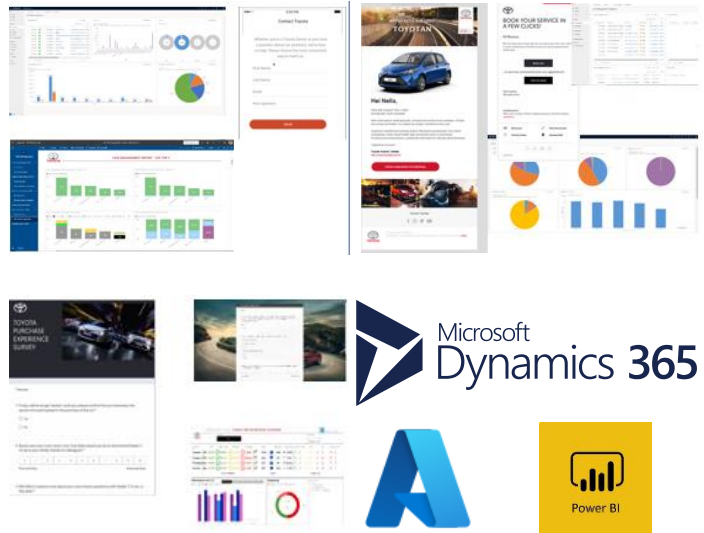
Deliver consistent & personalized customer experiences across the entire lifecycle of marketing, customer service, sales and aftersales, ...



40

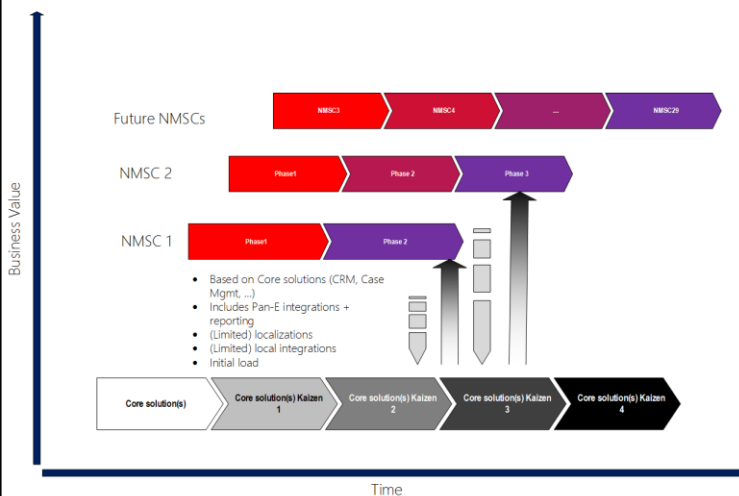
Pan-European CRM platform

- Multiple role-specific business applications on top of **Dynamics 365 CE** to support customer-centric processes (lead management, case management, marketing automation, voice of the customer, etc ...) rolled out to different NMSCs within Europe.
- **Azure** as integration and data platform
- **Power BI** for data visualization and reporting



41

THE CHALLENGE OF SCALE



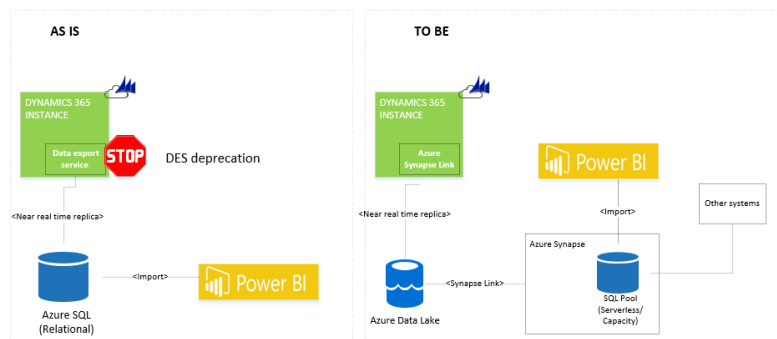
- Production instance per country (data separation) all within TME tenant
- Multiple production instances
- Dataverse holds Customer 360° for millions of customers



43

TRANSITIONING FROM DES TO AZURE SYNAPSE

- Started using Data Export Service for replication to Azure SQL in 2017
- Azure SQL replica used by Power BI, integrations and data pipelines from other teams.
- Deprecation announcement from Microsoft in October 2021 – deadline November 2022.
- Transitioned to Azure Synapse Link for Dataverse in 2022



44

Lessons learned

Initial setup

- Initial setup smoother than expected - successfully transitioned 9 production environments to use Azure Synapse Link in a couple of months time.
- Microsoft support tickets required on multiple occasions
- Use AAD groups to help you set security on ADLS
- No granular security possible in Synapse SQL Serverless
- OAuth2 authentication might be a blocker for other systems.
- Data Factory pipelines and Azure SQL still required.
- ALM for Synapse is quite different from Azure or Power Platform artifacts



45

Lessons learned

*Microsoft support
experience*

- Multiple request to move Dynamics instances to correct region
- Synchronization stopped "Block size exceeded" for large "customer address" table - > 24 mio records. Back-end fix required from Synapse team
- Partitioned views not getting updated correctly – stability improved after fixes by product team in November 2022
- No "partitioned view" available for empty tables – product fix expected May 2023
- ...



46

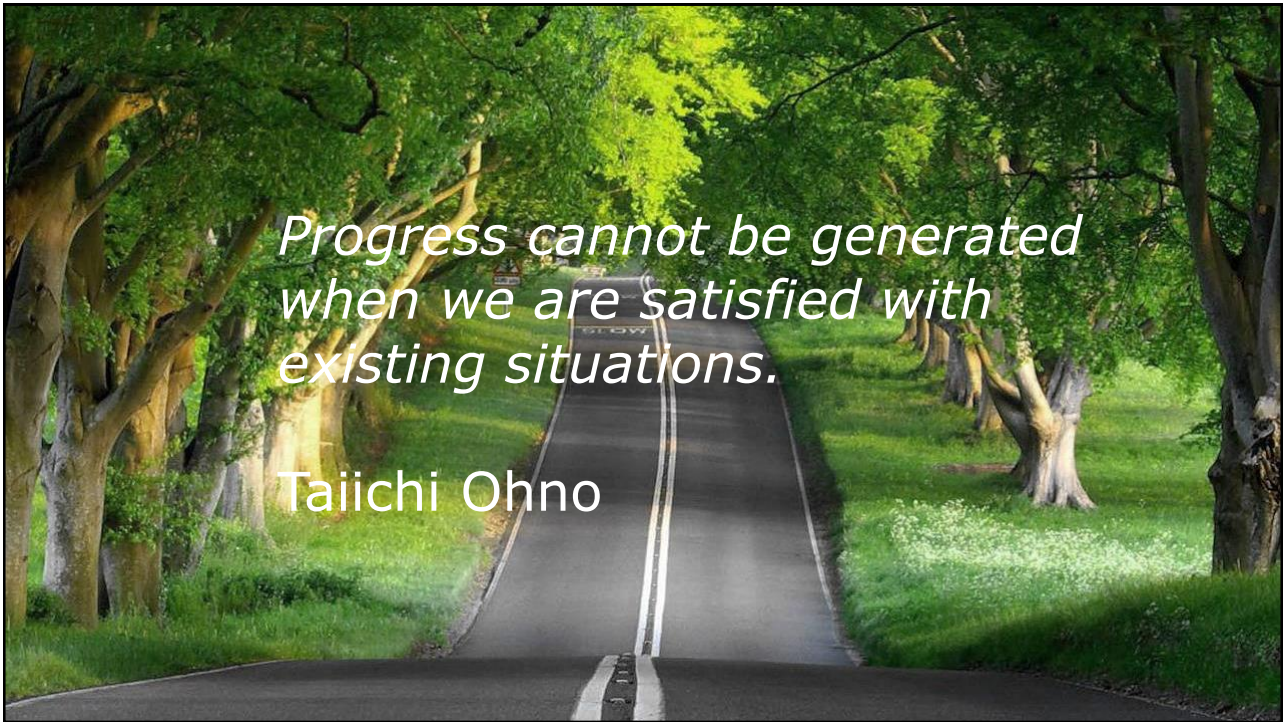
Lessons learned

*Operating and monitoring
the system*

- Monitoring and alerting needs maturing – keep an eye on the release planner <https://aka.ms/releaseplanner> for Synapse status notifications <https://bit.ly/3O7kgQ2>
- Current workaround – select query on sys.views
- Cost forecasting/estimation is difficult
- Optimization required – move from CSV to Parquet



47



48

THANK YOU

- Slides to be published on <https://www.github.com/jorisp> in the Presentation repo
- Follow me on @jopxtwits for notifications on series of blog posts about "Azure Synapse for Dataverse" (or other stuff)



49