

# Intro synapses

Thursday, 15 December 2022 1:52 PM



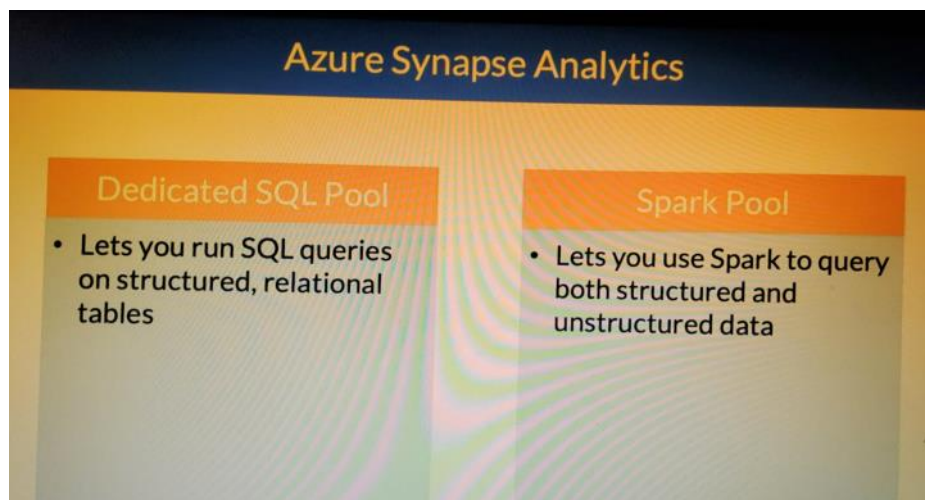
**Guy Hummel**  
Microsoft Certified Azure Cloud Architect and Data Engineer

 [linkedin.com/in/guy-hummel](https://linkedin.com/in/guy-hummel)  
 @GuyRHummel  
 [support@cloudacademy.com](mailto:support@cloudacademy.com)

and I'm a Microsoft Certified Azure Expert. If you have any questions, feel free to connect

Intro to Synapse.

Data Warehouse	Data Lake
<ul style="list-style-type: none"><li>• Structured data (relational tables)</li><li>• Collect data from many databases</li><li>• Designed for reporting</li><li>• Query using SQL</li></ul>	<ul style="list-style-type: none"><li>• Unstructured data (documents, etc.) and structured data</li><li>• Collect all kinds of data in one place</li><li>• Designed for big data analytics</li><li>• Query using Apache Spark</li></ul>



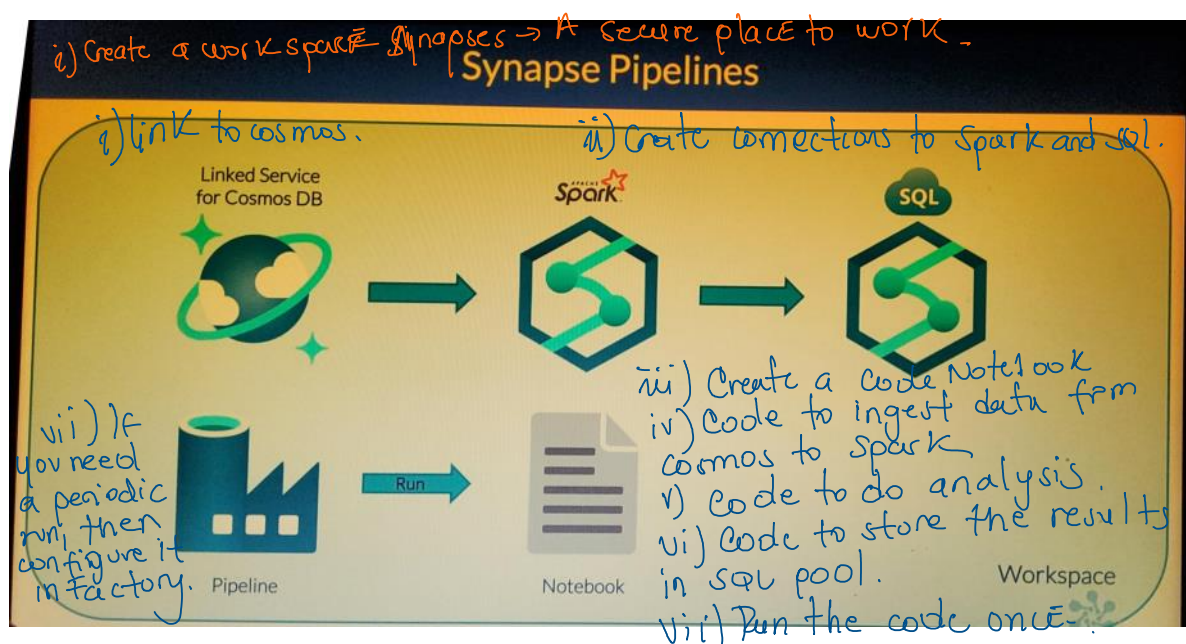
Synapse offers solutions to work with both worlds: unstructured and structured data, in only one service.

→ Synapse also provides a sophisticated tool for getting data from other Azure services and transforming it.

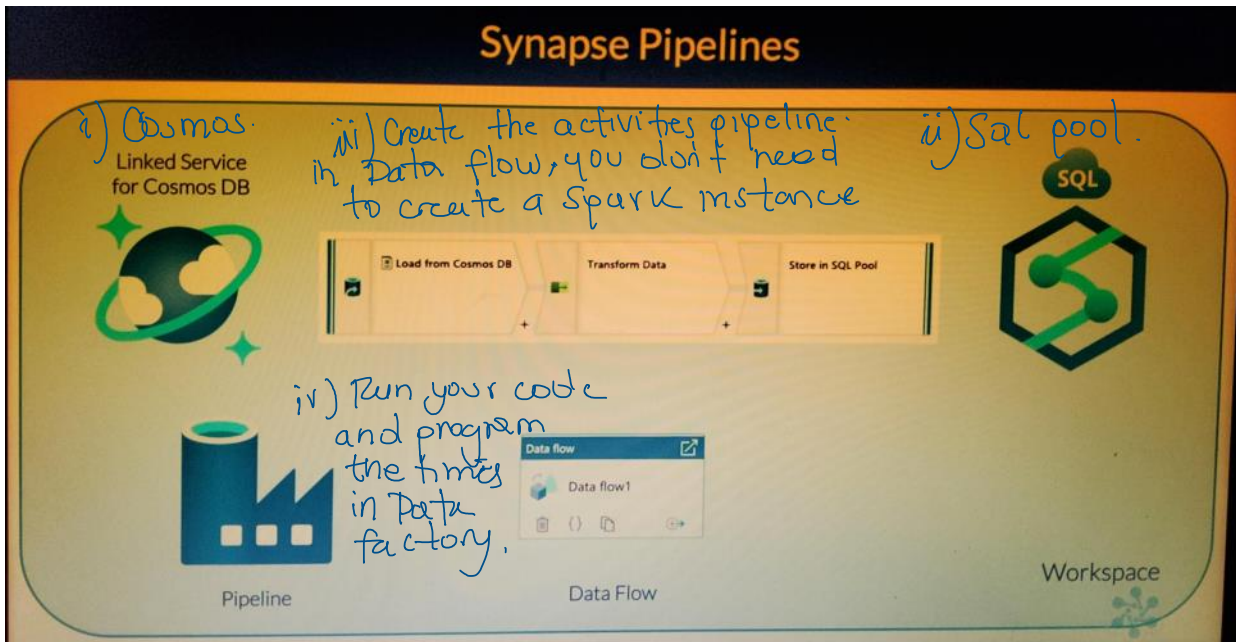
→ Azure Synapse Pipeline is a stripped down version of Data Factory (lets you create data pipelines)

→ 2 approaches → with code  
→ without code.

i) Code option.

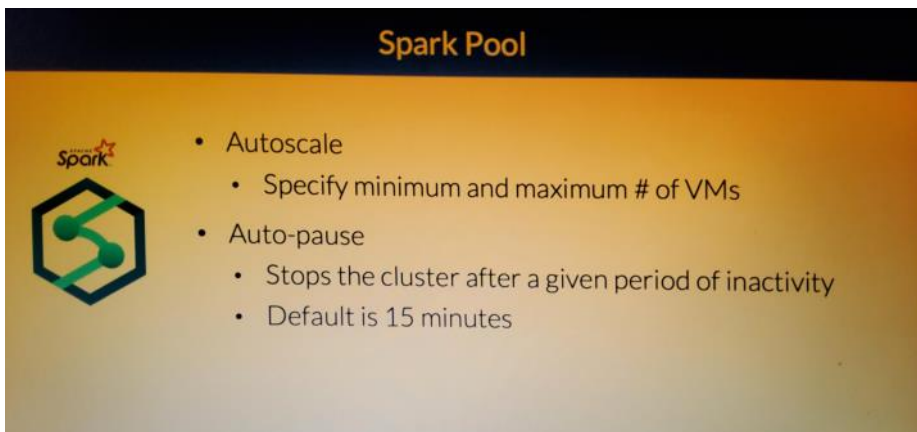


ii) Without code option



Spark and SQL pools → Both are clusters of VM  
 → Cost out a lot of money. Be careful.  
 → Different underlying architecture.

Mechanisms given for Microsoft to manage.  
Spark and SQL pool clusters



Spark

SQL  
 .com has dedicated and serverless pools.



SQL  
→ SQL has dedicated and serverless pools.

## Dedicated Pools

→ New name for Azure SQL Data Warehouse.

### Dedicated SQL Pool



- Formerly known as Azure SQL Data Warehouse
- Provides a compute cluster and storage
- DWUs (Data Warehousing Units) set the amount of CPU, memory, and I/O in the compute cluster
- You can only increase or decrease the # of DWUs manually because there's no autoscaling feature
- Storage space is provided by Azure Storage, so it scales independently from the compute cluster
- You can manually pause it when it's not in use
- When it's paused, you won't pay for the compute cluster, but you'll still pay for the storage

cloud academy

### Serverless SQL Pool



- Don't have their own storage
- Don't have access to data in dedicated SQL pools
- Can query data in other places, such as
  - CSV, Parquet, or JSON files in Azure Storage
  - Azure Open Datasets (publics)
  - External tables in Azure Storage created by Spark pools
- You don't have to pay for the compute resources in them
- You only have to pay for the amount of data processed by your queries
- When you create an Azure Synapse Workspace, it will automatically create a serverless SQL pool

every time they are shutdown