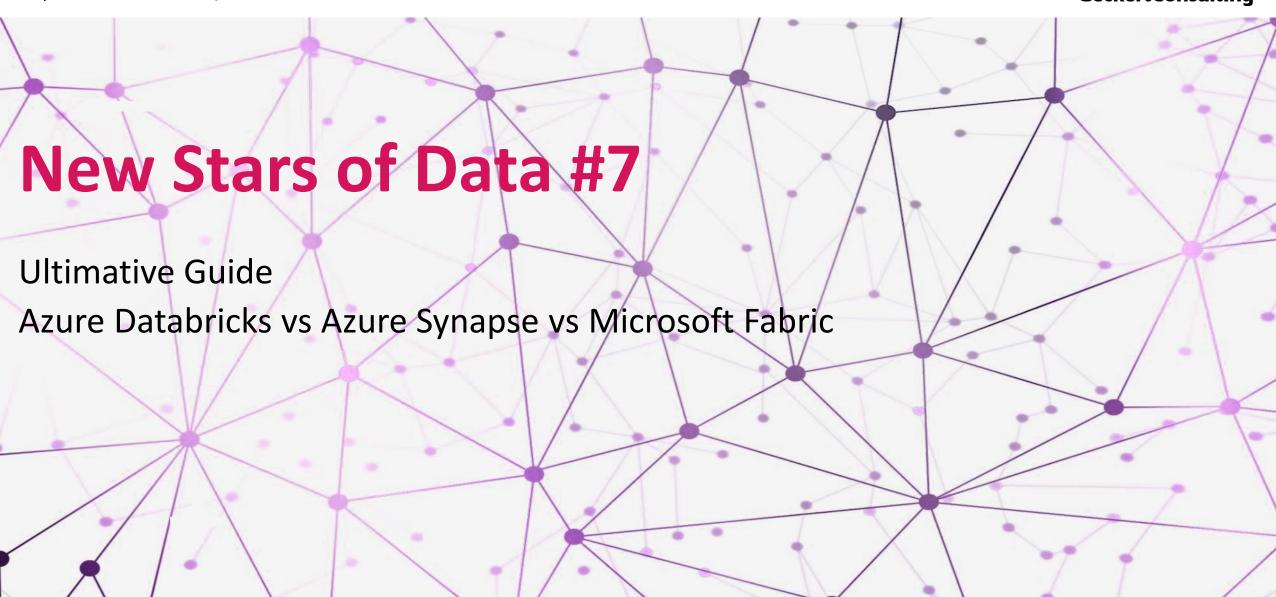
Christopher Geckert





Introduction

- Freelancer
- English/German





Christopher Geckert

Data Solution Architect



Timeline

G

- Look and Feel
- Environment
- Security
- Data Integration
- Data Transformation

Live Demo Databricks



Live Demo Synapse



Live Demo Fabric





• How easy is it to create an environment?

• How can I integrate this into the corporate infrastructure?

• What do I need to do to maintain the infrastructure?









Simplicity

















G

Simplicity – Background info

- DB 2/5
 - No Native ADLS Gen 2 Connection
 - Maintenance of Secret Scope not available via GUI
 - Networking could be complex
 - Entra ID Support not 100% e.g. nested groups
- Synapse 3/5
 - Shared web endpoint could lead to issues
 - Selfhosted IR for internal resources
- Fabric 4/5
 - Pricing
 - Scaling the correct capacity hard to guess

G

Integration – Background info

- **DB** 5/5
 - Supports Vnet, DNS etc.
- Synapse 3/5
 - additional Software required "Selfhosted IR", which is complex to scale
 - because of Managed Vnet loosing control over Network traffic
- Fabric 2/5
 - No Network Integration possible, but has at least data Gateway

G

Operation & Maintenance – Background info

- DB 2/5
 - no native support of system assigned managed identities
 - so needs password/secret maintenance
 - sometimes bad error messages
 - very unhandy user management
- Synapse 3/5
 - bad and sometimes even wrong error messages
 - no native Email support in case a pipeline fails
- Fabric 3/5
 - bad and sometimes even wrong error messages
 - no system assigned managed identities for lower tier workspaces
 - (no native Email support in case a pipeline fails)



• How does network security work for the platform?

What's the user management like?

• Which permission model is available? What's the granularity of the permissions?

• How good is the usability of all these security related modules?









Network







User











Usability (Security)





G

Network – Background info

- DB 5/5
 - Supports Vnet, DNS, Route Tables etc.
- Synapse 3/5
 - Managed Vnet, makes it difficult to control Network traffic
 - Within Azure okay because of Managed PE
- Fabric 1/5
 - No Network Security Available

G

User Management – Background info

- DB 2/5
 - No native Entra support
 - No Nested Entra Groups
- Synapse 5/5
 - Entra fully supported
- Fabric 5/5
 - Entra fully supported

G

Permission Model – Background info

- DB 5/5
- Synapse 3/5
 - No Object/Objecttype level permissions
- Fabric 4/5
 - Roles on Workspacelevel need more granularity

G

Usability (Security) – Background info

- DB 3/5
 - 2 Step user management unhandy
 - can't check key vault connections via GUI
- Synapse 4/5
 - Can't manage SQL Instance Permissions within Synapse
- Fabric 3/5
 - For each Object Permissions but not for Object Types

Data Integration



Can I handle very large source datasets?

Is there a way to save costs on small data?

• How can I connect to the different data sources?

Is there any way to connect to On-Prem data sources?

How is the handling of all this?

Data Integration



Scalability (upscaling)

Scalability (downscaling)

Connectors

On Prem Access

Usability (Integration)

















Scalability (upscaling) – Background info

- **DB** 5/5
- Synapse 5/5
- Fabric 1/5
 - Capacity doesn't scale
 - Can't select Cluster/Capacity for Processing e.g. Transactional Sources Tables



Scalability (downscaling) – Background info

- DB 5/5,
- Synapse 3/5
 - Data flow 1/5 even the smallest Data flow uses 8 vCores
 - Notebook 4/5 smallest node size is still has 4 vCores and 32 GB RAM,
- Fabric 1/5
 - Capacity doesn't scale
 - Can't select Cluster/Capacity for Processing e.g. small excel files

G

Connectors – Background info

- **DB** 5/5
- Synapse 4/5
 - Copy Data Task and Data flows have a preset of inbuild connectors. Can't access Systems that are not supported (Except with Selfhosted IR and additional ODBC drivers)
- Fabric 4/5
 - Fabric Dataflows have a preset of inbuild connectors. Can't access Systems that are not supported

G

On Prem Access – Background info

- DB 5/5
- Synapse 3/5
 - Synapse has Selfhosted IR but this doesn't autoscale
 - Also it still pushes data via the Internet
 - Notebooks can't access On Prem Data
- Fabric 3/5
 - Dataflow can use On-Premise Gateway, but still there is Traffic over the internet
 - Notebooks can't access On Prem Data



Usability (Integration) – Background info

- DB 3/5
 - You can't open several Notebooks via GUI, always new Browser Tabs
- Synapse 4/5
 - Data flows don't support copy paste
 - especially large Dataflows are super unhandy to use
- Fabric 5/5

Data Transformation



Could I process large Facts?

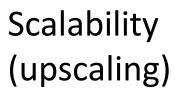
Is there a way to save costs on smaller dimensions?

Can I perform all the tasks I need?

• How complex is this platform to use?

Data Transformation

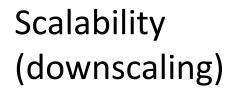


















Function Volume







Usability (Transformation









Scalability (upscaling) – Background info

- DB 5/5,
- Synapse 5/5,
- Fabric 1/5
 - Capacity doesn't scale
 - Can't select Cluster/Capacity for Processing e.g. Transactional Data



Scalability (downscaling) - Background info

- DB 5/5,
- Synapse 3/5
 - Data flow 1/5 even the smallest Data flow uses 8 vCores
 - Notebook 4/5 smallest node size is still uses 4 vCores and 32 GB RAM
- Fabric 1/5
 - Capacity doesn't scale
 - Can't select Cluster/Capacity for Processing e.g. Master Data

G

Function Volume – Background info

- DB 5/5,
- Synapse 3/5
 - Data flows have limited amount of Transformations
 - Notebooks mssparkutil has less functions than dbutil
- Fabric 4/5
 - Data flows Gen 2 have some limitations with exotic files
 - Notebooks mssparkutil has less functions than dbutil



Usability (Transformation) – Background info

- DB 2/5
 - always open new Browser Tab
 - Handling between, SQL Warehouse and Notebooks completely different
- Synapse 4/5
 - Mapping Dataflows especially big data flows get super hard to use
- Fabric 4/5
 - currently no Folders

Conclusion









Environment	10/15	9/15	10/15
Security	14/20	15/20	13(15)/20
Data Integration	23/25	19/25	14(22)/25
Data Transformation	17/20	15/20	10(18)/20
Sum	64/80	58/80	47(65)/80

Best feature

G

- Databricks (Delta Live Tables)
- Synapse (Serverless SQL Pool)
- Fabric (Cluster Startup Time)

Worst 'feature'



- Databricks (Can't Open multiple Notebooks at once)
- Synapse (No SQL Project for SQL Files)
- Fabric (No folders*)

Q&A





Thank you, for your attention



Please evaluate this session



Feel free to contact me

christopher.geckert@geckertconsulting.com



https://www.GeckertConsulting.com