Dave Ruijter

Azure Solution Architect

Macaw Netherlands

@DaveRuijter

ModernData.ai

Battle Of Modern Data Architectures

Have you had that moment, where you are in doubt which Azure service to use for your 'Modern Data Warehousing' solution? So many good options..
Like the Mapping/Wrangling Data Flows capabilities in Azure Data Factory, or the Delta feature in Databricks!

In this session we will look at the different services, compare them using real use-cases, and learn how to choose the best fit for each scenario.

dataMinds Connect

Our Partners

KOHERA
element 61
experience & expertise
Microsoft
REALDOLMEN
a Gfi Group company
Sparkle
cegeka
IN CLOSE COOPERATION
SentryOne®
LOOP1
solarwinds
U2U
ORDINA

Datawarehousing in the cloud…

Load and ingest

Transfer and store

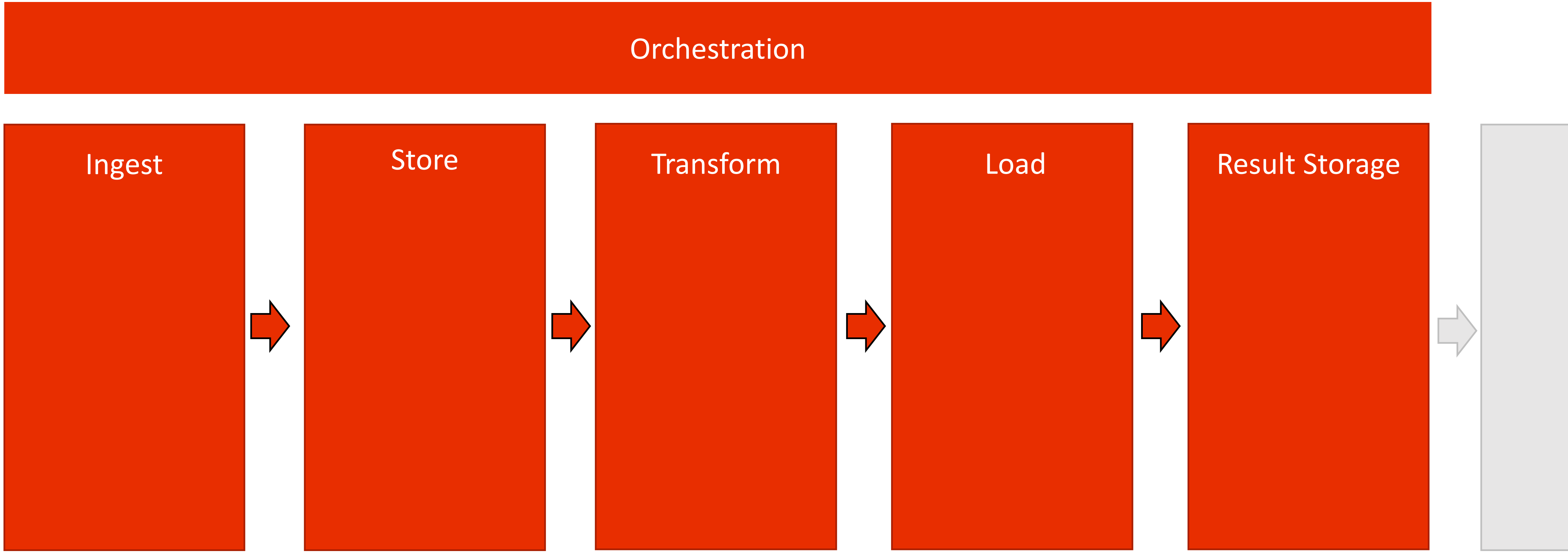
Process

Process and clean

Serve

Serve and analyze

Datawarehousing in the cloud…



Orchestration

Ingest

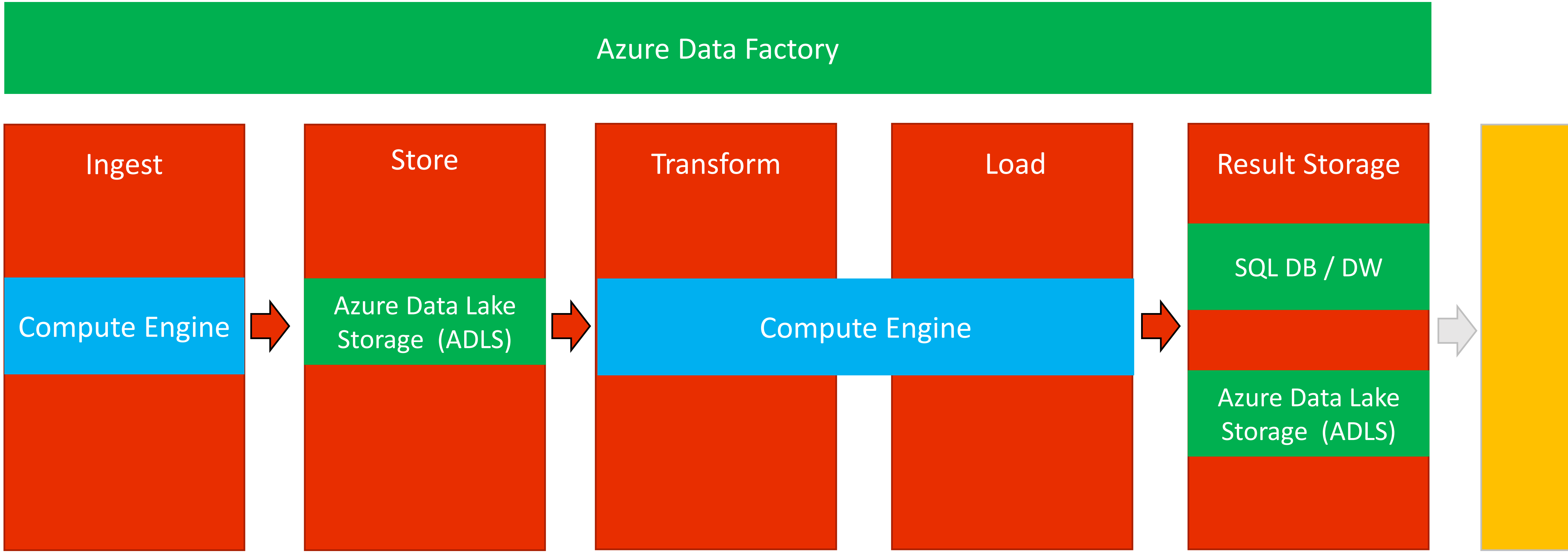
Store

Transform

Load

Result Storage

Datawarehousing in the cloud…



Azure Data Factory

Ingest

Compute Engine

Store

Azure Data Lake Storage (ADLS)

Transform

Load

Compute Engine

Result Storage

SQL DB / DW

Azure Data Lake Storage (ADLS)

Why is it so complicated?

- Various sources/formats
- Schema drift
- Scalability / “big data”
- Monitoring & Auditing
- Keep up with the business needs
- Version control
- ALM / DevOps

Datawarehousing in the cloud…

Logs, files, and media (unstructured)
Business/custom apps (structured)
Ingest
Azur
Azure Analysis Services
4
Power BI
IT DEPENDS

Datawarehousing in the cloud…

- Azure SQL /w PolyBase
- HDInsight
- Azure Data Lake Analytics
- Azure SSIS Integration Runtime
- Azure Data Factory (Mapping Data Flows)
- Azure Databricks
- Power BI dataflows

Datawarehousing in the cloud…

- Azure SQL /w PolyBase
- HDInsight
- Azure Data Lake Analytics
- Azure SSIS Integration Runtime
- Azure Data Factory (Mapping Data Flows)
- Azure Databricks
- Power BI dataflows

Datawarehousing in the cloud…

- ~~Azure SQL /w PolyBase~~
- HDInsight
- Azure Data Lake Analytics
- Azure SSIS Integration Runtime
- Azure Data Factory (Mapping Data Flows)
- Azure Databricks
- Power BI dataflows

Datawarehousing in the cloud…

- ~~Azure SQL /w PolyBase~~
- ~~HDInsight~~
- Azure Data Lake Analytics
- Azure SSIS Integration Runtime
- Azure Data Factory (Mapping Data Flows)
- Azure Databricks
- Power BI dataflows

Datawarehousing in the cloud…

- ~~Azure SQL /w PolyBase~~
- ~~HDInsight~~
- ~~Azure Data Lake Analytics~~
- Azure SSIS Integration Runtime
- Azure Data Factory (Mapping Data Flows)
- Azure Databricks
- Power BI dataflows

Datawarehousing in the cloud…

- ~~Azure SQL /w PolyBase~~
- ~~HDInsight~~
- ~~Azure Data Lake Analytics~~
- Azure SSIS Integration Runtime
- Azure Data Factory (Mapping Data Flows)
- Azure Databricks
- Power BI dataflows

Datawarehousing in the cloud…

- ~~Azure SQL /w PolyBase~~
- ~~HDInsight~~
- ~~Azure Data Lake Analytics~~
- Azure SSIS Integration Runtime
- Azure Data Factory (Mapping Data Flows)
- Azure Databricks
- Power BI dataflows

Datawarehousing in the cloud…

- ~~Azure SQL /w PolyBase~~
- ~~HDInsight~~
- ~~Azure Data Lake Analytics~~
- Azure SSIS Integration Runtime
- Azure Data Factory (Mapping Data Flows)
- Azure Databricks (Delta)
- Power BI dataflows

Datawarehousing in the cloud…

- ~~Azure SQL /w PolyBase~~
- ~~HDInsight~~
- ~~Azure Data Lake Analytics~~
- Azure SSIS Integration Runtime
- Azure Data Factory (Mapping Data Flows)
- Azure Databricks (Delta)
- ~~Power BI dataflows~~

The battle!

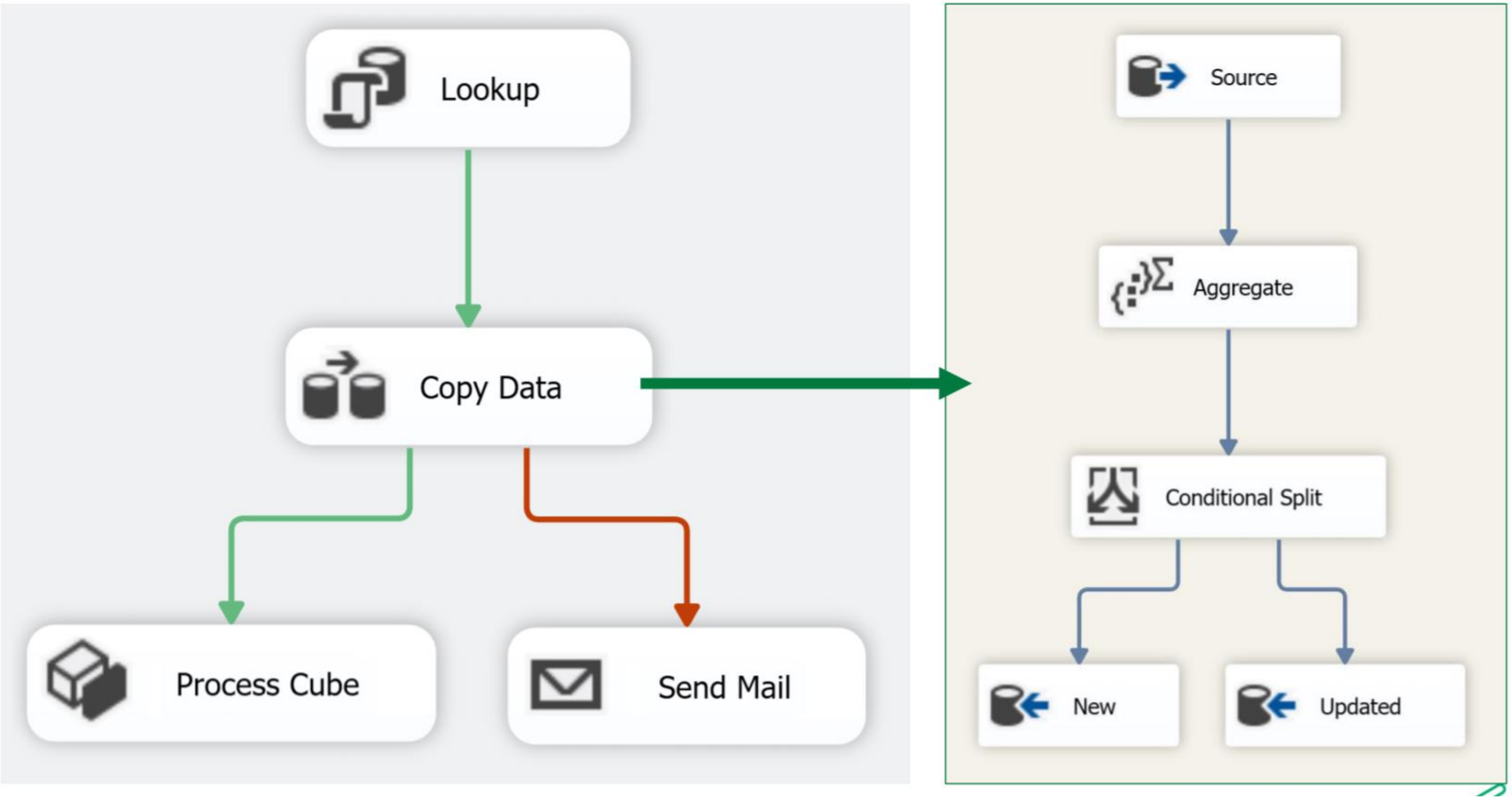
- round #1: capabilities
- round #2: developer experience
- round #3: operator experience
- round #4: security
- round #5: roadmap / future readiness
- round #6: coolness

01

capabilities

round.set(1)

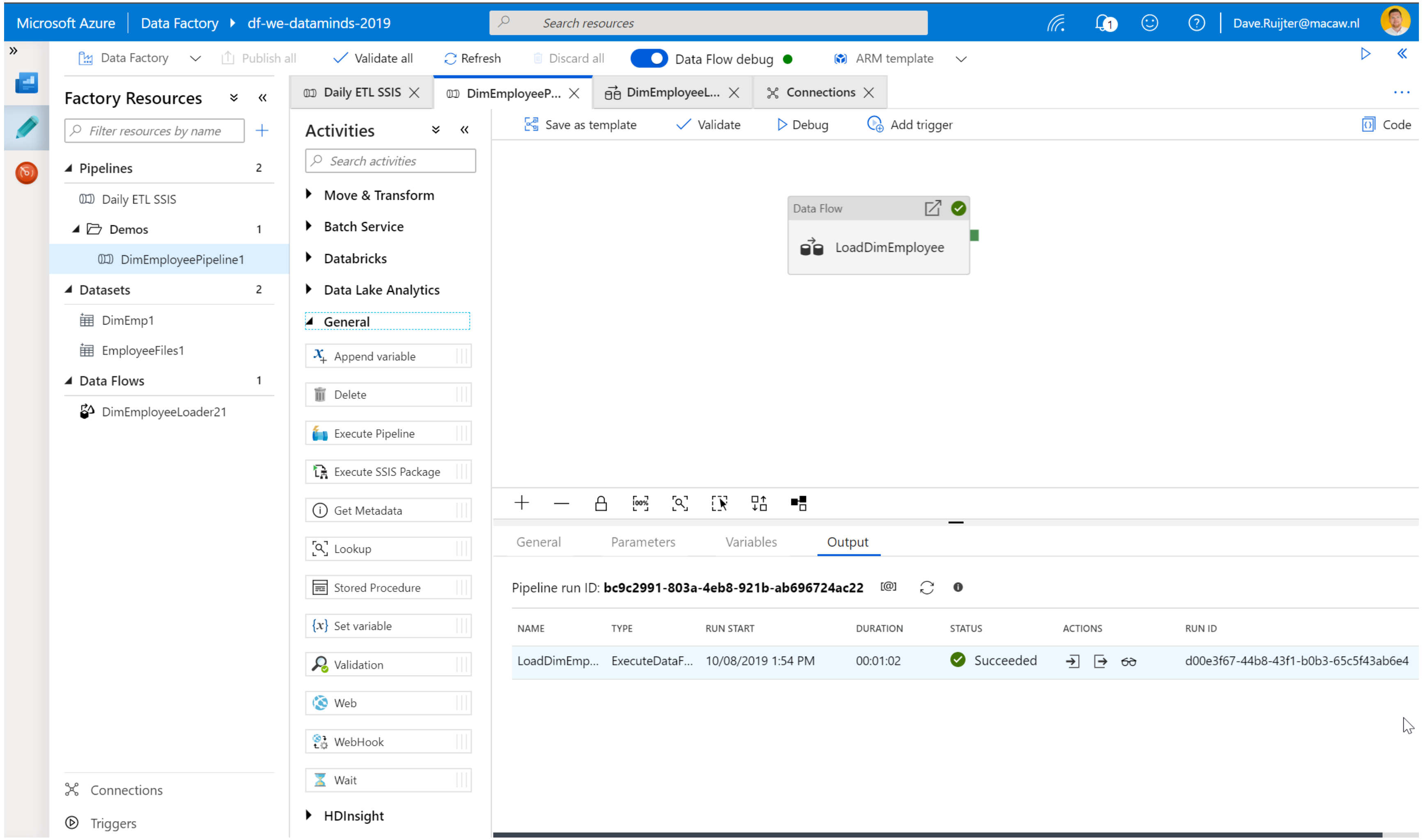
Capabilities – SSIS in Azure



Capabilities – SSIS in Azure

- Pros
  - Combination of simple & advanced options
  - Continue with existing solution / less initial investment
  - Mature tooling for meta-driven generation
  - Git integration in VS
- Cons:
  - Not serverless
  - No further development (continuity?)
  - Integration with Azure PaaS (data Source connectivity)
  - No advanced analytics / data science
  - No streaming data support

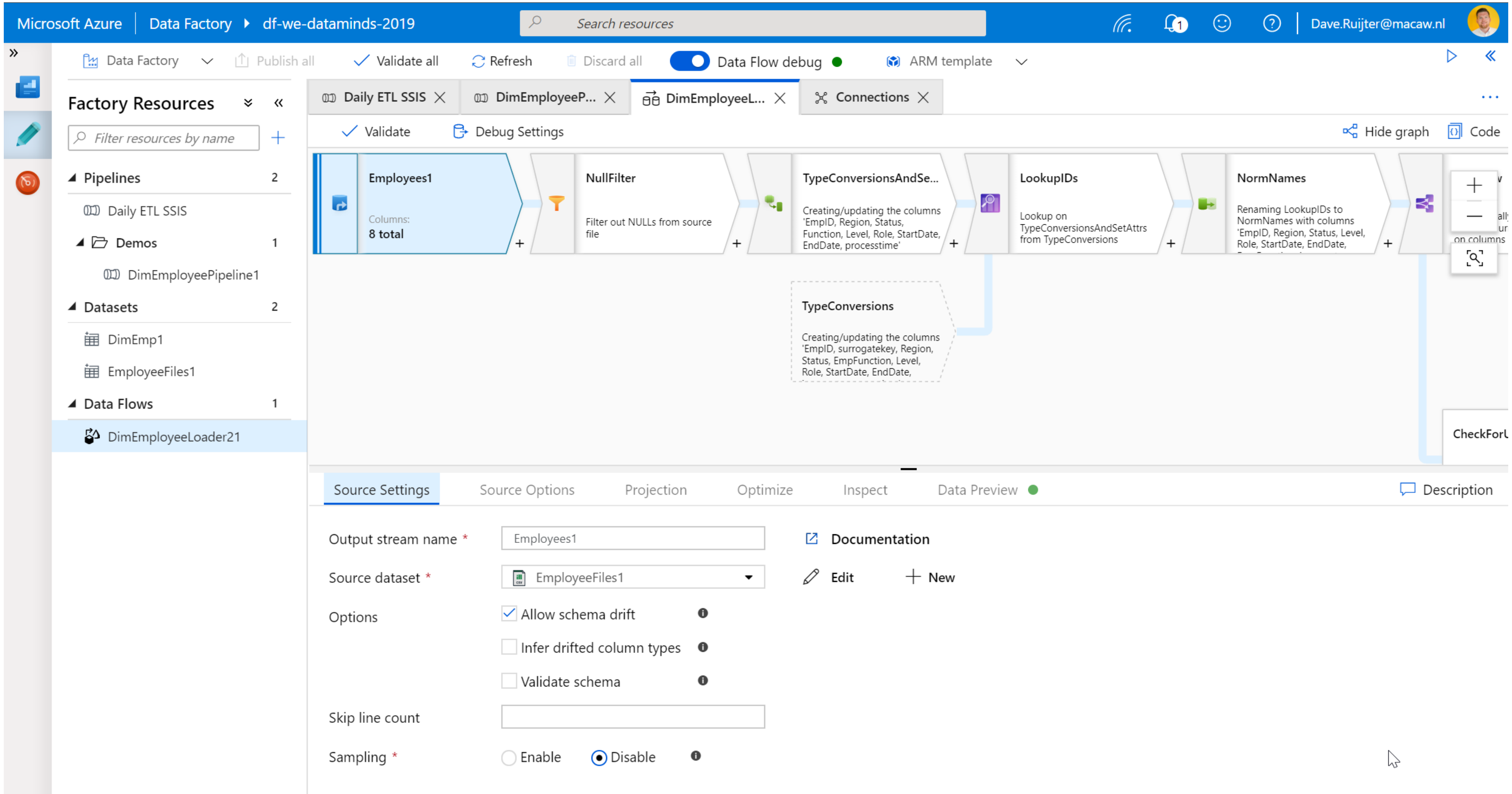
Capabilities - ADF (MDF!)



Pipeline run ID: **bc9c2991-803a-4eb8-921b-ab696724ac22**

NAME	TYPE	RUN START	DURATION	STATUS	ACTIONS	RUN ID
LoadDimEmp...	ExecuteDataF...	10/08/2019 1:54 PM	00:01:02	✓ Succeeded		d00e3f67-44b8-43f1-b0b3-65c5f43ab6e4

Capabilities - ADF (MDF!)



Employees1

Columns:
8 total

NullFilter

Filter out NULLs from source file

TypeConversionsAndSe...

Creating/updating the columns 'EmpID, Region, Status, Function, Level, Role, StartDate, EndDate, processtime'

LookupIDs

Lookup on TypeConversionsAndSetAttrs from TypeConversions

NormNames

Renaming LookupIDs to NormNames with columns 'EmpID, Region, Status, Level, Role, StartDate, EndDate,

TypeConversions

Creating/updating the columns 'EmpID, surrogatekey, Region, Status, EmpFunction, Level, Role, StartDate, EndDate,

Output stream name * Employees1

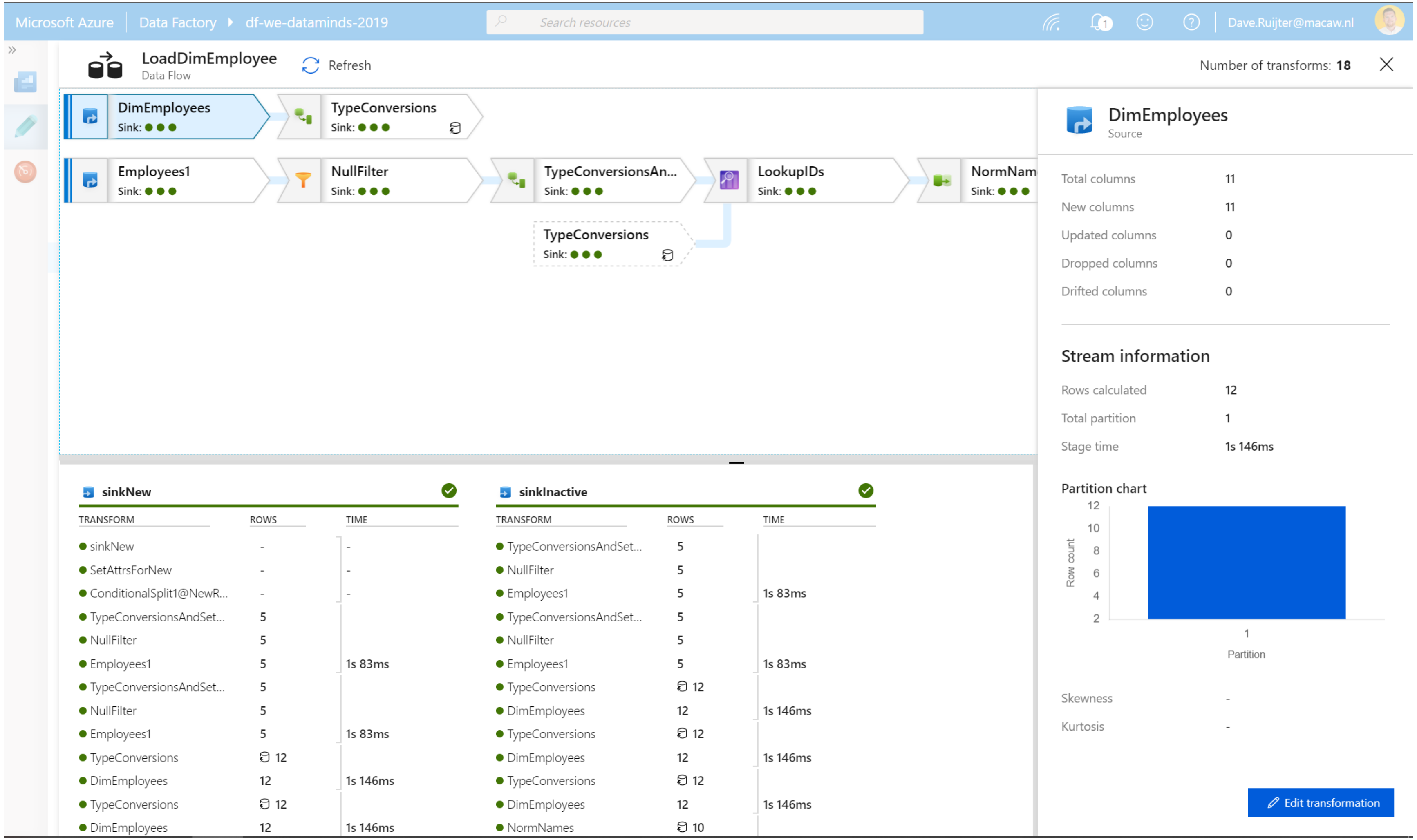
Source dataset * EmployeeFiles1

Options
☑ Allow schema drift
□ Infer drifted column types
□ Validate schema

Skip line count

Sampling * ○ Enable ● Disable

Capabilities - ADF (MDF!)



Number of transforms: **18**

DimEmployees
Source

Total columns	11
New columns	11
Updated columns	0
Dropped columns	0
Drifted columns	0

Stream information

Rows calculated	12
Total partition	1
Stage time	1s 146ms

Partition chart

Skewness	-
Kurtosis	-

sinkNew ✓

TRANSFORM	ROWS	TIME
● sinkNew	-	-
● SetAttrsForNew	-	-
● ConditionalSplit1@NewR...	-	-
● TypeConversionsAndSet...	5	
● NullFilter	5	
● Employees1	5	1s 83ms
● TypeConversionsAndSet...	5	
● NullFilter	5	
● Employees1	5	1s 83ms
● TypeConversions	12	
● DimEmployees	12	1s 146ms
● TypeConversions	12	
● DimEmployees	12	1s 146ms

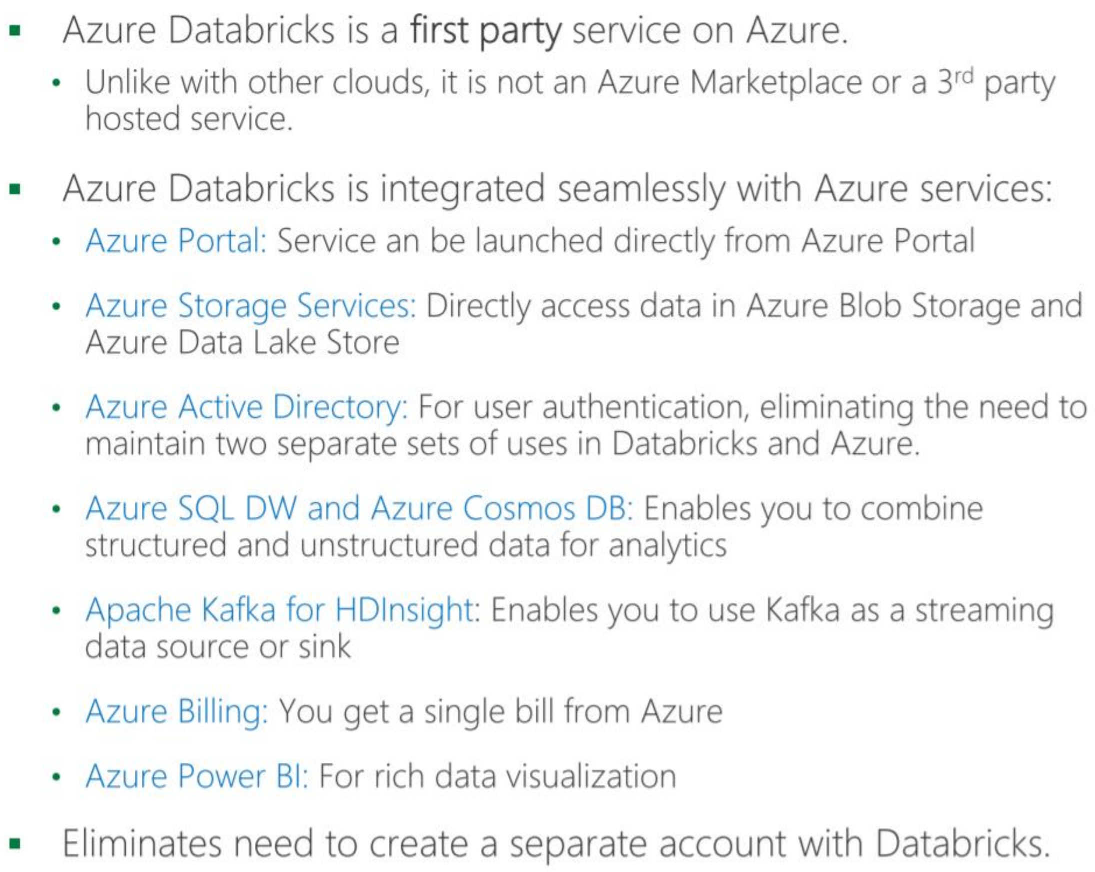
sinkInactive ✓

TRANSFORM	ROWS	TIME
● TypeConversionsAndSet...	5	
● NullFilter	5	
● Employees1	5	1s 83ms
● TypeConversionsAndSet...	5	
● NullFilter	5	
● Employees1	5	1s 83ms
● TypeConversions	12	
● DimEmployees	12	1s 146ms
● TypeConversions	12	
● DimEmployees	12	1s 146ms
● TypeConversions	12	
● DimEmployees	12	1s 146ms
● NormNames	10	

Capabilities - ADF (MDF!)

- Pros
  - Simplicity / easy to understand / no programming required
  - Low time to value
  - DevOps ready (bit clumsy design)
- Cons:
  - New / immature
  - Poor generation using meta-data driven frameworks
  - Limited set of (advanced) transformations
  - Not suited for complex tasks, fallback to functions/notebooks
  - Limited v-net support. MDF / Web Activities can’t access Key Vault.
  - Publishing is a manual action in the browser
  - No advanced analytics / data science
  - No streaming data support

Capabilities - Azure Databricks

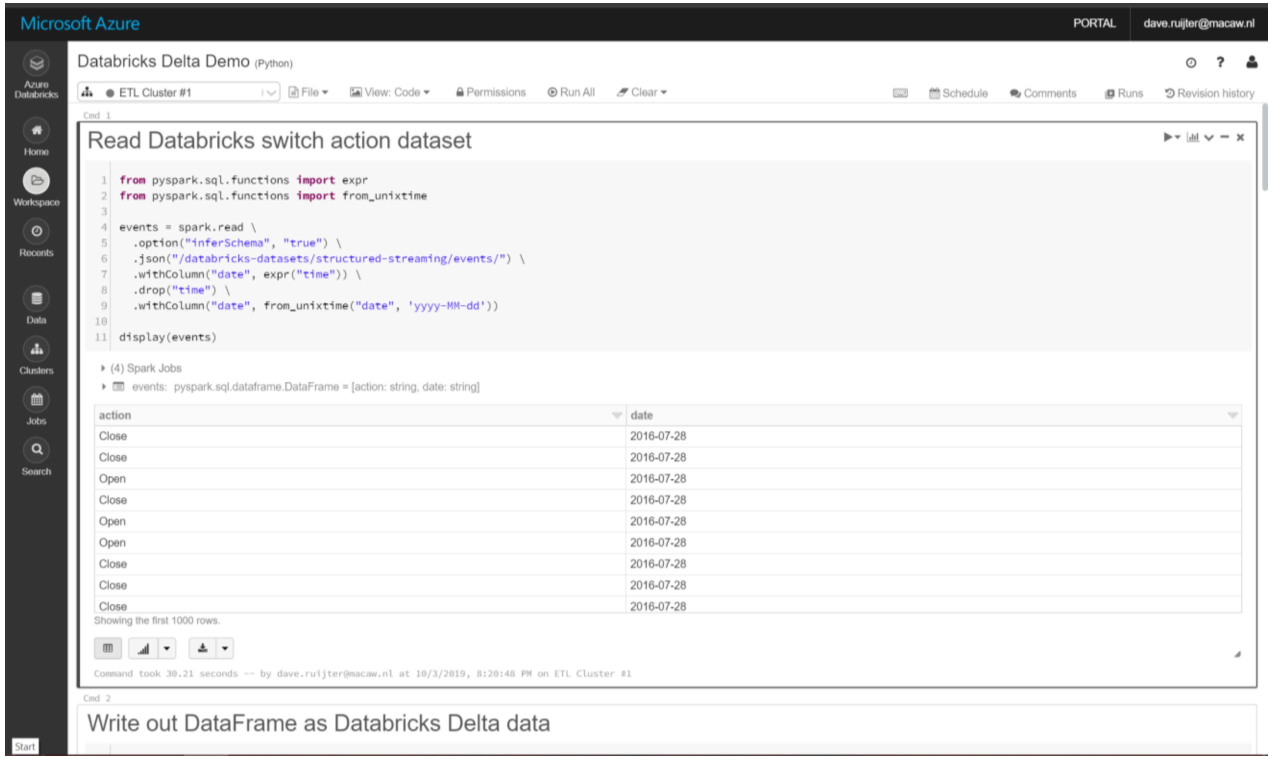


- Azure Databricks is a first party service on Azure.
  - Unlike with other clouds, it is not an Azure Marketplace or a 3rd party hosted service.
- Azure Databricks is integrated seamlessly with Azure services:
  - Azure Portal: Service an be launched directly from Azure Portal
  - Azure Storage Services: Directly access data in Azure Blob Storage and Azure Data Lake Store
  - Azure Active Directory: For user authentication, eliminating the need to maintain two separate sets of uses in Databricks and Azure.
  - Azure SQL DW and Azure Cosmos DB: Enables you to combine structured and unstructured data for analytics
  - Apache Kafka for HDInsight: Enables you to use Kafka as a streaming data source or sink
  - Azure Billing: You get a single bill from Azure
  - Azure Power BI: For rich data visualization
- Eliminates need to create a separate account with Databricks.

Capabilities – Databricks Delta

Essentially, it’s an optimized Spark table with SQL-like features:

- **ACID transactions**
- **DELETES**
  / UPDATES
  / UPSERTS
- **Statistics,**
  data skipping
  and ZORDER clustering



Read Databricks switch action dataset

```
from pyspark.sql.functions import expr
from pyspark.sql.functions import from_unixtime

events = spark.read \
  .option("inferSchema", "true") \
  .json("/databricks-datasets/structured-streaming/events/") \
  .withColumn("date", expr("time")) \
  .drop("time") \
  .withColumn("date", from_unixtime("date", 'yyyy-MM-dd'))

display(events)
```

(4) Spark Jobs

events: pyspark.sql.dataframe.DataFrame = [action: string, date: string]

action	date
Close	2016-07-28
Close	2016-07-28
Open	2016-07-28
Close	2016-07-28
Open	2016-07-28
Open	2016-07-28
Close	2016-07-28
Close	2016-07-28
Close	2016-07-28

Showing the first 1000 rows.

Command took 30.21 seconds -- by dave.ruijter@macaw.nl at 10/3/2019, 8:20:48 PM on ETL Cluster #1

Write out DataFrame as Databricks Delta data

Capabilities - Azure Databricks

- Pros
  - Extremely versatile and scalable
  - Easily add streaming data
  - Not only applicable for data engineering “unified analytics”
  - Interactive notebook experience
  - Cloud agnostic / open source
- Cons:
  - Steep learning curve
  - Not serverless (don't underestimate cluster management)
  - Poor Git integration
  - Longer time to value
  - Poor Service Principal support

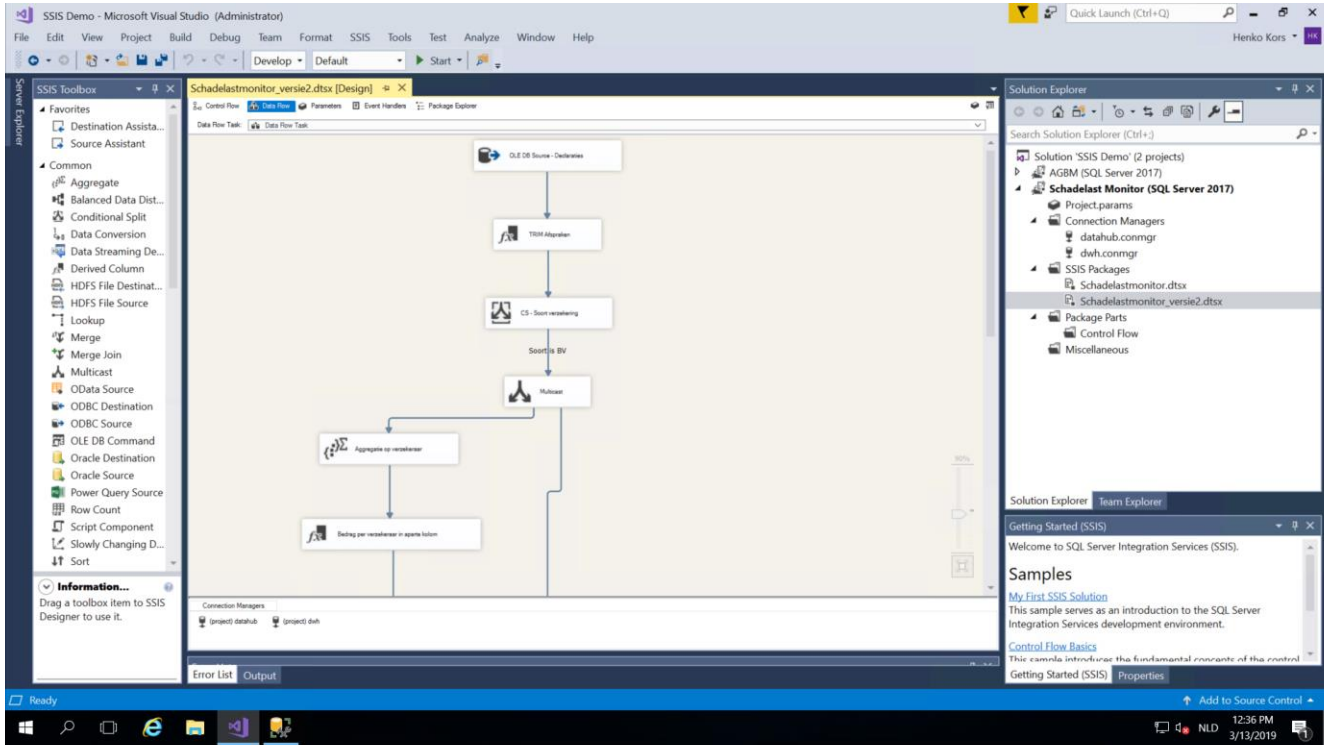
02

developer experience

round.set(2)

Developer experience - SSIS

- Tool: Visual Studio (crashes / manual updates)
- Infrastructure-as-a-Service: Virtual Machine
- Good options to generate code
- Poor collaboration
- Testing via dataviewers
- Disconnected from ADF
- Source code XML
- Schema drift



Drag a toolbox item to SSIS Designer to use it.

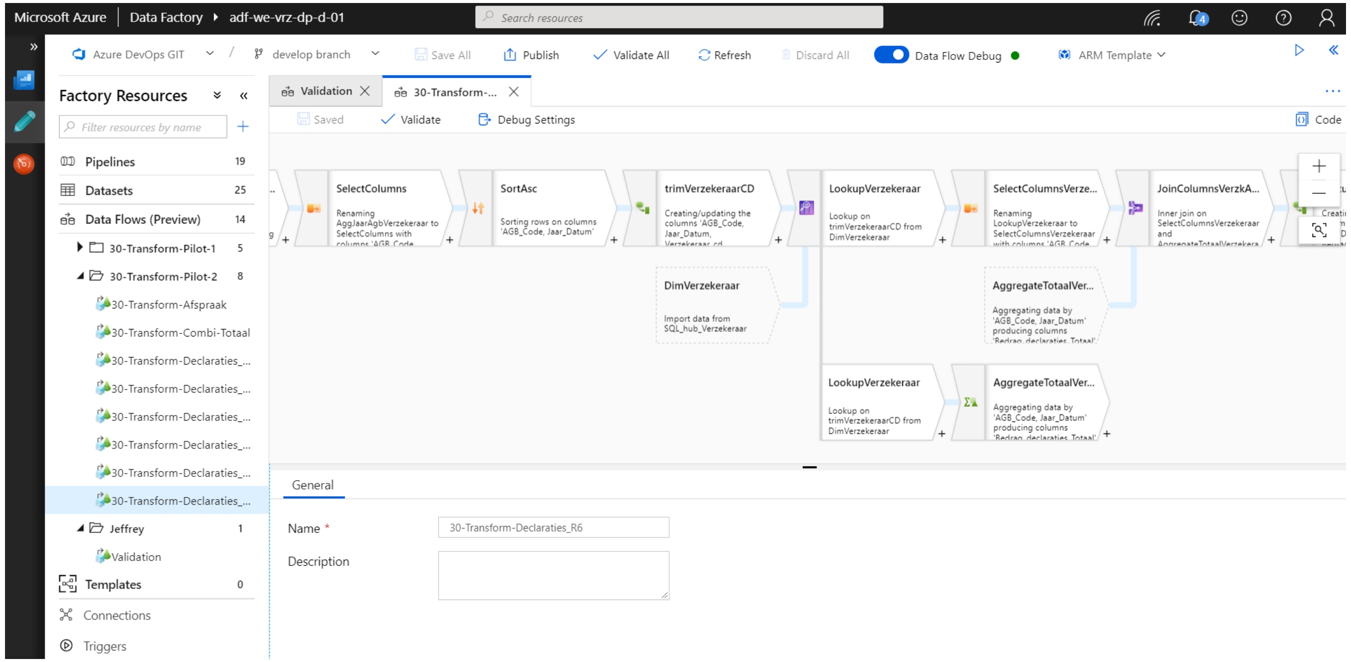
Welcome to SQL Server Integration Services (SSIS).

This sample serves as an introduction to the SQL Server Integration Services development environment.

This sample introduces the fundamental concepts of the control

Developer experience – ADF MDF

- Tool: Browser
- Platform-as-a-Service: Azure Portal
- Simple to start
- Can feel limited
- Breaking changes
- Poor collaboration
- Seamless integration ADF
- Source code is JSON
- Testing via debug-mode
- Schema drift

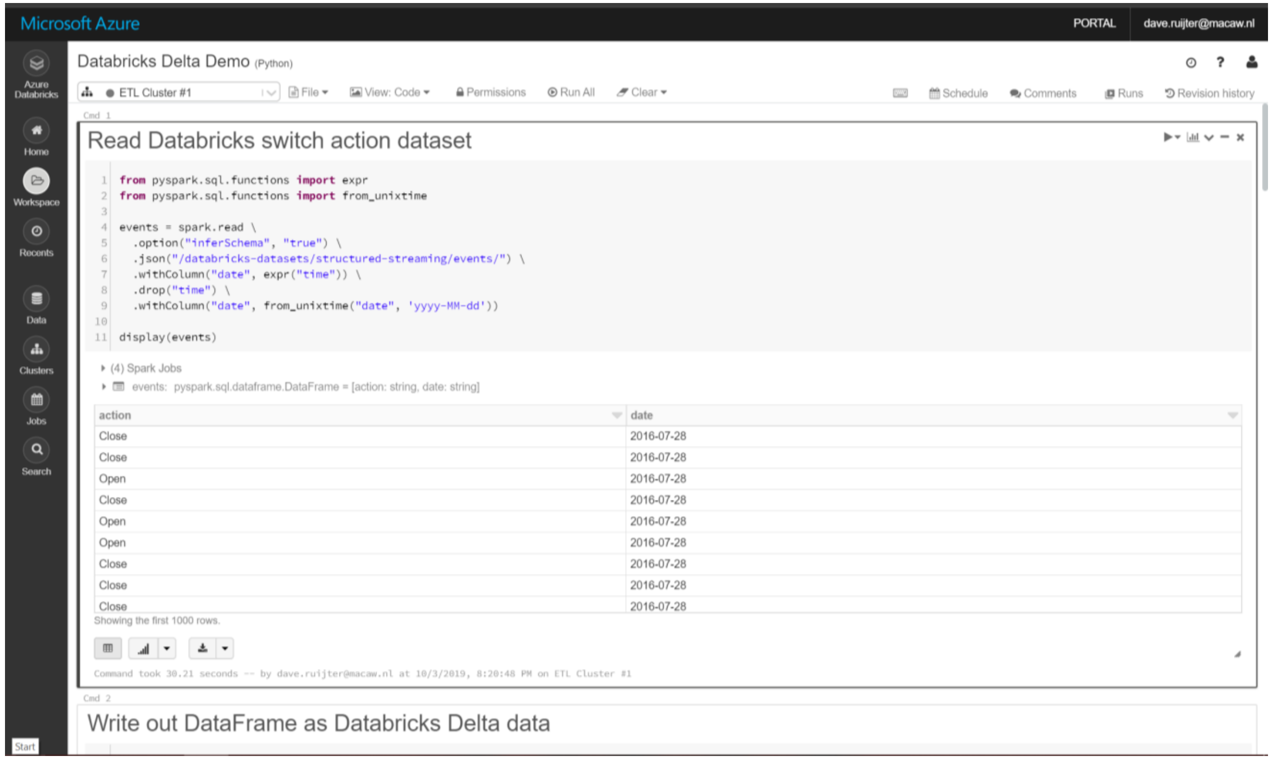

Microsoft Azure
Data Factory
adf-we-vrz-dp-d-01
Search resources
Azure DevOps GIT
develop branch
Save All
Publish
Validate All
Refresh
Discard All
Data Flow Debug
ARM Template
Factory Resources
Filter resources by name
Pipelines 19
Datasets 25
Data Flows (Preview) 14
30-Transform-Pilot-1 5
30-Transform-Pilot-2 8
30-Transform-Afspraak
30-Transform-Combi-Totaal
30-Transform-Declaraties_...
30-Transform-Declaraties_...
30-Transform-Declaraties_...
30-Transform-Declaraties_...
30-Transform-Declaraties_...
30-Transform-Declaraties_...
Jeffrey 1
Validation
Templates 0
Connections
Triggers
Validation
30-Transform-...
Saved
Validate
Debug Settings
Code
SelectColumns
Renaming AggJaarAgbVerzekeraar to SelectColumns with columns 'AGB_Code
SortAsc
Sorting rows on columns 'AGB_Code, Jaar_Datum'
trimVerzekeraarCD
Creating/updating the columns 'AGB_Code, Jaar_Datum, Verzekeraar_cd
DimVerzekeraar
Import data from SQL_hub_Verzekeraar
LookupVerzekeraar
Lookup on trimVerzekeraarCD from DimVerzekeraar
SelectColumnsVerze...
Renaming LookupVerzekeraar to SelectColumnsVerzekeraar with columns 'AGB_Code
JoinColumnsVerzkA...
Inner join on SelectColumnsVerzekeraar and AggregateTotaalVerzekera
AggregateTotaalVer...
Aggregating data by 'AGB_Code, Jaar_Datum' producing columns 'Bedrag_declaraties_Totaal'
LookupVerzekeraar
Lookup on trimVerzekeraarCD from DimVerzekeraar
AggregateTotaalVer...
Aggregating data by 'AGB_Code, Jaar_Datum' producing columns 'Bedrag_declaraties_Totaal'
General

Name * 30-Transform-Declaraties_R6

Description

Developer expierience – Databricks

- Tool: Browser
- Platform-as-a-Service: Azure Portal
- Just code
- Good collaboration
- Almost anything is possible
- Multilingual
- Future VS Code support
- Schema drift



Read Databricks switch action dataset

```python
from pyspark.sql.functions import expr
from pyspark.sql.functions import from_unixtime

events = spark.read \
  .option("inferSchema", "true") \
  .json("/databricks-datasets/structured-streaming/events/") \
  .withColumn("date", expr("time")) \
  .drop("time") \
  .withColumn("date", from_unixtime("date", 'yyyy-MM-dd'))

display(events)
```

```
(4) Spark Jobs
events:  pyspark.sql.dataframe.DataFrame = [action: string, date: string]
```

action	date
Close	2016-07-28
Close	2016-07-28
Open	2016-07-28
Close	2016-07-28
Open	2016-07-28
Open	2016-07-28
Close	2016-07-28
Close	2016-07-28
Close	2016-07-28

Showing the first 1000 rows.

```
Command took 30.21 seconds -- by dave.ruijter@macaw.nl at 10/3/2019, 8:20:48 PM on ETL Cluster #1
```

Write out DataFrame as Databricks Delta data

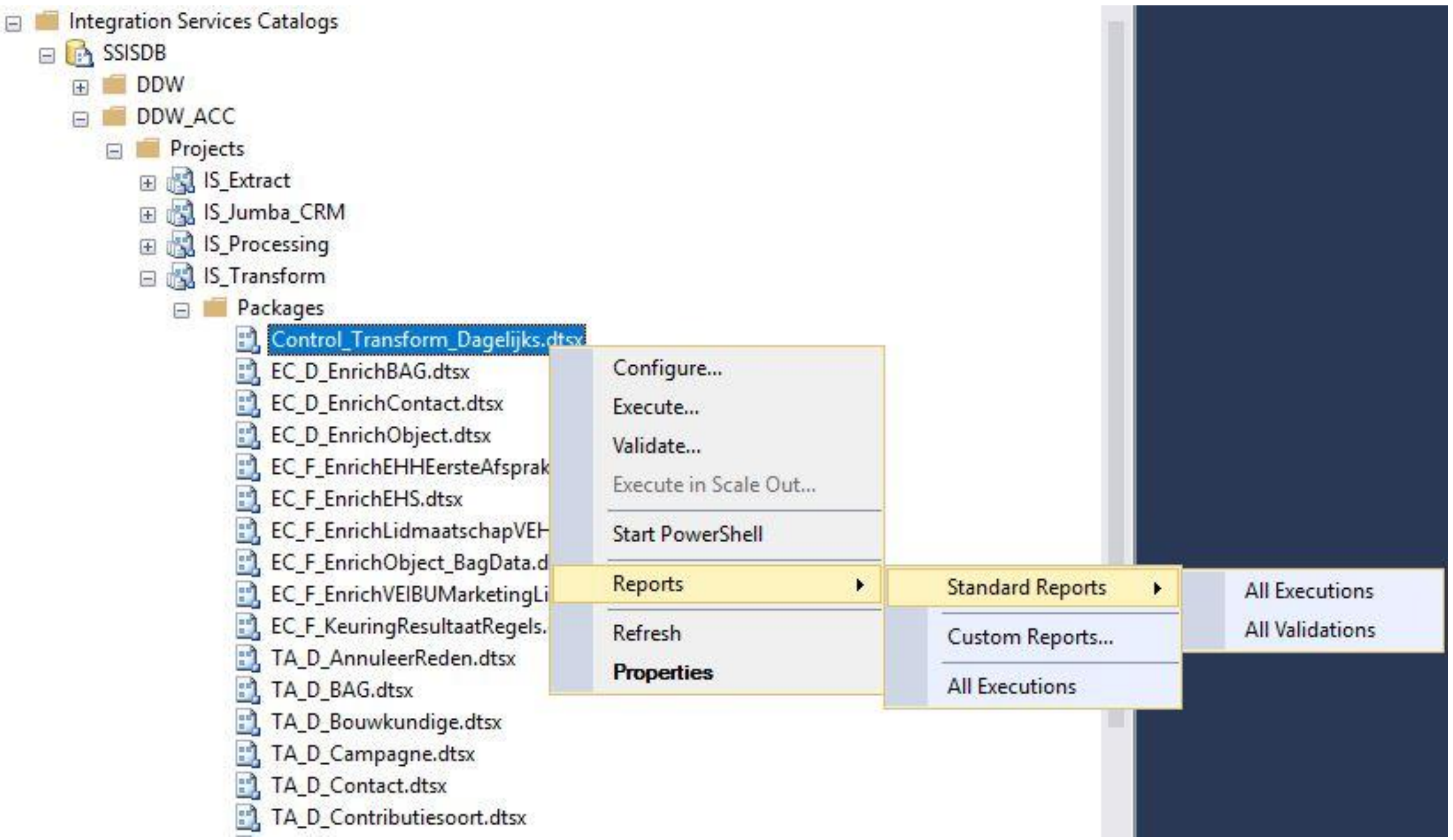
03

operator expierience

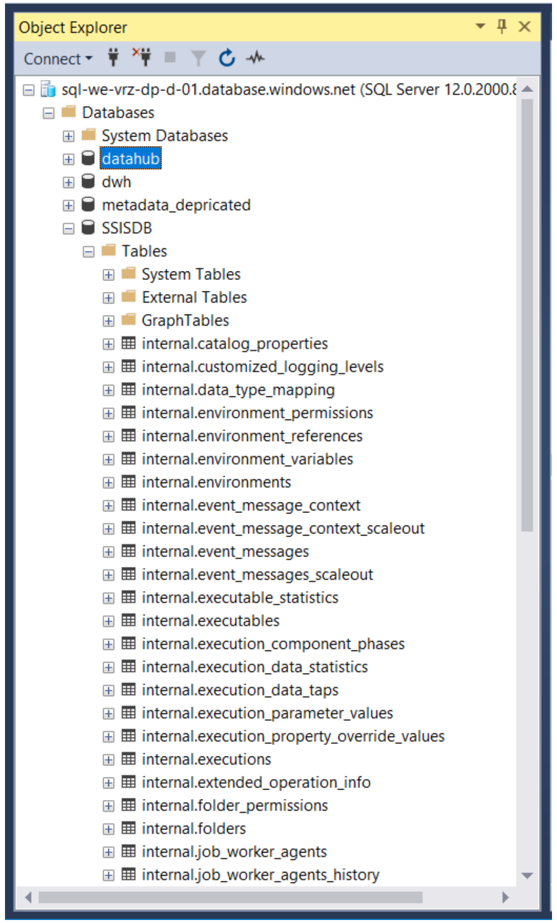
round.set(3)

Operator experience – SSIS in Azure

- SQL Server Management Studio (SSMS)



```
Integration Services Catalogs
  SSISDB
    DDW
    DDW_ACC
      Projects
        IS_Extract
        IS_Jumba_CRM
        IS_Processing
        IS_Transform
          Packages
            Control_Transform_Dagelijks.dtsx
            EC_D_EnrichBAG.dtsx
            EC_D_EnrichContact.dtsx
            EC_D_EnrichObject.dtsx
            EC_F_EnrichEHHEersteAfsprak
            EC_F_EnrichEHS.dtsx
            EC_F_EnrichLidmaatschapVEH
            EC_F_EnrichObject_BagData.d
            EC_F_EnrichVEIBUMarketingLi
            EC_F_KeuringResultaatRegels.
            TA_D_AnnuleerReden.dtsx
            TA_D_BAG.dtsx
            TA_D_Bouwkundige.dtsx
            TA_D_Campagne.dtsx
            TA_D_Contact.dtsx
            TA_D_Contributiesoort.dtsx
```



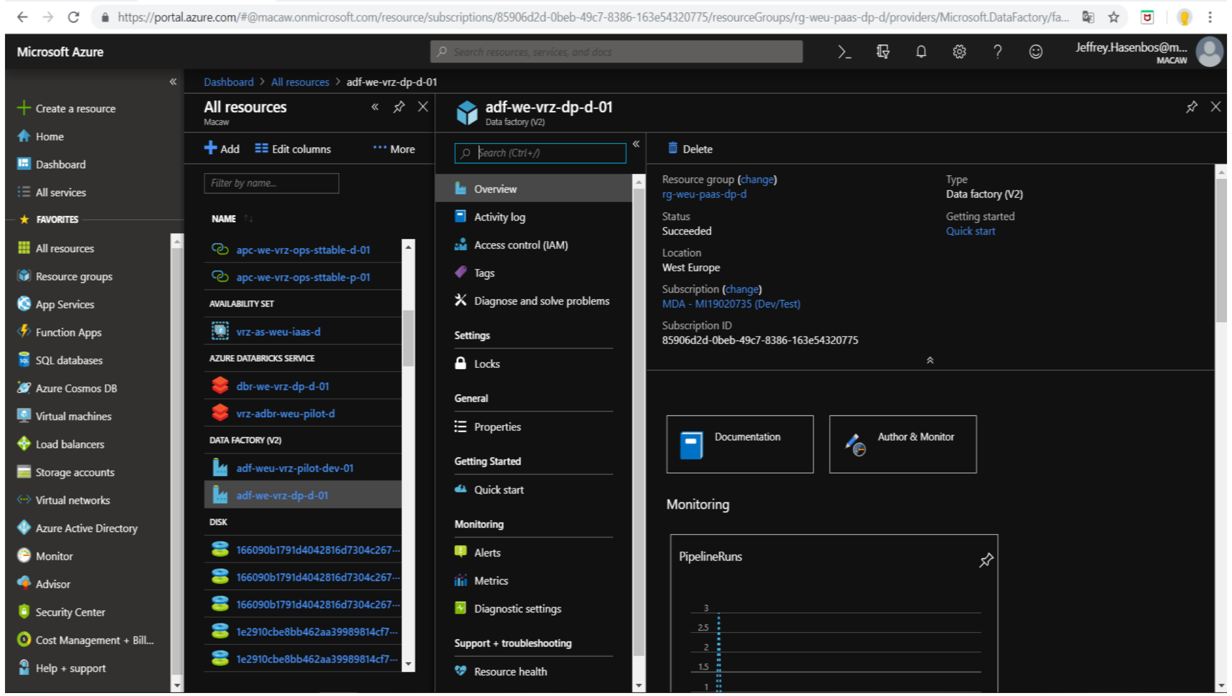
```
sql-we-vrz-dp-d-01.database.windows.net (SQL Server 12.0.2000.8
  Databases
    System Databases
    datahub
    dwh
    metadata_depricated
    SSISDB
      Tables
        System Tables
        External Tables
        GraphTables
        internal.catalog_properties
        internal.customized_logging_levels
        internal.data_type_mapping
        internal.environment_permissions
        internal.environment_references
        internal.environment_variables
        internal.environments
        internal.event_message_context
        internal.event_message_context_scaleout
        internal.event_messages
        internal.event_messages_scaleout
        internal.executable_statistics
        internal.executables
        internal.execution_component_phases
        internal.execution_data_statistics
        internal.execution_data_taps
        internal.execution_parameter_values
        internal.execution_property_override_values
        internal.executions
        internal.extended_operation_info
        internal.folder_permissions
        internal.folders
        internal.job_worker_agents
        internal.job_worker_agents_history
```

Operator experience – SSIS in Azure

- Deployment can be complicated
- Debugging / troubleshooting can be intimidating
- Limited integration of monitoring (ADF / Databricks)

Operator experience - MDF

- Internet Browser (Azure Portal)
- (Azure) PowerShell



Resource group (change)
rg-weu-paas-dp-d

Type
Data factory (V2)

Status
Succeeded

Getting started
Quick start

Location
West Europe

Subscription (change)
MDA - MI19020735 (Dev/Test)

Subscription ID
85906d2d-0beb-49c7-8386-163e54320775

Operator experience - Databricks

- Deployment complicated (clusters/notebooks)
- Internet Browser (Azure Portal)
- (Azure) PowerShell
- Debugging / troubleshooting can be intimidating
- Limited integration of monitoring (ADF / Databricks)

04

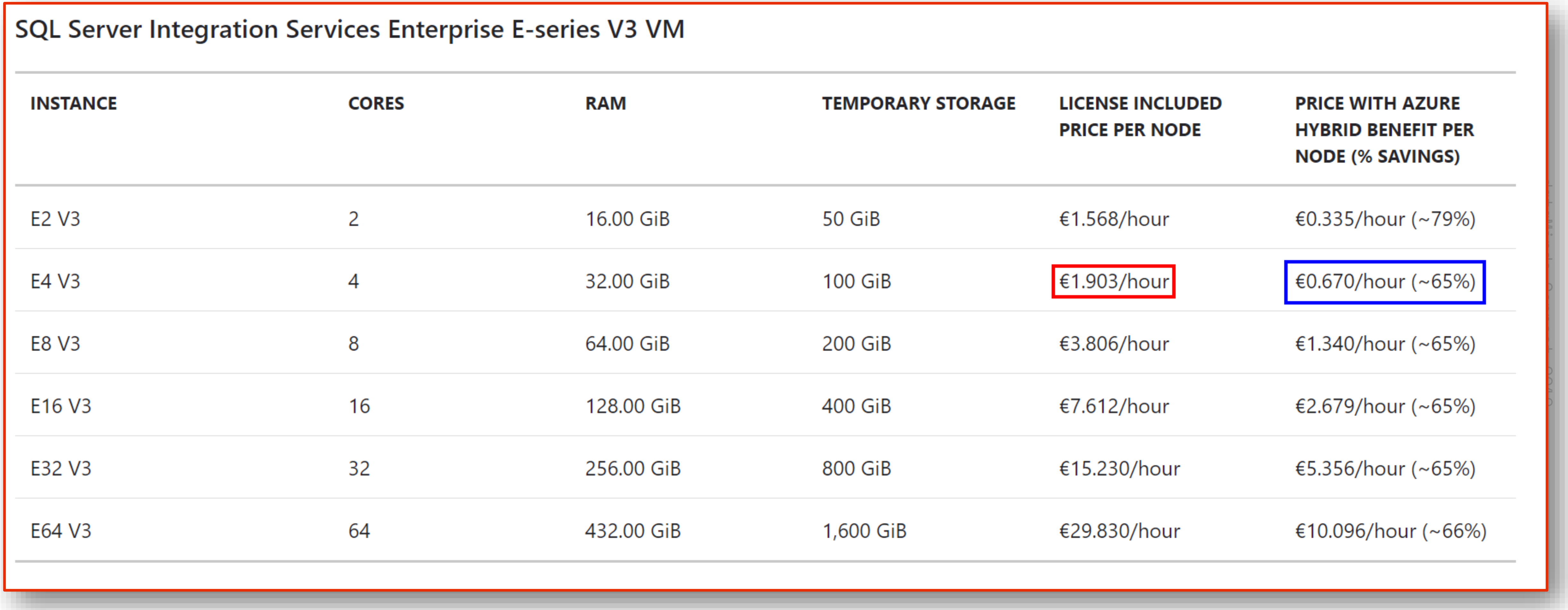
pricing

round.set(4)

Pricing – SSIS in Azure

- Integration Runtime costs (“it depends”)
- Benefit from existing SQL Server licensing

Pricing – SSIS in Azure



SQL Server Integration Services Enterprise E-series V3 VM

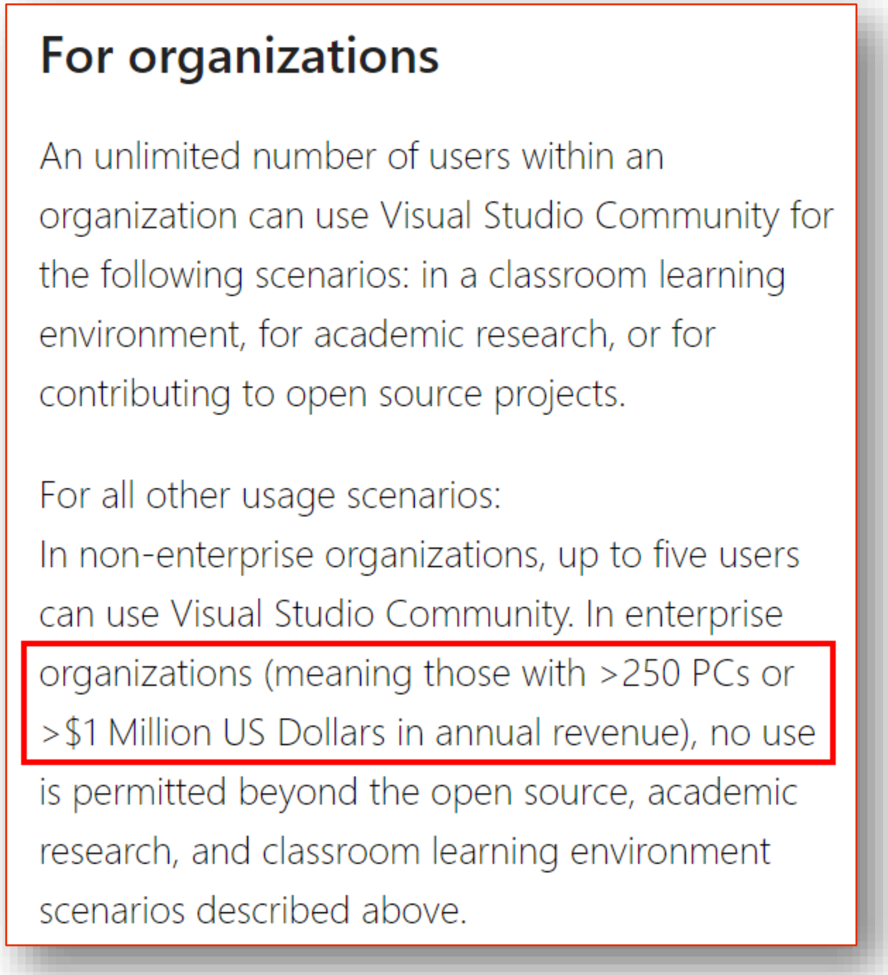
INSTANCE	CORES	RAM	TEMPORARY STORAGE	LICENSE INCLUDED PRICE PER NODE	PRICE WITH AZURE HYBRID BENEFIT PER NODE (% SAVINGS)
E2 V3	2	16.00 GiB	50 GiB	€1.568/hour	€0.335/hour (~79%)
E4 V3	4	32.00 GiB	100 GiB	€1.903/hour	€0.670/hour (~65%)
E8 V3	8	64.00 GiB	200 GiB	€3.806/hour	€1.340/hour (~65%)
E16 V3	16	128.00 GiB	400 GiB	€7.612/hour	€2.679/hour (~65%)
E32 V3	32	256.00 GiB	800 GiB	€15.230/hour	€5.356/hour (~65%)
E64 V3	64	432.00 GiB	1,600 GiB	€29.830/hour	€10.096/hour (~66%)

https://azure.microsoft.com/en-us/pricing/details/data-factory/ssis/

Pricing – SSIS in Azure

- License Visual Studio

(don’t forget about the DevOps server)



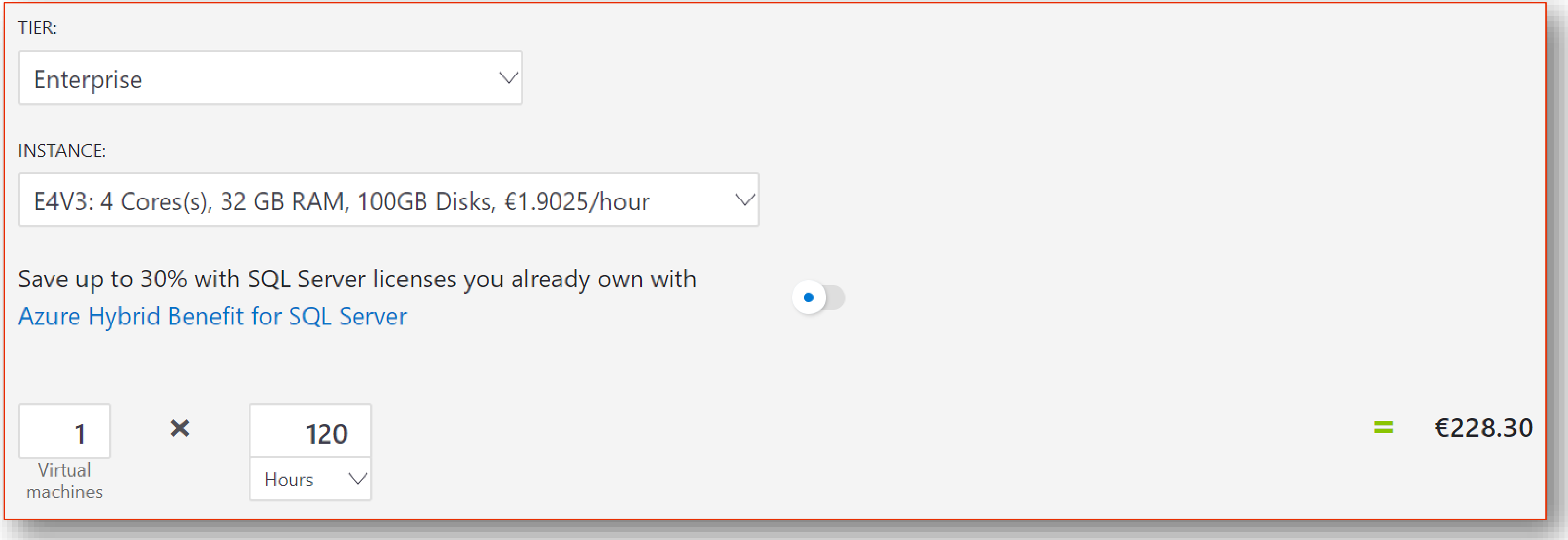
For organizations

An unlimited number of users within an organization can use Visual Studio Community for the following scenarios: in a classroom learning environment, for academic research, or for contributing to open source projects.

For all other usage scenarios:
In non-enterprise organizations, up to five users can use Visual Studio Community. In enterprise organizations (meaning those with >250 PCs or >$1 Million US Dollars in annual revenue), no use is permitted beyond the open source, academic research, and classroom learning environment scenarios described above.

Pricing – SSIS in Azure

- Development workload (3 days a week)



TIER:

Enterprise

INSTANCE:

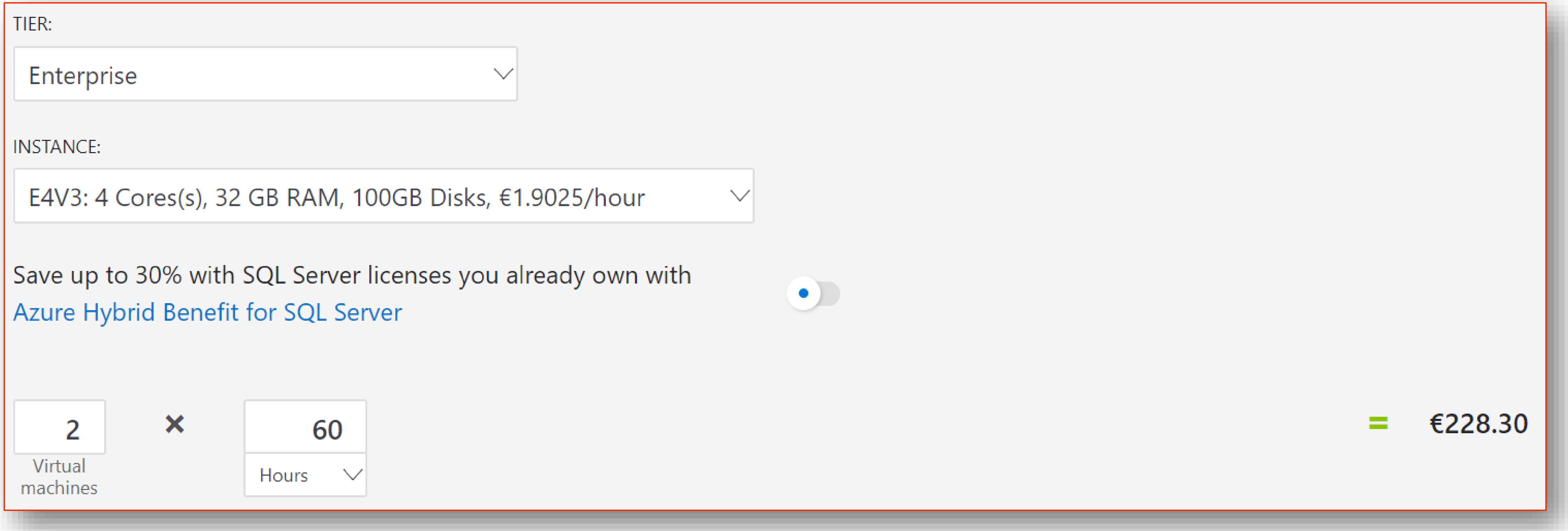
E4V3: 4 Cores(s), 32 GB RAM, 100GB Disks, €1.9025/hour

Save up to 30% with SQL Server licenses you already own with Azure Hybrid Benefit for SQL Server

1 Virtual machines × 120 Hours = €228.30

Pricing – SSIS in Azure

- Production workload (2h a day)



TIER:

Enterprise

INSTANCE:

E4V3: 4 Cores(s), 32 GB RAM, 100GB Disks, €1.9025/hour

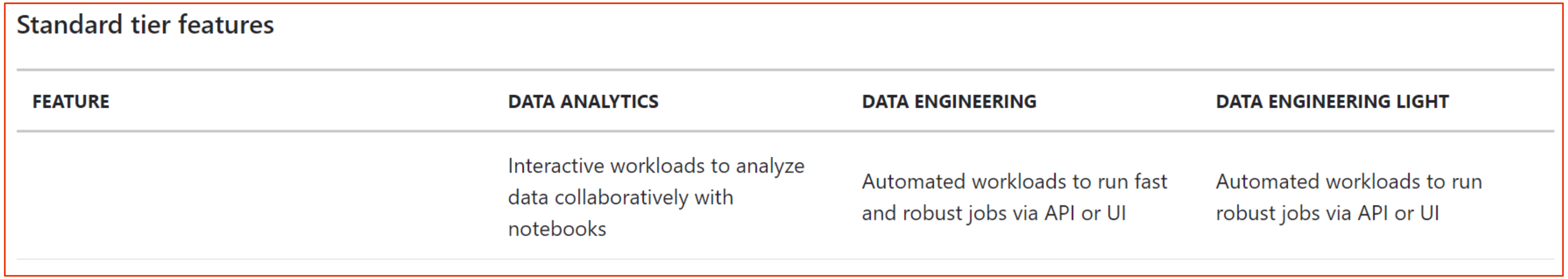
Save up to 30% with SQL Server licenses you already own with Azure Hybrid Benefit for SQL Server

2 Virtual machines × 60 Hours = €228.30

ADF MDF - Pricing

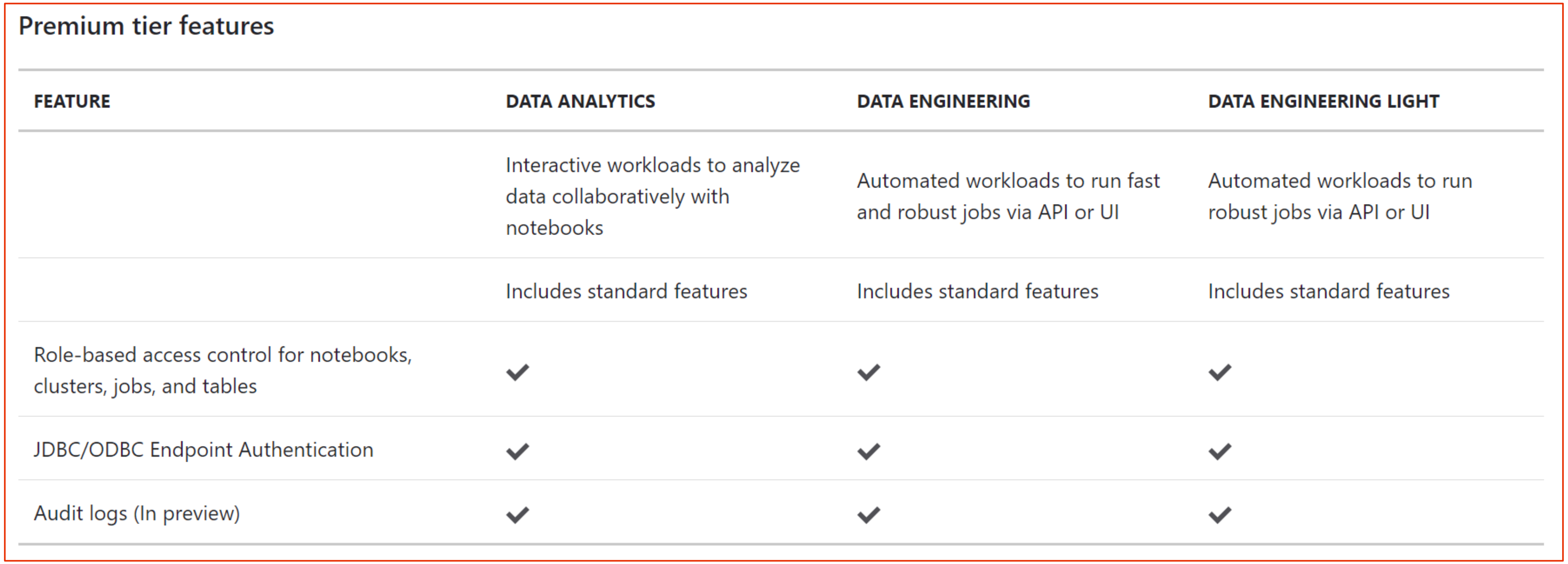
- Debug Mode:
  - “Preview Pricing”
  - 8 cores default
  - $0.112 / hour
  - 60 minutes default Time To Live (TTL)
  - Example dev-day: 10h. x 8 (cores) x $0.112 = **$8.96**
- Transform data in Blob Store (scheduled):
  - “Preview Pricing”
  - 8 cores default
  - $0.112 / hour
  - 10 minutes default Time To Live (TTL)
  - Example: 10m. compute + 10m. TTL = 0,33h. x 8 (cores) x $0.112 = **$0.299**

Azure Databricks - Pricing



Standard tier features

FEATURE	DATA ANALYTICS	DATA ENGINEERING	DATA ENGINEERING LIGHT
	Interactive workloads to analyze data collaboratively with notebooks	Automated workloads to run fast and robust jobs via API or UI	Automated workloads to run robust jobs via API or UI



Premium tier features

FEATURE	DATA ANALYTICS	DATA ENGINEERING	DATA ENGINEERING LIGHT
	Interactive workloads to analyze data collaboratively with notebooks	Automated workloads to run fast and robust jobs via API or UI	Automated workloads to run robust jobs via API or UI
	Includes standard features	Includes standard features	Includes standard features
Role-based access control for notebooks, clusters, jobs, and tables	✓	✓	✓
JDBC/ODBC Endpoint Authentication	✓	✓	✓
Audit logs (In preview)	✓	✓	✓

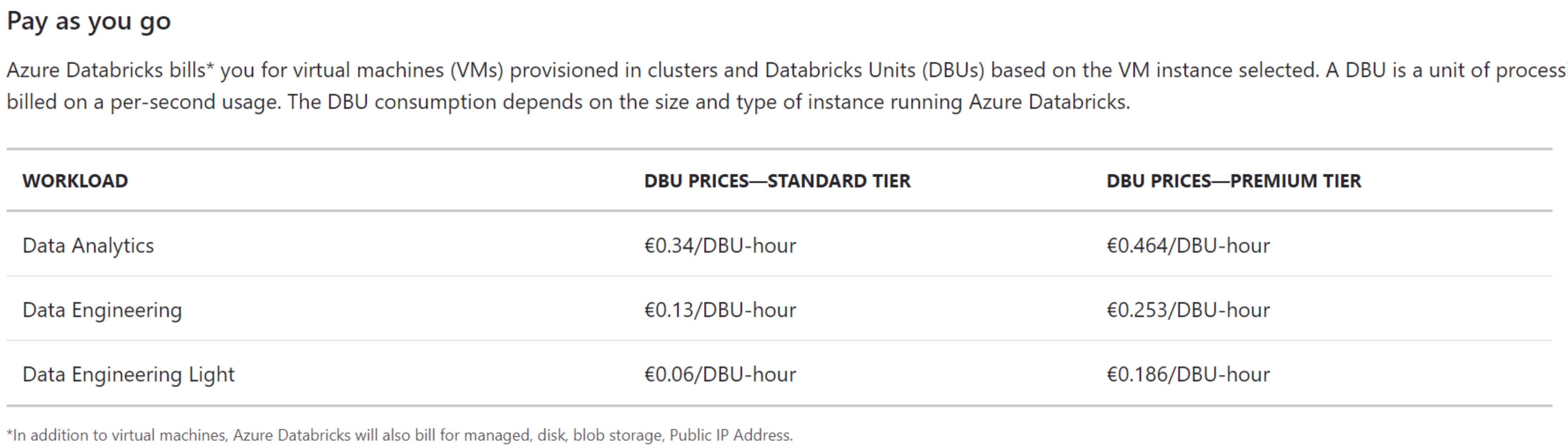
Azure Databricks - Pricing

- Data Analytics:
  - Interactive Clusters only here
  - Power BI connection to data in cluster
  - Notebook collaboration experience
- ‘Data Engineering Light’
  - Delta not available
  - Notebooks not available (also no scheduling of notebooks)

- Premium:
  - Role-based access control for notebooks, clusters, jobs, and tables
  - Audit Logs (preview)
  - JDBC/ODBC Endpoint Authentication

Azure Databricks - Pricing

- Development: Premium Tier - Data Analytics
- Production: Premium Tier – Data Engineering



Pay as you go

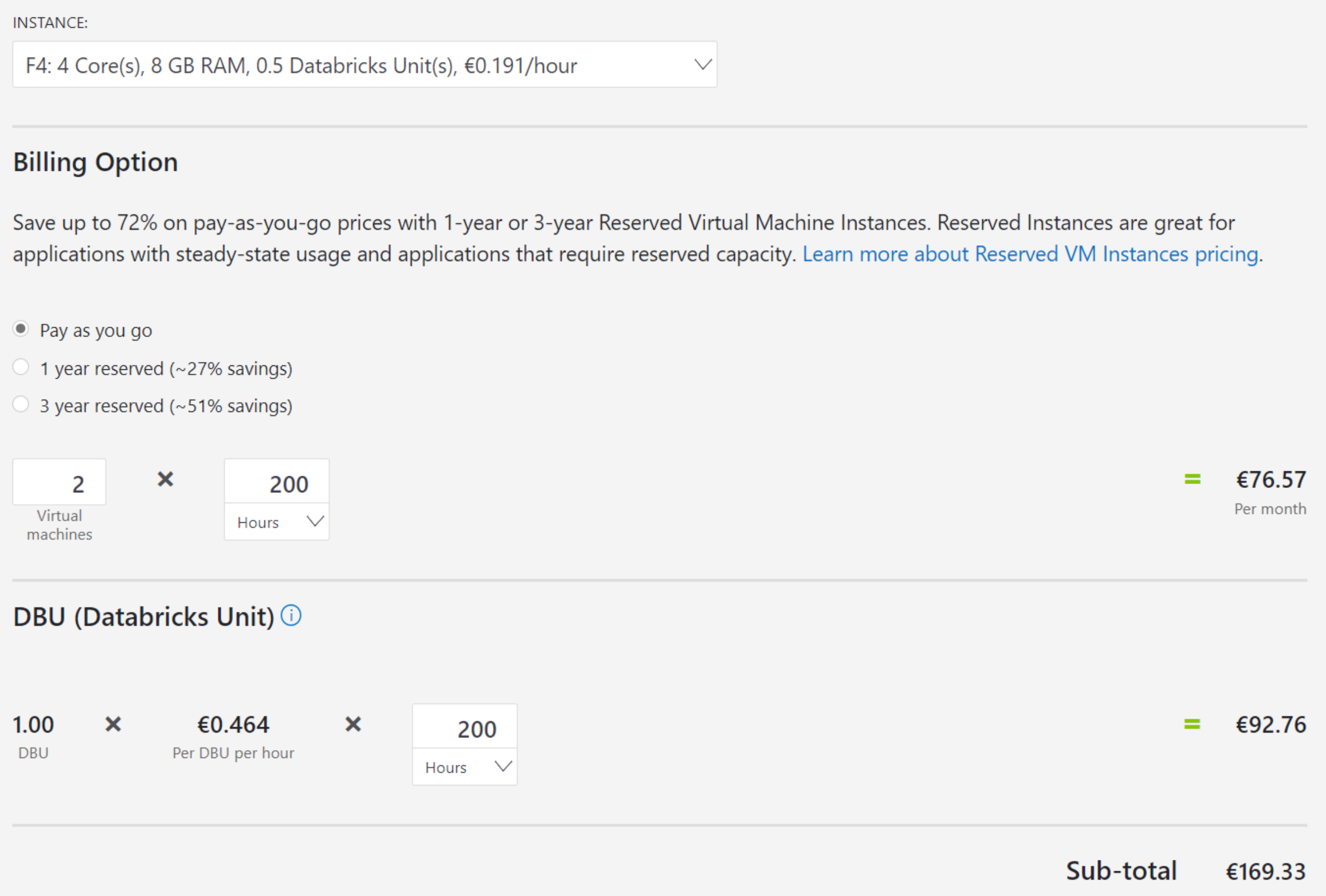
Azure Databricks bills* you for virtual machines (VMs) provisioned in clusters and Databricks Units (DBUs) based on the VM instance selected. A DBU is a unit of processing capability, billed on a per-second usage. The DBU consumption depends on the size and type of instance running Azure Databricks.

WORKLOAD	DBU PRICES—STANDARD TIER	DBU PRICES—PREMIUM TIER
Data Analytics	€0.34/DBU-hour	€0.464/DBU-hour
Data Engineering	€0.13/DBU-hour	€0.253/DBU-hour
Data Engineering Light	€0.06/DBU-hour	€0.186/DBU-hour

*In addition to virtual machines, Azure Databricks will also bill for managed, disk, blob storage, Public IP Address.

Azure Databricks - Pricing

- Development: Premium Tier - Data Analytics
- Workload: 5 days a week



INSTANCE:

F4: 4 Core(s), 8 GB RAM, 0.5 Databricks Unit(s), €0.191/hour

Billing Option

Save up to 72% on pay-as-you-go prices with 1-year or 3-year Reserved Virtual Machine Instances. Reserved Instances are great for applications with steady-state usage and applications that require reserved capacity. Learn more about Reserved VM Instances pricing.

● Pay as you go

○ 1 year reserved (~27% savings)

○ 3 year reserved (~51% savings)

2 Virtual machines × 200 Hours = €76.57 Per month

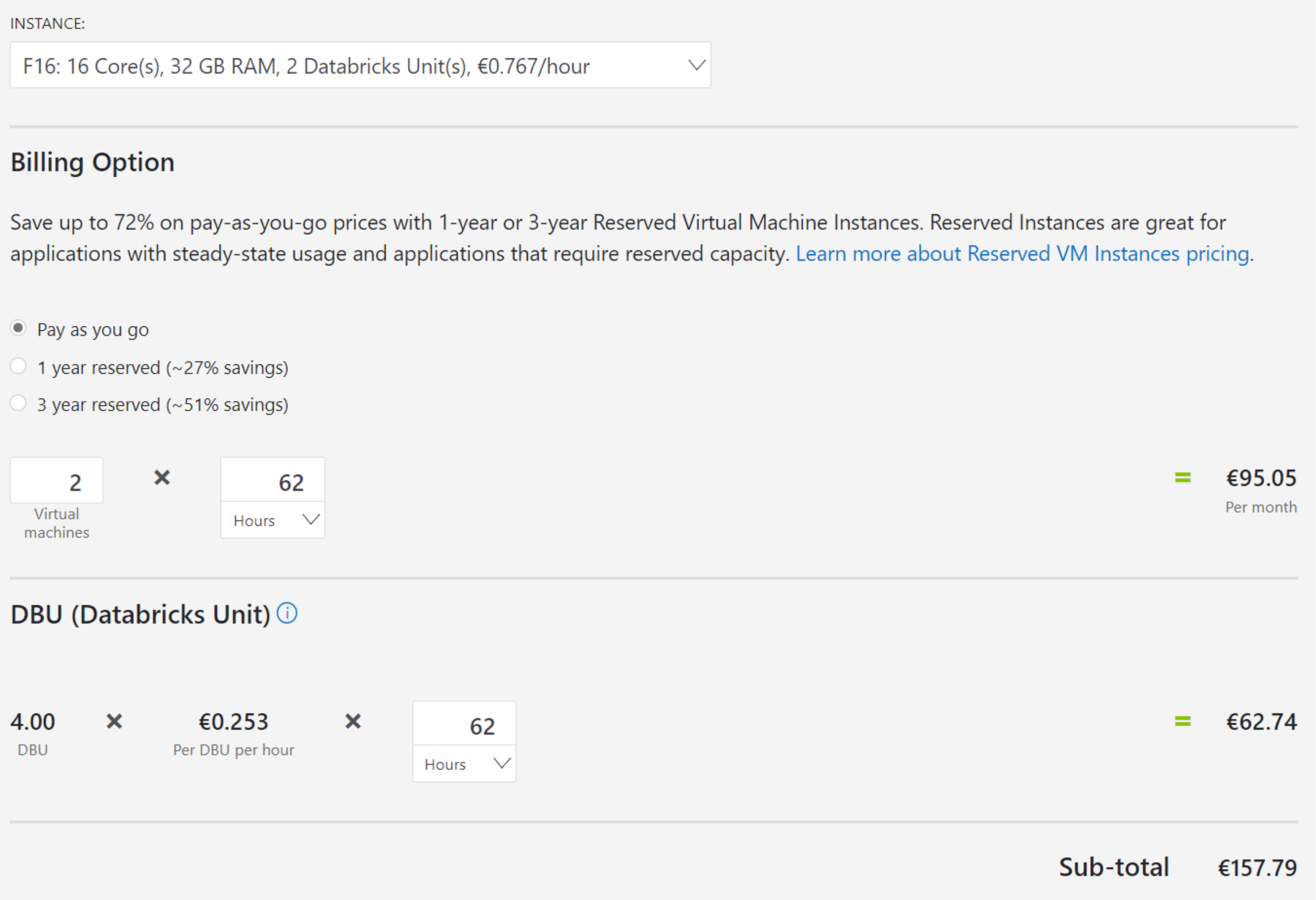
DBU (Databricks Unit)

1.00 DBU × €0.464 Per DBU per hour × 200 Hours = €92.76

Sub-total €169.33

Azure Databricks - Pricing

- Production : Premium Tier – Data Engineering
- Workload: 2 hours a day



INSTANCE:

F16: 16 Core(s), 32 GB RAM, 2 Databricks Unit(s), €0.767/hour

Billing Option

Save up to 72% on pay-as-you-go prices with 1-year or 3-year Reserved Virtual Machine Instances. Reserved Instances are great for applications with steady-state usage and applications that require reserved capacity. Learn more about Reserved VM Instances pricing.

● Pay as you go
○ 1 year reserved (~27% savings)
○ 3 year reserved (~51% savings)

2 Virtual machines × 62 Hours = €95.05 Per month

DBU (Databricks Unit)

4.00 DBU × €0.253 Per DBU per hour × 62 Hours = €62.74

Sub-total €157.79

10/8/2019

dataMinds Connect 2019

05

roadmap

round.set(5)

Roadmap – SSIS in Azure

- ?

Roadmap – ADF MDF

- Active Monitoring (watch progress live)

Roadmap – Databricks

- C# as notebook language
- Integration with Visual Studio Code

06

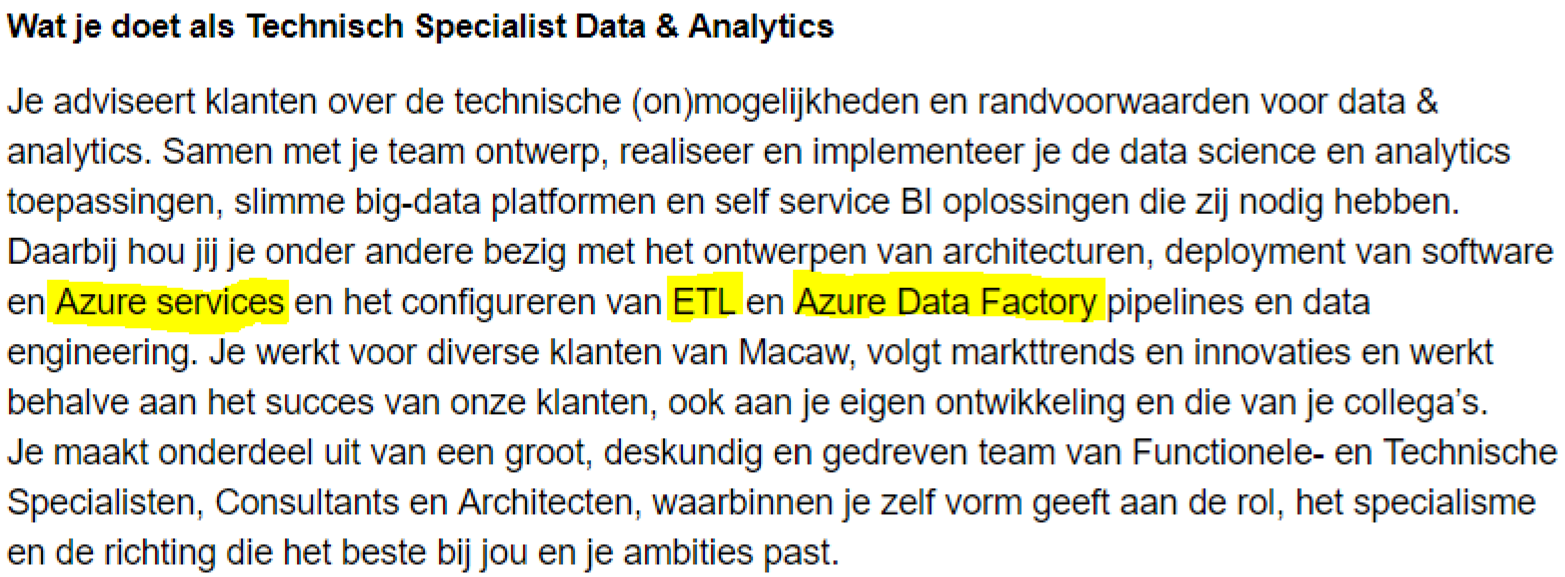
coolness

round.set(6)

Recruiting

- People are increasingly looking for new tooling in job offers:
    - Azure
    - Azure Databricks
    - Azure Data Factory
    - Data Lake
    - Datawarehouse
    - DevOps

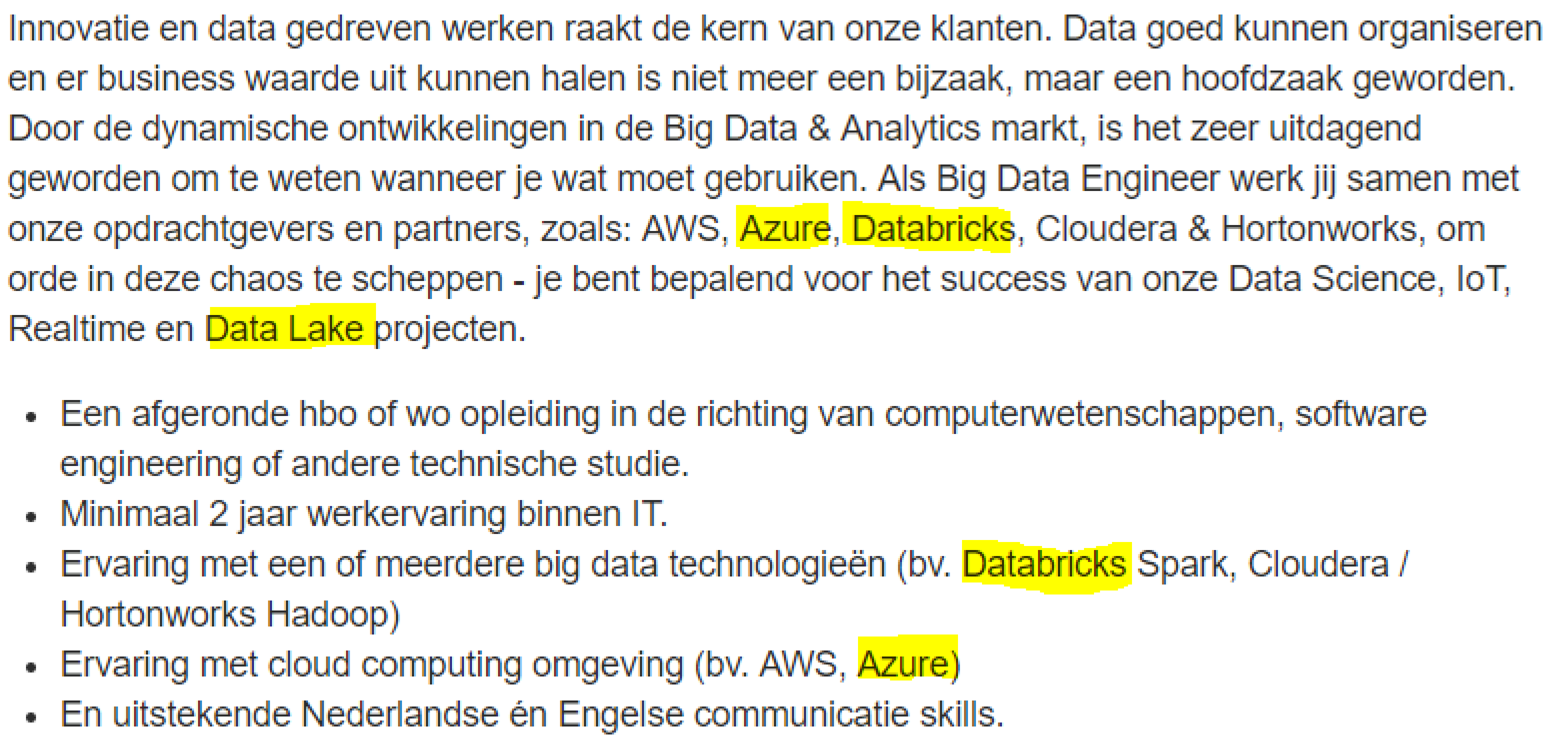
Job offers - examples



Wat je doet als Technisch Specialist Data & Analytics

Je adviseert klanten over de technische (on)mogelijkheden en randvoorwaarden voor data & analytics. Samen met je team ontwerp, realiseer en implementeer je de data science en analytics toepassingen, slimme big-data platformen en self service BI oplossingen die zij nodig hebben. Daarbij hou jij je onder andere bezig met het ontwerpen van architecturen, deployment van software en Azure services en het configureren van ETL en Azure Data Factory pipelines en data engineering. Je werkt voor diverse klanten van Macaw, volgt markttrends en innovaties en werkt behalve aan het succes van onze klanten, ook aan je eigen ontwikkeling en die van je collega's.
Je maakt onderdeel uit van een groot, deskundig en gedreven team van Functionele- en Technische Specialisten, Consultants en Architecten, waarbinnen je zelf vorm geeft aan de rol, het specialisme en de richting die het beste bij jou en je ambities past.

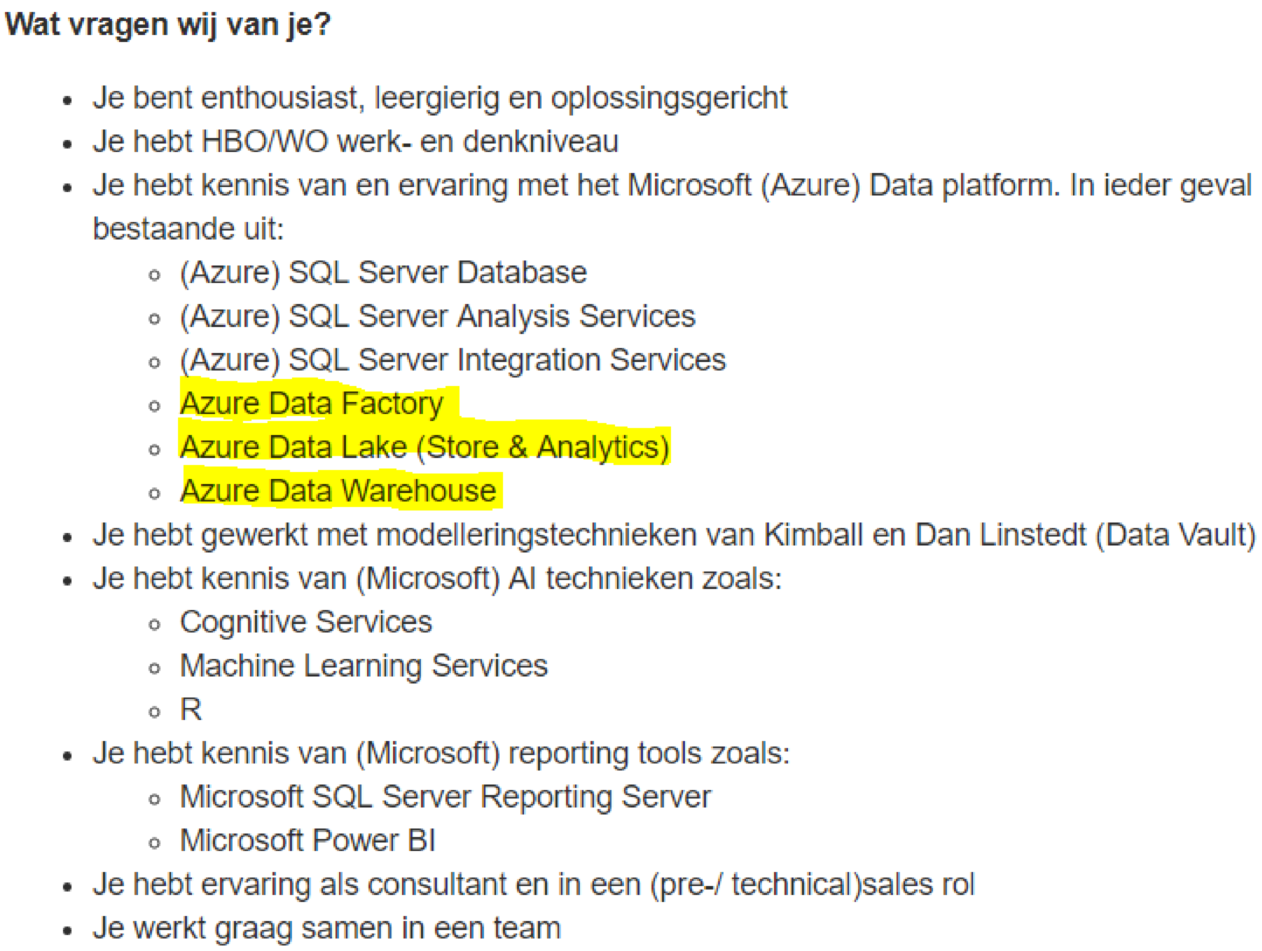
Job offers - examples



Innovatie en data gedreven werken raakt de kern van onze klanten. Data goed kunnen organiseren en er business waarde uit kunnen halen is niet meer een bijzaak, maar een hoofdzaak geworden. Door de dynamische ontwikkelingen in de Big Data & Analytics markt, is het zeer uitdagend geworden om te weten wanneer je wat moet gebruiken. Als Big Data Engineer werk jij samen met onze opdrachtgevers en partners, zoals: AWS, Azure, Databricks, Cloudera & Hortonworks, om orde in deze chaos te scheppen - je bent bepalend voor het success van onze Data Science, IoT, Realtime en Data Lake projecten.

- Een afgeronde hbo of wo opleiding in de richting van computerwetenschappen, software engineering of andere technische studie.
- Minimaal 2 jaar werkervaring binnen IT.
- Ervaring met een of meerdere big data technologieën (bv. Databricks Spark, Cloudera / Hortonworks Hadoop)
- Ervaring met cloud computing omgeving (bv. AWS, Azure)
- En uitstekende Nederlandse én Engelse communicatie skills.

Job offers - examples



Wat vragen wij van je?

- Je bent enthousiast, leergierig en oplossingsgericht
- Je hebt HBO/WO werk- en denkniveau
- Je hebt kennis van en ervaring met het Microsoft (Azure) Data platform. In ieder geval bestaande uit:
    - (Azure) SQL Server Database
    - (Azure) SQL Server Analysis Services
    - (Azure) SQL Server Integration Services
    - Azure Data Factory
    - Azure Data Lake (Store & Analytics)
    - Azure Data Warehouse
- Je hebt gewerkt met modelleringstechnieken van Kimball en Dan Linstedt (Data Vault)
- Je hebt kennis van (Microsoft) AI technieken zoals:
    - Cognitive Services
    - Machine Learning Services
    - R
- Je hebt kennis van (Microsoft) reporting tools zoals:
    - Microsoft SQL Server Reporting Server
    - Microsoft Power BI
- Je hebt ervaring als consultant en in een (pre-/ technical)sales rol
- Je werkt graag samen in een team

The battle!

- round #1: capabilities
- round #2: developer experience
- round #3: operator experience
- round #4: security
- round #5: roadmap / future readiness
- round #6: coolness

Thank You

What do you think?

1. Open the form
2. Provide constructive feedback
3. Be eligible for an amazing prize!

http://bit.ly/dataMindsConnectSession

bit.ly is CASE SENSITIVE!

99

Q&A

Our Partners

KOHERA
element 61
experience & expertise
Microsoft
REALDOLMEN
a Gfi Group company
Sparkle
cegeka
IN CLOSE COOPERATION
SentryOne
LOOP1
solarwinds
U2U
ORDINA