# Notebooks 101 for SQL People

Database Professionals Virtual Meetup Group – July 28, 2021



Program Manager, Microsoft jukoesma@microsoft.com



#### **Barbara Valdez (@barbaracodes)**

Software Engineer, Microsoft bavaldez@microsoft.com





# Our learning journey today

Notebooks, eh?

Why Notebooks?

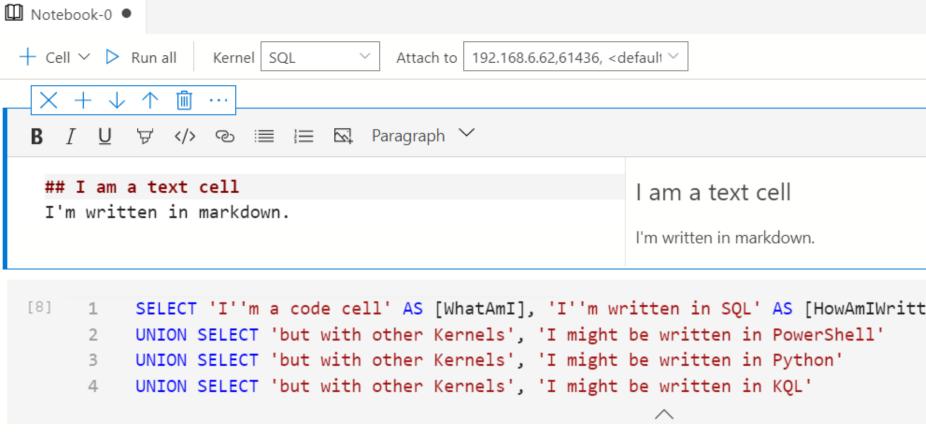
**Getting started with Notebooks** 



## **Notebooks**

# Text cells Code cells

- SQL
- PowerShell
- Python
- KQL
- .NET interactive

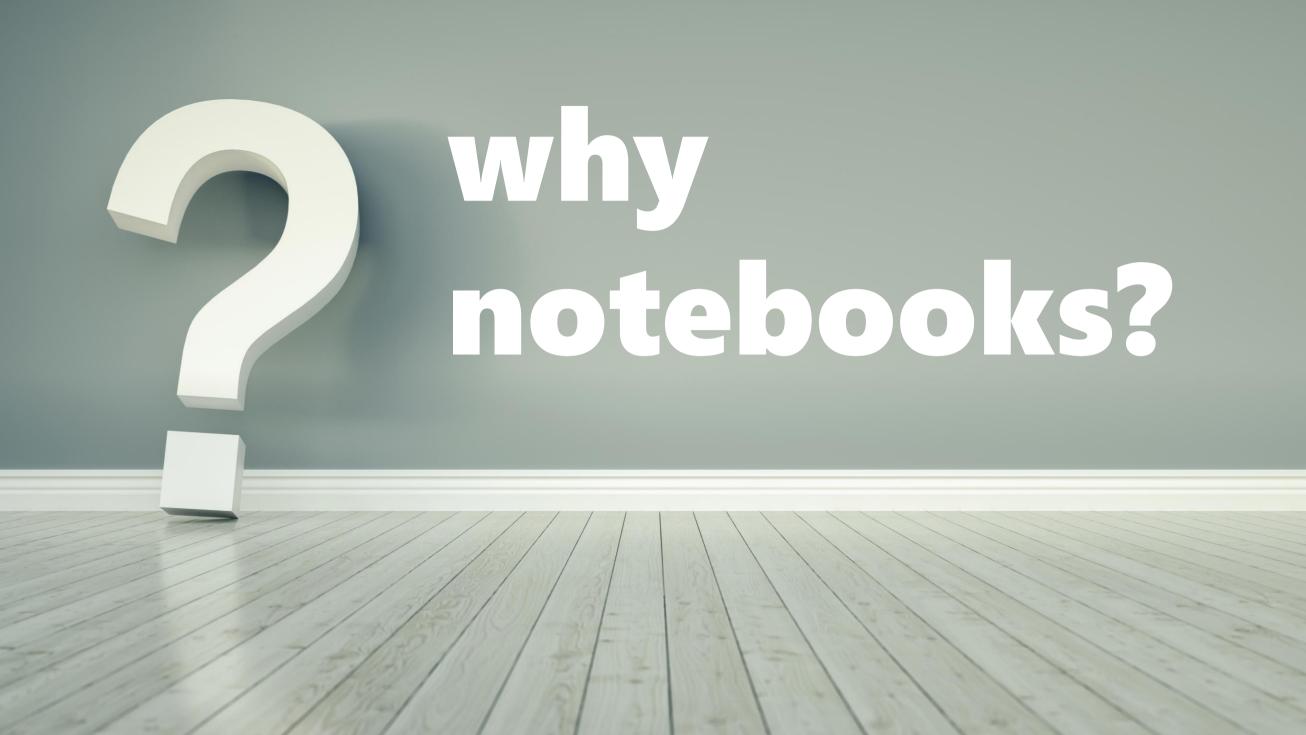


(4 rows affected)

Total execution time: 00:00:00.013

#### 

	WhatAmI	HowAmIWritten
1	but with other Kernels	I might be written in KQL
2	but with other Kernels	I might be written in PowerS
3	but with other Kernels	I might be written in Python
4	I'm a code cell	I'm written in SQL



I need to know which product has the fastest turnaround ..

Let me query that!

Here you go...  $\left[ egin{pmatrix} \egin{pmatrix} egin{pmatrix} egin{pmatrix} egin{pmatrix} egin{pmatrix} \egin{pmatrix} \$ 

*Oh.*. I need the product category..

Sure..

Here you go again...  $[ ]_{J}$ 

Thanks! This is for the last quarter, right?

Do you mean sold products?

```
-- Prepared by Cookie Monster
 -- 2020.03.01
 -- Question from Marketing about how much we
 sell and what's the shorter manufacturing time.
SELECT TOP 10 *
 FROM Production Product
ORDER BY Name ASC:
 == Get Product list withprostctanutegopyewdaysfast
 mahaf€acture days
 USP Naventer Ewodks 2020 Pe
 GOp. Product Number
                                                                                                                            X 1
 SEPECISTPRICE AS Price
        p. NaysToMapudaGtuCategory
  FROMN Binedustion of the From Binedustion o
 WHERE rod Day Number ufacture < 4
 GOp ListPrice AS Price
         p.DaysToManufacture
 FROM Production. Product p
 INNER JOIN Production. ProductSubcategory s
 ON s.ProductSubcategoryID = p.ProductSubcategoryID
 INNER JOIN Production.ProductCategory c
 ON c.ProductCategoryID = s.ProductCategoryID
 WHERE p.DaysToManufacture < 4
 GO
```

I need to know which product has the fastest turnaround ..

Let me query that!

Here you go... 🗓

*Oh.*. I need the product category..

Sure..

Here you go again...  $[ \bigcup_j$ 

Thanks! This is for the last quarter, right?

Do you mean sold products?

Here you go again...  $[l]_{J}$ 

You are a Rockstar! Thanks.

```
-- Get Product list and product category with
fast manufacture days
USE AdventureWorks2017;
GO
SELECT
  c.Name AS ProductCategory
  p.Name as ProductName
  p.ProductNumber
  p.ListPrice AS Price
  p.DaysToManufacture
FROM Production. Product p
INNER JOIN Production. ProductSubcategory s
ON s.ProductSubcategoryID =
p.ProductSubcategoryID
INNER JOIN Production.ProductCategory c
ON c.ProductCategoryID = s.ProductCategoryID
WHERE p.DaysToManufacture < 4
AND EXISTS (SELECT TOP 1 1
FROM Sales Sales Order Detail sod
INNER JOIN Sales.SalesOrderHeader soh
ON soh.SalesOrderID = sod.SalesOrderID
WHERE sod.ProductID = p.ProductID
AND soh.OrderDate BETWEEN '20140101' AND
'20140331')
G<sub>0</sub>
```

```
-- Get Product list and product category with fast manufacture days
                                                                                               USE AdventureWorks2017;
 -- Prepared by Cookie Monster
                                                                                               GO
                                                                                              ⊟SELECT
 -- Question from Marketing about how much we sell and what's the shorter manufacturing time.
                                                                                                      c.Name AS ProductCategory
                                                                                                    , p.Name as ProductName
SELECT TOP 10 *
                                                                                                    . p.ProductNumber
 FROM Production.Product
                                                                                                    . p.ListPrice AS Price
 ORDER BY Name ASC:
                                                                                                    , p.DaysToManufacture
                                                                                                FROM Production. Product p
 -- Get Product List with fast manufacture days.
                                                                                                    INNER JOIN Production.ProductSubcategory s
ON s.ProductSubcategoryID = p.ProductSubcategoryID
      p.Name as ProductName
                                                                                                    INNER JOIN Production.ProductCategory c
   , p.ProductNumber
                                                                                                        ON c.ProductCategoryID = s.ProductCategoryID
   , p.ListPrice AS Price
                                                                                                WHERE p.DaysToManufacture < 4
    , p.DaysToManufacture
                                                                                                AND EXISTS (SELECT TOP 1 1
 FROM Production.Product p
 WHERE p.DaysToManufacture < 4
                                                                                                                 FROM Sales.SalesOrderDetail sod
                                                                                                                     INNER JOIN Sales.SalesOrderHeader soh
                                                                                                                         ON soh.SalesOrderID = sod.SalesOrderID
 -- Get Product list and product category with fast manufacture days
                                                                                                                 WHERE sod.ProductID = p.ProductID
 USE AdventureWorks2017:
                                                                                                                 AND soh.OrderDate BETWEEN '20140101' AND '20140331')
= SELECT
      c.Name AS ProductCategory
     , p.Name as ProductName
                                                                                               -- Returns total sales and the discounts for each product. In the second example, the total revenue is calculated for each product.
     . p.ProductNumber
                                                                                               USE AdventureWorks2017;
    , p.ListPrice AS Price
    , p.DaysToManufacture
 FROM Production.Product p
                                                                                              SELECT p.Name AS ProductName
     INNER JOIN Production.ProductSubcategory s
                                                                                               NonDiscountSales = SUM((OrderOty * UnitPrice)),
        ON s.ProductSubcategoryID = p.ProductSubcategoryID
                                                                                               Discounts = SUM(((OrderOty * UnitPrice) * UnitPriceDiscount))
     INNER JOIN Production.ProductCategory c
                                                                                                FROM Production. Product AS p
        ON c.ProductCategorvID = s.ProductCategorvID
                                                                                                INNER JOIN Sales.SalesOrderDetail AS sod
 WHERE p.DaysToManufacture < 4
                                                                                               ON p.ProductID = sod.ProductID
                                                                                                GROUP BY p.Name
 -- Get Product list and product category with fast manufacture days
                                                                                               ORDER BY ProductName DESC:
 USE AdventureWorks2017:

□ SELECT

                                                                                               -- This is the query that calculates the revenue for each product in each sales order.
      c.Name AS ProductCategory
    , p.Name as ProductName
                                                                                               USE AdventureWorks2017:
    , p.ProductNumber
    , p.ListPrice AS Price

∃SELECT 'Total income is', SUM(((OrderOty * UnitPrice) * (1.0 - UnitPriceDiscount))), ' for ',

     , p.DaysToManufacture
                                                                                               p.Name AS ProductName
 FROM Production.Product p
                                                                                               FROM Production. Product AS p
    INNER JOIN Production.ProductSubcategory s
                                                                                               INNER JOIN Sales.SalesOrderDetail AS sod
        ON s.ProductSubcategoryID = p.ProductSubcategoryID
     INNER JOIN Production.ProductCategory c
                                                                                               ON p.ProductID = sod.ProductID
        ON c.ProductCategoryID = s.ProductCategoryID
                                                                                               GROUP BY p.Name
 WHERE p.DaysToManufacture < 4
                                                                                               ORDER BY ProductName ASC;
 -- Get Product list and product category with fast manufacture days
 USE AdventureWorks2017:
⊟SELECT
      c.Name AS ProductCategory
     , p.Name as ProductName
     , p.ProductNumber
    , p.ListPrice AS Price
```

, p.DaysToManufacture FROM Production.Product p

INNER JOIN Production.ProductSubcategory s

INNER JOIN Production.ProductCategory c

ON s.ProductSubcategoryID = p.ProductSubcategoryID

## My dirty little scripts in the closet ....



You know that list that you sent me last time about that fast manufacturer... Can refresh it, please?

What list?

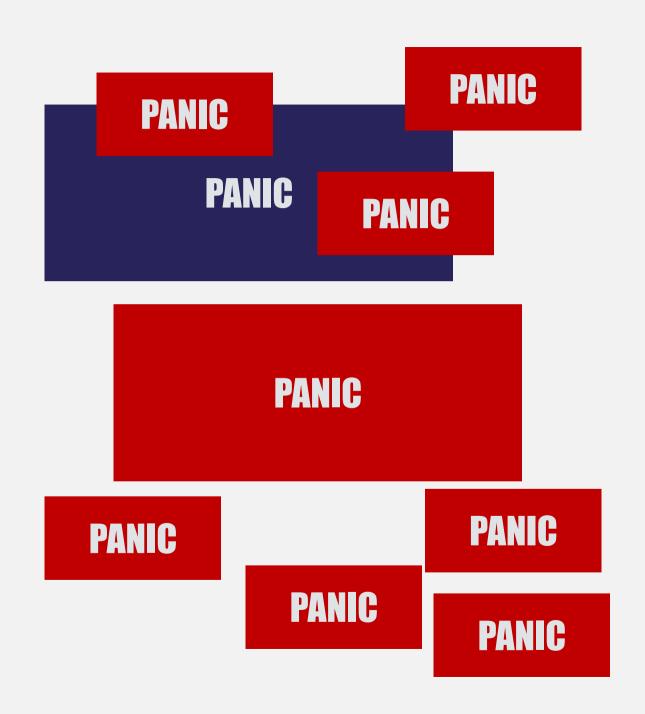
Oh.. That list! Yea yea...

Here you go again... 🗓

I want to compare what the result back then and what it is now; if we have had any changes...

Was the logic the same as what you did last time?

Hello?







When scripts and output are separated ...

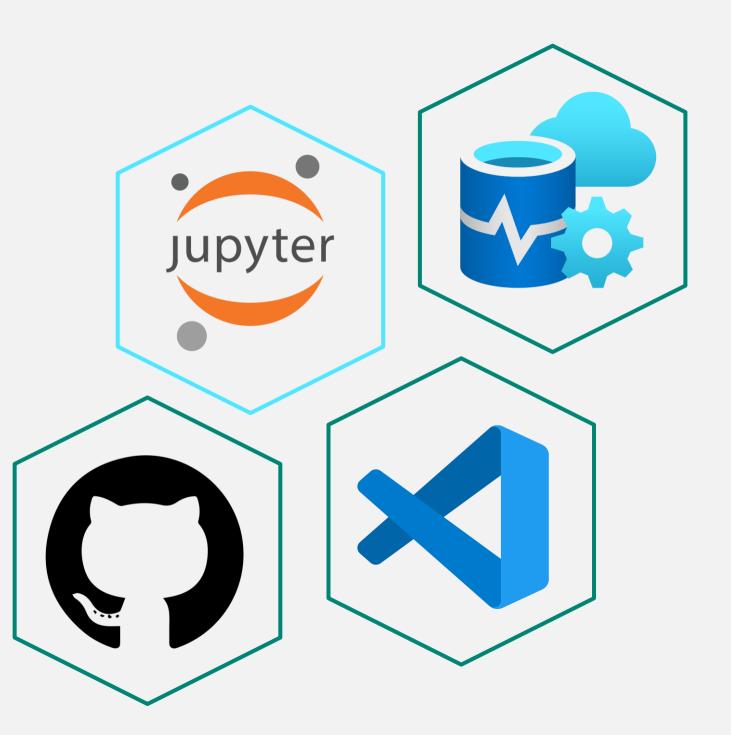
- 1. Hard to reproduce based on output (reproducibility)
- 2. Code looks good, but is it returning the right results? (code review)
- 3. I executed this script so I guessed it worked last time? (auditing)
- 4. What is this script about? (context)
- 5. For Troubleshooting guide, lots of copying and pasting code (executable)



the scientific method			
Question	research/observation		
hypothesis			
experiment			
res	Procedure  1. 2. 3. 4. 5.		
conclusions			
new questions/notes for next time			
© FEELSLIKEHOMEBLOG.COM			

# Remember our School Science Project?





### Jupyter Notebook

Jupyter Notebook open source format

Jupyter Notebook dev tools – examples:

- Azure Data Studio
- VSCode
- Jupyter Lab

Jupyter Notebook viewers – examples:

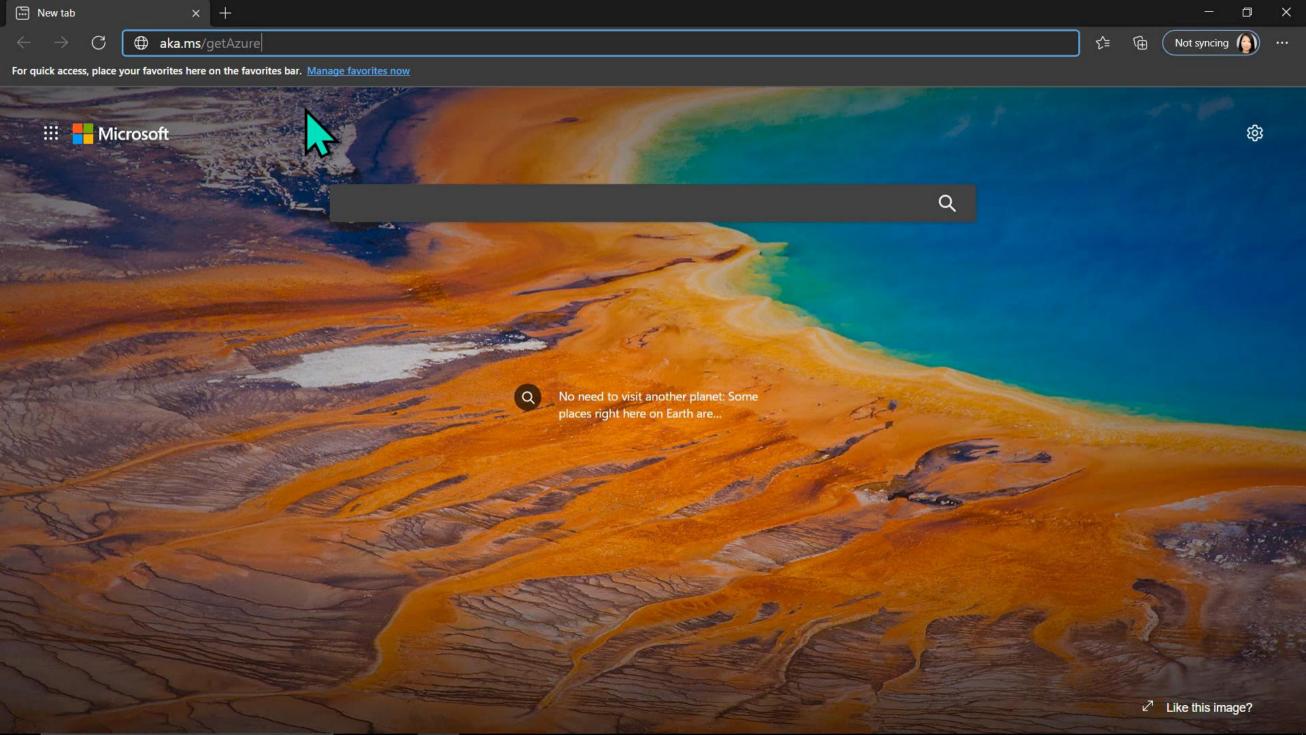
- GitHub native notebook viewer
- Azure DevOps Jupyter Preview

### **Azure Data Studio**

#### **Modern Data Experiences**

- Open Source
- Cross Platform client tool (Windows, MacOS, Linux)
- Support Git
- Rich data-oriented kernels:
   SQL, KQL, PowerShell
- Extensible framework









### **Notebook Development and Sharing**

- Write Notebooks in Azure Data Studio
  - Troubleshooting Guide template
- Bundle and organize your notebooks as Jupyter Books
- · Use source control e.g. GitHub, Azure DevOps
- Ship it!
  - For external facing use GitHub release features, example: <a href="https://github.com/MsSQLGirl/jubilant-data-wizards/releases/tag/v0.0.0">https://github.com/MsSQLGirl/jubilant-data-wizards/releases/tag/v0.0.0</a>
  - · Ship as an Azure Data Studio extension to launch a Jupyter Book

# Jump start with your existing scripts

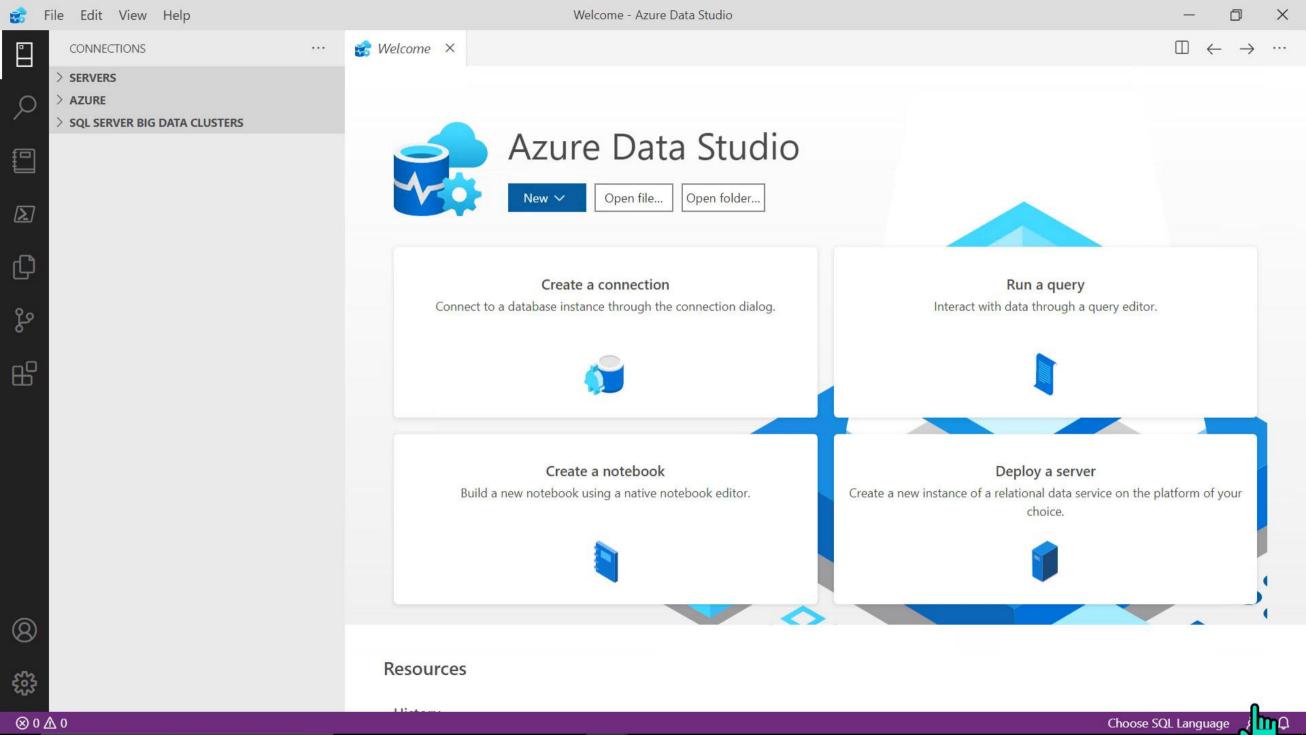


- SQL scripts to Notebook:
   ConvertTo-SqlNotebook
- Powershell scripts to Notebook:
   ConvertTo-PowerShellNoteBook

https://github.com/dfinke/PowerShellNotebook

# new deployment with notebooks









Notebook as UX

Deployment Troubleshooting

## Notebook Use Cases



Notebooks as Documentation

Documentation

Education, Blogging

Feature Discovery



Notebooks for Collaboration

Collaboration

Reporting

Data Viz



Notebooks for Data Workflows

**ML Notebooks** 

Data Science

Data Engineering



**Operational Notebooks** 

Execute as Job, Testing

Command Line Execution

Incident Response Workflow



#### Useful resources

http://aka.ms/AzureDataStudio | http://aka.ms/getAzureDataStudio | http://aka.ms/AzureDataStudioBlog

Tiger Toolbox:

SQL Assessment Toolkit: <a href="https://github.com/microsoft/sql-server-samples/tree/master/samples/manage/sql-assessment-api/notebooks">https://github.com/microsoft/sql-server-samples/tree/master/samples/manage/sql-assessment-api/notebooks</a>

SQL Hybrid Toolkit: <a href="https://github.com/microsoft/tigertoolbox/tree/master/SQL-Hybrid-Cloud-">https://github.com/microsoft/tigertoolbox/tree/master/SQL-Hybrid-Cloud-</a>

Toolkit/content

Glenn Berry's Diagnostic Notebooks for SQL Server and Azure SQL DB:

https://glennsqlperformance.com/resources/

Rob Sewell's Notebook <a href="http://sqldbawithabeard.com">http://sqldbawithabeard.com</a><a href="https://github.com/SQLDBAWithABeard/JupyterNotebooks">https://github.com/SQLDBAWithABeard/JupyterNotebooks</a>

The SQL Diagnostic (Jupyter) Book by Emanuele Meazzo <a href="https://tsql.tech/the-sql-diagnostic-jupyter-book/">https://tsql.tech/the-sql-diagnostic-jupyter-book/</a>

## Key take-aways

#### Notebooks for **SQL** people use cases

- Reproducible research, aka. Data analysis and data exploration.
- Troubleshooting guides for incident response
- Tutorial / How-to's
- Data Engineering
- Deployment UI\*
- Feature Discovery UI\*

#### **Language (Kernel) supported in Azure Data Studio:**

- SQL, KQL
- PowerShell
- Python net interactive
- PySpark
- Spark Scala
- Spark R



Try it out today: <a href="http://aka.ms/getAzureDataStudio">http://aka.ms/getAzureDataStudio</a>



