Notebooks in Microsoft Fabric

Jens Vestergaard

Jens Vestergaard (he/him)

Fabric Dude @ CatMan Solution, DK

In BI space since 2003 which means SQL Server 2000 -> … -> Azure SQL Database and SQL Server Analysis Services 2000 -> … -> Analysis Services Tabular Engine, but also ASP .Net 1.0 -> … -> .Net Core 8

So, in essence: SQL, C#, Pascal, Python, Basic, Prolog, MDX, Java Script, PowerShell, Visual Basic + more

Microsoft Data Platform MVP since 2017

8 active MS Certifications (Associate & Expert level)

jv@t-sql.dk

t-sql.dk

@vestergaardj

/in/jvestergaard

CatMan Solution

Notebooks in Microsoft Fabric

Introduction

Notebooks in Microsoft Fabric

- Welcome!
- The importance of Spark/PySpark/Pandas in data processing.
  - Distributed (scalability)
  - In-Memory (+ cache)
  - Python (+ ecosystem)
  - Unified processing engine
  - Fault tolerance (distributed)
  - Versatility (batch, real time/micro batches…)
- Outline the agenda for the session →

Agenda

- 📣 Introduction
- 📣 Overview of Spark, PySpark & Pandas
- 📣 So what's Microsoft's twist on this?
- 😳 Enter Notebook
- ⌨️ Hands on Experience
- 📣 Conclusion
- 🗣️ QnA

Spark, PySpark & Pandas

Notebooks in Microsoft Fabric

- This is the content

=

PySpark

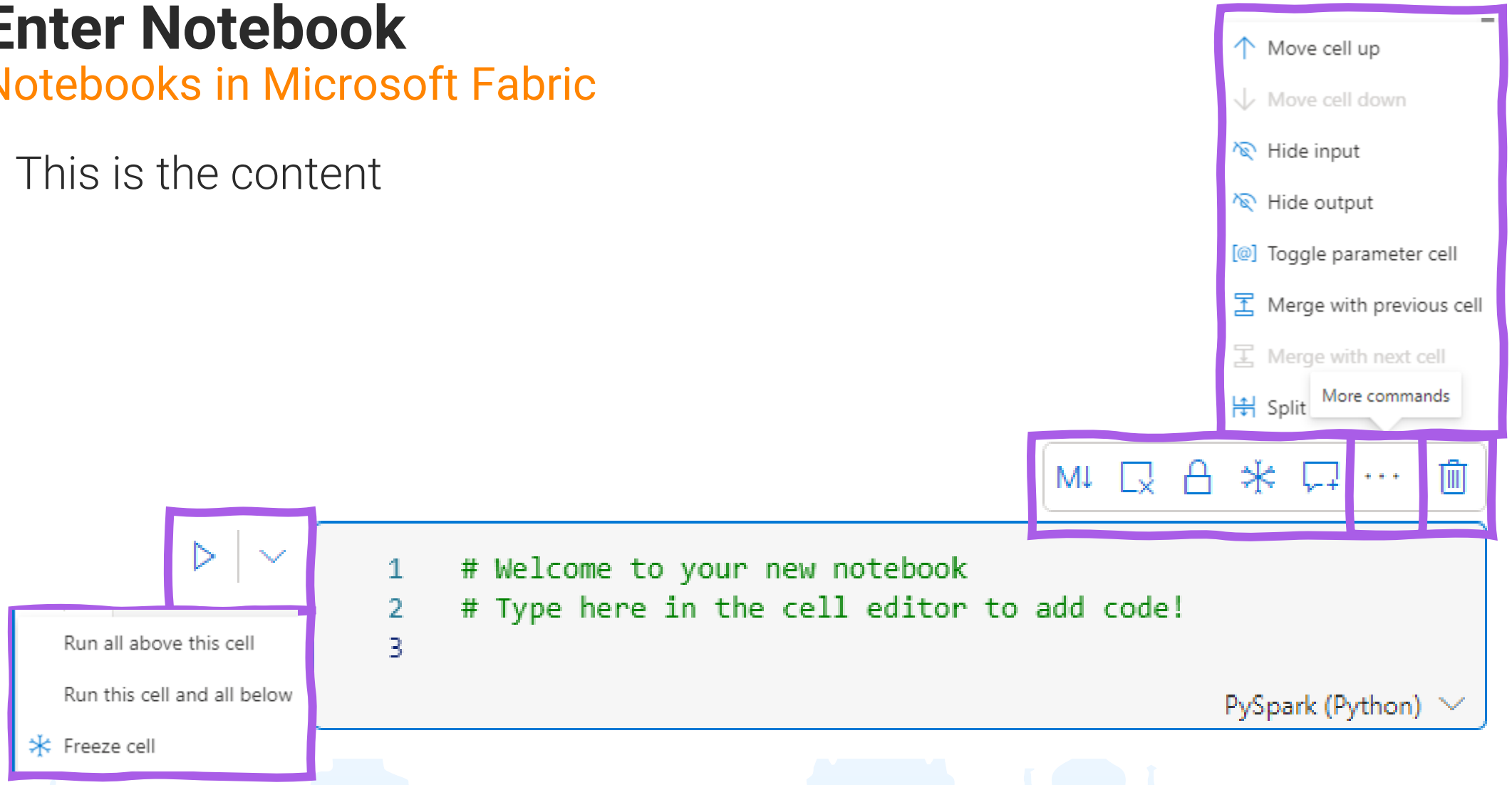
Well played Microsoft, well played…

Notebooks in Microsoft Fabric

Enter Notebook

Notebooks in Microsoft Fabric

- This is the content



Hands on Experience

Notebooks in Microsoft Fabric

Please proceed to the nearest Fabric Portal

(Synapse Data Engineering (microsoft.com))

Conclusion - PySpark

Notebooks in Microsoft Fabric

- **Why and How:**
  - Designed for distributed computing,
  - Suitable for large-scale data processing in parallel across clusters.
  - Utilize DataFrame and Spark SQL APIs
  - Leverage distributed computing for efficiency.
- **Benefits over SQL Server:**
  - Scalability
  - Parallel processing
  - Seamless integration with big data ecosystem.
- **Less ideal for:**
  - When you are dealing with a large number of small files
  - Interactive data exploration

Conclusion - Pandas

Notebooks in Microsoft Fabric

- **Why and How:**
  - Rich set of data structures and functions.
  - Use DataFrames for loading, cleaning, and analyzing smaller datasets
  - Flexibility in data manipulation.
- **Benefits over SQL Server:**
  - Ease of use
  - Flexibility
  - Rapid prototyping for data scientists and analysts.
- Less ideal when:
  - Memory constraints
  - Limited scalability
  - Streaming

Conclusion - Fabric

Notebooks in Microsoft Fabric

- Why & How:
  - Data Engineering: It's right next to your other favorite tools, like Power BI 😄
  - Notebooks: Lots of languages, frameworks, community contributions
  - Cost
  - ML/AI integration
  - Data integration
- Benefits over SQL Server:
  - What SQL Server?
  - Collaboration build-in
  - Open standard as opposed to proprietary format
- Less ideal when:
  - Don't know Python, but there is a fix for that 🫵

Feedback

Please rate this session!

Your feedback will help with

- speaker evaluation
- content relevance
- decision making for future events
- quality improvement

Q&A

Thank you!