

Notebooks a jejich použití na SQL Serveru

Mgr. Vladimír Mužný

MVP: Data Platform | MCSE: Data Management and Analytics | MCT

vladimir.muzny@dropman.cz

 @VladimirMuzny

Agenda

- Notebook není laptop, aneb jak s tím zacházet
- Notebooky v některých technologiích Microsoft
- Návrhy na praktické použití pro SQL odborníky



Notebook není laptop...

Bylo, nebylo...

- Notebook přichází s BigData clustery jako pohodlná možnost vývoje
 - Jupyter
 - Zeppelin
- Užitečné pro sofistikované analýzy dat
- Užitečné pro vysvětlení nebo pro ukázky
- Adoptovány mnoha technologiemi
 - Databricks
 - Azure Machine Learning
 - Visual Studio Code
 - Azure Data Studio
 - Azure Synapse
 - A jistě i další

Projekt Jupyter

- <https://jupyter.org/>
- Vznik v r. 2014 jako „obal“ kolem prostředí IPython
- Motivace:
 - Vývoj open-source platformy pro pokrytí velké škály různých jazyků
 - ◆ Jedna z velkých výhod notebooků dnes
 - Např.: DBC notebooky umožňují psaní v Python, R, Scala a dalších
- Web-based
 - Jupyter Server
- Vývoj
 - JupyterHub
 - JupyterLab

Struktura notebooku

- (Default) kernel
 - Runtime pro vykonávání spustitelného kódu
 - V některých alternativách notebooků tzv. výchozí kernel, který lze změnit
- Connection
 - Server, na kterém je zvolený runtime
- Text cell
 - Není spustitelný
 - Obsahuje formátovaný text
- Code cell
 - Spustitelný kód

Markdown

- Formátovaný text, mnohem čitelnější než komentáře jazyka
- Má svou syntaxi:
 - # (i několikrát): Nadpisy
 - **text**: Kurzíva
 - ****text****: tučné písmo
 - *: odrážky
 - 1.: číslování
 - [Jupyter](<https://jupyter.org>): Popsaný odkaz (bez mezery mezi závorkami)
 - ![Muj pes](https://iot02vladimir.blob.core.windows.net/data/IMG_1117_b.jpg)
- *Pozn: Dá se „obelstít“ HTML značkami*

Magic commands

- Pokud to varianta notebooku podporuje, lze každý code cell spustit v jiném kernelu
- Nastavení kernelu pomocí tzv. magic command, např.:
 - `%%python`
 - `%%r`
 - `%%scala`
 - `%%sql`
 - `%%powershell`
 - `%%bash`
 - A samozřejmě `%%md` (markdown; popis)
 - *Pozn.: některé varianty notebooků používají jen jeden symbol %*



Notebooky v technológiách Microsoftu

Notebooky v Azure

- HD Insight
 - Starší způsob hostování big data clusterů v Azure
 - Některé clustery (např. Apache Spark) obsahují Jupyter server a notebooky
- Azure Databricks
 - Technologie pro analýzu a zpracování velkých dat
 - Obálka nad Apache Spark
 - Notebooky v formátu DBC
 - Ukázka
- Azure Machine Learning
 - Ukázka

Notebooky lokálně

- Visual Studio Code
 - Podpora např. PowerShell
- Azure Data Studio
 - Multiplatformní (postupně i pro různé databáze), volně ke stažení
 - Zatím hlavně SQL práce
 - Zvláštnost: obsah proměnných se neudrží mezi dvěma code cell
 - Ukázka
 - ◆ Krátké představení ADS
 - ◆ Notebooky v ADS

Praktické použití pro SQL odborníky

Užití notebooků v praxi

- V Azure Data Studiu jsme schopni vytvářet „manuály“
- Příklady z praxe:
 - Vysvětlení nové funkcionality na SQL Serveru
 - Postup při obnovení databáze
 - Vysvětlení nastavení permissions
 - Oprava vlastnictví objektů ve schématu
 - A mnohé další

Shrnutí

- SSMS
 - Rutinní práce s databázemi
 - Mnoho dialogových oken
 - Chybějící možnost práce s jinými prostředími
 - Chybí napojení na Git
- ADS
 - Výborné pro dobrou dokumentaci, vysvětlení
 - Vizualizace dat (v plenkách)
 - Napojení na Git(hub)
 - ◆ Korektní zobrazení