

# Intelligent Query Processing

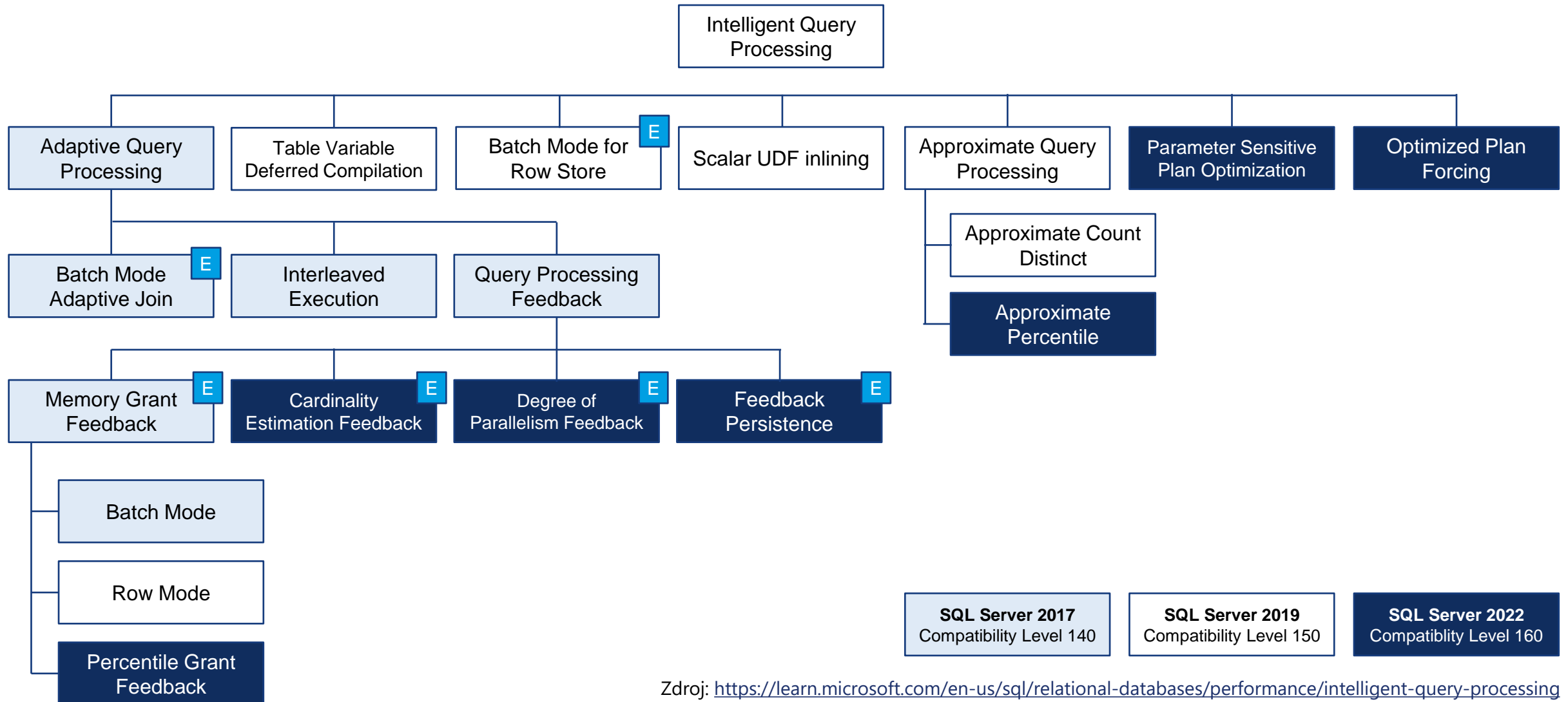
**RNDr. David Gešvindr, Ph.D.**

MVP: Data Platform | MCSE: Data Platform | MCT

[david@wug.cz](mailto:david@wug.cz)

 @gesvindr

# Intelligent Query Processing



# Memory Grant Feedback

Pouze Enterprise edice

- Dle exekučního plánu dotaz pro některé operace potřebuje paměť
  - Např.: Hash nebo Sort
- Množství paměti se spočítá při sestavení exekučního plánu a **je dále neměnné**
  - Pokud je paměti nedostatek, dojde k disk spill a využije se tempdb (pomalé)
  - Pokud je paměti zbytečně mnoho, alokujeme zbytečně paměť, kterou by využili ostatní

# Memory Grant Feedback

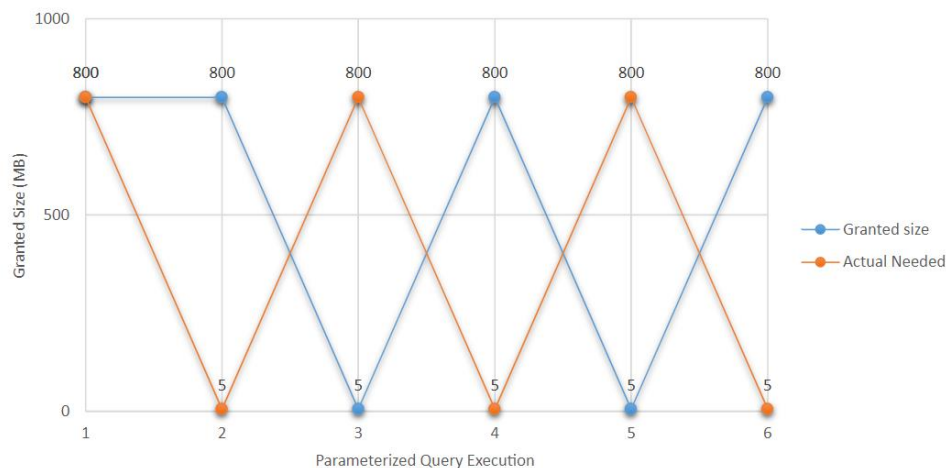
Pouze Enterprise edice

- Memory Grant Feedback zohlední stav využití paměti při exekuci dotazu a na základě zpětné vazby upraví velikost alokované paměti v exekučním plánu pro budoucí exekuce
- Od SQL Serveru 2017 (compatibility level 140)
  - Podporován pro batch processing mode (vyžadovány column store indexy)
- Od SQL Serveru 2019 (compatibility level 150)
  - Podporován i pro běžné dotazy využívající zpracování dat po řádku

# Memory Grant Feedback

Pouze Enterprise edice

- Před SQL Serverem 2022 se informace o posledním využití paměti dotazem uchovávala pouze v cache exekučního plánu
- SQL Server 2022 využívá Query Store pro trvalé uložení informace o memory grant feedbacku (*sys.query\_store\_plan\_feedback*) a zohledňuje nejen pouze předchozí exekuci dotazu
  - Předchází se tak problémům při oscilující alokaci paměti třeba vlivem parametrizace dotazu



# Degree of Parallelism (DOP) Feedback

Pouze Enterprise edice

- Cílem této technologie je identifikovat neefektivní využití paralelizmu při exekuci dotazů
- Pokud je detekováno neefektivní využití paralelizmu, tak začne SQL Server snižovat míru paralelizmu a ověřovat, jestli se zvyšuje efektivita dotazu
  - Minimální úroveň paralelizmu, kam DOP Feedback smí zajít je 2
  - Nemůže vynutit úplné zablokování paralelizmu na dotazu

# Cardinality Estimation Feedback

Pouze Enterprise edice

- Optimalizátor má různé modely odhadu kardinality, pomocí kterých odhaduje ze statistik, kolik řádků bude zpracováváno v jednotlivých částech exekučního plánu
- Použité modely jsou ovlivněny volbou compatibility levelu databáze (případně nastavením Legacy Cardinality Estimation)
- SQL Server 2022 analyzuje použité předpoklady pro výpočet kardinality a pokud zjistí, že nevedou k efektivnímu exekučnímu plánu, tak je změní a ověří dopad této změny
  - Correlation model assumptions při odhadu selektivity podmínky
  - Containment model assumptions při odhadu kardinality joinu
  - Optimizer row goal při použití TOP, IN či EXISTS

# Batch Mode Adaptive Join

Pouze Enterprise edice

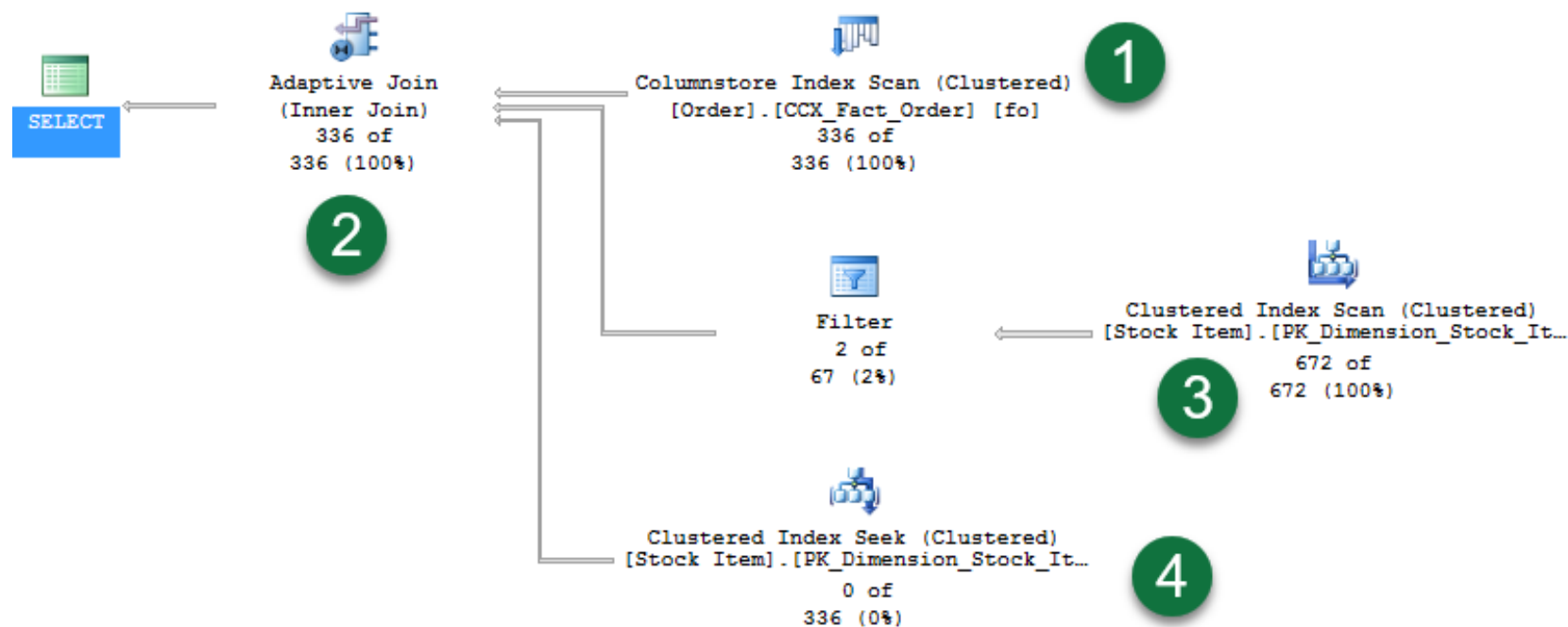
- Na základě odhadovaného počtu řádků se SQL Server rozhoduje v exekučním plánu mezi použitím algoritmu **nested loop** a **hash join**
- **Pokud se odhad liší od skutečnosti, může to mít fatální následky na výkon dotazu**
  - Problém i s parameter sniffingem, kdy se plán sestaví podle vstupních parametrů u první exekuce procedury



# Batch Mode Adaptive Join

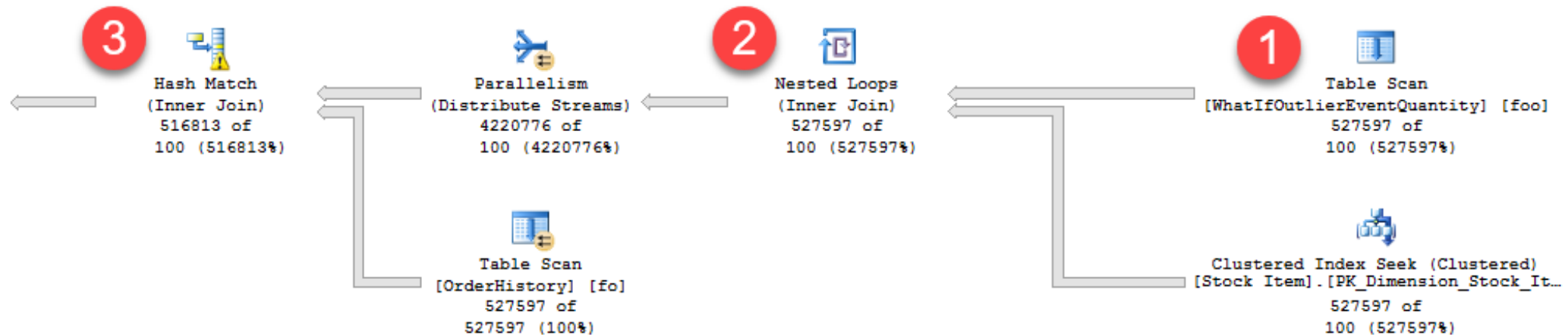
Pouze Enterprise edice

- Adaptive Join **nejprve ověří skutečný počet řádků na vstupu** a poté se rozhodne, který algoritmus pro join bude výhodnější



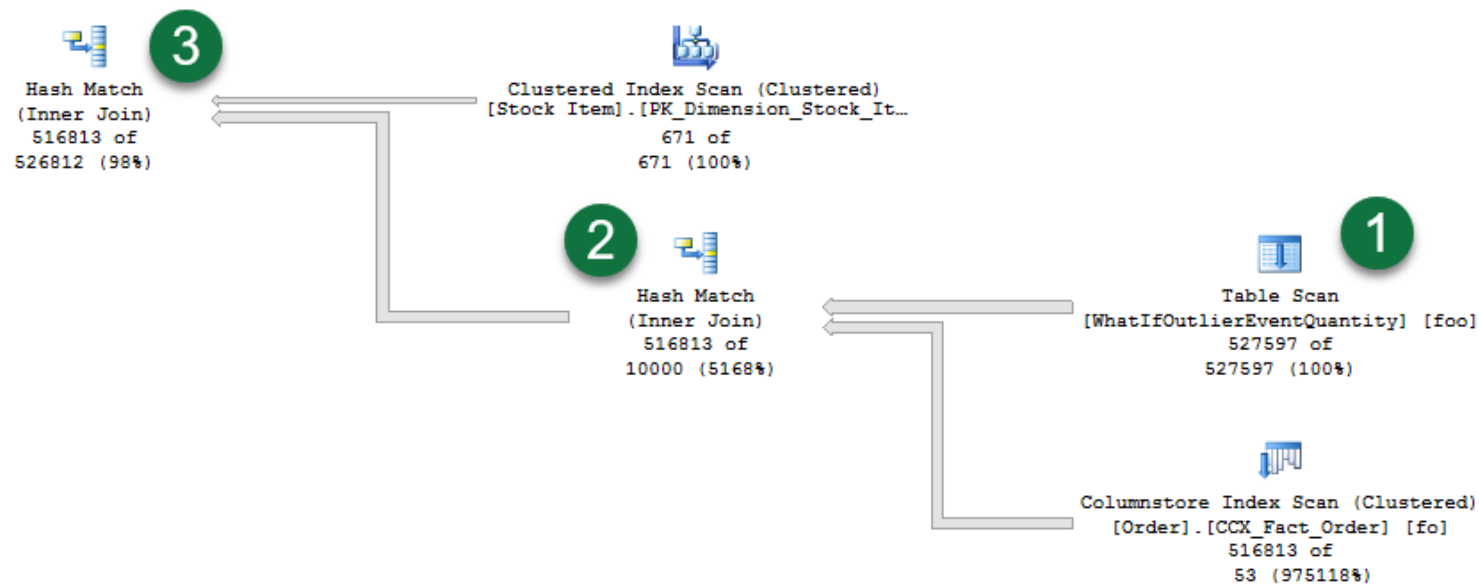
# Interleaved Execution

- Multi-statement Table Valued Functions mají kardinalitu v exekučním plánu pevně stanovenou na 100 (od SQL Serveru 2014, dříve 1)



# Interleaved Execution

- Interleaved Execution umí přerušit proces optimalizace dotazu, vyhodnotit neznámý podstrom a pro pokračování optimalizace použít skutečný počet řádků



# Table Variable Deferred Compilation

- Table Variables měly až doposud **pevně danou kardinalitu 1**
  - Způsobuje řadu problémů při optimalizaci dotazů
- Pokud je nastaven compatibility level 150, zjistí se skutečný počet řádků uložených v proměnné typu tabulka a ten se použije při sestavení exekučního plánu
  - Týká se pouze exekuce, při které se sestavuje exekuční plán (parameter sniffing)

# Batch Mode on Rowstore

Pouze Enterprise edice

- Row Processing Mode
  - Operátory v exekučním plánu zpracovávají data po individuálních řádcích
- Batch Processing Mode
  - Operátory v exekučním plánu zpracovávají data pro dávkách přibližně 900 řádků
  - Menší režie a podstatně rychlejší zpracování při velkých objemech dat (analytické dotazy)
  - V SQL Serveru 2017 a starších omezen pouze na column store indexy
  - Od SQL Serveru 2019 se použije i při práci s rowstore, **pokud se vyplatí**
    - ◆ Spojení a agregace dat, minimálně 100 000 řádků, velké vytížení CPU

# Scalar UDF inlining

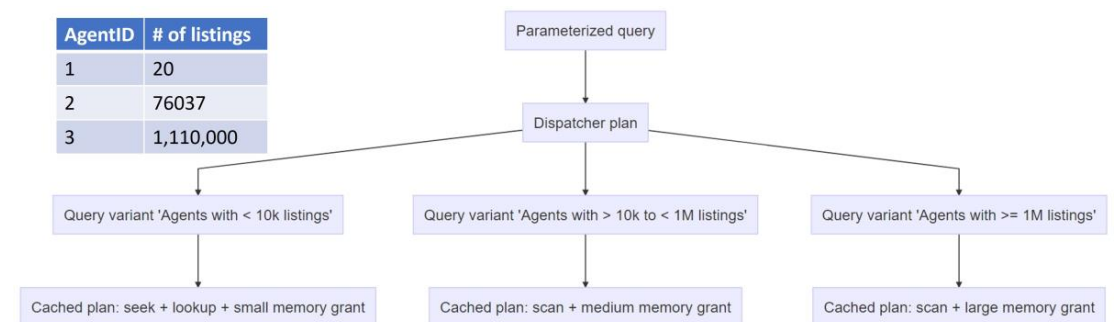
- Skalární funkce bývají častým výkonostním problémem
  - Nejsou součástí exekučního plánu dotazu
  - Volají se vždy nezávisle pro každou sadu vstupních parametrů
  - Nejsou oceněny (SQL Server neví, jak je exekuce funkce drahá)
  - Blokují paralelní zpracování
- Řešení? Nepoužívat skalární funkce nebo pořídit SQL Server 2019 😊

# Scalar UDF inlining

- SQL Server 2019 analyzuje skalární funkci a pokud je to možné její tělo automaticky vloží do samotného dotazu.
  - Zahrnuje komplexní transformace zdrojového kódu funkce (IF → CASE)
  - Odpadnou nevýhody spojené s exekucí skalární funkce (v dotazu prakticky neexistuje)
- Požadavky kladené na skalární funkci:
  - Funkce používá DECLARE, SET, SELECT, IF/ELSE, RETURN, další UDF
  - Nesmí záviset na čase, nesmí používat EXECUTE AS, nesmí používat TV/TVP

# Parameter Sensitive Plan (PSP) Optimization

- Parametrizované dotazy mohou trpět na problém známý jako **parameter sniffing**
  - Exekuční plán se vygeneruje výhodný pro sadu parametrů při prvním volání, uloží se do cache exekučních plánů a následně se využívá i pro další hodnoty parametrů, což nemusí být výhodné
- SQL Server 2022 umožní, aby jeden parametrizovaný dotaz měl v cache více exekučních plánů v závislosti na hodnotě parametru
- Při sestavování exekučního plánu se analyzují statistiky a pokud se identifikuje příliš nerovnoměrné rozložení hodnot, tak se vygeneruje více exekučních plánů





# Další informace

- [GOC631](#) Optimalizace, ladění a monitorování T-SQL dotazů
- [GOC278](#) Tvorba pokročilých T-SQL dotazů v prostředí Microsoft SQL Serveru
- [GOC628](#) Zpracování XML, JSON, hierarchických a geografických dat v Microsoft SQL Serveru
- [GOC212](#) Správa a provoz Microsoft SQL Server databází v Microsoft Azure

# Dotazy

**RNDr. David Gešvindr, Ph.D.**

MVP: Data Platform | MCSE: Data Platform | MCT

[david@wug.cz](mailto:david@wug.cz)

 @gesvindr