



## Extended Events

David Hlaváček

MCP, MCSE, MCSA, MCTS, MCITP, MCT

HLAVACEK *pro*

1

Základní přehled architektury Extended Events

Tvorba Session Step by Step

Monitorování dotazů a modulů

Speciální akce a filtry

Jak sledovat výkonově problematické události

Speciální eventy

Metadata

HLAVACEK *pro*

2

## Co jsou Extended Events ?

- Pokročilá monitorovací technika založena na sběru událostí
  - Uvedena v MS SQL Server 2008
  - GUI od SSMS 2012
- Hlavním cílem je zjednodušení identifikace problémů díky možnosti vysoce komplexní a sofistikované konfigurace
- Lightweight
- Umožňuje sledovat události, které jiné metody nedokáží nebo velmi obtížně
  - Page Split stránek indexů (alias „špatný“ page split)
  - Zaznamenat pouze prvních N výskytů události nebo jen každou N-tou
  - Wait statistiky na úrovni připojení
  - Použití plan\_handle a tsql\_stack

## Proč nepoužívat SQL Trace a Profiler?

- Nízká flexibilita monitorovací techniky, není rozvíjen
- Vysoký výkonnostní dopad na monitorovaný systém v závislosti na
  - Množství a typech trasovaných událostí v jediném trasování
  - Aktuální zátěži trasovaného serveru
  - Množství aktivních trasování
- Trasované události vždy generují všechny informace bez ohledu na definici trasování
  - Všechny trace columns
- Filtr trasování je aplikován až **po** zachycení dané události
- Pouze 2 typy I/O Provider
- Restart monitorované instance ukončí trasování

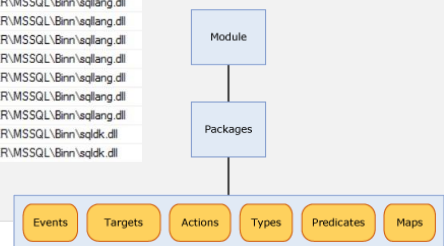
HLAVACEK *partner*

5

## Extended Events Objekty

- Package
  - Nejvyšší metadatový objekt, zaveden modulem při startu
  - Kontejner obsahující objekty a jejich vlastnosti
  - Není funkčně omezen
    - Objekty jednoho Package mohou být použity objekty jiného
  - Např. package0, sqlservr, sqllos, sqlclr

name	description	file_version	name
sqlservr	Extended events for Microsoft SQL Server	2015.130.1711.0	C:\Program Files\Microsoft SQL Server\MSSQL13.MSSQLSERVER\MSSQL\Binn\sqlmin.dll
sqlservr	Extended events for Microsoft SQL Server	2015.130.1711.0	C:\Program Files\Microsoft SQL Server\MSSQL13.MSSQLSERVER\MSSQL\Binn\sqlang.dll
SecAudit	Security Audit Events	2015.130.1711.0	C:\Program Files\Microsoft SQL Server\MSSQL13.MSSQLSERVER\MSSQL\Binn\sqlang.dll
ucs	Extended events for Unified Communications Stack	2015.130.1711.0	C:\Program Files\Microsoft SQL Server\MSSQL13.MSSQLSERVER\MSSQL\Binn\sqlang.dll
sqlclr	Extended events for SQL CLR	2015.130.1711.0	C:\Program Files\Microsoft SQL Server\MSSQL13.MSSQLSERVER\MSSQL\Binn\sqlang.dll
filestream	Extended events for SQL Server FILESTREAM and F...	2015.130.1711.0	C:\Program Files\Microsoft SQL Server\MSSQL13.MSSQLSERVER\MSSQL\Binn\sqlang.dll
SQL.Satellite	Extended events for SQL Satellite	2015.130.1711.0	C:\Program Files\Microsoft SQL Server\MSSQL13.MSSQLSERVER\MSSQL\Binn\sqlang.dll
sqlsni	Extended events for Microsoft SQL Server	2015.130.1711.0	C:\Program Files\Microsoft SQL Server\MSSQL13.MSSQLSERVER\MSSQL\Binn\sqlang.dll
sqlservr	Extended events for Microsoft SQL Server	2015.130.1711.0	C:\Program Files\Microsoft SQL Server\MSSQL13.MSSQLSERVER\MSSQL\Binn\sqlang.dll
package0	Default package. Contains all standard types, maps, c...	2015.130.1711.0	C:\Program Files\Microsoft SQL Server\MSSQL13.MSSQLSERVER\MSSQL\Binn\sqlidk.dll
sqlos	Extended events for SQL Operating System	2015.130.1711.0	C:\Program Files\Microsoft SQL Server\MSSQL13.MSSQLSERVER\MSSQL\Binn\sqlidk.dll

HLAVACEK *partner*

## Extended Events Objekty

### • Event

- Reprezentuje oblast/bod v kódu, který je vykonán SQL Serverem
  - Např. dokončení SQL dotazu, deadlock, čekání na zdroj atd.
- Obsahuje sadu základních informací, které event popisují
  - Tato sada je k dispozici okamžitě a bude zaznamenána vždy
- Může obsahovat volitelné elementy, které budou zaznamenány pouze na vyžádání

Global Fields (Actions) Filter (Predicate) Event Fields

Name	Description	Data Type
cpu_time	Indicates the CPU time (in microseconds)	uint64
duration	The time (in microseconds)	uint64
last_row_count	The last row count for this statement	uint64
line_number	The statement line number	int32
logical_reads	The number of logical pages read	uint64
offset	The statement start offset	int32
offset_end	The statement end offset	int32
<input type="checkbox"/> parameterized_plan_handle	The plan handle of the current statement	binary_data
physical_reads	The number of physical pages read	uint64
row_count	The number of rows that were returned	uint64
<input checked="" type="checkbox"/> statement	The text of the statement	unicode_string
writes	The number of page writes	uint64

Module

Packages

Events Targets Actions Types Predicates Maps

HLAVACEK *pro*

## Extended Events Objekty

### • Action

- Slouží k zajištění dodatečných informací k nastalé události
  - Informace, které event neobsahuje (session\_id, database\_name apod.)
  - Pokročilé informace jako memory dump, debug break apod.
- Jsou procesovány synchronně po vyhodnocení predikátu
  - Mají tedy vyšší výkonnostní nároky na thread, který událost spustil
- Actions mohou být použity u všech eventů, ale některá data nemusí být dostupná v dané části kódu, kde event nastane

Global Fields (Actions) Filter (Predicate) Event Fields

Name	Description
<input type="checkbox"/> database_id	Collect database ID
<input type="checkbox"/> database_name	Collect current database name
<input type="checkbox"/> debug_break	Break the process in the default debugger
<input type="checkbox"/> dump_rg_history_ring_buffer	Dump pool and group ring buffer to xevent
<input type="checkbox"/> event_sequence	Collect event sequence number
<input type="checkbox"/> is_system	Collect whether current session is system
<input type="checkbox"/> last_error	Collects the value of the thread's last Windows error
<input type="checkbox"/> nt_username	Collect NT username

Module

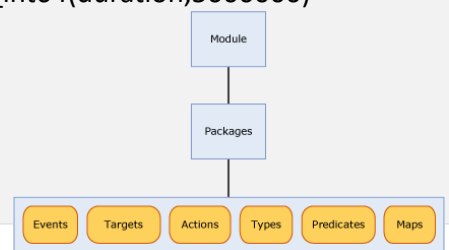
Packages

Events Targets Actions Types Predicates Maps

HLAVACEK *pro*

## Extended Events Objekty

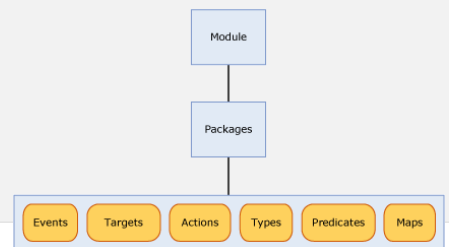
- Predicate
  - Boolean výraz definující podmínku, na základě které dojde k zaznamenání události
  - Lze použít aritmetické nebo textové operátory, možnost sestavení komplexních podmínek (LIKE, NOT, () atd.)
  - Filtrovat lze na základě event dat nebo globálních vlastností
  - Predicate může uložit svůj stav
    - Zachycení pouze prvních 10 výskytů nebo jen sudých
    - Zachycení duration vyšší než všechny předchozí dotazy
      - `duration > 5000000` vs `greater_than_max_int64(duration,5000000)`
  - Pořadí podmínek hraje roli



HLAVACEK *pro*

## Extended Events Objekty

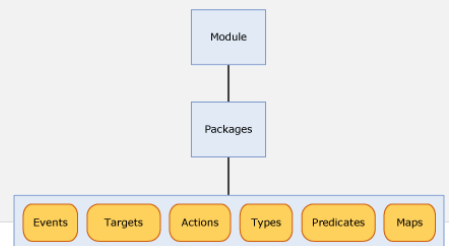
- Target
  - Slouží pro konzumaci (uložení) událostí
  - Jsou optimalizovány pro procesování jediné události nebo celého bufferu
  - Synchronní vs. Asynchronní
  - Základní
    - Ring Buffer
    - Event File
  - Pokročilé
    - Bucketizer (Histogram)
    - Event Counter
    - Event Pairing
    - ETW (Event Tracing for Windows)



HLAVACEK *pro*

## Extended Events Objekty

- Type
  - Udává datový typ sloupců v eventu, predikátů, actions atd.
- Map
  - Zjišťuje význam nebo textový popis na základě číselného identifikátoru
  - Např. wait\_type = 181 znamená WRITELOG



HLAVACEK *plus*

## Životní cyklus Extended Events



- Pokud při filtrování použijeme některou globální vlastnost (Action), musí dojít k synchronnímu volání funkce pro zjištění její hodnoty

HLAVACEK *plus*

## Extended Event Engine

- Operační prostředí nabízející funkce a služby pro jednotlivé moduly
  - Obsahuje seznam Event Session a kontroluje přístup k nim
  - Spravuje memory buffer pro všechny aktivní sessions
  - Spravuje vlákna, která umísťují informace z paměti do cílů (dispatchers)
- Neobsahuje metadata ani eventy
  - Jednotlivé moduly spolupracují s XE Engine a zajišťují metadata pro jednotlivé objekty

## Extended Event Session

- Objekt, obsahující konfiguraci událostí, které mají být monitorovány
  - Včetně všech souvislostí jako filtrování, informace o událostech, cíl pro uložení atd.
- Lze definovat „libovolné“ množství Session
  - Session je permanentním objektem (není nutné po restartu opět vytvářet)
  - Každou session lze zapnout/vypnout dle potřeby
    - Příprava session pro konkrétní situaci
    - Současně může být spuštěno několik session
  - Možnost automatického zapnutí po restartu služby
    - Neplánovaný restart služby nemusí přerušit monitorování
- Každá session má vlastní memory buffer, do kterého ukládá informace o událostech před umístěním do požadovaného cíle

## Vlastnosti Event Session

- MAX\_MEMORY
  - Maximální velikost paměti pro Session
  - Zaokrouhleno na 64 kB
  - Velikost ovlivňuje zápis do cíle typu file
  - Vyšší hodnota zlepšuje stabilitu streaming cíle
    - Watch Live Data
  - Standardně rozdělena na 3 buферы

Event retention mode:  Single event loss  
 Multiple event loss  
 No event loss (not recommended)

Maximum dispatch latency:  In seconds   
 Unlimited

Max memory size:  MB

Max event size:  MB

Memory partition mode:  None  
 Per node  
 Per CPU

Restore Defaults

- sys.dm\_xe\_sessions

name	total_regular_buffers	regular_buffer_size	total_large_buffers	large_buffer_size	total_buffer_size	buffer_policy_desc
Test01	3	1441587	0	0	4324761	drop_event
Test02	3	7012147	0	0	21036441	drop_event

HLAVACEK *plus*

15

## Vlastnosti Event Session

- MAX\_PARTITION\_MODE
  - Ovlivňuje počet memory buffers
- NONE
  - 3 buферы
  - MAX\_MEMORY rovnoměrně dělena na 3
- PER\_NODE
  - 3 buферы pro každý NUMA node
  - MAX\_MEMORY rovnoměrně mezi počet bufferů
- PER\_CPU
  - 2.5 bufferu pro každý jeden SQL scheduler
- Doporučení
  - Běžně používat PER\_NUMA, při vyšším očekávaném počtu událostí PER\_CPU

Event retention mode:  Single event loss  
 Multiple event loss  
 No event loss (not recommended)

Maximum dispatch latency:  In seconds   
 Unlimited

Max memory size:  MB

Max event size:  MB

Memory partition mode:  None  
 Per node  
 Per CPU

Restore Defaults

HLAVACEK *plus*

16



## Vlastnosti Event Session

- **MAX\_EVENT\_SIZE**
  - Zajišťuje tvorbu large buffers pro události, které jsou větší než standardní buffer
    - Eventy větší než buffer jsou standardně zahozeny
- **MAX\_EVENT\_SIZE >= MAX\_MEMORY**
- Při sledování událostí s velkým objemem dat je konfigurace nutná
  - sql\_text action
  - io\_data události file\_written

Event retention mode:  Single event loss  
 Multiple event loss  
 No event loss (not recommended)

Maximum dispatch latency:  In seconds   
 Unlimited

Max memory size:  MB

**Max event size:  MB**

Memory partition mode:  None  
 Per node  
 Per CPU

[Restore Defaults](#)

## Vlastnosti Event Session

- **EVENT\_RETENTION\_MODE**
  - Ovlivní možnost ztráty monitorovaných informací
  - Single event, entire memory buffer, no event loss
  - NO\_EVENT\_LOSS může zásadně degradovat výkon
  - Pokud není volné místo v bufferu, standardně dochází ke ztrátě jednoho daného eventu
- **MAX\_DISPATCH\_LATENCY**
  - Maximální čas, po který zůstane event v bufferu než je umístěn do asynchronního cíle
  - Ovlivní výkon vláken pokud je spuštěno větší množství session s nízkou dispatch latency

**Event retention mode:  Single event loss**  
 Multiple event loss  
 No event loss (not recommended)

Maximum dispatch latency:  In seconds   
 Unlimited

Max memory size:  MB

Max event size:  MB

Memory partition mode:  None  
 Per node  
 Per CPU

[Restore Defaults](#)



## Jak začít s Extended Events ?

- GUI
  - Pro SQL Server 2008 je dostupný add-in
    - [www.sqlskills.com/free-tools/sql-server-2012-extended-events-add-in/](http://www.sqlskills.com/free-tools/sql-server-2012-extended-events-add-in/)
  - Pro SQL Server 2012 a vyšší je GUI v Management Studiu
  - Doporučuji použít **SSMS 16.3 (August 2016)**
    - Extended Events Templates se stejnou funkcí jako Profiler
- T-SQL

```
CREATE EVENT SESSION [xeMonitor] ON SERVER
ADD ...
```



## Action : sql\_text

- Jedná se o InputBuffer dané session
  - Nikoliv text dotazu, který je vykonáván (statement)
  - Snadno tak dochází omylům
- InputBuffer je maskován při výskytu slov naznačující bezpečnostně citlivá data
  - Např. password
- S ohledem na velikost InputBuffer může být sql\_text náročný na paměť a další zpracování

name	timestamp
sp_statement_completed	2016-10-02 16:47:54

Event: sp\_statement\_completed (2016-10-02 16:47:54.9289008)

Field	Value
cpu_time	0
database_name	AdventureWorks2014
duration	11671663
last_row_count	504
line_number	5
logical_reads	17
nest_level	1
object_id	1623676832
object_name	WaitingProc
object_type	PROC
offset	84
offset_end	146
physical_reads	0
row_count	504
source_database_id	15
sql_text	EXEC WaitingProc
statement	SELECT * FROM Production.Product
writes	0

HLAVACEK *pro*

23

## Speciální filtry a vlastnosti eventů

- package0.counter
  - Lze použít v predikátu eventu pro zjištění počtu výskytu dané události
  - Predikát package0.counter <= 50
    - Zaznamená pouze prvních 50 výskytů události
- divides\_by\_uint64
  - Operátor predikátu
  - Umožňuje filtrovat pouze události, kdy je zbytek po celočíselném dělení 0
  - [package0].[divides\_by\_uint64]([package0].[counter],(2))
    - Zaznamená každou druhou událost

HLAVACEK *pro*

24

## Causality Tracking

Causality tracking:

 Track how events are related to one another.

- Usnadňuje identifikaci vztahu různých událostí
- Ve výstupu označí události identifikátorem `attach_activity_id.guid`
  - Události se stejným identifikátorem jsou ve vzájemném vztahu

name	timestamp	attach_activity_id.guid	statement
sql_statement_starting	2016-10-18 13:57:19.6919763	97625399-04BE-476A-B0A2-041A484BF249	SELECT * FROM Sales.SalesOrderDetail a JOIN Sales.SalesOrderDetail b O...
attention	2016-10-18 13:57:28.1591479	97625399-04BE-476A-B0A2-041A484BF249	NULL

HLAVACEK *pro*

25

## Recompile

- `sql_statement_starting` a `sp_statement_starting` obsahují sloupec **state**
  - **state** má 3 stavy Normal, Recompiled, Execution Plan Flush
  - Pokud dojde k Recompiled, událost starting bude ve výstupu opakovaně
    - Recompiled i Normal
    - Pozor na analýzu výstupu (Event Pairing)
- `sql_statement_recompile`
  - Nastává při recompile ad-hoc dotazu, procedury nebo triggeru
  - Sloupec `statement_recompile_cause` vysvětluje příčinu recompile

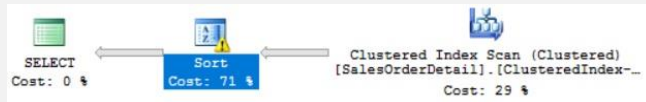
sql_statement_recompile	2016-10-19 10:18:09.7762442	Schema changed	SELECT * FROM Sales.SalesOrderDetail
sql_statement_recompile	2016-10-19 10:18:55.9361821	Option (recompile) requested	SELECT * FROM Sales.SalesOrderDetail
sql_statement_recompile	2016-10-19 10:24:34.1214104	Statistics changed	SELECT * FROM Sales.SalesOrderDetail
sql_statement_recompile	2016-10-19 10:24:21.3146208	Deferred compile	SELECT case when

HLAVACEK *pro*

26

## Dotazy s problémy v plánu

- sort\_warning / hash\_warning (spill operace)



### Warnings

Operator used tempdb to spill data during execution with spill level 1 and 1 spilled thread(s); Sort wrote 589 pages to and read 589 pages from tempdb with granted memory 6656KB and used memory 6656KB

- Chybějící JOIN predikát
  - missing\_join\_predicate

```
SELECT b.SalesOrderID, a.OrderQty, b.CustomerID
FROM Sales.SalesOrderDetail a, Sales.SalesOrderHeader b
WHERE a.ProductID = 927
```

## Dotazy s problémy v plánu

- Chybějící statistiky
  - missing\_column\_statistics

missing_column_statistics	2016-10-19 09:42:46.0324787	62A1BB1E-31BF-446C-A7AE-DBF55B8F475E	NULL
sql_statement_completed	2016-10-19 09:42:46.0507503	62A1BB1E-31BF-446C-A7AE-DBF55B8F475E	SELECT SalesOrderID, ProductID

- Chybějící JOIN predikát
  - missing\_join\_predicate

```
SELECT b.SalesOrderID, a.OrderQty, b.CustomerID
FROM Sales.SalesOrderDetail a, Sales.SalesOrderHeader b
WHERE a.ProductID = 927
```

## Dotazy s problémy v plánu

- Problematická konverze
  - plan\_affecting\_convert

plan_affecting_convert	2016-10-19 09:46:04.5344207	Cardinality Estimate	AFD50FD8-A4E7-4532-97A2-47859255F3EB	NULL
plan_affecting_convert	2016-10-19 09:46:04.5344211	Seek Plan	AFD50FD8-A4E7-4532-97A2-47859255F3EB	NULL
sql_statement_completed	2016-10-19 09:46:04.6335221	NULL	AFD50FD8-A4E7-4532-97A2-47859255F3EB	SELECT LoginID, JobTitle, OrganizationN

- Filtr neodpovídá datům filtrovaného indexu
  - unmatched\_filtered\_indexes

unmatched_filtered_indexes	2016-10-19 09:42:56.4302816	NULL	0649444E-A06B-40CF-AB97-1EBF0D1A6BF0	NULL
sql_statement_completed	2016-10-19 09:42:56.4359000	NULL	0649444E-A06B-40CF-AB97-1EBF0D1A6BF0	SELECT SalesOrderID, ProductID, UnitPri

## Identifikace vzniklých chyb

- error\_reported
  - Chyby vzniklé v průběhu vykonávání dotazu
  - I v případě, že jsou zachyceny TRY/CATCH
    - is\_intercepted označuje, zda byla chyba zachycena TRY/CATCH
  - Možnost filtrování error\_number, severity, state
- Příklady error\_number
  - 220 – Přetečení datového typu
  - 507 – DELETE řádku, na který se odkazuje FK
  - 2627 – INSERT duplicitního řádku do tabulky s UNIQUE
  - 8134 – Chyba dělení 0
  - 8152 – Truncate textové hodnoty

```

error_number      50000
is_intercepted   True
message          Nastala chyba
query_hash       0
severity         16
sql_text         BEGIN TRY  RAISERROR(N'Nastala chyba',16,1)  END TRY BEGIN CATCH  PRINT 'Osetrena'  END CATCH
state            1
tsql_stack       <frames><frame level=1' handle=0x0100010079F9871550BD0DB0F0200000000000000000000000000000000000000
user_defined     True

```

## Sledování čekání na dostupnost zdrojů

- wait\_info
  - Zachycuje čekání na Cooperative (non-preemptive) zdroje
  - Párový event (opcode start/end)
- wait\_info\_external
  - Zachycuje čekání na Preemptive zdroje
    - Zdroje mimo SQL Server (např. NTLM autentizace DC)
- wait\_completed
  - Stejný význam jako výše uvedení, ale není párový
  - Nastává při dokončení čekání na zdroj

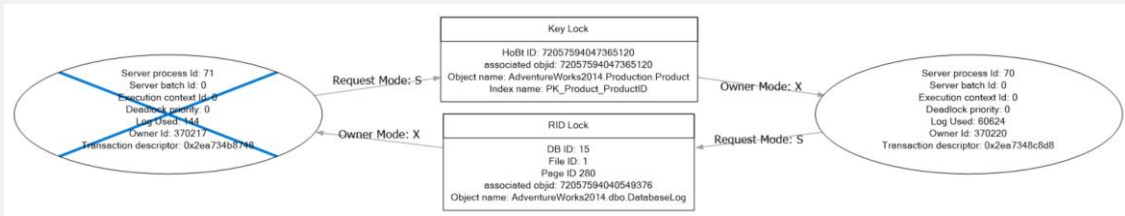
## Parameter Sniffing

- Výchozí chování u programových modulů (procedury, funkce, trigger)
- **Může** způsobovat výkonnostní problémy
- Použijeme tyto události :
  - sp\_statement\_completed
  - sql\_statement\_completed
  - U těchto událostí budeme sledovat standardní výkonové vlastnosti
    - cpu\_time, logical\_reads, duration atd.
    - Ideálně spolu s action query\_hash
- Na výsledek monitorován použijeme STDEV



## Lock, block, deadlock, timeout, cancel

- xml\_deadlock\_report



- attention
  - Předčasné ukončení operace klientem (cancel, ukončení připojení, chyba ovladače)

## Lock, block, deadlock, timeout, cancel

- lock\_timeout / lock\_timeout\_greater\_than\_0

Field	Value
associated_object_id	72057594040549376
attach_activity_id_guid	4464EB9D-7ADA-4DF9-8DD2-B99CF7330E5C
attach_activity_id_seq	4
database_id	15
database_name	
duration	5000000
lockspace_nest_id	1
lockspace_sub_id	1
lockspace_workspace_id	3206088617840
mode	S
object_id	0
owner_type	Transaction
resource_0	280
resource_1	1
resource_2	0
resource_description	
resource_type	RID
transaction_id	407412

- blocked\_process\_report
  - Zjistí blokováného a blokujícího, pokud délka čekání přesáhla definovaný čas
  - Délka blokování je konfigurována vlastností 'blocked process threshold'

## Speciální eventy

- query\_memory\_grant\_usage
  - Zachytí dotaz, jehož memory grant je vyšší než 5 MB
  - Slouží pro porovnání granted\_memory a used\_memory
- query\_thread\_profile
  - Umožňuje analyzovat každý iterátor v plánu a vlákna při paralelních operacích
- transaction\_log
  - Umožňuje sledovat zápisy do transakčního logu databáze
  - Ideální pro přesné monitorování „spatných“ page split
- inaccurate\_cardinality\_estimate
  - Identifikuje operátory v plánu, jejichž rows estimate se pětinašobně liší od actual

## Něco speciálně pro administrátory

- backup\_restore\_progress\_trace
  - Umožňuje detailně sledovat jednotlivé operace backup/restore procesu
- database\_backup\_restore\_throughput
  - Umožňuje detailně sledovat výkon (množství toku dat) při backup/restore
- perfobject\_logicaldisk
- perfobject\_proces
- perfobject\_processor
- perfobject\_system
  - Události nastávají každých 15 sec a sbírají informace z čítačů Performance monitoru
  - Bohužel se jedná o RAW data a bez Base hodnoty jsou **nepoužitelné**

---

## Extended Events metadata

- `sys.fn_xe_file_target_read_file()`
  - Funkce slouží pro čtení dat z event file výstupu (.xel soubory)
- `sys.dm_xe_session_targets`
  - Obsahuje informace o uložení dat
  - V případě paměťového cíle samotná data (`target_data`)
- `sys.dm_xe_session`
  - Informace o všech aktivních sessions (např. velikost memory bufferu)
- `sys.server_event_sessions`
  - Seznam a základní vlastnosti všech existujících event sessions