

# NOVINKY PS CLEMENTINE 5.0

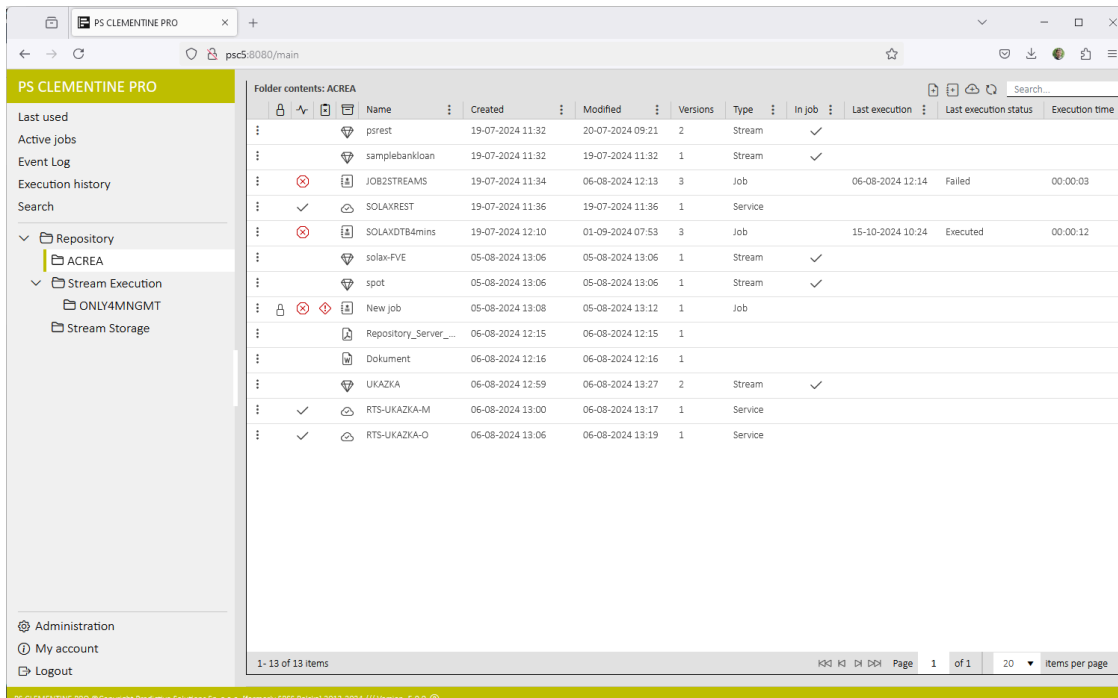
PS Clementine (PSC) je dataminingový nástroj pro přípravu a analýzu dat využívající nejnovější analytický engine IBM SPSS Modeler 18.5.

Jedná se o sofistikované, flexibilní, uživatelsky jednoduché, intuitivní a škálovatelné řešení. V rámci organizace ho lze nasadit a implementovat do firemních procesů na řešení a automatizaci jednoduchých rutinních i komplexních (analytických či datově náročných) úloh.

Aktuální verze přináší celou řadu novinek.

## Nové aplikační rozhraní

Nově je jediným rozhraním řešení PSC webový klient. Není tedy nutné instalovat další aplikaci a distribuce řešení business uživatelům je tak výrazně jednodušší. Pro obsluhu řešení PSC tak stačí pouze prohlížeč a uživatelské přihlášení, podle něhož je business uživatelům zpřístupněna určitá funkcionality a obsah analytického úložiště.



## Diagram úloh

Jednoduché grafické rozhraní pro organizaci běhu různých typů procesů v rámci jedné komplexní automatizované úlohy (jobu).

The screenshot displays the PS CLEMENTINE PRO web interface. On the left is a navigation sidebar with options like 'Last used', 'Active jobs', and 'Repository'. The main area shows a folder 'ACREA' containing a table of jobs. Below the table, a 'Job information' window is open for 'New job [Latest] as of [16-10-2024 10:27:38]', showing a flow diagram with nodes 'Start', 'psrest', 'samplebankloan', 'solax-FVE', and 'spot'.

Name	Created	Modified	Versions	Type	In job	Last execution	Last execution status	Execution time
psrest	19-07-2024 11:32	20-07-2024 09:21	2	Stream	✓			
samplebankloan	19-07-2024 11:32	19-07-2024 11:32	1	Stream	✓			
JOBZSTREAMS	19-07-2024 11:34	06-08-2024 12:13	3	Job		06-08-2024 12:14	Failed	00:00:03
SOLAXREST	19-07-2024 11:36	19-07-2024 11:36	1	Service				
SOLAXDTB4mins	19-07-2024 12:10	01-09-2024 07:53	3	Job		15-10-2024 10:24	Executed	00:00:12
solax-FVE	05-08-2024 13:06	05-08-2024 13:06	1	Stream	✓			
spot	05-08-2024 13:06	05-08-2024 13:06	1	Stream	✓			
New job	05-08-2024 13:08	16-10-2024 10:27	1	Job				

## Práce s novými formáty souborů

V rámci analytického úložiště mohou být uloženy a editovány skripty Python, PowerShell a windowsové bat soubory. Webové rozhraní zároveň nabízí jednoduchý editor programovacího kódu pro editaci a kontrolu správnosti, případně pro definování parametrů.

## Rozšířená správa virtuální adresářové struktury analytického úložiště

V rámci analytického úložiště lze exportovat/importovat objekty i celé adresáře, což výrazně usnadňuje práci se soubory při replikaci prostředí z jiných instancí PSC.

## Služby komunikující protokolem WebSocket

Platí pouze pro architekturu Central Processing with Server. Nad vybranými streamy jsou vytvořeny tzv. channels, které jsou volány pomocí komunikačních protokolů WebSocket a umožňují obousměrnou komunikaci v reálném čase mezi klientem a serverem. Tím byla prakticky odstraněna časová prodleva při inicializaci streamu a odezva je prakticky okamžitá. Těchto „channels“ může být otevřeno libovolné množství, nicméně je třeba brát v potaz na hardwarové možnosti, především RAM použitého stroje.

## Promotování objektů a úloh mezi prostředím

Platí pouze pro architekturu Central Processing. Nově byl přidán mechanismus pro přenos (promotování) objektů a úloh mezi různými prostředím, např. mezi testem a produkcí. Úlohy a všechny potřebné objekty (streamy, skripty, parametry, nastavení) mohou být přenášeny automaticky což výrazně usnadňuje a urychluje proces nasazení a výrazně snižuje riziko chyby.

## Verzování úloh v úložišti

Automatizované úlohy mohou být verzovány, což výrazně zjednodušuje týmovou spolupráci, sledování změn a nasazování v různých prostředích (dev/test/prod).

## Nové funkcionality v IBM SPSS Modeler 18.5

- **Date/Time funkce**

Funkce `date_add` a `time_add` nabízejí v prostředí Expression Builder více možností a flexibility při vytváření výrazů.

- **Nové možnosti s využitím jazyka Python**

Záložka Python ve vybraných uzlech umožňuje využití skriptovacího jazyka Python a možnost volání nativního Python API rozhraní pro interakci s IBM SPSS Modeler.

- **Vizualizace dat pomocí Python Spark**

Výstupní uzel pro Python Spark má nově záložku Output Graph, podobně jako pro prostředí jazyka R, umožňující vizualizaci dat pomocí jednoduchých skriptů.

- **Nová nastavení uzlů Auto Classifier/Numeric**

Byla přidána volba Calculate ensemble distribution graph určující zda (ne)budou tyto grafy zahrnuty v automatickém modelu. Defaultně je tato volba zakázána, což výrazně zrychluje tvorbu modelu. Zároveň byla defaultně zakázána volba Calculate predictor validity, což rovněž výrazně zefektivňuje auto-modelování.

- **Podpora nových operačních systémů a dalších technologií**

Podpora systémů RedHat9, macOS Sonoma 14.0 a SUSE 15. Kompatibilita s Oracle 21c, SQL Server 2022 a dB2 11.5.9. Podpora R 4.2.2, Apache Spark 3.4, Apache Hive 3.1.3, Apache Impala 4.0.0 a Python 3.10.7. Podpora hw Apple M1.

- **Rozšířené možnosti přístupu k souborům**

Schopnost přístupu k souborům uloženým v prostředí OneDrive. Podpora prohlížení souborů prostřednictvím protokolu WEBDAV.

