



Evropská unie  
Evropský sociální fond  
Operační program Zaměstnanost



**DRG**  
RESTART

# INSTALAČNÍ PŘÍRUČKA A MANUÁL PRO UŽIVATELE NÁSTROJE CZ-DRG PRE-GROUPER

---

Metodický materiál projektu DRG Restart

**Zpracoval autorský kolektiv pod vedením:** T. Pavlík, M. Bartůňková, P. Klika, J. Linda, L. Dušek

**Autoři:** P. Klika, K. Kupčák, T. Pavlík, Z. Bortlíček, M. Uher, J. Klika

Verze CZ-DRG: 7.0

Verze dokumentu: 2

Datum: 31.8. 2024



## Obsah

Účel nástroje.....	4
Předpoklady pro spuštění programu.....	4
Instalace .....	5
Struktura instalačního balíku.....	5
Nastavení pro spuštění v prostředí Linux.....	5
Spustitelné prostředí Javy .....	5
Nastavení dostupné paměti ve spouštěcím souboru.....	6
Instalace JDBC knihovny pro připojení do databáze .....	6
Vytvoření databázových tabulek.....	6
Konfigurace pre-grouperu.....	8
Parametry procesu .....	8
Zadání konfiguračních parametrů .....	10
Iniciální konfigurace .....	10
Konfigurační CSV soubory .....	10
Zadání konfiguračního parametru na příkazové řádce .....	10
Nastavení vstupů a výstupů pre-grouperu.....	10
Vstup z databáze .....	10
Vstup z textových souborů.....	11
Výstup do databáze.....	11
Výstup do textového souboru .....	12
Nastavení specifických parametrů pro JDBC.....	12
Spuštění pre-grouperu .....	13
Spuštění programu v iniciální konfiguraci .....	13
Spuštění s parametry definovanými v konfiguračním souboru .....	13
Spuštění s parametry definovanými na příkazové řádce .....	13
Předpřipravené spouštěcí skripty s ukázkovou konfigurací .....	14
Ukázka spuštění programu.....	14
Logování .....	15
Spuštění jako služba .....	15
Testování rychlosti pre-grouperu.....	17
Scénář „velká pojišťovna“ .....	18
Scénář „malá pojišťovna“ .....	18



Evropská unie  
Evropský sociální fond  
Operační program Zaměstnanost



**DRG**  
RESTART

Scénář „velká nemocnice“ .....	19
Scénář „malá nemocnice“ .....	19



## Účel nástroje

Program, označovaný jako pre-grouper, slouží k vytvoření podkladů pro klasifikační algoritmus systému CZ-DRG, tzn. vytvoření takzvaných vstupních datových vět grouperu (popsané v samostatném dokumentu DATOVÉ ROZHRANÍ NÁSTROJE CZ-DRG GROUPEU). Vstupem pre-grouperu jsou soubory nebo databázové tabulky (resp. pohledy) odpovídající datovému rozhraní pre-grouperu, které je popsáno v samostatné příloze DATOVÉ ROZHRANÍ NÁSTROJE CZ-DRG PRE-GROUPEU. Výstupem je soubor nebo databázová tabulka obsahující sestavené hospitalizační případy v podobě vstupních datových vět grouperu. Doplnující výstupy jsou rovněž popsány v datovém rozhraní pre-grouperu.

Pre-grouper implementuje postup popsáný v dokumentu METODIKA SESTAVENÍ HOSPITALIZAČNÍHO PŘÍPADU V SYSTÉMU CZ-DRG. Podrobně je algoritmus popsán v dokumentu TECHNICKÁ DOKUMENTACE NÁSTROJE CZ-DRG PRE-GROUPEU.

## Předpoklady pro spuštění programu

Nástroj pre-grouper je CLI (řádková) aplikace implementovaná v jazyku Java. Pre-grouper je tedy multiplatformní, přenositelný, snadno spravovatelný a bezpečný (za předpokladu pravidelné aktualizace systému, JRE a vlastního nástroje pre-grouper).

Základní požadavky na systém jsou shrnuty v následující tabulce:

Parametr / vlastnost	Hodnota	Poznámka
HW	Standardní HW umožňující běh aplikací v Javě, 128 MB volné RAM, 10 MB volného HDD pro instalaci + minimálně 4x velikost vstupních dat (v souboru CSV), alespoň 2 vlákna	Optimální parametry dle velikosti zpracovávaných dat, u velmi velkých souborů (pojištěn) je nutné i několik GB RAM
Operační systém	Standardní OS umožňující běh aplikací v Javě, Windows/UNIX/Linux	Ekvivalentní IR-DRG grouperu
Java	JRE (Java Runtime Environment) v <b>aktuální a podporované verzi - Java 8 (poslední free verze je 8u202), Java 11 nebo OpenJDK 11</b>	Ekvivalentní IR-DRG grouperu
Připojení k DB (volitelné)	Standardní JDBC knihovna pro použitou DB	Libovolná DB podporující JDBC rozhraní, např. MS SQL Server, Oracle, DB2, PostgreSQL, ...
Spuštění aplikace	Příkazová řádka OS (např. cmd, bash)	



## Instalace

Program je distribuován ve formě komprimovaného balíku (zip), který obsahuje programové balíky jazyka Java (jar), spouštěcí skripty (sh/bat), konfigurační soubory, číselníky a testovací data ve formátu CSV. Tento archiv je potřeba rozbalit do samostatného adresáře v cestě, ke které má uživatel právo pro čtení, zápis a spouštění programů.

### Struktura instalačního balíku

Název balíku ve formátu: **01\_\_CZ-DRG\_<verze\_programu>\_Pregrouper.zip**

např.: **01\_\_CZ-DRG\_v7\_r0\_Pregrouper.zip**

Seznam složek, které jsou součástí instalace:

- **ciselniky** – obsahuje aktuální podobu číselníků nutných pro běh pre-grouperu ve formě csv souborů, např. seznam kritických výkonů a zvlášť účtovaných položek
- **config** – obsahuje vzorové konfigurační soubory nástroje
- **db\_install** – vzorové DDL skripty pro vytvoření databázových tabulek a pohledů pro načítání a ukládání dat z/do databáze pro vybrané databázové servery
- **input** – složka pro ukládání vstupních CSV souborů dle DR pre-grouperu, obsahuje testovací soubory *pgp\_doklady02.csv*, *pgp\_doklady06.csv*, *pgp\_položky.csv*
- **lib** – obsahuje doplňující programové balíky nástroje bez knihoven JDBC pro připojení do databáze
- **log** – prázdná složka, slouží pro uložení logových souborů *pregrouper\_log.csv* a *error.log* (vytvoření této složky umožňuje zvýšit zabezpečení dat při rutinním provozu – hlavní složka pre-grouperu nemusí mít právo zápisu a u jednotlivých podsložek se může nastavit přístup pro zápis jen pro nezbytné složky, tzn. Input, log, output a temp)
- **output** – prázdná složka, slouží k uložení výstupního souboru
- **pgp\_service\_master** – obsahuje spouštěcí skripty (sh/bat), jar soubory procesu a xml soubor s nastavením kontextových proměnných - *default.properties* (je uložen na konci stromové struktury *pgp\_service\_master/pregrouper/pgp\_service\_master\_<verze\_programu>/contexts/*)
- **temp** – prázdná složka, slouží k uložení dočasných souborů

### Nastavení pro spuštění v prostředí Linux

Spouštěcím skriptům (\*.sh) je nutné přidat oprávnění pro spuštění. Obecně lze provést např. pomocí následujícího příkazu spuštěného v hlavním adresáři pre-grouperu (povolí spuštění nástroje všem uživatelům):

```
chmod -R a+x *.sh
```

### Spustitelné prostředí Javy

Od verze CZ-DRG 5.0 je podporováno spustitelné prostředí Javy (JRE) ve verzi 11 a zpětně i verzi 8 a to jak na oficiální distribuci Oracle, tak implementacích zdarma (OpenJDK). Distribuce Oracle těchto verzí již nejsou podporovány za bezplatných podmínek pro jiné než osobní použití. Doporučujeme tedy použití některé z volně dostupných distribucí JRE 11 nebo 8.



Distribuce řeší aktuální chybu knihovny log4j (viz <https://www.nukib.cz/cs/infoservis/hrozby/1781-upozorneni-na-zranitelnost-apache-log4j-log4shell/>). Je zahrnuta nová verze knihovny log4j a do spouštěcích skriptů je doplněn parametr `log4j2.formatMsgNoLookups=True`. Toto nastavení by mělo kritickou chybu eliminovat. Doporučujeme ověřit všechny spouštěcí skripty ve Vašem prostředí, zda obsahují uvedený parametr!

### Nastavení dostupné paměti ve spouštěcím souboru

Parametr `-Xmx` určuje maximální hodnotu alokované paměti pro data JVM (Java Virtual Machine). Hodnotu lze nastavit úpravou spouštěcího bat/sh souboru. Například hodnota `-Xmx10G` udává maximum možné paměti na 10 GB. Parametr je možné nastavit dle HW možností zařízení, na němž je program spouštěn, obvykle se nastavuje na polovinu až dvě třetiny dostupné operační paměti.

Parametr `-Xms` určuje iniciální hodnotu alokované paměti. Je doporučeno nastavit jeho hodnotu shodnou s parametrem `-Xmx`:

```
java -Xms1G -Xmx1G ... pregroup.pgp_service_master_0_1.pgp_service_master
...
```

Současně s tímto parametrem je nutné adekvátně upravit i maximální počet záznamů současně zpracovávaných v paměti (viz část **KONFIGURACE PRE-GROUPERU**). V případě některých databází je také třeba omezit počet záznamů, které se v jednom kroku načítají z databáze – viz část **NASTAVENÍ SPECIFICKÝCH PARAMETRŮ PRO JDBC**. V případě, kdy tyto parametry nejsou nastaveny odpovídajícími hodnotami, může dojít k pádu aplikace v důsledku nedostatku paměti. V takovém případě je nutné buď navýšit velikost dostupné paměti pro aplikaci, nebo omezit počet najednou zpracovávaných (resp. načítaných) záznamů.

### Instalace JDBC knihovny pro připojení do databáze

V případě čtení, resp. zápisu dat z/do databáze, je třeba instalovat do operačního systému příslušný JDBC ovladač odpovídající použitému typu databáze. V případě požadavku na využití některé databáze použijte ovladač dostupný na stránkách výrobce nebo dodavatele příslušné databáze. Můžete tento balík nahrát do standardní složky **lib** anebo do jiné cesty např. **libdb**, která bude při spuštění nástroje dostupná. Cestu k této mimo složku **lib** umístěné knihovně je potřeba doplnit do spouštěcího souboru sh/bat do parametru `-cp` (class search path):

```
java ... -cp ../libdb/mssql-jdbc-9.2.1.jre8.jar;../libdb/postgresql-42.2.21.jar;...
```

### Vytvoření databázových tabulek

Pro využití možnosti číst vstupní data z databáze je třeba v databázi vytvořit tabulky nebo pohledy, které odpovídají **DATOVÉMU ROZHRAŇÍ NÁSTROJE CZ-DRG PRE-GROUPER**. Rovněž pro výsledné záznamy, které mají být ukládány do databáze, je nutné vytvořit příslušnou tabulku odpovídající struktuře vstupní datové věty grouperu dle **DATOVÉHO ROZHRAŇÍ NÁSTROJE CZ-DRG GROUPER**, případně doplňujícím výstupním tabulkám datového rozhraní pre-grouperu.

Vytvořené tabulky musejí mít shodné názvy sloupců, pořadí sloupců a odpovídající datové typy (text, číslo, datum). Přesné datové typy jsou specifické pro konkrétní databázi. Vzorové příklady definice tabulek pro vybrané databáze jsou součástí distribuce pre-grouperu ve složce **db\_install**. V případě



využití jiné databáze může být nutné provést drobné úpravy datových typů, případně další syntaxe v definici tabulek.

Alternativně je možno definovat pouze pohledy (view) nad existujícími daty databáze, které transformují již uložená data do podoby, kterou vyžaduje pre-grouper. Výsledná datová struktura pohledů musí rovněž odpovídat struktuře datového rozhraní. Ve složce **db\_install** jsou vzorové pohledy, které načítají data z datového rozhraní Národního registru hrazených zdravotních služeb (NRHZS) a které mohou zdravotní pojišťovny použít pro testování funkce pre-grouperu nad svými daty.

Názvy tabulek a jejich umístění je možno konfigurovat v nastavení pre-grouperu (viz následující část [KONFIGURACE PRE-GROUPERU](#)).



## Konfigurace pre-grouperu

### Parametry procesu

Program pre-grouper je možno konfigurovat pomocí následujících parametrů:

Tabulka 1 Seznam konfiguračních parametrů nástroje pre-grouper

Název parametru	Popis
<b>buffer</b>	Udává maximální počet řádků současně zpracovávaných procesem v paměti při paměťově náročných operacích. Po dosažení nastaveného limitu jsou záznamy ukládány do dočasných souborů na disku. Parametr musí být přizpůsoben velikosti dostupné paměti pro JVM, nastavené pomocí parametru -Xmx. Přednastavená hodnota je určena na 250000 odpovídající parametru -Xmx 1GB RAM, při změně velikosti dostupné paměti by se měl přibližně ve stejném poměru upravit počet zpracovávaných záznamů.
<b>ciselniky</b>	Cesta (absolutní nebo relativní) ke složce s číselníky; musí být zakončena lomítkem
<b>context_file</b>	Určuje cestu (absolutní nebo relativní) a jméno externího souboru s nastavením konfiguračních proměnných (ve formě CSV s oddělovačem tabulátor)
<b>db_driver_class</b>	Název třídy s ovladačem pro danou databázi, například pro připojení do databáze PostgreSQL je to <i>org.postgresql.Driver</i> nebo <i>com.microsoft.sqlserver.jdbc.SQLServerDriver</i> pro připojení do databáze MS SQL – pro čtení vstupních dat z DB
<b>db_output</b>	Parametr určující, zda se má výstupní soubor uložit do databázové tabulky či nikoliv, povolené hodnoty jsou <i>true</i> nebo <i>false</i> . Přednastavená je hodnota <i>false</i> , která značí ukládání výstupu do souboru.
<b>db_password</b>	Heslo pro připojení do database – pro čtení vstupních dat z DB
<b>db_url</b>	JDBC url pro připojení do vybrané databáze (schématu) – viz část <a href="#">NASTAVENÍ SPECIFICKÝCH PARAMETRŮ PRO JDBC</a> – pro čtení vstupních dat z DB
<b>db_username</b>	Uživatelské jméno pro připojení do databáze – pro čtení vstupních dat z DB
<b>doklady02_db</b>	Název databázové tabulky nebo pohledu obsahující doklady typu 02 – pro čtení vstupních dat z DB
<b>doklady02_file</b>	Název vstupního souboru dokladů typu 02 včetně přípony – pro čtení vstupních dat ze souboru, načítá se ze složky určené parametrem <b>indir</b>
<b>dokladyOstatni_db</b>	Název databázové tabulky nebo pohledu obsahující doklady ostatních typů (06) – pro čtení vstupních dat z DB
<b>dokladyOstatni_file</b>	Název vstupního souboru dokladů ostatních typů (06) včetně přípony – pro čtení vstupních dat ze souboru, načítá se ze složky určené parametrem <b>indir</b>
<b>indir</b>	Cesta (absolutní nebo relativní) ke složce se vstupními soubory – pro čtení vstupních dat ze souboru; musí být zakončena lomítkem
<b>input_type</b>	Určení vstupu pro data, podporované hodnoty jsou buď <i>file</i> (čtení ze vstupního csv souboru), nebo <i>db</i> (čtení dat z databáze)





<b>outdir</b>	Cesta (absolutní nebo relativní) ke složce pro uložení výstupního souboru – pro čtení vstupních dat ze souboru; musí být zakončena lomítkem
<b>output_name_db</b>	Název databázové tabulky pro uložení výstupu z programu – pro uložení výsledných dat do DB
<b>output_name_file</b>	Název výstupního souboru programu – pro uložení výsledných dat do souboru, ukládá se do složky určené parametrem <b>outdir</b>
<b>polozky_db</b>	Název databázové tabulky nebo pohledu obsahující položky – pro čtení vstupních dat z DB
<b>polozky_file</b>	Název souboru položek včetně přípony – pro čtení vstupních dat ze souboru, načítá se ze složky určené parametrem <b>indir</b>
<b>tempdir</b>	Cesta (absolutní nebo relativní) ke složce pro uložení dočasných souborů pre-grouperu (význam jednotlivých temp výstupů je podrobně popsán v technické dokumentaci programu); musí být zakončena lomítkem
<b>temp_db</b>	Určuje, zda se mají dočasné soubory ukládat navíc do databáze (stejně jako vstupní/výstupní data). Možné hodnoty jsou „true“ nebo „false“
<b>doklady_spravne_napojene_db</b>	Tabulka, do které se zaznamenávají doklady přiřazené k případům
<b>terapeuticke_dny_db</b>	Seznam kalendářních dnů, kdy byl k hospitalizačnímu případu započítán uvedený terapeutický den
<b>doklady06_datum_vyrazene_db</b>	Doklady 06 nepřirazené k žádnému HP na základě data provedení prvního výkonu
<b>doklady06_extramuralni_vyrazene_db</b>	Doklady 06, které nebyly přiřazeny k danému HP jako extramurální péče
<b>doklady06_intramuralni_vyrazene_db</b>	Doklady 06, které nebyly přiřazeny k danému HP jako intramurální péče
<b>doklady06_nejasnePriraditelna_vyrazene_db</b>	Doklady 06, které nebyly přiřazeny k danému HP jako nejasně přiřaditelná péče
<b>runId</b>	ID spuštění, které se vloží na konec sestavených hospitalizačních případů do souboru vstupní datové věty grouperu při běhu pre-grouperu v dávkovém režimu (nikoliv jako služba). Případně zadané hodnoty parametru RUN_ID ve vstupních souborech jsou ignorovány. Při použití pre-grouperu v režimu služby se tento parametr ignoruje a vkládá se hodnota, předaná pomocí volání http požadavku (viz část <a href="#">SPUŠTĚNÍ JAKO SLUŽBA</a> ).
<b>tempdir_clear</b>	none – nesmaže žádný soubor z tempdir end – smaže soubory v tempdir stejně jako doposud (default hodnota při nevyplnění) progress - postupně promazává tempdir od souborů, které nejsou dále potřeba, oproti end promaže navíc i procesní soubory k0, které dosud zůstávaly i po ukončení programu Jakákoliv jiná hodnota se chová jako none.

Poznámky:



- kontextové parametry **indir**, **tempdir**, **outdir** a **ciselniky** udávající cestu ke složkám na disku musí končit lomítkem
- cesty v systémovém prostředí Windows lze zadat s využitím znaku lomítka (/)

## Zadání konfiguračních parametrů

Jednotlivé parametry je možno nastavit několika způsoby:

### Iniciální konfigurace

Základní konfigurace nástroje je popsána v konfiguračním souboru *Default.properties*, který je uložen v podadresáři složky **pgp\_service\_master** (pgp\_service\_master/pregrouper/pgp\_service\_master\_0\_1/contexts/). Obsahuje cesty ke složkám a souborům, které jsou součástí instalačního balíku. Tyto parametry jsou použity tehdy, pokud není použit žádný z dále uvedených způsobů jejich modifikace.

### Konfigurační CSV soubory

Konfigurační parametry je možné uložit do externího souboru typu CSV. Každý řádek souboru obsahuje vždy dvojici název parametru a jeho hodnota. Symbol pro oddělení dvojice je tabulátor (\t), každý řádek je ukončen znakem pro konec řádku (\n). Je možno vytvořit více konfiguračních souborů pro různá nastavení nástroje. Výběr konkrétního konfiguračního souboru je proveden při spuštění pre-grouperu (viz část **SPUŠTĚNÍ PRE-GROUPERU**).

Součástí instalace je několik ukázkových konfiguračních souborů, které demonstrují použití pre-grouperu s různými typy vstupů a výstupů.

### Zadání konfiguračního parametru na příkazové řádce

Poslední možností určení hodnoty konfiguračních parametrů je jejich zadání přímo v příkazové řádce při spuštění pre-grouperu. Takto je možno zadávat parametry, které se mění často, např. při každém spuštění nástroje (např. jméno vstupních souborů, pokud jsou vytvářeny vždy s novým jménem). Parametry zadané přímo na příkazové řádce mají přednost před hodnotami určenými některým z předešlých postupů.

### Nastavení vstupů a výstupů pre-grouperu

Proces je možné spouštět se vstupem z textových souborů typu csv anebo z databázových tabulek (resp. pohledů). Rovněž výstup může být uložen do souboru anebo do databáze. Typy vstupů a výstupů se mohou navzájem libovolně kombinovat (CSV -> CSV, CSV -> DB, DB -> CSV, DB -> DB). V závislosti na požadovaném vstupu a výstupu je třeba nastavit následující skupiny parametrů.

### Vstup z databáze

Je potřeba vyplnit všechny kontextové parametry pro připojení do databáze pomocí JDBC (viz sekce níže). Dále zadat parametry **doklady02\_db**, **doklady06\_db** a **polozky\_db** a nastavit parametr **input\_type** na hodnotu **db**. Proces nejprve načte data z databáze a uloží je do textových souborů ve složce **tempdir** a dále pokračuje shodně, jako v případě vstupu ze souboru (sám si vyplní hodnoty u parametrů s příponou **\_file** a parametru **indir**).



Tabulka 2 Ukázka konfigurace vstupu z DB

Název parametru	Hodnota
<b>db_driver_class</b>	com.microsoft.sqlserver.jdbc.SQLServerDriver
<b>db_url</b>	jdbc:sqlserver://localhost:1433;databaseName=pregrouper; authentication=SqlPassword;trustServerCertificate=true
<b>db_username</b>	pregrouper
<b>db_password</b>	pregrouper
<b>input_type</b>	db
<b>doklady02_db</b>	pgp_doklady02
<b>dokladyOstatni_db</b>	pgp_doklady06
<b>polozky_db</b>	pgp_polozky

### Vstup z textových souborů

Není nutné vyplňovat parametry pro připojení do databáze pomocí JDBC a parametry s příponou **\_db**. Je nutné nastavit parametr **indir** pro cestu ke složce se vstupními soubory a parametry **doklady02\_file**, **doklady06\_file** a **polozky\_file** určující názvy vstupních souborů pro doklady a položky. Parametr **input\_type** musí mít hodnotu **file**.

Tabulka 3 Ukázka konfigurace vstupu z CSV

Název parametru	Hodnota
<b>indir</b>	../input
<b>input_type</b>	file
<b>doklady02_file</b>	pgp_doklady02.csv
<b>dokladyOstatni_file</b>	pgp_doklady06.csv
<b>polozky_file</b>	pgp_polozky.csv

### Výstup do databáze

Vygenerované vstupní datové věty grouperu je možné uložit do databáze nastavením parametru **db\_output** na hodnotu **true**. Je nutné vyplnit parametry připojení přes JDBC (viz níže), parametry pro uložení výstupního csv souboru **outdir**, **output\_name\_file** a parametr **output\_name\_db** určující název databázové tabulky. Dále lze (volitelně) nastavit hodnotu **temp\_db** na **true**, která způsobí ukládání vybraných dočasných výstupů navíc do zadaných databázových tabulek.

Tabulka 4 Ukázka konfigurace výstupu do DB

Název parametru	Hodnota
<b>db_output</b>	true
<b>outdir</b>	../output/
<b>output_name_file</b>	vstupni_veta.csv
<b>output_name_db</b>	czdrg_vstupni_veta
<b>temp_db</b>	true
<b>doklady_spravne_napojene_db</b>	pgp_doklady_pripadu
<b>terapeuticke_dny_db</b>	pgp_terapeuticke_dny
<b>doklady06_datum_vyrazene_db</b>	pgp_doklady06_vyrazene_dle_data



doklady06_extramuralni_vyrazene_db	pgp_doklady06_vyrazene_extramuralni
doklady06_intramuralni_vyrazene_db	pgp_doklady06_vyrazene_intramuralni
doklady06_nejasnePriraditelna_vyrazene_db	pgp_doklady06_vyrazene_nejasnePriraditelna

### Výstup do textového souboru

Není nutné vyplňovat parametry pro připojení do databáze pomocí JDBC a parametry s příponou **\_db**. Je nutné nastavit parametr **outdir** pro cestu ke složce s výstupním souborem a parametr **output\_name\_file** určující název tohoto souboru. Parametr **db\_output** musí mít hodnotu *false*.

Tabulka 5 Ukázka konfigurace výstupu do CSV

Název parametru	Hodnota
db_output	false
outdir	../output
output_name_file	vstupni_veta.csv

### Nastavení specifických parametrů pro JDBC

Základní nastavení připojení do databáze se provádí pomocí dvou konfiguračních parametrů - **db\_url** a **db\_driver\_class**. Parametr **db\_url** obsahuje zejména specifikaci typu databáze, cestu k serveru a číslo portu, případně název databáze. V závislosti na typu připojované databáze může být nutné doplnit specifické parametry – například pro definici schématu (searchpath, currentSchema) nebo počtu přenášených záznamů (defaultRowFetchSize). Parametry jsou obvykle oddělovány středníkem (;) a mezi názvem a hodnotou parametru je znak = (rovnítko).

Příklad vyplnění parametru **db\_url** pro napojení do databáze PostgreSQL:

```
jdbc:postgresql://localhost:5432/db_name;currentSchema=schema_name;
defaultRowFetchSize=1000
```

Např. v tomto případě byl omezen počet záznamů načítaných v jednom bloku parametrem defaultRowFetchSize, aby nedošlo k vyčerpání dostupné paměti nástroje pre-grouper.

Další ukázka zobrazuje vzorové nastavení parametru pro připojení do databáze MS SQL Server:

```
jdbc:sqlserver://localhost:1433;databaseName=db_name;
authentication=SqlPassword;trustServerCertificate=true
```

Přesný popis jednotlivých parametrů pro připojení ke konkrétní databázi najdete v dokumentaci každé konkrétní databáze.



## Spuštění pre-grouperu

Pre-grouper se použít z příkazové řádky. Hlavní spouštěcí skripty jsou umístěny ve složce **pgp\_service\_master**, dle operačního systému se použít varianta **pgp\_service\_master\_run.sh** (UNIX/Linux) nebo **pgp\_service\_master\_run.bat** (Windows). V kořenové složce nástroje pre-grouper jsou dále vytvořeny pomocné spouštěcí skripty, které usnadňují spuštění nástroje pro vybrané varianty vzorové konfigurace (určené pomocí konfiguračních CSV souborů).

### Spuštění programu v iniciální konfiguraci

Pokud při spuštění nástroje nejsou uvedeny žádné konfigurační parametry, použijí se hodnoty ze souboru *default\_install.properties*, které jsou nastaveny pro ukázkový běh nad přiloženými testovacími daty. Spouští se skripty ve složce **pgp\_service\_master**:

- Linux

```
./pgp_service_master_run.sh
```

- Windows

```
pgp_service_master_run.bat
```

### Spuštění s parametry definovanými v konfiguračním souboru

Pro spuštění pre-grouperu s parametry, které byly určeny v konfiguračním CSV souboru, je třeba spustit spouštěcí skript s parametrem *context\_file*, kam se zadá cesta k tomuto souboru. Cesta může být absolutní nebo relativní. Parametr je zadáván ve tvaru:

```
--context_param context_file="cesta_a_nazev_souboru"
```

Např.:

- Linux

```
./pgp_service_master_run.sh --context_param context_file="../config/file.csv"
```

- Windows

```
pgp_service_master_run.bat --context_param context_file="C:/pregrouper/config/MSSQL.csv"
```

### Spuštění s parametry definovanými na příkazové řádce

Hodnoty kontextových parametrů lze měnit při spuštění programu zadáním v příkazové řádce. Přepisují se defaultní hodnoty nastavené některým z předchozích postupů. Každý parametr je zadáván ve tvaru:

```
--context_param nazev_parametru="hodnota_parametru"
```

Jednotlivé parametry se zapisují za sebe, např. (Linux):



```
./pgp_service_master_run.sh --context_param indir=" ../indir/" --context_param outdir=" ../outdir/" --  
context_param doklady02_file="doklady02.csv" --context_param  
dokladyOstatni_file="doklady06.csv" --context_param polozky_file="polozky.csv" --context_param  
input_type="file" --context_param db_output="false" --context_param  
output_name_file="output.csv"
```

### Předpřipravené spouštěcí skripty s ukázkovou konfigurací

V kořenovém adresáři nástroje pre-grouper jsou přiloženy spouštěcí skripty, které se spouštějí s předvybranou vzorovou konfigurací. Jsou zahrnuty ukázky pro načítání dat ze vzorových CSV souborů a vybrané databázové technologie, vždy ve variantě Windows (.bat) a UNIX/Linux (.sh). Tyto skripty ukazují možnost snadného spuštění často používaných konfigurací např.:

- Linux (pro čtení ze souborů)

`./pregrouper-file.sh`

- Windows (pro čtení z MS SQL databáze)

`pregrouper-mssql.bat`

### Ukázka spuštění programu

Obrázek 1 Ukázka spuštění CZ-DRG pre-grouperu v konfiguraci pro čtení ze souboru na Linuxu

```
klikap@clegane: /opt/CZ-DRG-Pregrouper-1.0
klikap@clegane:/opt/CZ-DRG-Pregrouper-1.0$ ./pregrouper-file.sh
CZ-DRG Pre-grouper 1.0
(C) 2018, ÚZIS ČR
Sestavení HP uspesne ukonceno.
klikap@clegane:/opt/CZ-DRG-Pregrouper-1.0$
```



## Logování

Každé spuštění pre-grouperu je zaznamenáno v souboru `pregrouper_log.csv`, umístěném v podsložce `log`. Je zde uveden jeden záznam při začátku spuštění procesu `pgp_service_master` a jeden na jeho konci, s časovou značkou, interním ID spuštění, názvem procesu (`pgp_service_master`), jeho verzí a stavem (`begin/end`). Koncový záznam je navíc doplněn o údaje o výsledku (`success/failure`) a době trvání. Ukázka souboru je uvedena v následující tabulce:

Tabulka 6 Ukázka záznamů o spuštění pre-grouperu

Datum a čas	Id spuštění	Proces	Verze	Stav	Výsledek	Trvání
2021-12-04 10:06:45	AacTtb	CZ-DRG Pregrouper 5.0 – 5r1-211210	0.1	begin		
2021-12-04 10:08:40	AacTtb	CZ- DRG Pregrouper 5.0 – 5r1-211210	0.1	end	success	13,607 sec

V případě výskytu neočekávané chyby a pádu aplikace je chybový výstup uložen do souboru `error.log` rovněž umístěném v podsložce `log`. Obsahuje chybové hlášení Java JRE o nastalé chybě (např. při nedostatku paměti).

## Spuštění jako služba

Nástroj je také možné spustit jako službu, která umožní rychlé spouštění sestavení hospitalizačních případů ze záznamů, vložených do databáze. Spuštěná služba naslouchá na zadaném portu a očekává na něm HTTP POST request s identifikátorem spuštění pre-grouperu (`RUN_ID`) v jeho těle. Pro připojení k databázi se vstupními a výstupními záznamy je používané nastavení sekce pro čtení a zápis z/do databáze v konfiguračním souboru. Čtení a ukládání záznamů do souboru není v tomto režimu podporováno.

Službu lze spustit například následujícím příkazem:

```
„pgp_service_master/pgp_service_master_run.bat --context_param service=true
--context_param http_port=8888 --context_param timer=10
--context_param context_file=\"../config/MSSQL.csv““
```

Ten spustí pre-grouper na daném počítači jako službu poslouchající na portu 8888 a čte/ukládá data z/do MSSQL databáze. Při spuštění pre-grouperu jako služby lze použít následující parametry v příkazové řádce:

Název parametru	Popis	Defaultní hodnota
<b>service</b>	Indikátor, zda má být spuštěn pre-grouper v režimu služby ( <code>true/false</code> )	<code>false</code>
<b>http_host</b>	Adresa serveru, na kterém bude služba pre-grouperu očekávat http požadavky	<code>localhost</code>
<b>http_port</b>	Číslo portu, na kterém bude služba pre-grouperu očekávat http požadavky	<code>8080</code>
<b>timer</b>	Interval v milisekundách mezi kontrolami na příchozí požadavky	<code>1</code>



Pro snadnější práci s pre-grouperem spuštěným v režimu služby jsou v kořenovém adresáři nástroje vytvořeny pomocné .bat a .sh skripty, které ukazují typické použití, např. **pregrouper-mssql-service-port.bat**.

Pro vytvoření HTTP POST requestu lze použít příkazovou řádku (Linux nebo Windows) s následujícím příkazem:

```
curl -i -X POST -d 1 http://localhost:8888/
```

Zde je "1" id konkrétního spuštění nástroje pre-grouper nad připravenými daty, kdy budou zpracovány pouze záznamy ze vstupních tabulek v databázi, které mají uvedenou tuto hodnotu ve sloupci RUN\_ID. Zadané záznamy musí již v tabulce existovat (mohou v ní být i další záznamy, které ale nebudou zpracovány).

Text "http://localhost:8888/" je URL adresa, kam se má požadavek odeslat, tedy adresa, kde běží spuštěná aplikace pre-grouperu v režimu služby. Odpovědí na tento požadavek je buď stavový HTTP kód 200, který reprezentuje průběh bez chyb, anebo kód 404, reprezentující prázdný výsledek sestavení (typicky pro zadané RUN\_ID nebyly nalezeny žádné záznamy). Při výskytu závažného problému při běhu programu (např. nepodaří se načíst/uložit data z/do DB) vrací výsledek 500. V případě úspěšného sestavení jsou záznamy uloženy do tabulky vstupní datové věty grouperu dle konfigurace nástroje.





## Testování rychlosti pre-grouperu

Uvedené hodnoty se týkají pre-grouperu verze 1.0, pro verzi 2.0 jsou údaje velice podobné. V rámci vývoje pre-grouperu byla provedena řada testovacích běhů nad testovacími daty Národního registru hrazených zdravotních služeb (NRHZS). Byla vybrána data dvou pojišťoven a dvou nemocnic, vždy jedna velká a jedna malá. Byly vybírány doklady za jeden kalendářní rok (podle roku dávky). V případě testování dat pojišťovny vždy doklady dané pojišťovny, v případě testování dat nemocnice doklady všech pojišťoven, které mají v hlavičce dávky uvedeno IČZ dané nemocnice (tj. v tomto případě není započítána extramurální péče). Byly vybírány pouze doklady 02 a 06 (všechny) a do položek všechny výkony, hromadně a individuálně vyráběná léčiva a zdravotnický materiál.

Testováno bylo na virtuálním serveru s procesory Intel(R) Xeon(R) CPU E5-2620 v4 @ 2.10GHz, který měl k dispozici až 24 GB RAM a 8 jader. Operační systém je Ubuntu Linux 16.04 a virtualizační platforma KVM. Testovací konfigurace byla nastavována na různé hodnoty dostupných prostředků, aby byla ověřena použitelnost nástroje pro různé účely. Na serveru běžela v danou chvíli pouze testovaná úloha, ale výsledky mohou být ovlivněny nastavením virtualizace nebo systémovými procesy operačních systémů. Data byla vždy načítána ze souboru.

Následující sekce zahrnují souhrn údajů o vybraných testovacích datech a výsledky z testování nad vybranými konfiguracemi. Na základě těchto výsledků je možné nastavit optimální konfiguraci pro jednotlivé uživatele tohoto nástroje a odhadnout časovou náročnost sestavení hospitalizačních případů.



## Scénář „velká pojišťovna“

Tabulka 7 Zpracovávaná data scénáře "velká pojišťovna"

Ukazatel	Hodnota
Počet dokladů 02	1 907 503
Počet dokladů 06	44 110 701
Počet položek	212 051 247
Celková velikost vstupních dat (MB)	12 510
Výsledný počet sestavených HP	1 197 611

Tabulka 8 Výsledky testování scénáře "velká pojišťovna"

Počet jader	Dostupná paměť	Počet v jednom bloku zpracovávaných záznamů	Čas trvání (min)
8	16 GB	4 000 000	129,1
8	8 GB	2 000 000	137,6
8	4 GB	1 000 000	155,9
4	4 GB	1 000 000	170,7

## Scénář „malá pojišťovna“

Tabulka 9 Zpracovávaná data scénáře "malá pojišťovna"

Ukazatel	Hodnota
Počet dokladů 02	118 822
Počet dokladů 06	3 043 646
Počet položek	14 354 580
Celková velikost vstupních dat (MB)	872
Výsledný počet sestavených HP	81 080

Tabulka 10 Výsledky testování scénáře "malá pojišťovna"

Počet jader	Dostupná paměť	Počet v jednom bloku zpracovávaných záznamů	Čas trvání (s)
8	8 GB	2 000 000	492
8	4 GB	1 000 000	508
8	1 GB	250 000	523
4	1 GB	250 000	594



## Scénář „velká nemocnice“

Tabulka 11 Zpracovávaná data scénáře "velká nemocnice"

Ukazatel	Hodnota
Počet dokladů 02	97 131
Počet dokladů 06	1 445 906
Počet položek	6 557 737
Celková velikost vstupních dat (MB)	408
Výsledný počet sestavených HP	70 693

Tabulka 12 Výsledky testování scénáře "velká nemocnice"

Počet jader	Dostupná paměť	Počet v jednom bloku zpracovávaných záznamů	Čas trvání (s)
8	4 GB	1 000 000	324
8	1 GB	250 000	361
8	256 MB	50 000	417
4	256 MB	50 000	491

## Scénář „malá nemocnice“

Tabulka 13 Zpracovávaná data scénáře "malá nemocnice"

Ukazatel	Hodnota
Počet dokladů 02	8 079
Počet dokladů 06	94 520
Počet položek	584 193
Celková velikost vstupních dat (MB)	34
Výsledný počet sestavených HP	6 343

Tabulka 14 Výsledky testování scénáře "malá nemocnice"

Počet jader	Dostupná paměť	Počet v jednom bloku zpracovávaných záznamů	Čas trvání (s)
8	4 GB	1 000 000	23
8	1 GB	250 000	25
8	128 MB	50 000	33
4	128 MB	50 000	39