



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

Metodický materiál systému CZ-DRG

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 2027

Verze dokumentu: 1

Datum: 15. 4. 2026

Obsah

Účel nástroje.....	3
Předpoklady pro spuštění programu.....	3
Instalace	4
Struktura instalačního balíku.....	4
Nastavení pro spuštění v prostředí Linux.....	4
Spustitelné prostředí Javy	5
Nastavení dostupné paměti ve spouštěcím souboru.....	5
Instalace JDBC knihovny pro připojení do databáze	5
Vytvoření databázových tabulek.....	5
Konfigurace pre-grouperu	7
Parametry procesu Sekce <program>	7
Sekce <databaseInput>	10
Sekce <databaseOutput>	10
Zadání konfiguračních parametrů	10
Iniciální konfigurace	11
Konfigurační CSV/xml soubory.....	11
Zadání konfiguračního parametru na příkazové řádce	11
Nastavení vstupů a výstupů pre-grouperu.....	11
Vstup z databáze	11
Vstup z textových souborů	12
Výstup do databáze	12
Výstup do textového souboru	12
Nastavení specifických parametrů pro JDBC.....	13
Spuštění pre-grouperu	14
Spuštění programu v iniciální konfiguraci	14
Nastavení dostupné paměti ve spouštěcím souboru.....	14
Spuštění s parametry definovanými v konfiguračním souboru	14
Spuštění s parametry definovanými na příkazové řádce	15
Předpřipravené spouštěcí skripty s ukázkovou konfigurací	15
Ukázka spuštění programu.....	16
Logování	17
Spuštění jako služba	18
Testování rychlosti pre-grouperu.....	19
Scénář „velká pojišťovna“.....	20

Scénář „malá pojišťovna“	20
Scénář „velká nemocnice“	21
Scénář „malá nemocnice“	21

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 ROZHRAŇÍ 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 ROZHRAŇÍ 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, 100 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 (pojišťoven) 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 aktuální a podporované verzi - Java 8 (poslední free verze je 8u202), Java 117 nebo OpenJDK 117 a vyšší	Ekvivalentní IR-DRG grouperu !!! POZOR !!! Je nutné přejít na verzi Java 17 s podporou LTS (Long Time Support) !!!
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_20267_r0_Pregrouper.zip**

Seznam složek, které jsou součástí instalace:

- **.idea** – obsahuje pomocné xml soubory
- **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_polozky.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.log~~ *log.txt* (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/*~~
- **settings** – obsahuje vzorový konfigurační soubor nástroje
- **talend** – prázdná složka, do které je možné pro porovnání shody výsledků nahrát výstupní soubor vytvořený s využitím dříve distribuované verze pre-grouperu vytvořený s využitím programovacího prostředí Talend (nyní distribuovaný program je přímo vytvořený v jazyku Java verze 17)
- **temp** – prázdná složka, slouží k uložení ~~dočasných~~ pomocných výstupních 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 vyšší verze.~~

~~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`, např. pro 4 GB paměti:~~

```
java -Xms14G -Xmx14G ...  
pregrouper.pgp_service_master_0_1.pgp_service_master CZ-DRG-Pregrouper-  
2027.jar cz.drg.pregrouper.Main ...
```

~~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

Program Pregrouper umožňuje nakonfigurovat množství hodnot pro každého uživatele jako je připojení k databázi, cestu k vstupním/výstupním csv souborům nebo databázovým tabulkám a další.

Proces Pregrouperu může být parametrizován pomocí řady konfiguračních parametrů, které jsou standardně uloženy v souboru **settings/programSettings.xml**. Jde o XML dokument, který je členěn do několika částí a popisuje základní chování nástroje, jeho vstupy a výstupy. Hodnoty parametrů se zapisují mezi zahajovací a ukončovací „tag“ příslušného elementu dle XML syntaxe:

`<nazevElementu>konfiguracni_hodnota</nazevElementu>`

Tedy např.:

`<csvInputDoklady02>input/pgp_doklady02.csv</csvInputDoklady02>`

Parametry procesu Sekce <program>

Program je možno konfigurovat pomocí následujících parametrů:

Tabulka 1 Seznam konfiguračních parametrů nástroje pre-grouper

Název parametru	Popis
buffer	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.
ciselniky	Cesta (absolutní nebo relativní) ke složce s číselníky; musí být zakončena lomítkem
context_file	Určuje cestu (absolutní nebo relativní) a jméno externího souboru s nastavením konfiguračních proměnných (ve formě CSV s oddělovačem tabulátor)
db_driver_class	Název třídy s ovladačem pro danou databázi, například pro připojení do databáze PostgreSQL je to <code>org.postgresql.Driver</code> nebo <code>com.microsoft.sqlserver.jdbc.SQLServerDriver</code> pro připojení do databáze MS-SQL – pro čtení vstupních dat z DB
db_output	Parametr určující, zda se má výstupní soubor uložit do databázové tabulky či nikoliv, povolené hodnoty jsou <code>true</code> nebo <code>false</code> . Přednastavená je hodnota <code>false</code> , která značí ukládání výstupu do souboru.
db_password	Heslo pro připojení do database – pro čtení vstupních dat z DB
db_url	JDBC url pro připojení do vybrané databáze (schématu) – viz část NASTAVENÍ SPECIFICKÝCH PARAMETRŮ PRO JDBC – pro čtení vstupních dat z DB
db_username	Uživatelské jméno pro připojení do databáze – pro čtení vstupních dat z DB
doklady02_db	Název databázové tabulky nebo pohledu obsahující doklady typu 02 – pro čtení vstupních dat z DB



doklady02_file	Název vstupního souboru dokladů typu 02 včetně přípony — pro čtení vstupních dat ze souboru, načítá se ze složky určené parametrem indir
dokladyOstatni_db	Název databázové tabulky nebo pohledu obsahující doklady ostatních typů (06) — pro čtení vstupních dat z DB
dokladyOstatni_file	Název vstupního souboru dokladů ostatních typů (06) včetně přípony — pro čtení vstupních dat ze souboru, načítá se ze složky určené parametrem indir
indir	Cesta (absolutní nebo relativní) ke složce se vstupními soubory — pro čtení vstupních dat ze souboru; musí být zakončena lomítkem
input_type	Určení vstupu pro data, podporované hodnoty jsou buď file (čtení ze vstupního csv souboru), nebo db (čtení dat z databáze)
outdir	Cesta (absolutní nebo relativní) ke složce pro uložení výstupního souboru — pro čtení vstupních dat ze souboru; musí být zakončena lomítkem
output_name_db	Název databázové tabulky pro uložení výstupu z programu — pro uložení výsledných dat do DB
output_name_file	Název výstupního souboru programu — pro uložení výsledných dat do souboru, ukládá se do složky určené parametrem outdir
polozky_db	Název databázové tabulky nebo pohledu obsahující položky — pro čtení vstupních dat z DB
polozky_file	Název souboru položek včetně přípony — pro čtení vstupních dat ze souboru, načítá se ze složky určené parametrem indir
tempdir	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
temp_db	Určuje, zda se mají dočasné soubory ukládat navíc do databáze (stejně jako vstupní/výstupní data). Možné hodnoty jsou „true“ nebo „false“
doklady_spravne_napojene_db	Tabulka, do které se zaznamenávají doklady přiřazené k případům
terapeuticke_dny_db	Seznam kalendářních dnů, kdy byl k hospitalizačnímu případu započítán uvedený terapeutický den
doklady06_datum_vyrazene_db	Doklady 06 nepřijížené k žádnému HP na základě data provedení prvního výkonu
doklady06_extramuralni_vyrazene_db	Doklady 06, které nebyly přiřazeny k danému HP jako extramurální péče
doklady06_intramuralni_vyrazene_db	Doklady 06, které nebyly přiřazeny k danému HP jako intramurální péče
doklady06_nejasnePriraditelna_vyrazene_db	Doklady 06, které nebyly přiřazeny k danému HP jako nejasně přiřaditelná péče
doklady_mimo_hp_db	Identifikace dokladů, na kterých je některá péče provedená v den, který nespadá do doby trvání hospitalizačního případu.
runId	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 SPUŠTĚNÍ JAKO SLUŽBA).

tempdir_clear	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 (/)

<csvInputDoklady02>

Cesta ke vstupnímu souboru s doklady 02 v případě puštění bez parametru **-input**. V případě nevyplnění hodnoty v konfiguračním souboru se použije defaultní hodnota **./input/pgp_doklady02.csv**.

<csvInputDoklady06>

Cesta ke vstupnímu souboru s doklady 06 v případě puštění bez parametru **-input**. V případě nevyplnění hodnoty v konfiguračním souboru se použije defaultní hodnota **./input/pgp_doklady06.csv**.

<csvInputPolozky>

Cesta ke vstupnímu souboru s doklady 02 v případě puštění bez parametru **-input**. V případě nevyplnění hodnoty v konfiguračním souboru se použije defaultní hodnota **./input/pgp_polozky.csv**.

<csvOutput>

Cesta pro cílový výstupní soubor v případě puštění bez parametru **-output**. V případě nevyplnění hodnoty v konfiguračním souboru se použije defaultní hodnota **./output/vstupni_veta.csv**.

<talendInput>

Cesta ke vstupnímu souboru s výstupními hodnotami z původního programu Talend pro porovnání výstupů obou programů v případě puštění s parametrem **-compareOutput**. V případě nevyplnění hodnoty v konfiguračním souboru se použije defaultní hodnota **./talend/vstupni_veta.csv**.

<tempFolder>

Cesta pro cílovou složku pro sekundární výstupy z programu. V případě nevyplnění hodnoty v konfiguračním souboru se použije defaultní hodnota **./temp**.

<evaluationBatchSize>

Nastavení počtu naráz zpracovávaných unikátních pojištěnců v jedné dávce dat rozdělené mezi paralelní vlákna/workery programu. Nevhodné nastavení v poměru k poskytnuté paměti pro běh programu může způsobit nestabilitu až nemožnost běhu programu. V případě nevyplnění hodnoty v konfiguračním souboru se použije defaultní hodnota **2000**.

Bezpečný poměr poskytnuté paměti a velikosti jedné dávky je kolem 2GB : 1000, kde vyšší počet workerů může vyžadovat úpravu této hodnoty.

<evaluationWorkerCount>

Nastavení počtu paralelních vláken/workeru programu. V kombinaci s <evaluationBatchSize> určuje celkovou rychlost zpracování dat. V případě nevyplnění hodnoty v konfiguračním souboru se použije defaultní hodnota **maximálního počtu jader nalezených běhovým prostředím Java**.

Doporučené množství workerů je zhruba počet jader procesoru nad kterým program běží.

Sekce <databaseInput>

Program je možno konfigurovat pomocí následujících parametrů:

<databaseConnection>

Povinné nastavení připojení pro cílový databázový server při spuštění programu s parametrem **-input database**.

<databaseUser>

Nastavení uživatelského jména pro přihlášení do cílového serveru.

<databasePassword>

Nastavení uživatelského hesla pro přihlášení do cílového serveru.

Nastavení vstupních databázových tabulek

Konfigurace	Default hodnota při nevyplnění
<doklady02Table>	pgp_doklady02
<doklady06Table>	pgp_doklady06
<polozkyTable>	pgp_polozky

Sekce <databaseOutput>

Program je možno konfigurovat pomocí následujících parametrů:

<databaseConnection>

Povinné nastavení připojení pro cílový databázový server při spuštění programu s parametrem **-output database**.

<databaseUser>

Nastavení uživatelského jména pro přihlášení do cílového serveru.

<databasePassword>

Nastavení uživatelského hesla pro přihlášení do cílového serveru.

Nastavení cílových databázových tabulek

Konfigurace	Default hodnota při nevyplnění
<outputTable>	czdrg_vstupni_veta
<vyrazene06DatumTable>	pgp_doklady06_vyrazene_dle_data
<vyrazene06ExtramuralniTable>	pgp_doklady06_vyrazene_extramuralni
<vyrazene06IntramuralniTable>	pgp_doklady06_vyrazene_intramuralni
<vyrazene06NejasneTable>	pgp_doklady06_vyrazene_nejasnepriraditelna
<dokladyPripaduTable>	pgp_doklady_pripadu
<terapeutickeDnyTable>	pgp_terapeuticke_dny

Zadání konfiguračních parametrů

Jednotlivé parametry je možno nastavit **několikadvěma** 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 CSVxml soubory

Konfigurační parametry je možné uložit do externího souboru typu **CSVxml**. 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 případně pro různé uživatele. Výběr konkrétního konfiguračního souboru je proveden při spuštění Pregrouperu (viz část **SPUŠTĚNÍ PRE-GROUPERU**).

Součástí instalace je několik jeden ukázkových konfiguračních souborů, které demonstrují použití pre-grouperu s různými typy vstupů a výstupů ve složce **./settings/settings.xml**.

Zadání konfiguračního parametru na příkazové řádce

Další možností určení hodnoty konfiguračních parametrů je jejich zadání přímo v příkazové řádce při spuštění Pregrouperu. 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 v konfiguračním souboru.

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
db_driver_class	com.microsoft.sqlserver.jdbc.SQLServerDriver
db_url	jdbc:sqlserver://localhost:1433;databaseName=pregrouper; authentication=SqlPassword;trustServerCertificate=true
db_username	pregrouper
db_password	pregrouper
input_type	db
doklady02_db	pgp_doklady02
dokladyOstatni_db	pgp_doklady06
polozky_db	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
indir	../input
input_type	file
doklady02_file	pgp_doklady02.csv
dokladyOstatni_file	pgp_doklady06.csv
polozky_file	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
db_output	true
outdir	../output/
output_name_file	vstupni_veta.csv
output_name_db	czdrg_vstupni_veta
temp_db	true
doklady_spravne_napojene_db	pgp_doklady_pripadu
terapeuticke_dny_db	pgp_terapeuticke_dny
doklady06_datum_vyrazene_db	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** ~~<databaseConnection>~~ v souboru settings/settings.xml 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
```

Příklad pro lokální mysql server:

```
jdbc:mysql://localhost:3306/drg_pregrouper
```

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

Spuštění pre-grouperu

Pro spuštění programu Pregrouper je nutné mít běhové prostředí Javy verze 17+. Nižší verze nejsou s programem kompatibilní. Pregrouper se použít z příkazové řádky. Hlavní spouštěcí skripty jsou umístěny ve kořenové složce ~~pgp_service_master~~pre-grouperu, dle operačního systému se použít varianta ~~pgp_service_master_runpregrouper.sh~~ (UNIX/Linux) nebo ~~pgp_service_master_runpregrouper.bat~~ (Windows). V kořenové složce nástroje pre-grouper jsou dále vytvořeny pomocné spouštěcí skripty, které ~~usnadňují~~umožňují spuštění nástroje pro vybrané varianty ~~vzorové konfigurace~~ (určené pomocí konfiguračních CSV souborů CSV anebo databáze).

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~~

Nastavení dostupné paměti ve spouštěcím souboru

Program lze spustit s pomocí standardního příkazu pro spuštění java .jar souborů s množstvím poskytnuté RAM paměti a cestou k používaným knihovnám. 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, např. základním příkazem pro spuštění se 4 GB paměti:

~~java -Xms4G -Xmx4G -cp lib/*;CZ-DRG-Pregrouper-2027.jar
cz.drg.pregrouper.Main~~

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:~~

Program umožňuje zadat alternativní konfigurační soubor, např. pro jednotlivé uživatele. Pokud parametr není uveden, konfigurace se načítá ze standardního konfiguračního souboru ./settings/programSettings.xml.

~~java ... - configFile ./settings/programSettings_novak.xml ...
-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_paramindir="../indir/" --context_paramoutdir="../outdir/" --
context_paramdoklady02_file="doklady02.csv" --context_param
dokladyOstatni_file="doklady06.csv" --context_parampolozky_file="polozky.csv" --context_param
input_type="file" --context_paramdb_output="false" --context_param
output_name_file="output.csv"~~

~~Program podporuje několik argumentů pro nastavení jeho běhu, které se přidávají na konec základního spouštěcího příkazu.~~

- ~~• **-input <typ_vstupu>** - <typ_vstupu> může být „csv“ anebo „database“, pro určení odkud má program brát vstupní data, při nevyplnění se použije „csv“~~
- ~~• **-output <typ_výstupu>** - <typ_výstupu> může být „csv“ anebo „database“, pro určení kam má program zapisovat výstupní data, při nevyplnění se použije „csv“~~
- ~~• **-configFile <jméno_souboru>** – pokud parametr není uveden, konfigurace se načítá ze standardního konfiguračního souboru settings/programSettings.xml~~
- ~~• **-service <port>** – použije se pro spuštění programu jako služba, ten pak nenačítá hromadně vstupní/výstupní data, ale otevře port, kde očekává http POST požadavek s číslem run_id v těle požadavku pro určení dat z databáze nad kterými se program spustí, při nevyplnění portu se použije 8888~~
- ~~• **-compareOutput** – použije se pro zapnutí porovnání výstupu programu s poskytnutým výstupem z Talendu, cesta k souboru je definovaná v konfiguračním souboru~~

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~~

~~pregrouper-db.bat/sh~~

~~pregrouper-db-to-file.bat/sh~~



pregrouper-file.bat/sh

pregrouper-file-to-db.bat/sh

pregrouper-file-settings.bat/sh

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, UZIS CR
Sestavení HP uspesne ukonceno.
klikap@clegane:/opt/CZ-DRG-Pregrouper-1.0$
```



```
C:\WINDOWS\system32\cmd. x + v
CZ-DRG Pregrouper 2027 - 2027r0-260415
(C) 2018 - 2026, UZIS CR
2026-04-21T13:02:26.887316900 - 1/4 Ordering input files.
2026-04-21T13:02:26.897545400 - Max Memory (MB): 8192
2026-04-21T13:02:26.898548500 - Total Memory (MB): 8192
2026-04-21T13:02:26.898548500 - Free Memory (MB): 8149
2026-04-21T13:02:26.898548500 - Used Memory (MB): 42
2026-04-21T13:02:26.900753600 - Sorting file input/pgp_doklady02.csv
2026-04-21T13:02:26.900753600 - Input file size (MB): 0
2026-04-21T13:02:26.903046200 - Number of parts for external merge sort: 7.4246196E-7
2026-04-21T13:02:26.904315300 - Small enough file - in memory sorting possible
2026-04-21T13:02:26.912137100 - Sorted input/pgp_doklady02.csv in 3 ms
2026-04-21T13:02:26.914179 - Sorting file input/pgp_doklady06.csv
2026-04-21T13:02:26.914179 - Input file size (MB): 0
2026-04-21T13:02:26.914179 - Number of parts for external merge sort: 2.1478154E-6
2026-04-21T13:02:26.914179 - Small enough file - in memory sorting possible
2026-04-21T13:02:26.917526300 - Sorted input/pgp_doklady06.csv in 3 ms
2026-04-21T13:02:26.920549400 - Sorting file input/pgp_polozky.csv
2026-04-21T13:02:26.920549400 - Input file size (MB): 0
2026-04-21T13:02:26.921547800 - Number of parts for external merge sort: 6.9135554E-6
2026-04-21T13:02:26.921547800 - Small enough file - in memory sorting possible
2026-04-21T13:02:26.939450600 - Sorted input/pgp_polozky.csv in 18 ms
2026-04-21T13:02:26.940444400 - 2/4 Preparing polozky offset shard metadata.
2026-04-21T13:02:27.094252900 - 3/4 Aggregating data to hospital cases.
2026-04-21T13:02:27.240295400 - 9/9 doklady02 processed.
2026-04-21T13:02:27.550928200 - 4/4 Final run summary.
2026-04-21T13:02:27.554644600 - Doklady02K0RowCount: 0
2026-04-21T13:02:27.555893100 - Doklady06K0RowCount: 0
2026-04-21T13:02:27.555893100 - PolozkyK0RowCount: 0
2026-04-21T13:02:27.558814900 - K1 error 0 count: 3
2026-04-21T13:02:27.560251700 - K1 error 1 count: 0
2026-04-21T13:02:27.560251700 - K1 error 2 count: 0
2026-04-21T13:02:27.560251700 - K1 error 3 count: 0
2026-04-21T13:02:27.561250100 - K1 error 4 count: 0
2026-04-21T13:02:27.561250100 - K1 error 5 count: 0
2026-04-21T13:02:27.561250100 - K1 error 6 count: 0
2026-04-21T13:02:27.561250100 - K1 error 7 count: 0
2026-04-21T13:02:27.562250900 - K1 error 8 count: 0
2026-04-21T13:02:27.562250900 - K1 error 9 count: 0
2026-04-21T13:02:27.562250900 - K1 error 10 count: 0
2026-04-21T13:02:27.562250900 - K1 error 11 count: 0
2026-04-21T13:02:27.562250900 - Program finished in 1321ms.
2026-04-21T13:02:27.563516500 - Program will shutdown.
Klasifikace všech záznamů byla úspěšně ukončena, zprávu a záznamy o případných chybách najdete v souboru log/log.txt.
Press any key to continue . . .
```

Logování

Každé spuštění pre-grouperu je zaznamenáno v souboru `pregrouper_log.csvtxt`, 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 informace o úspěšném výpočtu anebo 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\"“
pregrouper-db-service-port-8888.bat/sh
    
```

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

~~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 Pregrouperu 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 pojišťoven 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