



# Aplikace pro řízení paralelního zpracování dat

## K čemu aplikace slouží ?

Aplikace je určena pro paralelní zpracování velkého množství textových dat z projektu CommonCrawl zpracovávaných Výzkumnou skupinou znalostních technologií v rámci projektu Corpora Processing Software.

Aplikace pracuje se 2 typy uzlů – master a slave. Uzly typu slave slouží jako výpočetní uzly. Zpracování úlohy je řízeno z centrálního uzlu typu master. Master obsahuje webové uživatelské rozhraní, které umožňuje definovat nové úlohy ke zpracování a výstupní validace pro ověření úspěšného zpracování daného vstupu. U spuštěných úloh poskytuje rozhraní pro řízení a monitorování běžících úloh.

## Monitorování běhu úlohy

U běžící úlohy aplikace poskytuje:

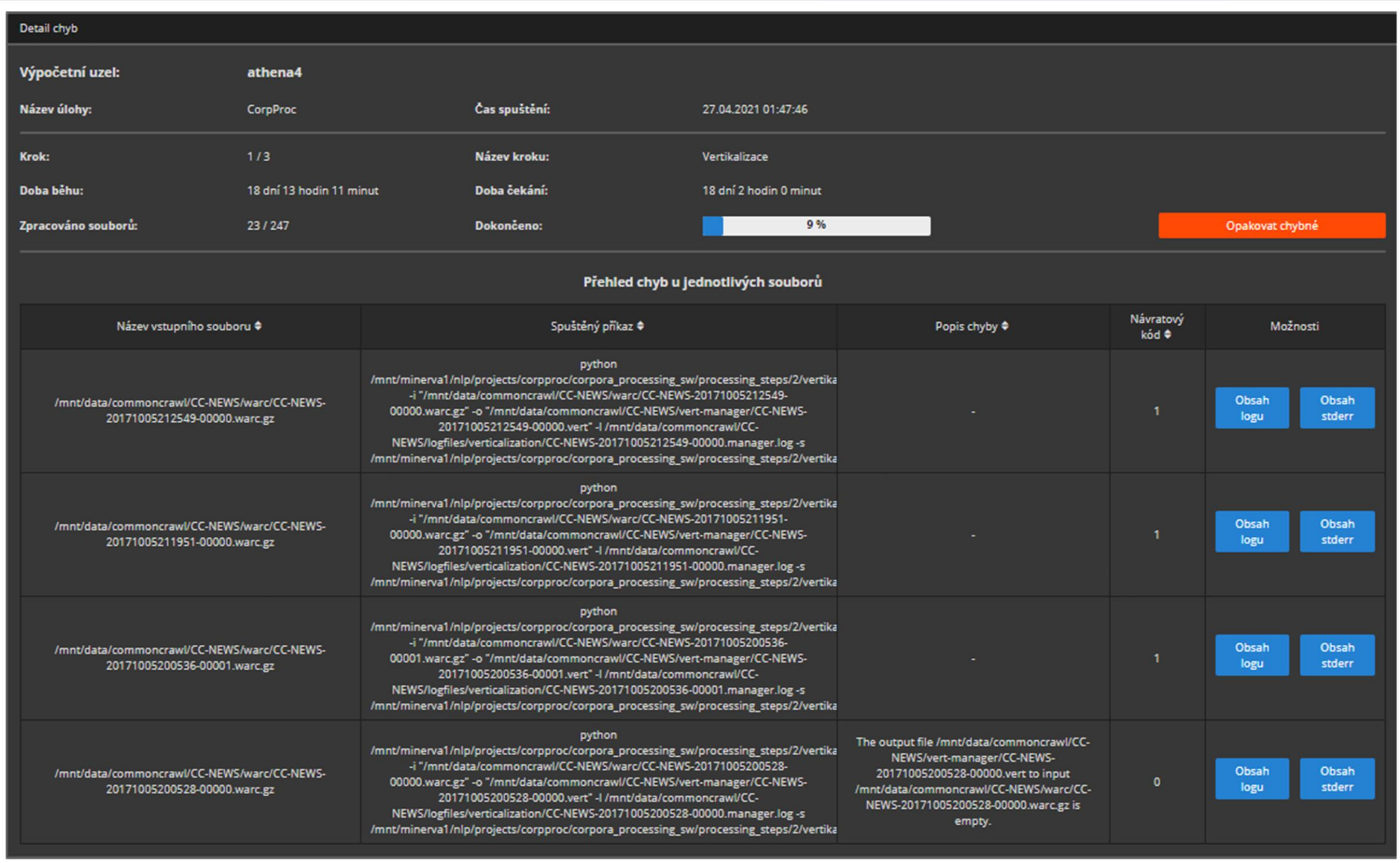
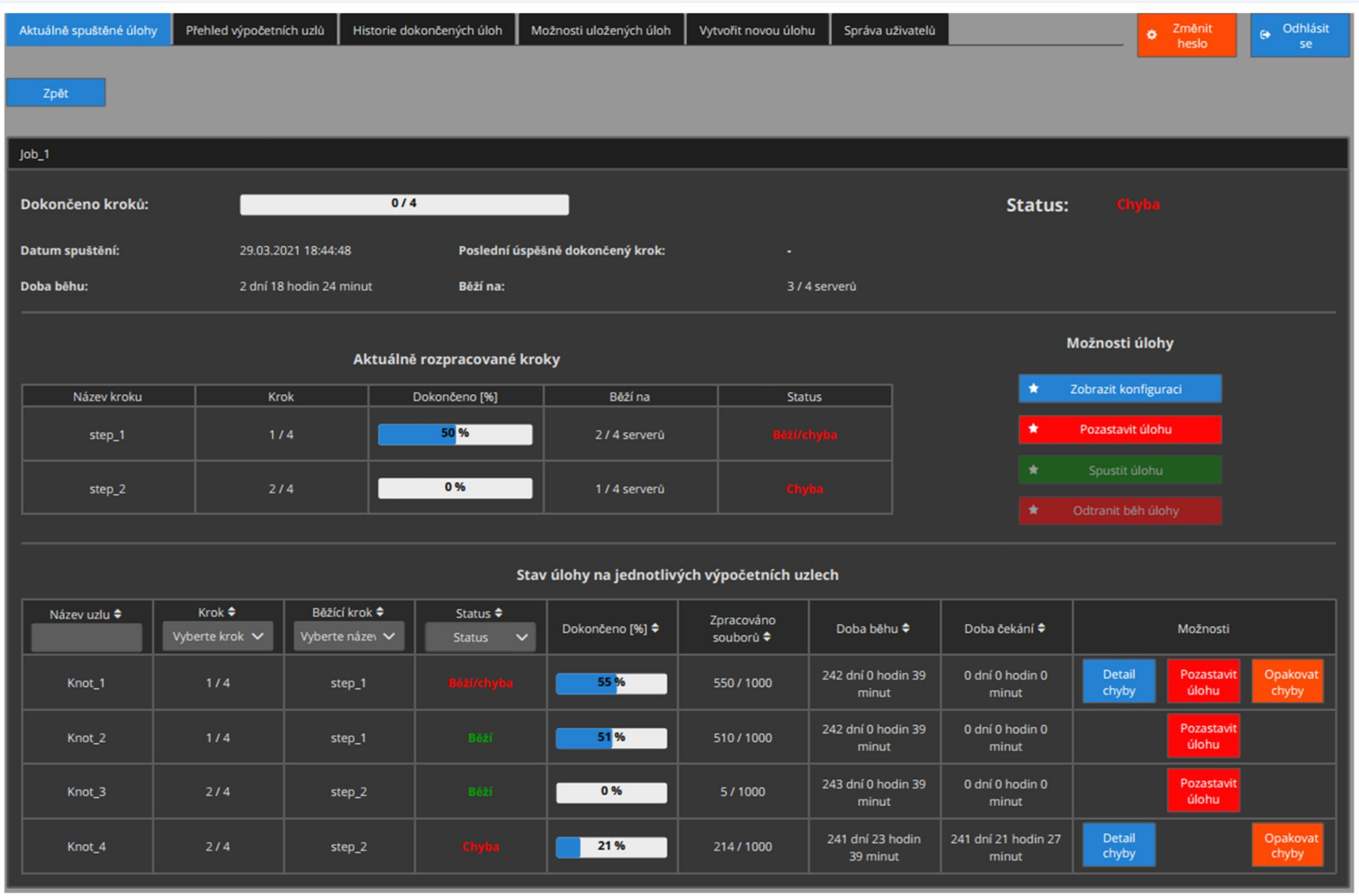
- Základní údaje o úloze
- Přehled o stavu všech kroků úlohy
- Přehled o stavu všech kroků úlohy na jednotlivých výpočetních uzlech
- Možnost zobrazit si přehled vstupů, jejichž zpracování skončilo s chybou
- Možnost zobrazit si konfiguraci úlohy
- Spustit/Pozastavit úlohu
- Spustit opětovné zpracování vstupů, které skončily s chybou

U každé chyby aplikace zobrazuje:

- Spuštěný příkaz
- Popis chyby
- Návrátový kód programu

Dále je možné v závislosti na konfiguraci úlohy:

- Zobrazit obsah definovaného logu
- Zobrazit obsah chybového výstupu



## Statistiky dokončených úloh

Během provádění úlohy aplikace ukládá statistiky o jejím průběhu. U každé dokončené úlohy je možné zobrazit statistiky o:

- Celkové době trvání úlohy
- Době zpracování daného vstupu
- Maximální/minimální době zpracování souboru
- Maximální/minimální době zpracování souboru normalizované velikostí vstupního souboru/složky
- Využití RAM
- Využití CPU

