

Zápis z konzultácie

Aplikácia pre paralelné spracovanie dat

KNOT

Katarína Galanská

28.06.2019

- vymyslieť názov aplikácie
- role
 - administrátor - mal by byť schopný resetovať heslá a zastaviť úlohy iných užívateľov
 - užívateľ
- malo by ísť pozastaviť úlohu, vynútenie pozastavenia úlohy, zastaviť úlohu úplne nemá zmysel
- pozastaviť úlohu všade na všetkých serveroch
- na výstup pridať RV s príponou .vert, .tagged, .parsed, náhrada zástupnej sekvencie
 - pomocou RV sa vyberie zo vstupu to, čo budem používať na výstupe (je mi jedno, či je to zložka alebo súbor)
 - do RV brať aj cestu k súboru, je to jednoduchšie na implementáciu
 - z RV dostanem list položiek, položky budem nahradzovať za \$1
 - aplikácia grúp RV aj do nových parametrov
 - `–out ../$1.vert`
 - `–warc ../$1.warc.gz`
- pre prácu s RV existuje knihovna, ktorá umožní získať list
- je potreba nahradzovať zástupné sekvencie v overení logu
- umožniť načítanie konfigurácie zo súboru
- priorita úlohy je definovaná podľa času spustenia

VSTUP	VÝSTUP
file	dir/file
dir of files	/mnt/data/out /mnt/data/out/\$1.vert

- jednotlivé kroky by mali mať svoj názov
- pridanie typu výstupu nám umožní povedať, že výstup je zložka a môžeme kontrolovať počet spracovaných súborov
- musí byť možné zmeniť a zobraziť konfiguráciu počas behu úlohy
- pri chybe je potreba vedieť na čom to spadlo, nestačí chybový kód, na akom príkaze to spadlo?, získať zo stderr
- vedieť ktorý krok ktorej úlohy spadol

Autentizácia

- iba IP adresa nestačí, je potreba generovať hash
- zahashuje sa náhodne vygenerovaný string skonkatenovaný s aktuálnym časom, pre toto riešenie je potreba mať synchronizovaný čas