

UZPD

HISTORICKÉ VLNY

Lubomír Bucek, Zuzana Richtrová

ČVUT 2018

Úvod

- Zadání a cíl projektu

Tvorba automatizovaného skriptu pro stahování dat o výškách vln.

Ukládání dat do NoSQL databázového systému Elasticsearch.

Zdrojová data

- NOAA National weather service
 - NCEP (National Centers for Environmental Prediction) – celosvětové modely počasí
- 62 mil. záznamů měsíčně
- Model Wavewatch III - rastrové soubory GRIB2(grided binary)
 - Production data – aktuální data
 - Historická data

Řešení problému

- formát dat GRIB2
 - Výstup z klimatických a jiných modelů
 - ecCodes – nástroj pro zpracování
 - Ovládání z příkazové řádky – Python knihovny nepoužitelné
 - Extrahování požadovaných informací

Řešení problému

- Python skripty – řízení běhu programu
 - `download.py` - Stahování dat na základě zadaného počátečního a koncového času
 - `process.py` – zpracovává a extrahuje stažená data pomocí nástroje `ecCodes`, nahrává data do databáze `Elasticsearch`
 - `download_process_past.py` – stahuje a zpracovává starší data

Řešení problému

- Bash skripty – řízení spouštění Python skriptů
 - download.sh – download.py
 - Přesměruje standardní výstup do .log souboru
 - elastic.sh – process.py
 - Čtení souboru s časy stažených dat
 - Spošťí process.py postupně pro časy stažených dat
 - Přesměrování výstupů do .log souboru
- Cron – spouštění bash skriptů pravidelně 4x denně

Elasticsearch

- Vyhledávací a úložný systém založený na Apache Lucene – full text
- Vhodné pro velké množství dat - rychlé vyhledávání oproti SQL - indexovací magie
- Odesílání požadavků a přijímání odpovědí jako JSON dokumenty přes API
- Velmi pomalý update, nelze joinovat

Závěr

- Úspěšně nasazeno a otestováno na Vesseltracker serveru

