

Meteorodáta v Pythone Efektívne

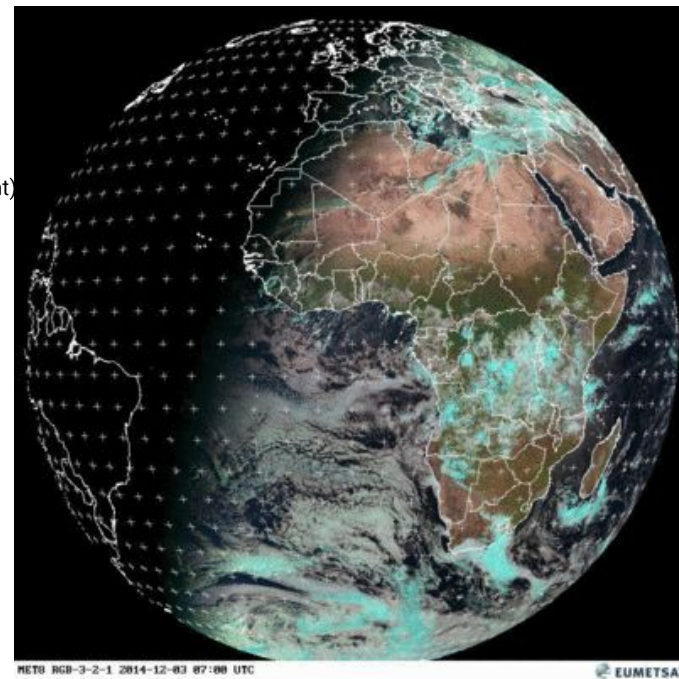
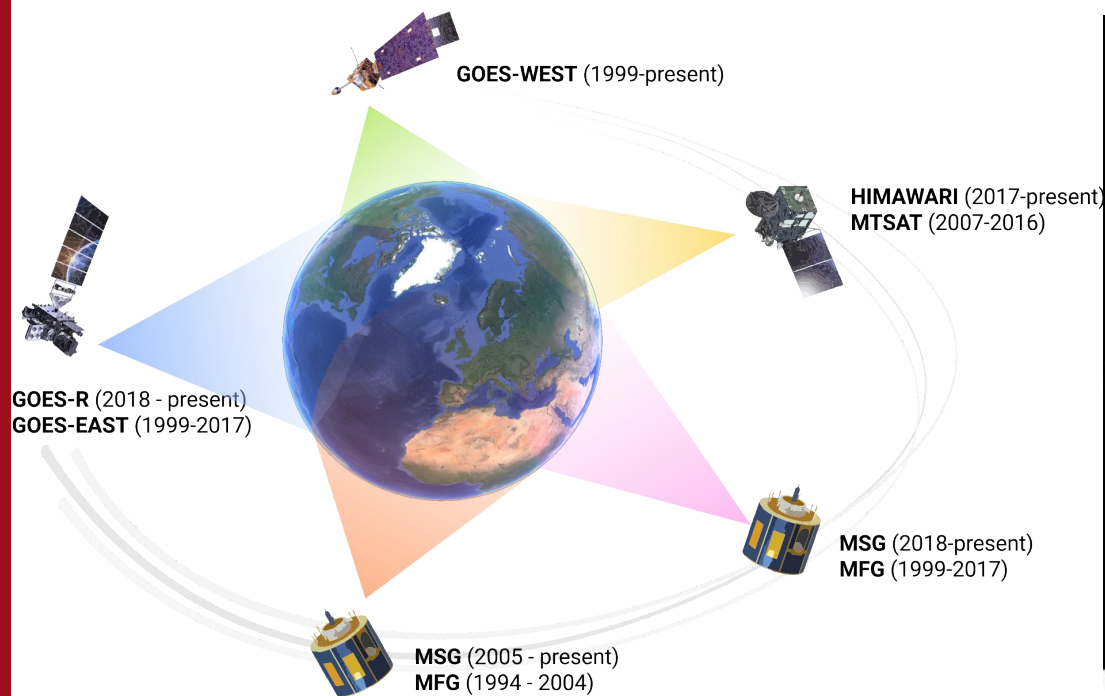
Miloš Korenčíak
Branislav Čief

SOLARGIS

O čom budeme hovoriť?

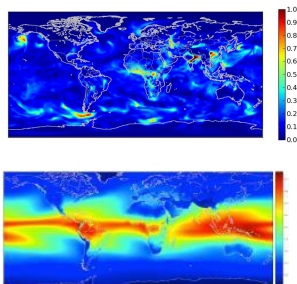
1. Čo sú metoodáta?
2. Formáty metoodát
3. Optimalizačné triky
4. Ukážky
5. Použitie metoodát
6. Otázky

Meteorodáta - vznik

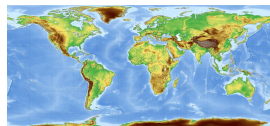


Meteorodáta - spracovanie

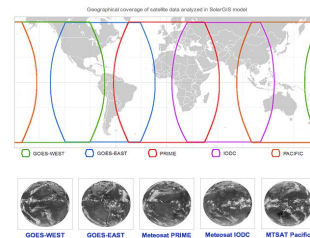
Atmosférické modely



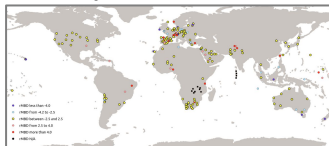
Terén a iné dáta



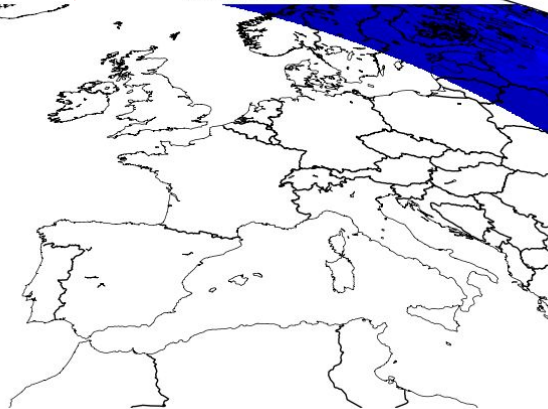
Meteorologické
satelity



Validácia pomocou
solárnych meraní



2017-07-22 05:15:00



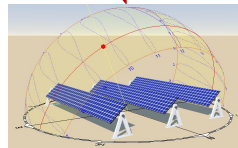
(c) Solargis s.r.o.

GHI
DNI
DIF

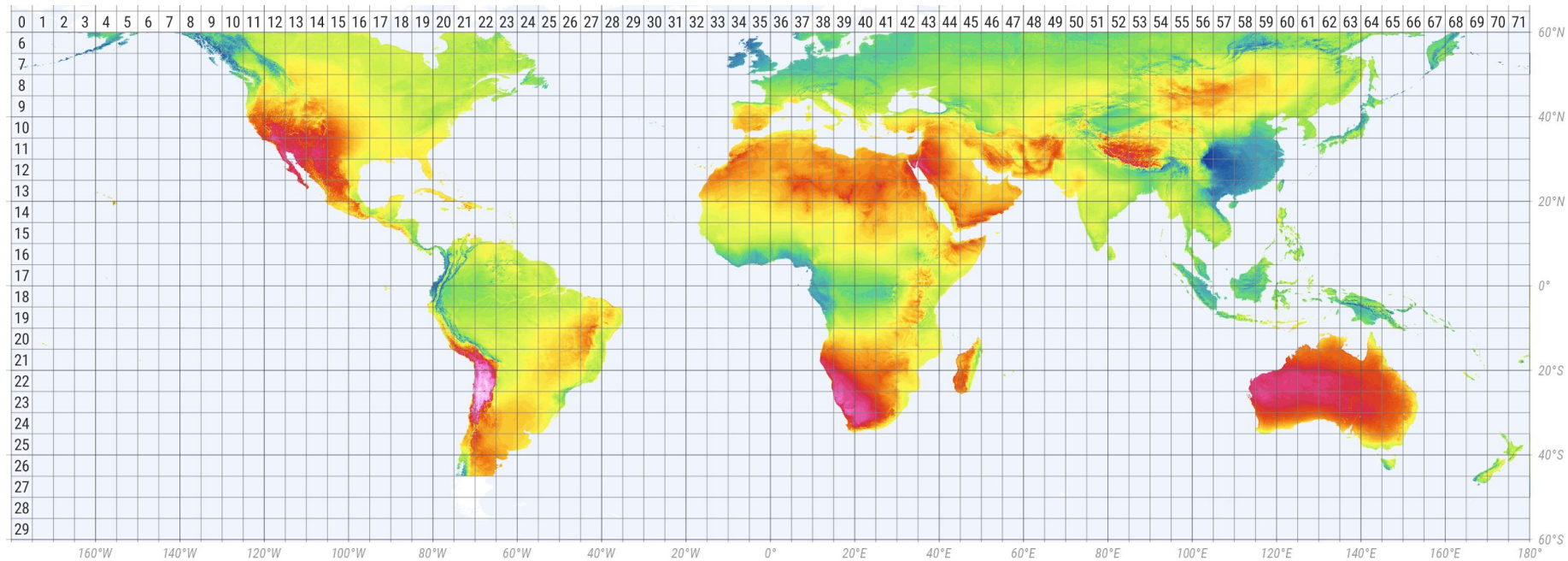
DNI
Direct Normal
Irradiance
(CSP elektrárne)



GTI
Global Tilted
Irradiance
(PV elektrárne)

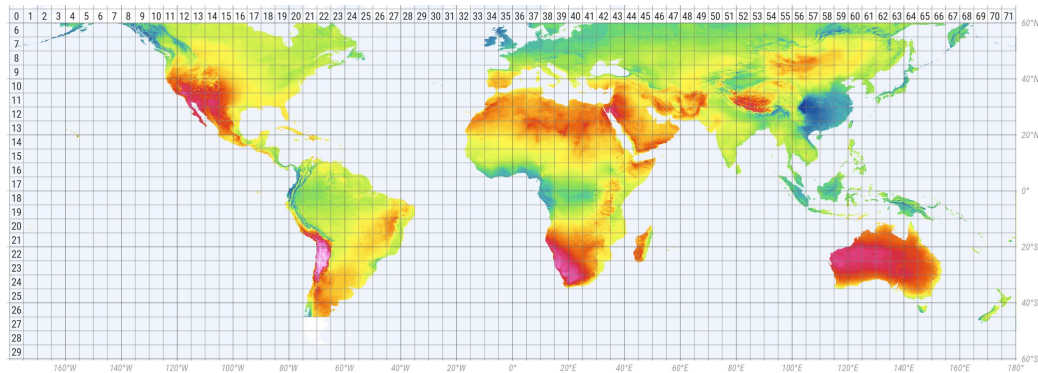


Meteodáta - výsledok



Akým IT výzvam čelíme?

- **Nutnosť segmentácie**
- Distribúcia medzi viacerými PC
- Nerovnomerný tok dát
- Obtiažna archivácia
- Krehkosť - Fault plány
- Špeciálny HW
- **Optimalizácia ukladania**



Úloha: Rychlo ukladat a loadovat velké data

Úloha: Rýchlo ukladať a loadovať veľké dáta

Aké veľké sú veľké dáta?

Úloha: Rýchlo ukladať a loadovať veľké dáta

- Pokrytie celého sveta min. 30 px/° (~ 14.5 Mpx)
- ~ 20 vstupných veličín / rastrových kanálov
- Každých 15 min počas 1 roka (35 040 časových bodov / rok)
- 10 -24 rokov histórie (podľa regiónu)

Aké veľké dáta?



Formáty meteodát

- grib
- hrit
- netCDF4
- geoTIFF
- area
- ...

netCDF4

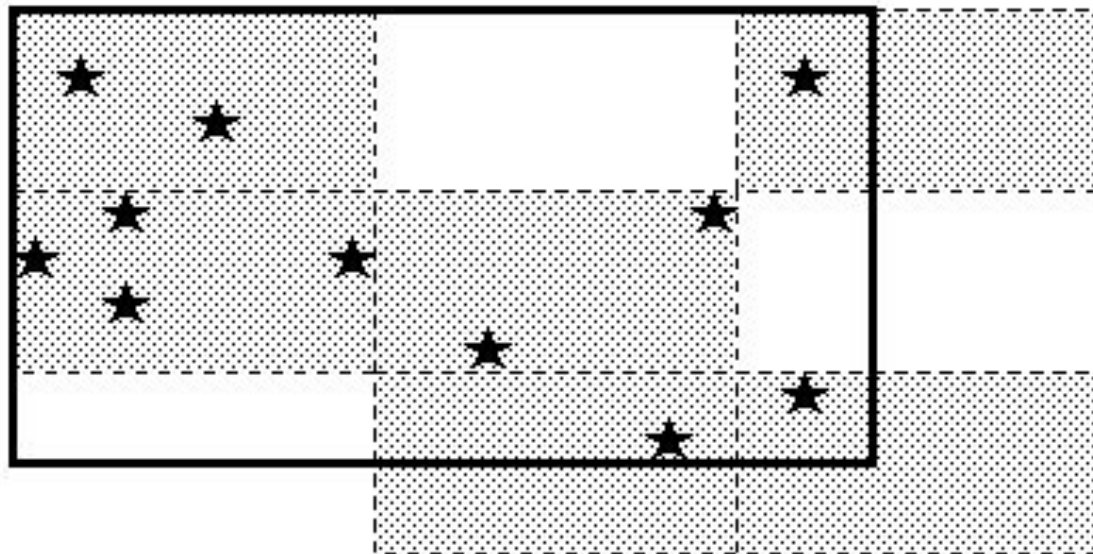
- Samopopisný
- Dobrá kompresia
- Binárny (= rýchly)
- Multi-dimenzionálny
- Flexibilný

Ako funguje netCDF4?

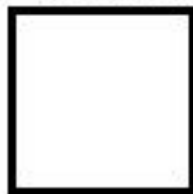
HDF5

netCDF4

CF1.6 compliant .nc



Dataset



Chunk



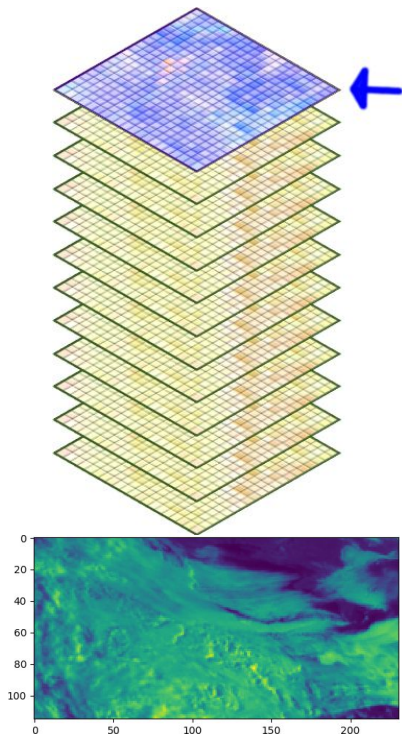
Point
Written



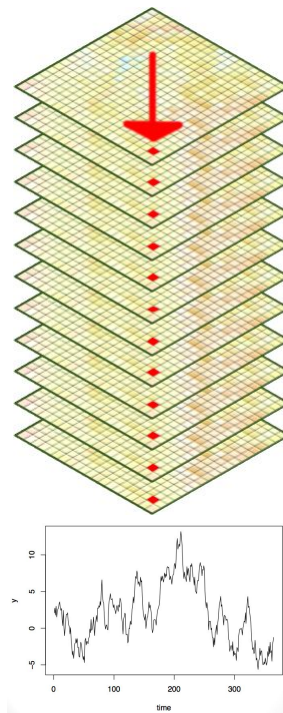
Dá sa to optimalizovať?

Dá sa to optimalizovať?

Priestorový prístup



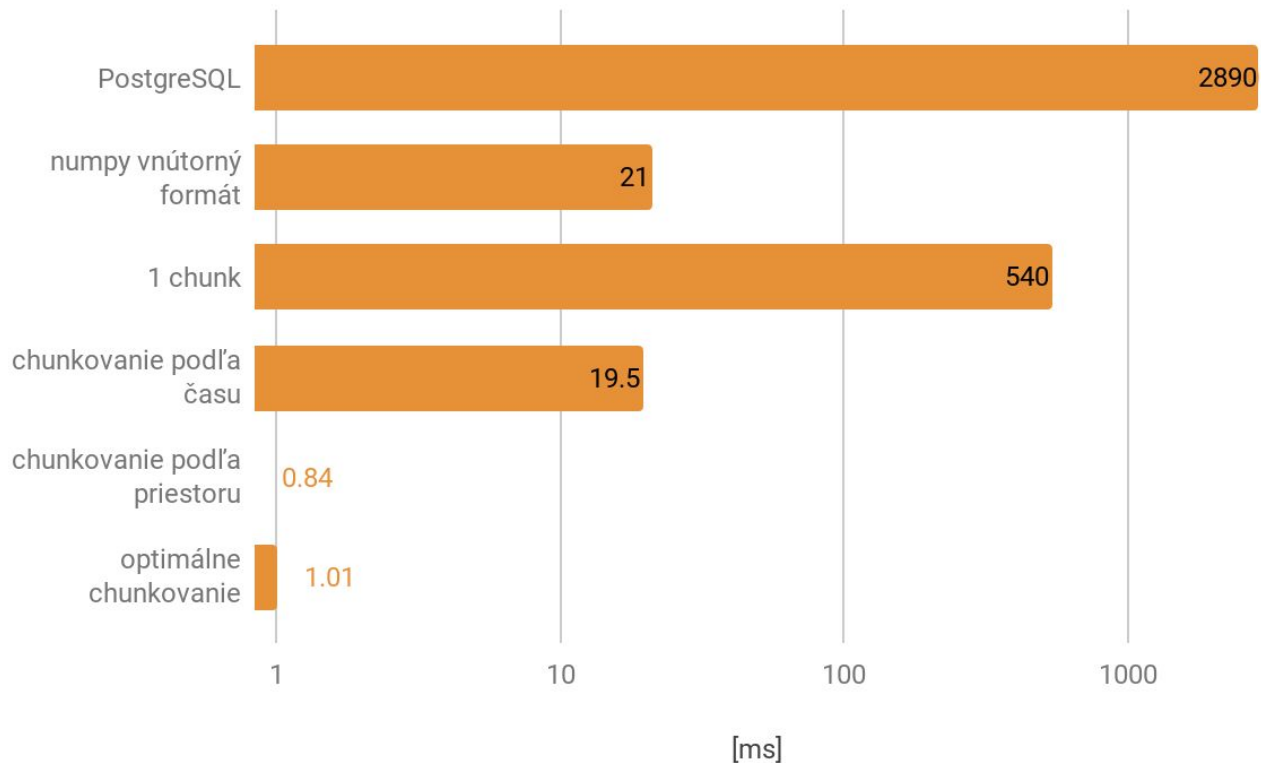
Bodový prístup



Prístup k dátam - priestorový

1. Postgres SQL
2. Numpy vnútorný formát
3. netCDF4 1 veľký chunk
4. netCDF4 chunkovaný podľa času
5. netCDF4 chunkovaný podľa priestoru
6. netCDF4 optimálne chunkovaný

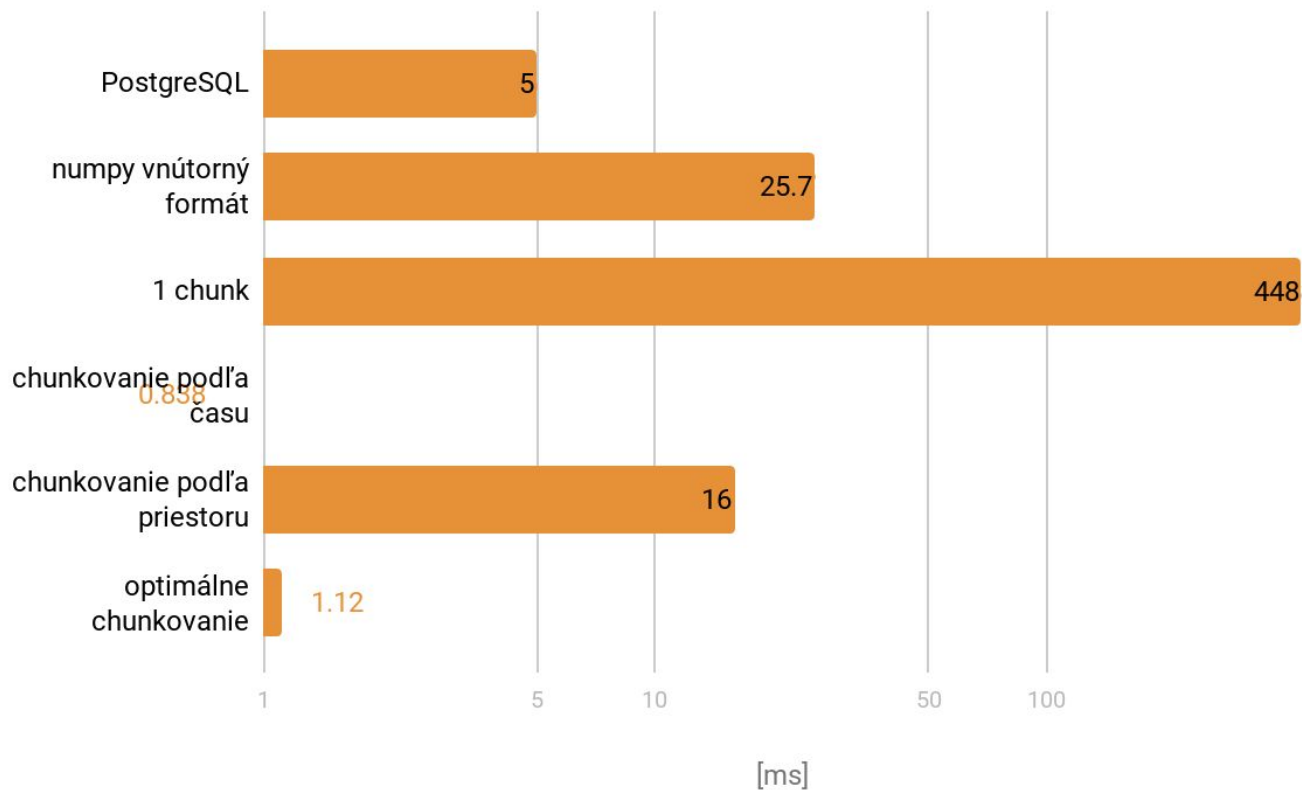
Prístup k dátam - priestorový



Prístup k dátam - bodový

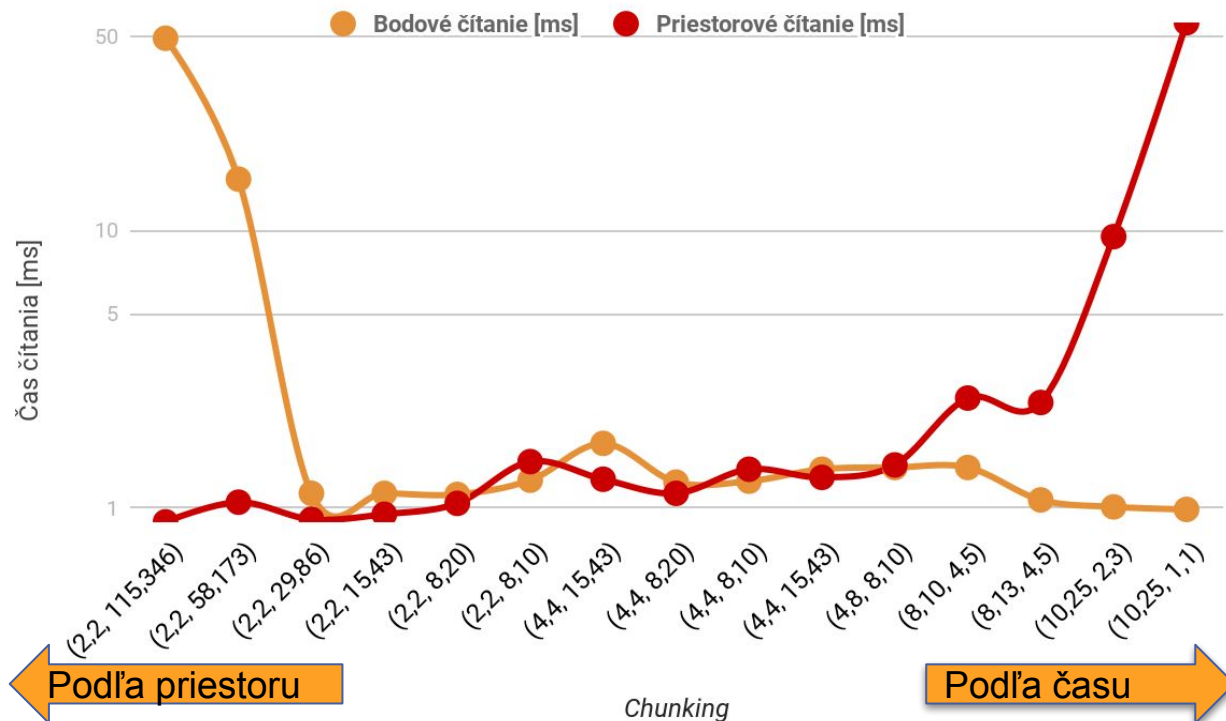
1. Postgres SQL
2. Numpy vnútorný formát
3. netCDF4 1 veľký chunk
4. netCDF4 chunkovaný podľa času
5. netCDF4 chunkovaný podľa priestoru
6. netCDF4 optimálne chunkovaný

Prístup k dátam - bodový



Porovnanie rôznych chunkovaní

Porovnanie rôznych chunkovaní



Poučenie

Dáta treba optimalizovať podľa toho, ako k nim pristupujeme.

Poučenie

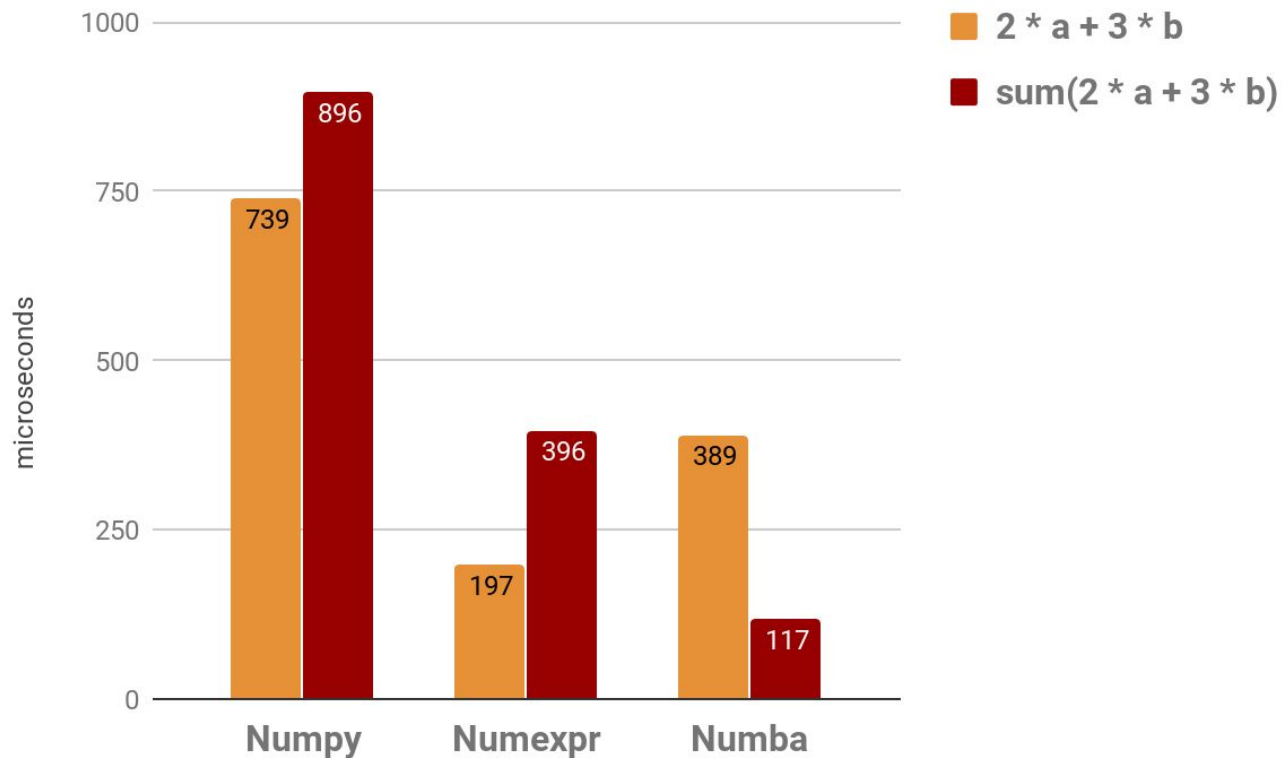
**Dáta treba optimalizovať podľa toho, ako
k nim pristupujeme.**

→ Dátová lokalita ←

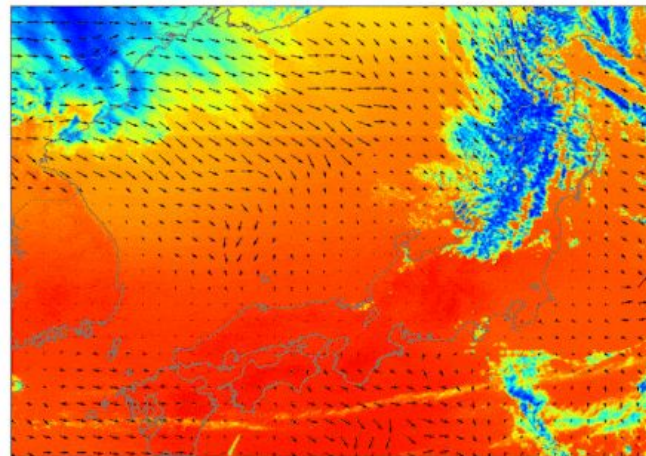
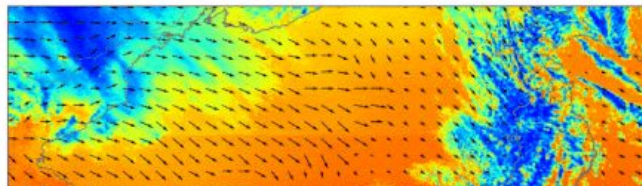
Porovnanie výkonu rôznych knižníc

- Numpy
- Numexpr
- Numba

Porovnanie výkonu rôznych knižníc



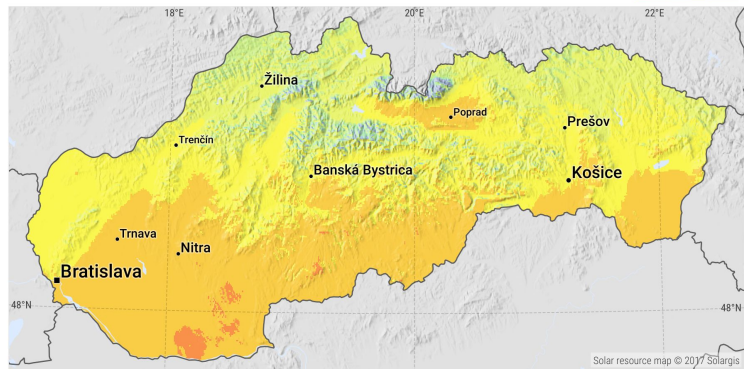
Čo ďalej s meteo dátami?



1/19

PHOTOVOLTAIC POWER POTENTIAL SLOVAKIA

SOLARGIS



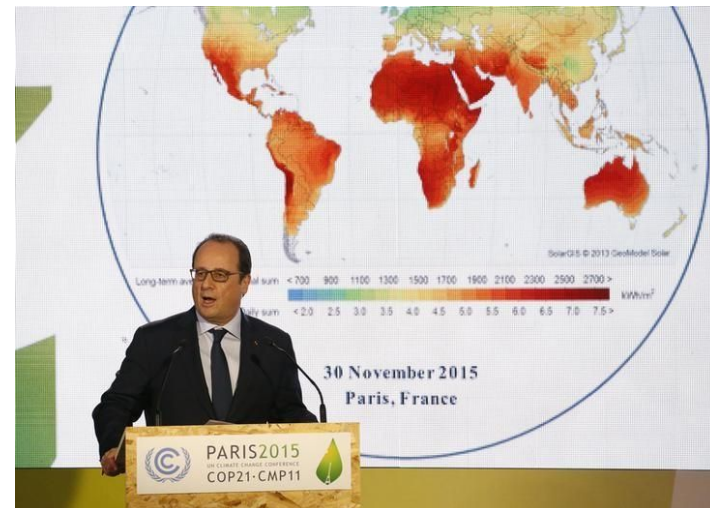
Average annual sum of PVOUT, period 1994-2016



This map is licensed by Solargis under the Creative Commons Attribution license (CC BY-SA 4.0). You are encouraged to use content of the map to benefit yourself and others in creative ways. For more information, please visit <http://solargis.com/download>.

SOLARGIS

Ako vplývame na svet?



WORLD BANK GROUP

THE WORLD BANK
IBRD • IDA

IFC | International
Finance Corporation

SOLARGIS

Ďakujeme za pozornosť