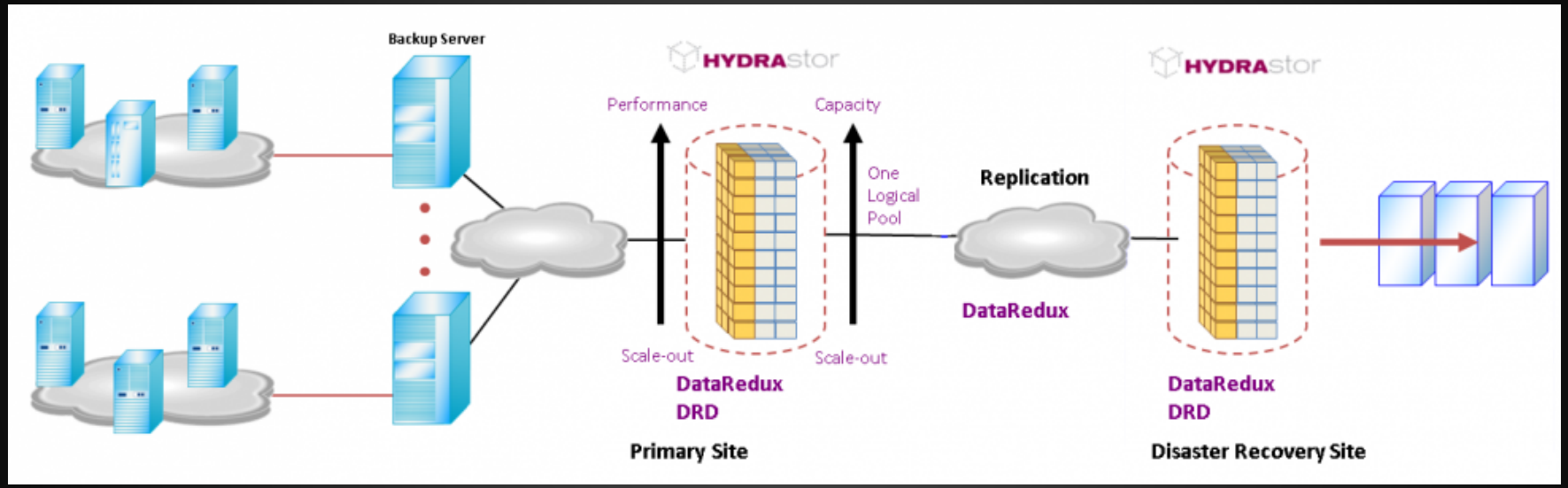
Optymalizacja ruchu sieciowego w symulowanym systemie rozproszonych kopii zapasowych

Autorzy:
Krzysztof Kiewicz
Paweł Kura
Dawid Łazarczyk
Marcel Zięba

Prowadzący:
dr Robert Dąbrowski

9LivesData

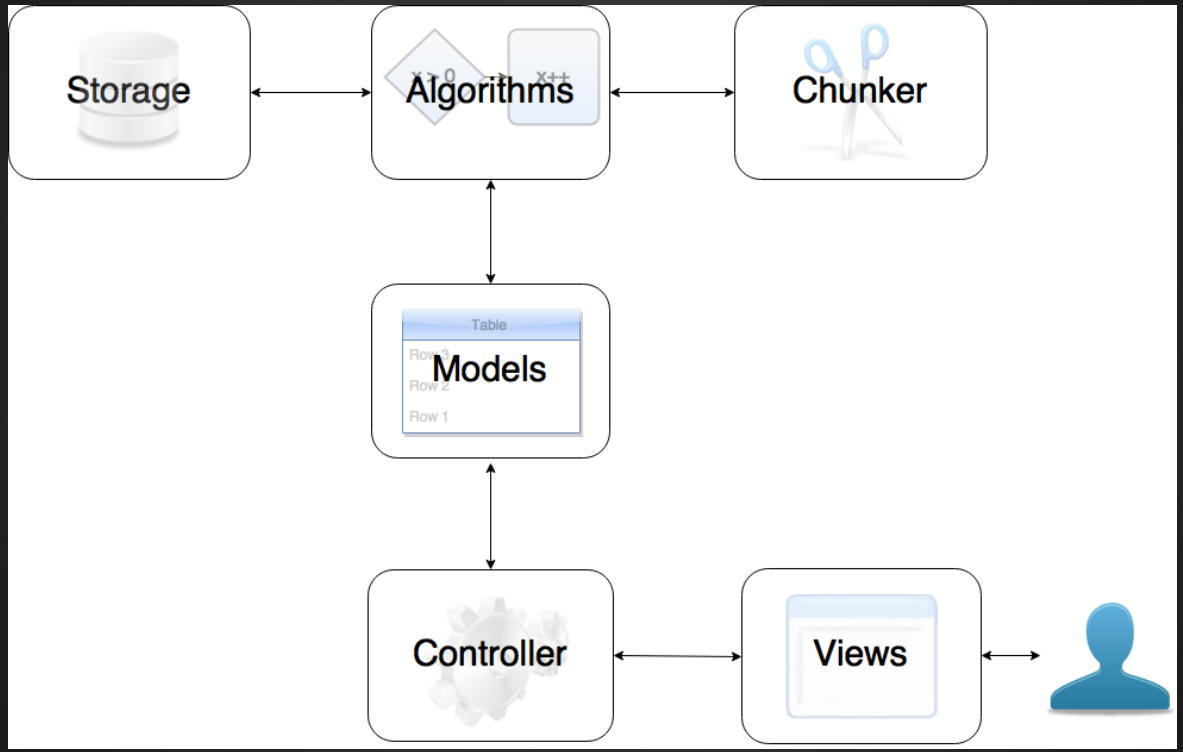
Wstęp

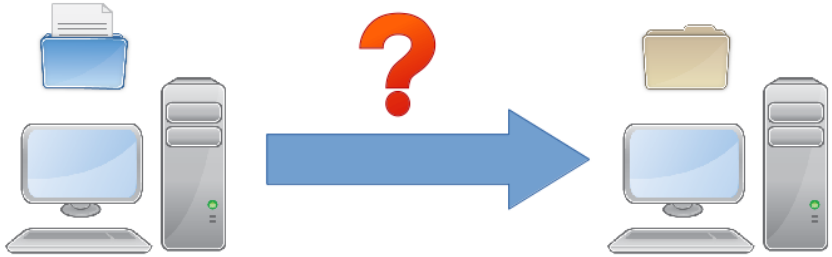

Backup Server
HYDRAstor
Performance
Capacity
One
Logical
Pool
Replication
DataRedux
Scale-out
Scale-out
DataRedux
DRD
Primary Site
HYDRAstor
DataRedux
DRD
Disaster Recovery Site

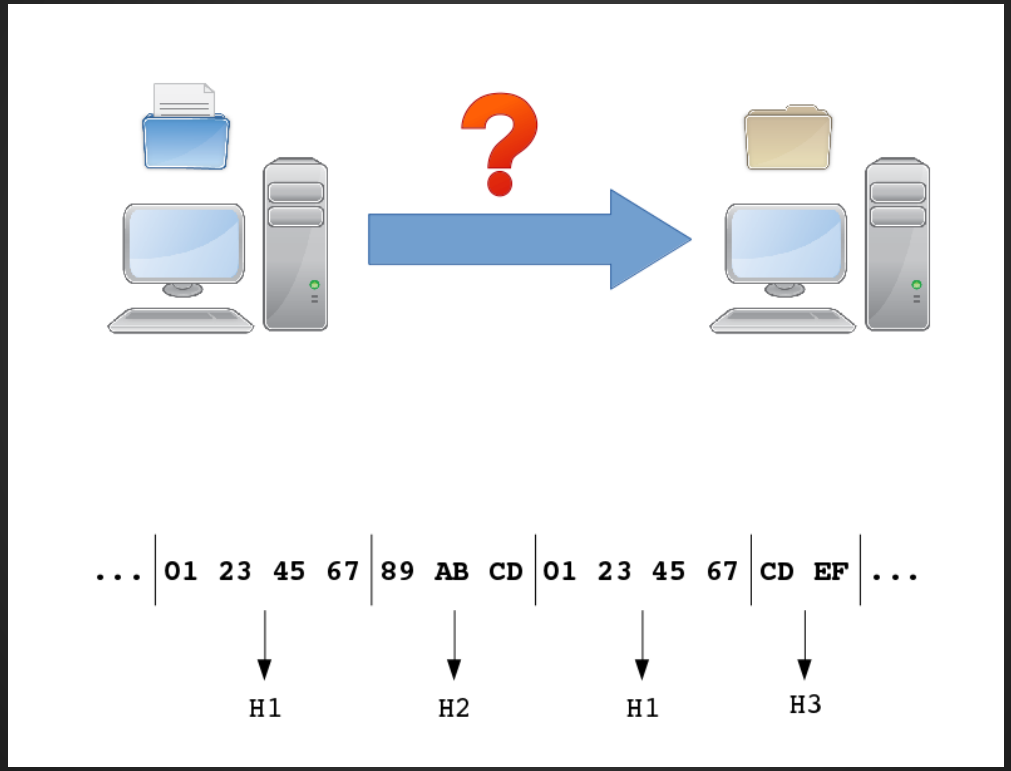
Cele projektu

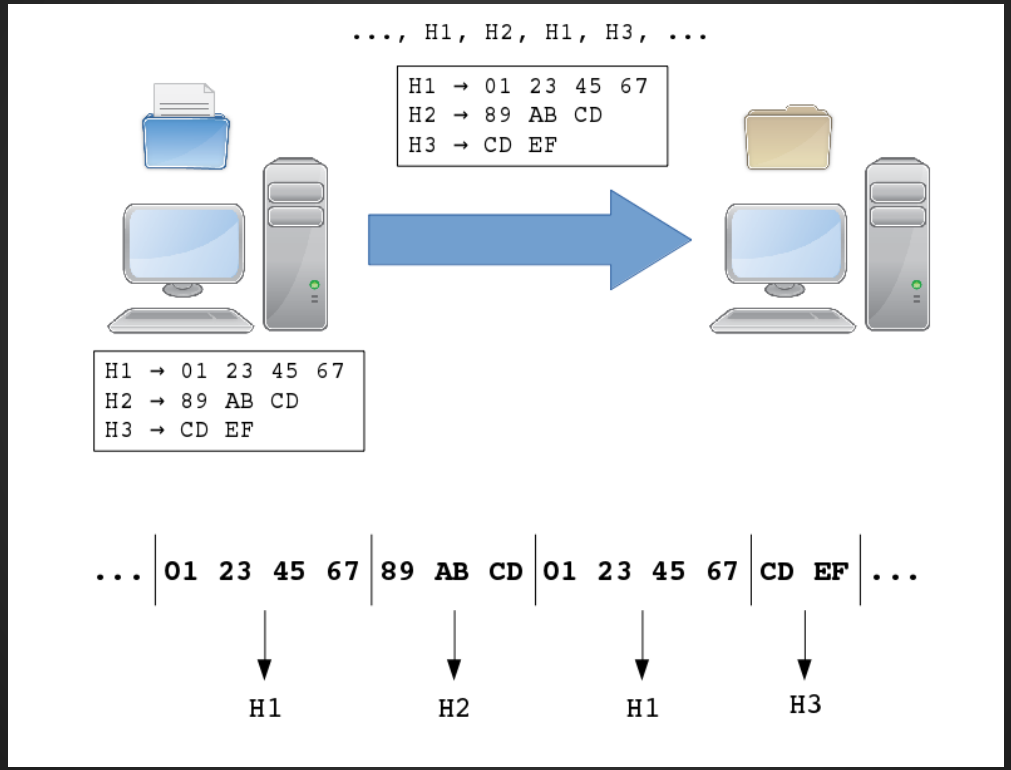
- Stworzenie symulatora
- Przygotowanie danych testowych
- Opracowanie wyników

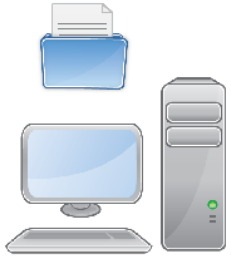
Architektura



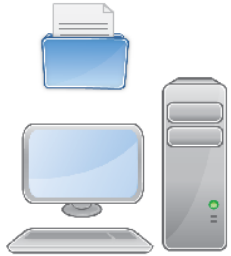



...
01 23 45 67
89 AB CD
01 23 45 67
CD EF
...
H1
H2
H1
H3

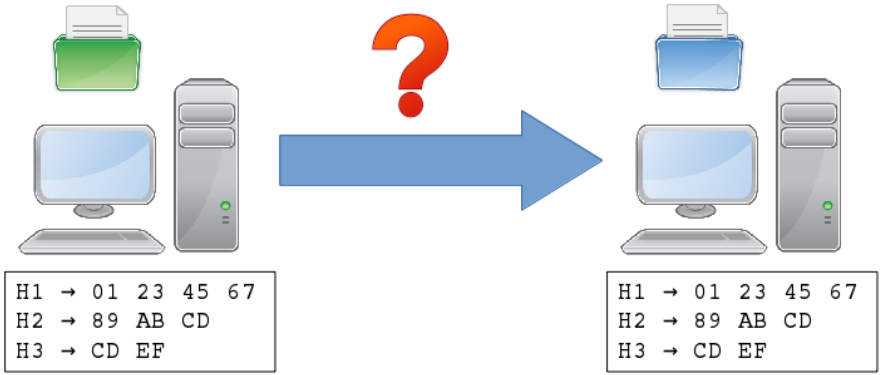

..., H1, H2, H1, H3, ...
H1 → 01 23 45 67
H2 → 89 AB CD
H3 → CD EF
H1 → 01 23 45 67
H2 → 89 AB CD
H3 → CD EF
... | 01 23 45 67 | 89 AB CD | 01 23 45 67 | CD EF | ...
H1
H2
H1
H3

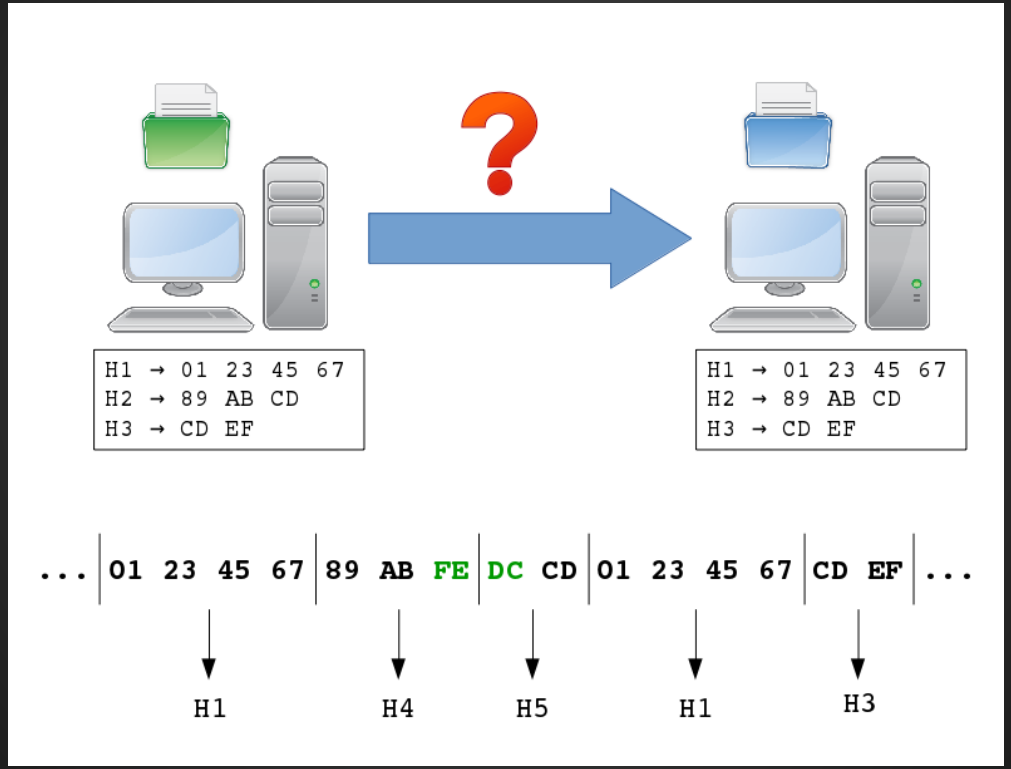


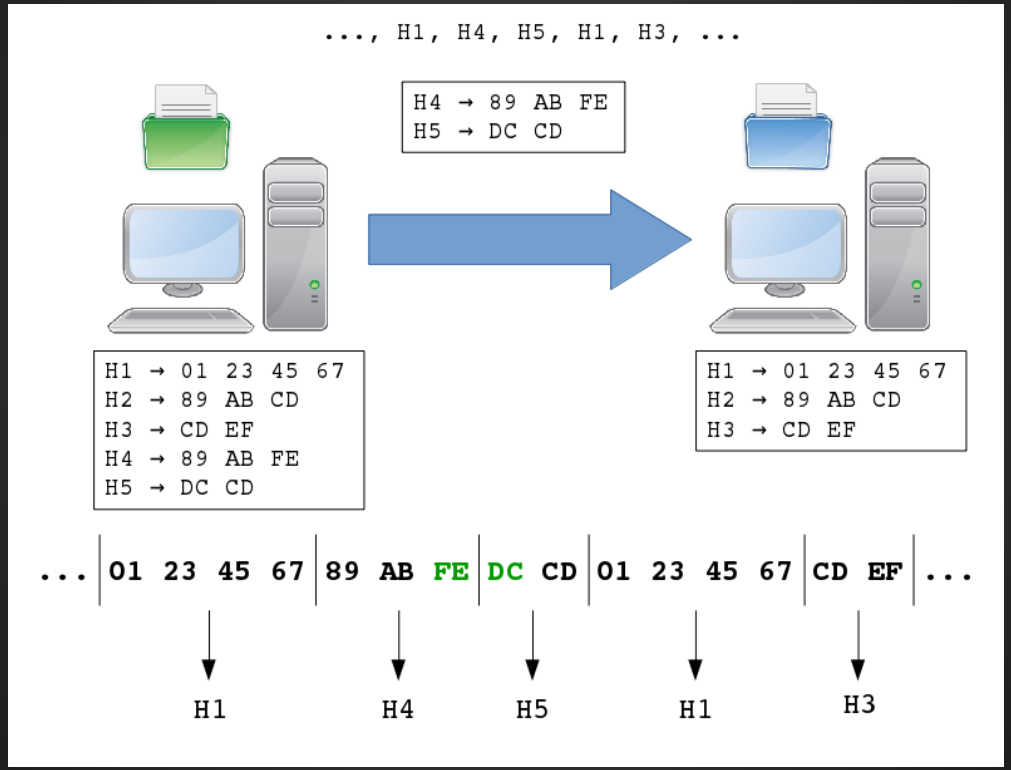
```
H1 → 01 23 45 67
H2 → 89 AB CD
H3 → CD EF
```

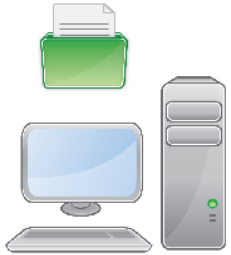


```
H1 → 01 23 45 67
H2 → 89 AB CD
H3 → CD EF
```

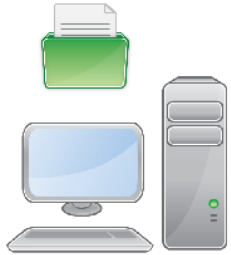

H1 → 01 23 45 67
H2 → 89 AB CD
H3 → CD EF
H1 → 01 23 45 67
H2 → 89 AB CD
H3 → CD EF


H1 → 01 23 45 67
H2 → 89 AB CD
H3 → CD EF
H1 → 01 23 45 67
H2 → 89 AB CD
H3 → CD EF
... 01 23 45 67 89 AB FE DC CD 01 23 45 67 CD EF ...
H1
H4
H5
H1
H3


..., H1, H4, H5, H1, H3, ...
H4 → 89 AB FE
H5 → DC CD
H1 → 01 23 45 67
H2 → 89 AB CD
H3 → CD EF
H4 → 89 AB FE
H5 → DC CD
H1 → 01 23 45 67
H2 → 89 AB CD
H3 → CD EF
... 01 23 45 67 89 AB FE DC CD 01 23 45 67 CD EF ...
H1
H4
H5
H1
H3



```
H1 → 01 23 45 67
H2 → 89 AB CD
H3 → CD EF
H4 → 89 AB FE
H5 → DC CD
```




```
H1 → 01 23 45 67
H2 → 89 AB CD
H3 → CD EF
H4 → 89 AB FE
H5 → DC CD
```

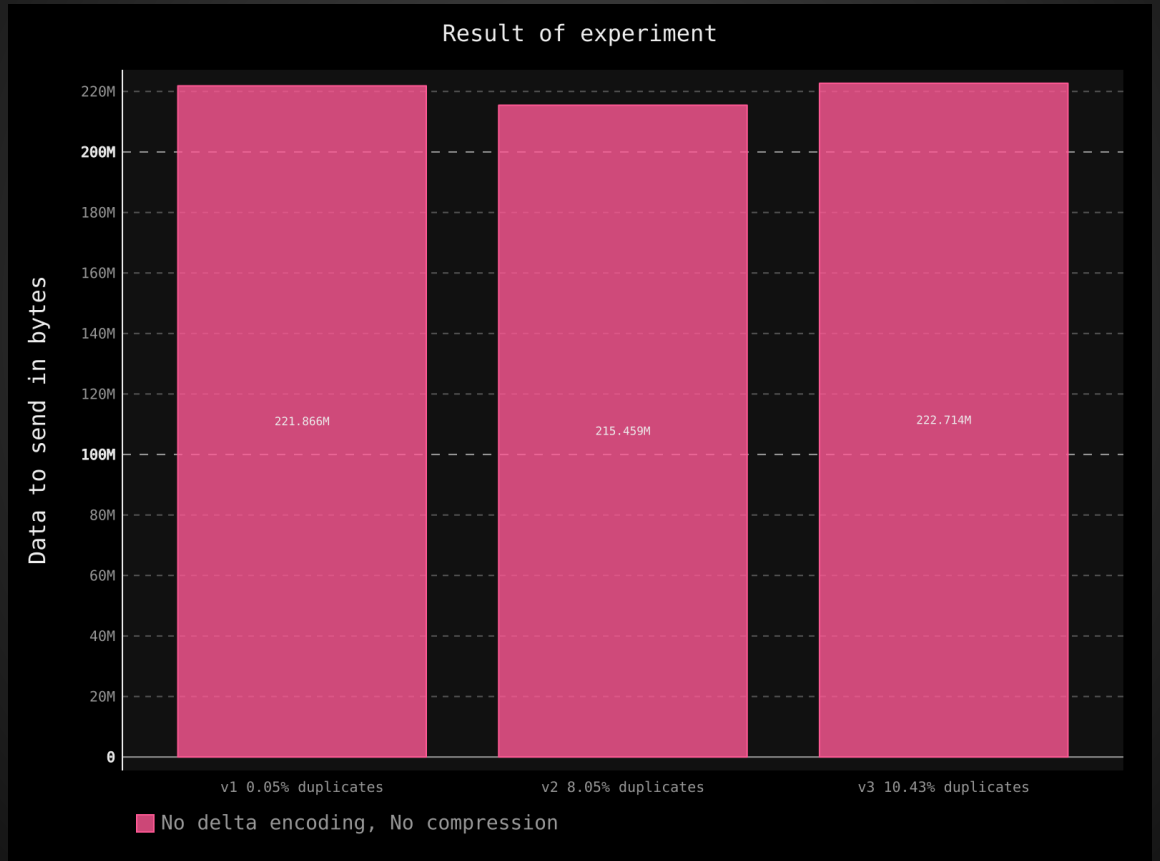
Testowane algorytmy

- Brak delta kodowania
- Optymalne delta kodowanie
- Indeks podobieństwa

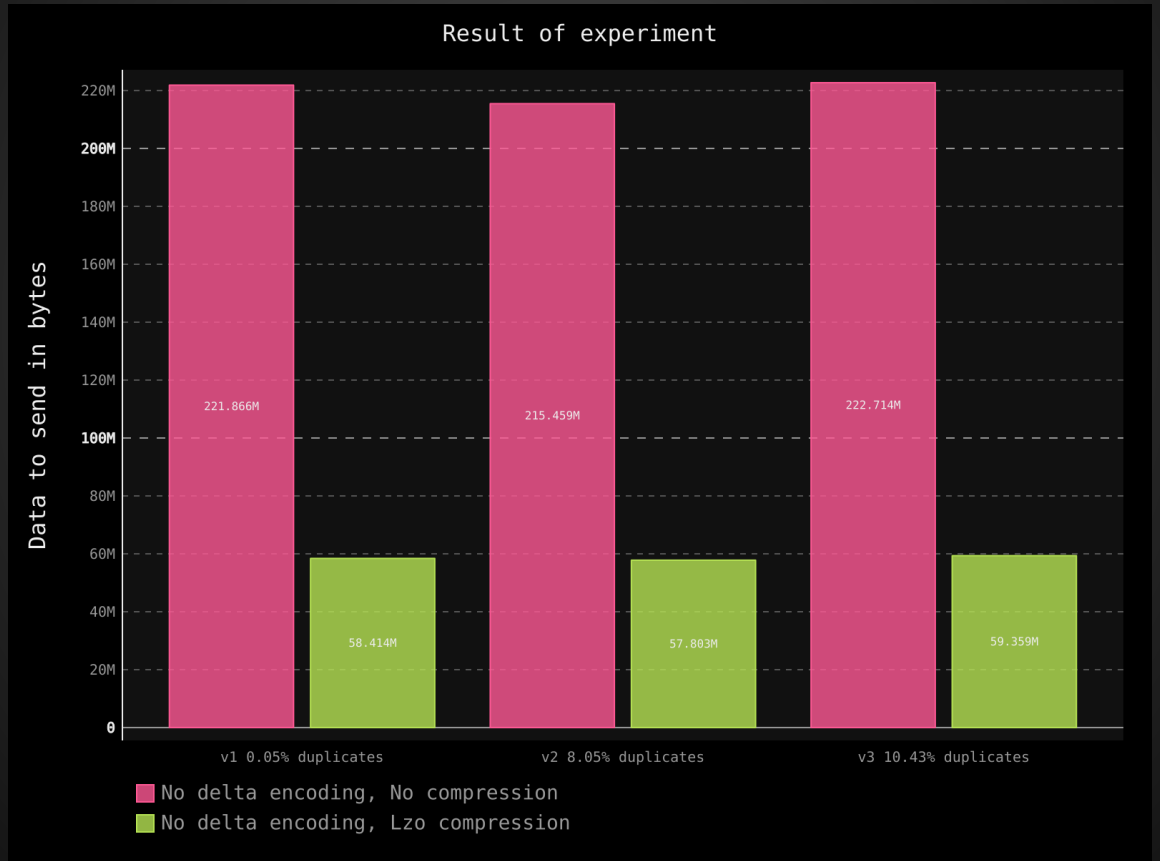
Dane testowe

- Źródła jądra Linuxa
- Kopie zapasowe katalogu domowego
- Bazy danych
- Strony internetowe

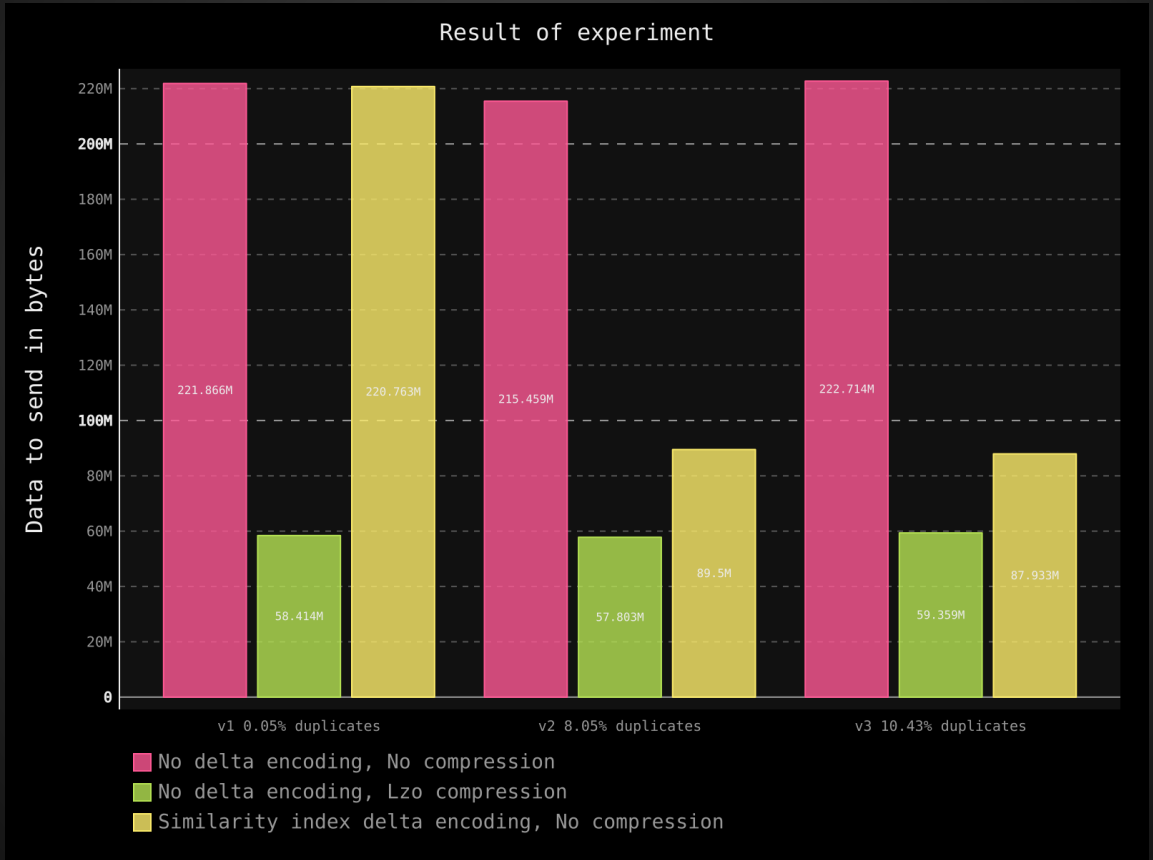
Źródła jądra Linuxa



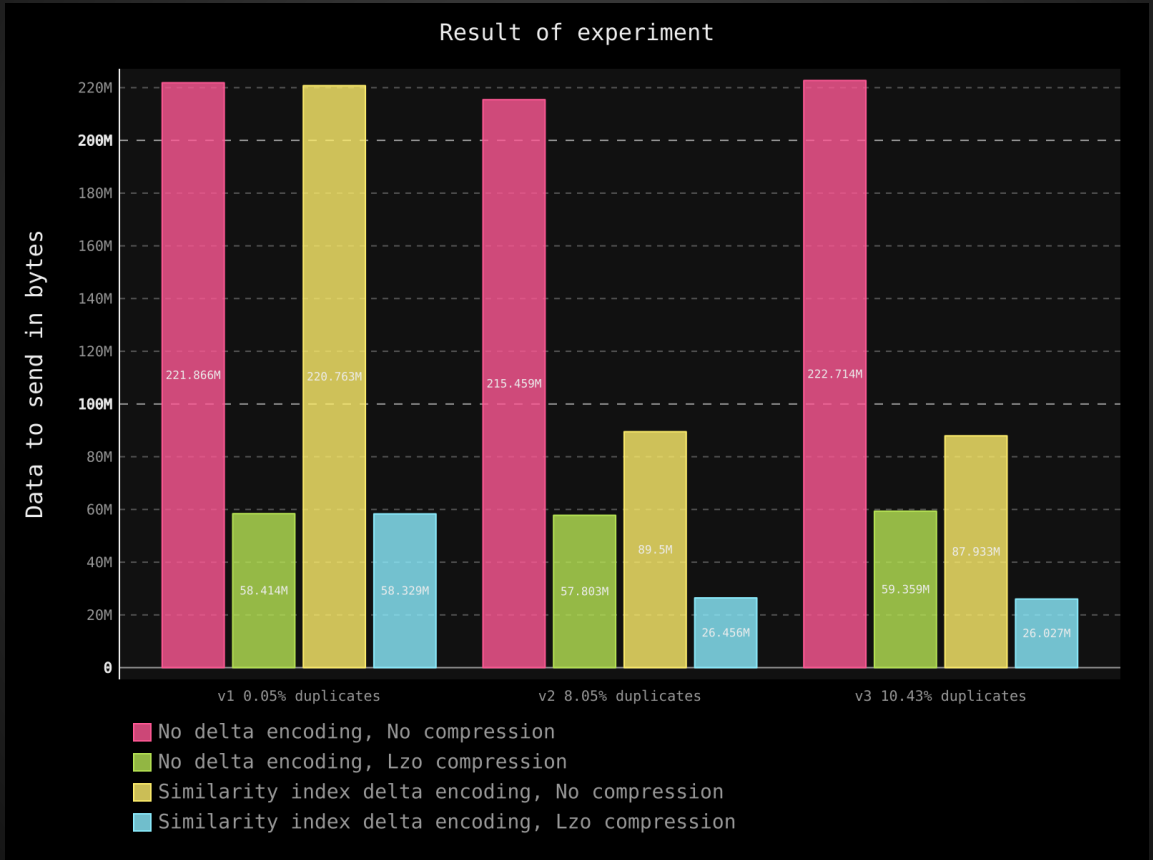
Źródła jądra Linuxa



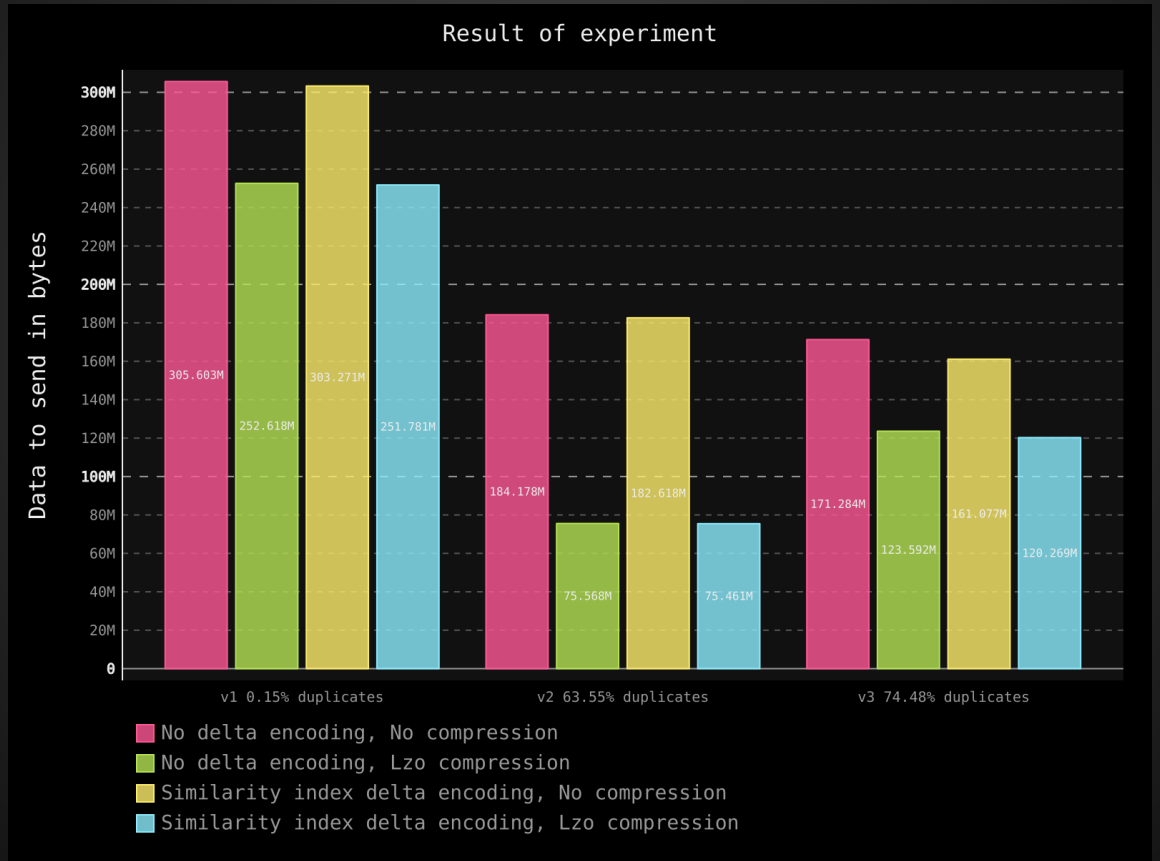
Źródła jądra Linuxa



Źródła jądra Linuxa



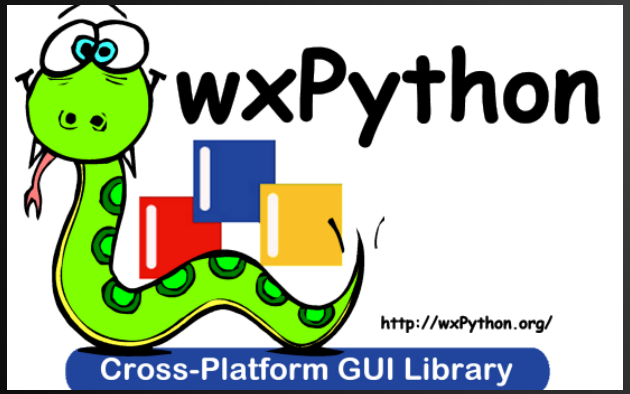
Wyniki dla katalogu domowego



Technologie

python™

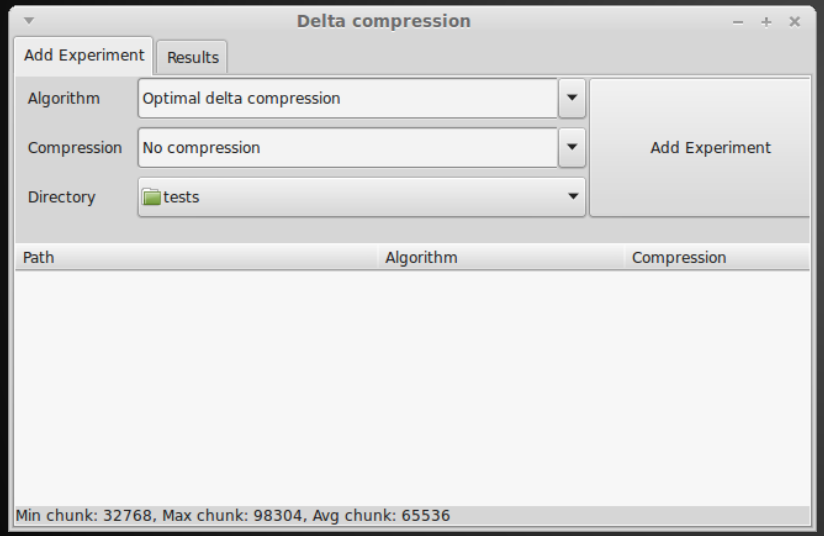
C++

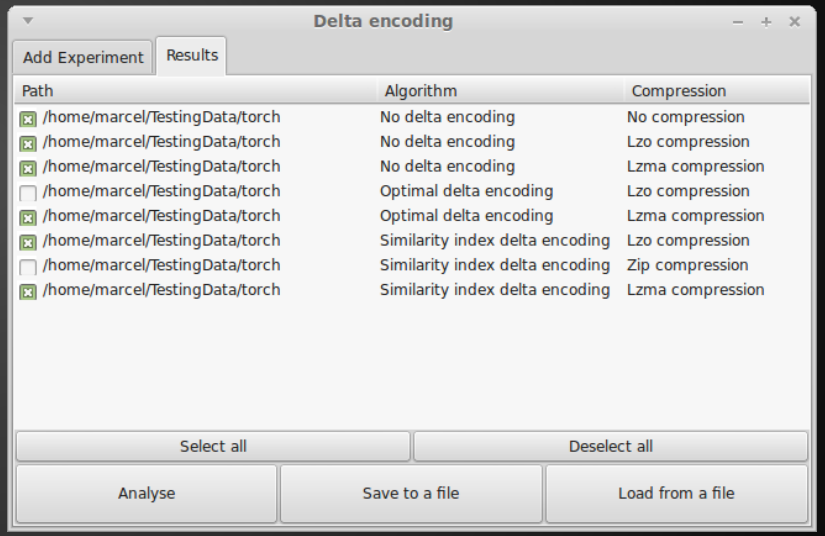


Organizacja pracy

- Agile
- Test Driven Development
- Inspekcja kodu (Code review)
- Github - repozytorium git
- Trello

Demo


Delta compression
Add Experiment
Results
Algorithm
Optimal delta compression
Compression
No compression
Directory
tests
Add Experiment
Path
Algorithm
Compression
Min chunk: 32768, Max chunk: 98304, Avg chunk: 65536


Delta encoding
Add Experiment
Results
Path
Algorithm
Compression
/home/marcel/TestingData/torch
No delta encoding
No compression
/home/marcel/TestingData/torch
No delta encoding
Lzo compression
/home/marcel/TestingData/torch
No delta encoding
Lzma compression
/home/marcel/TestingData/torch
Optimal delta encoding
Lzo compression
/home/marcel/TestingData/torch
Optimal delta encoding
Lzma compression
/home/marcel/TestingData/torch
Similarity index delta encoding
Lzo compression
/home/marcel/TestingData/torch
Similarity index delta encoding
Zip compression
/home/marcel/TestingData/torch
Similarity index delta encoding
Lzma compression
Select all
Deselect all
Analyse
Save to a file
Load from a file

Podsumowanie

- Symulator
- Algorytmy
- Dane testowe
- Wyniki

Dziękuję za uwagę