



POZNAN UNIVERSITY OF TECHNOLOGY

# Tematy projektów Edycja 2018

**Robert Wrembel**  
Poznan University of Technology  
Institute of Computing Science  
Robert.Wrembel@cs.put.poznan.pl  
[www.cs.put.poznan.pl/rwrembel](http://www.cs.put.poznan.pl/rwrembel)



## Analityka logów

- **Temat 1:** Parser zadań ETL w ABInitio (BZ WBK)
- **Temat 2:** Parser logów wykonania zadań ETL w ABInitio (BZ WBK)
- Założenia
  - wejście analizatora: plik tekstowy z danymi z logów ABInitio
  - parsowanie treści i zapisanie w bazie danych poszczególnych komponentów - sparsowanych planów wykonania i zawartości logów
  - baza danych relacyjna (np. Oracle) lub inna, taka aby można wykonywać analizy na treści
  - interfejs użytkownika z funkcjonalnością przeszukiwania i wizualizowania
    - technologia interfejsu dowolna, może być np. Tableau



## UDF w Netezza

---

⇒ **Temat 3:** Implementacja modelu predykcji w Netezza

⇒ Założenia:

- implementacja w Python
- przykładowe modele predykcji
  - random forrest, SVM epsilon regression, SVM nu regression
  - LS-SVM, feedforward neural network, bayesian additive regression trees, bayesian regularized neural networks
- przykładowe dane  
[https://trynthink.github.io/buildingsdatasets/show.html?title\\_id=long-term-energy-environment-data-for-ornl-research-house-3](https://trynthink.github.io/buildingsdatasets/show.html?title_id=long-term-energy-environment-data-for-ornl-research-house-3)



## Benchmark ETL

---

⇒ **Temat 4:** Benchmark ETL dla projektu "Balanced ETL optimization" (IBM)

- modyfikacja TPC-DS
- dane składowane w PostgreSQL i Hive
- dwa warianty
  - tabele z małą liczbą atrybutów
  - tabele z dużą liczbą atrybutów



## Składowanie danych medycznych

---

- **Temat 5:** Ocena technologii przetwarzania plików płaskich z badań klinicznych (Roche)
- Wymagania
  - ocena object file systems w składowaniu i przetwarzaniu danych z badań klinicznych
  - porównanie ofs z hdfs
- **Temat 6:** Budowa data lake danych medycznych w oparciu o hdfs (Roche)



## Składowanie danych medycznych

---

- **Temat 7:** Model standardu przechowywania danych klinicznych SDTM i jego implementacja w bazie danych (Roche)
- **Temat 8:** SPARQL Named Query Ontology and Service (Roche)
  - projekt typowo deweloperski, .Net, JS, Java
- **Temat 9:** Projekt i implementacja przykładowej bazy danych zgodnie se specyfikacją schematu OMOP (Roche)