

Wprowadzenie do Hurtowni Danych [223070-0115]

Zasady zaliczenia przedmiotu

Organizacyjnie

Prowadzący:

dr Mariusz Rafało

mrafalo@sgh.waw.pl

<http://mariuszrafalo.pl>

Zasady zaliczenia

1. Opracowanie rozwiązania problemu związanego z HD
2. Część technologiczna: implementacja
3. Część projektowa: dokument zawierający opis rozwiązania i zastosowanych technologii
4. Prezentacja wyników (ok 20 min.)

Przykładowe technologie

1. Integracja danych i ETL

- a) Talend
- b) Pentaho Kettle
- c) Apache NiFi
- d) Apache Kafka

2. Bazy danych

- a) Cassandra
- b) MongoDB

3. Raportowanie i analityka

- a) R
- b) Qlikview

Tematy projektów (przykłady)

1. Wersjonowanie danych za pomocą ETL Talend i Kettle
2. Hurtownia danych w architekturze big data
3. Hurtownia danych na platformie Apache Cassandra: wady i zalety
4. Rola bazy noSQL w architekturze HD na przykładzie MongoDB
5. Gromadzenie danych w czasie rzeczywistym za pomocą Apache Kafka