



Wyższa Szkoła Bankowa
Gdańsk Gdynia

Analiza danych w języku Java I

Zajęcia 7 – ETL

Anna Pakeizer

ETL

ETL [\[edytuj\]](#)

ETL ([ang.](#) *Extract, Transform and Load*) – narzędzia wspomagające proces pozyskania danych dla [baz danych](#), szczególnie dla [hurtowni danych](#).

Zadaniem narzędzi ETL jest:

- pozyskanie danych ze źródeł zewnętrznych,
- przekształcenie danych,
- załadowanie danych do bazy danych (zazwyczaj będącej hurtownią danych).

Zobacz też [\[edytuj\]](#) [| edytuj kod](#)]

- migracja danych
- OLAP

Kategoria: [Bazy danych](#)

<https://pl.wikipedia.org/wiki/ETL>

czasownik

1. **wydobyć, wyselekcjonować (np. dane)**

extract /ɪk'strækt/ **

czasownik

1. **wyciągać, wrywać** formal [TRANSITIVE]

Stop extracting your hair! (Przestań wrywać swoje włosy!)

The dentist extracted my wisdom tooth. (Dentysta wyrwał mojego zęba mądrości.)

2. **wydobywać, uzyskiwać (np. substancję)** [TRANSITIVE]

3. **wyciągać, wyduszać (np. informację)** [TRANSITIVE]

She extracted this information out of me. (Ona wyciągnęła ze mnie tę informację.)

Don't extract this from me, it's a secret. (Nie wyciągaj tego ze mnie, to tajemnica.)

4. **wykorzystywać, korzystać** [TRANSITIVE]

5. **wyciągać (pierwiastek matematyczny)** [TRANSITIVE]

He can't extract square roots without a calculator. (On nie potrafi wyciągać pierwiastków kwadratowych bez kalkulatora.)

[Słownik matematyczny](#)

S | ETL

- . S | ETL
- . Source | Extract Transform Load

Source

- Source – pozyskanie danych surowych ze źródła
- Jeżeli dane dla nas pozyskuje zewnętrzny system to rzeczywiście z naszego punktu widzenia tkwimy w ETL
- Jeżeli sami pozyskujemy dane to czynności dla Source i Extraction są inne, stąd warto je rozgraniczyć

Extraction

- . PreExtraction
- . Extraction
- . PostExtraction

PreExtraction

- Walidacja, czy dostaliśmy to czego oczekiwaliśmy
- Jeżeli dostaliśmy dane ale w złym formacie to możemy to tutaj zmienić, np. z JSONa do csv

Extraction

- Extraction fields – pola, które chcemy wyciągnąć z danych wejściowych
- Extraction fields to jest słownik np. {„imię”: „Jan”, „Nazwisko”: „Kowalski”, } // oczywiście w zależności od implementacji może być inaczej nazwany

Extraction – Transformation mapping

- . Transformation ma swoje Transformation fields
- . W tym miejscu mapujemy extraction fields z transformation fields
- . Nie wszystkie extraction fields muszą się znaleźć w transformation fields

Transformation

- . Przekształcanie transformation fields do postaci, jakiej oczekuje nasz distribution system

Transformation fields

- pola mogą być:
 - niezmienione, czyli równać się extraction fields;
 - zmienione/przekształcone;
 - wyliczone (derived) – są to pola, które w ogóle nie występowały w danych wejściowych, np. na wejściu mamy datę urodzenia, polem derived może być wiek, gdyż sami jesteśmy w stanie to sobie wyliczyć.

Transformation – Distribution mapping

- Łączymy transformation fields z distribution fields
- Nie wszystkie transformation fields muszą znaleźć się w mapowaniu
- Niektóre transformation fields są używane jako pomocnicze
- Niektóre transformation fields mogą być użyte tylko w niektórych scenariuszach

Distribuion

- . Wysyłanie danych do systemu zewnętrznego w takiej postaci oraz w taki sposób jaki ten system oczekuje

Narzędzia ETL

- . Microsoft SSIS
- . Apache Nifi