



Data Engineering

Ionian IEEE SB

Agenda



01. Introduction

- What is Data Engineering
- Data Engineering needs

02. APIs

- What is an API
- REST API
- API in Data Engineering

03. Data Processing

- Dataframes
- Data Normalization
- Data Cleaning

04. Databases

- Relational Databases
- Exporting to a Database

05. Data Visualisation

- Matplotlib / Seaborn
- Plot types

Data Science

- Data Scientist:
 - Επίπεδα 1 - 3
- Data Engineer:
 - Επίπεδα 4 - 5

THE DATA SCIENCE HIERARCHY OF NEEDS

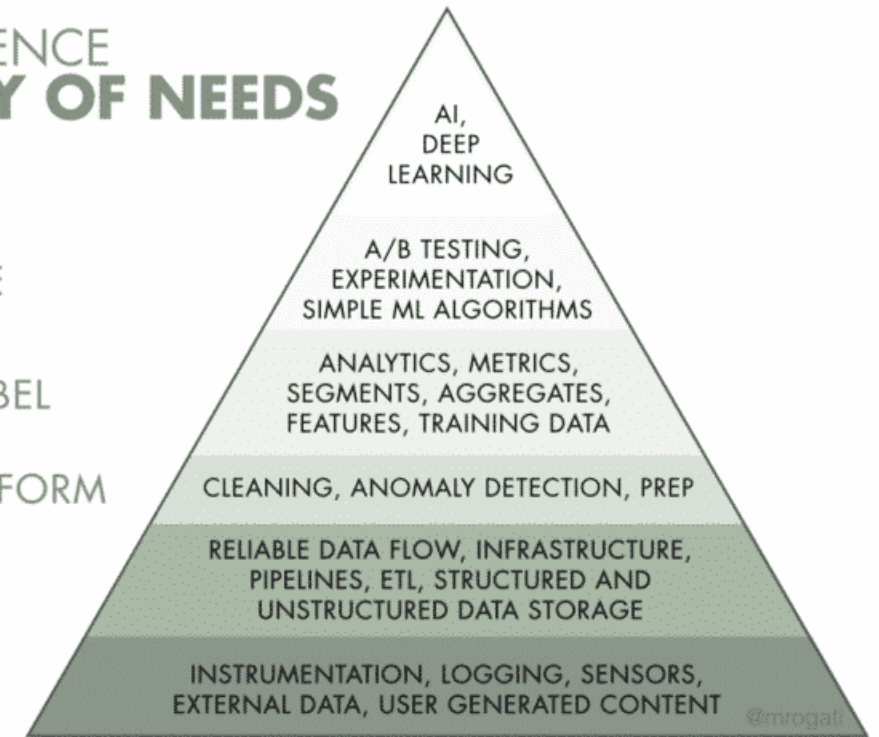
LEARN/OPTIMIZE

AGGREGATE/LABEL

EXPLORE/TRANSFORM

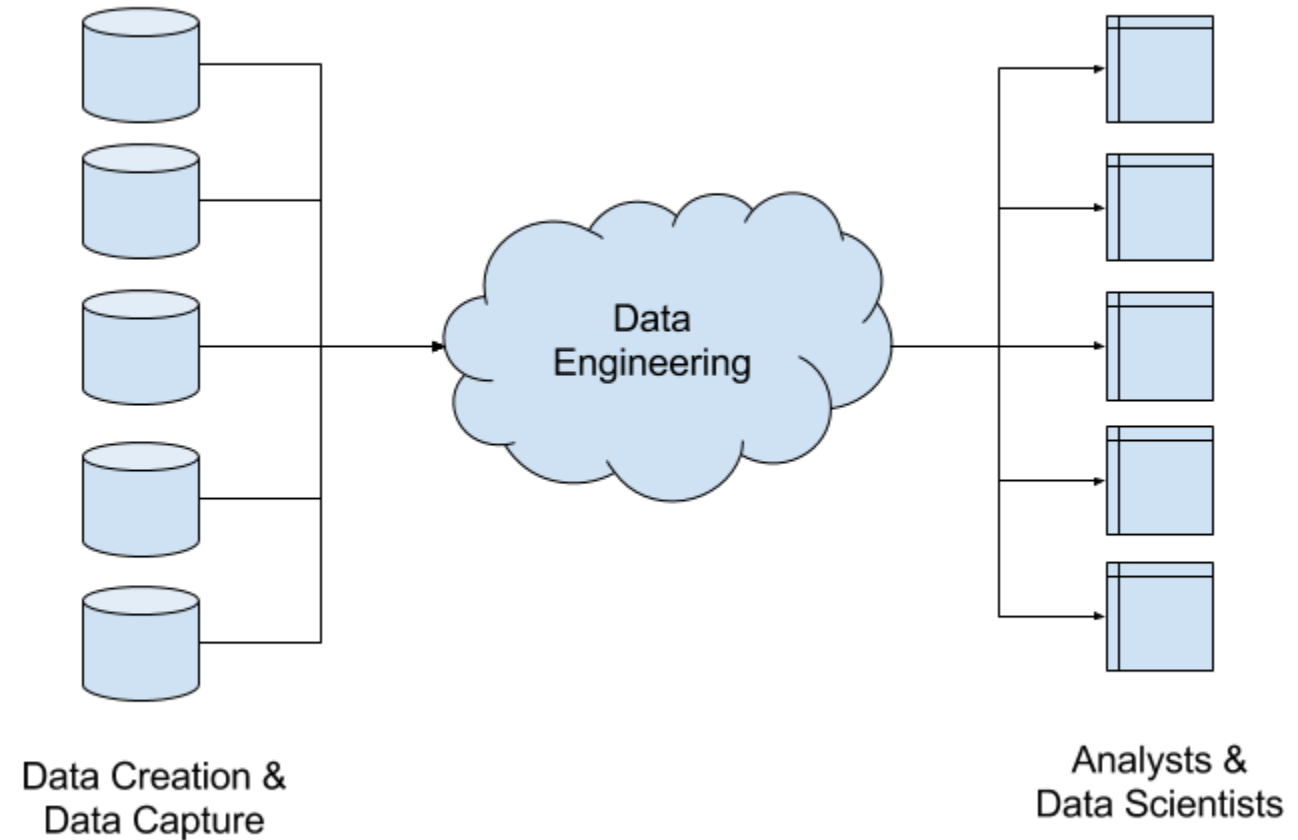
MOVE/STORE

COLLECT



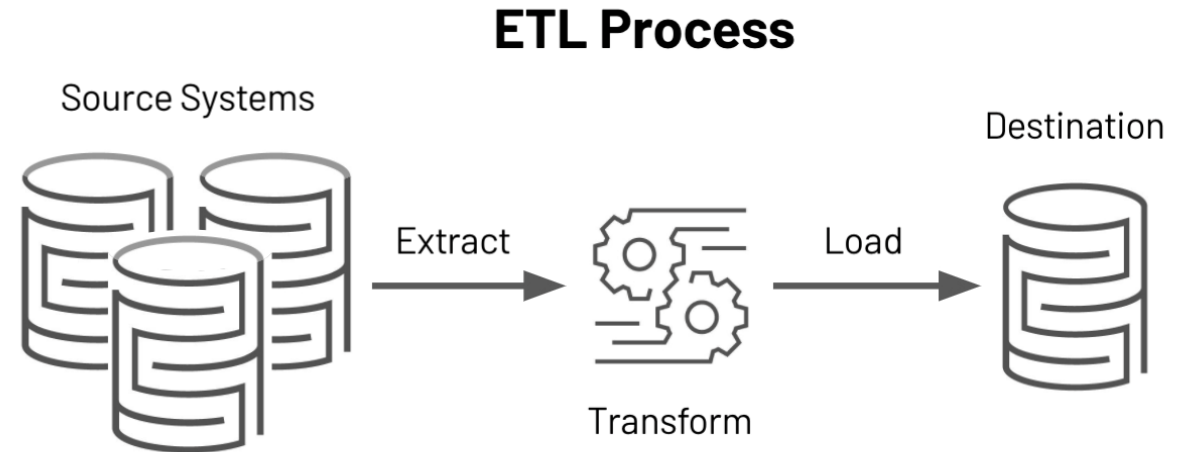
Data Engineering

- Μετατροπή και καθαρισμός δεδομένων
- Αποθήκευση δεδομένων σε βάσεις
- Δημιουργία Pipeline για την επεξεργασία δεδομένων



ETL – Extract Transform Load

- «Ωμά» Δεδομένα
 - Κακή ποιότητα
- «Σουλούπωμα» Δεδομένων
 - Ομοιόμορφα και τακτοποιημένα δεδομένα



Χρήσιμες Τεχνολογίες

A solid green horizontal bar.

- Python
- SQL
- REST APIs
- Hadoop / Spark

Ανάπτυξη Project



- Πρακτική εφαρμογή της θεωρίας
- Χρήσιμο για βιογραφικό
- Ανάπτυξη Programming Skills

Resources

- [Data Engineering](#)
- [Data Engineering IBM Course](#)
- [List of \(a lot of\) APIs](#)
- [Python Programming](#)
- [SQL Programming](#)