Kpler LNGmarket

Python, SQLAlchemy and Scrapy at Kpler

Jean Maynier
CTO

Kpler LNGmarket

Business : commodities markets

- **Domain**: trading / maritime transport
- **Customers**: physical traders / analysts
- **Needs**: follow cargoes flows and prices impact

Kpler LNGmarket

The first "**Cargo-Tracking**" system, bridging the gap between LNG **Shipping** and **Trading.**

Methodology : 4 **data layers**

- **"Meta" AIS** (10 networks)
- **Ports** Authorities
- Gas **Inventories**
- **Commercial** Database

Kpler LNGmarket

Step 1

Why python ?

- Simple, fast to prototype
- Data libraries (sqlAlchemy, Pandas, Scikit-learn)
- Data oriented community
- Scrapy: best scraping framework

Kpler LNGmarket

Step 2

Scrapy

- Simplify crawling
- Reusable extractors (xpath, css selectors)
- PaaS: Scrapinghub
- Blacklist: Tor, crawlera for IP rotation

Kpler LNGmarket

Step 3

Process raw data

- Clean and normalize data
  - Positions
  - Port calls
  - Contracts
  - Prices
- Python PubSub system
- Persist data : sqlAlchemy & Postgresql

Kpler LNGmarket

Step 4

Enrich data

Done

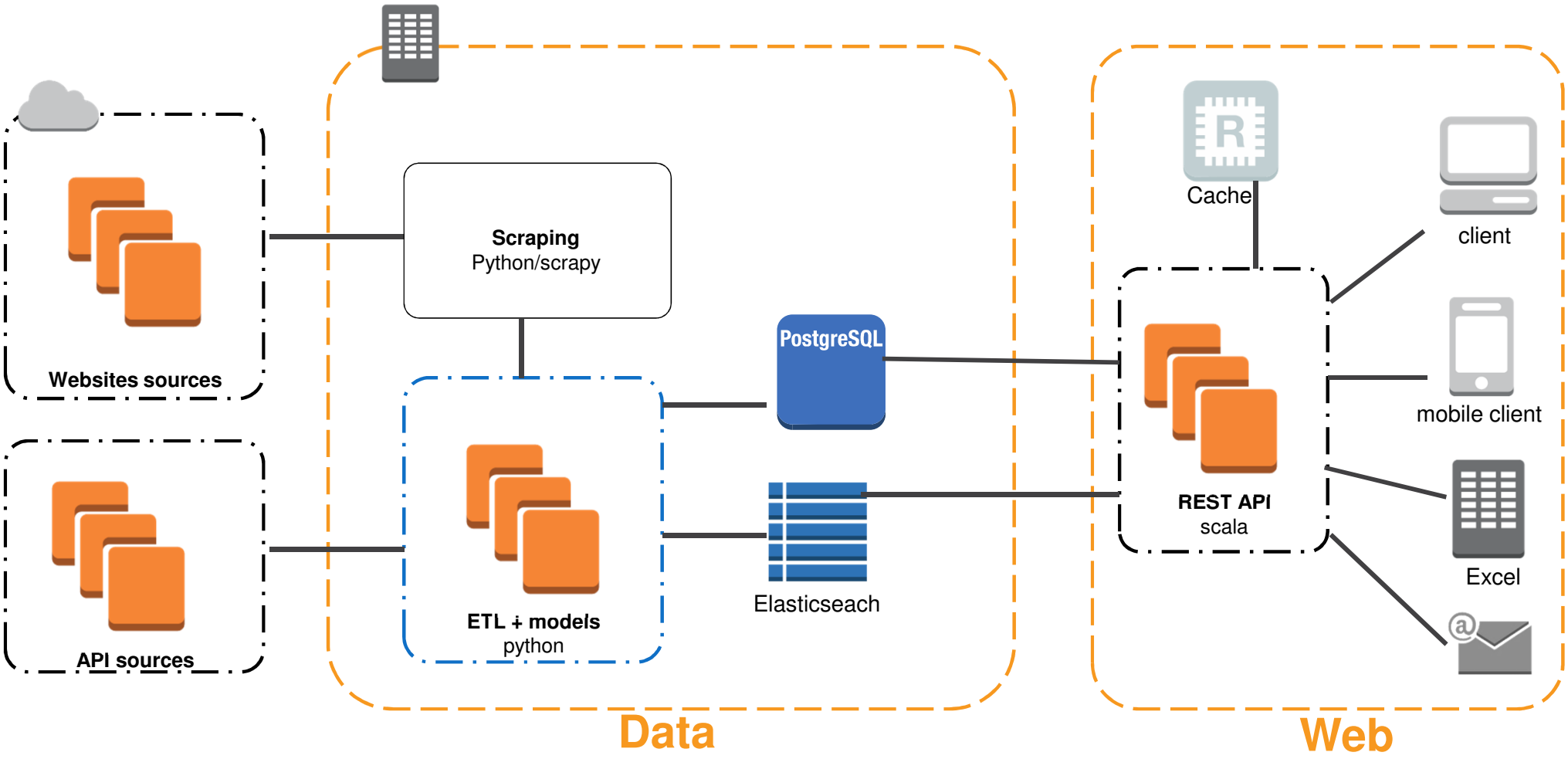
- Vessel dock/undock events
- Predict next destination
- Find buyer/seller
- Basic routing

Todo / WIP

- Estimate price
- Routing with weather
- Portfolio optimisation
- Fleet optimisation

Kpler LNGmarket

Architecture



Kpler LNGmarket

Metrics trends

Present

- DBs < 100Gb
- 50 sources
- 500 vessels
- Positions every 3min

Future

- 1-10 Tb
- 100 sources
- 10k to 100k vessels
- Position every 30s

Performance problem !

Kpler LNGmarket

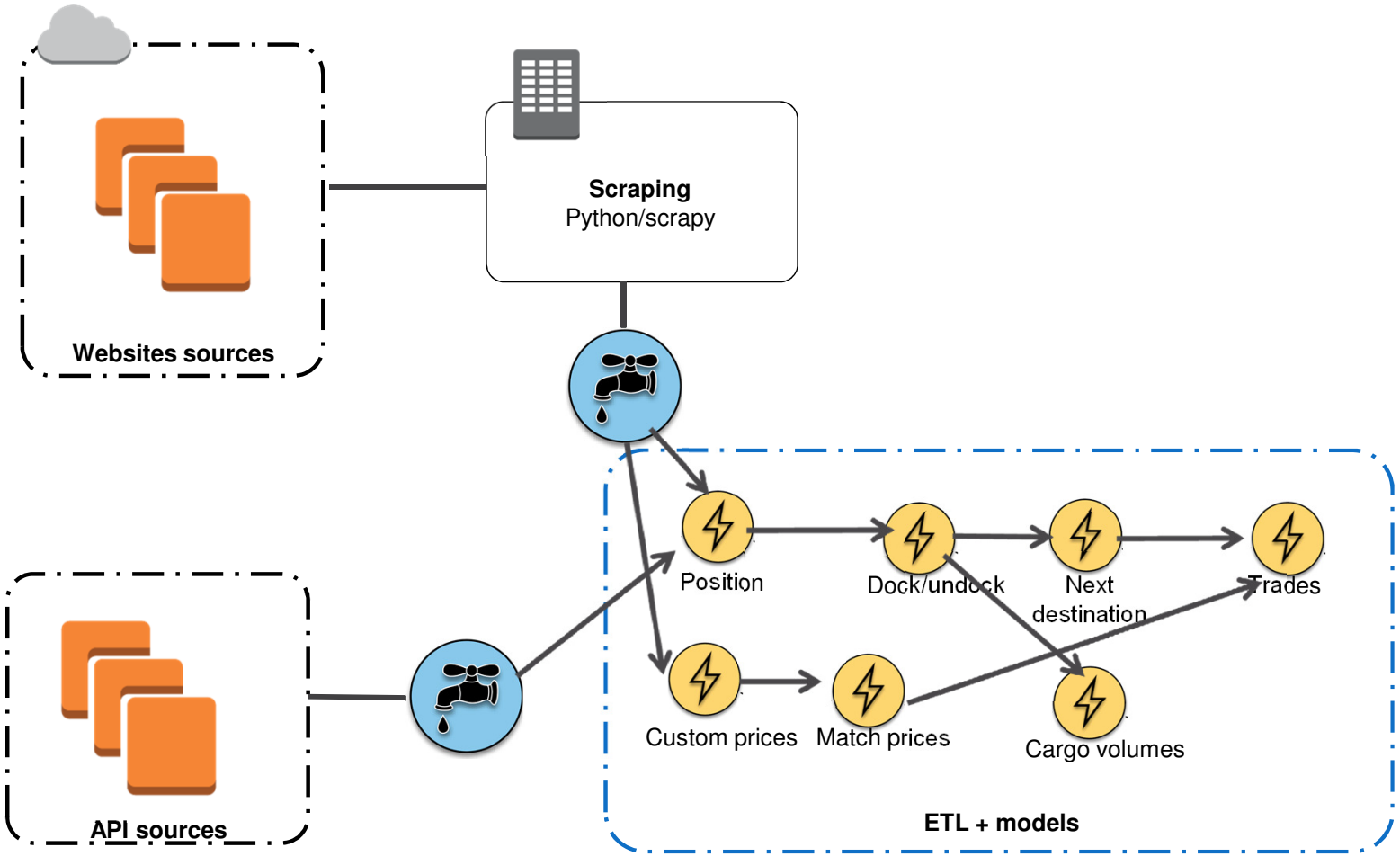
Step 5

Batch to data streaming

- Parallelization
- Granularity : item vs source
- Handle failure, no data loss, monitoring
- Akka streaming, Spark streaming, Celery, **Storm**

Kpler LNGmarket

Storm at Kpler: POC



Kpler LNGmarket

We are recruiting, join us !

jobs@kpler.com