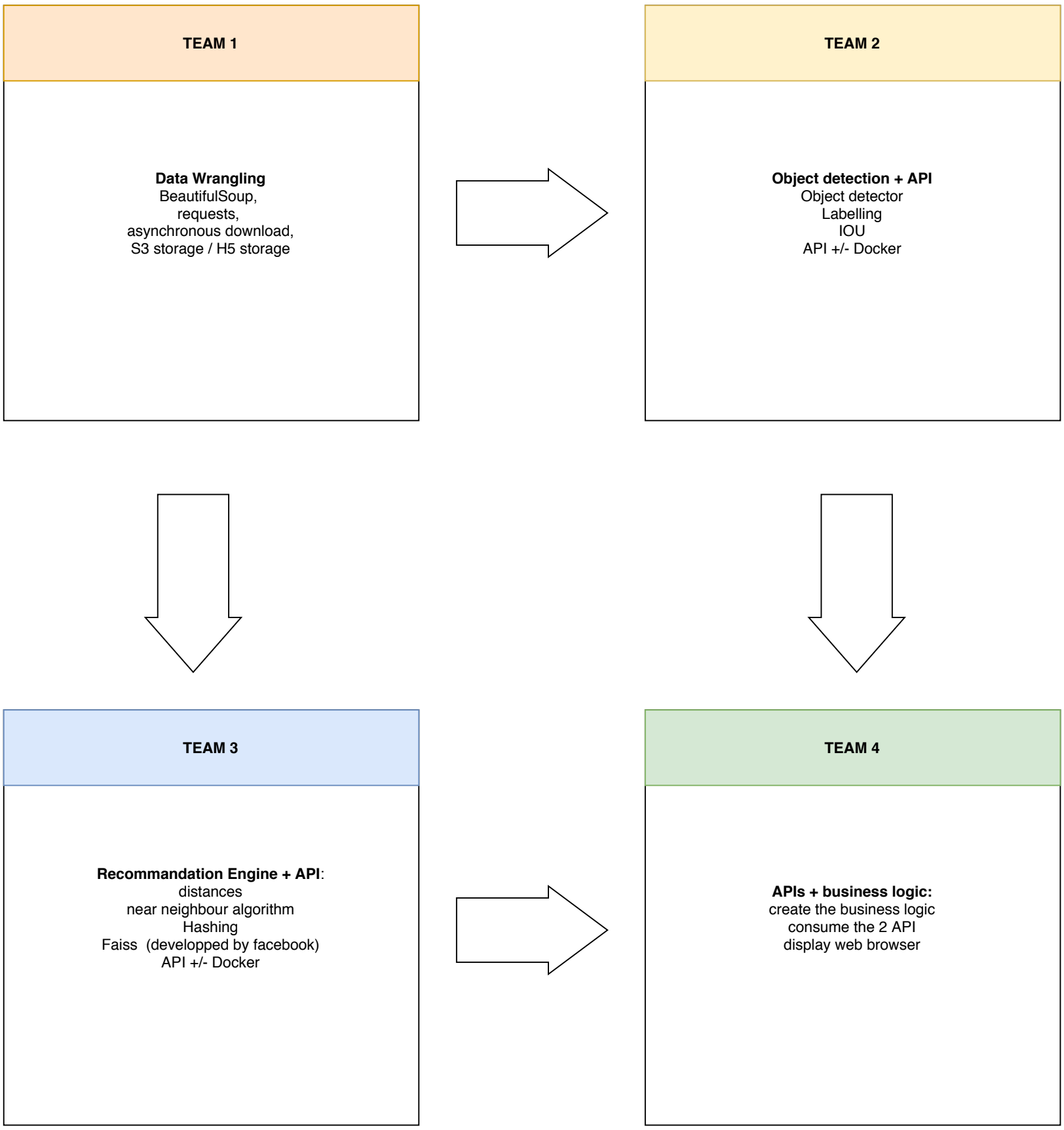
ARCHITECTURE



## PRODUCT OWNER

As described in the Scrum Guide, a Scrum Product Owner is responsible for maximizing the value of the product resulting from the work of the Development Team. How this is done may vary widely across organizations, Scrum Teams, and individuals.
The Product Owner is the sole person responsible for managing the Product Backlog. Product Backlog management includes:
Clearly expressing Product Backlog items.
Ordering the items in the Product Backlog to best achieve goals and missions.
Optimizing the value of the work the Development Team performs.
Ensuring that the Product Backlog is visible, transparent, and clear to all, and shows what the Scrum Team will work on next.
Ensuring the Development Team understands items in the Product Backlog to the level needed.

## DATA ENGINEER

Data Engineers are the data professionals who prepare the "big data" infrastructure to be analyzed by Data Scientists. They are software engineers who design, build, integrate data from various resources, and manage big data. Then, they write complex queries on that, make sure it is easily accessible, works smoothly, and their goal is optimizing the performance of their company's big data ecosystem.
They might also run some ETL (Extract, Transform and Load) on top of big datasets and create big data warehouses that can be used for reporting or analysis by data scientists. Beyond that, because Data Engineers focus more on the design and architecture, they are typically not expected to know any machine learning or analytics for big data.

Data Engineer skills

Database architectures

Hadoop-based technologies (MapReduce, Hive…)
Data modeling tools
Computer code: Python, C/C++ Java, Perl…
SQL technologies, NoSQL technologies
Data mining
Artificial Intelligence, Machine learning and Deep Learning: It is important they understand the algorithms to work with Data Scientists but they obviously do not need to master it like Data Scientists.
UNIX, Linux, Solaris and MS Windows

## SCRUM MASTER

**The Scrum Master serves the Product Owner in several ways, including:**

Ensuring that goals, scope, and product domain are understood by everyone on the Scrum Team as well as possible.
Finding techniques for effective Product Backlog management.
Helping the Scrum Team understand the need for clear and concise Product Backlog items.
Understanding product planning in an empirical environment.
Ensuring the Product Owner knows how to arrange the Product Backlog to maximize value.
Understanding and practicing agility.

**The Scrum Master serves the Development Team in several ways, including:**

Coaching the Development Team in self-organization and cross-functionality.
Helping the Development Team to create high-value products.
Removing impediments to the Development Team's progress.
Facilitating Scrum events as requested or needed.
Coaching the Development Team in organizational environments in which Scrum is not yet fully adopted and understood.

## DATA SCIENTIST

The data scientist rephrases a business issue (**predictive sales**, **predictive maintenance**, **customers segmentation**, **fraud detection**) into a scientific issue (supervised/non-supervised classification, regression, recommendation, optimization).

The tasks of a data scientist are complex and strategic, which is the reason why he or she needs to have **strong skills in computing, mathematics and statistics**. Moreover, curiosity, an engineering background and high-level technical skills (R language, Pyhton and other **Big Data technologies** like Spark, Hadoop), are also important ingredients.

Moreover, in the visualizing and analyzing phases, the data scientist has to take into account both raw data and his knowledge of business in order to extract the most representative features (« feature engineering »). **It is necessary to have knowledge of products and business issues and challenges to obtain relevant and reliable predictions.**

The last (but not least) skill of a data scientist is communication. Indeed, he or she has to present results to IT managers, and even top management. He or she has to be able to communicate not only the predictions themselves but also the hypothesis on which they are based.