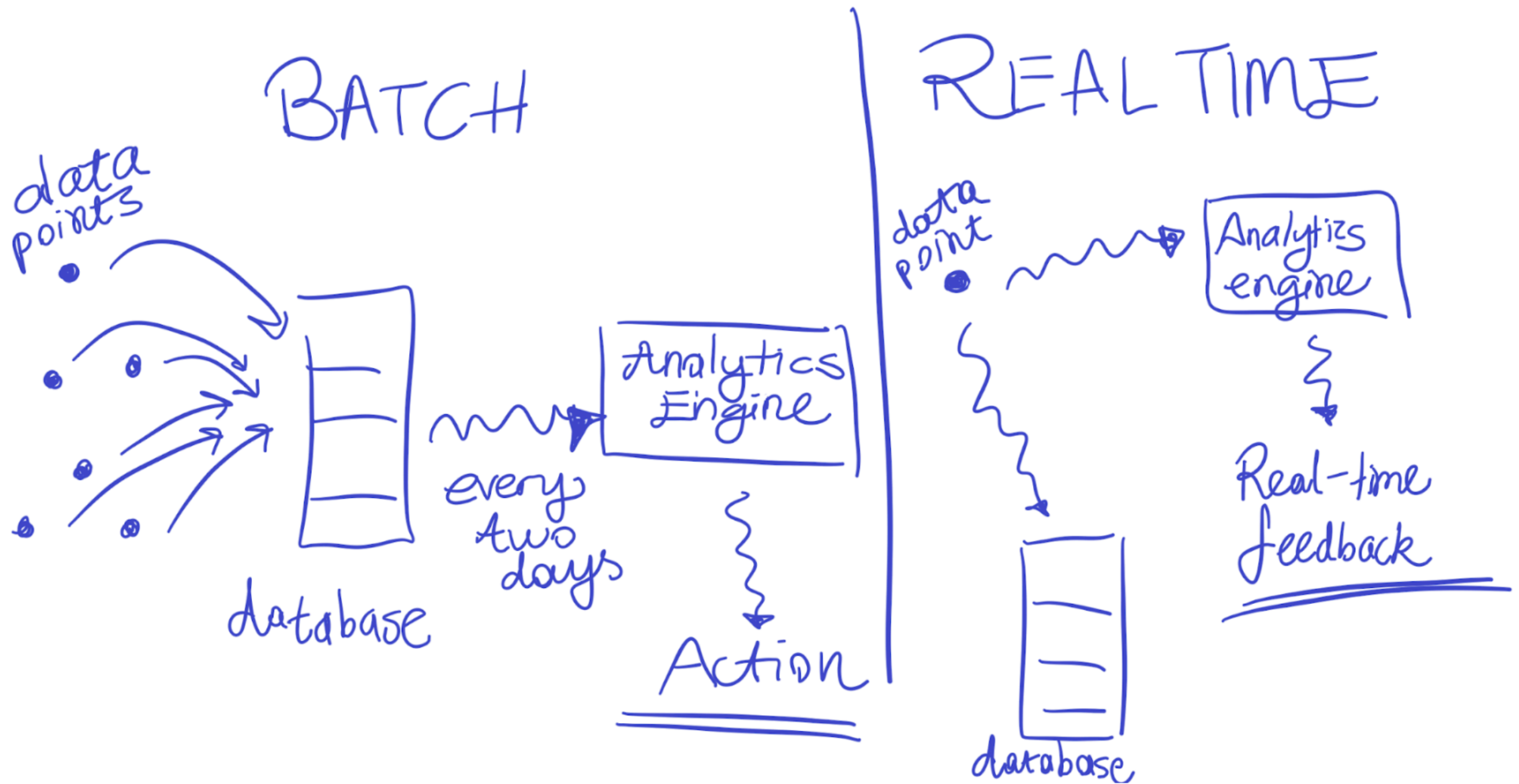


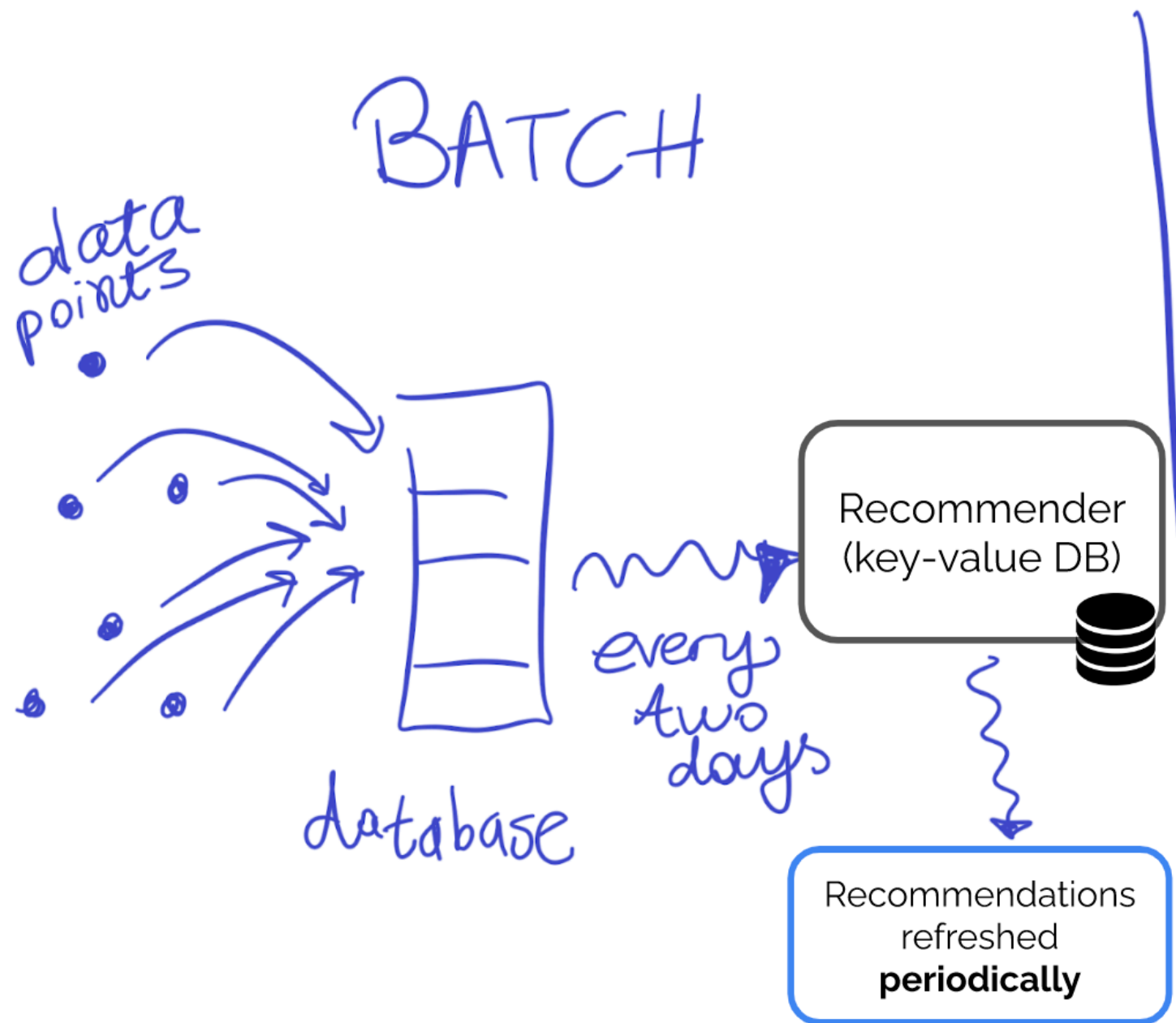
Recommender Systems in Industry: Focus on the System Design

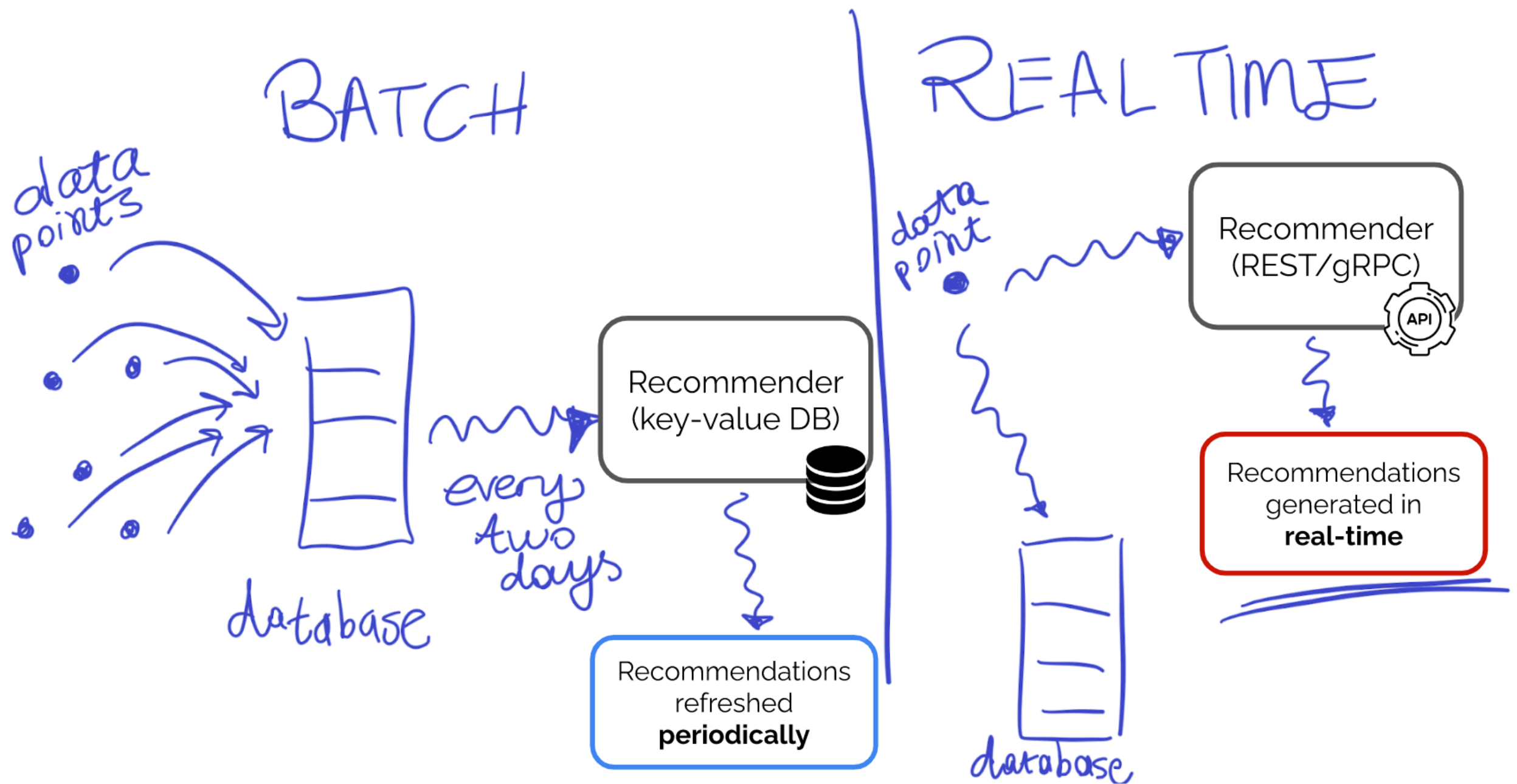
Marie Al-Ghossein

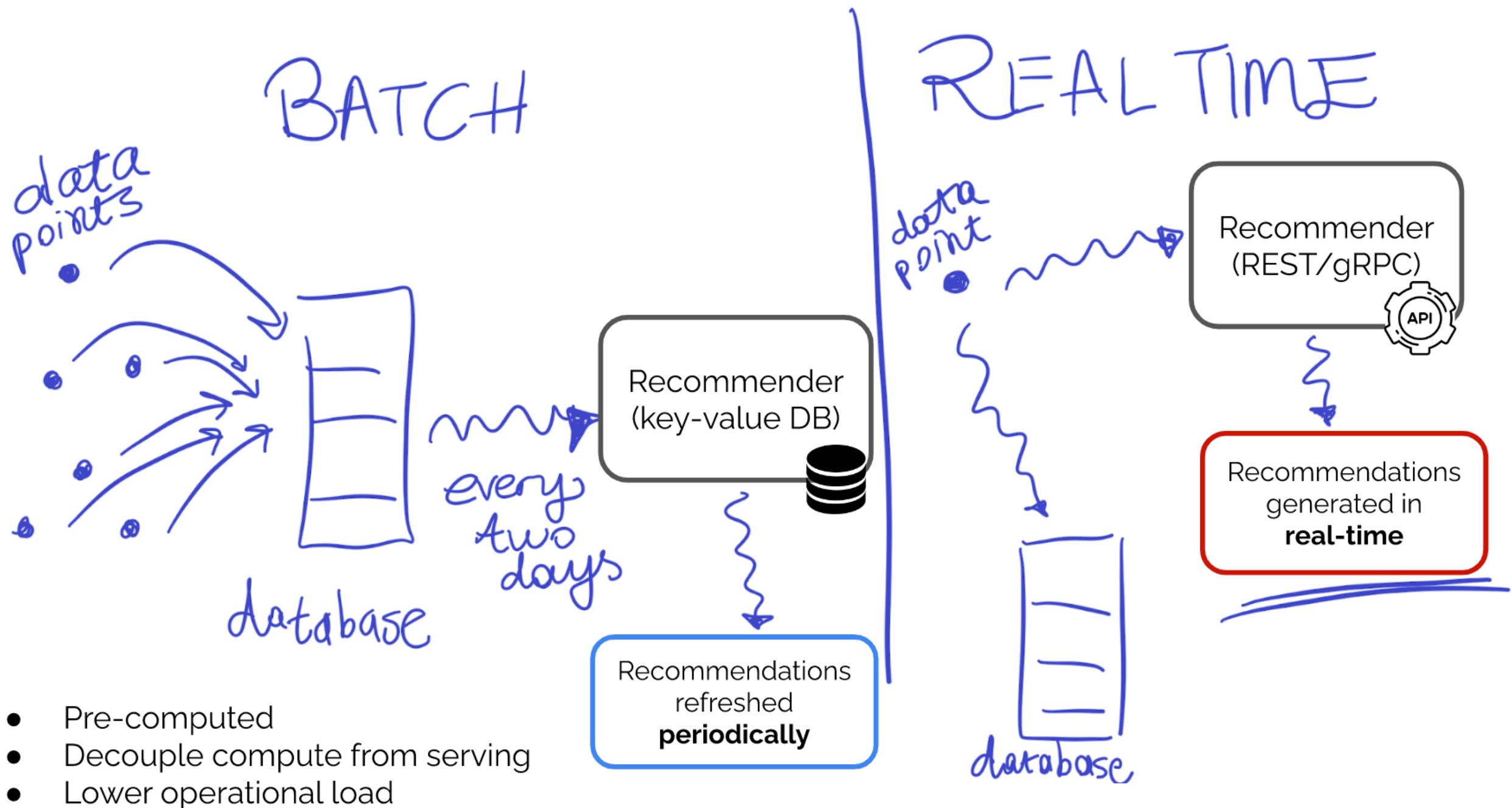


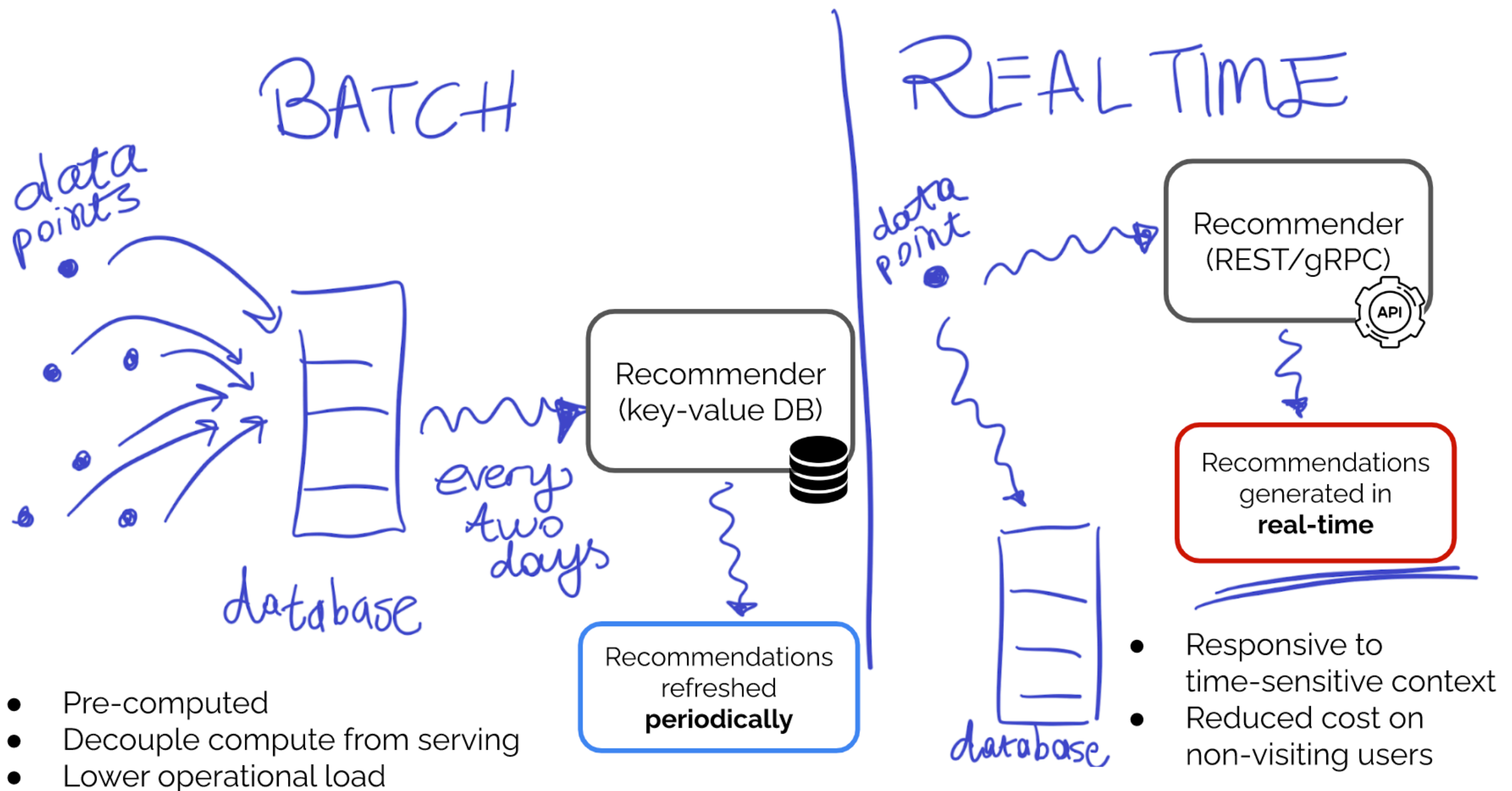
June 2, 2025











Online



Offline

Online

Offline

- Host batch processes such as training, index/graph building
- Load data into feature stores

Online

- Uses artifacts from offline environment to serve requests
- **Candidate retrieval** and **ranking**

Offline

- Host batch processes such as training, index/graph building
- Load data into feature stores

Retrieval

Ranking

- Fast but coarse
- Searches millions of items to get hundreds of candidates
- Approx NN, Graphs, etc.

Retrieval

Ranking

- Fast but coarse
- Searches millions of items to get hundreds of candidates
- Approx NN, Graphs, etc.

- Slower but more precise
- Ranks hundreds of candidates
- Adds more features
- Classification or learning-to-rank

Retrieval

Ranking

