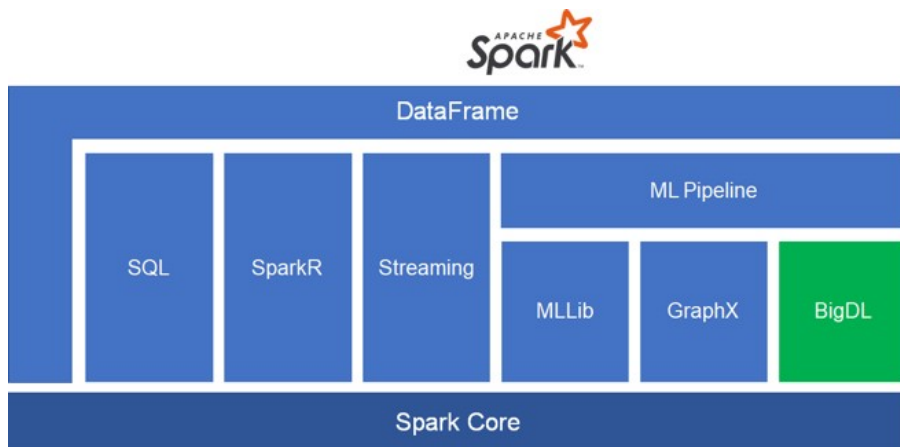




BigDL: A Distributed Deep Learning Framework for Big Data

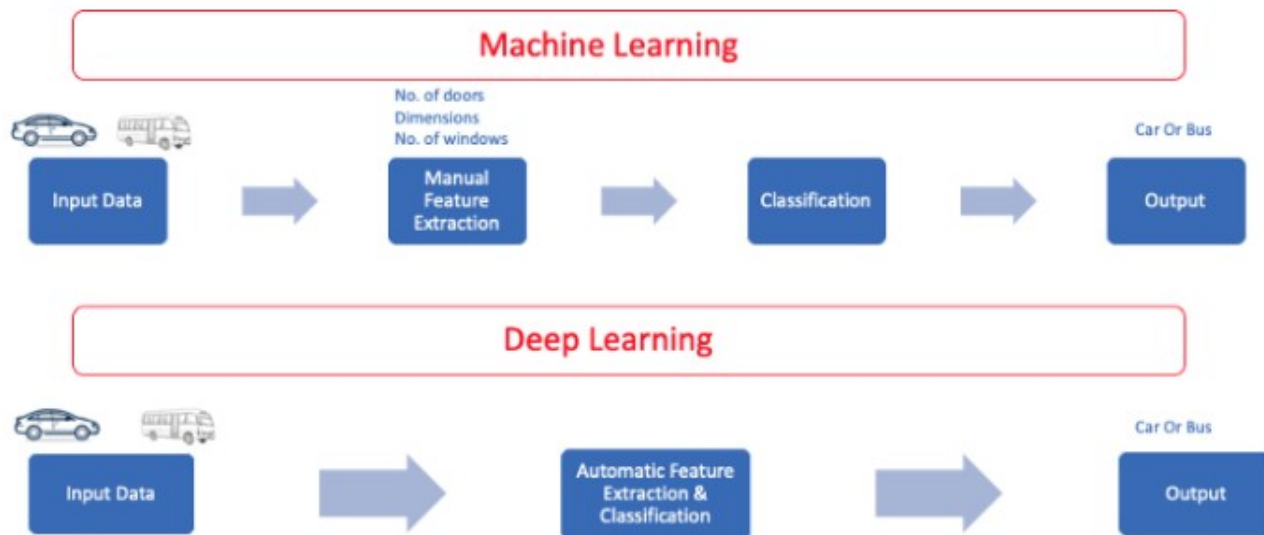
BigDL

- BigDL is a distributed deep learning library for Apache Spark; with BigDL, users can write their deep learning applications as standard Spark programs, which can directly run on top of existing Spark or Hadoop clusters.

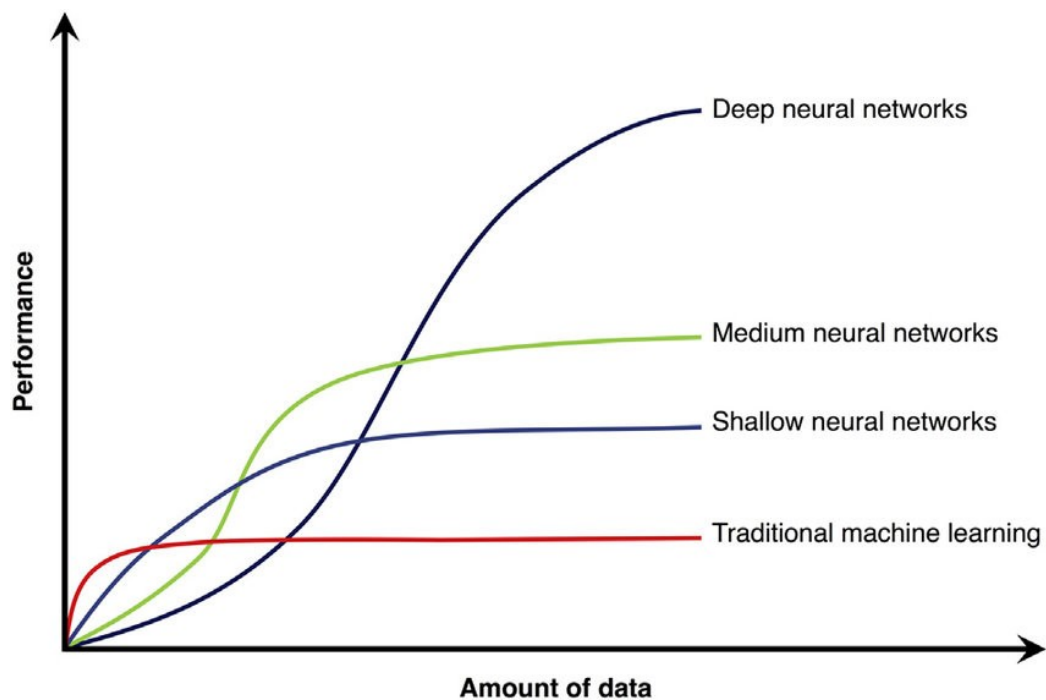


Why Deep Learning

- High level feature extraction
- Neural networks



Why Deep Learning



Big Data problem

- High model accuracy = bigger training datasets
- Bigger datasets = more complex pipelines
- Most of real-world data is non-labeled and not pre-processed

Big Data solutions

- Separate Apache Spark and GPU clusters
- Not preferred
- Back-and-forth development and debugging, incremental model update, deployment

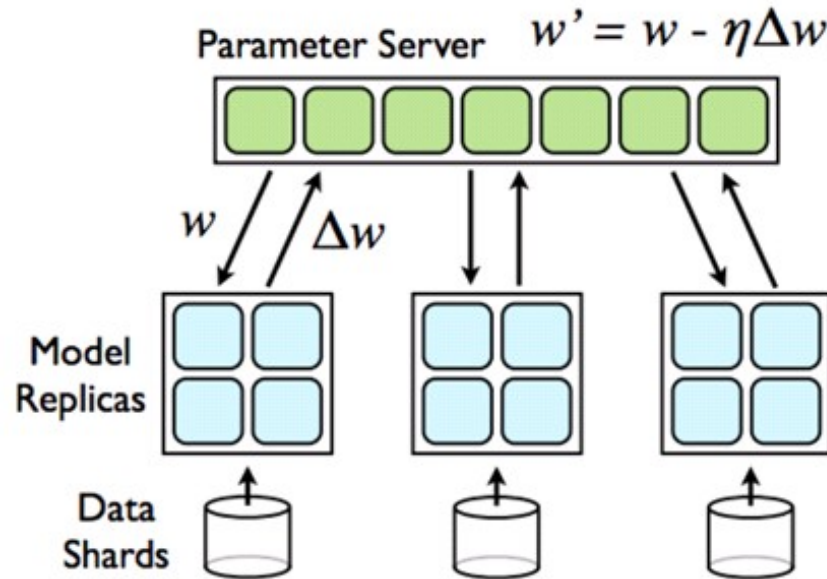
Big Data solutions

- “Connector approach” – interface between the two frameworks
- Problem – huge overhead
- Impedance mismatch

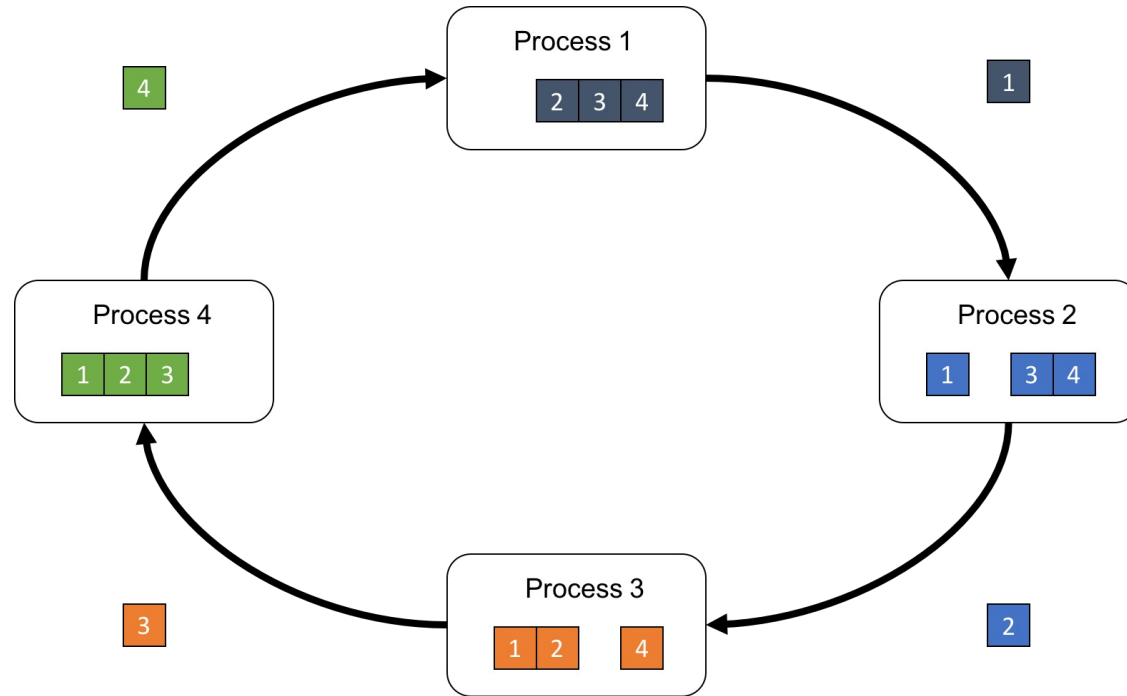
Big Data solutions

- BigDL approach - directly implements the distributed deep learning support in the big data system
- Implements parameter server style architecture - AllReduce
- Depends on Spark operations and data models
- Eliminates impendence mismatch problems

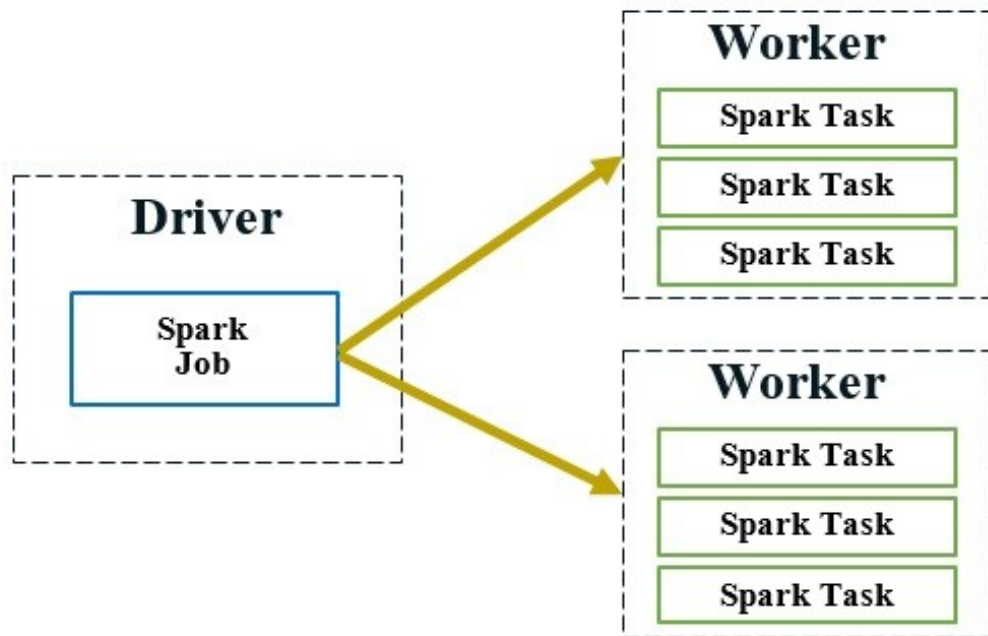
Distributed deep learning – Param server



Distributed deep learning – Ring AllReduce



Spark execution model

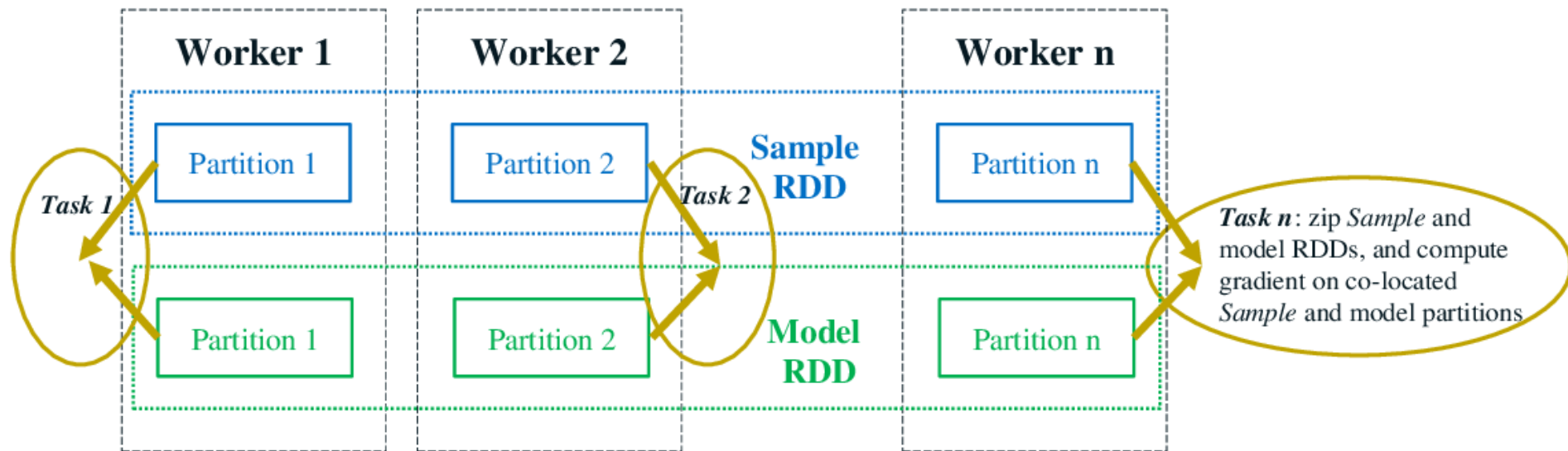


BigDL distributed deep learning

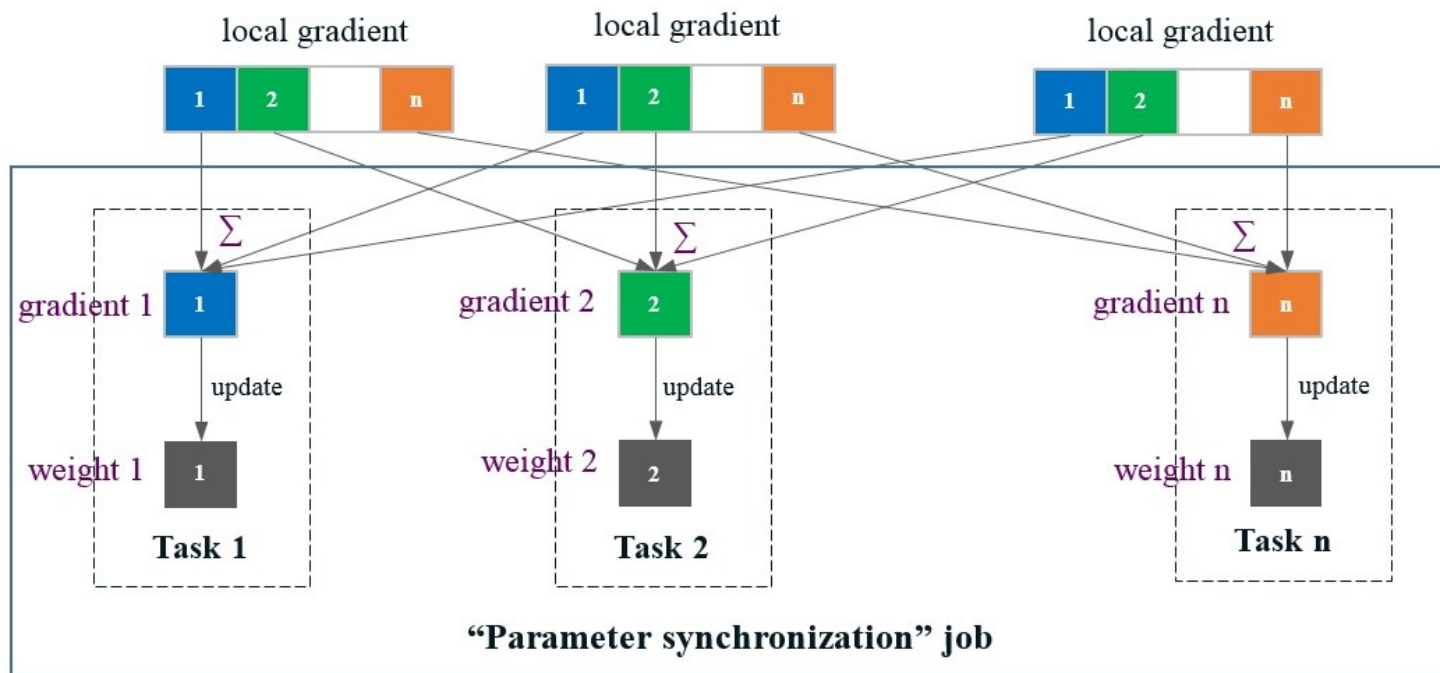
Algorithm 1 Data-parallel training in BigDL

```
1: for  $i = 1$  to  $M$  do
2:   //"model forward-backward" job
3:   for each task in the Spark job do
4:     read the latest weights;
5:     get a random batch of data from local Sample partition;
6:     compute local gradients (forward-backward on local model
       replica);
7:   end for
8:   //"parameter synchronization" job
9:   aggregate (sum) all the gradients;
10:  update the weights per specified optimization method;
11: end for
```

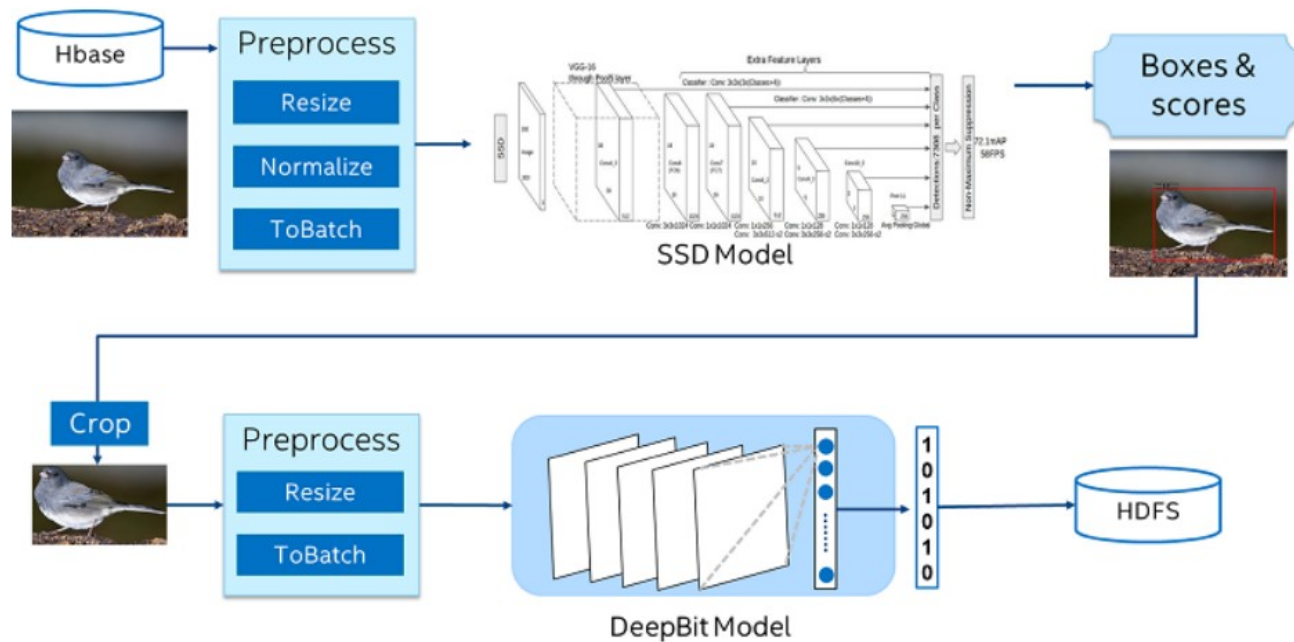
BigDL distributed deep learning



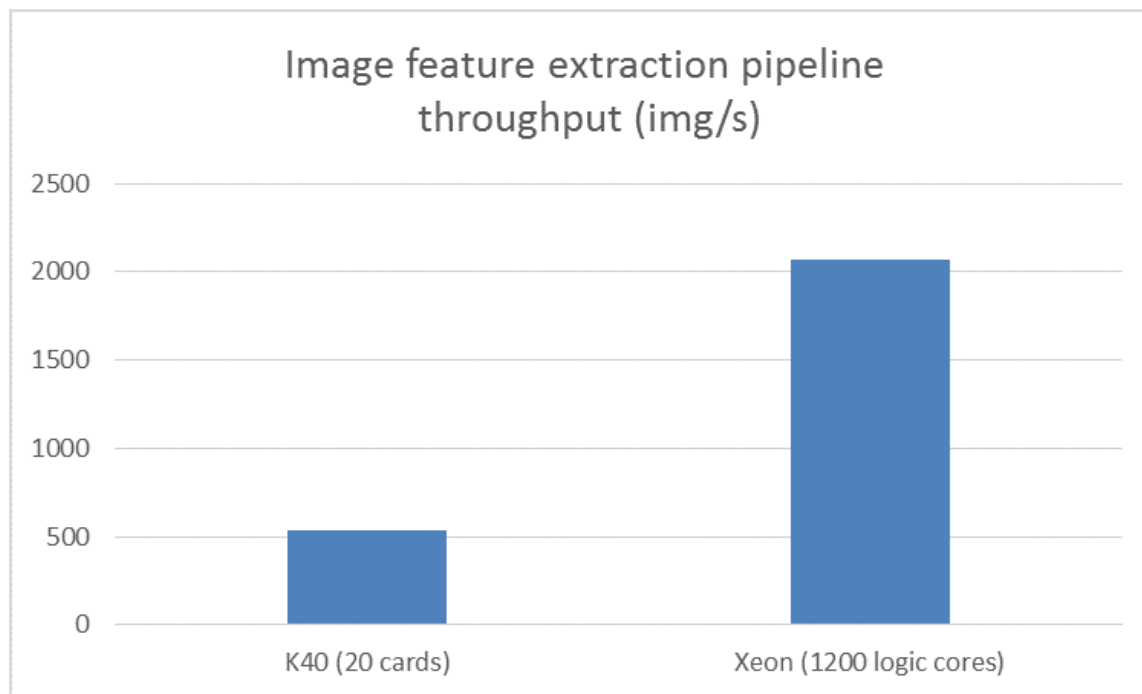
“Parameter synchronization” job



BigDL use case – JD.com



Performance differences



Hvala na pažnji!

- Diskusija