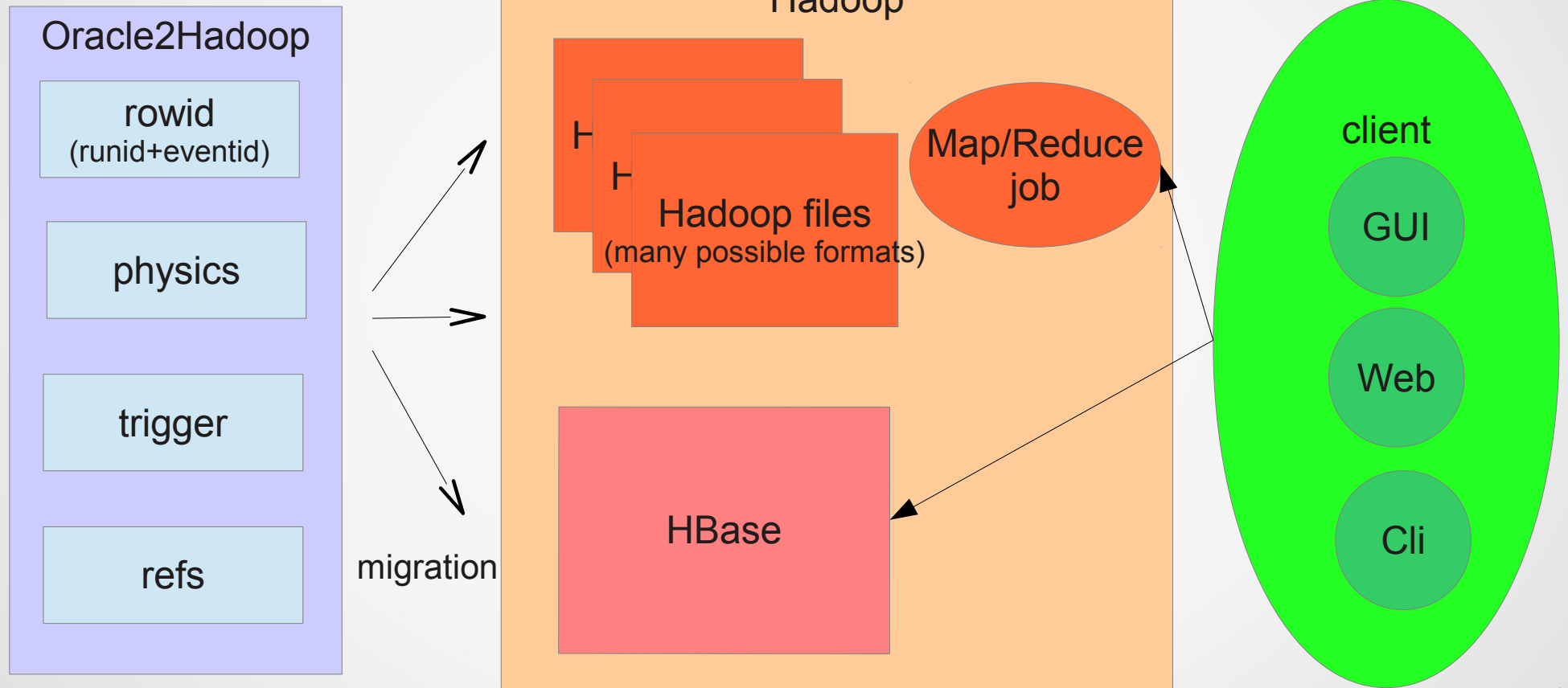


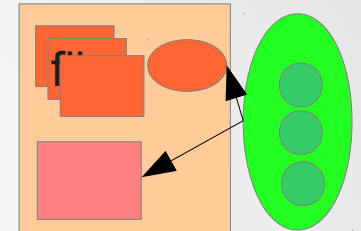
Hadoop/HBase evaluation/migration



Hadoop/HBase evaluation/migration

- several possibilities:

- HBase
 - “interactive”, but less control over storage & Map/Reduce
- direct HDFS files
 - batch-like, but choice of file formats and powerful Map/Reduce
 - easy upload/replication (files could be created outside and copied in)
 - many file formats directly supported, other could be added (maybe even Root files ?)
 - e.g. TFile supports per-collection compression, independent C++ IO
- have to evaluate size/speed/access
- all data may not be in the same technology
 - more index-like data in HBase, other in files
 - HBase may be updated with feedback about usage (a bit like a cache with recent results)
- should use the code running on the Hadoop (Map/Reduce)
- many interface language available



Hadoop/HBase evaluation/migration

- let's spend next 3-4 month to do a “wide evaluation”
 - try different architectures, file formats, languages,... on relatively small data
- then concentrate on one-or-two best candidates and do a real-life-like evaluation
 - try to optimize it
- then (if success) make it production quality
- in parallel, I'd like to try to put in our (existing) monitoring logs
 - which are already quite big and slow to analyse on Oracle
 - ideal case for Hadoop as you don't search on them, but execute Map/Reduce-like tasks already now
 - may include creation and storage of profiling histograms
- ServiceCatalog DB already in HBase (not yet on the IT servers) - very small, but more complex - useful to try interfaces

