

Ideas for PROOF

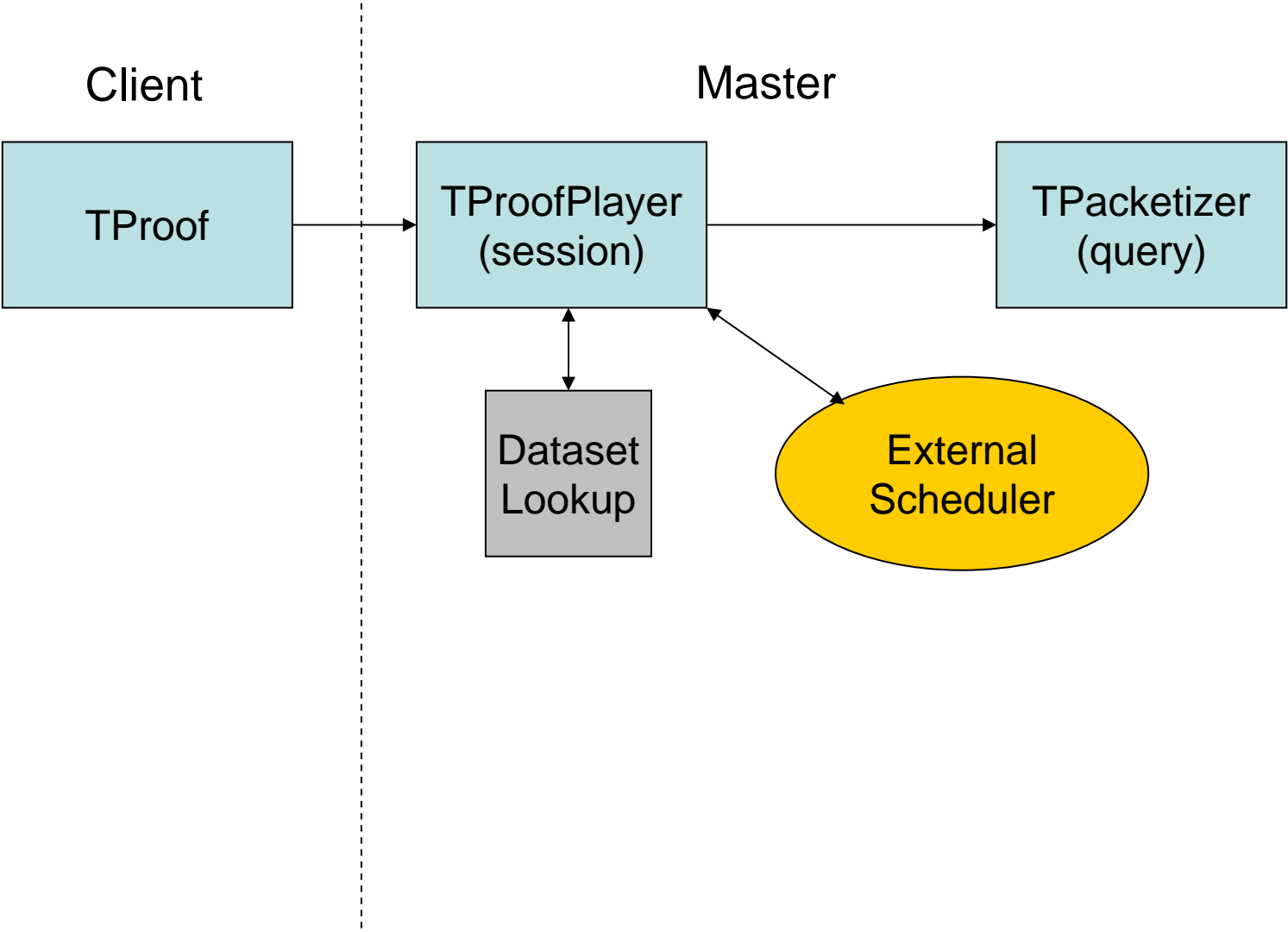
Jan Iwaszkiewicz

- Adding a scheduler to PROOF
- Idea for a more flexible architecture
 - exploiting client machines
 - Using multiple clusters

Why scheduling?

- Controlling resources and how they are used
- Improving efficiency
 - assigning to a query those nodes that have data which needs to be analyzed.
- Implementing different scheduling policies
 - e.g. providing equal access
- Efficient use even in case of saturation

Starting a query



Changes in PROOF

- Adding and removing slaves during a session (when starting a query)
 - Problems with remembering all the actions of a user
- Dataset lookup has to be done before the query starts
- Scheduler shall assign nodes for a given query
- Integration with MAUI scheduler?

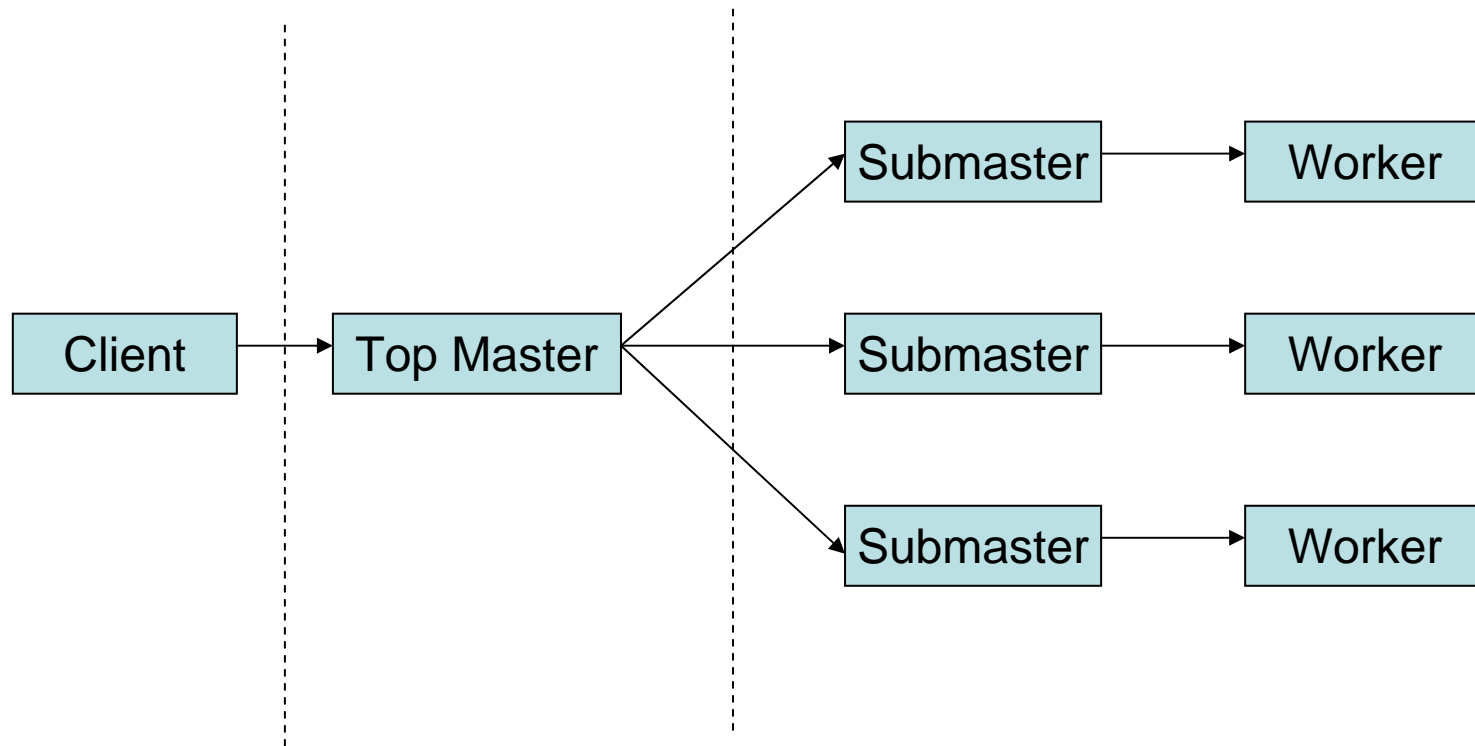
PROOF architecture

- Exploiting CPU power of local machine when working with PROOF
- Running PROOF on several clusters

3-level PROOF

- Top master
- Submasters for geographic domains
 - Packetizers are located on this level
- Workers

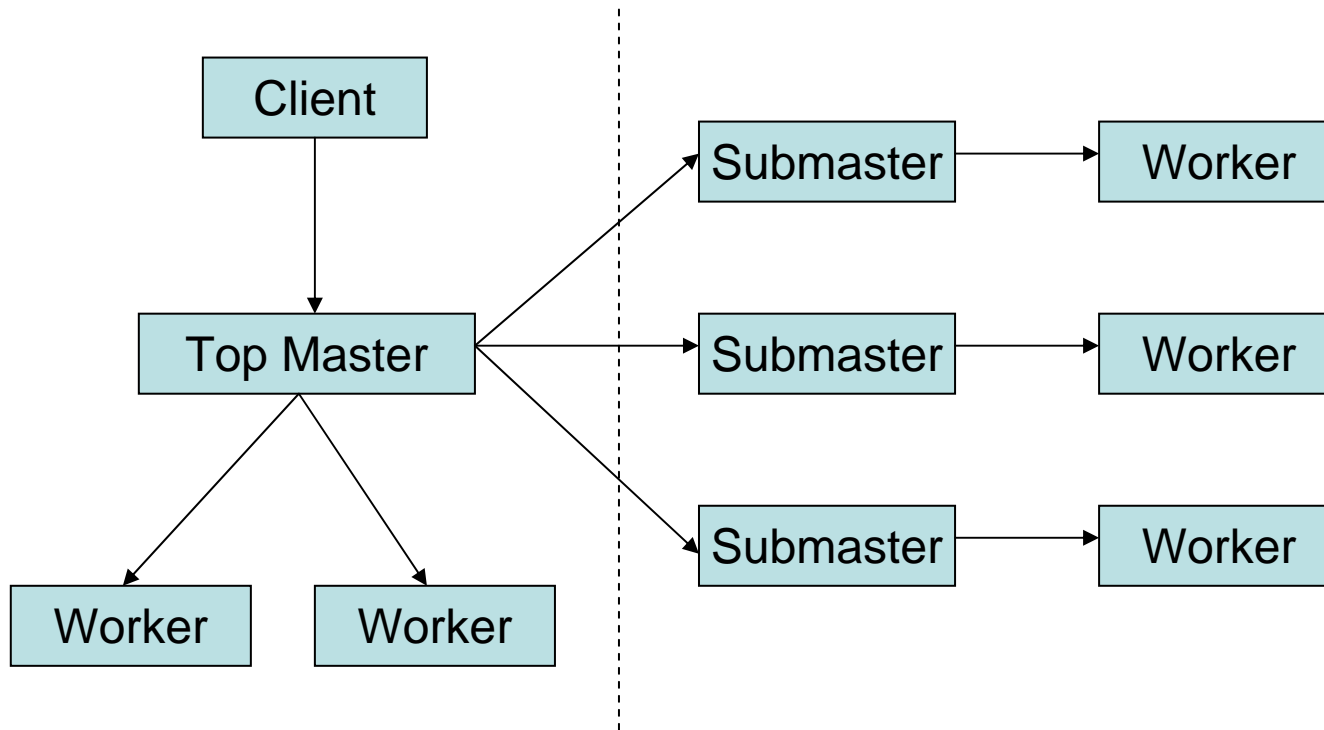
3 levels of PROOF



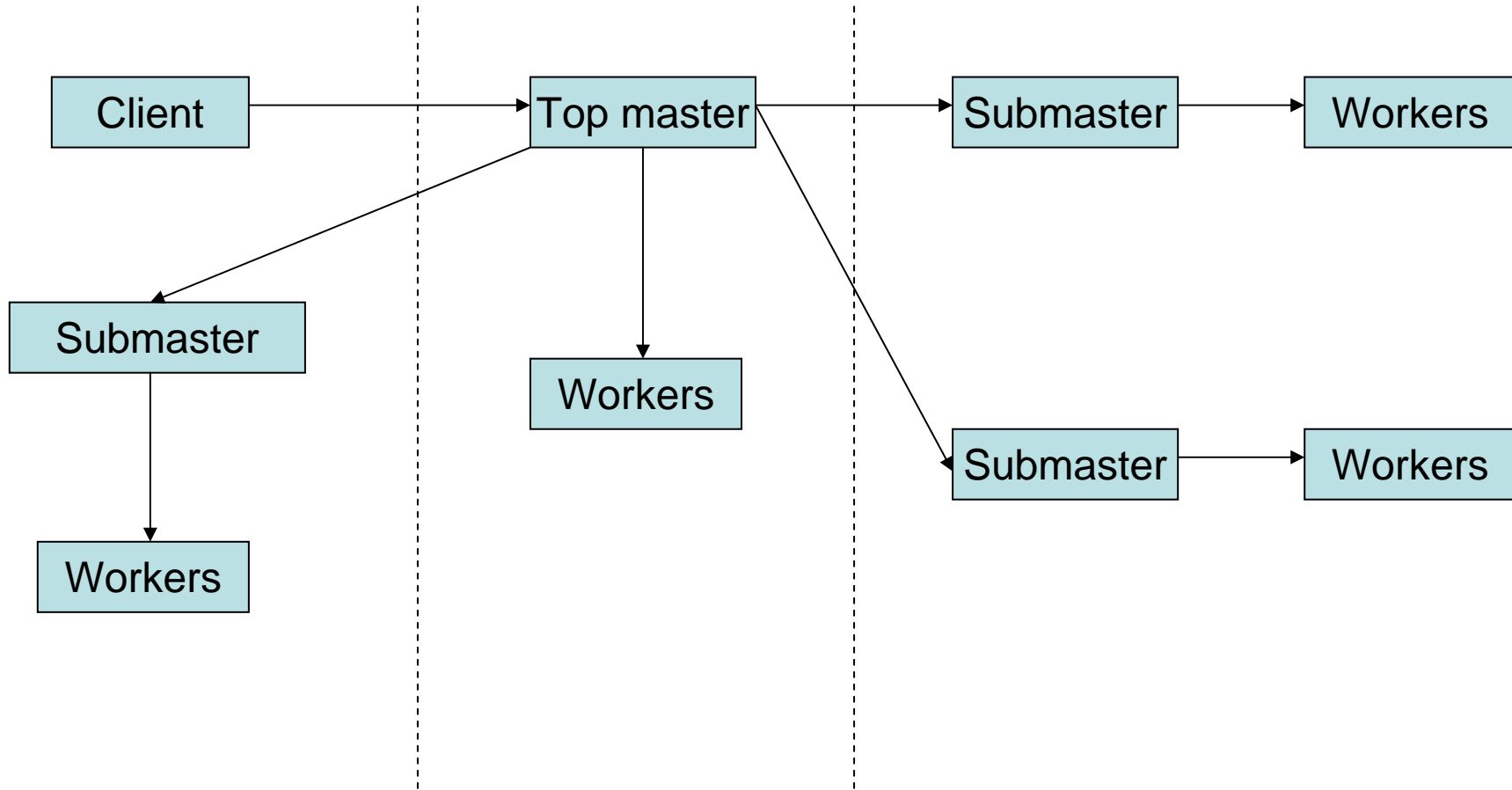
Tree-like Architecture

- Ideal: Any worker - leaf can be substituted by a subtree – master and a set of workers
- Problems:
 - Packetizers are designed to coordinate analysis of datasets. The datasets for different packetizer should be defined before the query starts

Proposed 1



Proposed 2



Changes to PROOF

- Packetizers
 - Needed additional level
- Dataset lookup
 - Will be moved outside packetizer anyway