A MapReduce Style Framework for Searches and Computations on Parallel Tree Structures

Abhinav Sarje Srinivas Aluru

February 6, 2009

Aim of this framework is to ...

- Provide a simple and generalized model for tree operations.
- Hide parallelism of the tree from the user.
- Minimize the number of user defined functions.

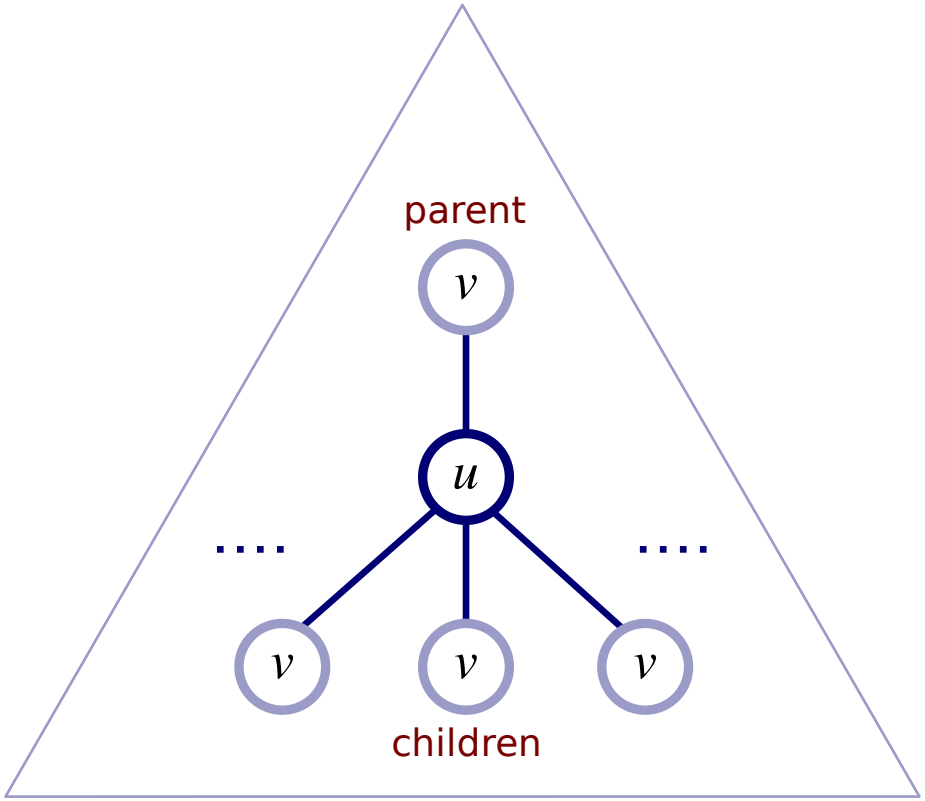
Google's MapReduce

The user defined functions:

$$\texttt{map}(k1, v1) \longrightarrow \text{list}(k2, v2)$$
$$\texttt{reduce}(k2, \text{list}(v2)) \longrightarrow \text{list}(v2)$$

- The input is a list of key-value pairs.
- `map` produces an intermediate list of key-value pairs.
- `reduce` merges all pairs with the same key into possibly smaller list of values.

A Tree Data Structure



$$u = \langle k_u, X_u \rangle$$

TreeSearch: searches on a tree structure

$$\texttt{treeSearch}(\text{list}(\mathcal{K})) \longmapsto \text{list}(\text{list}(v))$$

- $\text{list}(\mathcal{K}) = (\mathcal{K}_1, \mathcal{K}_2, ..., \mathcal{K}_n)$ is a list of n search items.
- $\text{list}(\text{list}(v)) = (\text{list}(v_1), ..., \text{list}(v_n))$ is the result
- $\text{list}(v_i)$ is the result for the search item $\mathcal{K}_i$

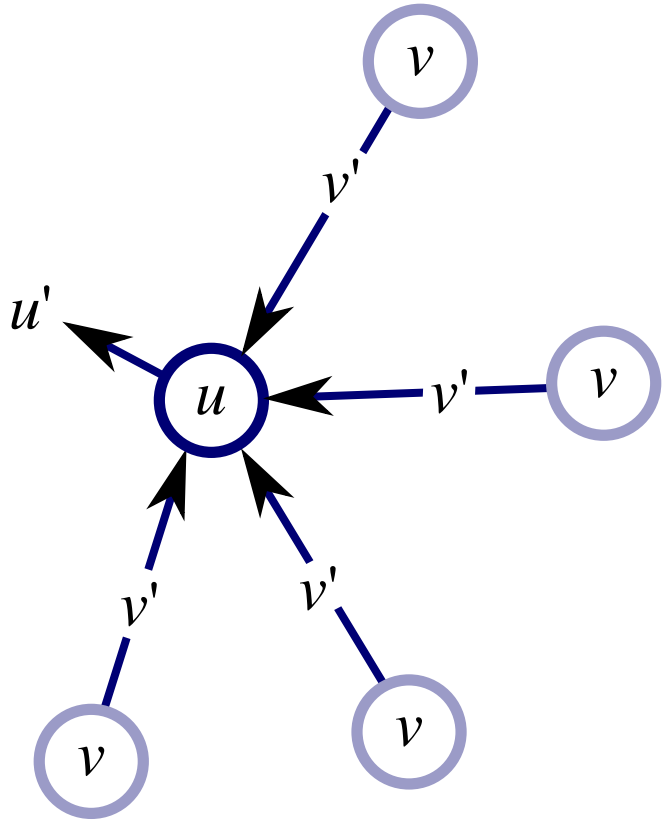
User defined function:

$$\texttt{select}(u, \mathcal{K}) \longrightarrow \text{list}(v) \mid \texttt{FOUND}$$

- $\mathcal{K}$ is the search item
- $\text{list}(v)$ is the list of desired children of the node u for the search to descend to
- `FOUND` specifies if u is a desired result node

TreeCompute: for computations on a tree structure

$$\texttt{treeCompute}(u) \longmapsto u'$$



TreeCompute: for computations on a tree structure

User defined functions:

$$\texttt{generate}(u) \longrightarrow \langle \text{list}(v), \texttt{DEPENDENCY} \rangle$$
$$\texttt{map}(u, v) \longrightarrow u'$$

- `map` is applied on each node from $\text{list}(v)$ to the updated node u in an arbitrary order.
- Primed variables represent updated values of their corresponding variables.

Generate

$$\texttt{parent}(u) \longmapsto v$$
$$\texttt{children}(u) \longmapsto \text{list}(v)$$

`DEPENDENCY` flag indicates if dependencies should be respected.