CLOUD COMPUTING APPLICATIONS

MapReduce Programming Model

Roy Campbell & Reza Farivar

What is MapReduce?

- MapReduce
  - Programming model from LISP
- Many problems can be phrased this way
- **Easy** to distribute
  - Hides difficulty of writing parallel code
  - System takes care of load balancing, dead machines, etc.
- Nice retry & failure semantics

Programming Concept

- Map
  - **Perform** a function on **individual values** in a data set to create a **new list** of values
- Reduce
  - **Combine** values in a data set to create a new **value**

SQUARE X = X * X **MAP** SQUARE [1,2,3,4,5] RETURNS [1,4,8,16,25]	SUM= [All Elements in Array, Total+=] REDUCE [1,2,3,4,5] RETURNS 15

MapReduce Programming Model

Input & Output: Each a set of key/value pairs

Programmer specifies two functions:

MAP (IN_KEY, IN_VALUE)

LIST(OUT_KEY, INTERMEDIATE_VALUE)

- Processes input key/value pair
- Produces intermediate pairs

MapReduce Programming Model

Input & Output: Each a set of key/value pairs

Programmer specifies two functions:

REDUCE (IN_KEY, IN_VALUE)

LIST(OUT_KEY, INTERMEDIATE_VALUE)

- Combines all intermediate values for a particular key
- Produces a set of merged output values (usually just one)