



CLOUD COMPUTING APPLICATIONS

MapReduce Introduction

Roy Campbell & Reza Farivar

Motivation: Large Scale Data Processing

- Many tasks composed of processing lots of data to produce lots of other data
- Large-Scale Data Processing
 - Want to use 1000s of CPUs
 - But don't want hassle of managing things

Motivation: Large Scale Data Processing

- MapReduce provides
 - User-defined functions
 - Automatic parallelization and distribution
 - Fault tolerance
 - I/O scheduling
 - Status and monitoring