

1. Word Count with MPI

An MPI implementation of Word Count problem

1.1. Introduzione

Word Count consiste nel determinare in un testo quante occorrenze ci sono di ogni parola presente. Generalmente si presenta in applicazioni in cui è necessario mantenere le parole inserite in limiti definiti (es. nella narrativa, il numero di parole definisce la categoria dello scritto).

Benchè estremamente semplice nella sua definizione, Word Count rappresenta una sfida nel campo della programmazione distribuita a causa dell'enorme taglia dei problemi che possono presentarsi: pensiamo infatti a situazioni in cui i documenti da esaminare hanno dimensioni arrivanti anche a 100 TB, in cui un'esecuzione sequenziale arriva a richiedere mesi se non anni di computazione.

E' naturale quindi pensare a Word Count come un problema di programmazione distribuita, in cui la distribuzione dell'input tra i nodi di un cluster può significativamente diminuire il tempo necessario, passando da anni al tempo per un caffè :coffee:

1.2. Soluzione proposta

1.2.1. Architettura

```
graph TD
  A[Christmas] -->|Get money| B[Go shopping]
  B --> C[Let me think]
  C -->|One| D[Laptop]
  C -->|Two| E[iPhone]
  C -->|Three| F[fa:fa-car Car]
```

1.3. Benchmark

1.3.1. Architettura di testing

1.3.2. Esperimenti

1.3.2.1. Weak Scalability

1.3.2.2. Strong Scalability

1.3.3. Risultati

1.4. Conclusioni