Typsicheres und generisches MapReduce

Jens Dörre

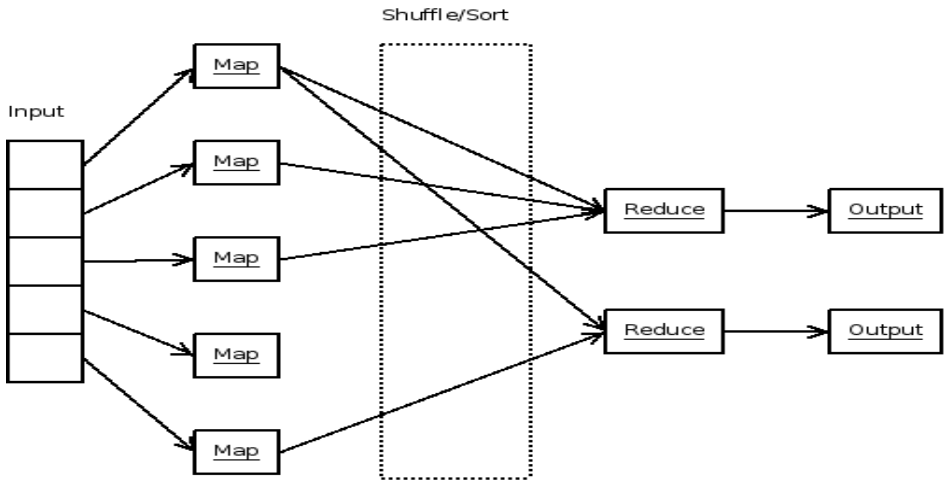
Universität Passau
Fakultät für Informatik und Mathematik
Lehrstuhl für Programmierung: Christian Lengauer

2009-10-12

Überblick

MapReduce-Beispiel

- Google: Erzeugung eines inversen Indexes
- Eingabe: sehr viele Textdokumente;
  Ausgabe: Index (Wort $\rightarrow$ DokumentIDs)
- Abarbeitung in zwei Phasen
  1. "Map" pro Dokument:
     Erzeuge Liste von Paaren (Wort, DokumentID)
  2. "Reduce" pro Wort:
     Erzeuge Paar (Wort, sortierte Liste aller zugehörigen DokumentIDs)



MapReduce

- Programmiermodell für massiv verteilte Ausführung
- Durch Google populär
- Beispiel-Anwendung: Generierung des Indexes für die Web-Suche
- Tausende weitere Anwendungen im Einsatz
  mit einem Durchsatz von Petabytes pro Tag
  - Clustering
  - Häufigste Anfragen
  - Extraktion semantischer Daten
- Datenparallelität als Basis
- Imperativ mit zwei Konzepten aus der Funktionalprogrammierung

Vorteile

- Einfaches Programmiermodell
  - Masse an Daten
  - Notwendige Flexibilität
- Sequenzielle (=einfache) Sicht auf paralleles und verteiltes System
- Akzeptanz bei Mainstream-Programmierern
- Prototyping einfach möglich
- Fehlertoleranz (bei geringem Anteil fehlerhafter *Daten*)

Nachteile

- Mangel an Typsicherheit
  - Späte Erkennung von Fehlern in *Programmen*
  - Probleme mit der Datensicherheit
- Eingeschränkte Anwendbarkeit
  - Maßstab
  - Art der Parallelität
- Generell wenig Systematik
  - Bezug zu Konzepten der Funktionalprogrammierung
  - Modell mit nur zwei Phasen – Praxis: längere Kette

Typisierung

- Statisches Typsystem
- Mehrphasiges Modell
  - $\boxed{M_1}$ - $\boxed{R_1 \text{ - } M_2}$ - $\boxed{R_2}$
- Theoretische (algebraische) Basis für
  - noch bessere Zerlegung der Programme
  - Optimierungen
- Herausforderung: Vorteile erhalten
  - Akzeptanz
  - Unterstützung für Prototyping
  - Fehlertoleranz
  - Einfachheit

Re-Skalierung

- Theoretisch: Grenzen nach oben und unten
- Praktisch: Fallstudie auf Multicores
- Vereinheitlichendes Konzept (und Interface?)

Vielen Dank für die Aufmerksamkeit!