

Map-Reduce

Lehrveranstaltung: „Schemalose Datenbanken“

vorgetragen von:

Moritz Kemp, Team „Hadoop / Hbase“

MapReduce: Simplified Data Processing on Large Clusters

*Jeffrey Dean and Sanjay Ghemawat,
Google Inc.*

Beispiel

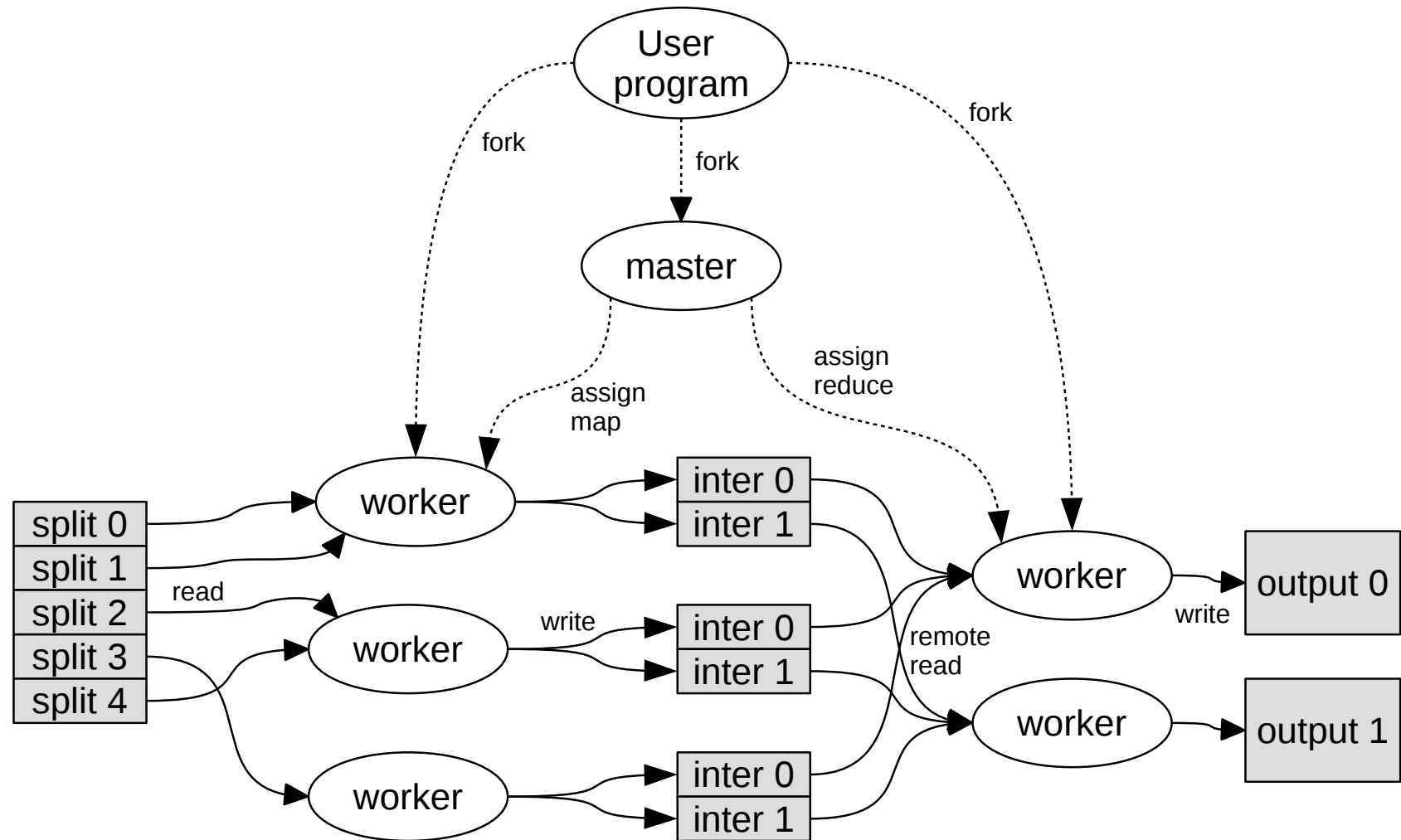
MAP

```
forall songs
  get artist_name
  emit <key: artist_name, value: '1'>
```

REDUCE

```
forall artists
  while artist
    artist_songs++
  emit <key: artist_name, value: artist_songs>
```

Framework



Input files

Map phase

Intermediate
files

Reduce phase

Output files

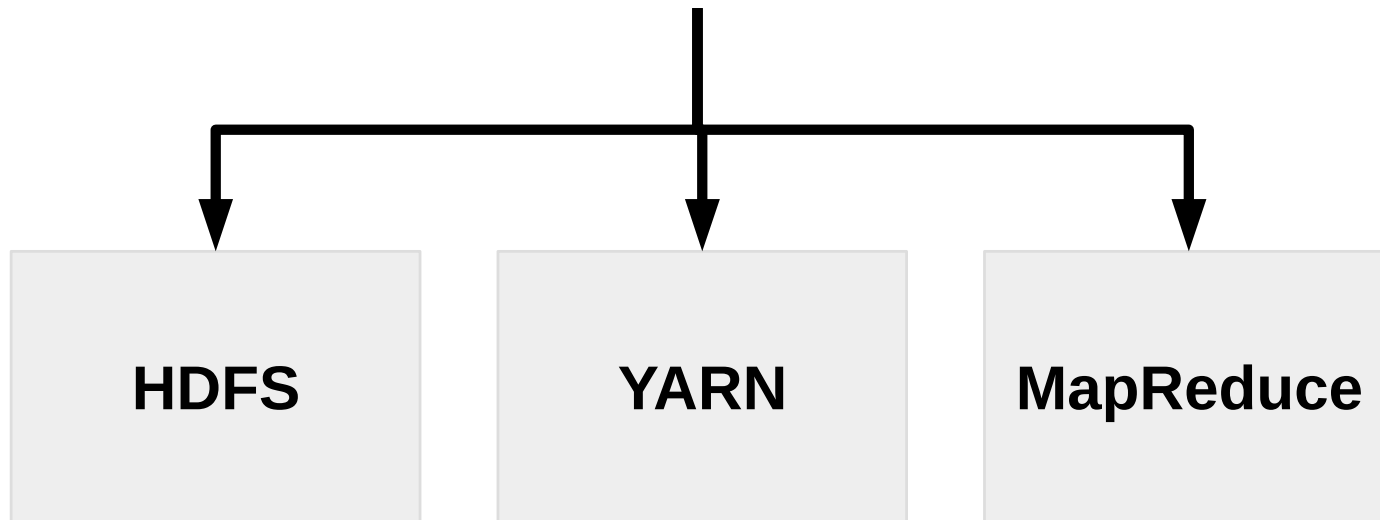
Datenbanken

- Statistische Analyse
- Erstellung von Graphen
- Re-Organisation
- Bildverarbeitung
- Filterung
- Suchen
- ...



Große Mengen an
(strukturierten) Daten

Ausblick



Danke für Ihre Aufmerksamkeit!