





## *Zwei praktische Aufgaben in einer verteilten Umgebung*

- **Transaktional:** NoSQL Anwendungsprogrammierung + Benchmarking
  - Backend eines typischen Webservices (Twitter-ähnlich)
  - Gegeben: Workload-Driver + Postgres-basierte Baselineimplementierung
- **Analytisch:** MapReduce/Hadoop
  - Praktische Implementierung von MapReduce Jobs
  - Vorgegebenes Datenset (WebTable Corpus) und Analyseziele

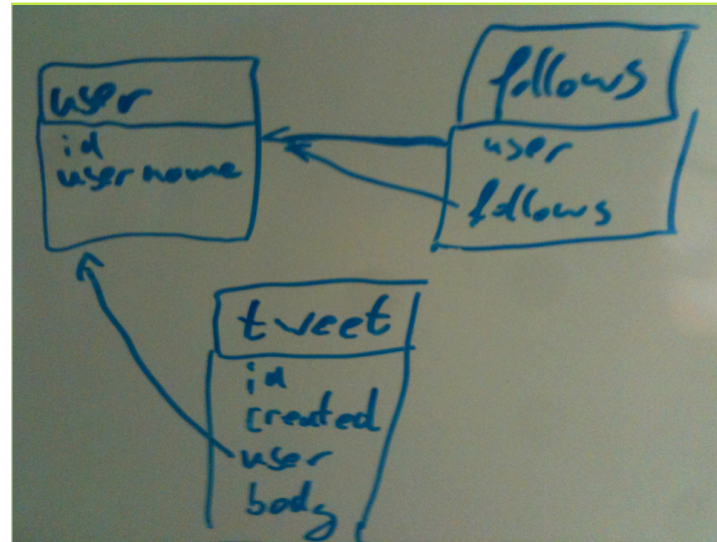
## *Lehrstuhleigene Wimpy-Node-Farm*

- 32 einfache Maschinen

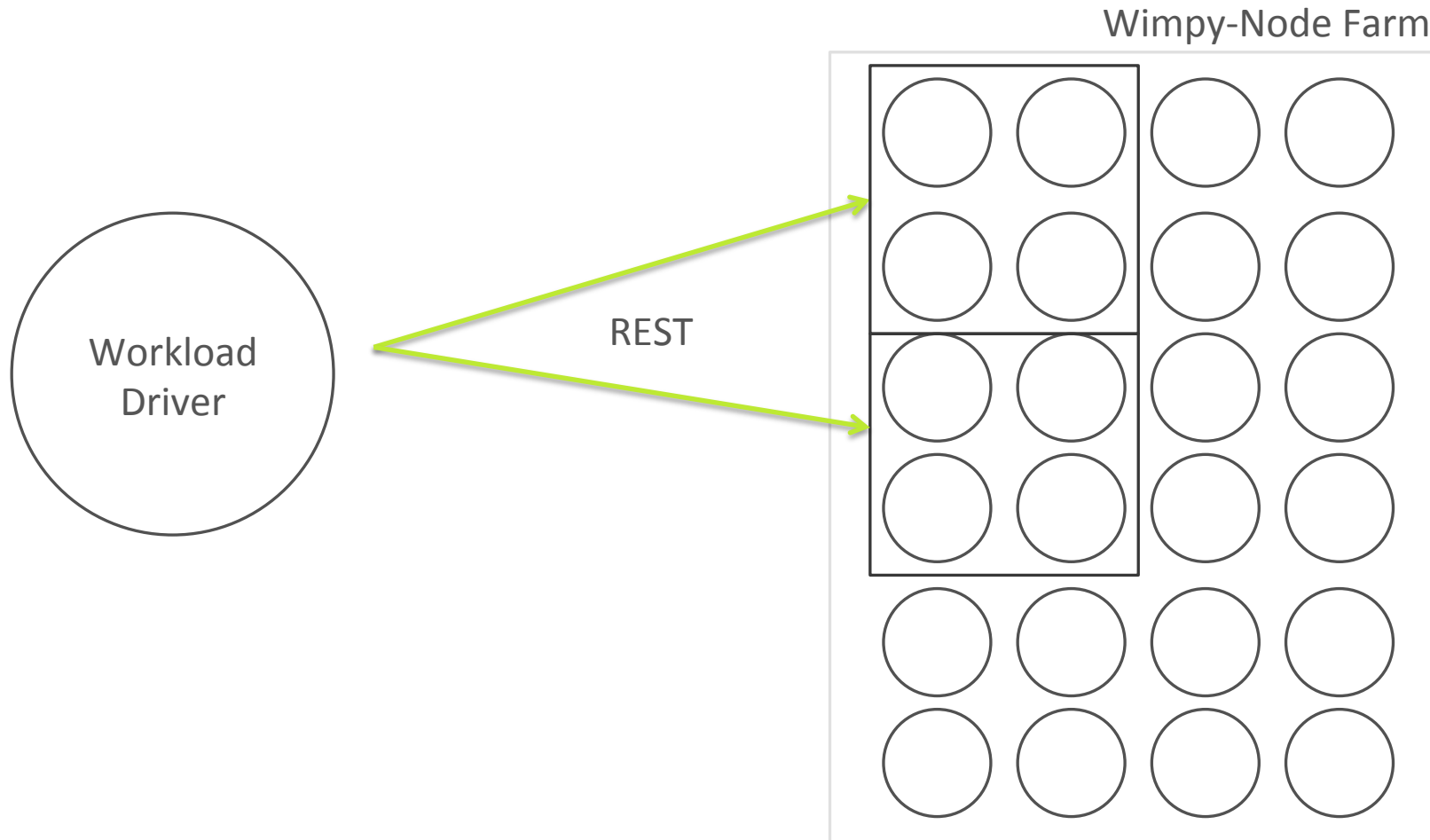


## Aufgabe + Szenario

- Arbeit in Teams, wöchentliche Tasks + Präsentation + Feedback
- Jedes Team nutzt eine andere NoSQL-DB
- Einfaches (?) Twitter-Szenario: User, Tweets, Follow-Beziehung

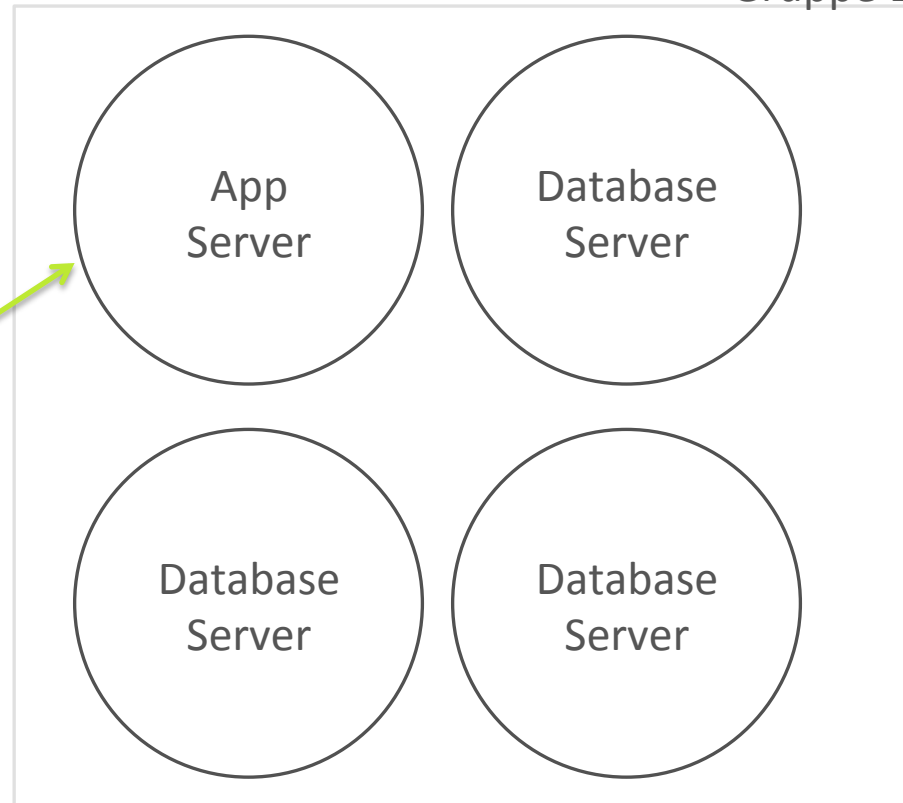


- REST-Interface ist zu Implementieren
  - Bulk-Import
  - Timeline Query
  - (+insert Tweet)
- Relationale Baseline-Implementierung in Postgres+Python ist zu schlagen



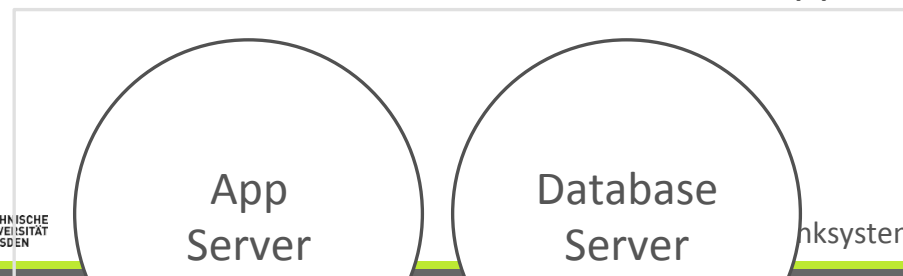


Gruppe 1



REST

Gruppe 2





## *Welche?*

- MongoDB (x)
- Cassandra (x)
- Redis
- HBase
- RethinkDB (x)
- Riak
- Neo4J
- CouchDB (x)
- FoundationDB (x)
- VoltDB

## *Bis nächste Woche 2-3 Slides + 3 min. Vortrag*

- Allgemeine Beschreibung
- Vergleich mit Relationalem Modell + ACID
- Scale-Out / Verteilung