

Web Programming with Python and JavaScript

Scalability



Server



Server

Benchmarking

Vertical Scaling



Server



Server

Horizontal Scaling

Server

Server

Server

```
graph TD; LB[Load Balancer] --- S1[Server]; LB --- S2[Server];
```

Load Balancer

Server

Server

Load Balancing

Load Balancing Methods

- Random Choice
- Round Robin
- Fewest Connections
- ...

Session-Aware Load Balancing

- Sticky Sessions
- Sessions in Database
- Client-Side Sessions
- ...

Autoscaling


```
graph TD; LB[Load Balancer] --- S1[Server]; LB --- S2[Server];
```

Load Balancer

Server

Server

```
graph TD; LB[Load Balancer] --- S1[Server]; LB --- S2[Server]; LB --- S3[Server];
```

Load Balancer

Server

Server

Server

```
graph TD; LB[Load Balancer] --- S1[Server]; LB --- S2[Server]; LB --- S3[Server]; LB --- S4[Server];
```

Load Balancer

Server

Server

Server

Server

```
graph TD; LB[Load Balancer] --- S1[Server]; LB --- S2[Server]; LB --- S3[Server];
```

Load Balancer

Server

Server

Server

```
graph TD; LB[Load Balancer] --- S1[Server]; LB --- S2[Server]; LB --- S3[Server]; LB --- S4[Server];
```

Load Balancer

Server

Server

Server

Server

```
graph TD; LB[Load Balancer] --- S1[Server]; LB --- S2[Server]; LB --- S3[Server]; LB --- S4[Server];
```

Load Balancer

Server

Server

Server

Server

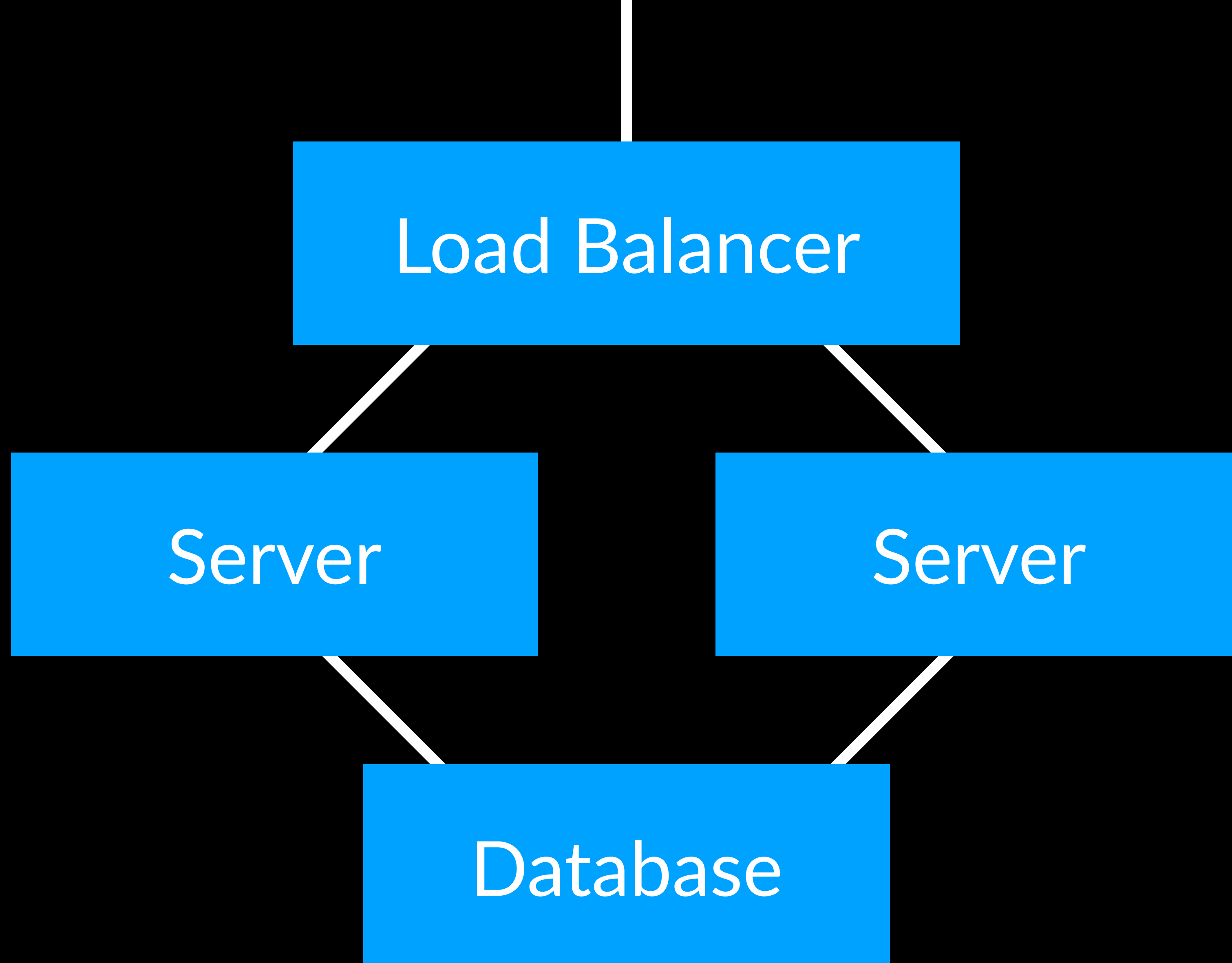
Scaling Databases

```
graph TD; LB[Load Balancer] --- S1[Server]; LB --- S2[Server];
```

Load Balancer

Server

Server



Database Partitioning

flights

id	origin	origin_code	destination	destination_code	duration
1	New York	JFK	London	LHR	415
2	Shanghai	PVG	Paris	CDG	760
3	Istanbul	IST	Tokyo	NRT	700
4	New York	JFK	Paris	CDG	435
5	Moscow	SVO	Paris	CDG	245
6	Lima	LIM	New York	JFK	455

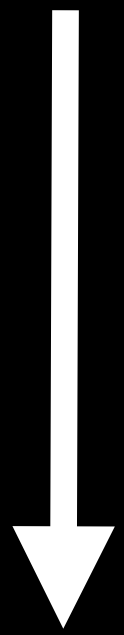
locations

id	code	name
1	JFK	New York
2	PVG	Shanghai
3	IST	Istanbul
4	LHR	London
5	SVO	Moscow
6	LIM	Lima
7	CDG	Paris
8	NRT	Tokyo

flights

id	origin_id	destination_id	duration
1	1	4	415
2	2	7	760
3	3	8	700
4	1	7	435
5	5	7	245
6	6	1	455

flights



flights_domestic

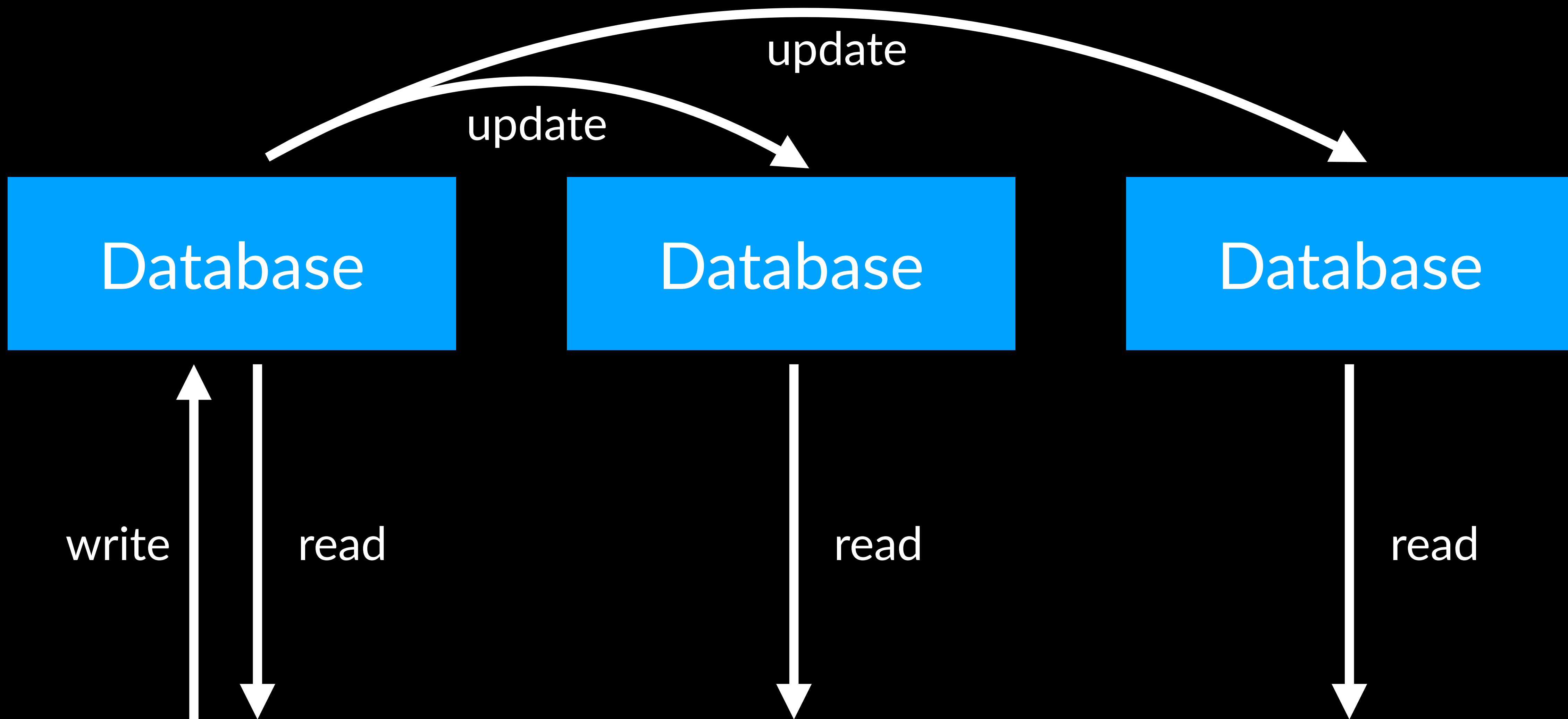
flights_international

Database Replication

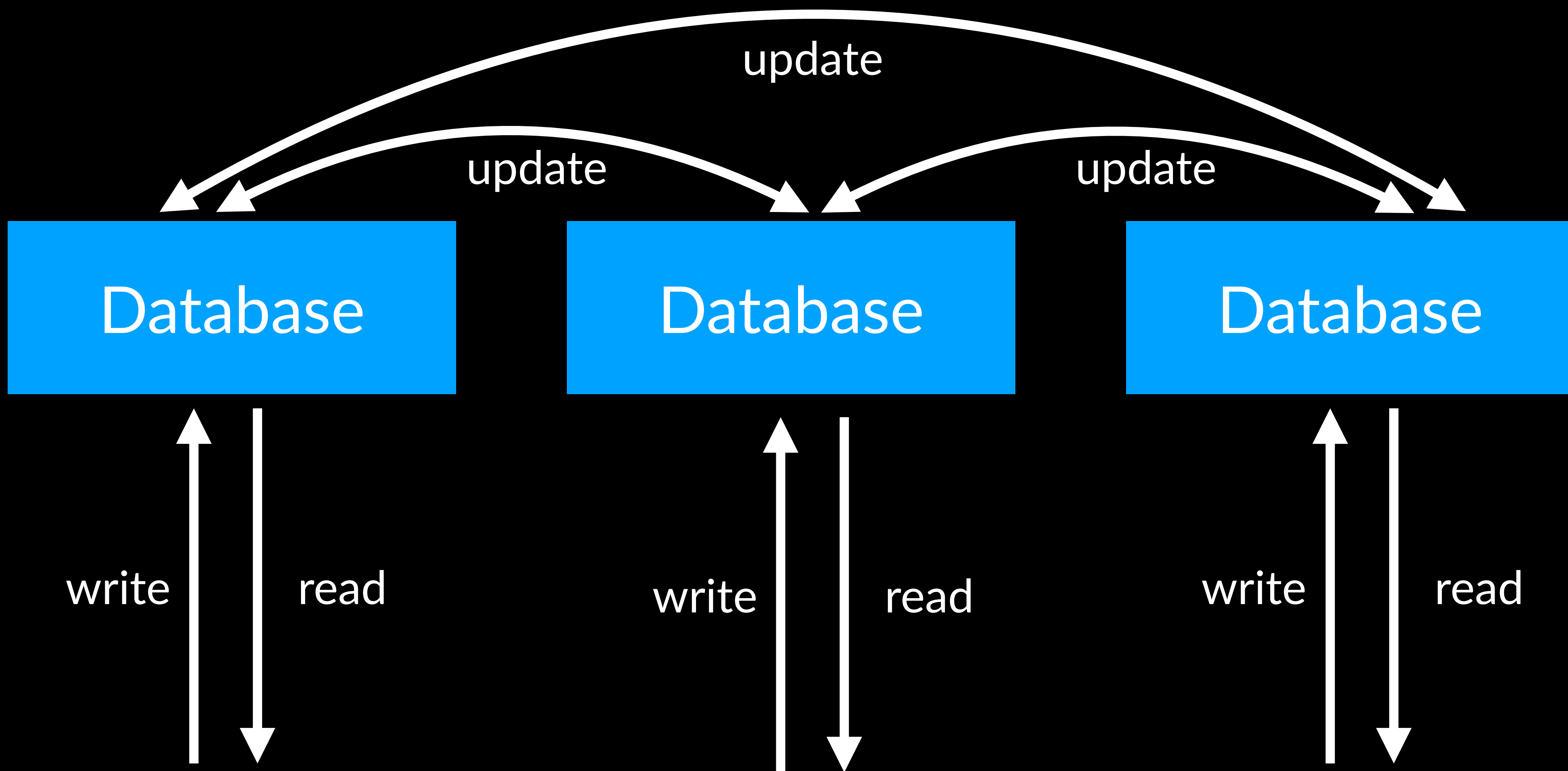
Database Replication

- Single-Primary Replication
- Multi-Primary Replication

Single-Primary Replication



Multi-Primary Replication



Caching

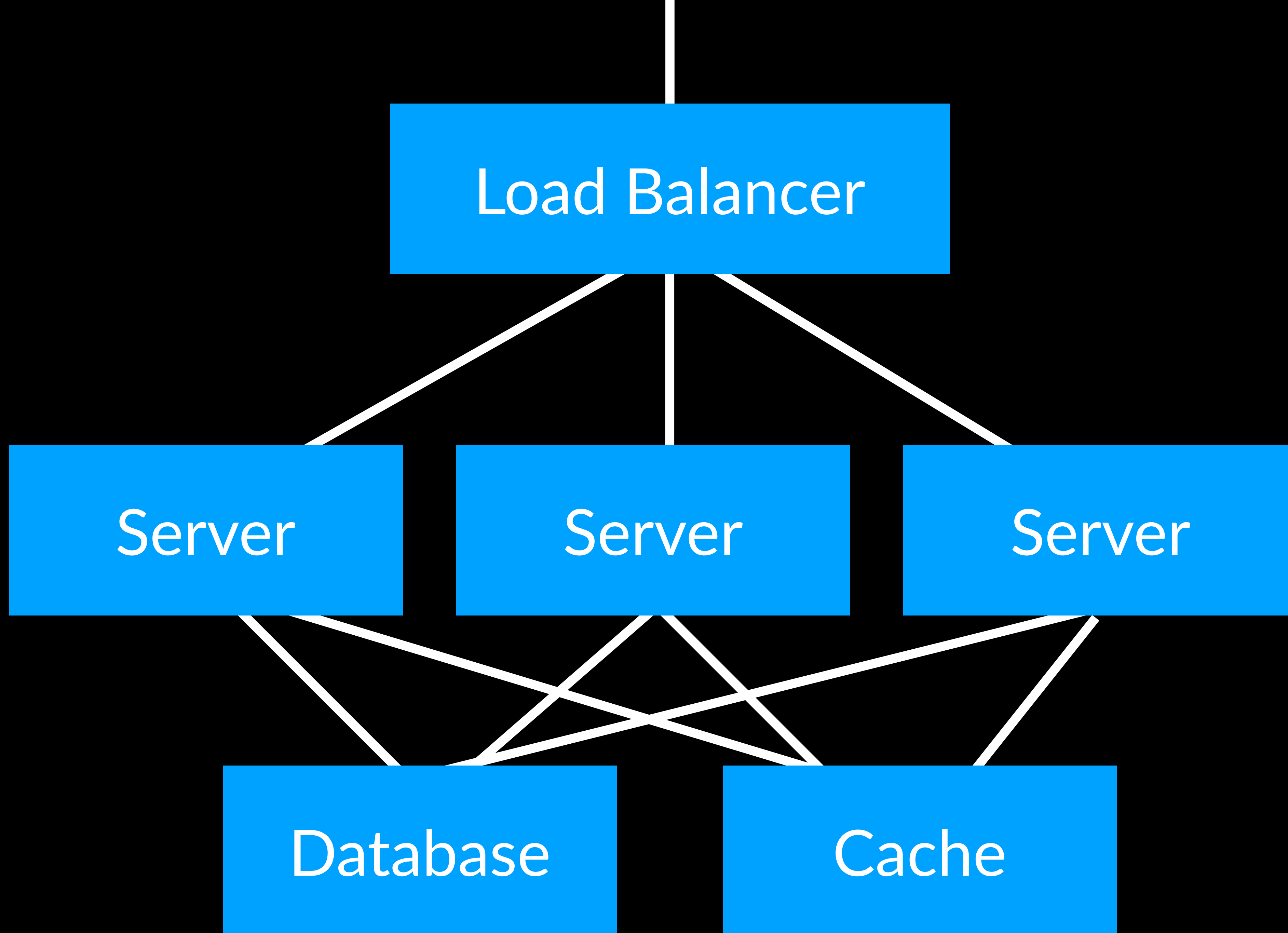
Client-Side Caching

Cache-Control: max-age=86400

Cache-Control: max-age=86400

ETag: "7477656E74796569676874"

Server-Side Caching



Web Programming with Python and JavaScript