

# CSE 344 Introduction to Data Management

Section 8: NoSQL and Parallel DBMS

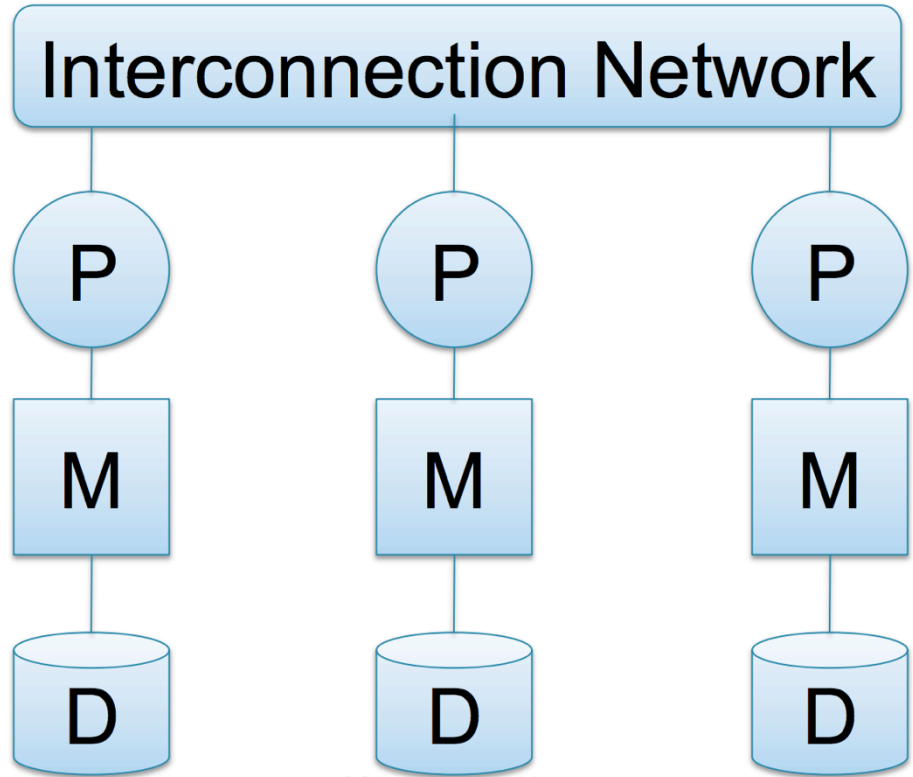
# Firestore Demo

# Architectures for Parallel DBMS

- Shared memory
- Shared disk
- Shared nothing

# Shared Nothing

- Clusters of machines
- Each machine has its own memory and disk
- Easy to maintain and scale
- Difficult to administer and tune

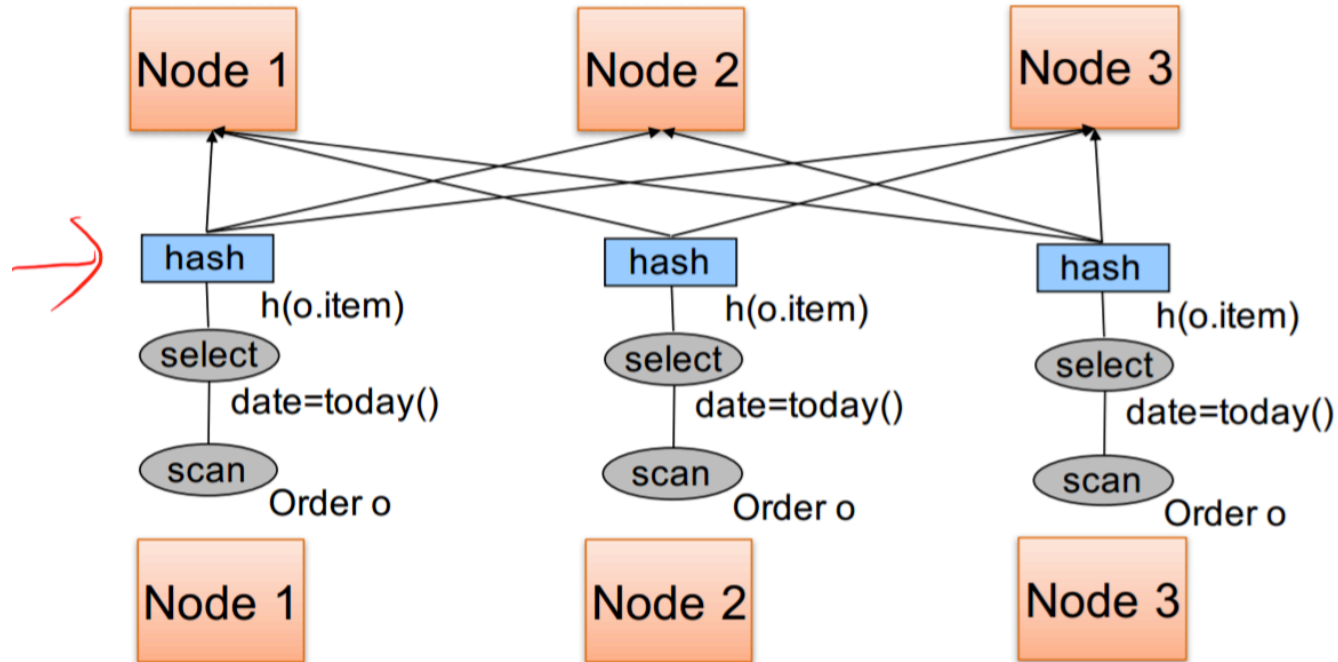
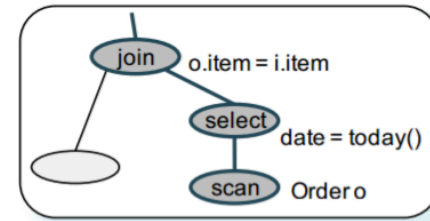


# Horizontal Data Partitioning

- Block Partition
  - Partition tuples arbitrarily s.t.  $\text{size}(R_1) \approx \dots \approx \text{size}(R_P)$
- Hash partitioned on attribute A
  - Tuple  $t$  goes to chunk  $i$ , where  $i = h(t.A) \bmod P + 1$
- Range partitioned on attribute A

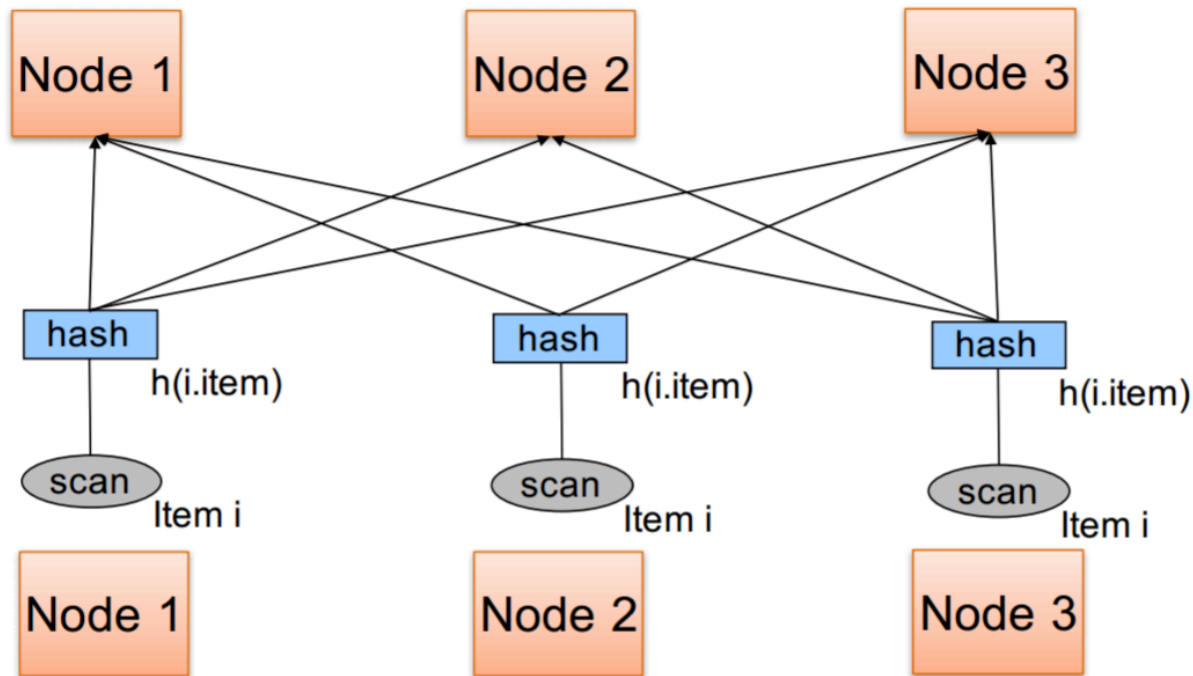
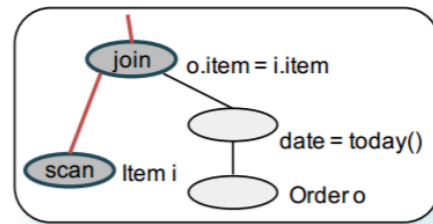
Order(oid, item, date), Line(item, ...)

# Example Parallel Query Plan



Order(oid, item, date), Line(item, ...)

# Example Parallel Query Plan



```
SELECT a, count(b) as countb  
      FROM R  
     WHERE a > 0  
     GROUP BY a
```

Machine 1

$\frac{1}{3}$  of R

Machine 2

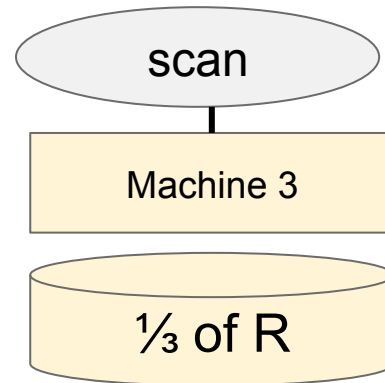
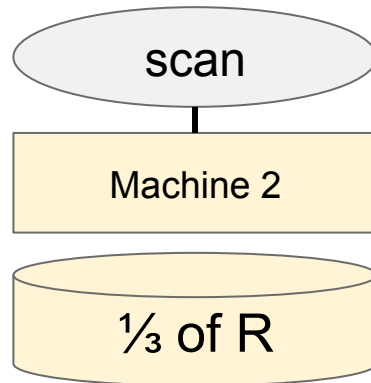
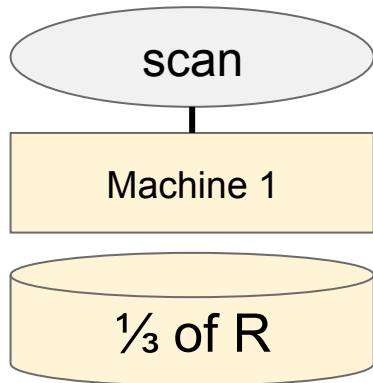
$\frac{1}{3}$  of R

Machine 3

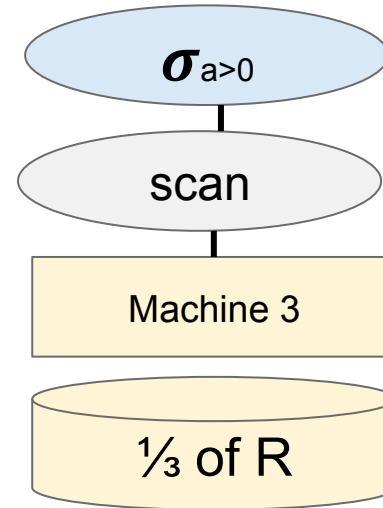
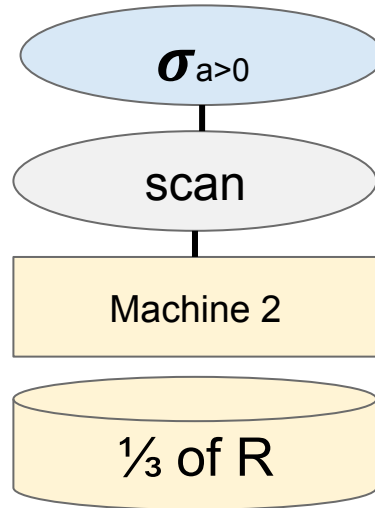
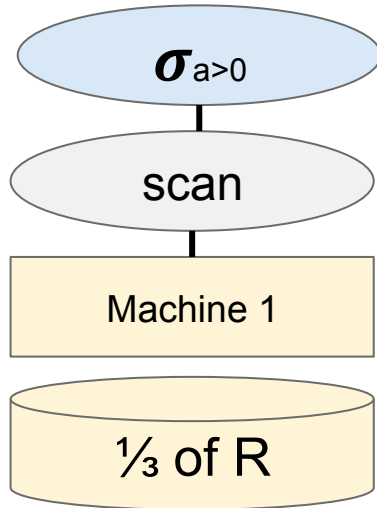
$\frac{1}{3}$  of R



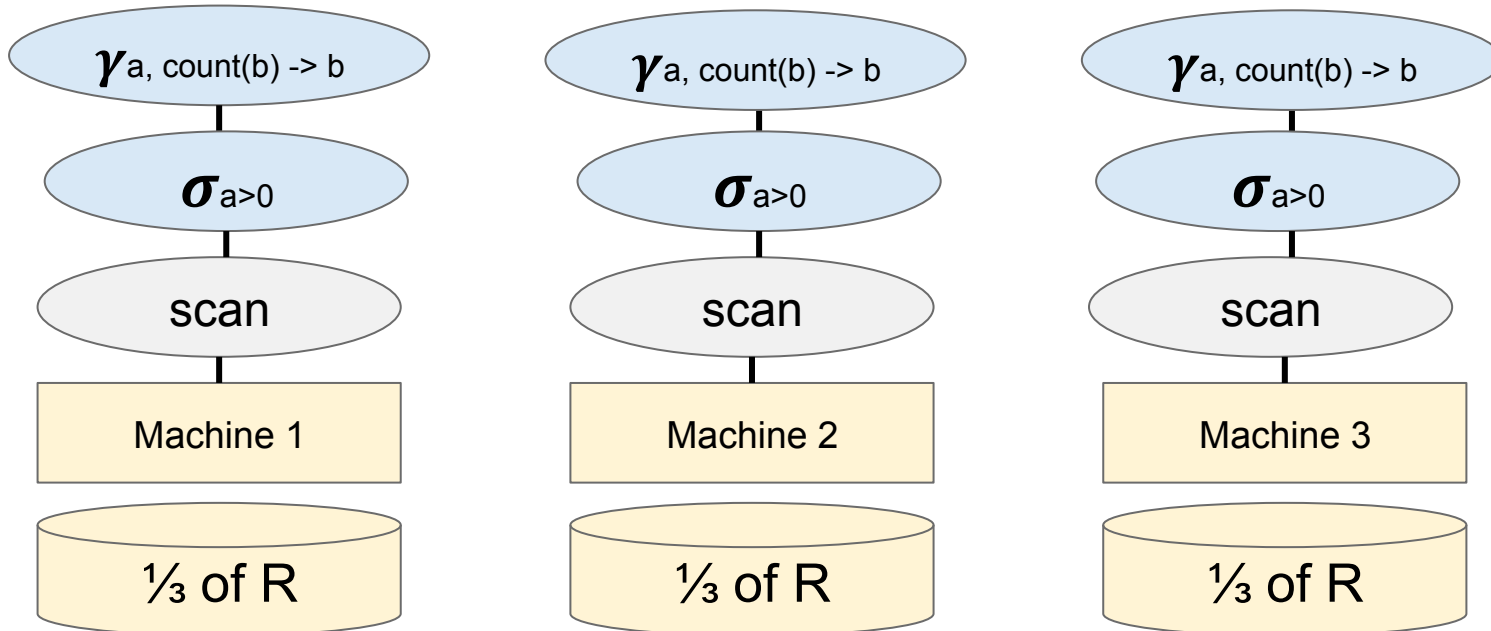
```
SELECT a, count(b) as countb  
FROM R  
WHERE a > 0  
GROUP BY a
```



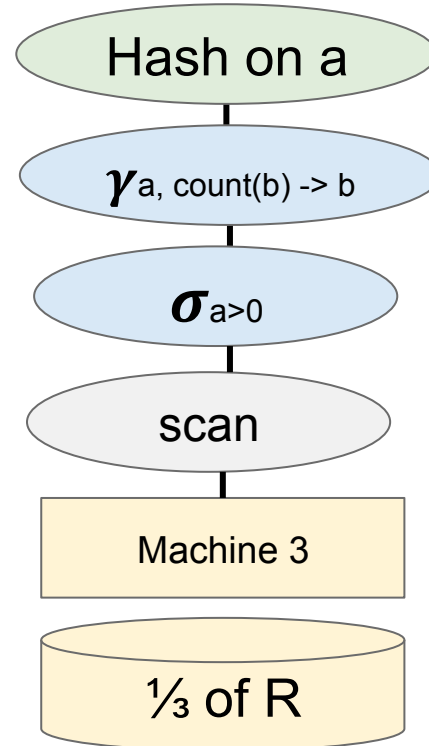
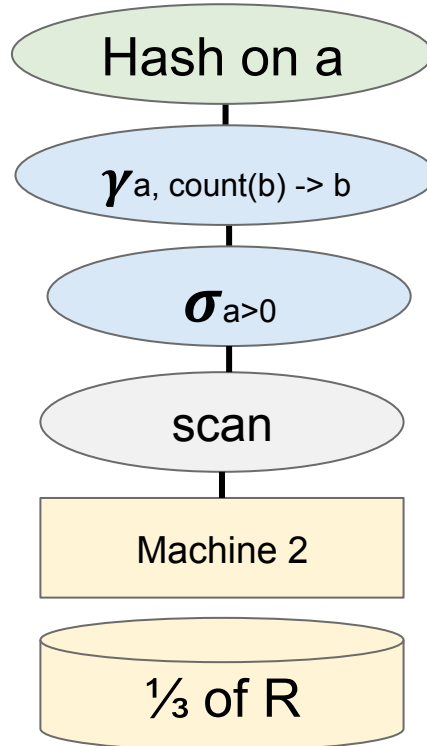
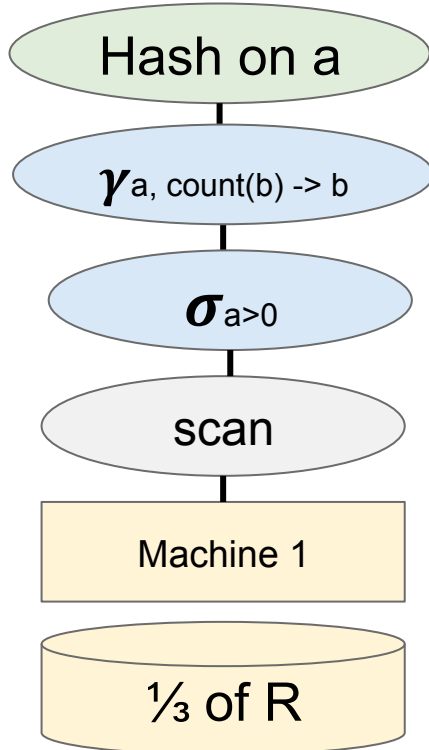
```
SELECT a, count(b) as countb
FROM R
WHERE a > 0
GROUP BY a
```



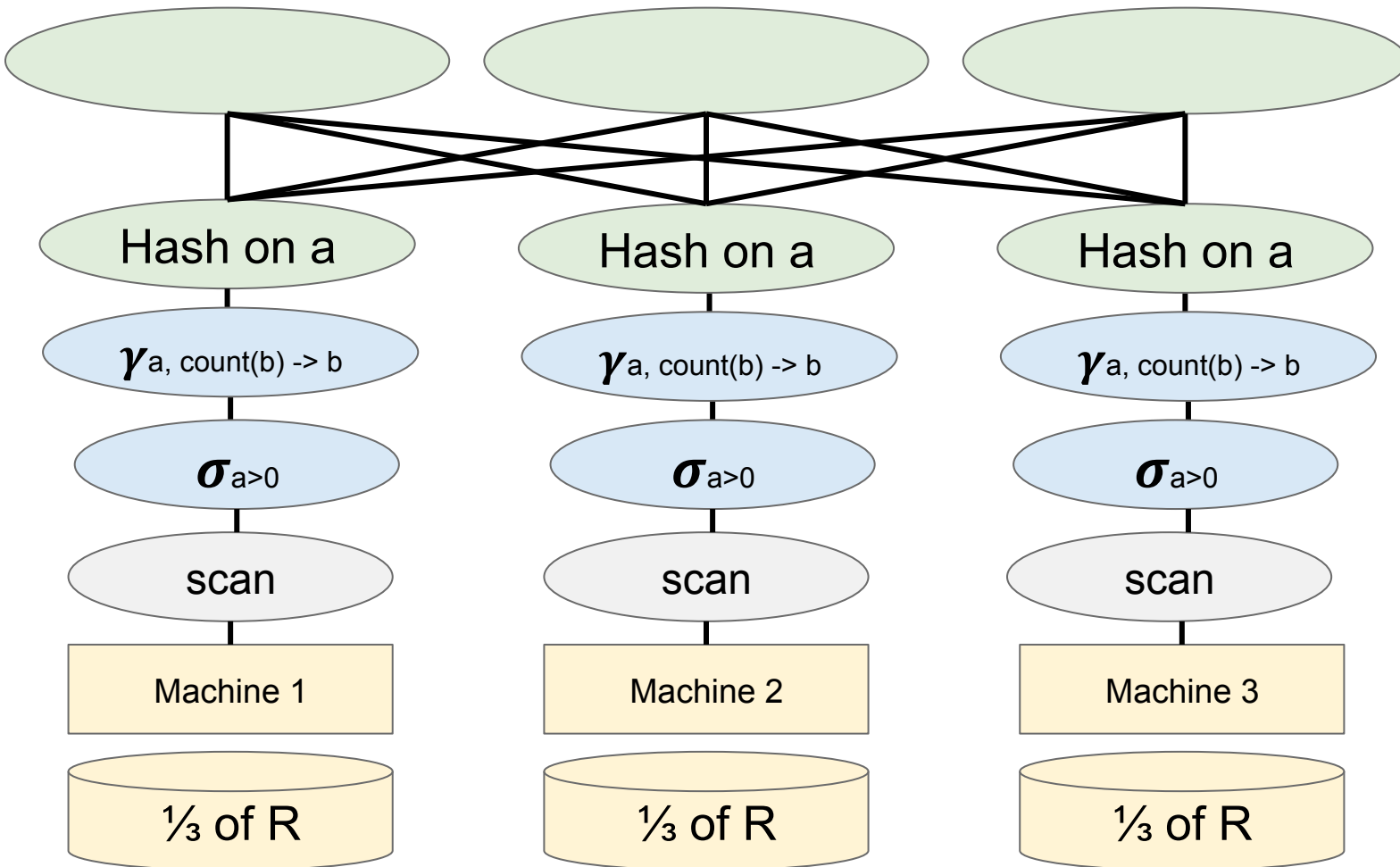
```
SELECT a, count(b) as countb
FROM R
WHERE a > 0
GROUP BY a
```



SELECT a, count(b) as countb  
FROM R WHERE a > 0 GROUP BY a



SELECT a,  
count(b) as countb  
FROM R  
WHERE a > 0  
GROUP BY a



SELECT a,  
count(b) as countb  
FROM R  
WHERE a > 0  
GROUP BY a

