Introduction to
Massively Parallel Databases

Wes Reing

- 10 + Years of Production Databases
- DataXu
- 100TB MPP Databases
- Twitter: @wreing
- Web: reing.com

What I Will Cover

- What purpose do MPPs serve
- How they work in theory
- Practical usage tips

Big Data

How Big is Big?

Bigger than a Single Postgres

Approximately 1 to 2 TB

Options

- Map Reduce / Distributed File System
- NoSQL
- Sharding
- MPP

Map Reduce / Distribute FS

- Runs great on commodity hardware
- Schemaless
- SQL support is not great
- Hadoop, MapR
- SQL support with Hive, Impala

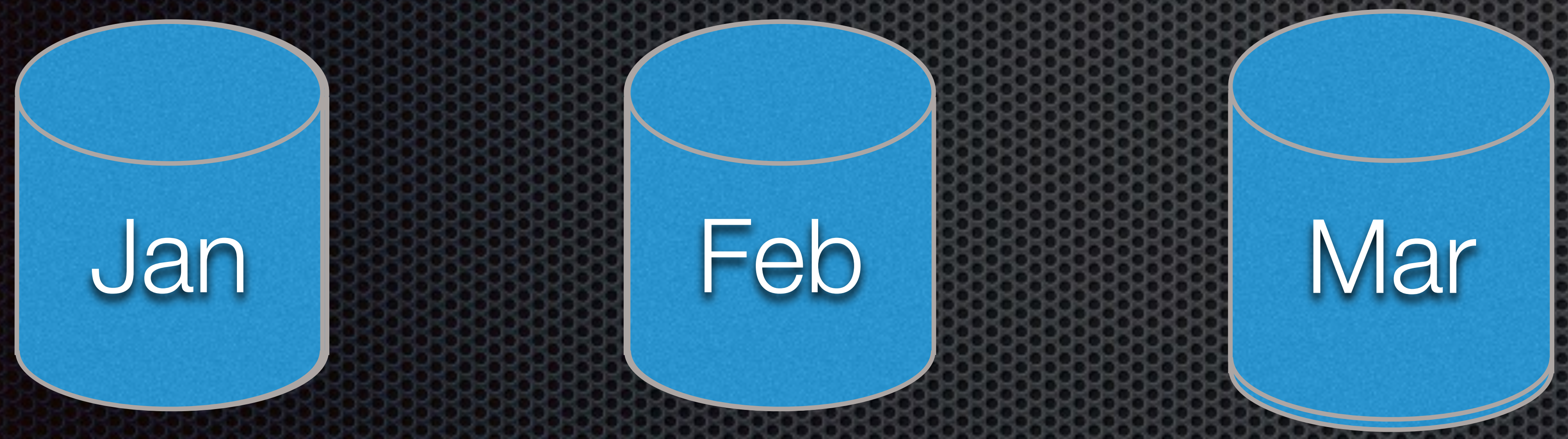
NoSQL

- Scale to Multiple Servers
- Key Value Storage
- Non-Relational
- Limited
- Limited Transaction Support
- MongoDB, FoundataionDB, Spanner, Riak

Sharding

Split the data on a key

- Company
- Date



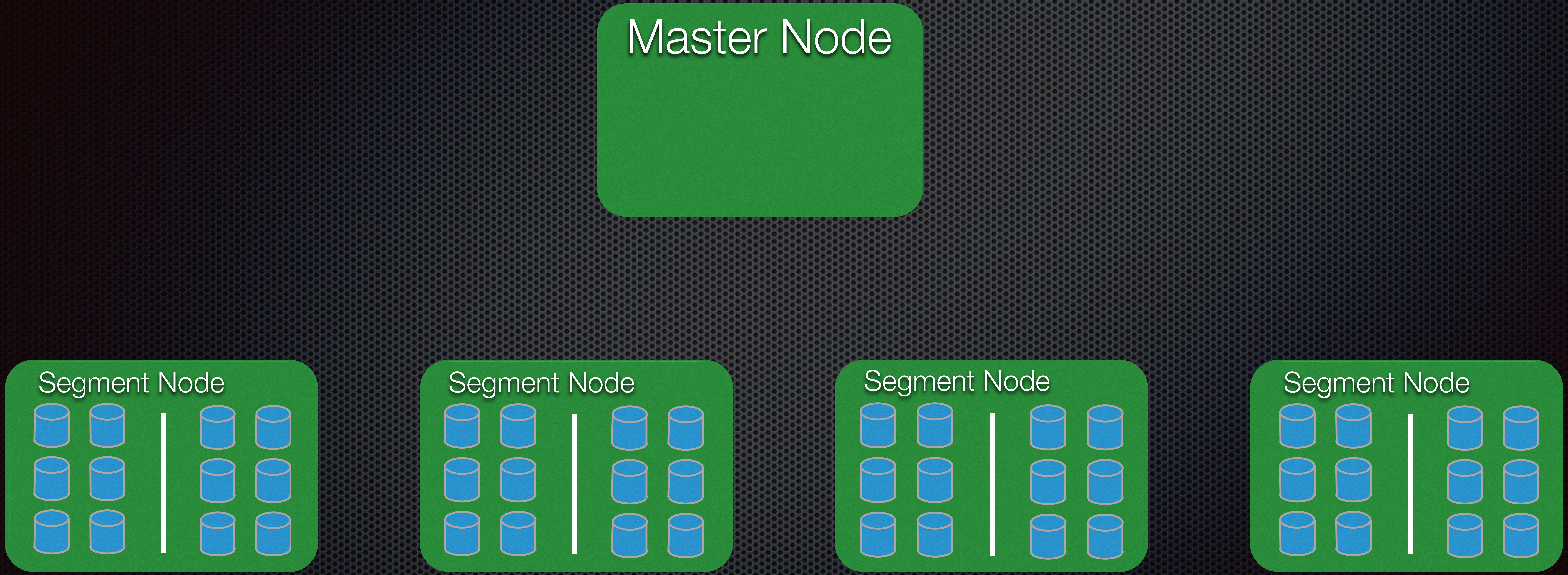
MPP

- A Master node acts like a traditional DB
- Lots of segment nodes split up the work
- Can Support Transactions and Indexes
- Many of the pros and cons of traditional DBs

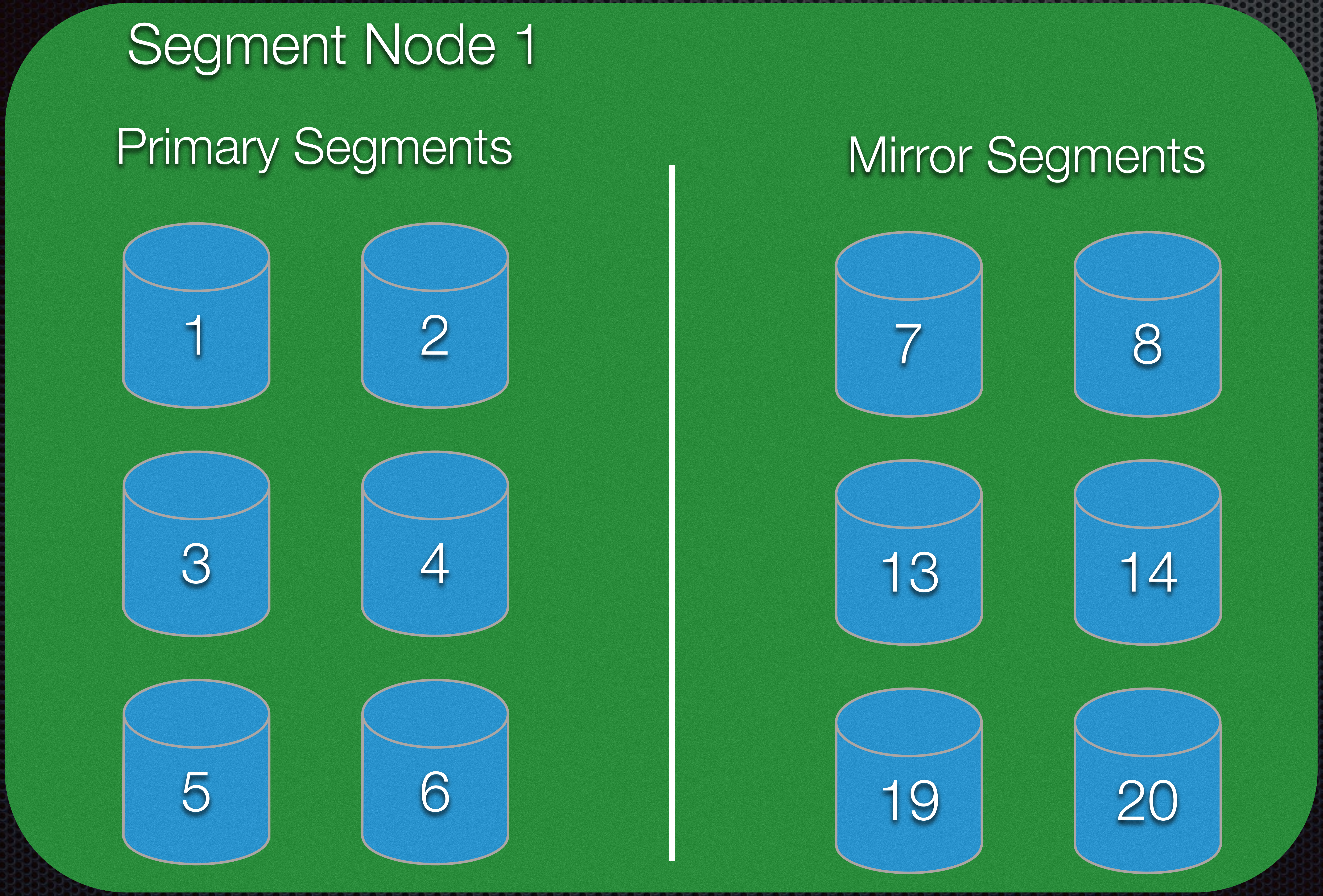
MPP

- No foreign Keys
- No functions that access tables

Greenplum


Master Node
Segment Node
Segment Node
Segment Node
Segment Node

Segment Node



Columnar Databases

- MPP Does not require Columnar Data Stores
- Most vendor implementations do use Columnar DBs
- Redshift, Greenplum, Vertica
- Greenplum allows both

Columnar Databases

- Imagine each column is a separate table
- Especially good for warehouse applications
- Not good for applications with large numbers of updates

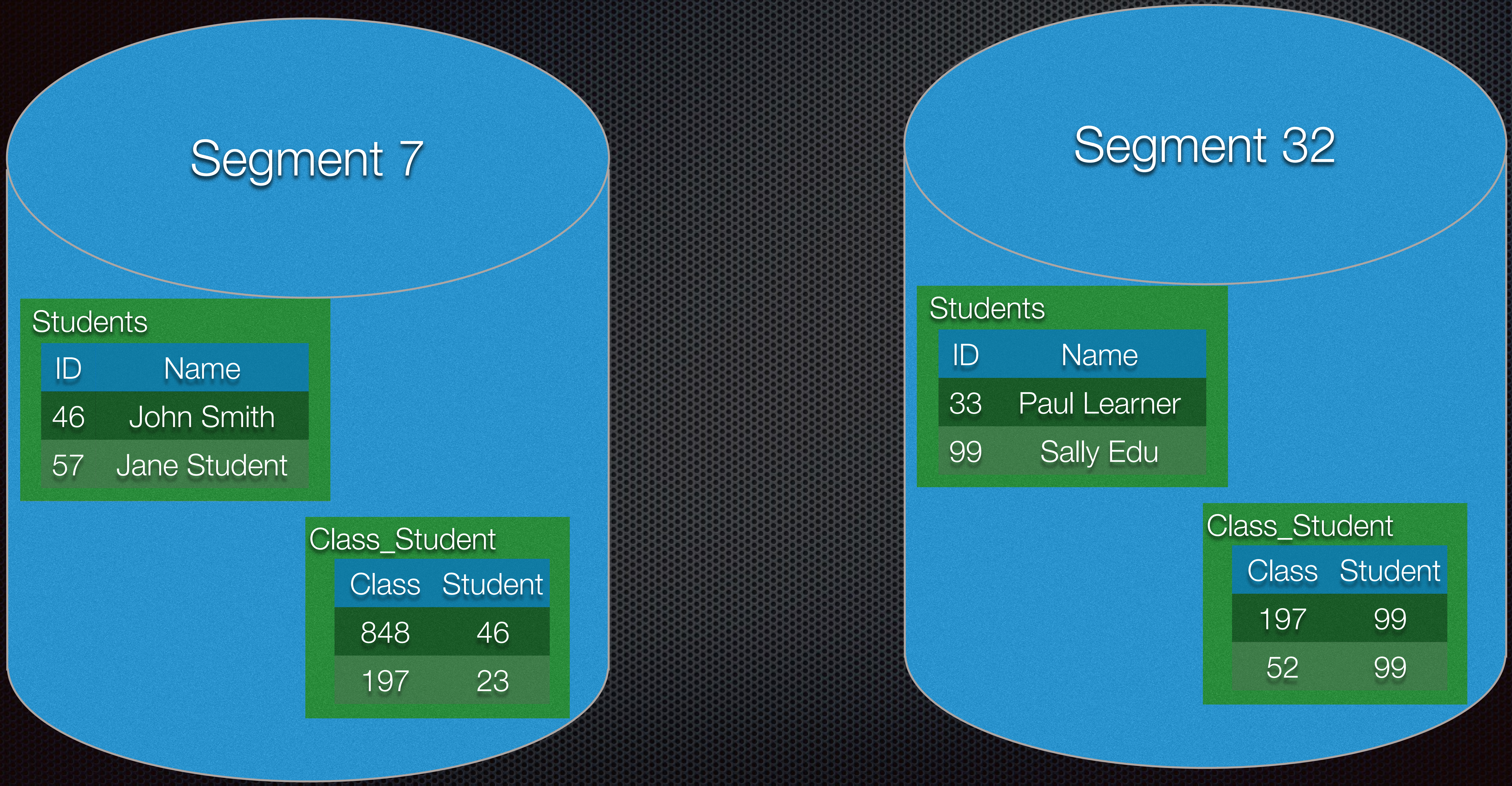
Distributing the data

- Distributed by Key
  - (Key % # of segments) = Segment
- Distributed Randomly

Choosing a good key

- Minimize Skew
- Distribute fact tables that will be joined with the same key
  - Employee ID
  - User ID
  - Order ID

Distribution Keys



Segment 7

Students

ID	Name
46	John Smith
57	Jane Student

Class_Student

Class	Student
848	46
197	23

Segment 32

Students

ID	Name
33	Paul Learner
99	Sally Edu

Class_Student

Class	Student
197	99
52	99

Distribution - Greenplum

```
CREATE TABLE students
(id      INT PRIMARY KEY,
 name    CHAR(50),
 address CHAR(200),
 gpa     NUMERIC,
 enrolled_on DATA
) DISTRIBUTE BY (id);
```

Partitioning

Orders

ID	Date	Item

Orders_2014

ID	Date	Item
95	2014	USB
87	2014	Headphone

Orders_2013

ID	Date	Item
52	2013	Cord
43	2013	Laptop

Orders_2012

ID	Date	Item
23	2012	TV
16	2012	USB

Partitioning

- Works in addition to distribution
- Supported in Greenplum, Vertica
- Not Supported in Redshift

Partitioning - Greenplum

- Defined in the create table statement
- Two levels of partitioning supported
- Keep the number of total partition down

Partitioning - Greenplum

```
CREATE TABLE students
(id          INT PRIMARY KEY,
 name        CHAR(50),
 address     CHAR(200),
 gpa         NUMERIC,
 enrolled_on DATE
)
DISTRIBUTE BY (id)
PARTITION BY RANGE(enrolled_on)
(START (date ‘2010-01-01’) INCLUSIVE
 END   (date ‘2015-01-01’) EXCLUSIVE
 EVERY (INTERVAL ‘1 month’);
```

Partitioning - Vertica

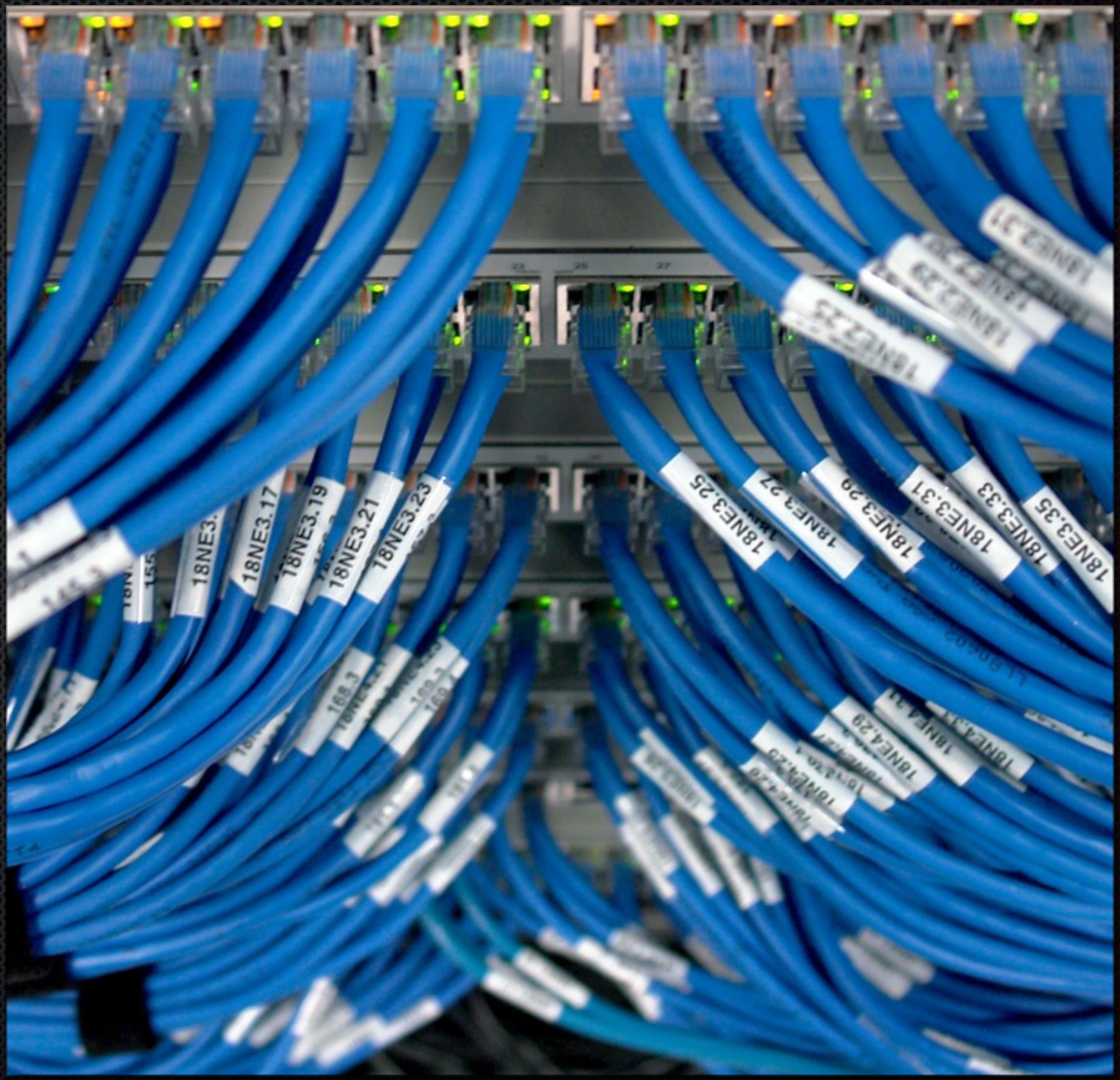
- Defined in the Create Statement
- Partition by Expression
- No more than 12 partitions per table

System Design

- Network Bandwidth
- Disk IO
- Processors

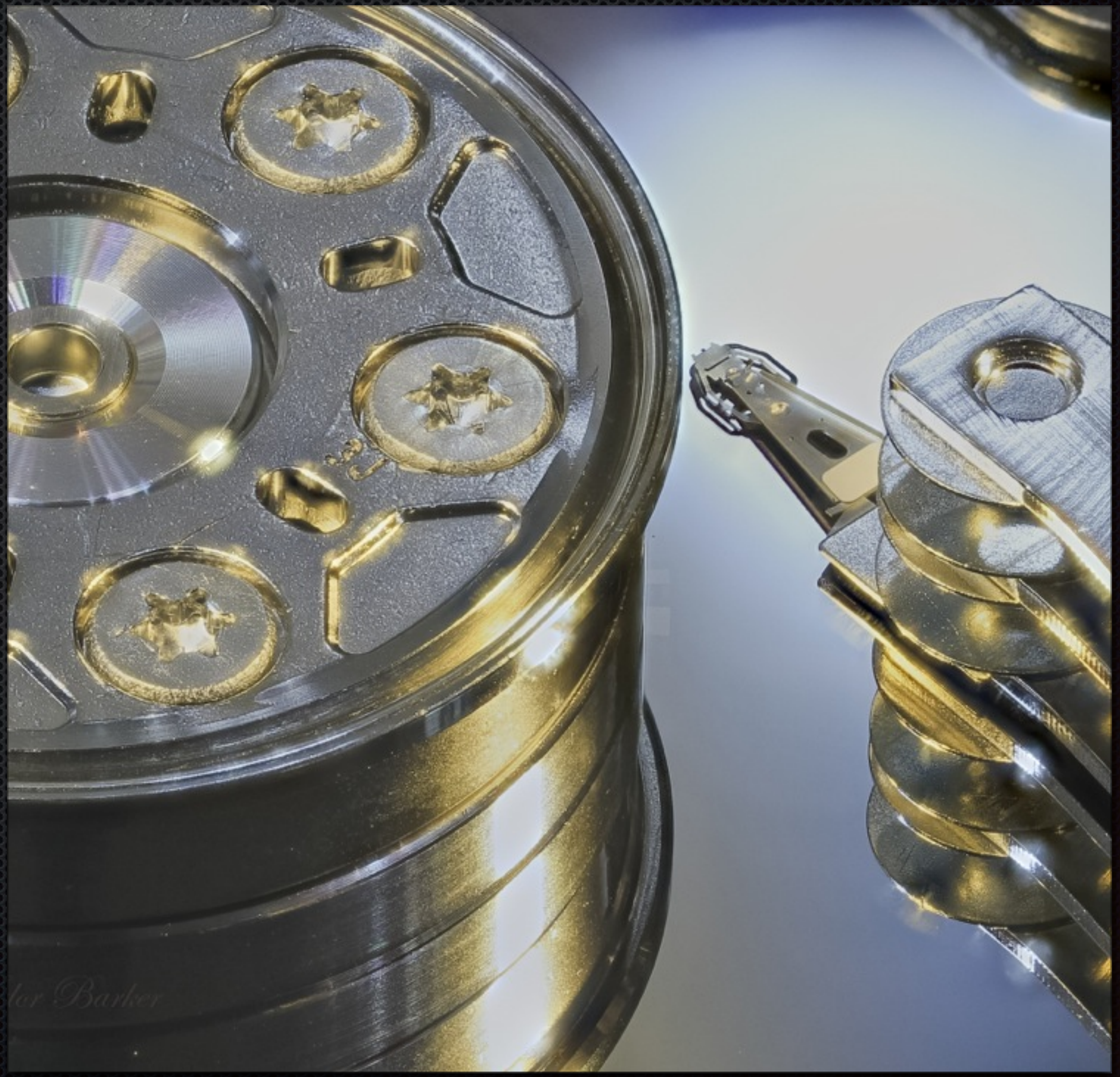
Network

- 10 Gigabit
- Segregated from other Traffic
- gpcheckperf



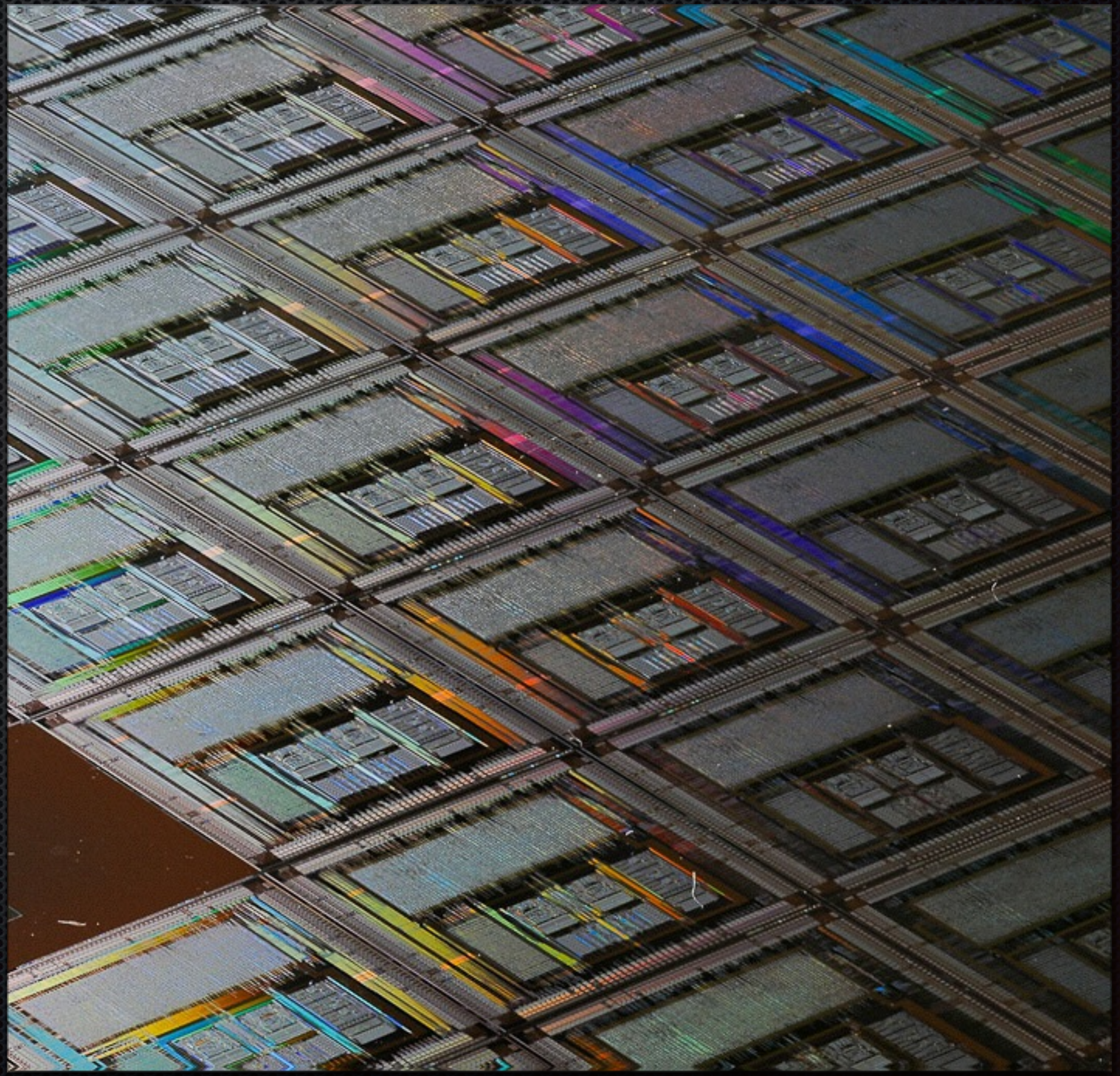
Disk IO

- SSD - RAID - 10k Magnetic
- Need to Balance Speed, Reliability and Cost
- Faliures



Processors and Memory

- Memory. You will need a lot
- CPUs are the largest factor in choosing Segments per Node
- One Core per Segment



Comparison - Greenplum

- Very full featured SQL
- Available as an appliance, and as software only
- Very sensitive to hardware

Comparison - Vertica

- Columnar from the ground up
- Projections

Comparison - Redshift

- Based on Paraccel
- Lots of Progress
- Limited SQL

Tips - Greenplum

Distinct Can Be Slow

```
SELECT DISTINCT classes
FROM students;
```

```
SELECT classes
FROM students
GROUP BY classes;
```

Tips - Greenplum

```
CREATE TABLE temporary_users
as
SELECT id, town, income
FROM users
where income > 20,000;
```

Photo Credits

- Factory Machinery - Daniel Foster - https://flic.kr/p/8cBdxe
- amagasaki-factory-20130227 - kenmainr - https://flic.kr/p/e6ihKW
- Water under glass - TheTallest - https://flic.kr/p/8tcjF
- Inside a Hard Disc Drive - Tudor Barker - https://flic.kr/p/4jcppM
- Switch - Andrew Hart - https://flic.kr/p/dmjkSk
- CPU Wafer Stack - Mark Sze - https://flic.kr/p/7rTz1h

Questions?