

Redshift & Final Project

Monday, March 20, 2017



Agenda

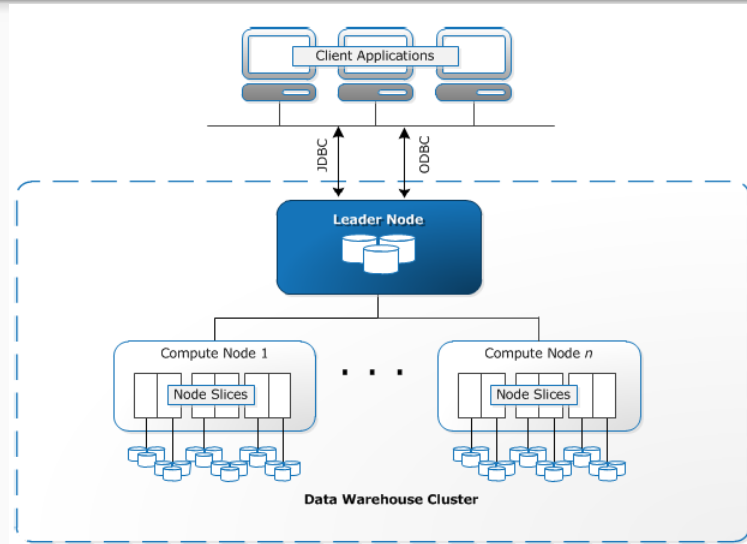
- Announcements
- Redshift Primer
- Final Project Specs
- Final Project Setup Session

Announcements

- Midterm exam grades
- No reading quiz today
- Next class: **guest lecture**

Redshift Primer

- Amazon's data warehouse database
- Derived from Postgres
- Does **not** support B-tree indexes
- Uses parallel processing
- Uses data compression (~3x)
- Uses columnar storage
- Distribution styles (Even, Key, All)
- Sort styles (Single, Compound, Interleaved)
- COPY, ANALYZE, VACUUM commands
- Workload Management (WLM)
- Denormalization (materializing views not supported)



Source: Redshift Developer Guide

What happened to referential integrity?

```
1 create table users(  
2     userid integer not null distkey sortkey,  
3     username char(8),  
4     firstname varchar(30),  
5     lastname varchar(30),  
6     city varchar(30),  
7     state char(2),  
8     email varchar(100),  
9     phone char(14),  
10    likesports boolean,  
11    liketheatre boolean,  
12    likeconcerts boolean,  
13    likejazz boolean,  
14    likeclassical boolean,  
15    likeopera boolean,  
16    likerock boolean,  
17    likevegas boolean,  
18    likebroadway boolean,  
19    likemusicals boolean);
```

```
21 create table venue(  
22     venueid smallint not null distkey sortkey,  
23     venuename varchar(100),  
24     venuecity varchar(30),  
25     venuestate char(2),  
26     venuseats integer);  
27  
28 create table category(  
29     catid smallint not null distkey sortkey,  
30     catgroup varchar(10),  
31     catname varchar(10),  
32     catdesc varchar(50));  
33  
34 create table date(  
35     dateid smallint not null distkey sortkey,  
36     caldate date not null,  
37     day character(3) not null,  
38     week smallint not null,  
39     month character(5) not null,  
40     qtr character(5) not null,  
41     year smallint not null,  
42     holiday boolean default('N'));
```

```
44 create table event(  
45     eventid integer not null distkey,  
46     venueid smallint not null,  
47     catid smallint not null,  
48     dateid smallint not null sortkey,  
49     eventname varchar(200),  
50     starttime timestamp);  
51  
52 create table listing(  
53     listid integer not null distkey,  
54     sellerid integer not null,  
55     eventid integer not null,  
56     dateid smallint not null sortkey,  
57     numtickets smallint not null,  
58     priceperticket decimal(8,2),  
59     totalprice decimal(8,2),  
60     listtime timestamp);  
61  
62 create table sales(  
63     salesid integer not null,  
64     listid integer not null distkey,  
65     sellerid integer not null,  
66     buyerid integer not null,  
67     eventid integer not null,  
68     dateid smallint not null sortkey,  
69     qtysold smallint not null,  
70     pricepaid decimal(8,2),  
71     commission decimal(8,2),  
72     saletime timestamp);
```

Demo 1: Launching a Redshift Cluster

Final Project Specs

<http://www.cs.utexas.edu/~scohen/assignments/final-project.pdf>

Demo 2: Connecting to Redshift