

En primer lugar se crea un cluster de Redshift Serverless usando la configuración por defecto. Esto nos crea un namespace y workgroup con la configuración básica necesaria.

Amazon Redshift Serverless > Get started with Amazon Redshift Serverless

Get started with Amazon Redshift Serverless

To start using Amazon Redshift Serverless, set up your serverless data warehouse and create a database.

When you create and use your serverless data warehouse for the first time, a \$300 credit toward Redshift Serverless usage is applied to the account.

Configuration

☒ Use default settings

Default settings have been defined to help you get started. You can change them at any time later.

☐ Customize settings

Customize your settings for your specific needs.

Namespace

Namespace is a collection of database objects and users. Data properties include database name and password, permissions, and encryption and security.

Your data is encrypted by default with an AWS owned key. To choose a different key, choose **Customize settings**.

Target namespace

default-namespace

- A **Namespace** is a collection of database objects and users. Namespaces group together all of the resources you use in Redshift Serverless, such as schemas, tables, users, datashares, and snapshots.
- A **Workgroup** is a collection of compute resources. Workgroups house compute resources that Redshift Serverless uses to run computational tasks.

Esperamos unos minutos a que cargue el cluster...

Create serverless

It may take a few minutes to complete. After completing the setup, you can work with your data.

Setting up your Amazon Redshift Serverless.

0%

Ease of use with Amazon Redshift Serverless

Access and analyze data without the need to set up, tune, and manage Amazon Redshift clusters.

El namespace y workgroup están listos para ser utilizados.

Amazon Redshift Serverless

Serverless dashboard

Query data

Create workgroup

Namespace overview

Filter namespace

All namespaces

Total snapshots

0

Datashares in my account

0

Datashares requiring authorization

0

Datashares from other accounts

0

Datashares requiring association

0

Namespaces / Workgroups

Namespace	Status	Workgroup	Status
default-namespace	Available	default-workgroup	Available

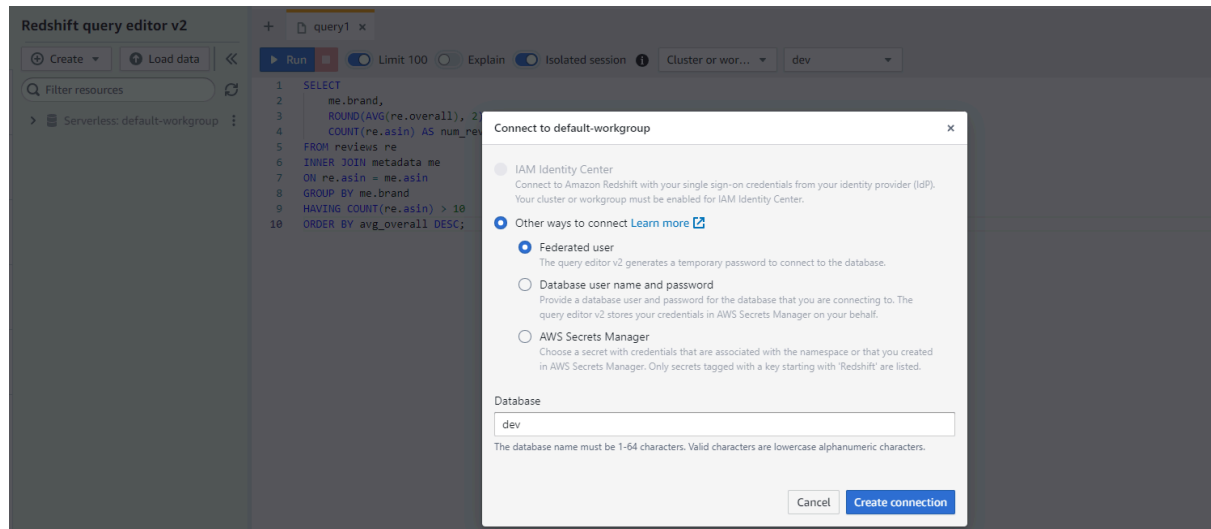
Total compute usage - new

Choose a workgroup

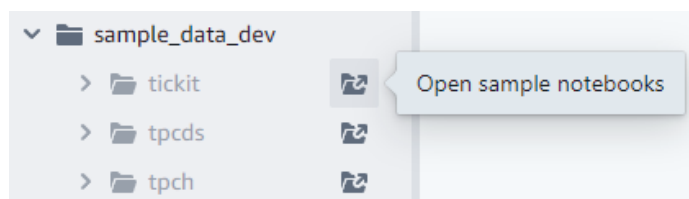
Last hour

To visualize the costs of your total compute usage, go to AWS Cost Explorer

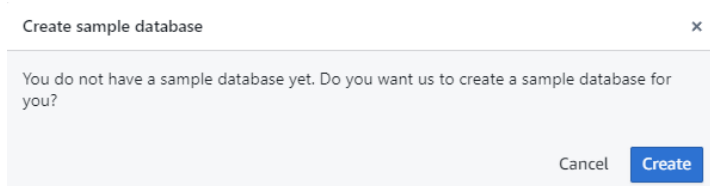
Por medio de la opción *query data* accedemos al editor de Redshift. Al entrar por primera vez se debe establecer la conexión, y en este caso se elige la forma Federated user. Cabe mencionar que también es posible crear unas credenciales o traerlas desde el servicio Secrets Manager de AWS.



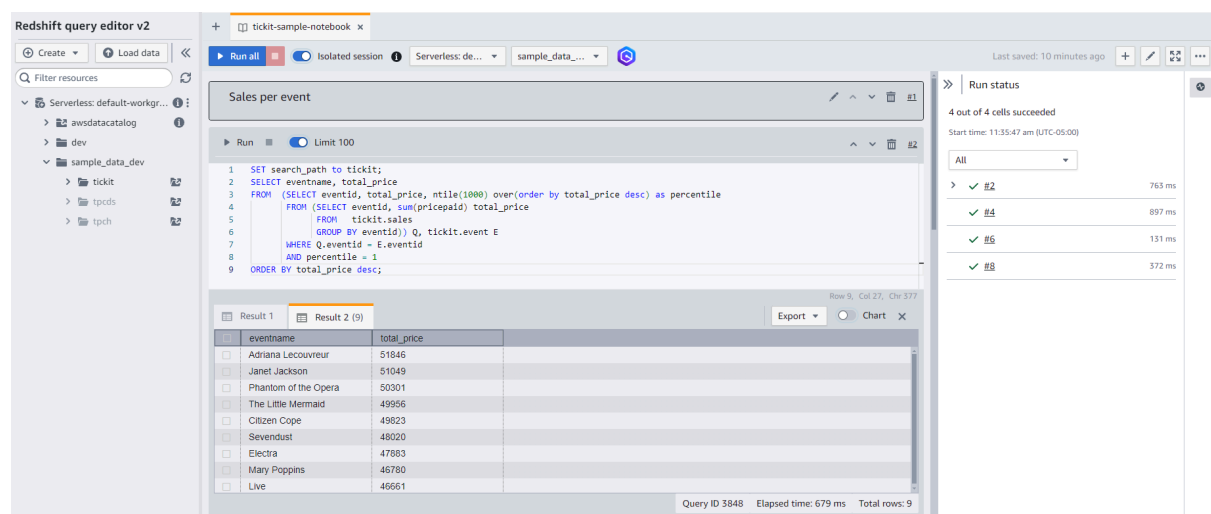
En este punto aparece una carpeta con datos de prueba que puede ser consultada en cualquier momento. Elegimos el sample llamado tickit.



Se crea la base de datos correspondiente.



Una vez creada, se nos proporciona unos queries de prueba incluyendo algunas tablas de la data.



Ahora, vamos a cargar los datos desde un bucket de S3. El bucket ya cuenta con los archivos correspondientes.

Amazon S3 > Buckets > tickit-sample-data

tickit-sample-data info

Objects Properties Permissions Metrics Management Access Points

Objects (7) info

Copy Copy S3 URI Copy URL Download Open Delete Actions Create folder Upload

Find objects by prefix

	Name	Type	Last modified	Size	Storage class
<input type="checkbox"/>	allevents_pipe.txt	txt	September 6, 2024, 12:18:53 (UTC-05:00)	435.4 KB	Standard
<input type="checkbox"/>	allusers_pipe.txt	txt	September 6, 2024, 12:18:54 (UTC-05:00)	5.6 MB	Standard
<input type="checkbox"/>	category_pipe.txt	txt	September 6, 2024, 12:18:55 (UTC-05:00)	465.0 B	Standard
<input type="checkbox"/>	date2008_pipe.txt	txt	September 6, 2024, 12:18:55 (UTC-05:00)	14.2 KB	Standard
<input type="checkbox"/>	listings_pipe.txt	txt	September 6, 2024, 12:18:57 (UTC-05:00)	11.0 MB	Standard
<input type="checkbox"/>	sales_tab.txt	txt	September 6, 2024, 12:18:58 (UTC-05:00)	10.7 MB	Standard
<input type="checkbox"/>	venue_pipe.txt	txt	September 6, 2024, 12:18:59 (UTC-05:00)	7.8 KB	Standard

Luego vamos a la configuración del namespace y seleccionamos la opción de crear un IAM role.

Amazon Redshift Serverless > Namespace configuration > default-namespace > Manage IAM roles

Manage IAM roles

Permissions

Associate an IAM role so that your serverless endpoint can LOAD and UNLOAD data. You can create an IAM role as the default for this configuration that has the AmazonRedshiftAllCommandsFullAccess policy attached. This policy includes permissions to run SQL commands to COPY, UNLOAD, and query data with Amazon Redshift Serverless. This policy also grants permissions to run SELECT statements for related services, such as Amazon S3, Amazon CloudWatch logs, Amazon SageMaker, and AWS Glue. You won't be able to run these SQL commands without an IAM role attached to your namespace.

Associated IAM roles (1)

Create, associate, or remove an IAM role. You can associate up to 50 IAM roles. You can also choose default.

Search for associated IAM role by name, status, or role type

Set default Manage IAM roles Associate IAM roles Create IAM role Remove IAM roles

1

<input type="checkbox"/>	IAM roles	Status	Role type
<input type="checkbox"/>	AmazonRedshift-CommandsAccessRole-20231126T121240	in-sync	Default

Se selecciona solo el bucket ya creado.

Create the default IAM role

Specify an S3 bucket for the IAM role to access

To create a new bucket, visit S3

☐ No additional S3 bucket

Create the IAM role without specifying S3 buckets.

☐ Any S3 bucket

Allow users that have access to your Redshift Serverless data to also access any S3 bucket and its contents in your AWS account.

☒ Specific S3 buckets

Specify one or more S3 buckets that the IAM role being created has permission to access.

S3 buckets (1/1)

Find bucket

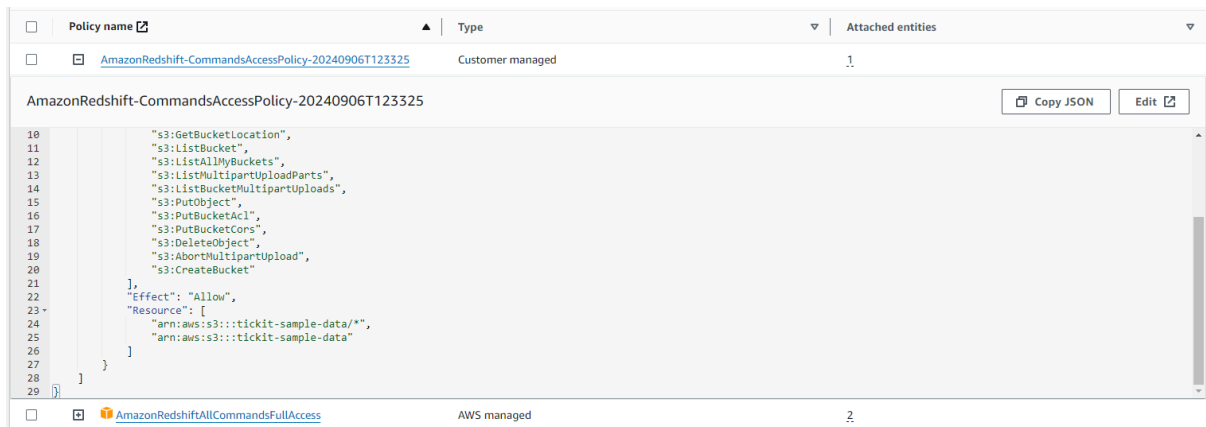
1

<input checked="" type="checkbox"/>	Name
<input checked="" type="checkbox"/>	tickit-sample-data

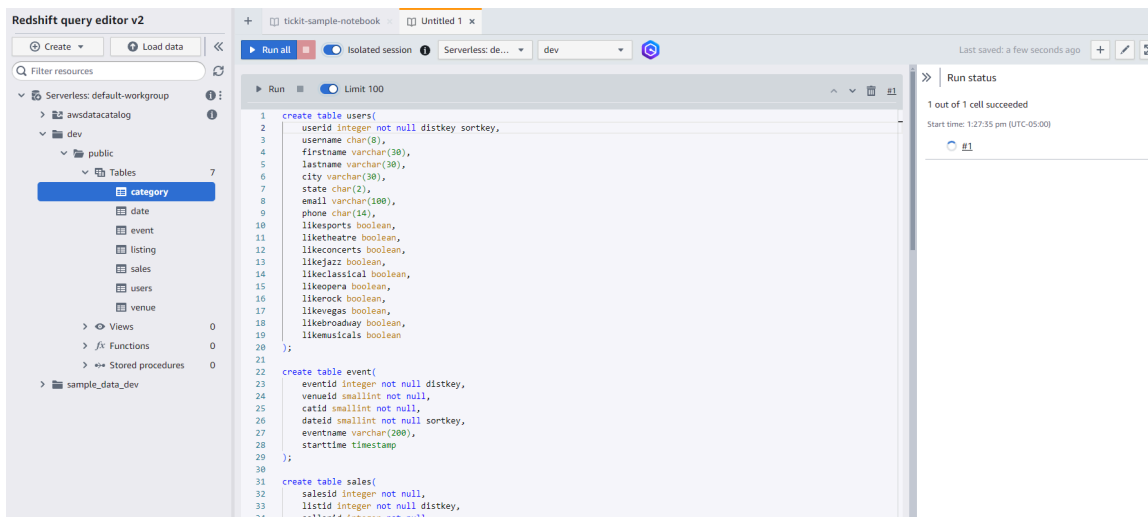
An IAM role AmazonRedshift-CommandsAccessRole-20231126T121240 is currently assigned as the default role. If you create an IAM role as default, this new IAM role replaces AmazonRedshift-CommandsAccessRole-20231126T121240 as the default role.

Cancel Create IAM role as default

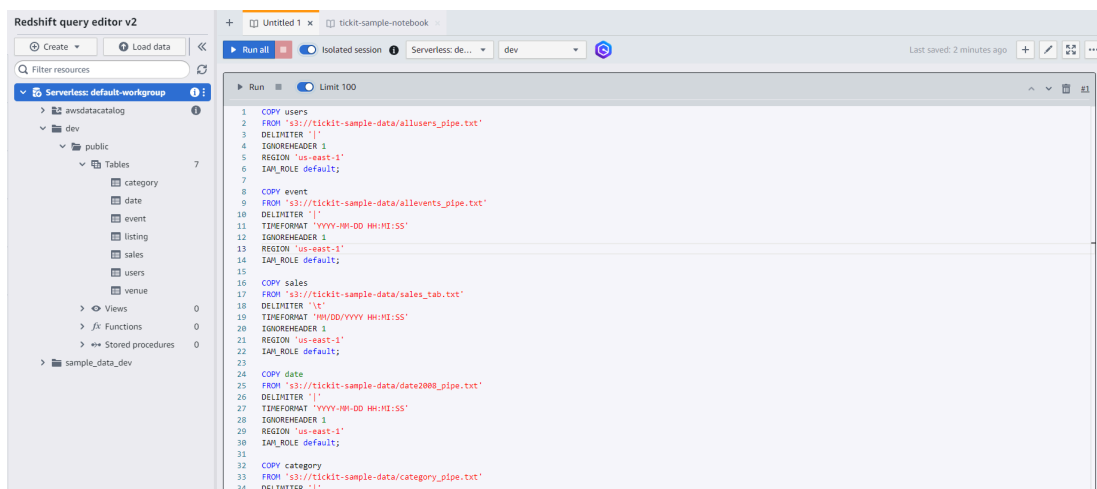
Como se puede ver, el rol queda apuntando al bucket con los permisos correspondientes, como también con el acceso total a Redshift.



Vamos al query editor y creamos las 7 tablas en un script de sql.



Una vez creadas, podemos ingresar los datos a través del comando COPY, apuntando a cada archivo en el bucket de S3.



Algunos comandos a destacar son: DELIMITER (caracter con el que se delimita cada columna), IGNOREHEADER 1 (no tomar como header la 1ra fila del archivo).

Row 3, Col 9, Chr 165

Result 1

Export Chart

Summary

Info:

- Load into table 'date' completed, 364 record(s) loaded successfully.

Returned rows: 0  
Query ID: 7437  
Elapsed time: 3.8s  
Result set query:

Respondemos las preguntas propuestas.

Run Limit 100

```

1 -- ¿Cuántos Usuarios gustan del Jazz?
2 SELECT COUNT(1) FROM USERS
3 WHERE likejazz = TRUE;

```

Result 1 (1)

count
12441

Run Limit 100

```

1 -- b. ¿Cuántos Usuarios gustan de la ópera y del rock al mismo tiempo?
2 SELECT COUNT(1) FROM USERS
3 WHERE likeopera = TRUE AND likerock = TRUE;

```

Result 1 (1)

count
3113

Run Limit 100

```

1 -- c. ¿Cuál es el promedio, moda y mediana del total de Ventas?
2 SELECT
3     AVG(pricepaid) AS promedio_ventas,
4     PERCENTILE_CONT(0.5) WITHIN GROUP (ORDER BY pricepaid) AS mediana_ventas,
5     (
6         SELECT pricepaid
7         FROM SALES
8         GROUP BY pricepaid
9         HAVING COUNT(1) = (
10            SELECT MAX(price)
11            FROM (
12                SELECT pricepaid, COUNT(1) as price
13                FROM SALES
14                GROUP BY pricepaid
15            )
16        )
17     ) AS moda_ventas
18 FROM SALES;

```

Result 1 (1)

promedio_ventas	mediana_ventas	moda_ventas
642.28	386	144

Run Limit 100

```

1 -- d. ¿Cuál el promedio de ventas de usuarios que gustan del rock, pero NO del Jazz?
2 SELECT AVG(s.pricepaid) AS promedio_ventas
3 FROM USERS u
4 INNER JOIN (SELECT buyerid, pricepaid FROM SALES) s
5 ON u.userid = s.buyerid
6 WHERE u.likerock = TRUE AND u.likejazz = FALSE;

```

Result 1 (1)

promedio_ventas
648.86

Realizamos la recopilación de datos usando diferentes tablas y almacenamos todo en una única tabla.

The screenshot shows the Redshift query editor v2 interface. On the left, a sidebar displays the database structure with a tree view showing 'Serverless: default-workgroup', 'awsdatacatalog', 'dev', 'public', and 'Tables'. The 'Tables' section is expanded, showing a list of tables including 'category', 'date', 'event', 'full\_data', 'listing', 'sales', 'users', and 'venue'. The main editor area contains a SQL query that creates a table named 'full\_data' and inserts data from various tables. The query is as follows:

```
1 /*Nombre de usuario, Apellido de
2 usuario, Correo de usuario, Nombre del evento, lugar del evento, Fecha
3 del evento, Cantidad y Total vendidos*/
4
5 CREATE TABLE full_data AS (
6   SELECT
7     u.firstname AS nombre_usuario,
8     u.lastname AS apellido_usuario,
9     u.email AS correo_usuario,
10    e.eventname AS nombre_evento,
11    v.venue AS lugar_evento, -- v.venue AS lugar_evento
12    e.starttime AS fecha_evento,
13    s.qty AS cantidad,
14    s.pricepaid AS total
15  FROM USERS u
16  INNER JOIN (SELECT buyerid, eventid, qty, pricepaid FROM SALES) s
17  ON u.buyerid = s.buyerid
18  INNER JOIN (SELECT eventid, venueid, eventname, starttime FROM EVENT) e
19  ON s.eventid = e.eventid
20  INNER JOIN (SELECT venueid, venue AS lugar_evento FROM VENUE) v
21  ON e.venueid = v.venueid
22 );
```

Below the query, the 'Result 1' summary is displayed, showing 'Returned rows: 0', 'Query ID: 11101', 'Elapsed time: 2s', and 'Result set query:'.

The screenshot shows the Redshift query editor v2 interface with a SQL query that selects all data from the 'full\_data' table. The query is as follows:

```
1 SELECT * FROM FULL_DATA;
```

Below the query, the 'Result 1 (100)' is displayed, showing a table with 7 columns: 'nombre\_usuario', 'apellido\_usuario', 'email', 'nombre\_evento', 'lugar\_evento', 'fecha\_evento', and 'cantidad'. The table contains 100 rows of data, with the first 10 rows visible in the screenshot.

	nombre_usuario	apellido_usuario	email	nombre_evento	lugar_evento	fecha_evento	cantidad
<input type="checkbox"/>	Hyatt	Blair	risus.Nunc@sociosquadt...	The Glass Menagerie	Shubert Theatre	2008-09-14 19:30:00	1
<input type="checkbox"/>	Damian	Brock	odio@habitantmorbistrit...	Sarah Brightman	Heinz Field	2008-10-10 19:00:00	4
<input type="checkbox"/>	Kameko	Schwartz	a@nec.com	Jack Johnson	Mellon Arena	2008-10-23 14:00:00	4
<input type="checkbox"/>	Martina	Albert	blandit.mattis@consectet...	U2	Arrowhead Stadium	2008-12-25 15:00:00	3
<input type="checkbox"/>	Simone	Baird	Sed.auctor.odio@aliquam...	Built To Spill	BankAtlantic Center	2008-04-12 19:00:00	2
<input type="checkbox"/>	Bell	Spence	nibh.dolor.nonummy@eg...	Spring Awakening	Shubert Theatre	2008-06-26 15:00:00	4
<input type="checkbox"/>	Sonya	Dunn	in@Aliquamrutrumlorem...	Boris Godunov	Metropolitan Opera	2008-06-07 14:30:00	2
<input type="checkbox"/>	Xaviera	Walters	nibh@est.org	Sweet Honey in the Rock	Sommet Center	2008-06-24 19:30:00	4
<input type="checkbox"/>	Vance	Stephens	Aenean.gravida@sapien.ca	Celine Dion	Cleveland Browns Stadium	2008-06-22 19:00:00	4

At the bottom of the screenshot, the query details are displayed: 'Query ID 11047', 'Elapsed time: 2760 ms', and 'Total rows: 100'.

Para exportar esta tabla al bucket de S3 se usa el comando UNLOAD.

The screenshot shows the Redshift query editor v2 interface with a SQL query that unloads data from the 'full\_data' table to an S3 bucket. The query is as follows:

```
1 UNLOAD ('SELECT * FROM FULL_DATA')
2 TO 's3://ticket-sample-data/redshift/'
3 IAM_ROLE DEFAULT
4 HEADER
5 DELIMITER ','
6 PARALLEL OFF
7 GZIP;
```

Below the query, the 'Result 1' summary is displayed, showing 'Info:' and 'UNLOAD completed, 168808 record(s) unloaded successfully.'

Algunos comandos a destacar son: DELIMITER (caracter con el que se delimita cada columna), HEADER (indica que se debe incluir la primera fila como header), PARALLEL OFF (desactivar el procesamiento en paralelo, al ser un archivo pequeño), GZIP (comprime el archivo).

Los datos se ven de la siguiente forma.

	A	B	C	D	E	F	G	H
1	nombre_usuario	apellido_usuario	correo_usuario	nombre_evento	lugar_evento	fecha_evento	cantidad	total
2	Aaron	Garrison	nisl@duiFusce.ca	Carmen	Lyric Opera House	12/14/2008 15:00	1	552
3	Samson	Wiley	ante.iaculis@vitaedolorDonec.edu	Rigoletto	Metropolitan Opera	2/21/2008 19:00	1	275
4	Adrian	Bolton	ligula.consectetuer.rhoncus@mi.org	Regina Spector	Kauffman Stadium	2/1/2008 14:00	4	544
5	Garth	Frank	neque.In@maurisMorbinon.com	Tristan und Isolde	Lyric Opera House	2/28/2008 14:30	1	421
6	Kennan	Wagner	In.lorem.Donec@Maecenas.org	Chromeo	Sommet Center	3/15/2008 14:00	2	316
7	Burke	Suarez	lobortis.risus.In@mollisDuisit.org	Rusted Root	North Charleston Coliseum	5/1/2008 19:00	2	168
8	Price	Humphrey	Cras.interdum.Nunc@primisinfaucibus.com	Patty Griffin	Heinz Field	5/1/2008 19:30	1	62
9	Kelsey	Mcmahon	sagittis.Nullam@odiotristique.edu	Paul McCartney	Qwest Field	6/29/2008 14:00	2	264
10	Madonna	Rich	lorem@famesacturpis.org	Eddie Money	Reliant Stadium	7/4/2008 15:00	4	472
11	David	Maddox	posuere.cubilia@accumsan.com	Woyzeck	Majestic Theatre	8/20/2008 19:00	3	1269
12	Hayfa	Carpenter	Sed@ideratEtiam.com	Dirty Dancing	Studio 54	9/22/2008 14:00	1	160
13	Kelly	Rodgers	eu.placerat@dui.ca	Wicked	Charles Playhouse	11/1/2008 19:30	2	70
14	Thane	Houston	dolor.nonummy@mauriseuelit.com	Gypsy	Lincoln Center for the Performing Arts	2/1/2008 14:30	1	479
15	Vincent	Cannon	dolor.Fusce@Donecfelisorci.org	Othello	Cort Theatre	3/1/2008 19:00	1	98
16	Wendy	Sanford	lorem@inlobortis.org	L Elisir d Amore	Lyric Opera House	3/26/2008 19:30	1	494
17	Basil	Perry	at.velit.Cras@Cumsociisnatoque.ca	David Sanborn	Rogers Centre	3/27/2008 15:00	1	108
18	Noah	Roman	mus@eget.ca	Oasis	Fox Theatre	4/22/2008 19:30	1	69
19	Rhiannon	Dorsey	neque@penatibus.edu	Country Thunder USA	Hersheypark Stadium	6/17/2008 19:00	1	173
20	Drew	Holder	vulputate@sagittisfelisDonec.ca	Spring Awakening	August Wilson Theatre	6/5/2008 20:00	2	538
21	Eaton	Glover	ut@nunc.com	Linda Eder	Dodger Stadium	6/18/2008 19:00	2	632