



CS527 Project #1

PROF. SAED SAYAD

DEPARTMENT OF COMPUTER SCIENCE

SPRING 2020

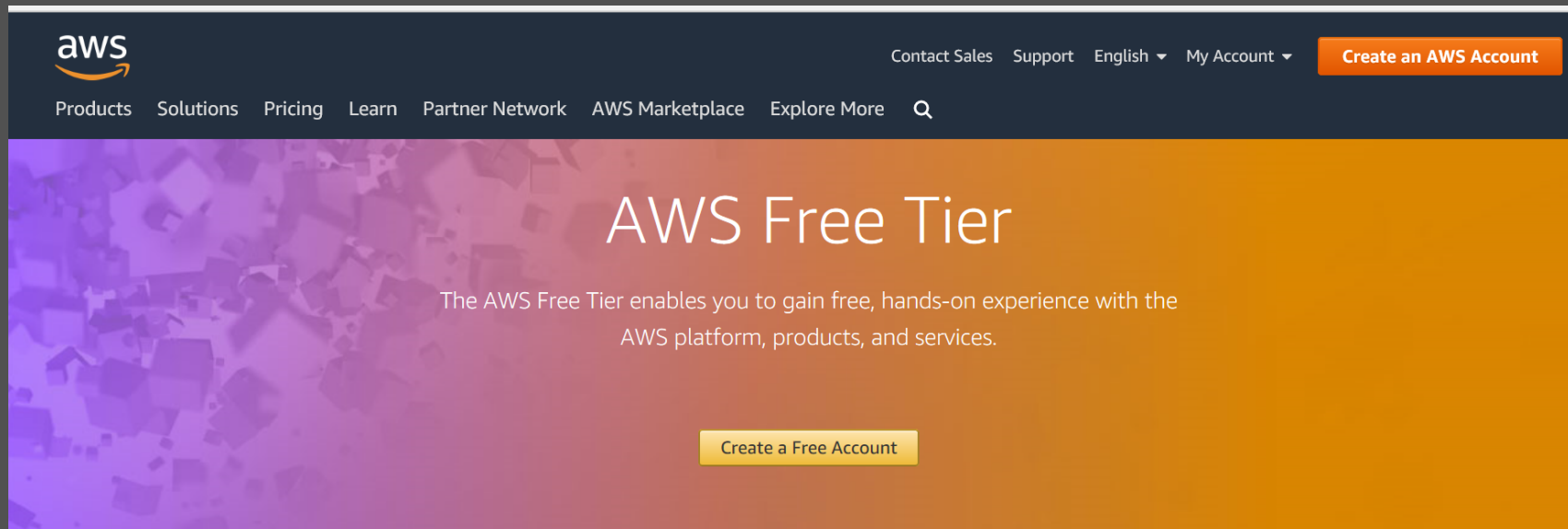
Build Your Team

1. Each team consists of 4-6 students.
2. Each team needs an administrator
3. Send your team info to Sakai.
4. Start Date: 2020-02-10
5. Presentation Date: 2020-03-02

Project – ETL (Extraction, Transformation and Loading)

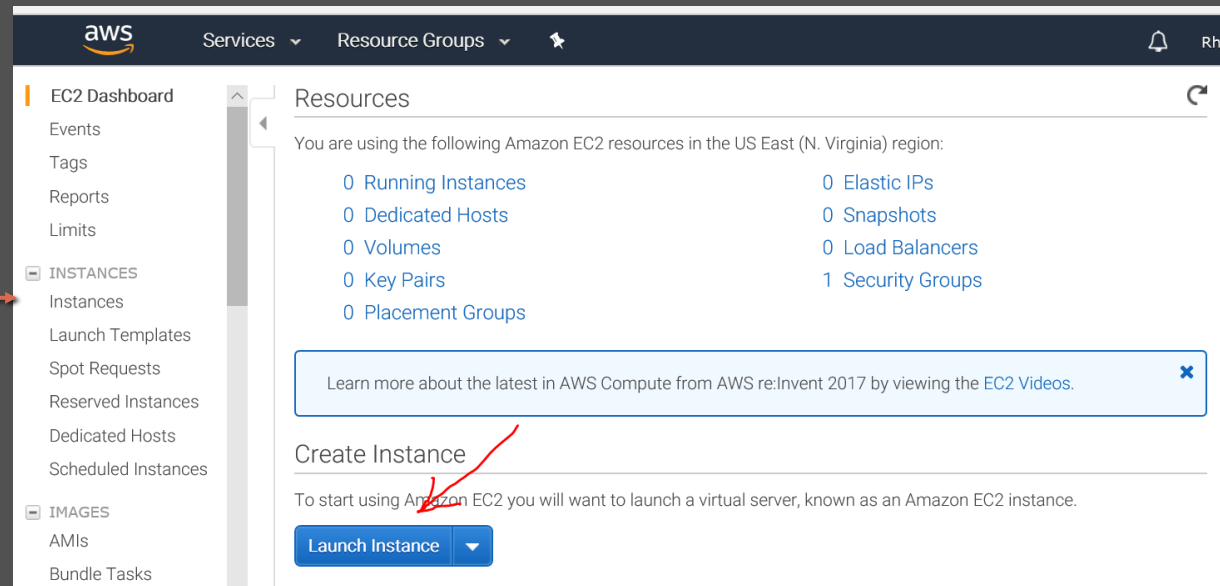
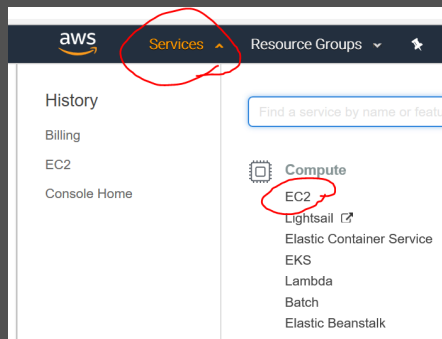
1. Create an AWS account (1)
2. Launch an EC2 micro instance (1)
3. Create a S3 Bucket (1)
4. Download Instacart (2)
5. Upload Instacart data to S3 (1)
6. Start a RDS (MySQL/ORACLE/SQL Server) Instance (1)
7. Export data from S3 to MySQL (3)
8. Start a Redshift Instance (1)
9. Export data from S3 to Redshift (3)

1. Create an AWS account

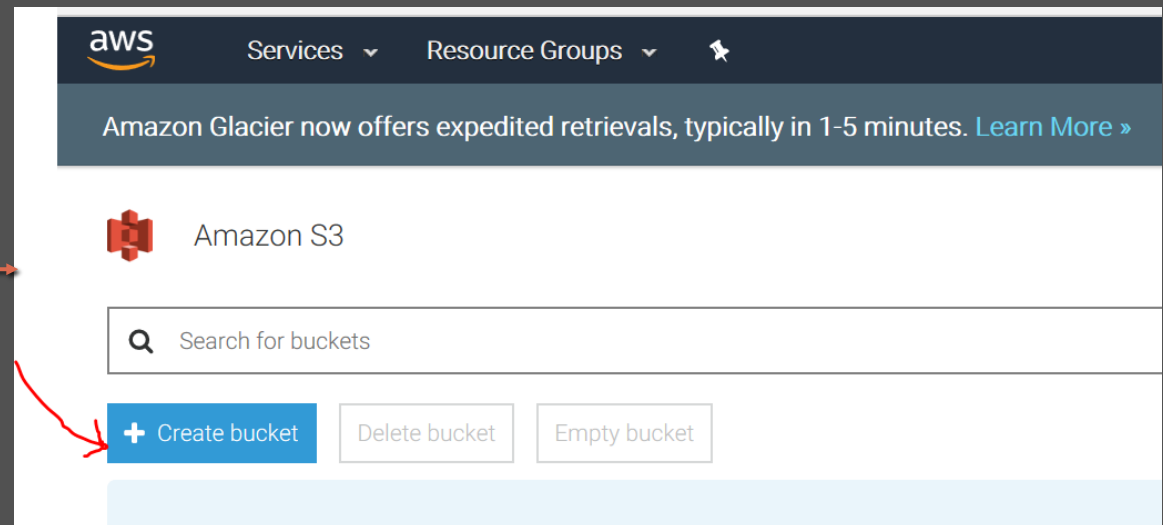
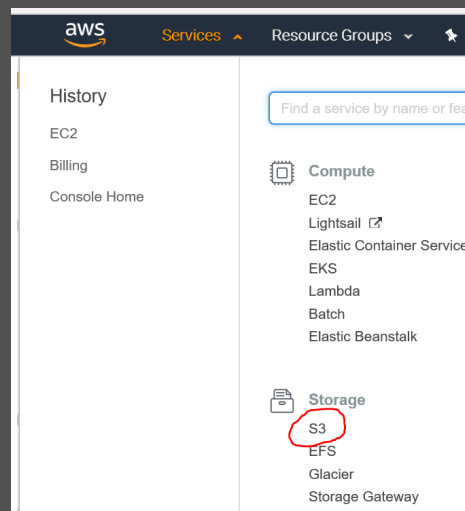


<https://aws.amazon.com/free/>

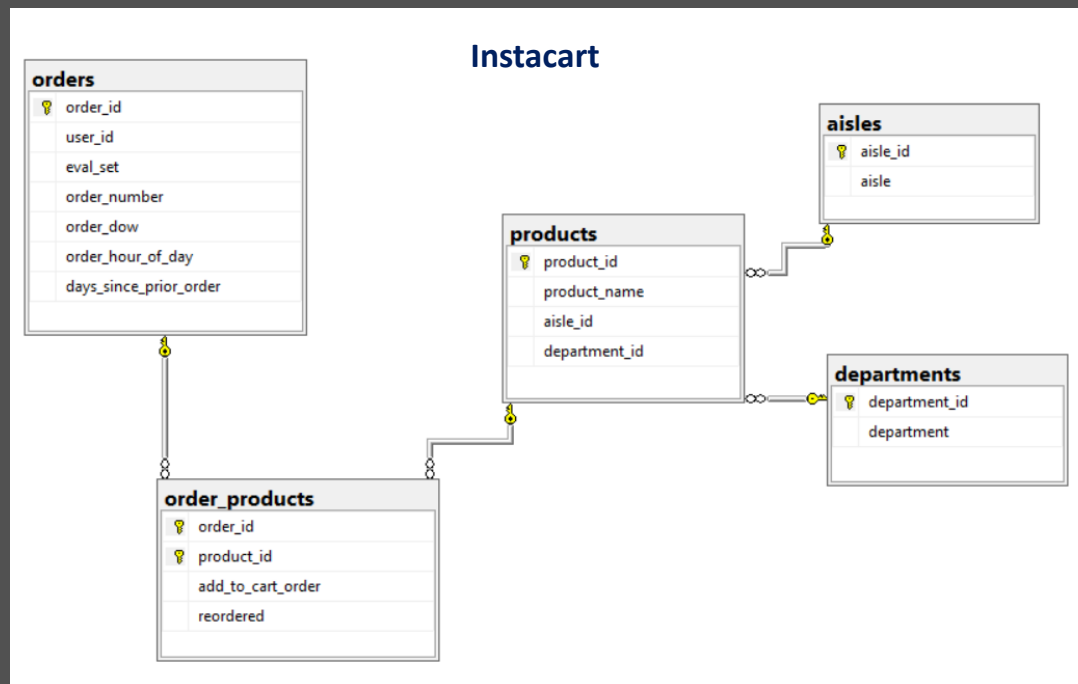
2. Launch an EC2 micro instance



3. Create a S3 Bucket



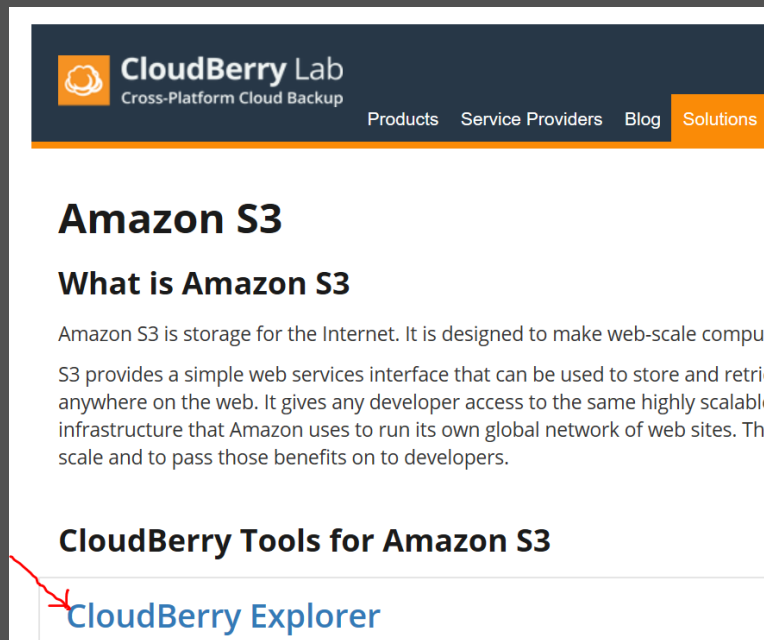
4. Download Data (Instacart & GEO)



aisles.csv
departments.csv
order_products.csv
orders.csv
products.csv

Instacart.zip

5. Upload Instacart data to S3



The screenshot shows the CloudBerry Lab website. The header includes the CloudBerry Lab logo and the tagline "Cross-Platform Cloud Backup". Navigation links for "Products", "Service Providers", "Blog", and "Solutions" are present, with "Solutions" highlighted in orange. The main content area is titled "Amazon S3" and "What is Amazon S3". It contains a paragraph describing Amazon S3 as storage for the Internet, designed for web-scale computing. Below this, another section is titled "CloudBerry Tools for Amazon S3", which includes a link to "CloudBerry Explorer" highlighted in blue. A red arrow points to this link.

CloudBerry Lab
Cross-Platform Cloud Backup

Products Service Providers Blog **Solutions**

Amazon S3

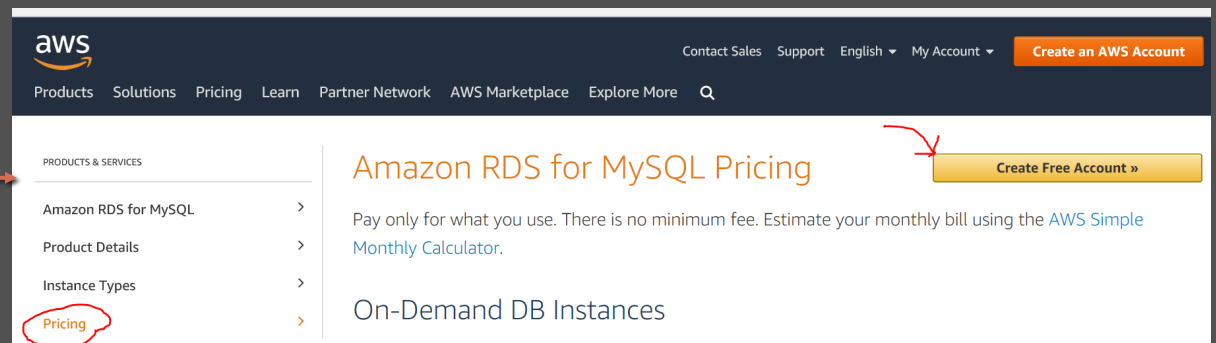
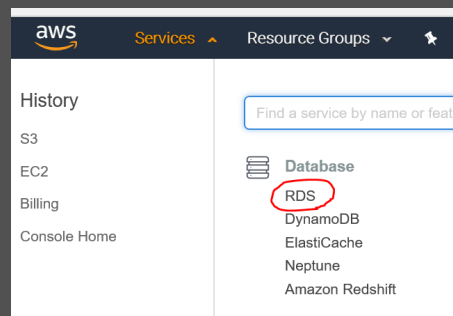
What is Amazon S3

Amazon S3 is storage for the Internet. It is designed to make web-scale computing simpler and more efficient. S3 provides a simple web services interface that can be used to store and retrieve anywhere on the web. It gives any developer access to the same highly scalable, infrastructure that Amazon uses to run its own global network of web sites. The scale and to pass those benefits on to developers.

CloudBerry Tools for Amazon S3

[CloudBerry Explorer](#)

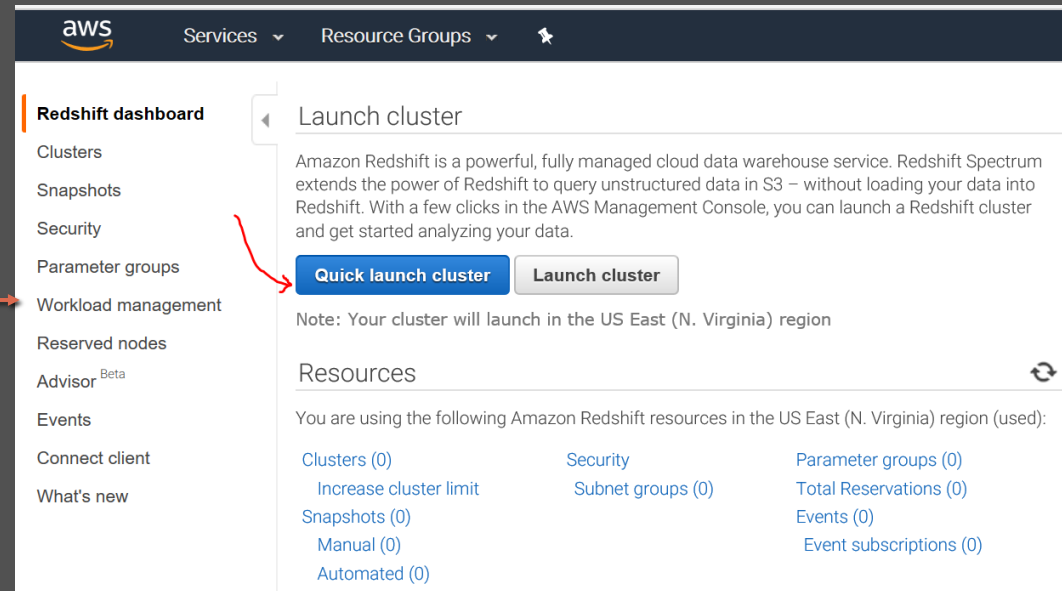
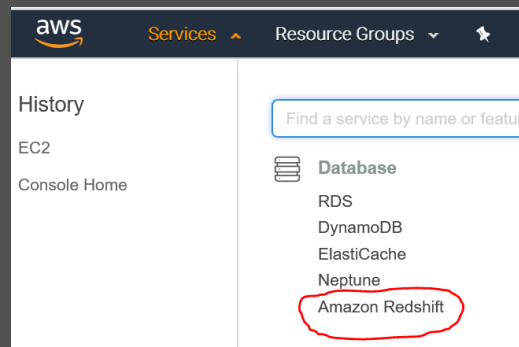
6. Start a RDS Instance



7. Export data from S3 to MySQL

- Use JDBC or ODBC Connectivity

8. Start a Redshift Instance



9. Export data from S3 to Redshift

- Use JDBC or ODBC Connectivity
- Use Redshift “COPY” function

[AWS Documentation](#) » [Amazon Redshift](#) » [Database Developer Guide](#) » [SQL Reference](#) » [SQL Commands](#) » COPY

COPY

Loads data into a table from data files or from an Amazon DynamoDB table. The files can be located in an Amazon Simple Storage Service (Amazon S3) bucket, an Amazon EMR cluster, or a remote host that is accessed using a Secure Shell (SSH) connection.

Project - Homework Assignments

- Practice the SQL scripts on the following website against your data on MySQL and Redshift.

SQL Tutorial > <https://www.w3schools.com/sql/default.asp>

- Students will be asked to write 10 different SQL codes (10)

Project - Web Application

- The web application should be able to connect to both MySQL and Redshift (4)
- The web app accepts SQL scripts as an input (1)
- The web app executes SQL scripts and shows the correct results (10)
- The web app shows the elapsed time (1)

Web Application

☒ MySQL ☐ Redshift

Query:

Run

Time Elapsed

Project - Marking

Project	Team	Individual
ETL Process	14	
Homework Assignments		10
Web Application	16	
<i>Total</i>	40	

Project #2



mongoDB



*Good
Luck!*