



# AWS Lake Formation

## Hands-On Lab

Sep, 2019

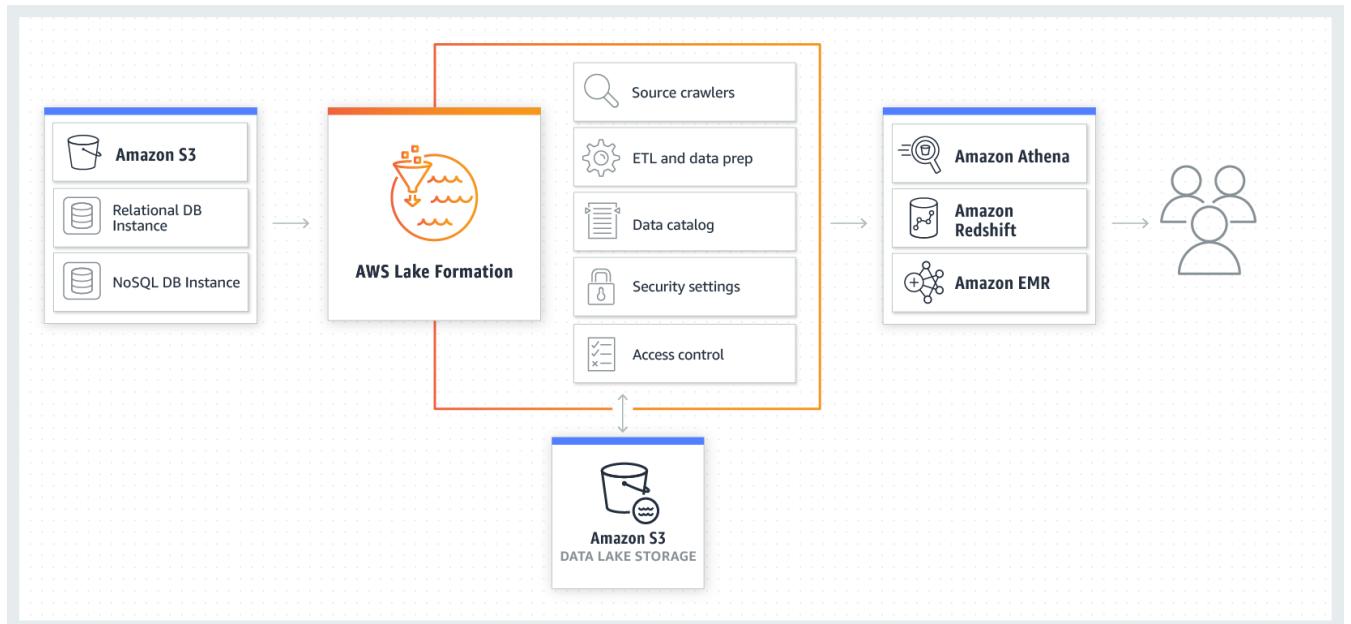
Amazon Web Services

## Table of Contents

AWS Lake Formation.....	3
Part 1. 환경 설정 .....	3
Task 1.1 : AWS Management Console 접속 및 리전 선택 .....	3
Task 1.2 : VPC 생성 .....	4
Task 1.3 : EC2 생성 .....	8
Task 1.4 : Database Table 구성 .....	15
Task 1.5 : Data Lake 를 위한 IAM Role 및 User 생성 .....	17
Part 2. JDBC Data Source 로부터 Data Lake 생성.....	32
Task 2.1 : AWS Glue Connection 설정 .....	32
Task 2.2 : Amazon S3 Bucket 생성.....	38
Task 2.3 : Data Lake Administrator 생성 .....	39
Task 2.4 : Amazon S3 경로 등록 .....	40
Task 2.5 : Data Location Permission 부여 .....	42
Task 2.6 : Data Catalog 의 Database 생성 .....	43
Task 2.7 : Data Permissions 부여 .....	44
Task 2.8 : ETL Workflow 생성 .....	46
Task 2.9 : SELECT Table 권한 부여 .....	50
Task 2.10 : Amazon Athena 로 Data Lake 조회 .....	52
Task 2.11 : Amazon Redshift Spectrum 으로 Data Lake 조회하기 .....	54
Part 3. CSV 파일로부터 Data Lake 생성 .....	62
Task 3.1 : Sample dataset 파일 업로드.....	62
Task 3.2 : Data Lake 설정하기 .....	63
Task 3.3 : Table 접근 권한 부여하기.....	71
Task 3.4 : Athena 로 데이터 조회하기 .....	72

## AWS Lake Formation

AWS Lake Formation 은 AWS Glue, Amazon Athena, Amazon Redshift, Amazon EMR(beta) 서비스와 통합되어 있으며, Data Catalog 에 대한 접근권한 관리를 중앙집중화 합니다.



이번 실습에서는 아래와 같은 사용자 및 역할을 가상으로 가정합니다.

사용자 및 역할	설명
IAM Administrator (Superuser)	IAM User/Role, S3 버킷을 생성하는 User AdministratorAccess 권한을 가짐
Data Lake Administrator	Data Catalog, Database 를 생성/접근할 수 있는 User Lake Formation 권한을 다른 User에게 부여할 수 있는 권한 소유 Data Lake 관리자(IAM Administrator 보다는 적은 권한 가짐)
Data Lake Analyst	Data Lake 를 조회만 할 수 있는 권한을 가진 User
Workflow role	ETL Workflow 를 수행하기 위한 권한을 가진 Role

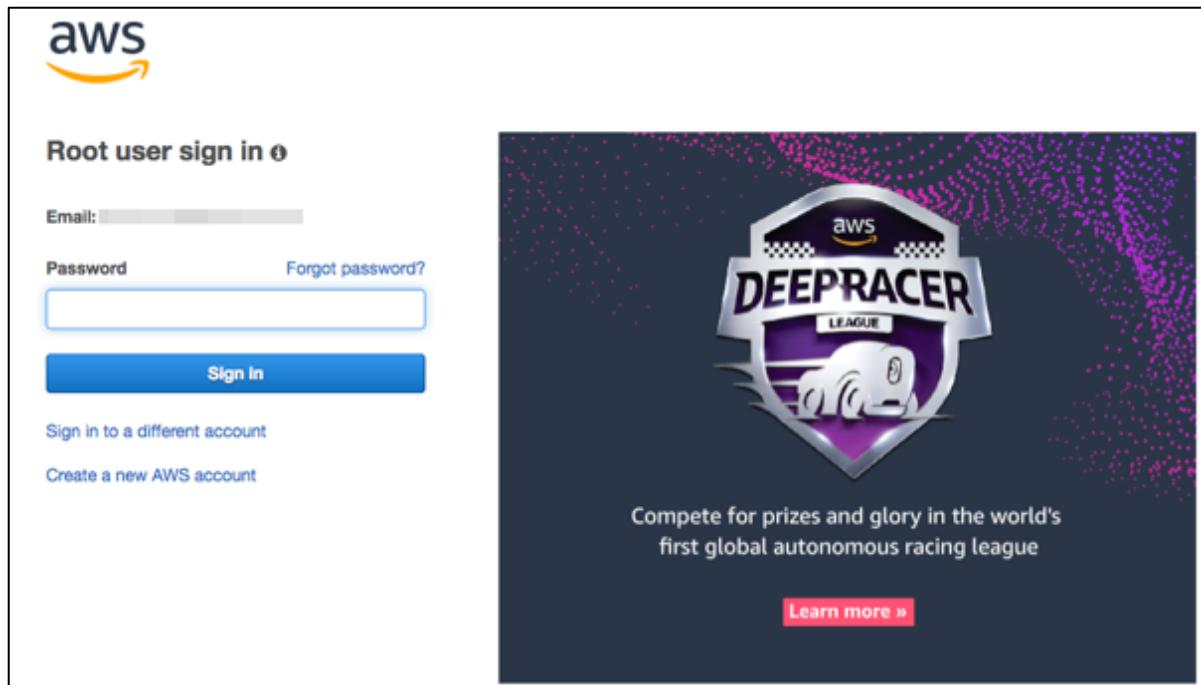
## Part 1. 환경 설정

Lake Formation 을 이용하여 Data Lake 를 생성하고, 사용자별 Data Lake 접근권한 설정에 대해 실습합니다.

먼저 Data Source 로서 Oracle Database 구성을 위한 환경설정을 합니다.

### Task 1.1 : AWS Management Console 접속 및 리전 선택

1. AWS Management Console 로 접속하십시오.(<https://console.aws.amazon.com>)



2. 오른쪽 상단에서 Region Name 을 “N. Virginia”로 선택합니다.

The screenshot shows the AWS Management Console homepage. On the right, there is a vertical navigation bar with a dropdown menu for 'Region'. The 'US East (N. Virginia)' option is highlighted with a red box. Other options listed include US East (Ohio), US West (N. California), US West (Oregon), Asia Pacific (Hong Kong), Asia Pacific (Mumbai), Asia Pacific (Seoul), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Tokyo), Canada (Central), EU (Frankfurt), EU (Ireland), EU (London), EU (Paris), EU (Stockholm), Middle East (Bahrain), and South America (São Paulo).

## Task 1.2 : VPC 생성

1. Elastic IP 를 생성합니다.

**VPC Dashboard**

Filter by VPC:

Select a VPC

**Launch VPC Wizard** **Launch EC2 Instances**

Note: Your Instances will launch in the US East (N. Virginia) region.

### Resources by Region

Refresh Resources

You are using the following Amazon VPC resources

Resource Type	Region	Count
VPCs	N. Virginia	1
Subnets	N. Virginia	6
Route Tables	N. Virginia	1
Internet Gateways	N. Virginia	1

**Elastic IPs**

Endpoints

Endpoint Services

NAT Gateways

**Allocate new address** Actions

Filter by tags and attributes or search by keyword

None found

You do not have any Addresses in this region

Click the Create Address button to create your first Address

**Allocate new address**

**Addresses > Allocate new address**

### Allocate new address

Allocate a new Elastic IP address by selecting the scope in which it will be used

Scope: VPC

IPv4 address pool:  Amazon pool  Owned by me

\* Required

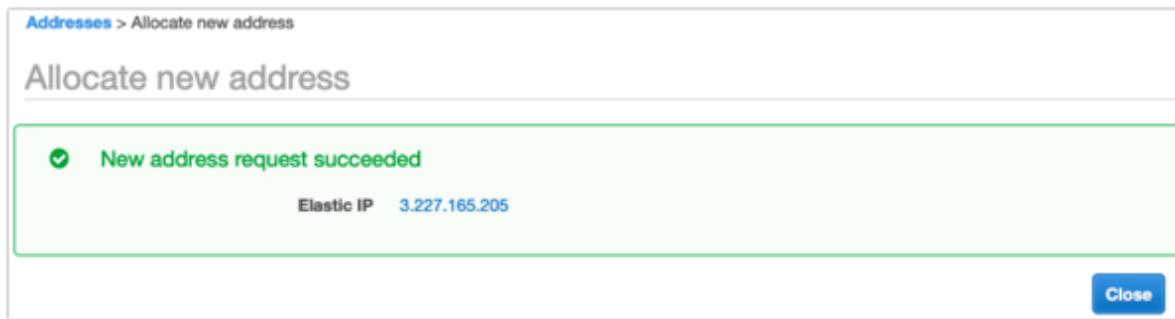
**Cancel** **Allocate**

**Addresses > Allocate new address**

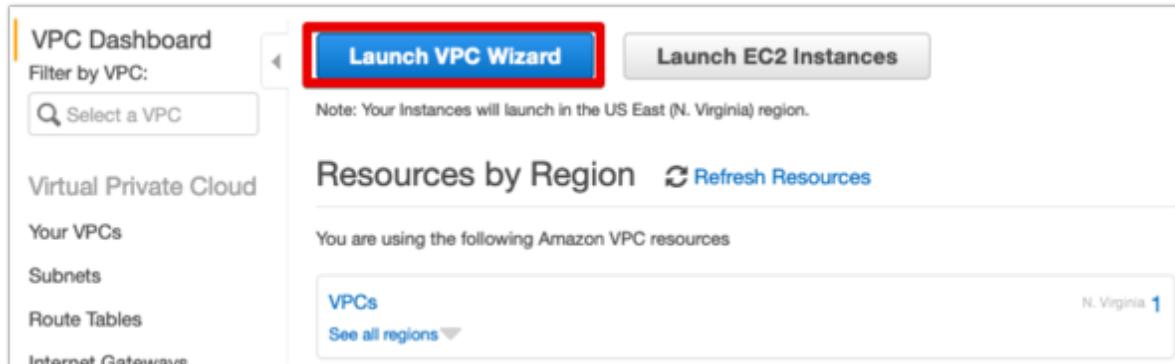
### Allocate new address

**New address request succeeded**

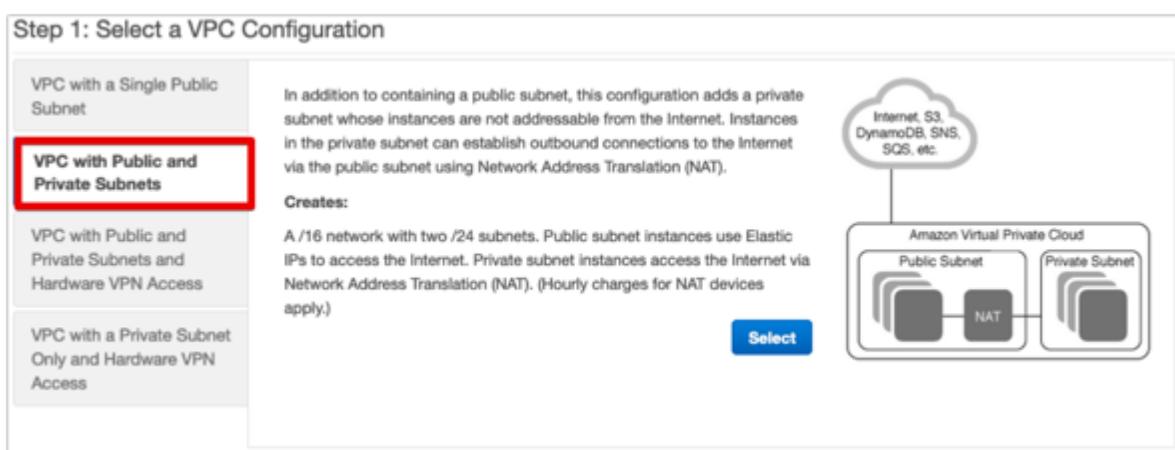
Elastic IP: 18.205.42.175



2. Launch VPC Wizard 를 선택합니다.



3. “Select a VPC Configuration”에서 “VPC with Public and Private Subnets”을 선택합니다.



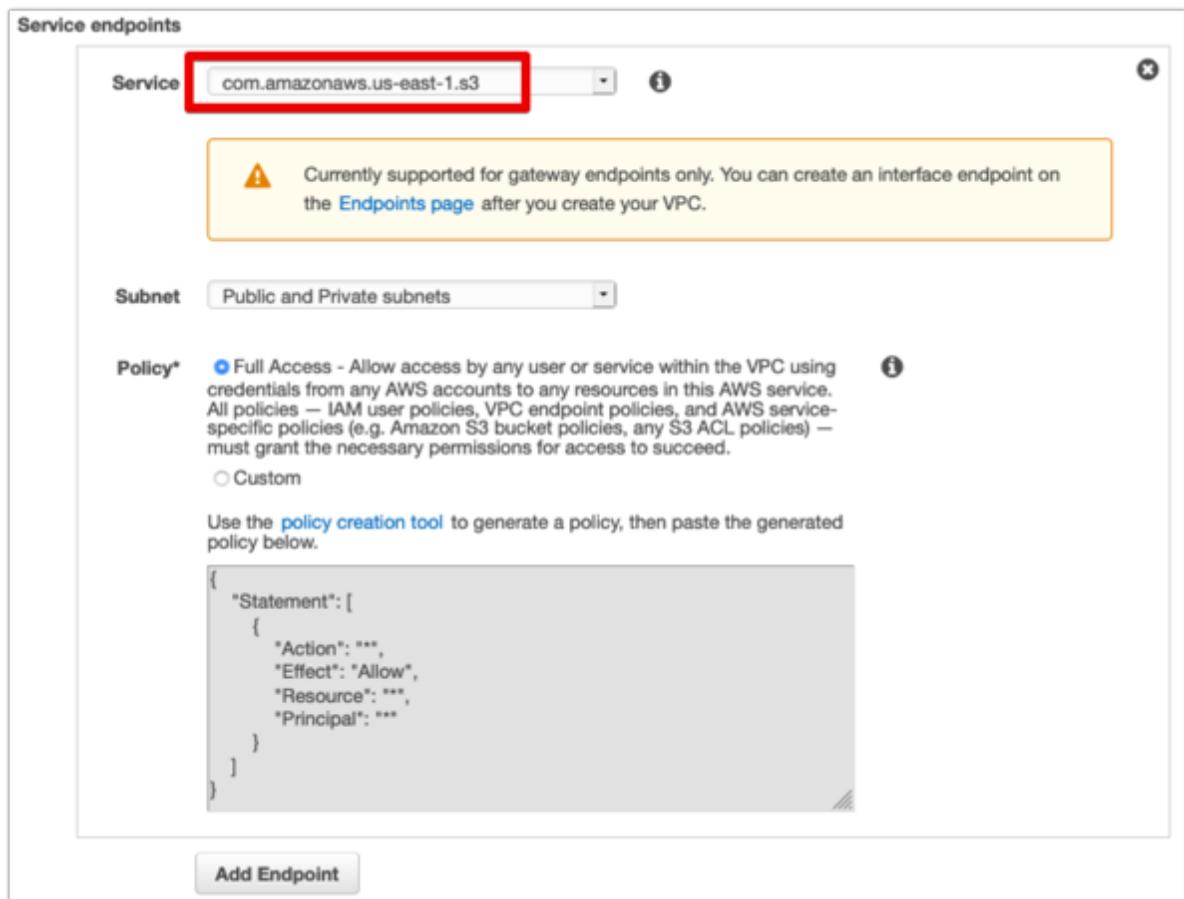
4. “VPC with Public and Private Subnets”에서 VPC 관련 정보를 입력합니다.

IPv4 CIDR	10.0.0.0/16
VPC name	DataLake-VPC
Public subnet	CIDR : 10.0.0.0/24 Availability Zone : us-east-1a Public subnet name : Public subnet
Private subnet	CIDR : 10.0.1.0/24 Availability Zone : us-east-1a Public subnet name : Private subnet
Elastic IP	미할당된 EIP 를 지정합니다.

Step 2: VPC with Public and Private Subnets

IPv4 CIDR block: 10.0.0.0/16 (65531 IP addresses available)
IPv6 CIDR block: <input checked="" type="radio"/> No IPv6 CIDR Block <input type="radio"/> Amazon provided IPv6 CIDR block
VPC name: DataLake-VPC
Public subnet's IPv4 CIDR: 10.0.0.0/24 (251 IP addresses available)
Availability Zone: us-east-1a
Public subnet name: Public subnet
Private subnet's IPv4 CIDR: 10.0.1.0/24 (251 IP addresses available)
Availability Zone: us-east-1a
Private subnet name: Private subnet
You can add more subnets after AWS creates the VPC.
Specify the details of your NAT gateway ( <a href="#">NAT gateway rates apply</a> ).
Elastic IP Allocation ID: eipalloc-01166dda0e2231acc
Service endpoints
<b>Add Endpoint</b>
Enable DNS hostnames: <input checked="" type="radio"/> Yes <input type="radio"/> No
Hardware tenancy: Default

5. “Add Endpoint”를 클릭하여, S3 를 위한 VPC Endpoint 를 생성합니다.  
S3 를 위한 VPC Endpoint 는 AWS Glue 에서 Source Data 를 S3 에 저장할 때 필요합니다.



모든 설정이 완료된 후, “Create VPC” 버튼을 클릭합니다.

VPC 생성이 완료된 후, Subnet, Route Table, Endpoint 등이 정상적으로 생성되었는지 확인합니다.

### Task 1.3 : EC2 생성

- EC2 를 생성하기 위해 EC2 Console 로 접속합니다.

여기서 생성하는 EC2 는 DataLake 의 Source 로서, Oracle Database 가 포함되어 있는 AMI 를 이용할 것입니다.

The screenshot shows the AWS EC2 Dashboard. On the left, there's a sidebar with links like 'EC2 Dashboard', 'Events', 'Tags', 'Reports', 'Limits', 'INSTANCES' (with sub-links for Instances, Launch Templates, Spot Requests, Reserved Instances, Dedicated Hosts, Scheduled Instances, Capacity Reservations), and 'IMAGES' (with sub-links for AMIs, Bundle Tasks). The main panel is titled 'Resources' and says 'You are using the following Amazon EC2 resources in the US East (N. Virginia) region:'. It lists '0 Running Instances', '0 Dedicated Hosts', '0 Volumes', '0 Key Pairs', and '0 Placement Groups'. Below this is a link to learn more about AWS Compute. A large blue button labeled 'Launch Instance' is prominently displayed.

## 2. AMI(ami-0a44c21c)를 선택합니다.

This screenshot shows the 'Step 1: Choose an Amazon Machine Image (AMI)' step of a wizard. At the top, it says '1. Choose AMI' is selected, along with other steps: '2. Choose Instance Type', '3. Configure Instance', '4. Add Storage', '5. Add Tags', '6. Configure Security Group', and '7. Review'. The main area has a search bar with 'ami-0a44c21c' typed in. To the left is a sidebar with 'Quick Start (0)', 'My AMIs (0)', 'AWS Marketplace (3637)', 'Community AMIs (1)', and a checkbox for 'Free tier only'. To the right, it says 'No results were found for "ami-0a44c21c" in the quick start catalog.' Below that, it says 'The following results for "ami-0a44c21c" were found in other catalogs:'. It shows '3637 results in AWS Marketplace' and '1 results in Community AMIs'. The 'Community AMIs' result is highlighted with a red box and says 'Community AMIs are AMIs that are shared by the general AWS community'.

This screenshot shows the same 'Step 1: Choose an Amazon Machine Image (AMI)' screen, but now the 'Community AMIs (1)' section is expanded. It shows one result: 'oracledb12c\_dms - ami-0a44c21c'. This result is also highlighted with a red box. To the right of the result, it shows 'Select' and '64-bit (x86)'. At the bottom right of the expanded section, there are navigation arrows for '1 to 1 of 1 AMIs'.

AMI : ami-0a44c21c 에는 Oracle Database 12C 가 포함되어 있습니다.

Step 2: Choose an Instance Type

General purpose	r5dn.24xlarge	96	768	4 x 900 (SSD)	Yes	100 Gigabit	Yes
<input checked="" type="checkbox"/> General purpose	m5.large	2	8	EBS only	Yes	Up to 10 Gigabit	Yes
General purpose	m5.xlarge	4	16	EBS only	Yes	Up to 10 Gigabit	Yes

[Cancel](#) [Previous](#) [Review and Launch](#) [Next: Configure Instance Details](#)

Instance Type 에는 m5.large 를 선택하고, “Next:Configure Instance Details”를 클릭합니다.

Step 3: Configure Instance Details

Configure the instance to suit your requirements. You can launch multiple instances from the same AMI, request Spot instances to take advantage of the lower pricing, assign an access management role to the instance, and more.

Number of instances	<input type="text" value="1"/>	<a href="#">Launch into Auto Scaling Group</a>
Purchasing option	<input type="checkbox"/> Request Spot instances	
Network	vpc-0dccb1d202a01ea3   DataLake-VPC	<a href="#">Create new VPC</a>
Subnet	subnet-05e2d991c296309e0   Private subnet   us-e	<a href="#">Create new subnet</a>
Auto-assign Public IP	<input type="checkbox"/> Use subnet setting (Disable)	
Placement group	<input type="checkbox"/> Add instance to placement group	
Capacity Reservation	<a href="#">Open</a>	<a href="#">Create new Capacity Reservation</a>
IAM role	<input type="text" value="None"/>	<a href="#">Create new IAM role</a>

Network 에는 이전 단계에서 생성한 VPC 를 선택하고, Subnet 에는 Private subnet 을 선택합니다.

EC2 Instance 에 System Manager 로 접속할 수 있는 권한을 부여하기 위해 “Create new IAM role”을 클릭합니다.

[Create role](#) [Delete role](#)

Showing 4 results

Role name	Description	Trusted entities
<input type="checkbox"/> AWSServiceRoleForOrganizations	Service-linked role used by AWS Organizations to enable integration...	AWS service: organizations (Service-Linked r...
<input type="checkbox"/> AWSServiceRoleForSupport	Enables resource access for AWS to provide billing, administrative a...	AWS service: support (Service-Linked role)
<input type="checkbox"/> AWSServiceRoleForTrustedAdvisor	Access for the AWS Trusted Advisor Service to help reduce cost, inc...	AWS service: trustedadvisor (Service-Linked ...
<input type="checkbox"/> OrganizationAccountAccessRole		Account: 967273450079

Create role

Select type of trusted entity

<b>AWS service</b> EC2, Lambda and others	<b>Another AWS account</b> Belonging to you or 3rd party	<b>Web identity</b> Cognito or any OpenID provider	<b>SAML 2.0 federation</b> Your corporate directory
--	---	---	--

Allows AWS services to perform actions on your behalf. [Learn more](#)

Choose the service that will use this role

<b>EC2</b> Allows EC2 instances to call AWS services on your behalf.	Lambda Allows Lambda functions to call AWS services on your behalf.	API Gateway	CodeBuild	EMR	Lambda	S3
---	--	-------------	-----------	-----	--------	----

\* Required

**Cancel** **Next: Permissions**

“Choose the service…”에 EC2 를 선택하고, “Next:Permissions”를 클릭합니다.

Create role

Attach permissions policies

Choose one or more policies to attach to your new role.

**Create policy**

**Filter policies** ▾  Showing 3 results

Policy name	Used as
<input type="checkbox"/> AmazonEC2RoleforAWSCodeDeploy	None
<input type="checkbox"/> AmazonEC2RoleforDataPipelineRole	None
<input checked="" type="checkbox"/> AmazonEC2RoleforSSM	None

\* Required

**Cancel** **Previous** **Next: Tags**

“Attach permissions policies”에서 “AmazonEC2RoleforSSM”을 검색하여 선택합니다.

Create role

Review

Provide the required information below and review this role before you create it.

Role name*	AmazonEC2RoleSSM
Use alphanumeric and '+-=_,@-_.' characters. Maximum 64 characters.	
Role description	Allows EC2 instances to call AWS services on your behalf.

**Cancel** **Previous** **Create role**

Role name 에는 ”AmazonEC2RoleforSSM”를 입력하고, “Create role”을 클릭합니다.

Step 3: Configure Instance Details

Placement group  Add instance to placement group

Capacity Reservation  Open

IAM role  **AmazonEC2RoleSSM**

CPU options  Specify CPU options

Shutdown behavior  Stop

EC2 생성 화면으로 되돌아 와서, 새로 생성한 IAM role 을 선택합니다.  
기존에 동일한 권한을 가진 Role 이 있다면, 그 Role 을 선택해도 됩니다.

Step 3: Configure Instance Details

Advanced Details

User data  As text  As file  Input is already base64 encoded

```
#!/bin/bash
cd /tmp
sudo yum install -y https://s3.amazonaws.com/ec2-downloads-windows/SSMAgent/latest/linux_amd64/amazon-ssm-agent.rpm
sudo systemctl start amazon-ssm-agent
```

User data 필드에 아래와 같이 System Manager Agent 를 Install 하도록 입력합니다.

```
#!/bin/bash
cd /tmp
sudo yum install -y https://s3.amazonaws.com/ec2-downloads-windows/SSMAgent/latest/linux_amd64/amazon-ssm-agent.rpm
sudo systemctl start amazon-ssm-agent
```

Step 4: Add Storage

Your instance will be launched with the following storage device settings. You can attach additional EBS volumes and instance store volumes to your instance, or edit the settings of the root volume. You can also attach additional EBS volumes after launching an instance, but not instance store volumes. [Learn more](#) about storage options in Amazon EC2.

Volume Type	Device	Snapshot	Size (GiB)	Volume Type	IOPS	Throughput (MB/s)	Delete on Termination	Enc
Root	/dev/xvda	snap-015919a1d3ea94bd9	40	General Purpose	120 / 3000	N/A	<input checked="" type="checkbox"/>	Not
EBS	/dev/sdb	snap-06f0a8524c8e01	40	General Purpose	120 / 3000	N/A	<input type="checkbox"/>	Not

Add Storage 에는 변경없이 그대로 둔 상태에서 “Next: Add Tags”를 클릭합니다.

1. Choose AMI   2. Choose Instance Type   3. Configure Instance   4. Add Storage   5. Add Tags   6. Configure Security Group   7. Review

### Step 5: Add Tags

A tag consists of a case-sensitive key-value pair. For example, you could define a tag with key = Name and value = Webserver. A copy of a tag can be applied to volumes, instances or both. Tags will be applied to all instances and volumes. [Learn more](#) about tagging your Amazon EC2 resources.

Key	(128 characters maximum)	Value	(256 characters maximum)	Instances	Volumes
Name		Oracle Database		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

Add another tag (Up to 50 tags maximum)

Cancel Previous Review and Launch Next: Configure Security Group

Add Tags 에는 Key/Value 에 Name/Oracle Database 를 입력합니다. 데이터 소스로 여러 데이터베이스를 선택할 수 있지만, 본 실습에서는 Oracle Database 를 이용할 것 입니다.

1. Choose AMI   2. Choose Instance Type   3. Configure Instance   4. Add Storage   5. Add Tags   6. Configure Security Group

### Step 6: Configure Security Group

A security group is a set of firewall rules that control the traffic for your instance. On this page, you can add rules to allow specific traffic to reach your instance. For example, if you want to set up a web server and allow Internet traffic to reach your instance, add rules that allow unrestricted access to the HTTP and HTTPS ports. You can create a new security group or select from an existing one below. [Learn more](#) about Amazon EC2 security groups.

Assign a security group:  Create a new security group  Select an existing security group

Security Group ID	Name	Description	Actions
sg-0360a160be52b04f7	default	default VPC security group	<a href="#">Copy to new</a>

Inbound rules for sg-0360a160be52b04f7 (Selected security groups: sg-0360a160be52b04f7)

Type	Protocol	Port Range	Source	Description
All traffic	All	All	sg-0360a160be52b04f7 (d)	

Cancel Previous Review and Launch

“Configure Security Group”에서 VPC 의 Default Security Group 을 선택합니다.

1. Choose AMI    2. Choose Instance Type    3. Configure Instance    4. Add Storage    5. Add Tags    6. Configure Security Group

### Step 7: Review Instance Launch

Please review your instance launch details. You can go back to edit changes for each section. Click **Launch** to assign a key pair to your instance and complete the launch process.

**AMI Details**

oracledb12c\_dms - ami-0a44c21c  
oracledb12c\_dms  
Root Device Type: ebs Virtualization type: hvm

**Instance Type**

Instance Type	ECUs	vCPUs	Memory (GiB)	Instance Storage (GB)	EBS-Optimized Available	Network Performance
t3.large	Variable	2	8	EBS only	Yes	Up to 5 Gigabit

**Security Groups**

Security Group ID	Name	Description
sg-0360a160be52b04f7	default	default VPC security group

**Launch**

EC2 Instance 의 설정 정보를 확인하고, Launch 를 클릭합니다.

1. Choose AMI    2. Choose Instance Type    3. Configure Instance    4. Add Storage    5. Add Tags    6. Configure Security Group

### Step 7: Review Instance Launch

Select an existing key pair or create a new key pair

A key pair consists of a public key that AWS stores, and a private key file that you store. Together, they allow you to connect to your instance securely. For Windows AMIs, the private key file is required to obtain the password used to log into your instance. For Linux AMIs, the private key file allows you to securely SSH into your instance.

Note: The selected key pair will be added to the set of keys authorized for this instance. Learn more about [removing existing key pairs from a public AMI](#).

Proceed without a key pair  
 I acknowledge that I will not be able to connect to this instance unless I already know the password built into this AMI.

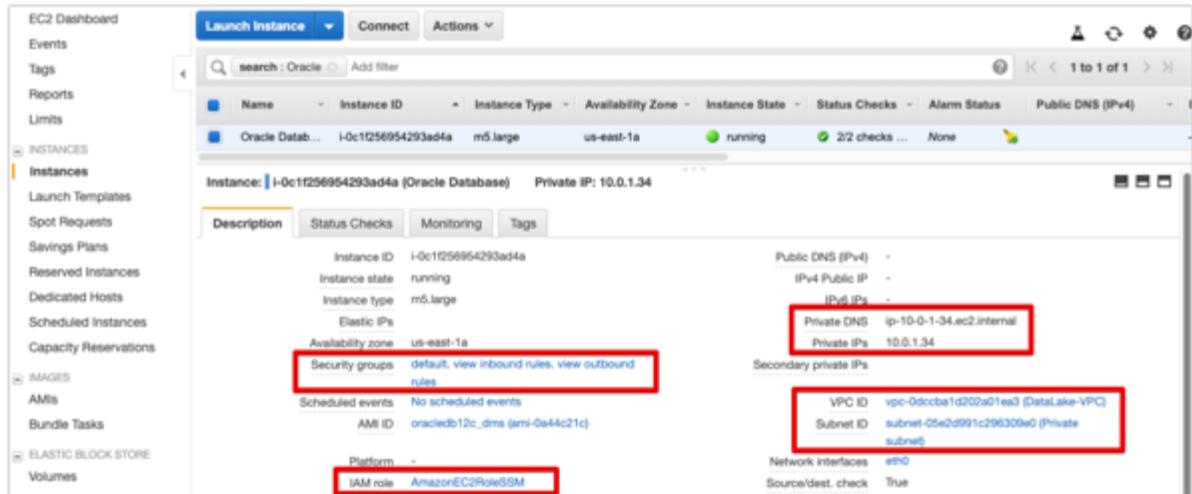
**Launch Instances**

Security Group ID	Name	Description
sg-0360a160be52b04f7	default	default VPC security group

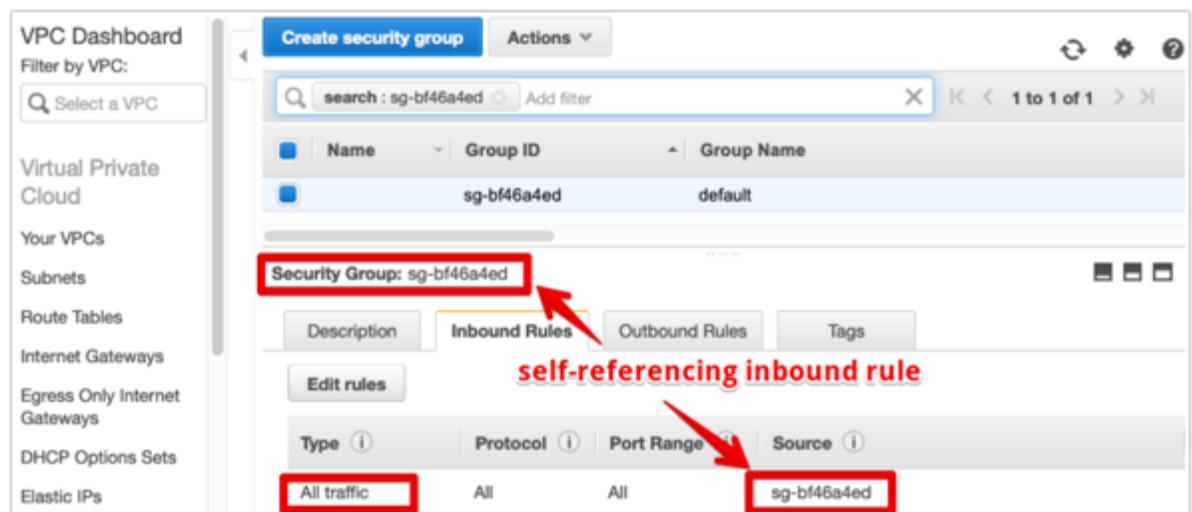
**Launch**

Key pair 항목에는 “Proceed without a key pair”를 선택하고, “I acknowledge that…”를 체크하고 “Launch Instance”를 클릭합니다.

이번 실습에서는 Key Pair 를 통해 EC2 에 접속하지 않고, System Manager 를 통해 접속하므로 별도로 Key Pair 를 만들지 않습니다.



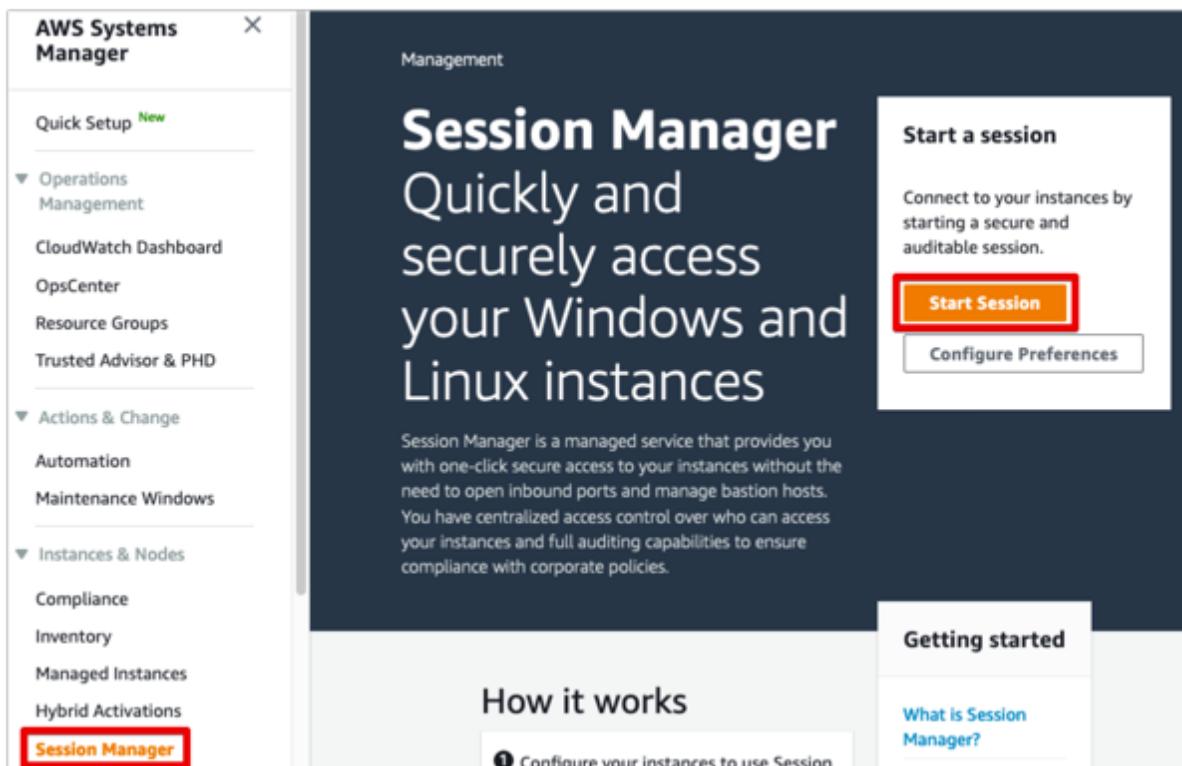
생성된 EC2 Instance 의 Status 가 정상인지 확인하고, Security group, Private IP, Subnet ID, IAM role 이 제대로 부여되었는지 확인합니다.



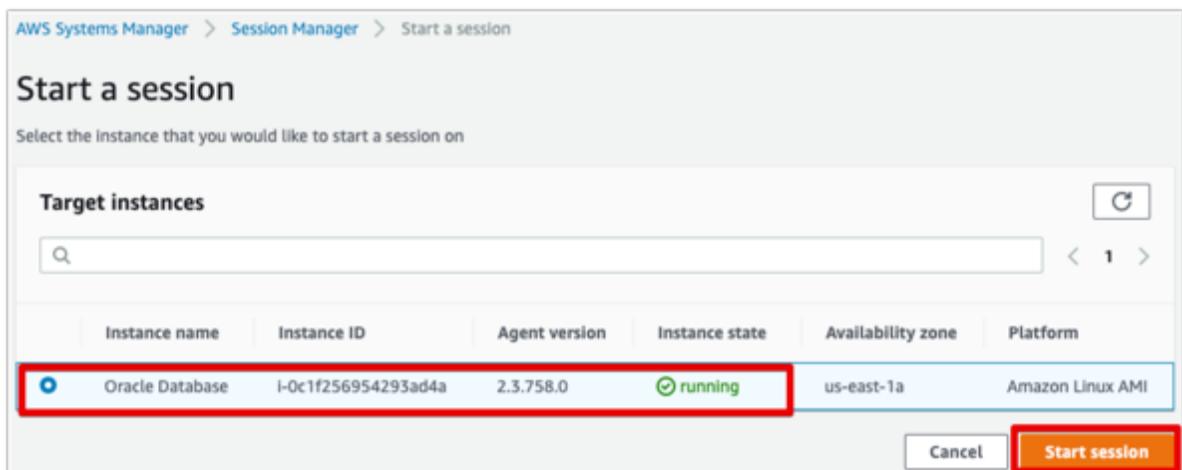
EC2 에 할당된 Security Group 의 Inbound Rule 에는 자기자신을 참조하는 self-referencing inbound rule 을 명시되어 있어야 합니다.

## Task 1.4 : Database Table 구성

1. AWS Systems Manager 의 Session Manager 로 접속합니다.



“Start Session”을 클릭합니다.



EC2 Instance에 “AmazonEC2RoleSSM”을 할당했으므로, Target instances에 표시됩니다.  
Instance를 선택하고, “Start session”을 클릭합니다.

## 2. Terminal에 접속합니다.

```
Session ID: root-07a936394afabb481           Instance ID: i-0c1f256954293ad4a   Terminate
sh-4.2$ sudo su - oracle
Last login: Thu Nov  7 14:11:55 UTC 2019 on pts/0
[oracle@ip-10-0-1-34 ~]$
```

Terminal에서 sudo su - oracle 명령을 실행하여 oracle User로 접속합니다.

## 3. demo\_schema.sql을 다운로드한 후, 실행하여 DEMO Schema를 생성합니다.

```
sudo su - oracle
curl 'https://for-dist-materials.s3.ap-northeast-2.amazonaws.com/oracle/demo_schema.sql' > demo_schema.sql
sqlplus / as sysdba
@demo_schema.sql
```

```
sh-4.2$ sudo su - oracle
Last login: Thu Nov  7 14:18:01 UTC 2019 on pts/0
[oracle@ip-10-0-1-34 ~]$ curl 'https://for-dist-materials.s3.ap-northeast-2.amazonaws.com/oracle/demo_schema.sql' > demo_schema.sql
% Total    % Received % Xferd  Average Speed   Time     Time  Current
          Dload  Upload Total Spent   Left Speed
100 50391  100 50391    0     0  45479      0  0:00:01  0:00:01 --::-- 45479
[oracle@ip-10-0-1-34 ~]$ sqlplus / as sysdba

SQL*Plus: Release 12.1.0.2.0 Production on Thu Nov 7 14:18:59 2019

Copyright (c) 1982, 2014, Oracle. All rights reserved.

Connected to:
Oracle Database 12c Enterprise Edition Release 12.1.0.2.0 - 64bit Production
With the Partitioning, OLAP, Advanced Analytics and Real Application Testing options

SQL> @demo_schema.sql

Session altered.

drop user demo cascade
*
ERROR at line 1:
ORA-01918: user 'DEMO' does not exist

User created.
```

DEMO User 로 접속하여, 아래와 같이 7 개의 Table 이 생성되었는지 확인합니다.

```
SQL>
SQL> conn demo/demo
Connected.
SQL> select table_name from user_tables;

TABLE_NAME
-----
JOB_HISTORY
EMPLOYEES
JOBS
DEPARTMENTS
LOCATIONS
COUNTRIES
REGIONS

7 rows selected.
```

DEMO User 의 Password 는 “demo”입니다.

### Task 1.5 : Data Lake 를 위한 IAM Role 및 User 생성

#### 1. Workflow 를 위한 IAM Role 생성

AWS Lake Formation 에서는 Workflow 를 이용하여 Data Import 를 Scheduling 합니다. 이를 위한 IAM Role 을 생성합니다.

IAM Console(<https://console.aws.amazon.com/iam>)에서 Role->Create role 을 선택합니다.

Create role

Select type of trusted entity

<b>AWS service</b> EC2, Lambda and others	Another AWS account Belonging to you or 3rd party	Web identity Cognito or any OpenID provider	SAML 2.0 federation Your corporate directory
--	--	--	---

Allows AWS services to perform actions on your behalf. [Learn more](#)

Choose the service that will use this role

<b>EC2</b>	Allows EC2 instances to call AWS services on your behalf.			
<b>Lambda</b>	Allows Lambda functions to call AWS services on your behalf.			
API Gateway	CodeBuild	EKS	Kinesis	S3
AWS Backup	CodeDeploy	EMR	Lambda	SMS
AWS Chatbot	CodeStar Notifications	ElastiCache	Lex	SNS
AWS Support	Comprehend	Elastic Beanstalk	License Manager	SWF
Amplify	Config	Elastic Container Service	Machine Learning	SageMaker
AppStream 2.0	Connect	Elastic Transcoder	Macie	Security Hub
AppSync	DMS	Elastic Load Balancing	MediaConvert	Service Catalog
Application Auto Scaling	Data Lifecycle Manager	Forecast	Migration Hub	Step Functions
Application Discovery Service	Data Pipeline	Global Accelerator	OpsWorks	Storage Gateway
Batch	DataSync	<b>Glue</b>	Personalize	Textract
Chime	DeepLens	Greengrass	QLDB	Transfer
CloudFormation	Directory Service	GuardDuty	RAM	Trusted Advisor
CloudHSM	DynamoDB	Inspector	RDS	VPC
CloudTrail	EC2	IoT	Redshift	WorkLink
CloudWatch Application	EC2 - Fleet	IoT Things Graph	Rekognition	WorkMail

\* Required

[Cancel](#) [Next: Permissions](#)

AWS service -> Glue -> Next: Permissions 를 선택합니다.

검색 필드에 “AWSGlueServiceRole”을 입력하고 검색한 후, “AWSGlueServiceRole” Policy 를 선택(체크)하고 Next: Tags 버튼을 누릅니다.

Create role

Attach permissions policies

Choose one or more policies to attach to your new role.

[Create policy](#)

Filter policies		Q AWSGlueServiceRole	Showing 1 result
	Policy name		Used as
<input checked="" type="checkbox"/>	AWSGlueServiceRole		Permissions policy (1)

필요하다면, Tag 에 원하는 값을 입력하고, Next: Review 버튼을 누릅니다.

Create role

Review

Provide the required information below and review this role before you create it.

Role name*	LakeFormationWorkflowRole
Use alphanumeric and '+-, @-_ ' characters. Maximum 64 characters.	
Role description	Allows Glue to call AWS services on your behalf.
Maximum 1000 characters. Use alphanumeric and '+-, @-_ ' characters.	
Trusted entities	AWS service: glue.amazonaws.com
Policies	AWSGlueServiceRole <a href="#">Edit</a>
Permissions boundary	Permissions boundary is not set

Role name에 “LakeFormationWorkflowRole”을 입력하고 “Create role”버튼을 누릅니다.

Role 페이지로 돌아와서, “LakeFormationWorkflowRole”을 검색한 후, 이름을 클릭합니다.

Roles > LakeFormationWorkflowRole

## Summary

Delete role

Role ARN	arn:aws:iam:::role/LakeFormationWorkflowRole <a href="#">Edit</a>
Role description	Allows Glue to call AWS services on your behalf.   <a href="#">Edit</a>
Instance Profile ARNs	<a href="#">Edit</a>
Path	/
Creation time	2019-11-24 12:00 UTC+0900
Last activity	Not accessed in the tracking period
Maximum CLI/API session duration	1 hour <a href="#">Edit</a>

Permissions Trust relationships Tags Access Advisor Revoke sessions

▼ Permissions policies (1 policy applied)

Attach policies [Add inline policy](#)

Policy name	Policy type
AWSGlueServiceRole	AWS managed policy

Summary 페이지에서 Permissions tab -> Add inline policy 를 클릭하고, 아래 Policy 를 추가합니다.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "Lakeformation",
      "Effect": "Allow",
      "Action": [
        "lakeformation:GetDataAccess",
        "lakeformation:GrantPermissions"
      ],
      "Resource": "*"
    }
  ]
}
```

Create policy

A policy defines the AWS permissions that you can assign to a user, group, or role. You can create and edit a policy in the visual editor and using JSON. [Learn more](#)

[Visual editor](#) [JSON](#) [Import managed policy](#)

```

1 {
2   "Version": "2012-10-17",
3   "Statement": [
4     {
5       "Sid": "VisualEditor0",
6       "Effect": "Allow",
7       "Action": [
8         "lakeformation:GrantPermissions",
9         "lakeformation:GetDataAccess"
10      ],
11      "Resource": "*"
12    }
13  ]
14 }
```

Create policy

Review policy

Before you create this policy, provide the required information and review this policy.

Name*	DatalakeDataAccess						
Maximum 128 characters. Use alphanumeric and '+,-,@,-' characters.							
Summary	<table border="1"> <thead> <tr> <th>Service</th> <th>Access level</th> <th>Resource</th> </tr> </thead> <tbody> <tr> <td>Lake Formation</td> <td>Limited: Write, Permissions management</td> <td>All resources</td> </tr> </tbody> </table>	Service	Access level	Resource	Lake Formation	Limited: Write, Permissions management	All resources
Service	Access level	Resource					
Lake Formation	Limited: Write, Permissions management	All resources					
* Required	<a href="#">Cancel</a> <a href="#">Previous</a> <a href="#">Create policy</a>						

Policy Name에 “DatalakeDataAccess”를 입력하고 “Create policy”버튼을 클릭합니다.

"LakeFormationWorkflowRole" Role 에 PassRole 권한을 부여하기 위해 아래와 같은 Policy 를 추가합니다.

The screenshot shows the 'Permissions' tab of an IAM role configuration. A red box highlights the 'Add inline policy' button in the top right corner of the policy list area. The list contains two items: 'AWSGlueServiceRole' (AWS managed policy) and 'DatalakeDataAccess' (Inline policy).

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "PassRolePermissions",
            "Effect": "Allow",
            "Action": ["iam:PassRole"],
            "Resource": [
                "arn:aws:iam::account-id:role/LakeFormationWorkflowRole"
            ]
        }
    ]
}
```

\*account-id 는 자신의 AWS Account number 로 변경합니다.

The screenshot shows the 'Create policy' screen with the 'JSON' tab selected. A red box highlights the 'arn:aws:iam::' part of the Resource ARN in the JSON code. The code is identical to the one shown above.

```

1 - {
2   "Version": "2012-10-17",
3   "Statement": [
4     {
5       "Sid": "PassRolePermissions",
6       "Effect": "Allow",
7       "Action": ["iam:PassRole"],
8       "Resource": [
9         "arn:aws:iam::             :role/LakeFormationWorkflowRole"
10      ]
11    }
12  ]
13 }
```

**Review policy**

Before you create this policy, provide the required information and review this policy.

Name*	DatalakePassRole															
Maximum 128 characters. Use alphanumeric and '+,-,@,_' characters.																
<b>Summary</b> <table border="1"> <thead> <tr> <th colspan="3">Service</th> <th>Access level</th> <th>Resource</th> </tr> </thead> <tbody> <tr> <td colspan="3">Allow (1 of 205 services) <a href="#">Show remaining 204</a></td> <td>Limited: Write</td> <td>RoleName   string like   LakeFormationWorkflowRole</td> </tr> <tr> <td colspan="5"><a href="#">IAM</a></td> </tr> </tbody> </table>		Service			Access level	Resource	Allow (1 of 205 services) <a href="#">Show remaining 204</a>			Limited: Write	RoleName   string like   LakeFormationWorkflowRole	<a href="#">IAM</a>				
Service			Access level	Resource												
Allow (1 of 205 services) <a href="#">Show remaining 204</a>			Limited: Write	RoleName   string like   LakeFormationWorkflowRole												
<a href="#">IAM</a>																
<a href="#">* Required</a> <a href="#">Cancel</a> <a href="#">Previous</a> <b>Create policy</b>																

Policy Name에 “DatalakePassRole”을 입력하고, “Create policy”버튼을 클릭합니다.

아래는 “LakeFormationWorkflowRole”에 부여된 Policy 결과입니다.

Permissions	Trust relationships	Tags	Access Advisor	Revoke sessions
<b>▼ Permissions policies (3 policies applied)</b>				
<a href="#">Attach policies</a>				<a href="#">+ Add inline policy</a>
Policy name	Policy type			
<a href="#">▶ AWSGlueServiceRole</a> <a href="#">▶ DatalakeDataAccess</a> <a href="#">▶ DatalakePassRole</a>	AWS managed policy Inline policy Inline policy			<a href="#">X</a> <a href="#">X</a> <a href="#">X</a>

## 2. Data Lake Administrator 를 위한 IAM User 생성

AWS Lake Formation 의 Data Lake 관리자를 위한 IAM User 를 생성합니다. Data Lake Administrator 는 Data Catalog 에 대한 모든 권한을 부여할 수 있는 권한을 가집니다.

IAM Console(<https://console.aws.amazon.com/iam>)에서 Users->Add user 을 클릭합니다.

Add user

Set user details

You can add multiple users at once with the same access type and permissions. [Learn more](#)

User name*	<input type="text" value="datalake_admin"/>
<a href="#">Add another user</a>	
Select AWS access type	
Select how these users will access AWS. Access keys and autogenerated passwords are provided in the last step. <a href="#">Learn more</a>	
Access type*	<input checked="" type="checkbox"/> <b>Programmatic access</b> Enables an <b>access key ID</b> and <b>secret access key</b> for the AWS API, CLI, SDK, and other development tools.
	<input checked="" type="checkbox"/> <b>AWS Management Console access</b> Enables a password that allows users to sign-in to the AWS Management Console.
Console password*	<input type="radio"/> Autogenerated password <input checked="" type="radio"/> Custom password  <password>.....</password> <input type="checkbox"/> Show password
Require password reset	<input type="checkbox"/> User must create a new password at next sign-in Users automatically get the <a href="#">IAMUserChangePassword</a> policy to allow them to change their own password.
Required <a href="#">Cancel</a> <a href="#">Next: Permissions</a>	

Add user 페이지에서 아래와 같이 선택합니다.

User name	datalake_admin
Access type	AWS Management Console access
Console password	Custom password
Require password reset	Uncheck

AWS managed policy에서 “AWSGlueConsoleFullAccess”와 “AmazonAthenaFullAccess” Policy를 선택합니다.

Set permissions

Add user to group	Copy permissions from existing user	Attach existing policies directly	
<a href="#">Create policy</a>			
<a href="#">Filter policies</a> <input type="text" value="AmazonAthenaFullAccess"/>			
	Policy name	Type	Used as
<input checked="" type="checkbox"/>	AmazonAthenaFullAccess	AWS managed	None

아래와 같은 inline policy를 추가합니다.

```
{  
    "Version": "2012-10-17",  
    "Statement": [  
        {  
            "Effect": "Allow",  
            "Action": [  
                "lakeformation:*",  
                "cloudtrail:DescribeTrails",  
                "cloudtrail:LookupEvents"  
            ],  
            "Resource": "*"  
        },  
        {  
            "Effect": "Deny",  
            "Action": [  
                "lakeformation:PutDataLakeSettings"  
            ],  
            "Resource": "*"  
        },  
        {  
            "Effect": "Allow",  
            "Action": "iam:CreateServiceLinkedRole",  
            "Resource": "*",  
            "Condition": {  
                "StringEquals": {  
                    "iam:AWSServiceName": "lakeformation.amazonaws.com"  
                }  
            }  
        },  
        {  
            "Effect": "Allow",  
            "Action": [  
                "iam:PutRolePolicy"  
            ],  
            "Resource": "arn:aws:iam::account-id:role/aws-service-  
role/lakeformation.amazonaws.com/AWSServiceRoleForLakeFormationDataAcces  
s"  
        }  
    ]  
}
```

\* account-id 는 자신의 AWS Account number 로 교체합니다.

Users > datalake\_admin

## Summary

[Delete user](#) [?](#)

User ARN	arn:aws:iam::[REDACTED]:user/datalake_admin	<a href="#">Edit</a>
Path	/	
Creation time	2019-11-24 13:11 UTC+0900	

Permissions Groups Tags Security credentials Access Advisor

▼ Permissions policies (2 policies applied)

[Add permissions](#) [+ Add inline policy](#)

Policy name	Policy type
AmazonAthenaFullAccess	AWS managed policy
AWSGlueConsoleFullAccess	AWS managed policy

Attached directly

1 Create policy 1 2

A policy defines the AWS permissions that you can assign to a user, group, or role. You can create and edit a policy in the visual editor and using JSON. [Learn more](#)

[Visual editor](#) [JSON](#) [Import managed policy](#)

```

1 {
2   "Version": "2012-10-17",
3   "Statement": [
4     {
5       "Effect": "Allow",
6       "Action": [
7         "lakeformation:*",
8         "cloudtrail:DescribeTrails",
9         "cloudtrail:LookupEvents"
10      ],
11      "Resource": "*"
12    },
13    {
14      "Effect": "Deny",
15      "Action": [
16        "lakeformation:PutDataLakeSettings"
17      ],
18    }
19  ]
20 }
```

Create policy

Review policy

Before you create this policy, provide the required information and review this policy.

Name*	DatalakeAdminBasic																								
Maximum 128 characters. Use alphanumeric and '+,-,@-_ ' characters.																									
<b>Summary</b> <table border="1"> <thead> <tr> <th colspan="3">Filter</th> </tr> <tr> <th>Service</th> <th>Access level</th> <th>Resource</th> </tr> </thead> <tbody> <tr> <td colspan="3"><b>Explicit deny (1 of 205 services)</b></td> </tr> <tr> <td>Lake Formation</td> <td>Limited: Permissions management</td> <td>All resources</td> </tr> <tr> <td colspan="3"><b>Allow (3 of 205 services) Show remaining 202</b></td> </tr> <tr> <td>CloudTrail</td> <td>Limited: List</td> <td>All resources</td> </tr> <tr> <td>IAM</td> <td>Limited: Write, Permissions management</td> <td>Multiple</td> </tr> <tr> <td>Lake Formation</td> <td>Full: List, Read, Write Limited: Permissions management</td> <td>All resources</td> </tr> </tbody> </table>		Filter			Service	Access level	Resource	<b>Explicit deny (1 of 205 services)</b>			Lake Formation	Limited: Permissions management	All resources	<b>Allow (3 of 205 services) Show remaining 202</b>			CloudTrail	Limited: List	All resources	IAM	Limited: Write, Permissions management	Multiple	Lake Formation	Full: List, Read, Write Limited: Permissions management	All resources
Filter																									
Service	Access level	Resource																							
<b>Explicit deny (1 of 205 services)</b>																									
Lake Formation	Limited: Permissions management	All resources																							
<b>Allow (3 of 205 services) Show remaining 202</b>																									
CloudTrail	Limited: List	All resources																							
IAM	Limited: Write, Permissions management	Multiple																							
Lake Formation	Full: List, Read, Write Limited: Permissions management	All resources																							

Policy Name에는 “DatalakeAdminBasic”을 입력합니다.

아래는 datalake\_admin User 의 Summary 화면입니다.

Users > datalake\_admin

**Summary**

User ARN: arn:aws:iam::[REDACTED]:user/datalake\_admin

Path: /

Creation time: 2019-11-24 13:11 UTC+0900

Permissions	Groups	Tags	Security credentials	Access Advisor
-------------	--------	------	----------------------	----------------

▼ Permissions policies (3 policies applied)

Add permissions      Add inline policy

Policy name	Policy type
AmazonAthenaFullAccess	AWS managed policy
AWSGlueConsoleFullAccess	AWS managed policy
DatalakeAdminBasic	Inline policy

Data Lake Administrator 가 “LakeFormationWorkflowRole”을 전달할 수 있도록 아래와 같이 inline policy 를 추가로 부여합니다.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "PassRolePermissions",
            "Effect": "Allow",
            "Action": ["iam:PassRole"],
            "Resource": [
                "arn:aws:iam::account-id:role/LakeFormationWorkflowRole"
            ]
        }
    ]
}
```

\*account-id 는 자신의 AWS Account number 로 변경합니다.

Create policy

A policy defines the AWS permissions that you can assign to a user, group, or role. You can create and edit a policy in the visual editor and using JSON. [Learn more](#)

[Visual editor](#) [JSON](#) [Import managed policy](#)

```

1  {
2      "Version": "2012-10-17",
3      "Statement": [
4          {
5              "Sid": "PassRolePermissions",
6              "Effect": "Allow",
7              "Action": ["iam:PassRole"],
8              "Resource": [
9                  "arn:aws:iam:::role/LakeFormationWorkflowRole"
10             ]
11         }
12     ]
13 }
```

Review policy

Before you create this policy, provide the required information and review this policy.

Name*	DatalakePassRole									
Maximum 128 characters. Use alphanumeric and '+,-,_,@-' characters.										
Summary	<table border="1"> <thead> <tr> <th>Service</th> <th>Access level</th> <th>Resource</th> </tr> </thead> <tbody> <tr> <td>Allow (1 of 205 services) <a href="#">Show remaining 204</a></td> <td>Limited: Write</td> <td>RoleName   string like   LakeFormationWorkflowRole</td> </tr> <tr> <td>IAM</td> <td></td> <td></td> </tr> </tbody> </table>	Service	Access level	Resource	Allow (1 of 205 services) <a href="#">Show remaining 204</a>	Limited: Write	RoleName   string like   LakeFormationWorkflowRole	IAM		
Service	Access level	Resource								
Allow (1 of 205 services) <a href="#">Show remaining 204</a>	Limited: Write	RoleName   string like   LakeFormationWorkflowRole								
IAM										
<a href="#">* Required</a> <a href="#">Cancel</a> <a href="#">Previous</a> <a href="#">Create policy</a>										

Policy Name 에 “DatalakePassRole”을 입력하고, “Create policy”버튼을 클릭합니다.

Users > datalake\_admin

## Summary

User ARN: arn:aws:iam:::user/datalake\_admin

Path: /

Creation time: 2019-11-24 13:11 UTC+0900

Permissions Groups Tags Security credentials Access Advisor

▼ Permissions policies (4 policies applied)

Add permissions Add inline policy

Policy name	Policy type
AmazonAthenaFullAccess	AWS managed policy
AWSGlueConsoleFullAcc...	AWS managed policy
DatalakeAdminBasic	Inline policy
DatalakePassRole	Inline policy

### 3. Data Analyst 를 위한 IAM User 생성

Data Analyst 를 위한 IAM User 를 생성합니다. 이 User 는 Data Lake 에 접속하여 조회만 가능한 권한을 부여받습니다.

IAM Console(<https://console.aws.amazon.com/iam>)에서 Users->Add user 을 클릭합니다.

Add user

Set user details

You can add multiple users at once with the same access type and permissions. [Learn more](#)

User name\*  datalake\_user

[Add another user](#)

Select AWS access type

Select how these users will access AWS. Access keys and autogenerated passwords are provided in the last step. [Learn more](#)

Access type\*  **Programmatic access**  
Enables an access key ID and secret access key for the AWS API, CLI, SDK, and other development tools.

**AWS Management Console access**  
Enables a password that allows users to sign-in to the AWS Management Console.

Console password\*  Autogenerated password  Custom password  
  
.....  
 Show password

Require password reset  User must create a new password at next sign-in  
Users automatically get the [IAMUserChangePassword](#) policy to allow them to change

[Cancel](#) [Next: Permissions](#)

Add user 페이지에서 아래와 같이 선택합니다.

User name	datalake_user
Access type	AWS Management Console access
Console password	Custom password
Require password reset	Uncheck

AWS managed policy에서 “AmazonAthenaFullAccess” Policy를 선택합니다.

Add user

1 2 3 4 5

▼ Set permissions

Add user to group  Copy permissions from existing user  Attach existing policies directly

Create policy

Filter policies  Showing 1 result

Policy name	Type	Used as
<input checked="" type="checkbox"/> AmazonAthenaFullAccess	AWS managed	Permissions policy (1)

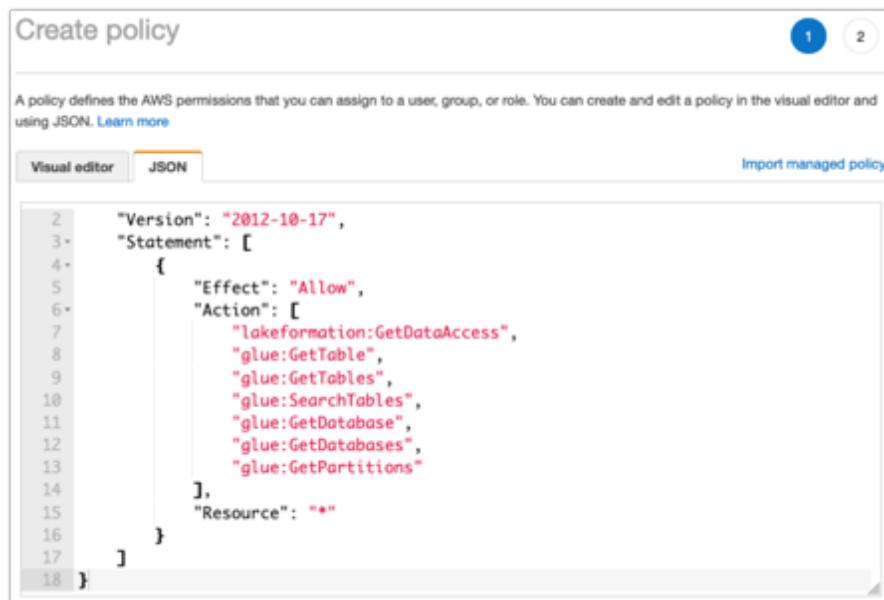
User를 생성한 후에, 아래와 같이 inline policy를 datalake\_user에게 추가로 부여합니다.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "lakeformation:GetDataAccess"
      ]
    }
  ]
}
```

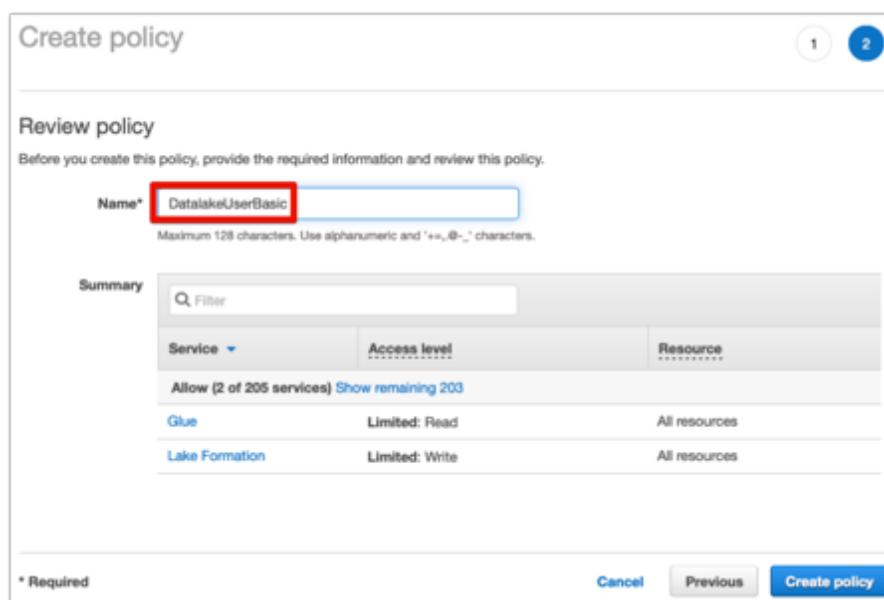
```

    "glue:GetTable",
    "glue:GetTables",
    "glue:SearchTables",
    "glue:GetDatabase",
    "glue:GetDatabases",
    "glue:GetPartitions"
  ],
  "Resource": "*"
}
]
}

```



Policy Name에는 “DatalakeUserBasic”을 입력합니다.



#### 4. IAM Administrator 를 위한 IAM User 생성

AdministratorAccess 권한을 가진 IAM User 를 생성합니다.

IAM Console(<https://console.aws.amazon.com/iam>)에서 Users->Add user 을 클릭합니다.

Add user

Set user details

You can add multiple users at once with the same access type and permissions. [Learn more](#)

User name\* admin

[Add another user](#)

Select AWS access type

Select how these users will access AWS. Access keys and autogenerated passwords are provided in the last step. [Learn more](#)

Access type  **Programmatic access**  
Enables an access key ID and secret access key for the AWS API, CLI, SDK, and other development tools.

**AWS Management Console access**  
Enables a password that allows users to sign-in to the AWS Management Console.

Console password  Autogenerated password  Custom password  
  
\*\*\*\*\*  
 Show password

Require password reset  User must create a new password at next sign-in  
Users automatically get the [IAMUserChangePassword](#) policy to allow them to change

\* Required [Cancel](#) [Next: Permissions](#)

Add user 페이지에서 아래와 같이 선택합니다.

User name	admin
Access type	AWS Management Console access
Console password	Custom password
Require password reset	Uncheck

AWS managed policy에서 “AdministratorAccess” Policy 를 선택합니다.

▼ Set permissions

Add user to group Copy permissions from existing user **Attach existing policies directly**

Create policy

Filter policies ▾  AdministratorAccess

Policy name	Type
<input checked="" type="checkbox"/> AdministratorAccess	Job function

## Part 2. JDBC Data Source로부터 Data Lake 생성

### Task 2.1 : AWS Glue Connection 설정

1. IAM Administrator User로 AWS Glue Console(<https://console.aws.amazon.com/glue/>)에 접속합니다.

**AWS Glue**

AWS Glue is a fully managed ETL (extract, transform, and load) service that makes it simple and cost-effective to categorize your data, clean it, enrich it, and move it reliably between various data stores.

**Build your AWS Glue Data Catalog**

AWS Glue automatically stores metadata in a central data catalog. It can create table definitions for many common data stores, including, S3 buckets, web logs, and AWS databases. AWS Glue recognizes, infers, organizes, and classifies your data.

[Learn more](#)

**Generate and edit transformations**

PySpark transformation scripts are auto generated using source and target metadata. You can store customized versions to transform your data to meet your business needs. AWS Glue provides an environment to modify your jobs.

[Learn more](#)

**Schedule and run your jobs**

AWS Glue runs your ETL jobs in a serverless environment. You don't need to set up the infrastructure, you just use Amazon's infrastructure and pay for the resources you use. You can define triggers to run jobs based on a schedule or event. AWS Glue enables you to monitor your jobs.

[Learn more](#)

AWS Glue Console에서 “Get started”를 클릭합니다.

2. Glue Connection을 설정합니다.(Data catalog -> Connections 를 클릭합니다.)

Name	Type	Date created	Last updated	Updated by
You don't have any connections yet. <a href="#">Add connection</a>				

“Add connection”을 클릭합니다.

**Add connection**

Set up your connection's properties.

For more information, see [Working with Connections](#).

Connection name	datalake-oracle-source
Connection type	JDBC
<input type="checkbox"/> Require SSL connection <small>Fail if unable to connect over SSL.</small>	
Description (optional)	
For Oracle	

**Next**

Connection name	datalake-oracle-source
Connection type	JDBC
Description	For Oracle

Oracle Database 의 Data 를 ETL 할 것이므로, Connection type 은 “JDBC”로 설정합니다.

**Add connection**

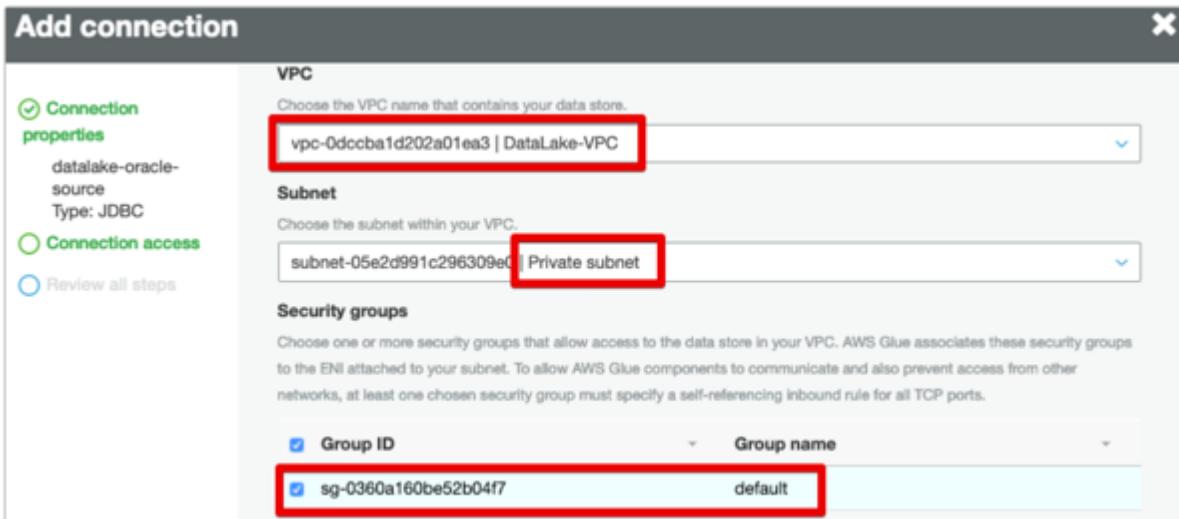
Set up access to your data store.

For more information, see [Working with Connections](#).

JDBC URL	jdbc:oracle:thin://@10.0.1.34:1521/CDB1
Username	demo
Password	****

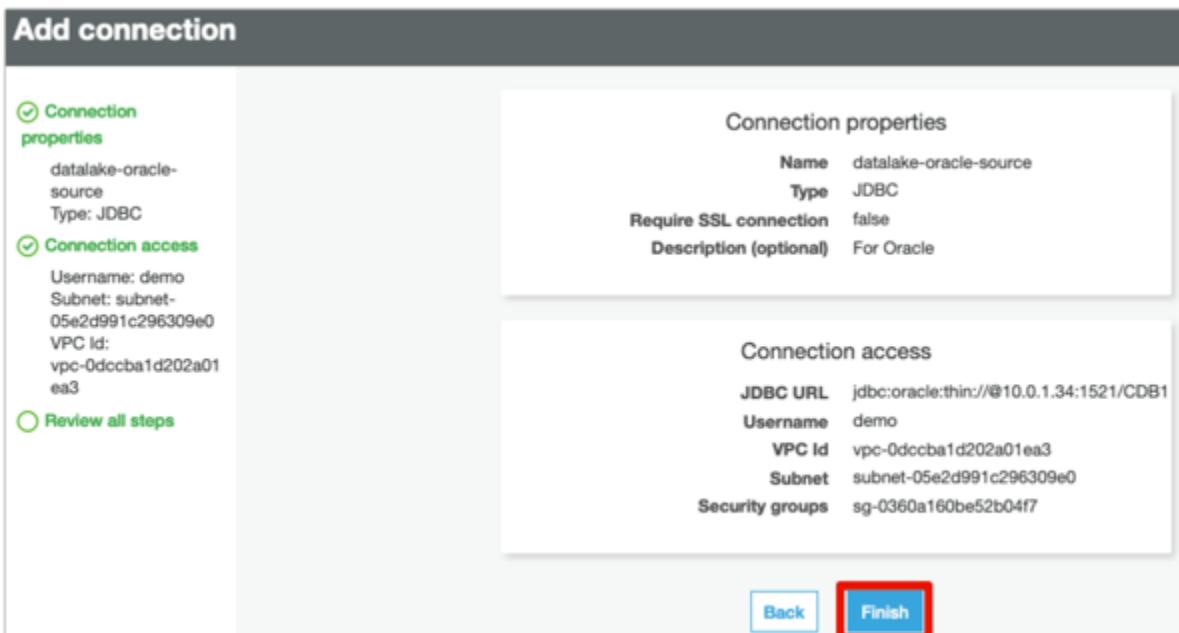
JDBC URL	형식) jdbc:oracle:thin://@<Private IP>:<Port>/<service_name> 예제) jdbc:oracle:thin://@10.0.1.34:1521/CDB1
	현재 생성된 Database Service Name 은 CDB1, Port 는 1521 입니다. <Private IP>는 Part 1에서 생성한 EC2 Instance 의 Private IP 입니다.
Username	demo
Password	demo

Oracle Database 가 아닌 경우에는 JDBC URL 의 형식이 다릅니다. 다른 데이터베이스가 Source 인 경우, 이 [링크](#)를 참조하여 설정합니다.



이전 단계에서 생성한 VPC, Private subnet, Security Group 을 선택합니다.

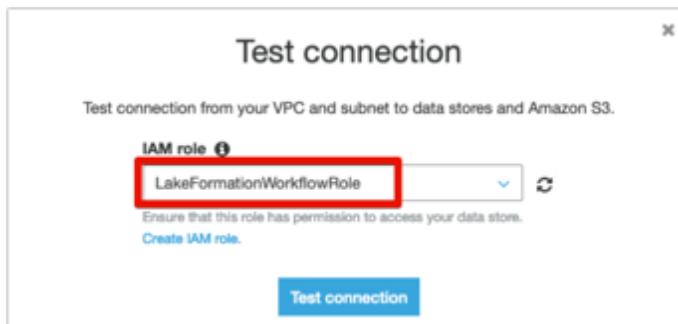
반드시 Private subnet 이어야 하며, self-referencing inbound rule 0| 포함되어 있는 Security Group 을 지정해야 합니다.



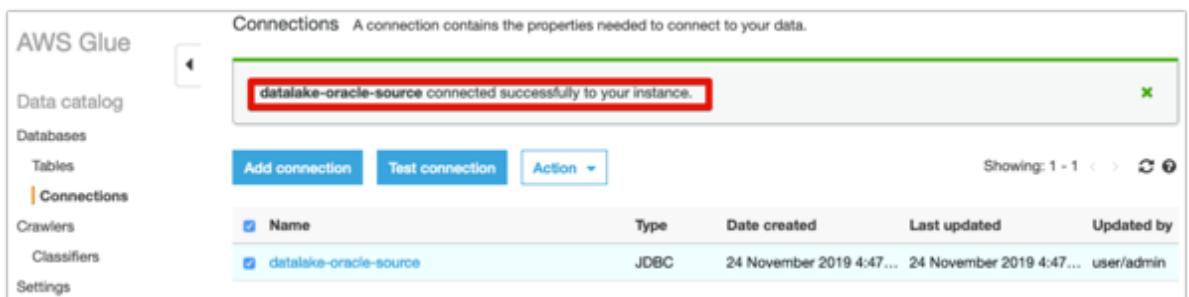
Finish 를 클릭하여 Connection 을 생성합니다.



생성한 Connection 이 정상적으로 작동하는지 확인하기 위해 “Test connection”을 클릭합니다.



Test connection 을 위해 필요한 IAM Role 에는 Part1에서 생성한 “LakeFormationWorkflowRole”을 선택하고, “Test connection”을 클릭합니다.

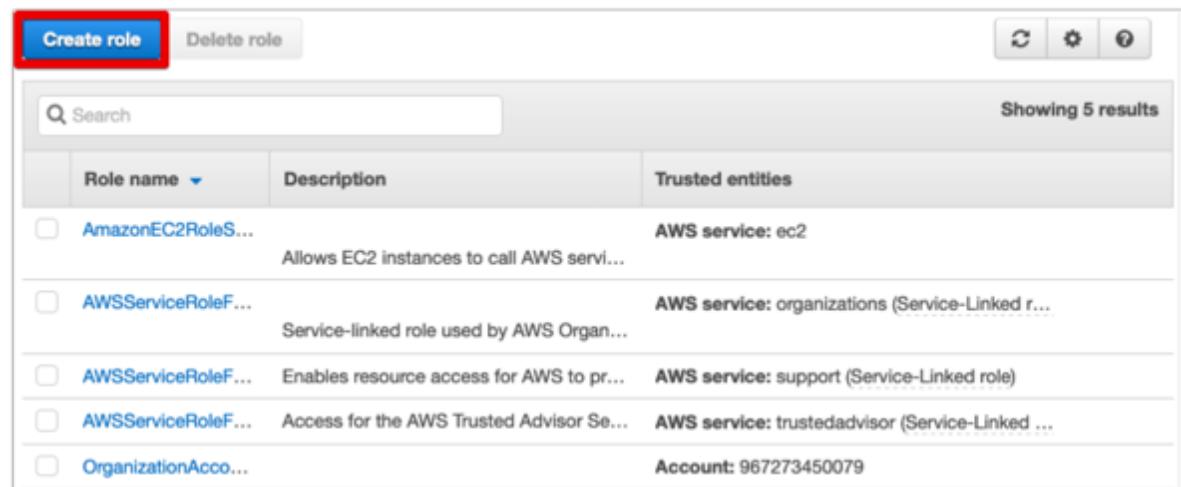


Test connection 이 성공적으로 수행되었는지 확인합니다.

=====삭제=====

Connection 을 Test 하기 위해서는 S3FullAccess 권한이 필요합니다.

처음 Connection 을 생성한 경우에는 아래와 같이 필요한 권한을 가진 IAM Role 을 생성합니다.



“Create role”을 클릭하여 새로운 Role 을 생성합니다.

Create role

Select type of trusted entity

<b>AWS service</b> EC2, Lambda and others	<b>Another AWS account</b> Belonging to you or 3rd party	<b>Web identity</b> Cognito or any OpenID provider	<b>SAML 2.0 federation</b> Your corporate directory
--	---	---	--

Allows AWS services to perform actions on your behalf. [Learn more](#)

Choose the service that will use this role

<b>EC2</b>	Allows EC2 instances to call AWS services on your behalf.			
<b>Lambda</b>	Allows Lambda functions to call AWS services on your behalf.			
API Gateway	CodeBuild	EMR	Lambda	S3
AWS Backup	CodeDeploy	ElasticCache	Lex	SMS
AWS Chatbot	Comprehend	Elastic Beanstalk	License Manager	SNS
AWS Support	Config	Elastic Container Service	Machine Learning	SWF
Amplify	Connect	Elastic Transcoder	Macie	SageMaker
AppStream 2.0	DMS	Elastic Load Balancing	MediaConvert	Security Hub
AppSync	Data Lifecycle Manager	Forecast	Migration Hub	Service Catalog
Application Auto Scaling	Data Pipeline	Global Accelerator	OpsWorks	Step Functions
Application Discovery Service	DataSync	<b>Glue</b>	Personalize	Storage Gateway

\* Required

[Cancel](#) [Next: Permissions](#)

“Glue”를 선택하고, “Next: Permissions”을 클릭합니다.

Create role

Attach permissions policies

Choose one or more policies to attach to your new role.

[Create policy](#)

Filter policies		Showing 2 results
	Policy name	Used as
<input type="checkbox"/>	AWSGlueServiceNotebookRole	None
<input checked="" type="checkbox"/>	AWSGlueServiceRole	None

Policy에 “AmazonS3FullAccess”와 “AWSGlueServiceRole”을 각각 선택합니다.

Create role

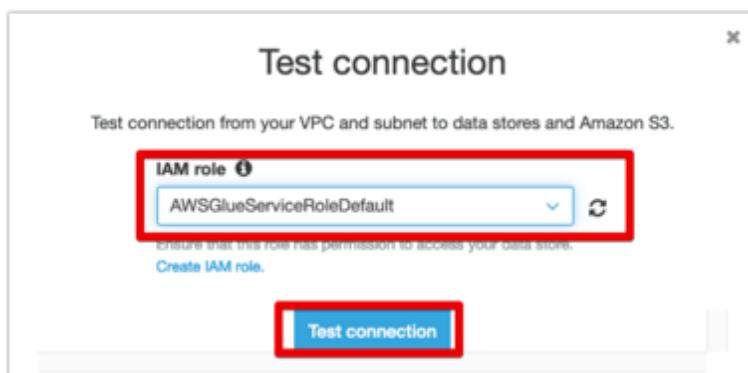
Review

Provide the required information below and review this role before you create it.

Role name*	AWSGlueServiceRoleDefault
Use alphanumeric and '+,-,@-_' characters. Maximum 64 characters.	
Role description	Allows Glue to call AWS services on your behalf.
Maximum 1000 characters. Use alphanumeric and '+,-,@-_' characters.	
Trusted entities	AWS service: glue.amazonaws.com
Policies	AmazonS3FullAccess AWSGlueServiceRole
<a href="#">Cancel</a> <a href="#">Previous</a> <a href="#" style="background-color: blue; color: white; border: 1px solid blue; padding: 2px;">Create role</a>	

Role name에 “AWSGlueServiceRoleDefault”를 입력하고 “Create role”을 클릭합니다.

Policies에 2개의 Policy가 할당되었는지 확인합니다.



Glue Connection Test 화면으로 되돌아와서, 생성한 IAM role을 지정하고, “Test connection”을 클릭합니다.

Connections A connection contains the properties needed to connect to your data.

datalake-oracle-source connected successfully to your instance.				
<a href="#">Add connection</a>	<a href="#">Test connection</a>	<a href="#">Action ▾</a>	Showing: 1 - 1	
<input checked="" type="checkbox"/> Name	Type	Date created	Last updated	Updated by
<a href="#">datalake-oracle-source</a>	JDBC	5 November 2019 ...	5 November 2019 ...	root

정상적으로 Data Source에 연결되는지 확인합니다.

=====삭제=====

## Task 2.2 : Amazon S3 Bucket 생성

Amazon Simple Storage Service(Amazon S3)는 Data Lake 의 root location 입니다.

- IAM Administrator User 로 Amazon S3 Console(<https://console.aws.amazon.com/s3/>)에 접속합니다.

The screenshot shows the Amazon S3 console interface. On the left, there's a sidebar with options like 'Buckets', 'Batch operations', 'Block public access (account settings)', and 'Feature spotlight'. The main area is titled 'S3 buckets' and contains a search bar, a dropdown for 'All access types', and a large red button labeled '+ Create bucket'. Below these, it says '0 Buckets' and '0 Regions'. A message at the bottom reads: 'You do not have any buckets. Here is how to get started with Amazon S3.'

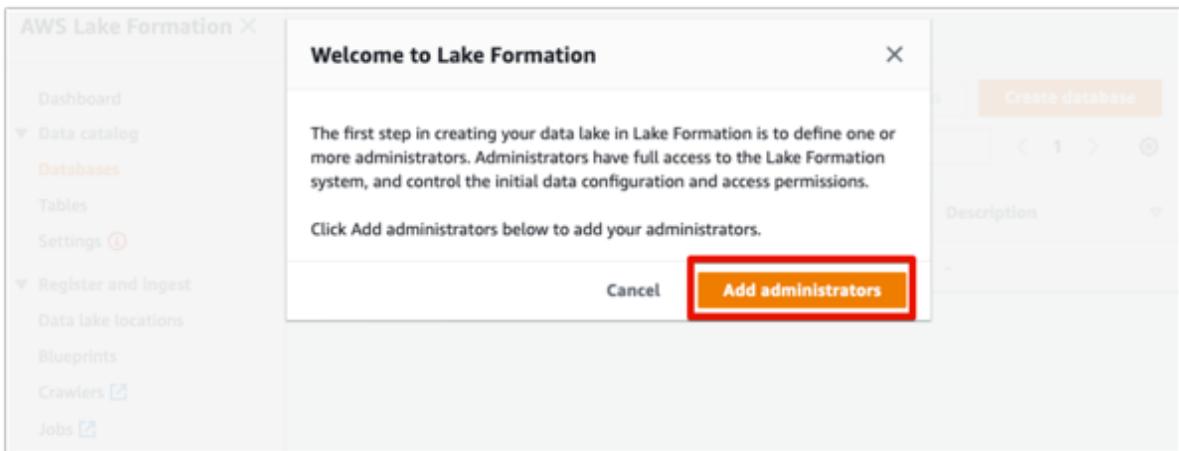
The screenshot shows the 'Create bucket' wizard. Step 1: Name and region. It asks for a 'Bucket name' (with 'ironpe-datalake-tutorials' entered) and a 'Region' (set to 'US East (N. Virginia)'). There's also a section for 'Copy settings from an existing bucket' which is empty. At the bottom, there's a 'Create' button (highlighted with a red box), a 'Cancel' button, and a 'Next' button.

Bucket name	<yourName>-datalake-tutorials 예)ironpe-datalake-tutorials
Region	US East (N. Virginia) Lake Formation 서비스와 동일한 Region 을 선택합니다.

### Task 2.3 : Data Lake Administrator 생성

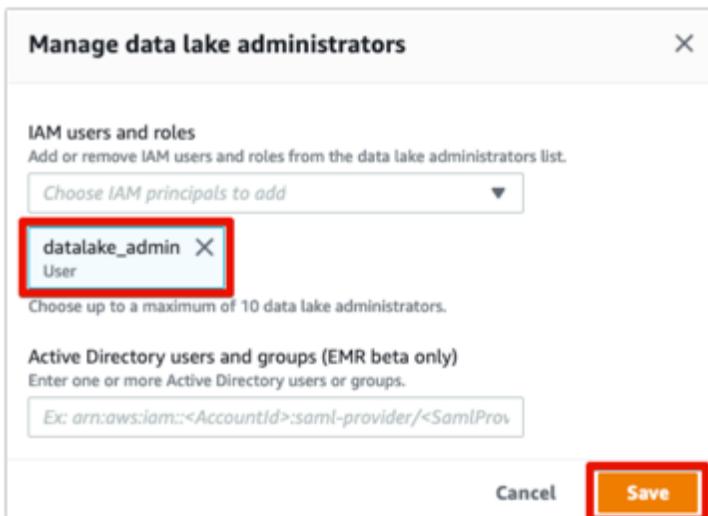
- IAM Administrator User 로 Amazon Lake Formation

Console(<https://console.aws.amazon.com/lakeformation/>)에 접속합니다.



AWS Lake Formation 에 처음 접속할 경우, 위와 같은 페이지가 표시됩니다. “Add administrators”를 클릭합니다.

- Data Lake Administrator 를 선택합니다.



IAM users and roles 에 Part1 에서 생성한 “datalake\_admin” User 를 선택합니다.

Data Lake Administrator 는 Data Catalog 의 모든 메타데이터를 조회할 수 있고, Data Resources 에 대한 권한을 부여하거나 회수할 수 있는 권한을 가집니다.

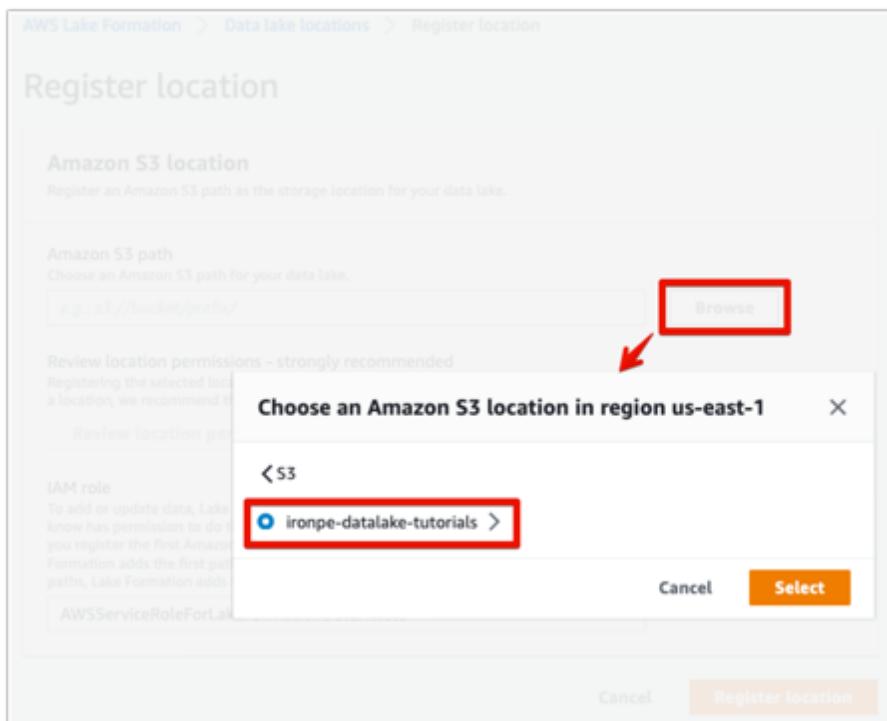
The screenshot shows the AWS Lake Formation console. On the left, the navigation menu is visible with several sections like Dashboard, Data catalog, Databases, Tables, Settings, Register and ingest, and Permissions. Under Permissions, the 'Admins and database creators' section is highlighted with a red box. The main content area shows the 'Data lake administrators (0/1)' section, which lists 'datalake\_admin' as an IAM user. Below it is the 'Database creators' section, which currently has no permissions.

## Task 2.4 : Amazon S3 경로 등록

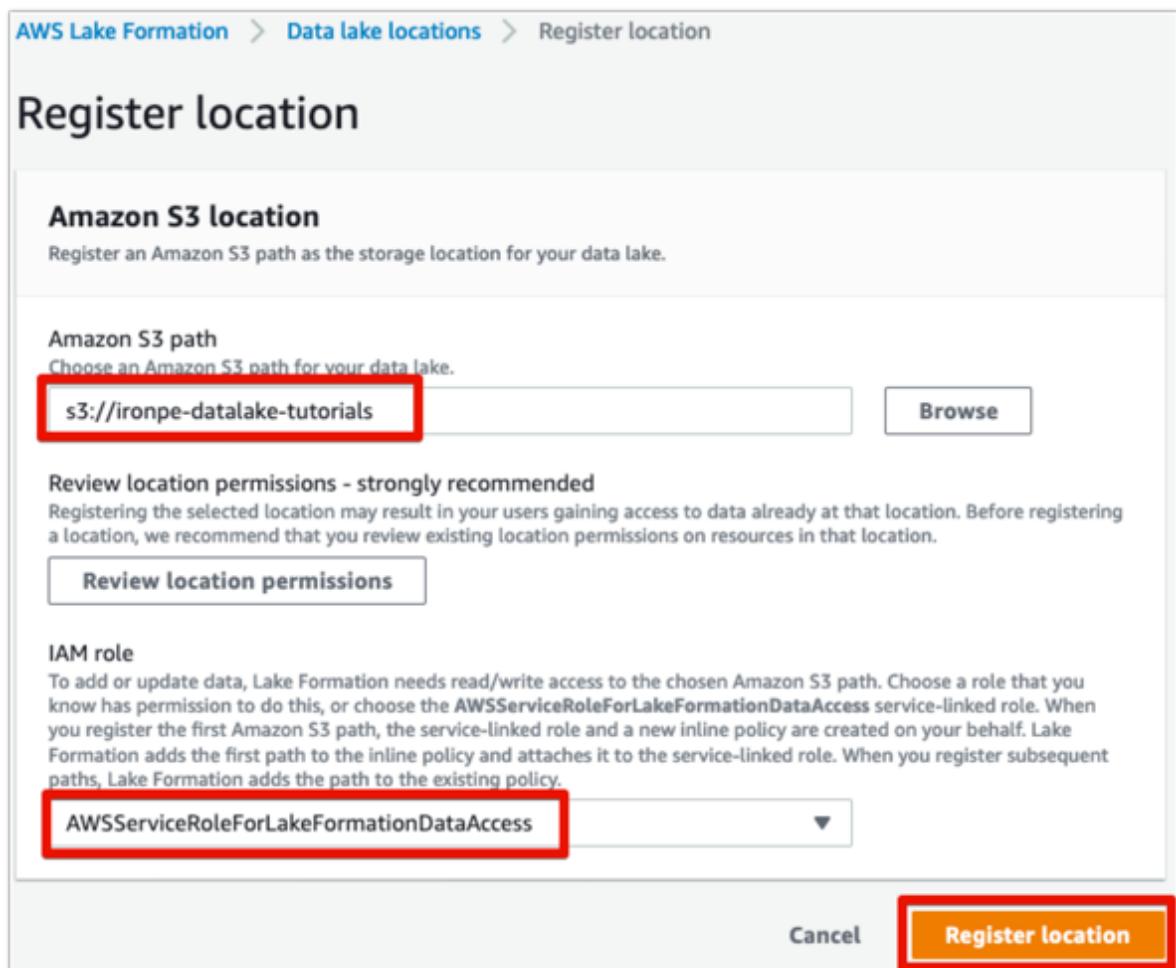
1. Data Lake Administrator User(datalake\_admin)로 Amazon Lake Formation Console(<https://console.aws.amazon.com/lakeformation/>)에 접속합니다.

The screenshot shows the AWS Lake Formation console again. The navigation menu is similar to the previous one, with the 'Data lake locations' section in the 'Register and ingest' category highlighted with a red box. The main content area shows the 'Data lake locations' section, which currently displays 'No Data lake storage'. A prominent orange 'Register location' button is located at the bottom right of this section.

Register and ingest -> Data lake locations -> Register location 을 클릭합니다.



Amazon S3 path에서 Browse를 클릭하고, 이전 단계에서 생성한 S3 Bucket을 선택합니다.



IAM role에는 "AWSServiceRoleForLakeFormationDataAccess"를 선택합니다.

AWS Lake Formation > Data lake locations

Data lake locations (0/1)

Actions Register location

Filter data lake storage

Amazon S3 path IAM role Last modified

<a href="s3://ironpe-datalake-tutorials">s3://ironpe-datalake-tutorials</a>	AWSServiceRoleFo...	Sun, Nov 24, 2019, 8:51 A...
---	---------------------	------------------------------

## Task 2.5 : Data Location Permission 부여

Workflow IAM Role(LakeFormationWorkflowRole)이 Data Location에 대한 권한을 가지도록 설정합니다.

1. Data Lake Administrator User(datalake\_admin)로 Amazon Lake Formation Console(<https://console.aws.amazon.com/lakeformation/>)에 접속합니다.
2. Data Location에 대한 권한을 부여합니다.

AWS Lake Formation > Data locations

Data locations

Choose a storage location for which to review, grant or revoke user permissions.

e.g.: s3://bucket/prefix/

Grant

Principal Principal type Resource Grantable

Enter an Amazon S3 location to view permissions.

Dashboard Data catalog Databases Tables Settings (1) Register and ingest Data lake locations Blueprints Crawlers (2) Jobs (2) Permissions Admins and database creators (1) Data permissions Data locations

Permissions -> Data locations -> Grant 를 클릭합니다.

**Grant permissions**

Add access permissions for specific storage locations.

**IAM users and roles**  
Add one or more IAM users or roles.

▾

LakeFormationWorkflowRole   
Role

**Active Directory users and groups (EMR beta only)**  
Enter one or more Active Directory users or groups.  
Ex: arn:aws:iam::<AccountID>:saml-provider/<SamProv>

**Storage locations**  
Choose one or more data lake locations.

e.g.: s3://bucket/prefix/

s3://ironpe-datalake-tutorials

Grantable

IAM users and roles	LakeFormationWorkflowRole
Storage locations	<yourName>-datalake-tutorials 예)ironpe-datalake-tutorials

**AWS Lake Formation**  Permission granted for: LakeFormationWorkflowRole to s3://ironpe-datalake-tutorials.

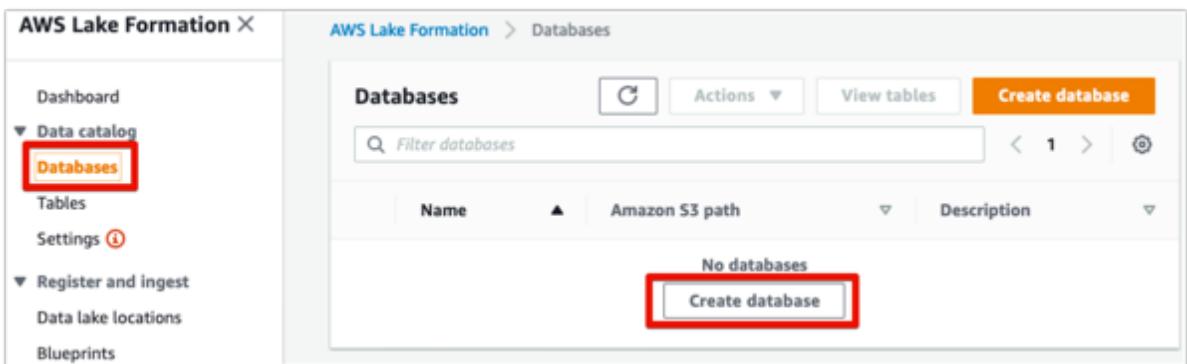
**AWS Lake Formation** > Data locations

**Data locations (1)**  
Choose a storage location for which to review, grant or revoke user permissions.

Principal	Principal type	Resource	Grantable
LakeFormationWorkflowRole	IAM role	s3://ironpe-datalake-tutorials	-

## Task 2.6 : Data Catalog 의 Database 생성

1. Data Lake Administrator User(datalake\_admin)로 Amazon Lake Formation Console(<https://console.aws.amazon.com/lakeformation/>)에 접속합니다.



Data catalog -> Databases -> Create database 를 클릭합니다.

## 2. Database 를 생성합니다.

AWS Lake Formation > Databases > Create database

## Create database

**Database details**  
Create a database in the Data Catalog.

Name  
 (Names may contain letters (A-Z), numbers (0-9), hyphens (-), or underscores (\_), and must be less than 256 characters long.)

Location - optional  
Choose an Amazon S3 path for this database, which eliminates the need to grant data location permissions on catalog table paths that are this location's children  
 Browse

Description - optional

Descriptions can be up to 2048 characters long.

**Default permissions for newly created tables**  
This setting maintains existing AWS Glue data catalog behavior. You can still set individual permissions, which will take effect when you revoke the Super permission from IAMAllowedPrincipals. See [Changing Default Settings for Your Data Lake](#).

Use only IAM access control for new tables in this database

Cancel Create database

Name 에는 “lakeformationTutorial”로 입력합니다.

## Task 2.7 : Data Permissions 부여

Data Catalog 에 메타데이터 테이블을 생성할 수 있는 권한을 부여합니다.

1. Data Lake Administrator User(datalake\_admin)로 Amazon Lake Formation Console(<https://console.aws.amazon.com/lakeformation/>)에 접속합니다.

The screenshot shows the AWS Lake Formation console with the 'Permissions' section selected. On the left sidebar, 'Data permissions' is highlighted with a red box. In the main area, the 'Data permissions (2)' section is displayed. The 'Grant' button at the top right of this section is also highlighted with a red box.

Principal	Principal type	Resource type	Resource	Permissions	Grantable
datalake_admin	IAM user	Database	lakeformation_tutorial	Alter, Create table, Drop	Alter, Create table, Drop
IAMAllowPrincipals	Group	Database	lakeformation_tutorial	Super	-

Permissions -> Data permissions -> Grant 를 클릭합니다.

2. Data Permission 을 부여합니다.

The screenshot shows the 'Grant permissions' dialog box. Under 'IAM users and roles', 'LakeFormationWorkflowRole' is selected and highlighted with a red box. Under 'Database', 'lakeformationTutorial' is selected and highlighted with a red box. In the 'Database permissions' section, the checkboxes for 'Create table', 'Alter', and 'Drop' are checked, while 'Super' is unchecked. At the bottom right, the 'Grant' button is highlighted with a red box.

IAM users and roles	LakeFormationWorkflowRole
Database	lakeformationTutorial
Database permissions	<input checked="" type="checkbox"/> Create table <input checked="" type="checkbox"/> Alter <input checked="" type="checkbox"/> Drop <input type="checkbox"/> Super
Grantable permissions	<input type="checkbox"/> Create table <input type="checkbox"/> Alter <input type="checkbox"/> Drop <input type="checkbox"/> Super

Database	lakeformation_tutorial
Database permissions	Create table, Alter, Drop : Check(선택) Super : Uncheck(해제)

## Task 2.8 : ETL Workflow 생성

AWS Lake Formation Workflow 는 Data Source 으로부터 Data Lake 에 Data 를 Ingest 하기 위해 AWS Glue job, crawler, trigger 생성을 자동화합니다.

AWS Lake Formation 은 Predefined Lake Formation Blueprint 를 제공합니다. 이번 실습에서는 “Database snapshot” Blueprint 를 이용하겠습니다.

1. Data Lake Administrator User(datalake\_admin)로 Amazon Lake Formation Console(<https://console.aws.amazon.com/lakeformation/>)에 접속합니다.

The screenshot shows the AWS Lake Formation console with the 'Blueprints' section selected in the sidebar. The 'Blueprint overview' section contains two main categories: 'Database blueprints' and 'Log file blueprints'. Below each category is a brief description and a 'Use blueprint' button, which is highlighted with a red box. The 'Workflows' section below shows a table with columns for Name, Created on, Last updated, and Last run status. A 'Use blueprint' button is located at the bottom of this section, also highlighted with a red box.

Register and ingest -> Blueprints -> Use blueprint 를 클릭합니다.

2. Blueprint Type 에 Database snapshot 을 선택합니다.

The screenshot shows the 'Use a blueprint' dialog box. Under the 'Blueprint type' heading, there are five options: 'Database snapshot' (selected and highlighted with a red box), 'Incremental database', 'AWS CloudTrail', 'Classic Load Balancer logs', and 'Application Load Balancer logs'. Each option has a brief description below it.

3. Import source 를 설정합니다.

Database connection 은 이전 단계에서 생성한 Glue connection name 을 입력합니다.

Source data path 에는 ETL 대상 테이블을 입력합니다.

**Import source**  
Configure the workflow source.

**Database connection**  
Choose the connection to the data source. [Create a connection in AWS Glue](#)

**Source data path**  
Enter the path from which to ingest data. For JDBC databases with schema support, enter database/schema/table. Substitute the percent (%) wildcard for schema or table.

Database connection	datalake-oracle-source
Source data path	<p>형식 : database/schema/table (Schema 를 지원하는 경우에만)</p> <p>*Oracle JDBC 는 Schema 를 지원하지 않습니다. 따라서 이 실습에서는 아래와 같이 입력합니다.(%는 wildcard 입니다.)</p> <p>예 : CDB1/EMPL%</p>

4. Import Target 을 설정합니다.

여기서는 Target 이 될 Glue Database, S3 위치, S3 에 저장될 데이터의 형식을 입력합니다.

**Import target**  
Configure the target of the workflow.

**Target database**  
Choose a database in the Data Catalog. [Create database](#)

**Target storage location**  
Choose a data lake location or other Amazon S3 path.

**Data format**  
Choose the output data format.

Target Database	lakeformationTutorial(임의의 값)
Target Source Location	<p>s3://&lt;yourName&gt;-datalake-tutorials(Unique name)</p> <p>예 : s3://ironpe-datalake-tutorials</p>

Data format	Parquet
-------------	---------

Import frequency 는 “Run on demand”로 선택합니다.

#### 5. Import Options 을 선택합니다.

Lake Formation 에서는 Glue workflow 를 통해 프로세스를 자동화합니다. 이에 대하 정보를 입력합니다.

The screenshot shows the 'Import options' configuration page. It includes fields for 'Workflow name' (set to 'lakeformation\_oracle'), 'IAM role' (set to 'LakeFormationWorkflowRole'), and 'Table prefix' (set to 'cdb1'). There are also optional fields for 'Maximum capacity' and 'Concurrency'. At the bottom right, there are 'Cancel' and 'Create' buttons, with 'Create' being highlighted.

\*Table prefix 에는 테이블을 구분하기 쉽도록 임의의 값을 입력합니다.

Workflow name	lakeformation_oracle(임의의 값)
IAM Role	LakeFormationWorkflowRole(이전 단계에서 생성한 Role)
Table prefix	cdb1(임의의 값)
Maximum capacity	최대 DPU 값
Concurrency	최대 동시 수행 가능 횟수(기본 값)

#### 6. Workflow 를 실행합니다.

The screenshot shows the 'Workflows (0/1)' list page. It displays a single workflow named 'lakeformation\_oracle'. The 'Actions' dropdown for this workflow has 'Start' highlighted. Other options in the dropdown include 'Delete' and 'View graph'.

Workflow 를 선택하고 Actions 에서 Start 를 선택합니다. Workflow 의 진행상태는 “Last run status”에서 확인할 수 있습니다.

Actions에서 “View graph”를 선택하면, 아래와 같이 AWS Glue Workflow의 상태를 확인할 수 있습니다.

The screenshot shows the AWS Glue Workflows console. On the left, there's a sidebar with various ETL-related options like Data catalog, Databases, Tables, Connections, Crawlers, Classifiers, Settings, and ETL. Under ETL, 'Workflows' is selected, showing a list of workflows. One workflow, 'lakeformation\_oracle', is highlighted with a red box and labeled 'Running'. Below the list is a 'Graph' tab, which is currently active, showing a detailed flowchart of the workflow's execution steps. A legend at the top of the graph area defines symbols for Start, Trigger, Job, Crawler, Incomplete, Error, and Deleting.

Workflow가 완료될 때까지 다소 시간이 소요됩니다. 소요시간은 데이터 양에 따라 달라집니다.

The screenshots show the AWS Lake Formation console. The top part shows the 'Blueprints' section. On the left, the sidebar includes 'Data catalog', 'Tables', 'Settings', 'Register and ingest', 'Blueprints' (which is selected), 'Crawlers', 'Jobs', 'Permissions', 'Admins and database creators', 'Data permissions', and 'Data locations'. The main area shows 'Database blueprints' and 'Log file blueprints' sections, with a 'Use blueprint' button. Below is a 'Workflows (0/1)' section with a table showing a single workflow named 'lakeformation\_tutorial' with a status of 'COMPLETED'. The bottom part shows the 'Tables' section. The sidebar here includes 'Tables' (which is selected) and 'Data locations'. The main area shows a table with three rows. The first row, 'cdb1\_cdb1\_demo\_employees', is highlighted with a red box. The columns include Name, Database, Location, Classification, and Last updated.

Name	Database	Location	Classification	Last updated
<b>cdb1_cdb1_demo_employees</b>	lakeformation_tutorial	s3://ironpe-datalake-tuto...	PARQUET	Sun, Nov 24, 2019
_temp_cdb1_cdb1_demo_employees	lakeformation_tutorial	s3://ironpe-datalake-tuto...	PARQUET	Sun, Nov 24, 2019
_cdb1_cdb1_demo_employees	lakeformation_tutorial	CDB1.DEMO.EMPLOYEES	oracle	Sun, Nov 24, 2019

Workflow가 완료된 후, Data Catalog가 생성된 것을 확인할 수 있습니다.

## Task 2.9 : SELECT Table 권한 부여

1. Data Lake Analyst User(datalake\_user)로 Amazon Lake Formation

Console(<https://console.aws.amazon.com/lakeformation/>)에 접속합니다.

Data catalog -> Tables 를 클릭합니다.

Name	Database	Location	Classification	Last updated
cdb1_cdb1_demo_employees	lakeformation_tutorial	s3://fronge-datalake-tutorials/cdb1...	PARQUET	Sun, Nov 24, 2019, 10:21 AM UTC
_temp_cdb1_cdb1_demo_...	lakeformation_tutorial	s3://fronge-datalake-tutorials/cdb1...	PARQUET	Sun, Nov 24, 2019, 10:20 AM UTC
_cdb1_cdb1_demo_employees	lakeformation_tutorial	CDB1.DEMO.EMPLOYEES	oracle	Sun, Nov 24, 2019, 10:05 AM UTC

2. Data Lake Analyst User(datalake\_user)가 Table에 대한 접근 권한이 있는지 확인합니다.

User: arn:aws:iam:::user:datalake\_user is not authorized to perform: glue:GetTableVersions on resource: arn:aws:glue:us-east-1:catalog

현재는 datalake\_user 는 Table 을 조회할 권한이 없다는 것을 확인할 수 있습니다.

3. Data Lake Administrator User(datalake\_admin)로 Amazon Lake Formation Console에 접속하여 권한을 부여합니다.

Permissions -> Data permissions 을 클릭합니다.

Principal	Principal type	Resource type	Resource	Permissions	Grantable
LakeFormationWorkFlowRole	IAM role	Database	lakeformation_tutorial	Alter, Create table, Drop	-
datalake_admin	IAM user	Database	lakeformation_tutorial	Alter, Create table, Drop	Alter, Create table, Drop
IAMAllowedPrincipals	Group	Database	lakeformation_tutorial	Super	-
LakeFormationWorkFlowRole	IAM role	Table	_cdb1_cdb1_demo_employees	Super, Alter, Delete, Drop, Insert	Super, Alter, Delete, Drop, Insert
LakeFormationWorkFlowRole	IAM role	Column	_cdb1_cdb1_demo_employees.*	Select	Select
LakeFormationWorkFlowRole	IAM role	Table	_temp_cdb1_cdb1_demo_employees	Super, Alter, Delete, Drop, Insert	Super, Alter, Delete, Drop, Insert
LakeFormationWorkFlowRole	IAM role	Column	_temp_cdb1_cdb1_demo_employees.*	Select	Select
LakeFormationWorkFlowRole	IAM role	Table	cdb1_cdb1_demo_employees	Super, Alter, Delete, Drop, Insert	Super, Alter, Delete, Drop, Insert
LakeFormationWorkFlowRole	IAM role	Column	cdb1_cdb1_demo_employees.*	Select	Select
<b>datalake_admin</b>	<b>IAM user</b>	<b>Table</b>	<b>cdb1_cdb1_demo_employees</b>	<b>Alter, Delete, Drop, Insert</b>	<b>Alter, Delete, Drop, Insert</b>

현재는 datalake\_admin이 Table에 대한 권한을 가진 것을 확인할 수 있습니다.

4. Data Lake Analyst User(datalake\_user)에 권한을 부여합니다.

Grant -> Grant permissions 을 클릭하여 Table 조회권한을 부여합니다.

**Grant permissions**

Choose the access permissions to grant. IAM permissions must also allow access.

**IAM users and roles**  
Add one or more IAM users or roles.

Choose IAM principals to add ▾

datalake\_user

User

**Active Directory users and groups (EMR beta only)**  
Enter one or more Active Directory users or groups.

Ex: arn:aws:iam::<AccountID>;saml-provider/<SAMLProviderName>;user/<User>

**Database**  
Add one or more databases.

Choose databases ▾

lakeformationTutorial

**Table - optional**  
Add one or more tables.

Choose tables ▾

cdb1\_cdb1\_demo\_employees

**Column - optional**  
Choose filter type

None

**Table permissions**  
Choose the specific access permissions to grant.

Alter  Insert  Drop  Delete  Select

Super  
This permission is the union of the individual permissions above and supersedes them. See here ↗

**Grantable permissions**  
Choose the permissions that may be granted to others.

Alter  Insert  Drop  Delete  Select

Super  
This permission allows the principal to grant any of the above permissions and supersedes those grantable permissions.

Cancel | **Grant**

권한을 선택한 후, “Grant”버튼을 클릭합니다.

IAM users and roles	datalake_user(Data Lake Analyst User)
Database	lakeformationTutorial(Glue database name)
Table - optional	대상 테이블 선택(여러 테이블 선택 가능)
Column - optional	Include/Exclude 규칙 적용
Table permissions	대상 테이블에 대한 권한 지정

Principal	Principal type	Resource type	Resource	Permissions	Grantable
LakeFormationWorkflow_owRole	IAM role	Table	cdb1_cdb1_demo_em playees	Super, Alter, Delete, Drop, Insert	Super, Alter, Delete, Drop, Insert
LakeFormationWorkflow_owRole	IAM role	Column	lakeformationTutorial .cdb1_cdb1_demo_em playees.*	Select	Select
datalake_admin	IAM user	Table	cdb1_cdb1_demo_em playees	Alter, Delete, Drop, Insert	Alter, Delete, Drop, Insert
datalake_admin	IAM user	Column	lakeformationTutorial .cdb1_cdb1_demo_em playees.*	Select	Select
<b>datalake_user</b>	IAM user	Table	cdb1_cdb1_demo_em playees	Super	-
<b>datalake_user</b>	IAM user	Column	lakeformationTutorial .cdb1_cdb1_demo_em playees.*	Select	-
IAMAllowedPrincipals	Group	Table	cdb1_cdb1_demo_em playees	Super	-

datalake\_user 에 지정된 Table 에 대한 Select 권한이 부여된 것을 확인할 수 있습니다.

이제 다시 datalake\_user 로 접속하여 Table 조회가 가능한지 확인해 보겠습니다.

## Task 2.10 : Amazon Athena 를 Data Lake 조회

1. Data Lake Analyst User(datalake\_user)로 Athena

Console(<https://console.aws.amazon.com/athena/>)에 접속합니다.

Settings 를 클릭하여 Athena Query Result 를 저장할 S3 위치를 지정합니다.

지정하는 S3 Bucket 에 대한 권한이 없는 경우 Query 수행시 오류가 발생합니다.

2. Athena Query Result 위치를 지정합니다.

Query result location : s3://aws-athena-query-results-<사용자명>/

(aws-athena-query-result-뒤에 임의의 사용자명을 입력합니다.)

### 3. Athena로 테이블을 조회합니다.

The screenshot shows the AWS Athena Query Editor interface. On the left, there's a sidebar with 'Data source' set to 'awsdatabatalog', 'Database' set to 'lakelformation\_tutorial', and a list of tables ('Tables (2)') and views ('Views (0)'). A red box highlights the 'Preview table' button next to the first table entry. The main area displays a query result table with columns: job\_id, employee\_id, salary, hire\_date, department\_id, last\_name, email, phone\_number, first\_name, commission\_pct, and male. The results show 10 rows of employee data from the table.

job_id	employee_id	salary	hire_date	department_id	last_name	email	phone_number	first_name	commission_pct	male
AD_PRES	100	24000.00	2003-06-17 00:00:00.000	90	King	SKING	515.123.4567	Steven		
AD_VP	101	17000.00	2005-09-21 00:00:00.000	90	Kochhar	NKOCHHAR	515.123.4568	Neena	100	
AD_VP	102	17000.00	2001-01-13 00:00:00.000	90	De Haan	LDEHAAN	515.123.4569	Lex	100	
IT_PROG	103	9000.00	2006-01-03 00:00:00.000	60	Hunold	AHUNOLD	590.423.4567	Alexander	102	
IT_PROG	104	6000.00	2007-05-21 00:00:00.000	60	Emat	BERNST	590.423.4568	Bruce	103	
IT_PROG	105	4800.00	2005-06-25 00:00:00.000	60	Austin	DAUSTIN	590.423.4569	David	103	
IT_PROG	106	4800.00	2006-02-05 00:00:00.000	60	Pataballa	VPATABAL	590.423.4560	Valli	103	
IT_PROG	107	4200.00	2007-02-07 00:00:00.000	60	Lorentz	DLORENTZ	590.423.5567	Diana	103	
FI_MGR	108	12000.00	2002-08-17 00:00:00.000	100	Greenberg	NGREENBE	515.124.4569	Nancy	101	
FI_ACCOUNT	109	9000.00	2003-08-18 00:00:00.000	100	Faviet	DFAVIET	515.124.4169	Daniel	108	

Table -> Preview table 를 클릭하여 Athena로 S3의 데이터를 조회합니다.

(참고)Athena Query Result 가 저장될 수 있는 S3 Bucket 은 AmazonAthenaFullAccess Policy에 포함되어 있으면, 이 Policy 는 datalake\_user 에 할당되어 있습니다. 만약, 다른 Bucket Name 을 사용하고자 한다면, 해당 Policy 에 S3 Bucket 을 추가해야 합니다.

The screenshot shows the AWS IAM User Management console. On the left, the navigation pane shows 'Identity and Access Management (IAM)' with 'Users' selected. The main area shows the 'Summary' for a user named 'datalake\_user'. The 'Permissions' tab is selected, showing two policies applied: 'AmazonAthenaFullAccess' (an AWS managed policy). The policy summary is displayed in JSON format, with a red box highlighting the 'Resource' section containing the ARN 'arn:aws:s3:::aws-athena-query-results-\*'.

```

{
  "Effect": "Allow",
  "Action": [
    "s3:GetBucketLocation",
    "s3:GetObject",
    "s3>ListBucket",
    "s3>ListBucketMultipartUploads",
    "s3>ListMultipartUploadParts",
    "s3 AbortMultipartUpload",
    "s3 CreateBucket",
    "s3 PutObject"
  ],
  "Resource": [
    "arn:aws:s3:::aws-athena-query-results-*"
  ]
}

```

## Task 2.11 : Amazon Redshift Spectrum 으로 Data Lake 조회하기

- IAM Administrator User 로 IAM Console(<https://console.aws.amazon.com/iam/>)에 접속합니다.
- Redshift 를 위한 IAM Policy 를 생성합니다.

The screenshot shows the AWS IAM Policies page. On the left, there's a sidebar with 'AWS Account' and 'Policies' selected. In the main area, there's a table titled 'Filter policies' with a search bar. A red box highlights the 'Create policy' button at the top left of the table area.

왼쪽 메뉴에서 Policies 를 선택한 후, “Create policy”버튼을 클릭합니다.

JSON 탭에서 아래 Policy 를 붙혀 넣고, “Create policy”를 클릭합니다.

The screenshot shows the 'Create policy' wizard, Step 1: Set permissions. It has two tabs: 'Visual editor' and 'JSON'. The 'JSON' tab is selected and contains the following policy document:

```

1  {
2   "Version": "2012-10-17",
3   "Statement": [
4     {
5       "Effect": "Allow",
6       "Action": [
7         "lakeformation:GetDataAccess",
8         "glue:GetTable",
9         "glue:GetTables",
10        "glue:SearchTables",
11        "glue:GetDatabase",
12        "glue:GetDatabases",
13        "glue:GetPartitions"
14      ],
15      "Resource": "*"
16    }
17  ]
  
```

At the bottom right, the 'Review policy' button is highlighted with a red box.

The screenshot shows the 'Create policy' wizard, Step 2: Review policy. The policy document is displayed in a large text area:

```

{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "lakeformation:GetDataAccess",
        "glue:GetTable",
        "glue:GetTables",
        "glue:SearchTables",
        "glue:GetDatabase",
        "glue:GetDatabases",
        "glue:GetPartitions"
      ],
      "Resource": "*"
    }
  ]
}
  
```

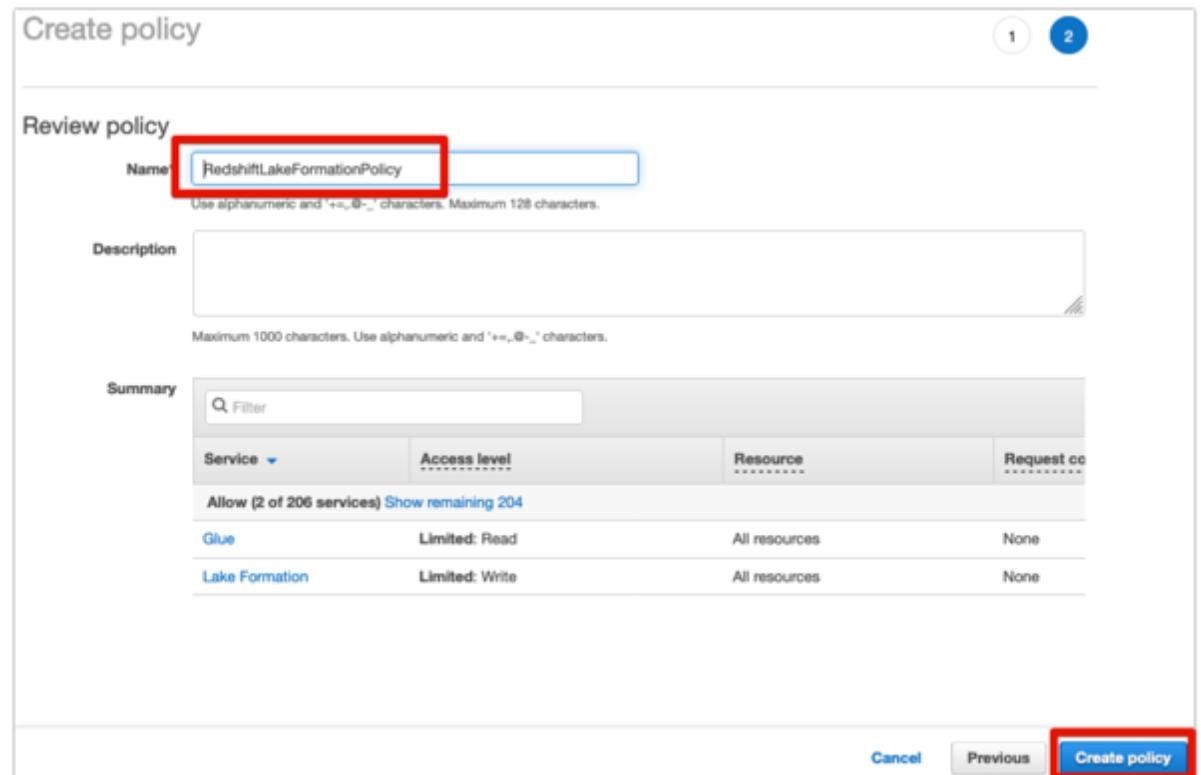
At the bottom right, the 'Review policy' button is highlighted with a red box.

```

        "Resource": "*"
    }
]
}

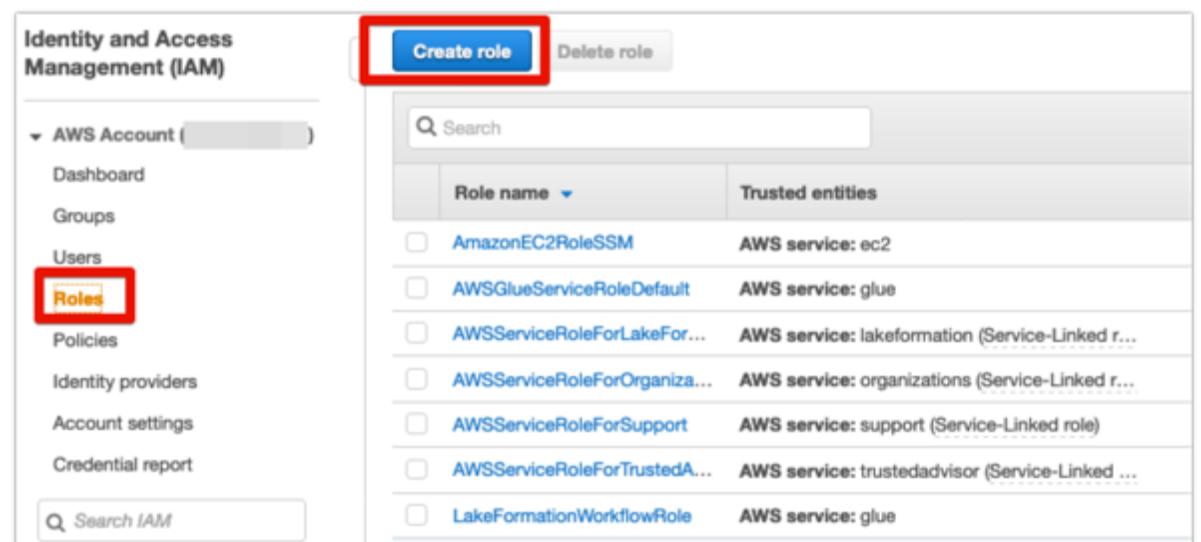
```

Policy Name 에 “RedshiftLakeFormationPolicy”를 입력합니다.



### 3. Redshift 를 위한 Role 을 생성합니다.

Roles -> Create role 버튼을 클릭합니다.



“Select type of trusted entity”에 AWS Service 를 선택하고, Choose the service…에는 Redshift 를 선택합니다.

Select your use case 에는 Redshift Customizable 을 선택하고, “Next: Permissions”를 클릭합니다.

Create role

Select type of trusted entity

<b>AWS service</b> EC2, Lambda and others	Another AWS account Belonging to you or 3rd party	Web identity Cognito or any OpenID provider	SAML 2.0 federation Your corporate directory
--	--	--	---

Allows AWS services to perform actions on your behalf. [Learn more](#)

Choose the service that will use this role

<b>EC2</b> Allows EC2 instances to call AWS services on your behalf.
<b>Lambda</b> Allows Lambda functions to call AWS services on your behalf.
API Gateway      CodeBuild      EKS      Kinesis      S3
AWS Backup      CodeDeploy      EMR      Lambda      SMS
AWS Chatbot      CodeStar Notifications      ElastiCache      Lex      SNS
AWS Support      Comprehend      Elastic Beanstalk      License Manager      SWF
Amplify      Config      Elastic Container Service      Machine Learning      SageMaker
AppStream 2.0      Connect      Elastic Transcoder      Macie      Security Hub
AppSync      DMS      Elastic Load Balancing      MediaConvert      Service Catalog
Application Auto Scaling      Data Lifecycle Manager      Forecast      Migration Hub      Step Functions
Application Discovery Service      Data Pipeline      Global Accelerator      OpsWorks      Storage Gateway
Batch      DataSync      Glue      Personalize      Textract
Chime      DeepLens      Greengrass      QLDB      Transfer
CloudFormation      Directory Service      GuardDuty      RAM      Trusted Advisor
CloudHSM      DynamoDB      Inspector      RDS      VPC
CloudTrail      EC2      IoT      Redshift      WorkLink
CloudWatch Application Insights      EC2 - Fleet      IoT Things Graph      Rekognition      WorkMail
CloudWatch Events      EC2 Auto Scaling      KMS      RoboMaker

Select your use case

<b>Redshift</b> Allows Redshift clusters to call AWS services on your behalf.
<b>Redshift - Customizable</b> Allows Redshift clusters to call AWS services on your behalf.

\* Required      Cancel      **Next: Permissions**

“RedshiftLakeFormationPolicy” Policy 를 검색하여, 체크하고 “Create role” 단계까지 진행합니다.

Create role

Attach permissions policies

Choose one or more policies to attach to your new role.

**Create policy**      **Filter policies**  Showing 1 result

Policy name	Used as
<input checked="" type="checkbox"/> RedshiftLakeFormationPolicy	None

Create role

Review

Provide the required information below and review this role before you create it.

Role name*	RedshiftLakeFormationRole
Use alphanumeric and '+-, @-' characters. Maximum 64 characters.	
Role description	Allows Redshift clusters to call AWS services on your behalf.
Maximum 1000 characters. Use alphanumeric and '+-, @-' characters.	
Trusted entities	AWS service: redshift.amazonaws.com
Policies	<a href="#">RedshiftLakeFormationPolicy</a>
Permissions boundary	Permissions boundary is not set
* Required	<a href="#">Cancel</a> <a href="#">Previous</a> <a href="#" style="border: 2px solid red; padding: 2px;">Create role</a>

Role name에 “RedshiftLakeFormationRole”을 입력하고, “Create role”버튼을 클릭합니다.

4. Data Lake Administrator User(datalake\_admin)로 Lake Formation Console(<https://console.aws.amazon.com/lakeformation/>)에 접속합니다.
5. Redshift Role에 SELECT Table 권한을 부여합니다.

Permissions -> Data permissions -> Grant를 클릭합니다.

AWS Lake Formation X

AWS Lake Formation > Permissions

Data permissions (21)

Choose a database or table for which to review, grant or revoke user permissions.

Principal	Principal type	Resource type	Resource	Permissions	Grantable
LakeFormationWorkflowRole	IAM role	Database	lakeformation_tutorial	Alter, Create table, Drop	-
datalake_admin	IAM user	Database	lakeformation_tutorial	Alter, Create table, Drop	Alter, Create table, Drop
datalake_admin	IAM user	Database	sampledb	Alter, Create table, Drop	Alter, Create table, Drop
IAMAllowedPrincipals	Group	Database	lakeformation_tutorial	Super	-
IAMAllowedPrincipals	Group	Database	sampledb	Super	-
LakeFormationWorkflowRole	IAM role	Table	_cdb1_cdb1_demo_employees	Super, Alter, Delete, Drop, Insert	Super, Alter, Delete, Drop, Insert

Grant 페이지에서 아래 정보를 입력합니다.

IAM users and roles	RedshiftLakeFormationRole
Database	lakeformation_tutorial
Table(optional)	xxx_employees
Table Permission	Select

**Grant permissions**

Choose the access permissions to grant. IAM permissions must also allow access.

**IAM users and roles**  
Add one or more IAM users or roles.

Choose IAM principals to add ▾

RedshiftLakeFormationRole X  
Role

**Active Directory users and groups (EMR beta only)**  
Enter one or more Active Directory users or groups.  
Ex: arn:aws:iam:<AccountID>:saml-provider/<SAMLProviderName>:user/<User>

**Database**  
Add one or more databases.

Choose databases ▾

lakeformationTutorial X

**Table - optional**  
Add one or more tables.

Choose tables ▾

cdb1\_cdb1\_demo\_employees X

**Column - optional**  
Choose filter type

None

**Table permissions**  
Choose the specific access permissions to grant.

Alter    Insert    Drop    Delete    Select

Super  
This permission is the union of the individual permissions above and supersedes them. [See here](#) ↗

**Grantable permissions**  
Choose the permissions that may be granted to others.

Alter    Insert    Drop    Delete    Select

Super  
This permission allows the principal to grant any of the above permissions and supersedes those grantable permissions.

Cancel   **Grant**

## 6. Amazon Redshift Cluster 를 생성합니다.

IAM Administrator User 로 Redshift Console(<https://console.aws.amazon.com/redshift/>)에 접속합니다.

The screenshot shows the Amazon Redshift dashboard. On the left, there's a sidebar with links like Clusters, Query editor, Saved queries, etc. In the center, there's a 'Quick launch' section with a large 'Launch cluster' button. A red box highlights this button. To the right, there's a 'Getting Started' section with links to various Redshift resources.

Quick launch 페이지에서 아래 정보를 입력합니다.

<b>Cluster identifier</b>	redshift-lakeformation-demo
<b>Database name</b>	dev
<b>Database port</b>	5439
<b>Master user name</b>	awsuser
<b>Master user password</b>	(Choose and confirm a password)
<b>Available IAM roles</b>	RedshiftLakeFormationRole

The screenshot shows the 'Launch your Amazon Redshift cluster - Quick launch' page. It includes fields for Node type (dc2.large), Nodes (2), Cluster identifier (redshift-lakeformation-demo), Database name (dev), Master user name (awsuser), Master user password, and Available IAM roles (RedshiftLakeFormationRole). The 'Launch cluster' button is highlighted with a red box.

Redshift Cluster 가 생성되는데, 몇 분 정도 소요됩니다.

7. Data Lake Analyst User 가 Redshift Cluster 에서 Query 를 수행할 수 있도록 권한을 부여합니다.

IAM Administrator User 로 IAM Console(<https://console.aws.amazon.com/iam/>)에 접속합니다.

User name	Groups	Access key age	Password age	Last activity	MFA
admin	None	None	4 days	Today	Not enabled
datalake_admin	None	None	4 days	Today	Not enabled
<b>datalake_user</b>	None	None	4 days	Today	Not enabled

Users 메뉴에서 datalake\_user 를 클릭합니다.

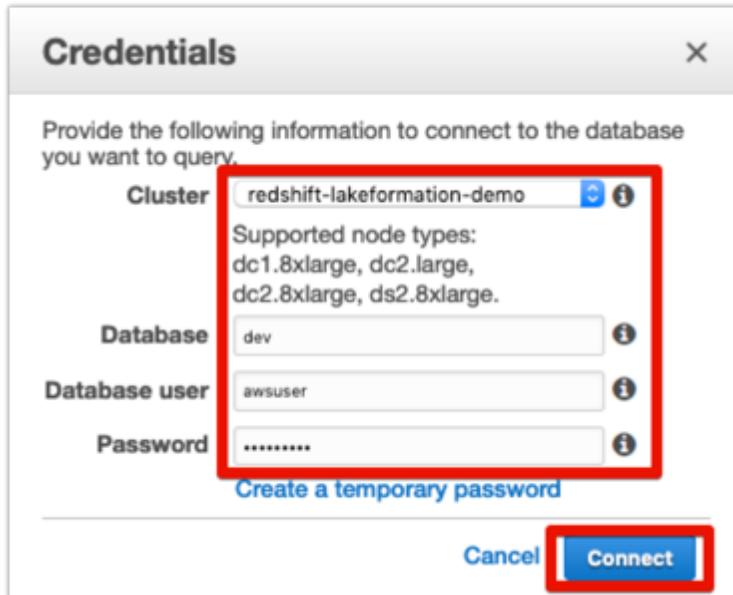
AmazonRedshiftReadOnlyAccess, AmazonRedshiftQueryEditor Policy 를 추가합니다.

Policy name	Type	Used as
AmazonRedshiftQueryEditor	AWS managed	None

## 8. Data Lake Analyst User 로 Redshift Query 를 수행합니다.

Data Lake Analyst User(datalake\_user)로 Redshift

Console(<https://console.aws.amazon.com/redshift/>)에 접속합니다.



Query editor 창에서 아래와 같이 External Schema 를 생성하는 Query 를 수행합니다.

```
create external schema if not exists oracle_cdb1 from DATA CATALOG database 'lakeformationTutorial' iam_role 'arn:aws:iam::account_id:role/RedshiftLakeFormationRole' region 'us-east-1';
```

`create external schema if not exists oracle_cdb1 from DATA CATALOG database`

`'lakeformationTutorial' iam_role`

`'arn:aws:iam::account_id:role/RedshiftLakeFormationRole' region 'us-east-1';`

`account_id`는 실제 본인의 AWS Account ID 입니다.

정상적으로 권한이 할당되었으면, 위와 같이 테이블이 표시됩니다.

The screenshot shows the AWS Redshift Spectrum Query Editor interface. On the left, there's a sidebar with 'Cluster redshift-lakeformation-demo', 'Database dev', 'Database user owneruser', 'Schema oracle\_cdb1', and 'Tables' section showing 'Showing 3 of 3 tables'. Below this is a 'Filter tables' input field. In the main area, there are two tabs: 'New Qu...'. The current tab contains the SQL query: `select * from oracle_cdb1..cdb1_cdb1_demo_employees`. Below the query is a 'Run query' button and other options like 'Save as', 'Save', and 'Clear'. To the right of the query is a note: 'Use Ctrl + Enter to run query, Ctrl + Space to autocomplete' and a 'Send feedback' link. The results section starts with 'Query results' and 'Query completed in 8.629 seconds'. It includes a 'Download CSV' button, 'Showing rows[1 - 18]', and a 'View execution' button. The results table has columns: job\_id, employee\_id, salary, hire\_date, department\_id, last\_name, and email. The data is as follows:

	job_id	employee_id	salary	hire_date	department_id	last_name	email
1	AD_PRES	100	24000.00	2003-06-17 00:00:00.000000	90	King	SKING
2	AD_VP	101	17000.00	2005-09-21 00:00:00.000000	90	Kochhar	NKOCHHAR
3	AD_VP	102	17000.00	2001-01-13 00:00:00.000000	90	De Haan	LDEHAAN
4	IT_PROG	103	9000.00	2006-01-03 00:00:00.000000	60	Hunold	AHUNOLD
5	IT_PROG	104	6000.00	2007-05-21 00:00:00.000000	60	Ernst	BERNST
6	IT_PROG	105	4800.00	2006-06-25 00:00:00.000000	60	Austin	DAUSTIN
7	IT_PROG	106	4800.00	2006-02-05 00:00:00.000000	60	Pataballa	VPATABAL
8	IT_PROG	107	4200.00	2005-05-07 00:00:00.000000	60	Lorentz	DLORENTZ
9	FI_MGR	108	12008.00	2002-08-17 00:00:00.000000	100	Greenberg	NGREENBR
10	FI_ACCOUNT	109	9000.00	2002-08-16 00:00:00.000000	100	Fiatel	DFAWNET

테이블 오른쪽 옆의 “눈”표시를 클릭하면, Redshift Spectrum 을 통해 S3 에 있는 데이터가 조회됩니다.

지금까지 AWS Lake Formation 을 이용하여 Relational Database 의 데이터를 Data Lake 로 옮기고, 테이블에 대한 권한설정 방법, Athena 및 Redshift Spectrum 을 통해 데이터를 조회하는 방법을 살펴보았습니다.

## Part 3. CSV 파일로부터 Data Lake 생성

### Task 3.1 : Sample dataset 파일 업로드

1. Sample dataset 을 다운로드합니다.

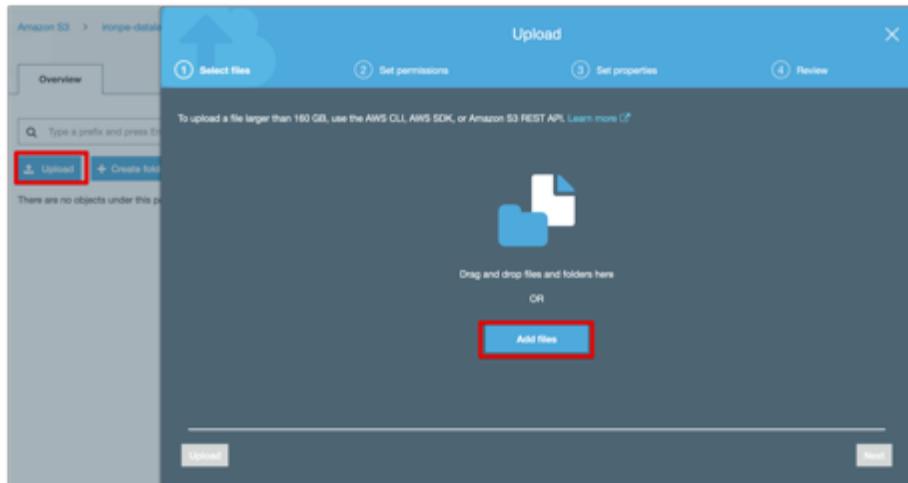
New York [Demographics Statistics by Zip](#) 링크에서 Sample dataset 을 다운로드합니다.  
다운로드되는 파일명은 Demographic\_Statistics\_By\_Zip\_Code.csv 입니다.

2. IAM Administrator User 로 접속하여, Sample dataset 을 업로드할 S3 폴더를 생성합니다.

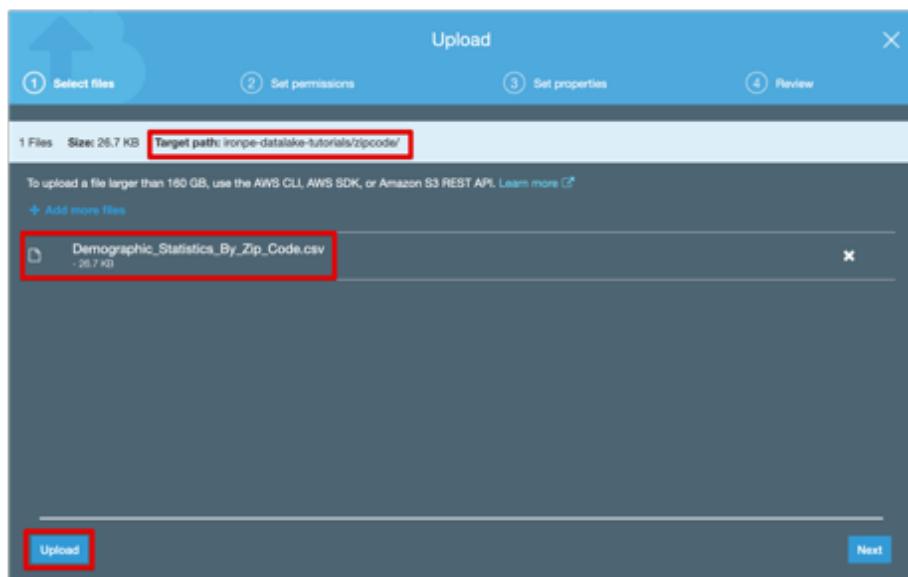
The screenshot shows the AWS S3 console interface. At the top, it says 'Amazon S3 > ironpe-datalake-tutorials'. Below this is a navigation bar with 'Overview', 'Properties', 'Permissions', and 'Management' tabs. The 'Overview' tab is selected. In the center, there's a search bar with placeholder text 'Type a prefix and press Enter to search. Press ESC to clear.' and an 'Actions' dropdown menu with options 'Upload', '+ Create folder', 'Download', and 'Actions'. Below this is a table with a single row. The first column has a checkbox and the text 'Name ▾'. The second column contains a folder icon followed by the text 'zipcode'. A red box highlights this folder name. Below the table is a note: 'When you create a folder, S3 console creates an object with the above name appended by suffix "/" and that object is displayed as a folder in the S3 console. Choose the encryption setting for the object:'. There are three radio button options: 'None (Use bucket settings)' (selected), 'AES-256' (Use Server-Side Encryption with Amazon S3-Managed Keys (SSE-S3)), and 'AWS-KMS' (Use Server-Side Encryption with AWS KMS-Managed Keys (SSE-KMS)). At the bottom are 'Save' and 'Cancel' buttons, with a red box highlighting the 'Save' button.

Data Lake S3 Bucket 내에 “zipcode” Name 으로 폴더를 생성합니다.

- Sample dataset 을 업로드합니다.

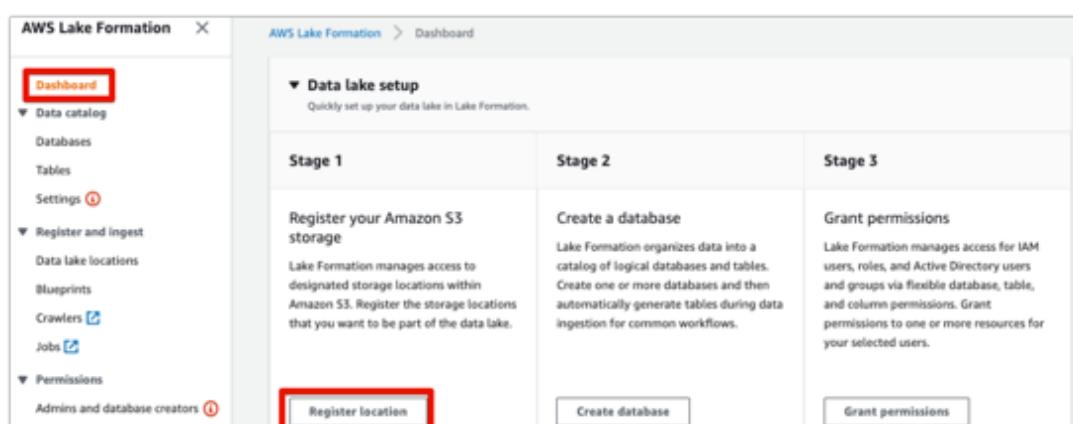


다운로드한 Demographic\_Statistics\_By\_Zip\_Code.csv 파일을 업로드합니다.



## Task 3.2 : Data Lake 설정하기

- Data Lake Administrator User(datalake\_admin)로 Lake Formation Console(<https://console.aws.amazon.com/lakeformation/>)에 접속합니다.
- Amazon S3 location 을 등록합니다.



Lake Formation 의 Dashboard -> Register location 을 클릭합니다.

The screenshot shows the 'Register location' step in the AWS Lake Formation console. It's a two-step process:

- Step 1: Register location (Main Page)**
  - Amazon S3 location:** A section for registering an Amazon S3 path as storage. It includes a text input field for the path (e.g., s3://bucket/prefix/) and a 'Browse' button (highlighted with a red box).
  - Review location permissions:** A link to review existing permissions on resources in the selected location.
  - IAM role:** A dropdown menu for selecting an IAM role. It currently shows 'AWSServiceRoleForLakeFormationDataAccess' (highlighted with a red box).
- Step 2: Choose an Amazon S3 location in region us-east-1**
  - A modal dialog titled 'Choose an Amazon S3 location in region us-east-1'. It lists paths under 'S3 / ironpe-datalake-tutorials':
    - 'cdb1\_cdb1\_demo\_employees'
    - 'zipcode' (highlighted with a red box)
  - Buttons: 'Cancel' and 'Select' (highlighted with a red box).

zipcode 폴더를 선택합니다.

AWS Lake Formation > Data lake locations > Register location

## Register location

**Amazon S3 location**

Register an Amazon S3 path as the storage location for your data lake.

**Amazon S3 path**

Choose an Amazon S3 path for your data lake.

**Browse**

**Review location permissions - strongly recommended**

Registering the selected location may result in your users gaining access to data already at that location. Before registering a location, we recommend that you review existing location permissions on resources in that location.

**Review location permissions**

**IAM role**

To add or update data, Lake Formation needs read/write access to the chosen Amazon S3 path. Choose a role that you know has permission to do this, or choose the AWSServiceRoleForLakeFormationDataAccess service-linked role. When you register the first Amazon S3 path, the service-linked role and a new inline policy are created on your behalf. Lake Formation adds the first path to the inline policy and attaches it to the service-linked role. When you register subsequent paths, Lake Formation adds the path to the existing policy.

**AWSServiceRoleForLakeFormationDataAccess**

**Cancel** **Register location**

### 3. Database 를 생성합니다.

AWS Lake Formation > Dashboard

**Dashboard** **Create database**

**Data lake setup**

Quickly set up your data lake in Lake Formation.

Stage 1	Stage 2	Stage 3
Register your Amazon S3 storage Lake Formation manages access to designated storage locations within Amazon S3. Register the storage locations that you want to be part of the data lake.	Create a database Lake Formation organizes data into a catalog of logical databases and tables. Create one or more databases and then automatically generate tables during data ingestion for common workflows.	Grant permissions Lake Formation manages access for IAM users, roles, and Active Directory users and groups via flexible database, table, and column permissions. Grant permissions to one or more resources for your selected users.

Dashboard 페이지에서 Create database 를 클릭합니다.

AWS Lake Formation > Databases > Create database

## Create database

**Database details**

Create a database in the Data Catalog.

Name: zipcode-db

Names may contain letters (A-Z), numbers (0-9), hyphens (-), or underscores (\_), and must be less than 256 characters long.

Location - optional: s3://ironpe-datalake-tutorials/zipcode

Description - optional: Enter a description

Descriptions can be up to 2048 characters long.

Default permissions for newly created tables: This setting maintains existing AWS Glue data catalog behavior. You can still set individual permissions, which will take effect when you revoke the Super permission from IAMAllowedPrincipals. See [Changing Default Settings for Your Data Lake](#).

Use only IAM access control for new tables in this database

Cancel **Create database**

Name	zipcode-db
Location	zipcod 파일을 업로드한 폴더 지정 예) s3://ironpe-datalake-tutorials/zipcode

#### 4. 권한을 부여합니다.

Dashboard 페이지에서 Grant permissions 을 클릭합니다.

AWS Lake Formation > Dashboard

**Dashboard**

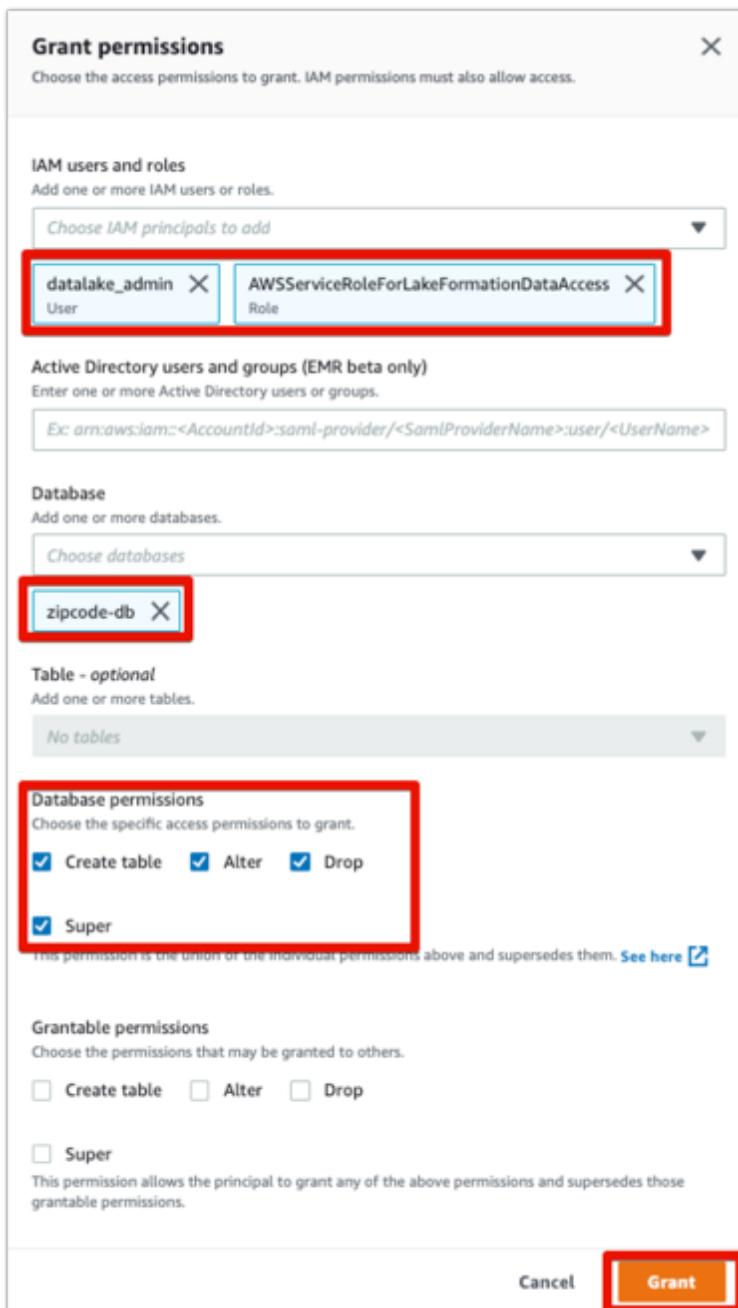
- Data catalog
- Databases
- Tables
- Settings
- Register and ingest
- Data lake locations
- Blueprints
- Crawlers
- Jobs
- Permissions
- Admins and database creators

**Data lake setup**

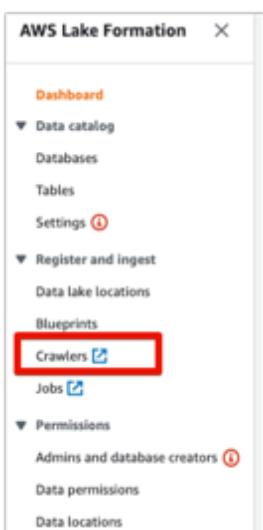
Quickly set up your data lake in Lake Formation.

Stage 1	Stage 2	Stage 3
Register your Amazon S3 storage	Create a database	Grant permissions
Lake Formation manages access to designated storage locations within Amazon S3. Register the storage locations that you want to be part of the data lake.	Lake Formation organizes data into a catalog of logical databases and tables. Create one or more databases and then automatically generate tables during data ingestion for common workflows.	Lake Formation manages access for IAM users, roles, and Active Directory users and groups via flexible database, table, and column permissions. Grant permissions to one or more resources for your selected users.

Register location **Create database** **Grant permissions**



## 5. Crawler 를 설정하고 실행합니다.



AWS Lake Formation 의 Register and ingest 에서 Crawlers 를 클릭합니다.

The screenshot shows the AWS Glue Data Catalog interface. On the left sidebar, 'Crawlers' is selected and highlighted with a red box. In the main area, there is a table with one row labeled 'lakeformation\_oracle\_discoverer\_e307b242'. At the top of the main area, there is a blue 'Add crawler' button, which is also highlighted with a red box.

Add crawler 를 클릭하고, S3 에 업로드된 데이터에서 Data Catalog 를 생성합니다.

This is the first step of the 'Add crawler' wizard. It asks for 'Add information about your crawler'. The 'Crawler name' field contains 'zipcodecrawler' and is highlighted with a red box. Below it is a note: 'Tags, description, security configuration, and classifiers (optional)'. At the bottom right is a blue 'Next' button, which is also highlighted with a red box.

Crawler name 0|| zipcodecrawler 를 입력합니다.

This is the second step of the 'Add crawler' wizard. It asks 'Specify crawler source type'. Under 'Crawler source type', the 'Data stores' option is selected and highlighted with a red box. Below it is another option 'Existing catalog tables'. At the bottom right are two buttons: 'Back' and 'Next', with 'Next' highlighted with a red box.

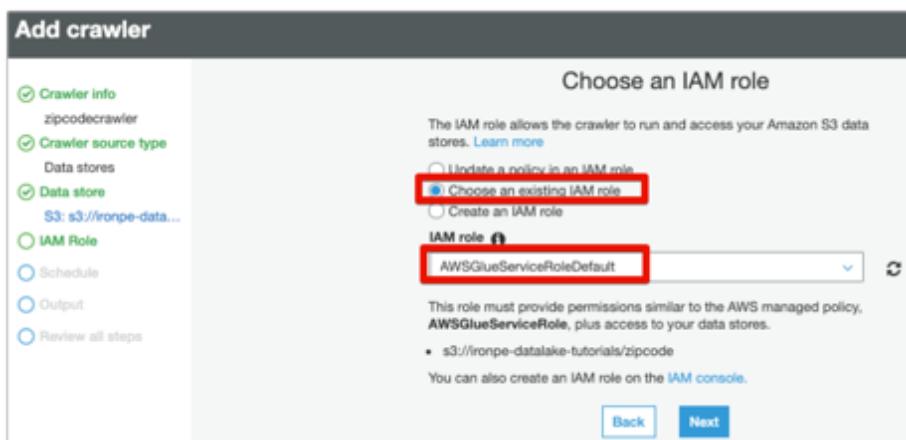
Data store 를 S3 로 지정하고, Sample dataset 을 업로드한 폴더를 지정합니다.

This is the third step of the 'Add crawler' wizard. It asks 'Add a data store'. Under 'Choose a data store', 'S3' is selected and highlighted with a red box. Below it is a note: 'Crawl data in Specified path in my account' (radio button selected). Under 'Include path', the value 's3://ironpe-datalake-tutorials/zipcode' is entered and highlighted with a red box. At the bottom right are 'Back' and 'Next' buttons, with 'Next' highlighted with a red box. At the very bottom of the page, there is a table with 'Choose a data store' set to 'S3'.

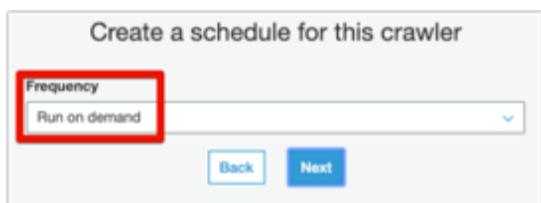
Include path	Sample dataset 을 업로드한 폴더 지정 예)s3://ironpe-datalake-tutorials/zipcode
--------------	---



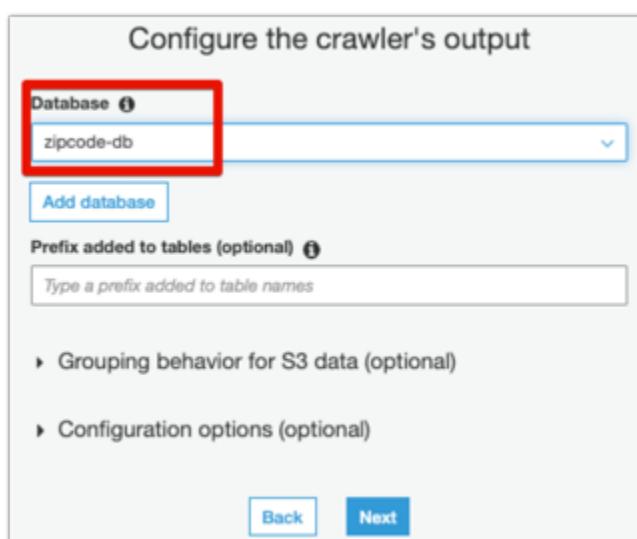
Add another data store 에는 No 를 선택합니다.



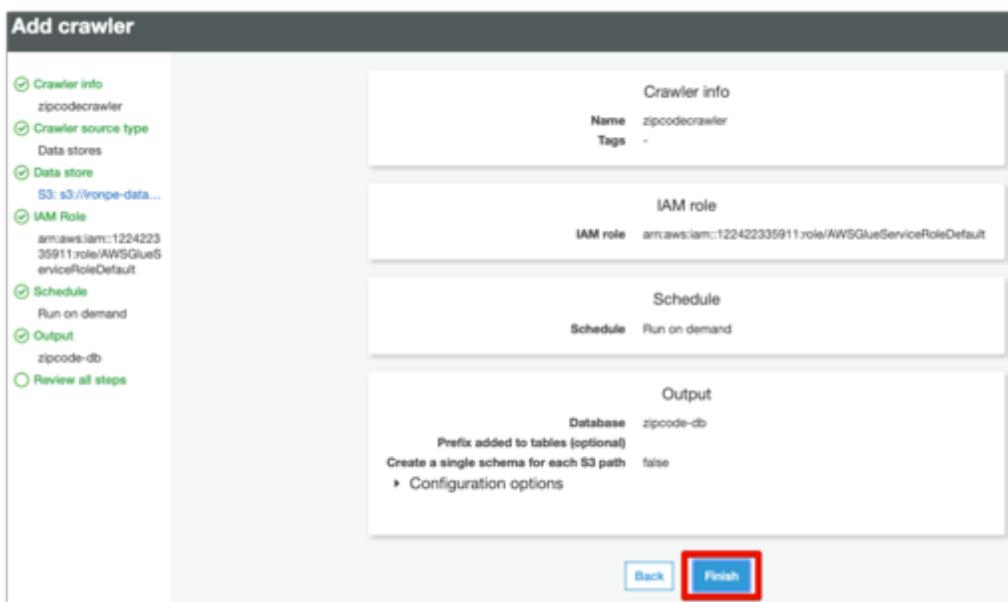
IAM role 에는 AWSGlueServiceRoleDefault 를 선택합니다.



Frequency 에는 Run on demand 를 선택합니다.



Database 는 zipcode-db 를 선택합니다.



입력된 정보를 확인하고, Finish 를 클릭합니다.

Name	Schedule	Status	Logs	Last runtime	Median runtime	Tables updated	Tables added
lakeformation_oracle_discoverer_e307b...		Ready	Logs	4 mins	4 mins	0	1
zipcodecrawler		Ready		0 secs	0 secs	0	0

Run it now?를 클릭합니다. Data Catalog 가 생성될 때까지 1~2 분정도 소요됩니다.

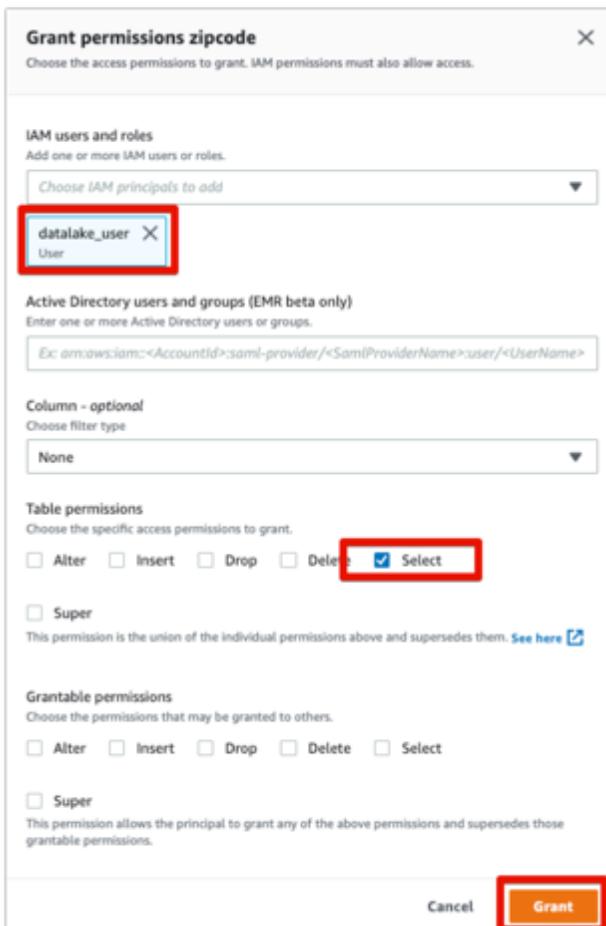
Name	Schedule	Status	Logs	Last runtime	Median runtime	Tables updated	Tables added
lakeformation_oracle_discoverer_e307b...		Ready	Logs	4 mins	4 mins	0	1
zipcodecrawler		Ready	Logs	1 min	1 min	0	1

Crawler 가 완료되면, Table 1 개가 추가됩니다. Database -> Tables 를 클릭하여 생성된 Table 을 확인합니다.

### Task 3.3 : Table 접근 권한 부여하기

1. Data Lake Administrator User(datalake\_admin)로 Lake Formation Console(<https://console.aws.amazon.com/lakeformation/>)에 접속합니다.
2. Data Lake Analyst User(datalake\_user)에게 Table에 대한 조회권한을 부여합니다.

Table 을 선택하고 Actions에서 Grant 를 선택합니다.



## Task 3.4 : Athena 로 데이터 조회하기

1. Data Lake Analyst User(datalake\_user)로 Athena

Console(<https://console.aws.amazon.com/athena/>)에 접속합니다.

2. zipcode Table 을 선택하고 Preview table 을 클릭합니다.

	jurisdiction name	count participants	count female	percent female	count male	percent male	count gender unknown	percent gender unknown	count
2	10002	35	19	54.00%	16	46.00%	0	0.00%	35
3	10003	1	1	100.00%	0	0.00%	0	0.00%	1
4	10004	0	0	0.00%	0	0.00%	0	0.00%	0
5	10005	2	2	100.00%	0	0.00%	0	0.00%	2
6	10006	6	2	33.33%	4	66.67%	0	0.00%	6
7	10007	1	0	0.00%	1	100.00%	0	0.00%	1
8	10009	2	0	0.00%	2	100.00%	0	0.00%	2
9	10010	0	0	0.00%	0	0.00%	0	0.00%	0
10	10011	3	2	66.67%	1	33.33%	0	0.00%	3

수고하셨습니다!

AWS Lake Formation Hand-On Lab 이 끝났습니다.

END OF DOCUMENT