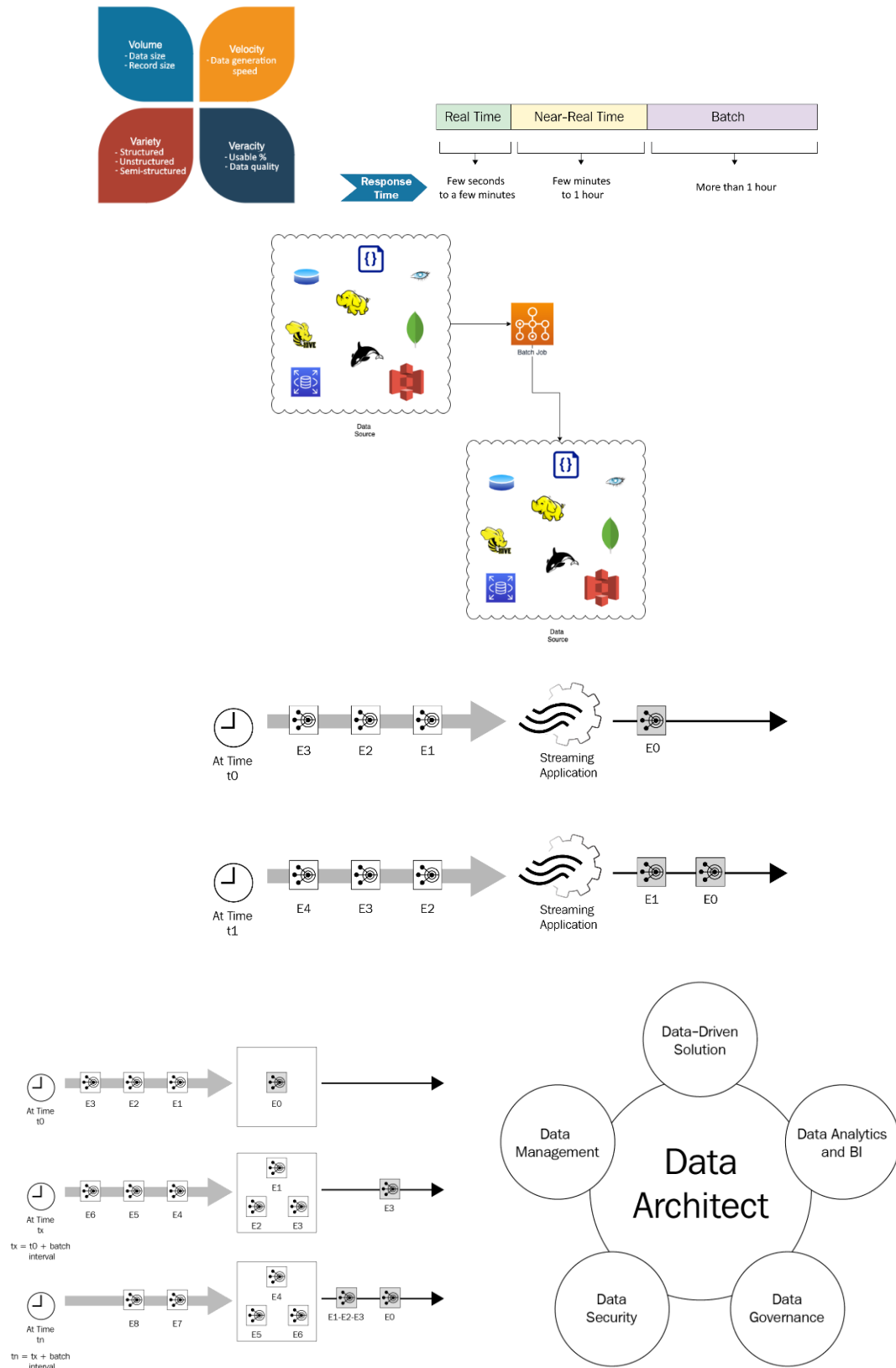
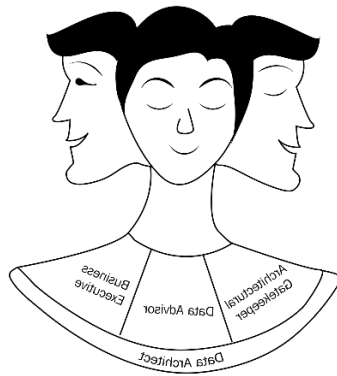
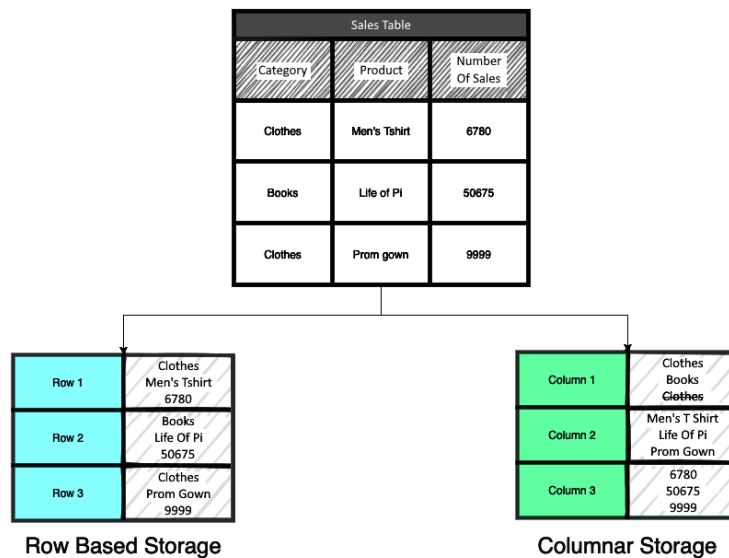
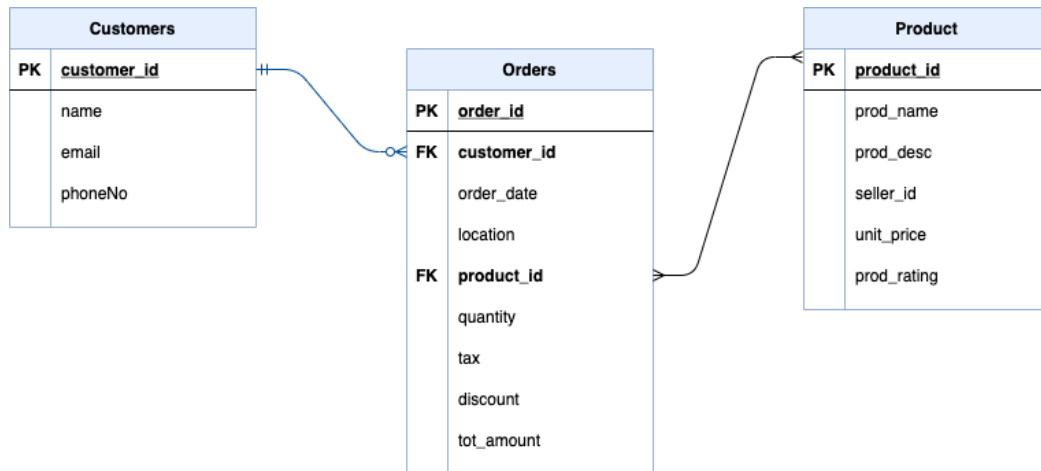


# Chapter 1: Basics of Modern Data Architecture

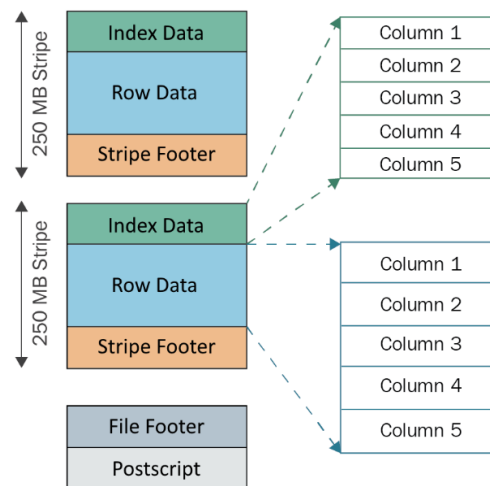
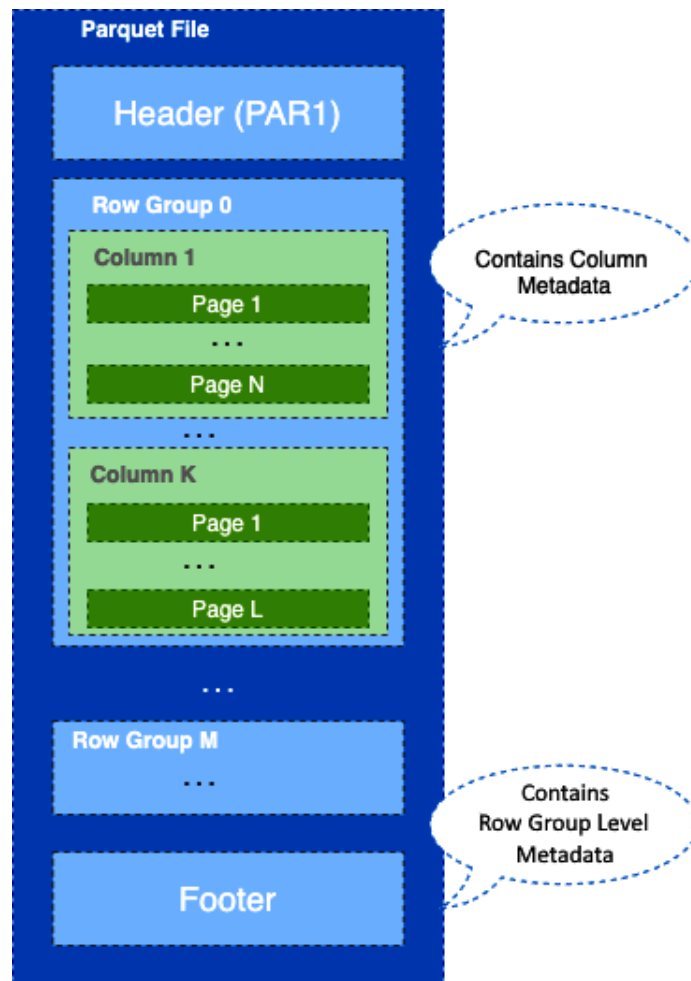




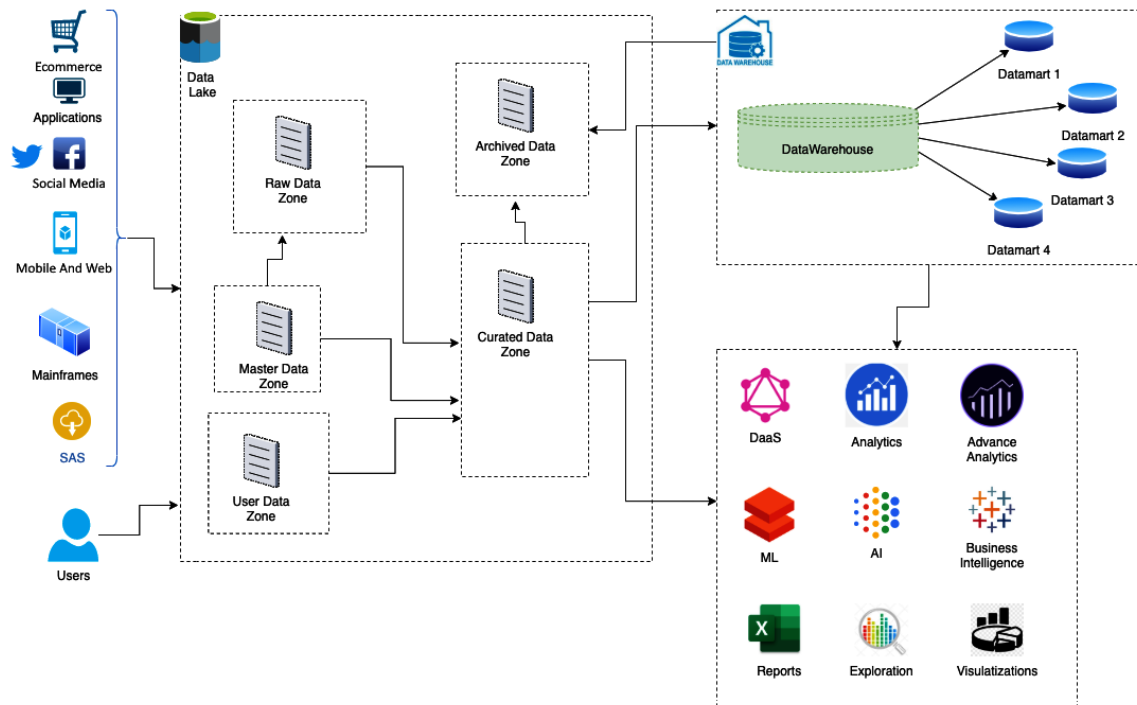
## Chapter 2: Data Storage and Databases



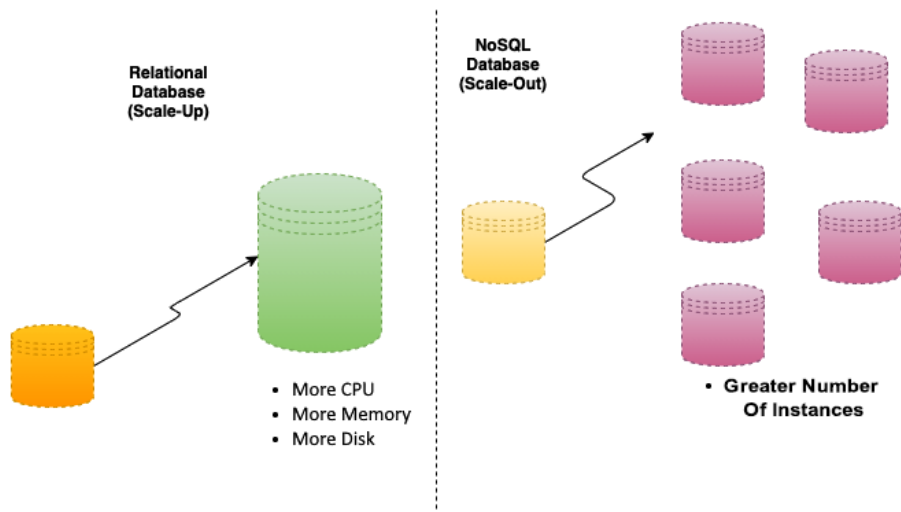
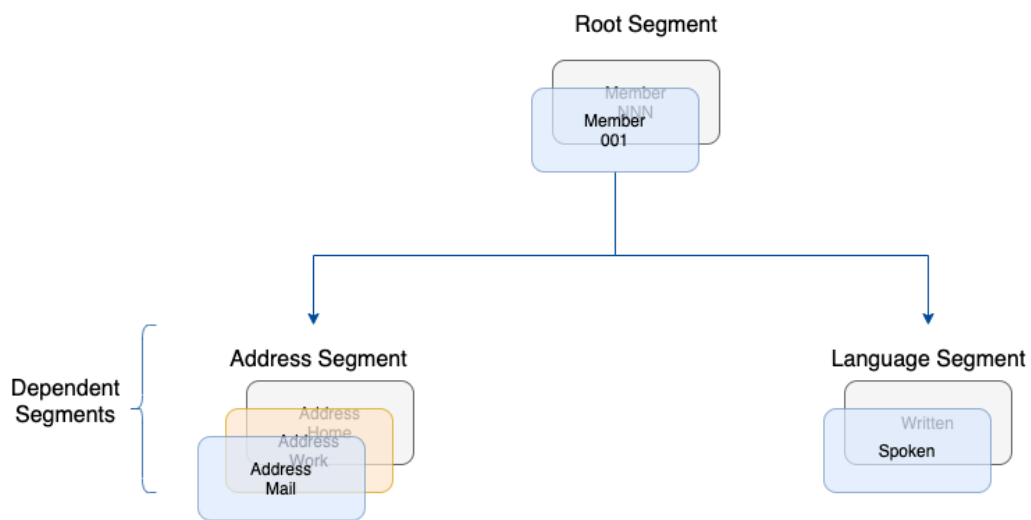
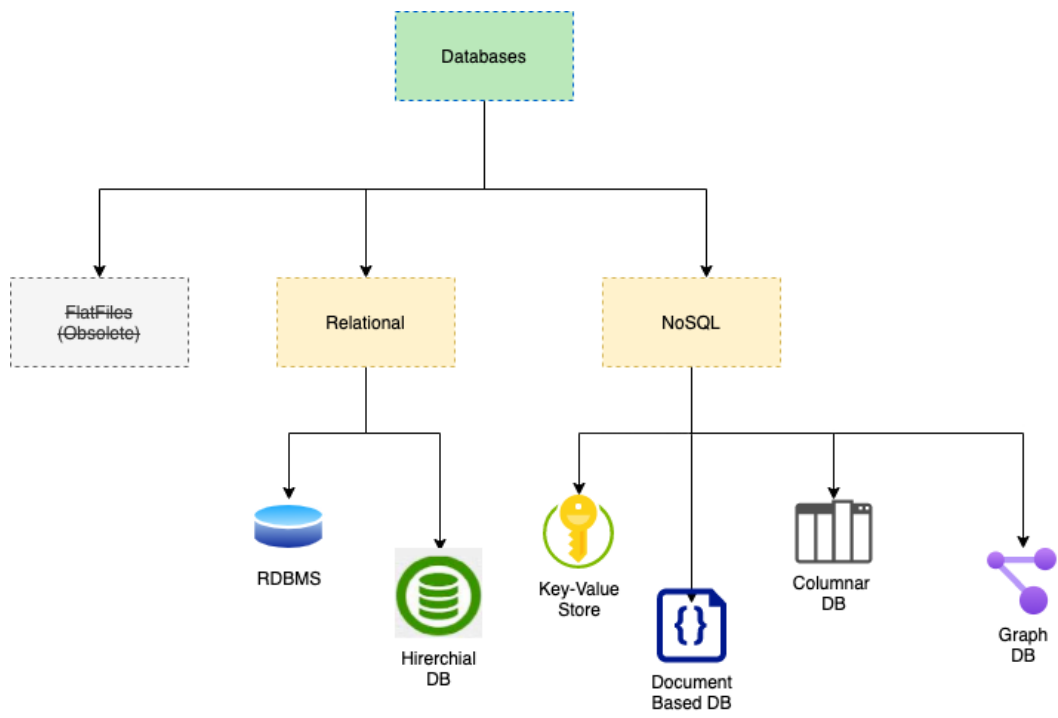
Name	Lossless Compress	Compression Ratio	Splitable	Compression Speed	Decompress Speed
Gzip	Yes	2.7x-3x	No	100 MBps	440 MBps
Snappy	Yes	2x	No	580 MBps	2020 MBps
LZ4	Yes	2.5x	No	800 MBps	4220 MBps
Zstd	Yes	2.8x	Yes	530 MBps	1360 MBps



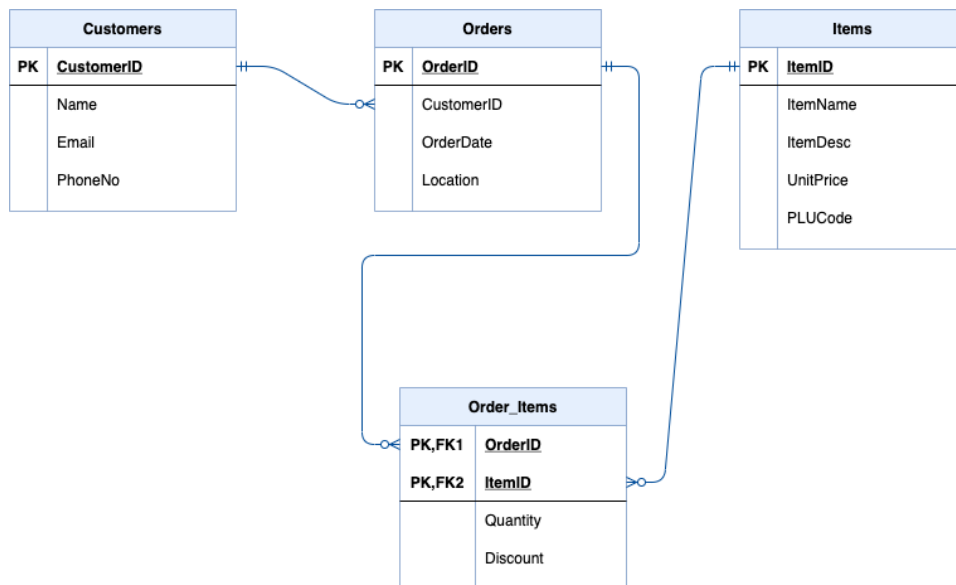
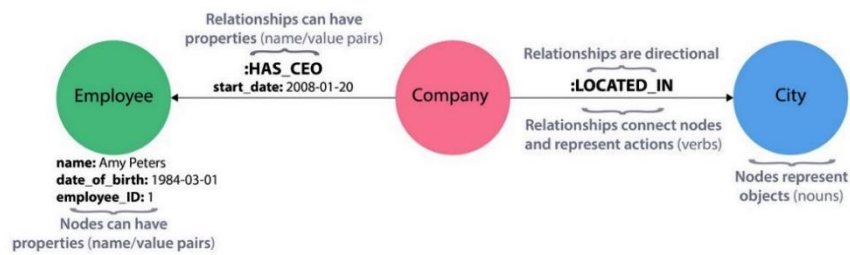




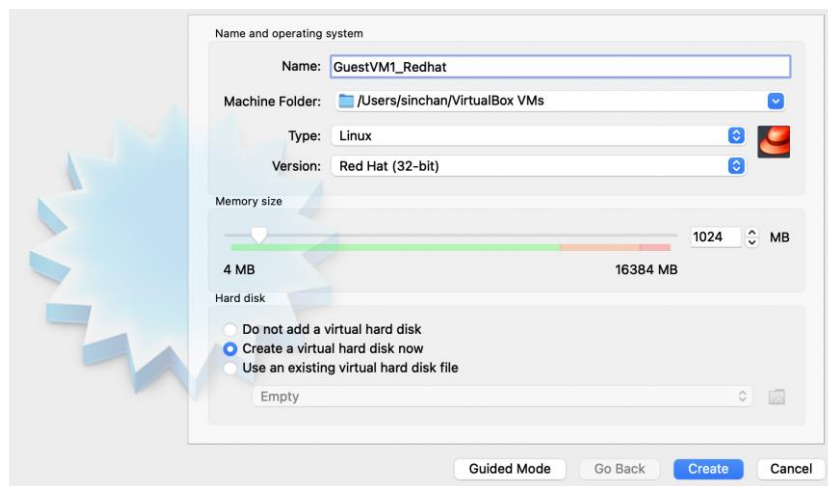
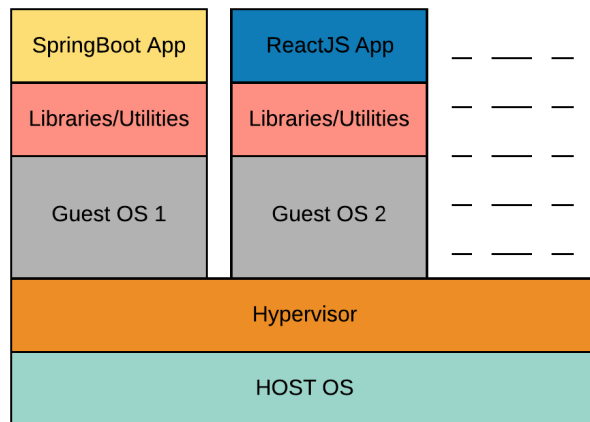
Characteristics	Data Lake	Data warehouse
<b>Load Pattern</b>	ELT (Extract, Load, and Transform)	ETL (Extract, Transform, and Load)
<b>Type Of Data Stored</b>	Structured, semi-structured and unstructured	Structured
<b>Analysis Pattern</b>	Acquire, analyze, and then determine structure of curated data	Create the structure first and then acquire the data for insights
<b>Data Ingestion Pattern</b>	Batch processing, real-time, near real-time processing	Batch processing
<b>Schema Application Time</b>	Schema-on-read i.e., schema is applied while reading the data	Schema-on-write i.e., schema is determined and is available when data is written

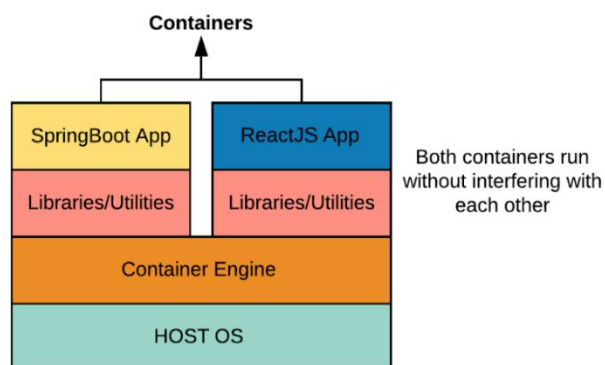
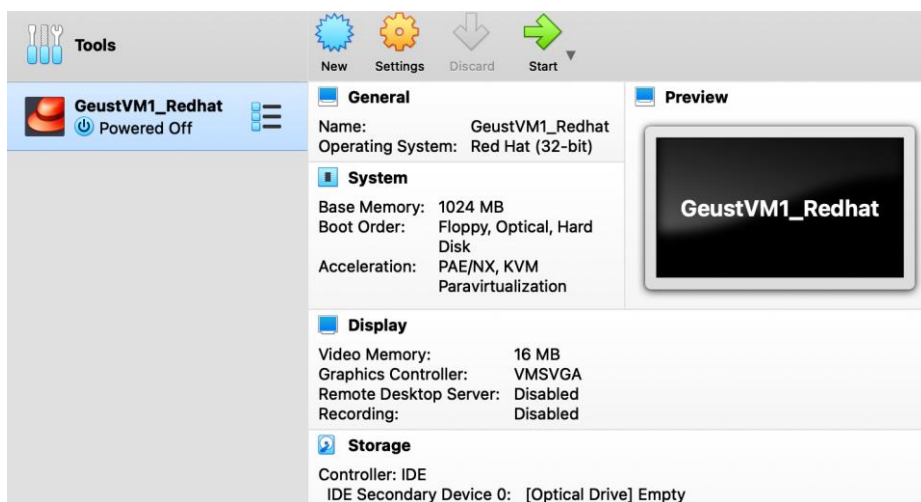
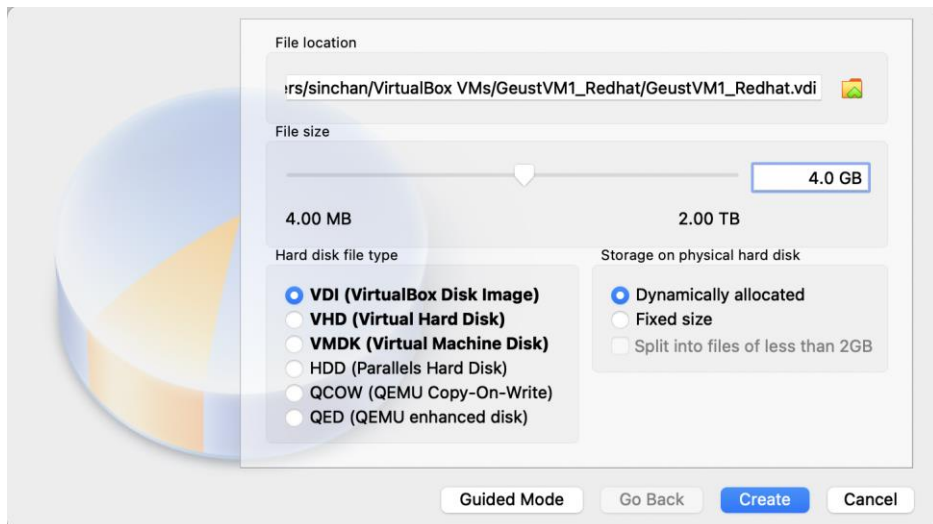


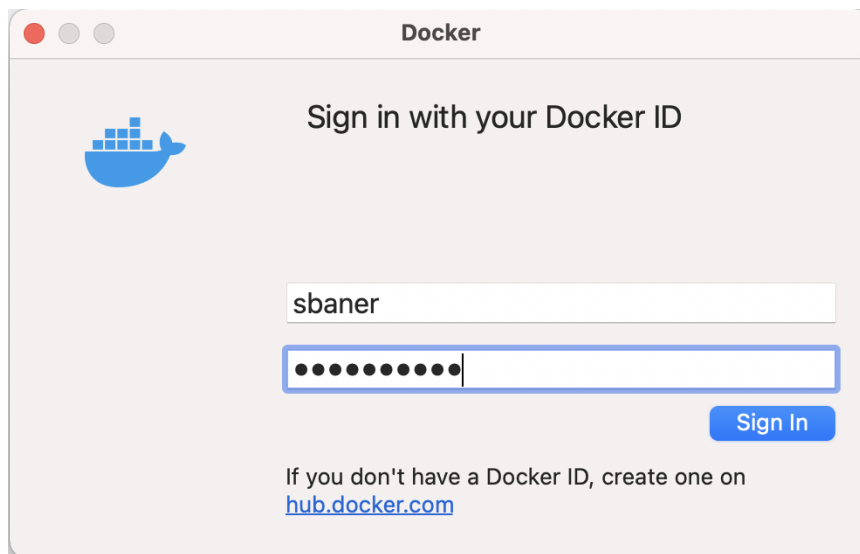
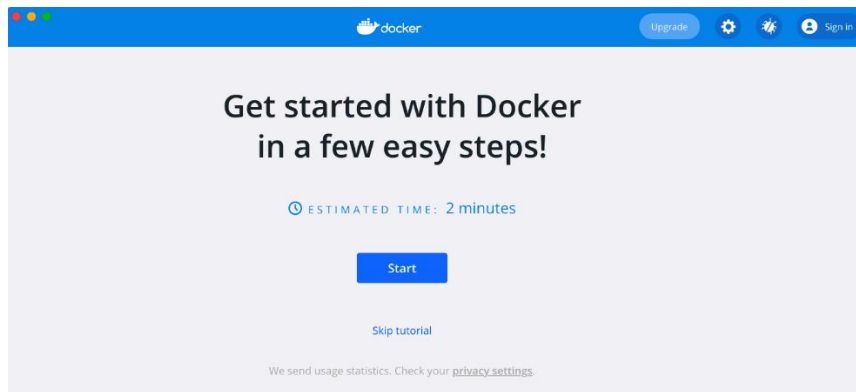
Column Family 1				Column Family 2		
Region 1		cf1 :col-A	cf1 :col-B	cf2:col-Foo	cf2:col-XYZ	cf2:foobar
	row-1					
	row-10					
	row-18	A18 - v1	B18 - v3	Foo18 - v1	XYZ18 - v2	foobar18 - v1
Region 2	row-2					
	row-5					
	row-6					
	row-7					



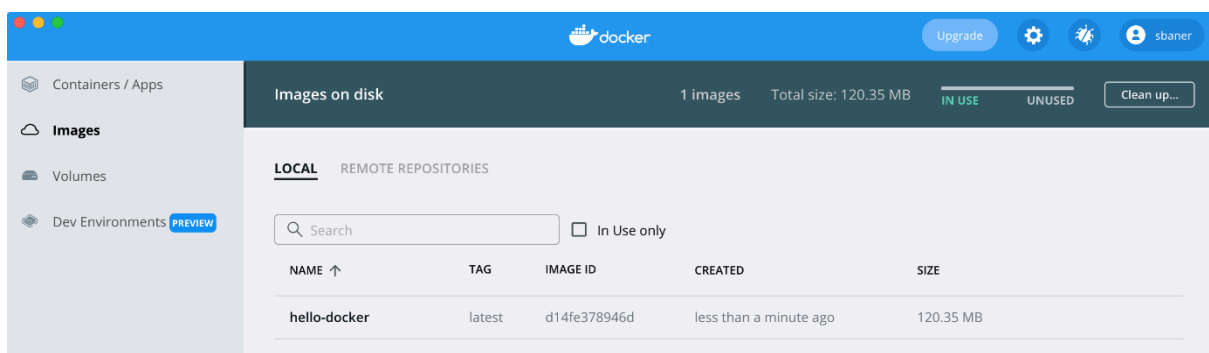
## Chapter 3: Identifying the Right Data Platform

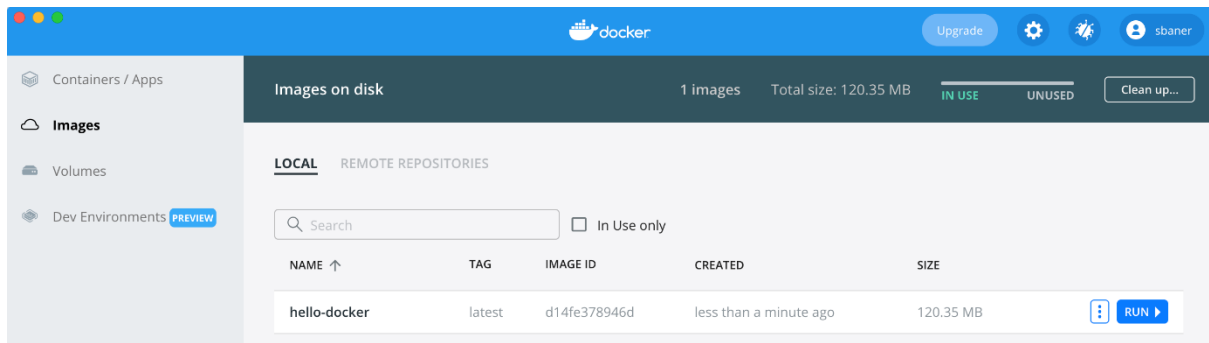






Command	Description
FROM	To specify the base image.
EXPOSE	To define which port will be used to access your container application.
ENTRYPOINT	A command that will always be executed when the container starts. If not specified, the default is /bin/sh -c
CMD	Arguments passed to the entrypoint. If ENTRYPOINT is not set (defaults to /bin/sh -c), the CMD will be the command the container executes.
COPY	To copy over files or directories from a specific location.
ADD	The same as COPY, but also able to handle remote URLs and unpack compressed files.
RUN	To install any applications and packages required for your container.
WORKDIR	To set the working directory for any commands that follow in the Dockerfile.





 **hello\_container** hello-docker:latest  
CREATION IN PROGRESS PORT: 8080

Optional Settings

**Container Name**

hello\_container

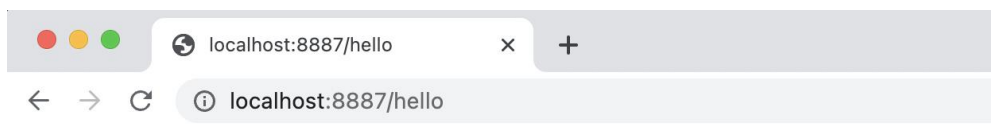
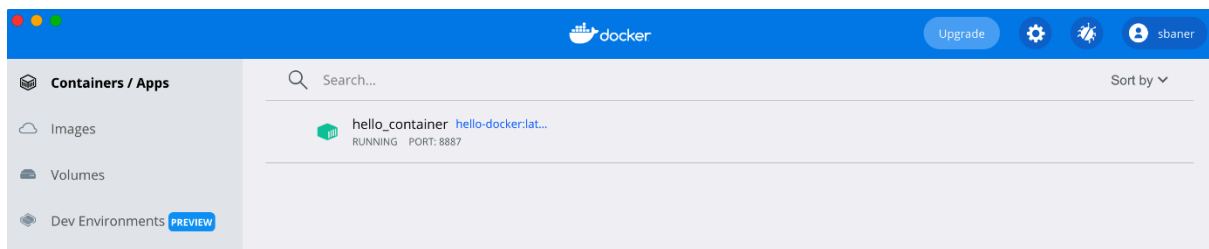
**Ports**

Local Host: 8887 Container Port: 8080/tcp

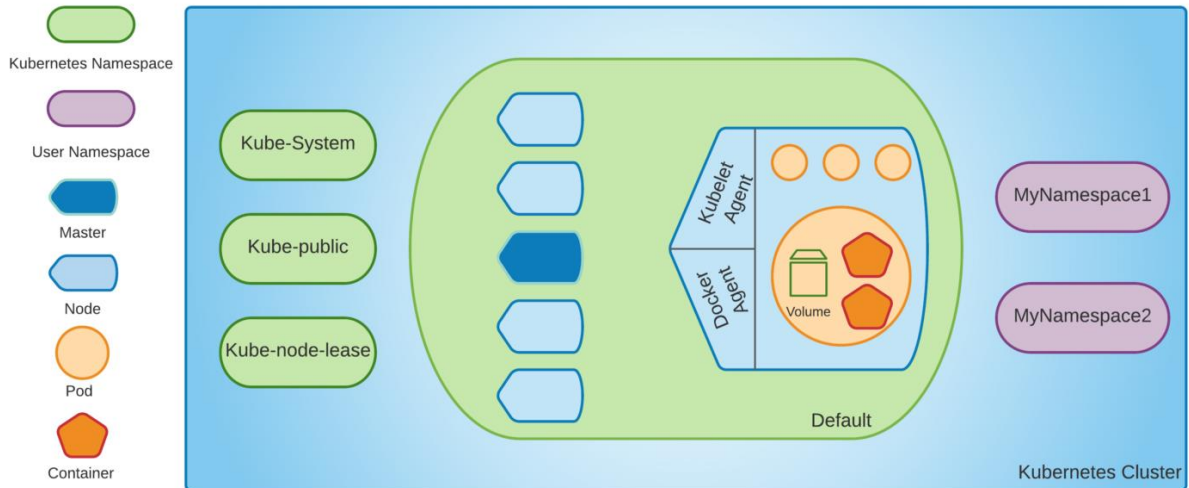
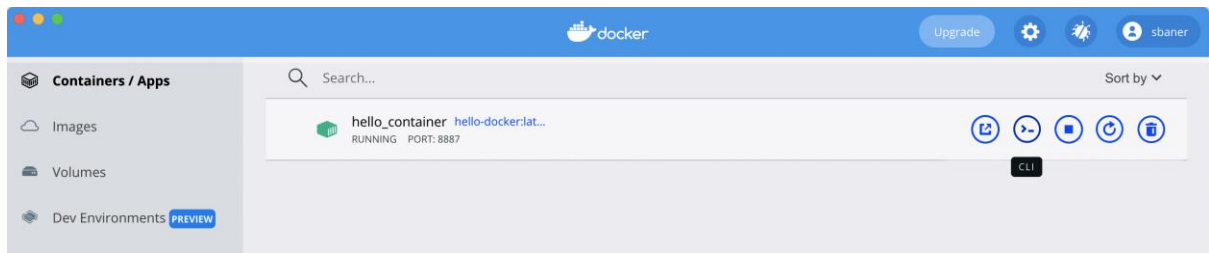
**Volumes**

Host Path: Container Path:

Cancel Run



Hello from the Container!

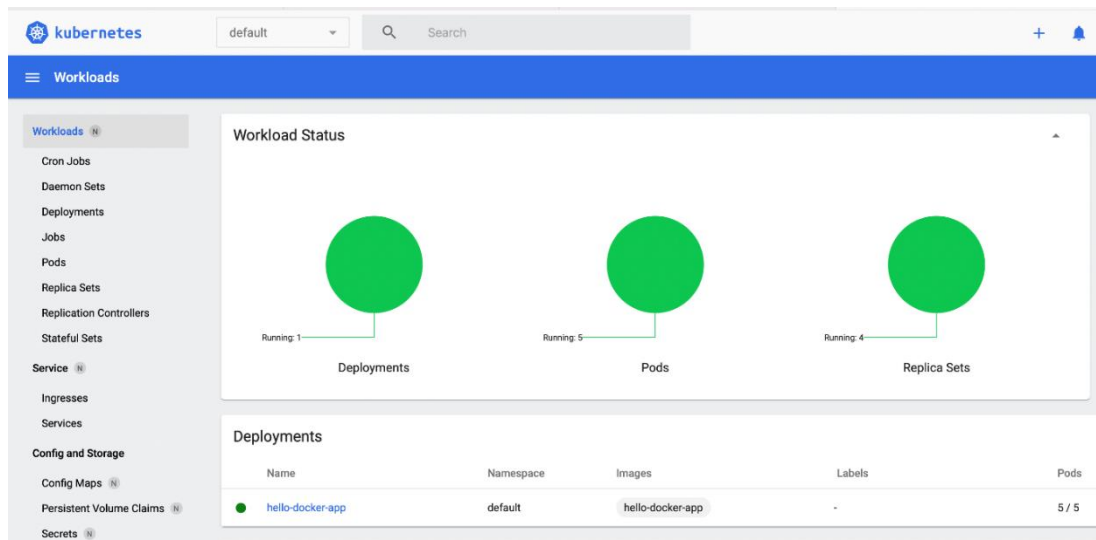


```
[sinchan@Sinchans-MBP-2 Downloads % minikube start
minikube v1.23.2 on Darwin 11.2 (arm64)
Automatically selected the docker driver. Other choices: virtualbox, ssh
Starting control plane node minikube in cluster minikube
Pulling base image ...
Downloading Kubernetes v1.22.2 preload ...
> preloaded-images-k8s-v13-v1...: 541.26 MiB / 541.26 MiB 100.00% 18.98 MiB
> gcr.io/k8s-minikube/kicbase: 321.23 MiB / 321.23 MiB 100.00% 8.45 MiB p/
Creating docker container (CPUs=2, Memory=1988MB) ...
Preparing Kubernetes v1.22.2 on Docker 20.10.8 ...
  ■ Generating certificates and keys ...
  ■ Booting up control plane ...
  ■ Configuring RBAC rules ...
Verifying Kubernetes components...
  ■ Using image gcr.io/k8s-minikube/storage-provisioner:v5
Enabled addons: storage-provisioner, default-storageclass
Done! kubectl is now configured to use "minikube" cluster and "default" namespace by default
```

```
[sinchan@Sinchans-MacBook-Pro-2 DockerExample % kubectl get services
NAME                TYPE          CLUSTER-IP      EXTERNAL-IP      PORT(S)          AGE
hello-docker-service LoadBalancer  10.99.213.136    <pending>        8080:30036/TCP   42m
kubernetes           ClusterIP     10.96.0.1        <none>           443/TCP          178m

[sinchan@Sinchans-MacBook-Pro-2 DockerExample % kubectl get deployments
NAME                READY   UP-TO-DATE   AVAILABLE   AGE
hello-docker-app    5/5     5            5           42m
```

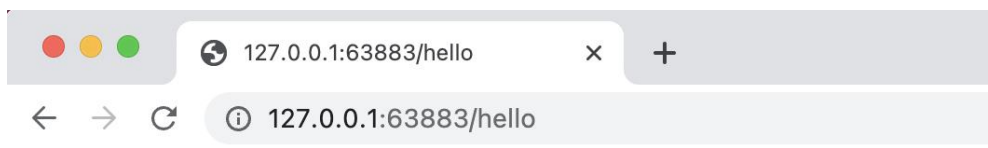




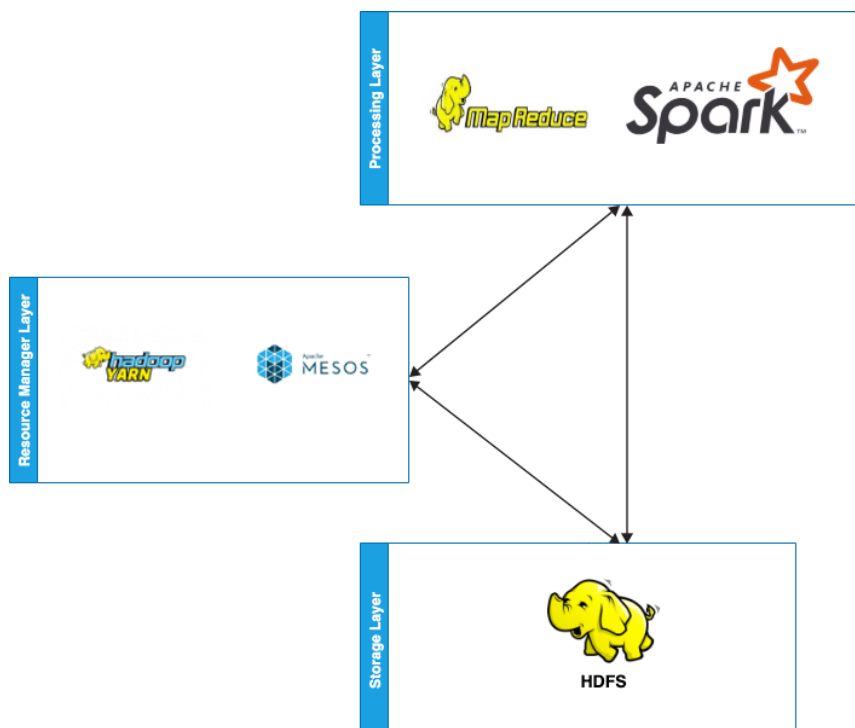
```
[sinchan@Sinchans-MacBook-Pro-2 DockerExample % minikube service --url hello-docker-service
Starting tunnel for service hello-docker-service.
```

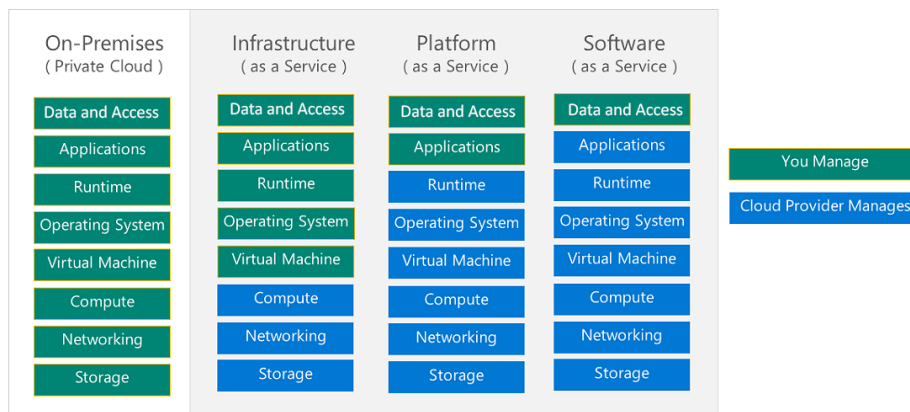
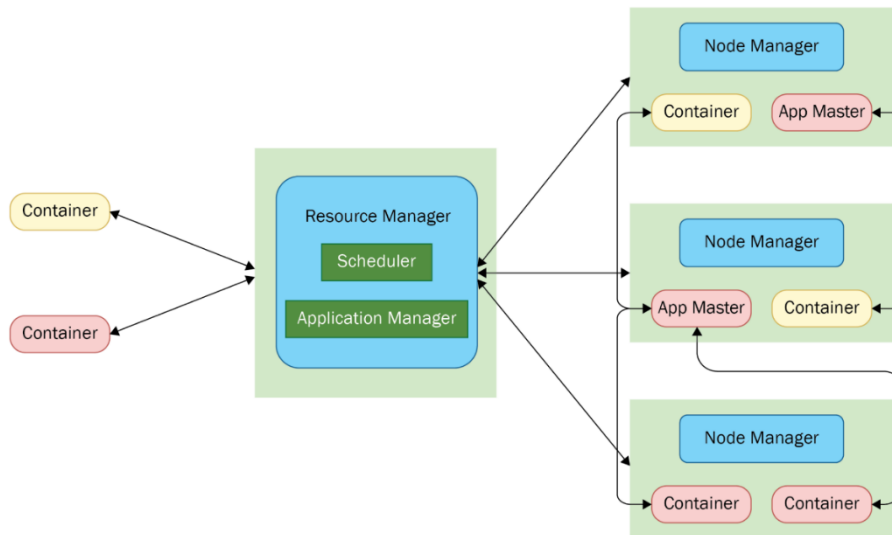
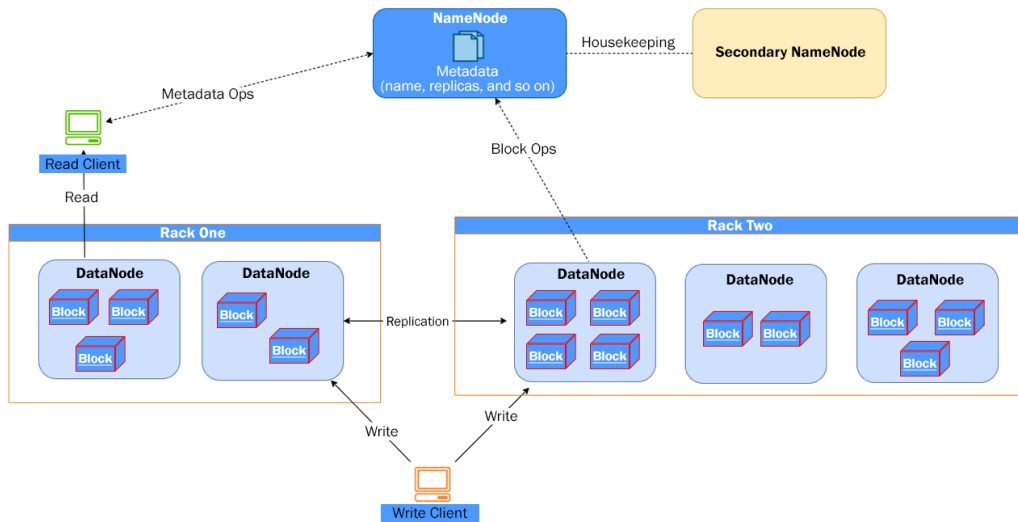
NAMESPACE	NAME	TARGET PORT	URL
default	hello-docker-service		http://127.0.0.1:63883

http://127.0.0.1:63883



Hello from the Container!

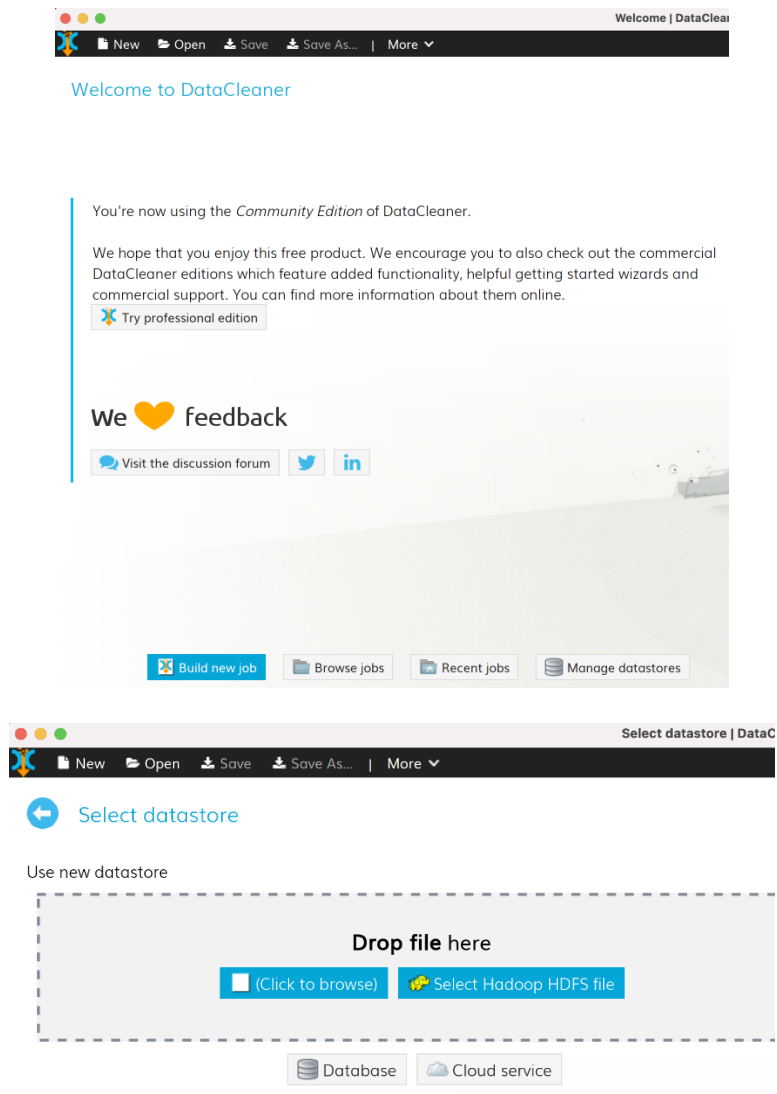




	<b>AWS</b>	<b>Azure</b>	<b>GCP</b>
Services	Huge range of services	Good range of services available. Exceptional services in AI/ML.	Limited services are available.
Maturity	Most mature	Catching up with AWS.	Still relatively less mature than the other two.
Marketplace	All vendors make their products available	Good vendor support but less than AWS.	
Reliability	Excellent	Excellent.	Excellent.
Security	Excellent	Excellent.	Fewer notches than AWS and GCP.
Cost	Varies	Most cost-efficient.	Varies.

	<b>AWS</b>	<b>Azure</b>	<b>GCP</b>
Support	Paid dev/enterprise support	Paid dev/enterprise support. More support options than AWS.	Paid dev/premium support. Costlier support than the other two.
Hybrid Cloud Support	Limited	Excellent.	Good.
Special Notes	More compute capacity versus Azure and GCP	Easy integration and migrations for existing Microsoft services.	Excellent support for containerized workloads. Global fiber network.

## Chapter 4: ETL Data Load - A Batch-Based Solution to Ingest Data in a Data Warehouse



inputData.csv | Analysis Job | Data

New Open Save Save As... More

information\_schema  
Book-Java\_Data\_Architect  
inputData.csv  
incidentNumber  
deviceSerialNum  
eventCode  
loggedTime  
closureTime  
status  
assignedTo  
resolutionComment  
default\_table

Library  
Transform  
Improve  
Analyze  
Date and time  
Machine learning  
Visualization  
Boolean analyzer  
Character set distribution  
Completeness analyzer  
Fill pattern  
Mark rows  
Number analyzer  
Pattern finder  
Reference data matcher  
Referential integrity  
String analyzer  
Unique key check

inputData.csv

String analyzer

Ready to execute

Click the 'Execute' button in the upper-right corner when you're ready to run the job.

String analyzer | DataCleaner

String analyzer

Documentation Rename

Input columns

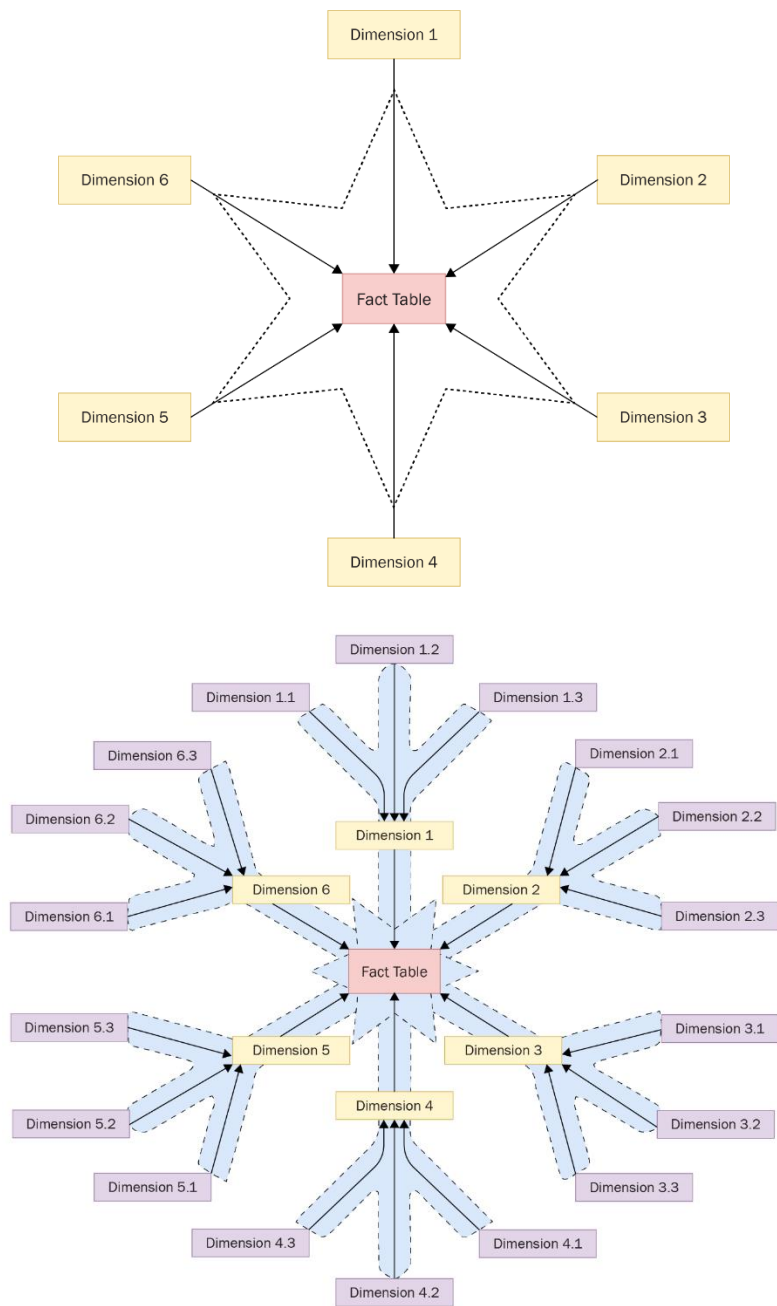
Columns:

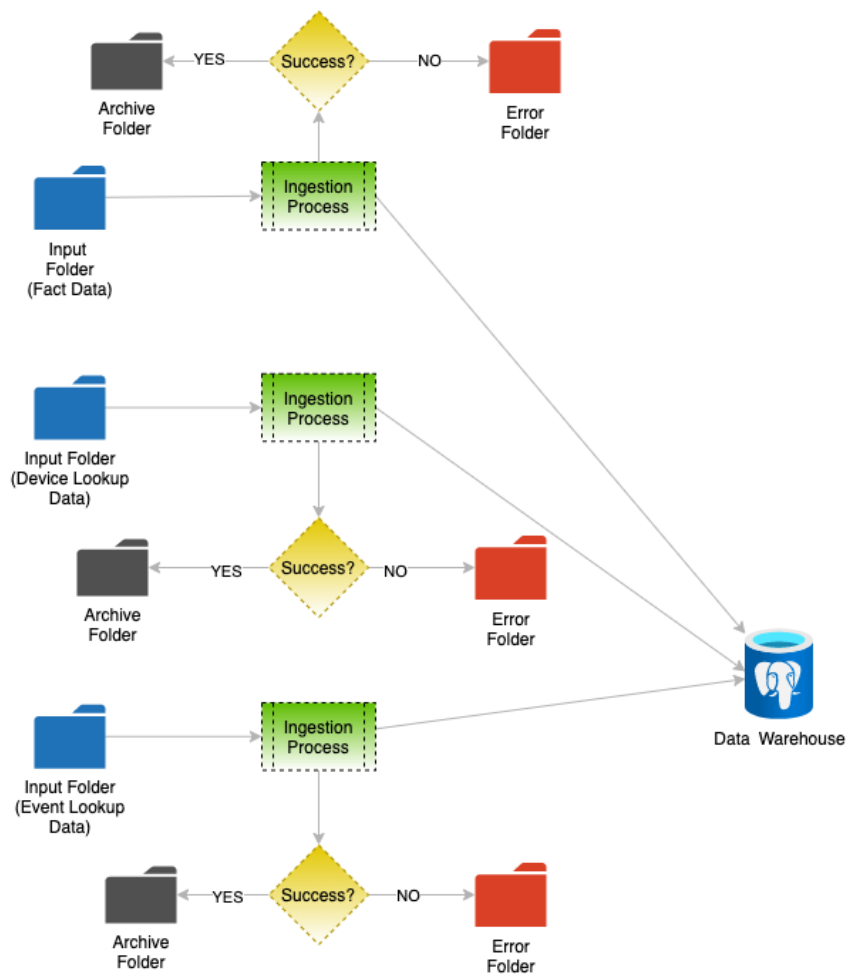
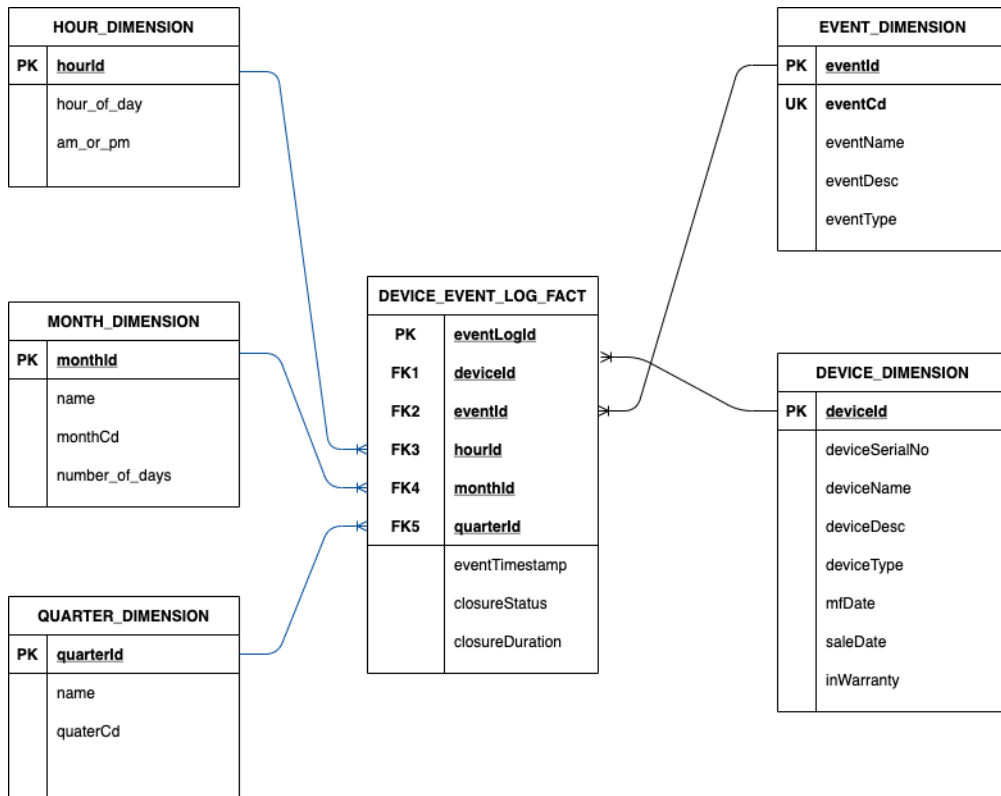
Select all Select none

☒ incidentNumber  
☒ deviceSerialNum  
☒ eventCode  
☒ loggedTime  
☒ closureTime  
☒ status  
☒ assignedTo  
☒ resolutionComment

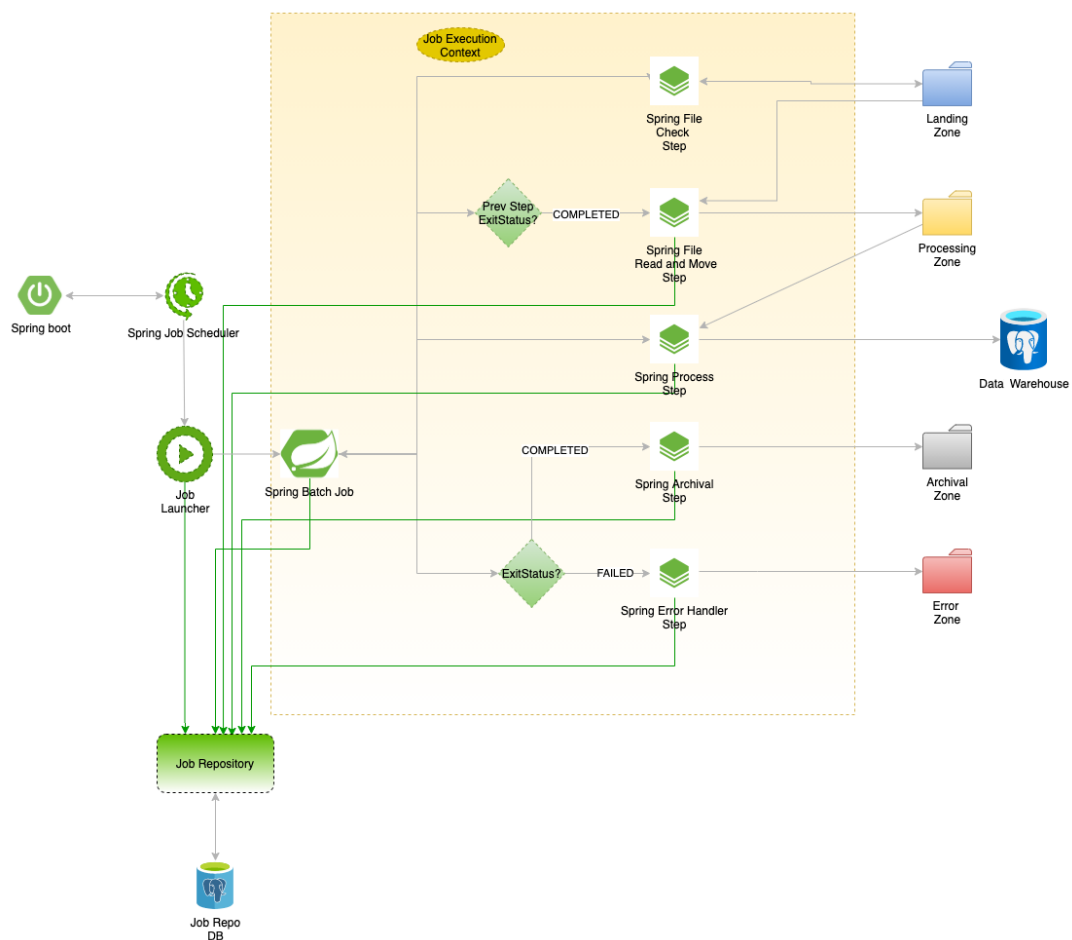
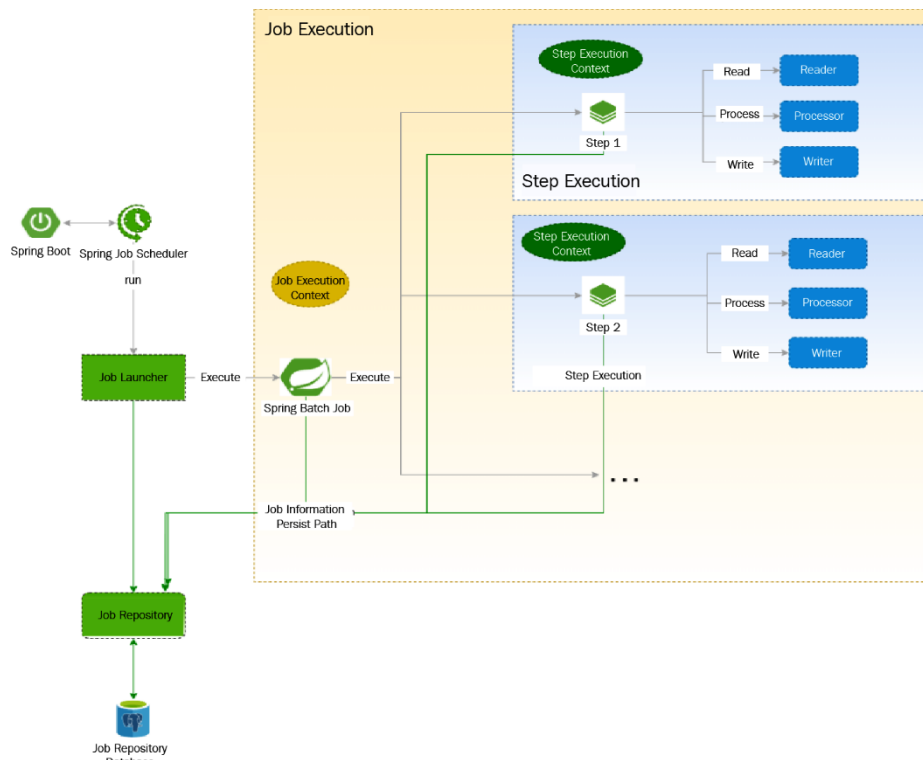


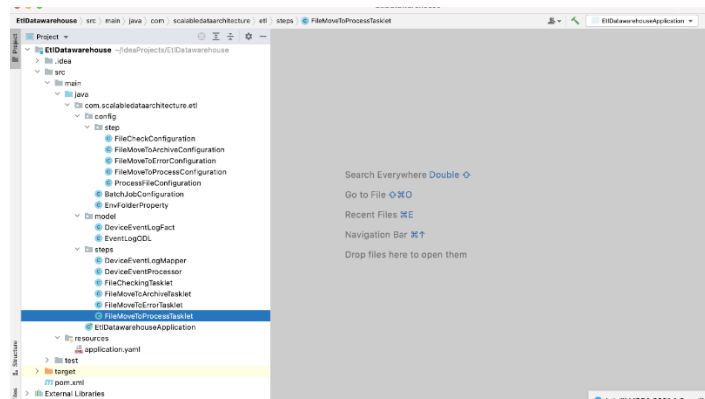
inputData.csv   Analysis results   DataCleaner										
Analysis results   inputData.csv										
String analyzer										
(8 columns)										
	incidentNumber	deviceSerialNum	eventCode	loggedTime	closureTime	status	assignedTo	resolutionComment		
Row count	300	300	300	300	300	300	300	300		300
Null count	0	0	0	0	0	0	0	0		0
Blank count	0	0	0	0	53	0	0	0		0
Entirely uppercase count	300	0	300	0	0	300	0	0		75
Entirely lowercase count	0	300	0	0	0	0	0	0		75
Total char count	1800	10800	1583	5700	4693	1694	1791			600
Max chars	6	36	6	19	19	6	10			2
Min chars	6	36	5	19	0	4	2			2
Avg chars	6	36	5.277	19	15.643	5.647	5.97			2
Max white spaces	0	0	0	1	1	0	0			1
Min white spaces	0	0	0	1	0	0	0			0
Avg white spaces	0	0	0	1	0.823	0	0			0.027
Uppercase chars	900	0	683	0	0	1694	302			297
Uppercase chars (excl. first letters)	600	0	383	0	0	1394	2			147
Lowercase chars	0	3480	0	0	0	0	1487			295
Digit chars	900	6120	900	4200	3458	0	0			0
Diacritic chars	0	0	0	0	0	0	0			0
Non-letter chars	900	7320	900	5700	4693	0	2			8
Word count	300	300	300	600	494	300	300			300
Max words	1	1	1	2	2	1	1			1
Min words	1	1	1	2	0	1	1			1











sinchan/postgres@localhost

Query Editor Query History Scratch Pad

```
1 SELECT * FROM chapter4.device_event_log_fact
2 ORDER BY eventlogid ASC
```

Data Output Explain Messages Notifications

	eventlogid [PK] text	deviceid integer	eventid integer	hourid integer	monthid integer	quarterid integer	eventtimestamp timestamp without time zone	closurestatus boolean	closureduration bigint
1	INC100	217	5	18	2	1	2021-02-03 17:34:27	true	23770800000
2	INC101	206	1	7	2	1	2021-02-01 06:15:49	true	10983600000
3	INC102	228	7	22	4	2	2020-04-20 21:00:49	true	5976000000
4	INC103	272	4	6	2	1	2021-02-01 05:26:51	true	17204400000
5	INC104	286	1	11	10	4	2021-10-24 10:52:21	false	-1
6	INC105	255	7	3	8	3	2020-08-23 02:10:24	true	5544000000
7	INC106	203	1	20	6	2	2020-06-12 19:00:22	true	21362400000
8	INC107	218	3	16	3	1	2021-03-02 15:57:18	true	10551600000
9	INC108	276	5	9	2	1	2021-02-01 08:28:56	true	23252400000
10	INC109	272	4	10	8	3	2020-08-13 09:41:02	true	10562400000
11	INC110	227	2	2	1	1	2021-01-20 01:05:34	true	12276000000
12	INC111	265	3	15	4	2	2021-04-16 14:33:54	true	21362400000
13	INC112	250	3	22	7	3	2020-07-12 21:59:45	true	18856800000
14	INC113	207	1	20	7	3	2021-07-20 19:08:43	true	29822400000
15	INC114	277	2	17	6	2	2021-06-24 16:03:51	true	22140000000
16	INC115	300	6	8	10	4	2020-10-02 07:46:55	true	23169600000

sinchan/postgres@localhost

Query Editor Query History Scratch Pad

```
1 SELECT * FROM public.batch_job_execution
2 ORDER BY job_execution_id ASC
```

Data Output Explain Messages Notifications

	job_execution_id [PK] bigint	version bigint	job_instance_id bigint	create_time timestamp without time zone	start_time timestamp without time zone	end_time timestamp without time zone	status character varying (10)	exit_code character
1	57	2	52	2021-11-09 00:56:43.907	2021-11-09 00:56:43.927	2021-11-09 00:56:43.99	COMPLETED	COMPLE
2	58	2	53	2021-11-09 00:57:00.014	2021-11-09 00:57:00.022	2021-11-09 00:57:00.055	COMPLETED	COMPLE
3	59	2	54	2021-11-09 00:58:00.011	2021-11-09 00:58:00.017	2021-11-09 00:58:01.064	COMPLETED	COMPLE
4	60	2	55	2021-11-09 00:59:00.01	2021-11-09 00:59:00.014	2021-11-09 00:59:00.038	COMPLETED	COMPLE

sinchan/postgres@localhost

Query Editor Query History Scratch Pad

```
1 SELECT * FROM public.batch_step_execution
2 ORDER BY step_execution_id ASC
```

Data Output Explain Messages Notifications

	step_execution_id [PK] bigint	version bigint	step_name character varying (100)	job_execution_id bigint	start_time timestamp without time zone	end_time timestamp without time zone	status character varying (1)	commit_count bigint	read_count bigint
	100	3	fileCheck	57	2021-11-09 00:56:43.942	2021-11-09 00:56:43.987	COMPLETED		1
	101	3	fileCheck	58	2021-11-09 00:57:00.036	2021-11-09 00:57:00.051	COMPLETED		1
	102	3	fileCheck	59	2021-11-09 00:58:00.031	2021-11-09 00:58:00.046	COMPLETED		1
	103	3	fileMoveToProcess	59	2021-11-09 00:58:00.064	2021-11-09 00:58:00.085	COMPLETED		1
	104	303	processFile	59	2021-11-09 00:58:00.115	2021-11-09 00:58:01.056	COMPLETED	301	
	105	3	fileMoveToArchive	59	2021-11-09 00:58:01.06	2021-11-09 00:58:01.063	COMPLETED		1
	106	3	fileCheck	60	2021-11-09 00:59:00.025	2021-11-09 00:59:00.034	COMPLETED		1

# Chapter 5: Architecting a Batch Processing Pipeline

S3 Intelligent - Tiering

▼ S3 Intelligent-Tiering (S3 INT)

Info

The Amazon S3 Intelligent-Tiering storage class is designed to optimize storage costs by automatically moving data to the most cost-effective access tier when access patterns change. S3 Intelligent-Tiering automatically stores objects in three access tiers: one tier optimized for frequent access, a lower-cost tier optimized for infrequent access, and a very-low-cost tier optimized for rarely accessed data. For a small monthly object monitoring and automation charge, S3 Intelligent-Tiering moves objects that have not been accessed for 30 consecutive days to the Infrequent Access tier and after 90 days of no access they're moved to the Archive Instant Access tier. If the objects are accessed later, S3 Intelligent-Tiering moves the objects back to the Frequent Access tier. To save more on data that doesn't require immediate retrieval, you can activate the optional asynchronous Archive Access and Deep Archive Access tiers. There are no retrieval fees when using the S3 Intelligent-Tiering storage class, and no additional tiering fees when objects are moved between access tiers within S3 Intelligent-Tiering.

S3 INT storage

13

TB per month

Percentage of Storage in INT-Frequent Access Tier

Assumes no range GETs and that objects are only retrieved once

15

%

Percentage of Storage in INT-Infrequent Access Tier (% of storage that hasn't been accessed in the last 30 days)

Assumes no range GETs and that objects are only retrieved once

15

%

Percentage of Storage in INT-Archive Instant Access Tier (% of storage that hasn't been accessed in the last 90 days)

Assumes no range GETs and that objects are only retrieved once

%

Percentage of Storage in INT-Archive Access Tier (% of storage that hasn't been accessed for a minimum of 90 days)

Assumes no range GETs and that objects are only retrieved once

90

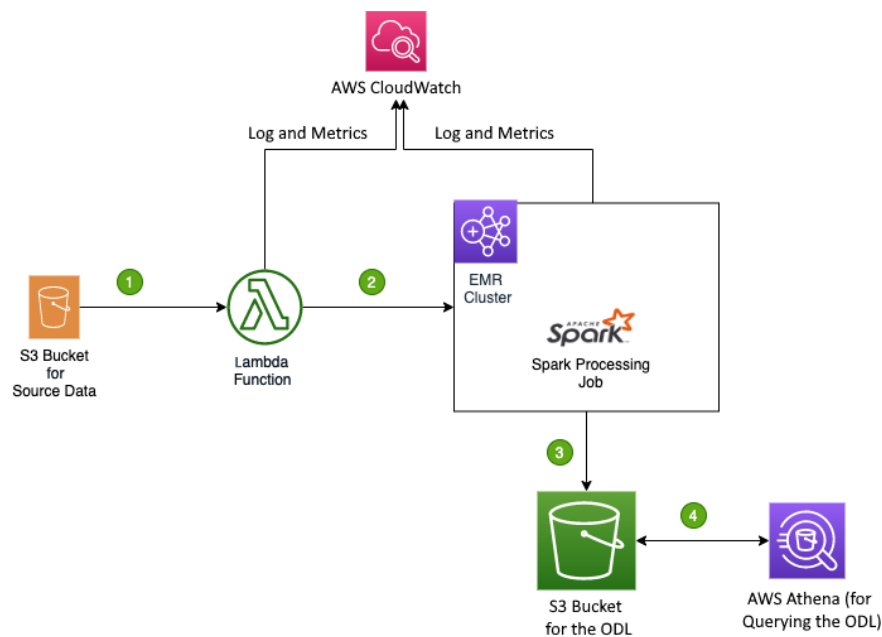
%

Percentage of Storage in INT-Deep Archive Access Tier (% of storage that hasn't been accessed for a minimum of 180 days)

Assumes no range GETs and that objects are only retrieved once

20

%



Files and folders (1 Total, 5.3 GB)

RemoveAdd filesAdd folder

All files and folders in this table will be uploaded.

Find by name
 < 1 >

<input type="checkbox"/>	Name	Folder	Type	Size
<input type="checkbox"/>	2019-Oct.csv	-	text/csv	5.3 GB

Destination

Destination  
[s3://scalabledataarch](#)

▶ Destination details  
 Bucket settings that impact new objects stored in the specified destination.

▶ Permissions  
 Grant public access and access to other AWS accounts.

▶ Properties  
 Specify storage class, encryption settings, tags, and more.

Cancel
 Upload

Connect to new dataset

File upload
 Data lake/data store
 Amazon S3
 Database connections
 Amazon Redshift
 JDBC
 AWS Glue Data Catalog
 Data Catalog S3 tables
 Data Catalog Redshift tables
 Data Catalog RDS tables
 All AWS Glue tables
 Others
 Amazon AppFlow
 AWS Data Exchange

Enter your source from S3  
 For you to select a folder, all files in the folder need to share the same file type. If there are different schemas, they will be merged.  

  
 Format is: s3://bucket/prefix

2019-Oct.csv is selected

S3 Buckets > scalabledataarch
 Select the entire folder
 ↻

Search S3 objects by name

< 1 > ⚙️

	Name	Size	Last updated
<input checked="" type="radio"/>	2019-Oct.csv	5.67 GB	December 4, 2021, 1:01:21 pm

DataBrew > Datasets > 2019-Oct

2019-Oct
 S3 2019-Oct.csv 5.7 GB
 Run data profile
 Create project wi

Dataset preview
 Data profile overview
 Column statistics
 Data quality rules
 Data lineage

Summary
 TOTAL ROWS: 10,000
 TOTAL COLUMNS: 10
 DATA SIZE: 10 GB

Correlations
 Correlation coefficient (r) defines how closely two variables are related. It ranges from -1.0 to +1.0, where 0 means there is no relationship between the variables.

Create data profile
 Examine and collect summaries of statistics about data in terms of structure, content, relationships,
 Run data profile

DATASETS

PROJECTS

RECIPES

DQ RULES

**JOBS**

WHAT'S NEW

DataBrew > Jobs > Create job

Create job

Job details

Job name

Identifier for the jobs

2019-Oct profile job

The job name must contain 1-240 characters. Valid characters are alphanumeric (A-Z, a-z, 0-9), hyphen (-), period (.), and space.

Job run sample

A job can be run on the entire dataset or a custom sample of the dataset.


Data sample

Define the scope of the dataset to run the job on

☒ Full dataset

☐ Custom sample

Job type

 Profile job

A profile job generates summary and statistics that give you the shape of your data.

DATASETS

PROJECTS

RECIPES

DQ RULES

**JOBS**

WHAT'S NEW

Associated dataset

2019-Oct

S3 | s3://scalabledataarch/2019-Oct.csv

Job output settings

Running a job generates output files at specified file destinations.

File type

Output format

JSON

S3 location

Format is: s3://bucket/folder/

s3://scalablearch-dataprof

Browse

Encryption

☐ Enable encryption for job output file

Encrypt the job output file using SSE-S3 or AWS KMS

Dataset level configurations

Choose the dataset-level statistics that you want to include in your generated data profile.

Default

Duplicate values

Enabled

Advanced statistics

PII statistics

Disabled

**Correlations widget**

Enabled, custom columns

Correlations widget

High impact

☒ Enable correlations matrix

Matrix visualization of correlation coefficient (r). Correlation coefficient (r) defines how closely two variables are related. It ranges from -1. relationship between the variables.

Correlation algorithm

Algorithm used to identify correlations

Pearsons correlation

Selected columns

Select the columns to calculate correlations for

☐ All columns

☐ First 10 numeric columns

☒ Custom columns

☒ Column name

☐ RegEx match

Enter column name

Add

You can add a maximum of 25 column names or regular expressions

Name: product\_id X

Name: category\_id X

Name: brand X

### ► Associated schedules - *optional*

You can associate up to 2 schedules to automate your job.

### ► Tags - *optional*

Metadata that you can define and assign to AWS resources. Each tag is a simple label consisting of a customer-defined key (name) and an optional value. Using tags can make it easier for you to organize resources by purpose, owner, environment, or other criteria.

## Permissions

DataBrew needs permission to connect to data on your behalf. Use an IAM role with the [required policy](#) attached.

### Role name

Choose the role that has access to connect to your data. Refresh to see the latest updates.

Create new IAM role



### New IAM role suffix

Your role will be prefixed with "AWSGlueDataBrewServiceRole-"

datapipeline

By clicking "Create job" you are authorizing creation of this role.

DATASETS

PROJECTS

RECIPES

Q RULES

**JOBS**

DataBrew > Jobs

Recipe jobs

**Profile jobs**

Schedules

Profile jobs (1)

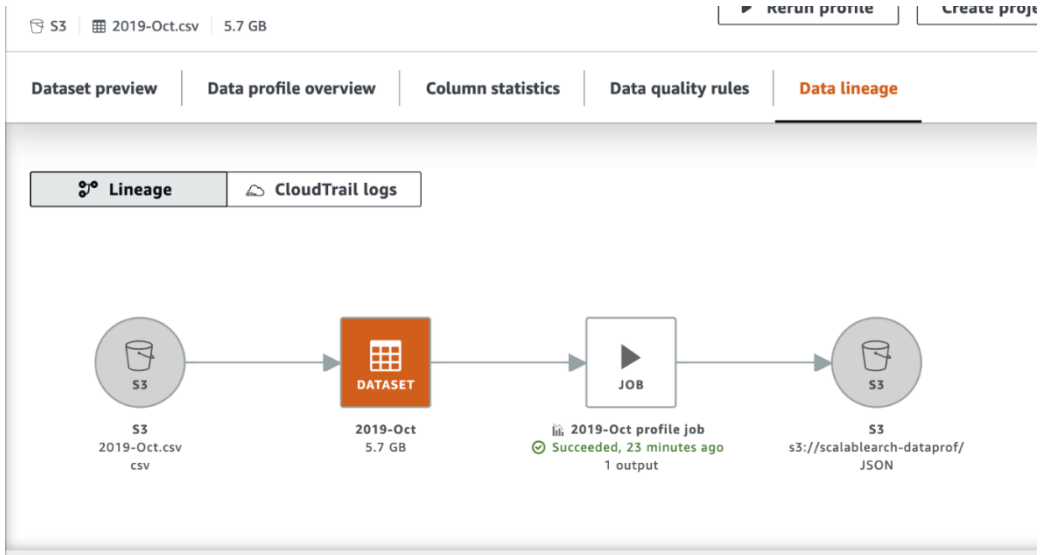
Find jobs

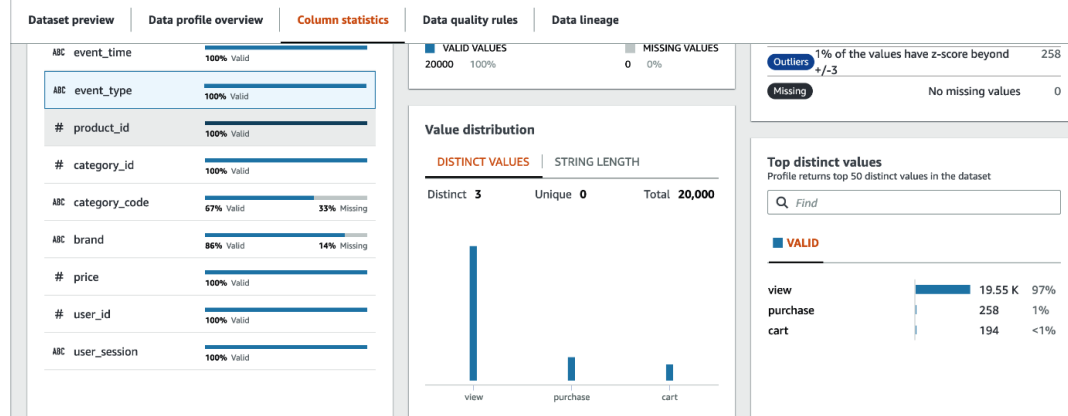
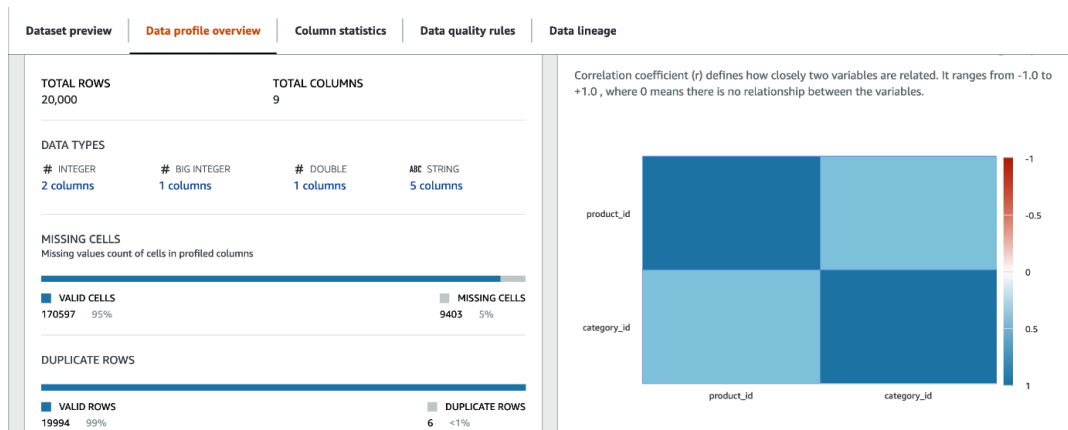
Show all

< 1 >

⌂

<input checked="" type="checkbox"/>	Job name	Last job run status	Dataset	Data profile	Last run	Created on	Created by	Tags
<input checked="" type="checkbox"/>	2019-Oct profile job	-	2019-Oct	<a href="#">View data profile</a>	-	2 minutes ago December 4, 2021, 9:26:57 am		





## Create Cluster - Quick Options [Go to advanced options](#)

### General Configuration

Cluster name

☒ Logging ⓘ

S3 folder  ⓘ

Launch mode ☐ Cluster ⓘ ☒ Step execution ⓘ

### Add steps

A step is a unit of work submitted to an application running on your EMR cluster. EMR programmatically installs the applications needed to execute the added steps. [Learn more](#) ⓘ

Step type

### Software configuration

Release  ⓘ

Applications Hadoop 3.2.1 ⓘ

### Hardware configuration

Instance type  ⓘ

The selected instance type adds 64 GiB of GP2 EBS storage per instance by default. [Learn more](#) ⓘ

Add step

Step type

Spark application

Run Spark application using spark-submit. [Learn more](#)

Name

Spark application

Deploy mode

Cluster

Run your driver on a slave node (cluster mode) or on the master node as an external client (client mode).

Spark-submit options

```
--class
com.scalabledataarchitecture.bigdata.EcomAnalysisDriver
```

Specify other options for spark-submit.

Application location\*

s3://jarandconfigs/EcommerceAnalysis-1.0-SNAPSHOT.jar

Path to a JAR with your application and dependencies (client deploy mode only supports a local path).

Arguments

s3://chapter5input/2019-Oct.csv
s3://scalabledataarch/output

Specify optional arguments for your application.

Action on failure

Terminate cluster

What happens if the step fails

Cancel

Add

## General Configuration

Cluster name

ManualCluster

☒ Logging

?

S3 folder

s3://aws-logs-627443126298-us-east-2/elasticmapreduce-logs-2019-10-07-15-00-00

Launch mode

☐ Cluster

?

☒ Step execution

?

## Add steps

A step is a unit of work submitted to an application running on your EMR cluster. EMR programmatically installs the applications needed to execute the added steps. [Learn more](#)

Name	Action on failure	JAR location	Arguments
Spark application	Terminate cluster	command-runner.jar	spark-submit --deploy-mode cluster --class com.scalabledataarchitecture.bigdata.EcomAnalysisDriver s3://jarandconfigs/EcommerceAnalysis-1.0-SNAPSHOT.jar s3://chapter5input/2019-Oct.csv s3://scalabledataarch/output

Step type

Spark application

Configure

- Summary
- Application user interfaces
- Monitoring
- Hardware
- Configurations
- Events
- Steps
- Bootstrap actions

Concurrency: 1 [Change](#)

After last step completes: Cluster auto-terminates

Add step

Clone step


Cancel step




View

Filter: 

All steps

Filter steps ...



 2 steps (all loaded) 

	ID	Name	Status	Start time (UTC-5) 	Elapsed time	Log files
 	s-2020P85RZS6TN	Spark application	Completed	2021-12-07 01:22 (UTC-5)	8 minutes	<a href="#">View logs</a>



## Summary

Delete role




Role ARN	arn:aws:iam::627443126298:role/lambda2emrtrigger_role 
Role description	Allows Lambda functions to call AWS services on your behalf.   <a href="#">Edit</a>
Instance Profile ARNs	
Path	/
Creation time	2021-12-07 01:35 EST
Last activity	Not accessed in the tracking period
Maximum session duration	1 hour <a href="#">Edit</a>

Permissions Trust relationships Tags (1) Access Advisor Revoke sessions

## ▼ Permissions policies (3 policies applied)

Attach policies

[+ Add inline policy](#)

Policy name ▼	Policy type ▼	
▶  AmazonS3FullAccess	AWS managed policy	✕
▶  AmazonElasticMapReduceforEC2Role	AWS managed policy	✕
▶  AmazonElasticMapReduceFullAccess	AWS managed policy	✕

Create function [Info](#)

Choose one of the following options to create your function.

**Author from scratch** ☒

Start with a simple Hello World example.

**Use a blueprint** ☐

Build a Lambda application from sample code and configuration presets for common use cases.

**Container image** ☐

Select a container image to deploy for your function.

**Browse serverless app repository** ☐

Deploy a sample Lambda application from the AWS Serverless Application Repository.

## Basic information

**Function name**  
Enter a name that describes the purpose of your function.

s3-lambda-trigger-emr

Use only letters, numbers, hyphens, or underscores with no spaces.

**Runtime** [Info](#)  
Choose the language to use to write your function. Note that the console code editor supports only Node.js, Python, and Ruby.

Java 11 (Corretto) ▼

**Architecture** [Info](#)  
Choose the instruction set architecture you want for your function code.

☒ x86\_64  
☐ arm64

## ▼ Change default execution role

**Execution role**  
Choose a role that defines the permissions of your function. To create a custom role, go to the [IAM console](#).

☐ Create a new role with basic Lambda permissions  
☒ Use an existing role  
☐ Create a new role from AWS policy templates

**Existing role**  
Choose an existing role that you've created to be used with this Lambda function. The role must have permission to upload logs to Amazon CloudWatch Logs.

lambda2emrtrigger\_role ▼ 

[View the lambda2emrtrigger\\_role role on the IAM console.](#)

## ► Advanced settings

Cancel

Create function

Lambda > Add trigger

Add trigger

Trigger configuration

S3

aws storage

Bucket

Please select the S3 bucket that serves as the event source. The bucket must be in the same region as the function.

chapter5input

Event type

Select the events that you want to have trigger the Lambda function. You can optionally set up a prefix or suffix for an event. However, for each bucket, individual events cannot have multiple configurations with overlapping prefixes or suffixes that could match the same object key.

All object create events

Prefix - optional

Enter a single optional prefix to limit the notifications to objects with keys that start with matching characters.

e.g. images/

Suffix - optional

Enter a single optional suffix to limit the notifications to objects with keys that end with matching characters.

e.g. .jpg

Successfully updated the function s3-lambda-trigger-emr.

S3

+ Add trigger

Layers

(0)

+ Add destination

Last modified

1 hour ago

Function ARN

arn:aws:lambda:us-east-2:627443126298:function:s3-lambda-trigger-emr

Code

Test

Monitor

Configuration

Aliases

Versions

Code source

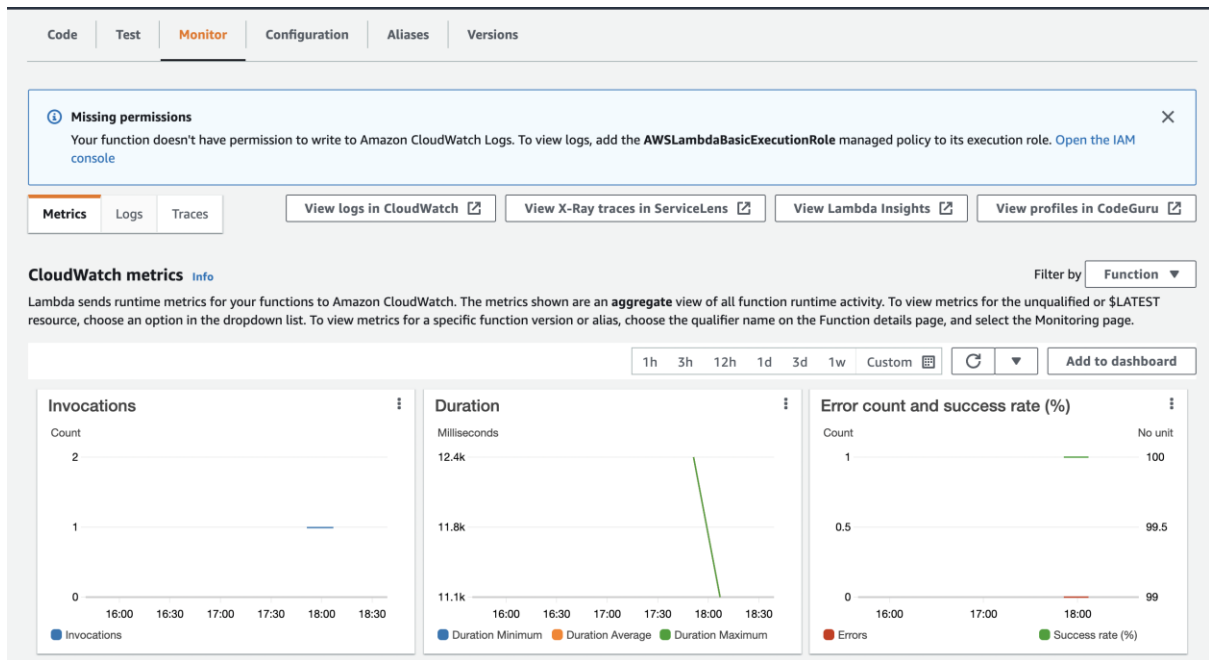
Info

Upload from

.zip or .jar file

Amazon S3 location

The deployment package of your Lambda function "s3-lambda-trigger-emr" is too large to enable inline code editing. However, you can still invoke your function...



Search for services, features, blogs, docs, and more [Option+S]

Clone Terminate AWS CLI export

**Cluster: chap5\_test\_auto** **Running** Running step

Summary | Application user interfaces | **Monitoring** | Hardware | Configurations | Events | Steps | Bootstrap actions

**Summary**

ID: j-34XX7EVW20X7R

Creation date: 2021-12-30 21:21 (UTC-5)

Elapsed time: 4 minutes

After last step completes: Cluster auto-terminates

Termination protection: Off [Change](#)

Tags: -- [View All](#) / [Edit](#)

Master public DNS: ec2-3-15-144-150.us-east-2.compute.amazonaws.com [Connect to the Master Node Using SSH](#)

**Configuration details**

Release label: emr-6.4.0

Hadoop distribution: Amazon

Applications: Spark 3.1.2

Log URI: s3://aws-logs-627443126298-us-east-2/elasticmapreduce/ [View](#)

EMRFS consistent view: Disabled

Custom AMI ID: --

**Application user interfaces**

Persistent user interfaces [View](#): [Spark history server](#), YARN timeline server

On-cluster user interfaces [View](#): Not Enabled [Enable an SSH Connection](#)

**Network and hardware**

Availability zone: us-east-2a

Subnet ID: [subnet-e927d782](#)

Master: **Running** 1 m5.xlarge

Core: **Running** 2 m5.xlarge

Status: SUCCEEDED

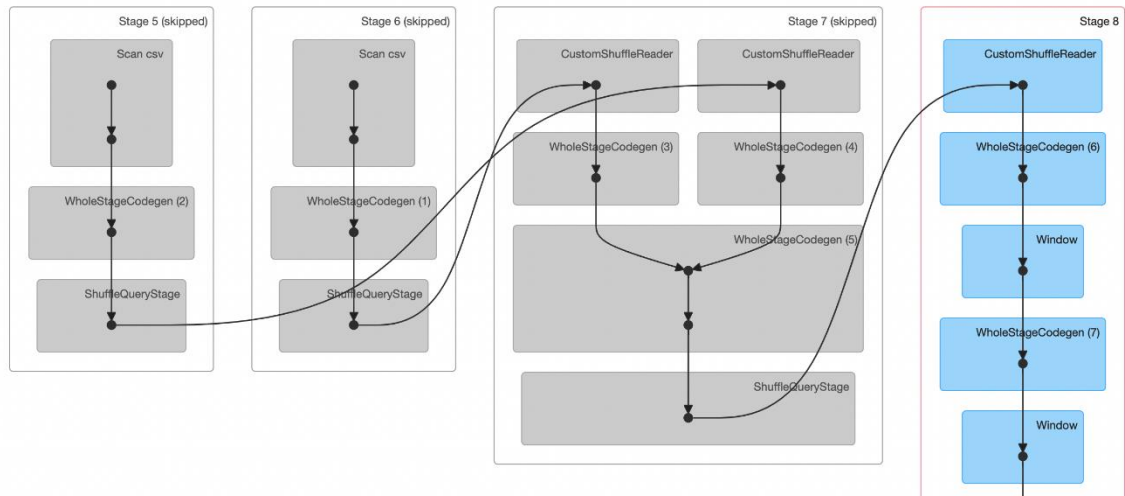
Associated SQL Query: 4

Completed Stages: 1

Skipped Stages: 3

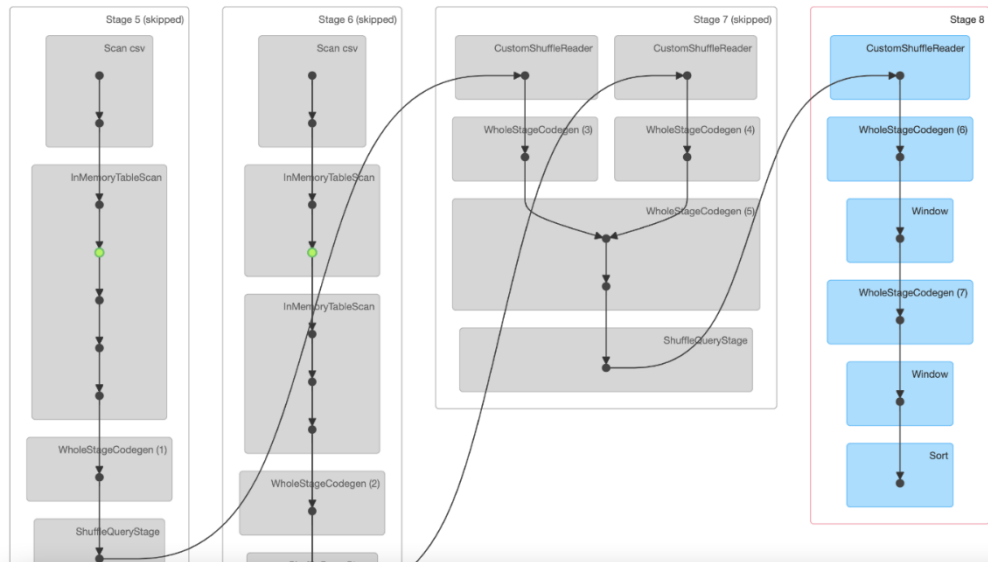
▶ Event Timeline

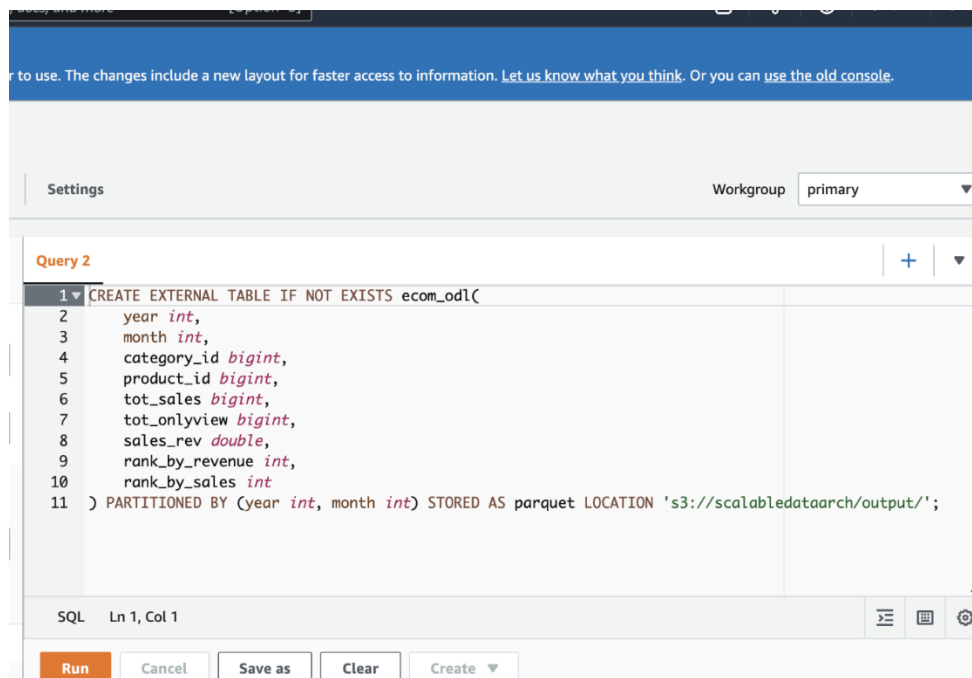
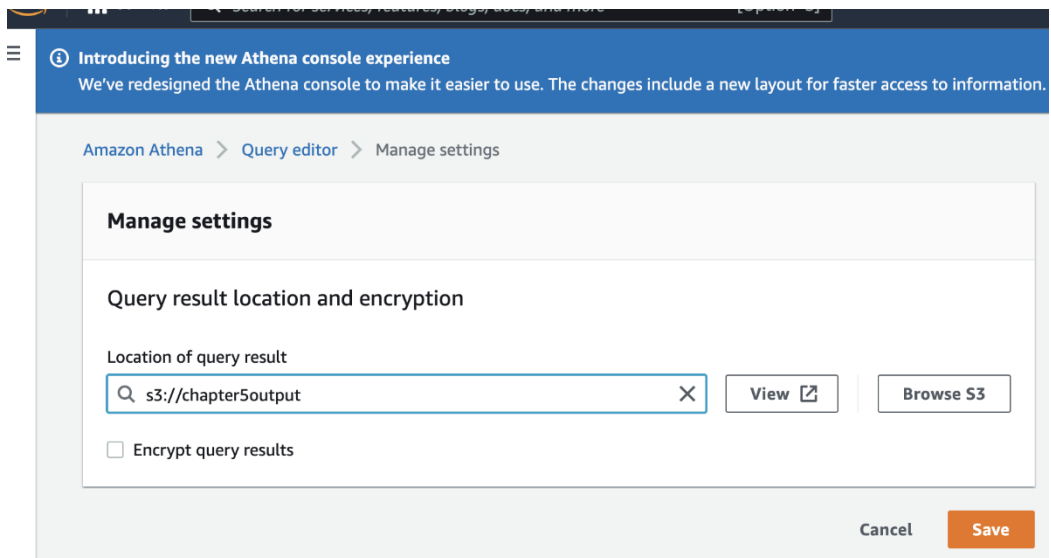
▼ DAG Visualization



▶ Event Timeline

▼ DAG Visualization





Query 2

1

```
select year,month,category_id,product_id,rank_by_revenue from ecom_odl where rank_by_revenue <=3 and year=2019 and month=10 order by category_id,rank_by_revenue;
```

SQLLn 1, Col 125

Run again

Cancel

Save as

Clear

Create

Completed

Time in queue: 0.299 sec

Run time: 0.832 sec

Data scanned: 891.08 KB

Results (100+)

Copy

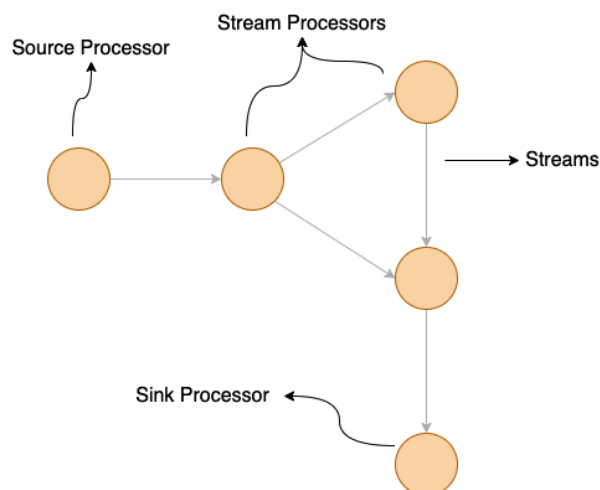
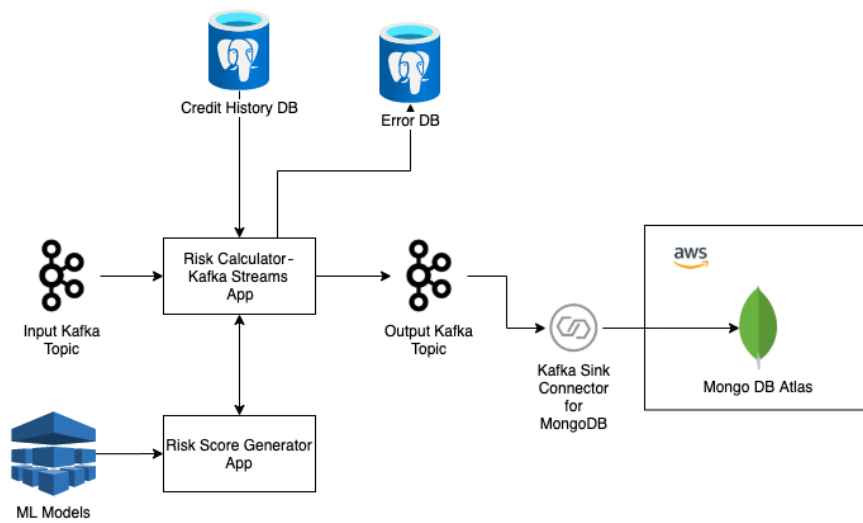
Download results

Search rows

< 1 ... >

year	month	category_id	product_id	rank_by_revenue
2019	10	2053013552226107603	8600067	1
2019	10	2053013552226107603	8600175	2
2019	10	2053013552226107603	8600140	3
2019	10	2053013552259662037	8500140	1
2019	10	2053013552259662037	8500139	2
2019	10	2053013552259662037	8500088	3

## Chapter 6: Architecting a Real-Time Processing Pipeline



```
import static org.hamcrest.MatcherAssert.assertThat;
import static org.hamcrest.Matchers.equalTo;
import static org.hamcrest.Matchers.is;
import static org.junit.Assert.assertEquals;

@RunWith(MockitoJUnitRunner.class)
public class CreditRiskCalculatorTests {

    private final Properties config = new Properties();

    public CreditRiskCalculator() {
        config.setProperty(StreamConfig.APPLICATION_ID_CONFIG, "CreditRiskCalculator");
        config.setProperty(StreamConfig.BOOTSTRAP_SERVERS_CONFIG, "localhost:9092");
        config.setProperty(StreamConfig.ZOOKEEPER_CONNECT_CONFIG, "localhost:2181");
        config.setProperty(StreamConfig.DEFAULT_KEY_SERIALIZER_CONFIG, "org.apache.kafka.common.serialization.StringSerializer");
        config.setProperty(StreamConfig.DEFAULT_VALUE_SERIALIZER_CONFIG, "org.apache.kafka.common.serialization.StringSerializer");
    }

    @Mock
    JdbcTemplate jdbcTemplate;

    @Mock
    RestTemplate restTemplate;

    // ... other test methods ...

}
```

The image shows a code editor with a context menu open over the `CreditRiskCalculatorTests` class. The menu options include:

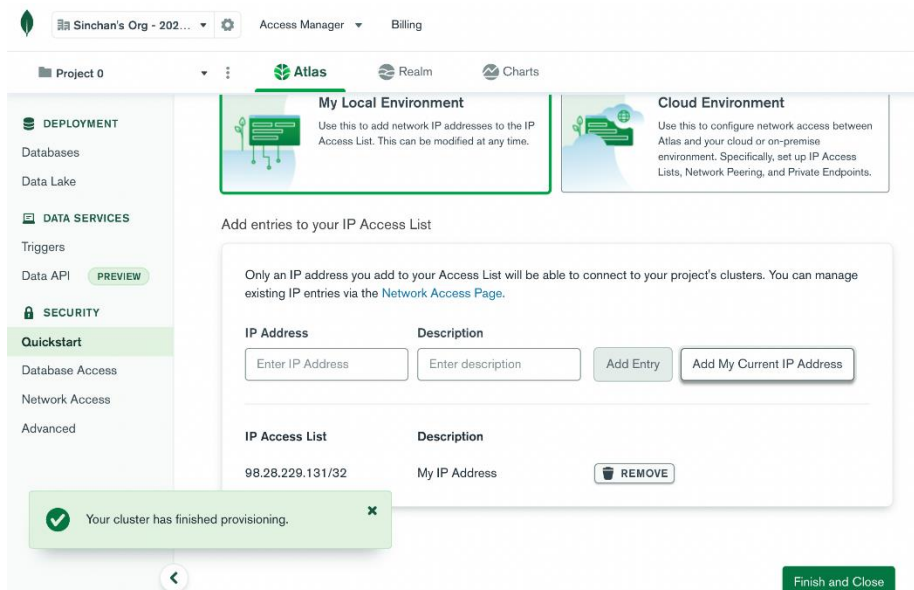
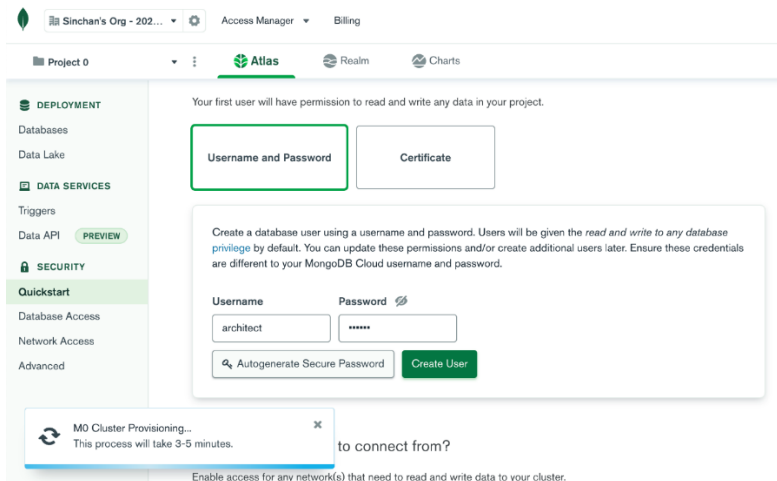
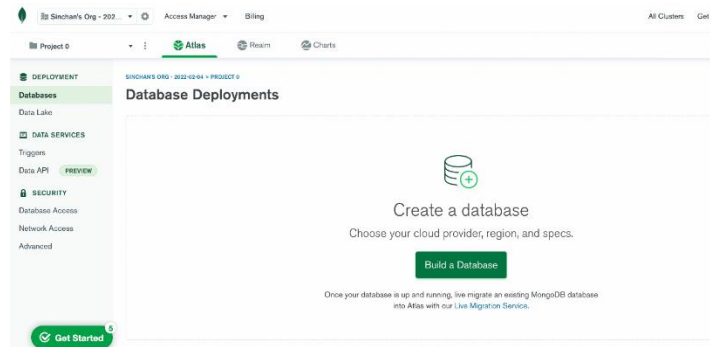
- Show Context Actions
- Paste
- Copy / Paste Special
- Column Selection Mode
- Find Usages
- Refactor
- Folding
- Analyze
- Go To
- Generate...
- Run 'CreditRiskCalculatorTests'
- Debug 'CreditRiskCalculatorTests'
- Run 'CreditRiskCalculatorTests' with Coverage
- Modify Run Configuration...
- Open In
- Local History
- Compare with Clipboard
- Create Gist

```
42 config.setProperty(StreamsConfig.BOOTSTRAP_SERVERS_CONFIG, "test:1234");
43 config.setProperty(StreamsConfig.DEFAULT_KEY_SERDE_CLASS_CONFIG, Serdes.String().getClass().getName());
44 config.setProperty(StreamsConfig.DEFAULT_VALUE_SERDE_CLASS_CONFIG, Serdes.String().getClass().getName());

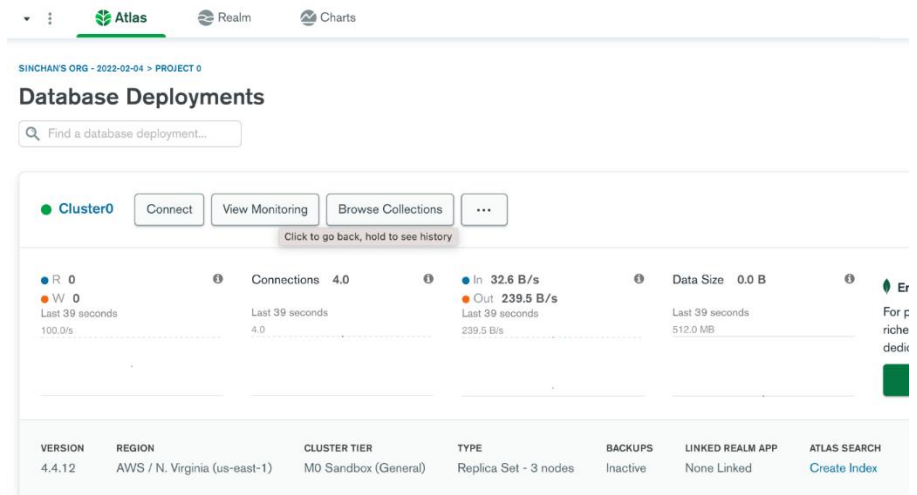
Run: CreditRiskCalculatorTests
Tests passed: 1 of 1 test - 1sec 41ms

CreditRiskCalculatorTests 1sec 41ms
creditRiskStreamsTest 1sec 41ms

/opt/homebrew/Cellar/openjdk@11/11.0.10/libexec/openjdk.jdk/Contents/Home/bin/java ...
17:46:39.869 [main] DEBUG org.apache.kafka.streams.kstream.internals.InternalStreamsBuilder - Adding nodes to topology
StreamsGraphNode{nodeName='root', buildPriority=null, hasWrittenToTopology=false, keyChangingOperation=false,
valueChangingOperation=false, mergeNode=false, parentNodes=[]} child nodes [StreamSourceNode{topicNames=[testInputTopic],
topicPattern=null, consumedInternal=org.apache.kafka.streams.kstream.internals.ConsumedInternal@e1781}
StreamsGraphNode{nodeName='KSTREAM-SOURCE-0000000000', buildPriority=0, hasWrittenToTopology=false, keyChangingOperation=false,
valueChangingOperation=false, mergeNode=false, parentNodes=[root]}]
17:46:39.871 [main] DEBUG org.apache.kafka.streams.kstream.internals.InternalStreamsBuilder - Adding nodes to topology
StreamSourceNode{topicNames=[testInputTopic], topicPattern=null, consumedInternal=org.apache.kafka.streams.kstream.internals
```







Atlas Realm Charts

Overview Real Time Metrics **Collections** Search Profiler Performance Advisor Online Archive

DATABASES: 0 COLLECTIONS: 0 VIEW VISI

## Explore Your Data

- **Find:** run queries and interact with documents
- **Indexes:** build and manage indexes
- **Aggregation:** test aggregation pipelines
- **Search:** build search indexes

Load a Sample Dataset Add My Own Data

[Learn more in Docs and Tutorials](#)

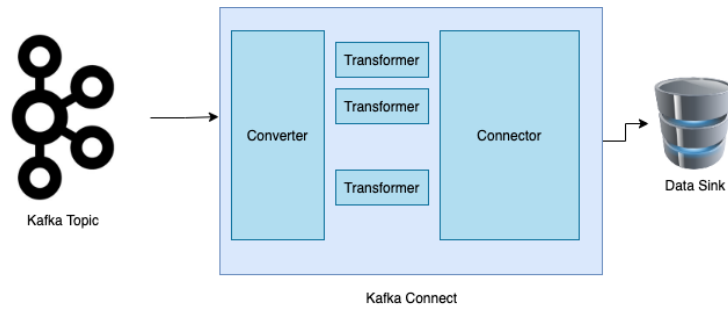
</

<

>

Documents

<



```

kafka_2.12-2.8.1 — java -Xmx512M -server -XX:+UseG1GC -XX:MaxGCPauseMillis=20 -XX:Initi...
...skCalculator — -zsh ... ...ookeeper.properties ...server.properties ...r localhost:9092 +

sinchan@Sinchans-MacBook-Pro-2 kafka_2.12-2.8.1 % bin/kafka-console-producer.
sh --topic landingTopic1 --bootstrap-server localhost:9092
>{"id":"5008804","applicationId":"CT20210809","flagOwnRealty":"Y","nameEduc
ationType":"Higher education","flagMobil":1,"flagWorkPhone":1,"nameIncomeType":
"Working","nameFamilyStatus":"Civil marriage","nameHousingType":"Rented apart
ment","cntChildren":0,"occupationType":"","genderCode":"M","flagEmail":0,"amt
IncomeTotal":427500.0,"daysBirth":-12005,"daysEmployed":-4542,"flagPhone":0,"
cntFamMembers":2,"flagOwnCar":"Y"}
  
```

```

...uorum.QuorumPeerMain config/zookeeper.properties ... ../libs/connect-kafka.Kafka config/server.properties ~/Documents/kafka_2.12-2.8.1 — -zsh ...c landingTopic1 --bootstrap-se

sinchan@Sinchans-MacBook-Pro-2 kafka_2.12-2.8.1 % bin/kafka-console-producer.sh --topic landingTopic1 --bootstrap-server localhost:9092
>{"id": 5008804, "genderCode": "M", "flagOwnCar": "Y", "flagOwnRealty": "Y", "cntChildren": 0, "amtIncomeTotal": 427500, "nameIncomeType": "Higher e
ducationType": "Higher education", "flagMobil": 1, "flagWorkPhone": 1, "flagPhone": 0, "flagEmail": 0, "occi
amMembers": 2 }
  
```

Project 0 Atlas Realm Charts

DEPLOYMENT

Databases

Data Lake

DATA SERVICES

Triggers

Data API PREVIEW

SECURITY

Database Access

Network Access

Advanced

CRRD

newloanrequest

CRRDB

Find Indexes Schema Anti-Patterns Aggregation Search Indexes

INSEI

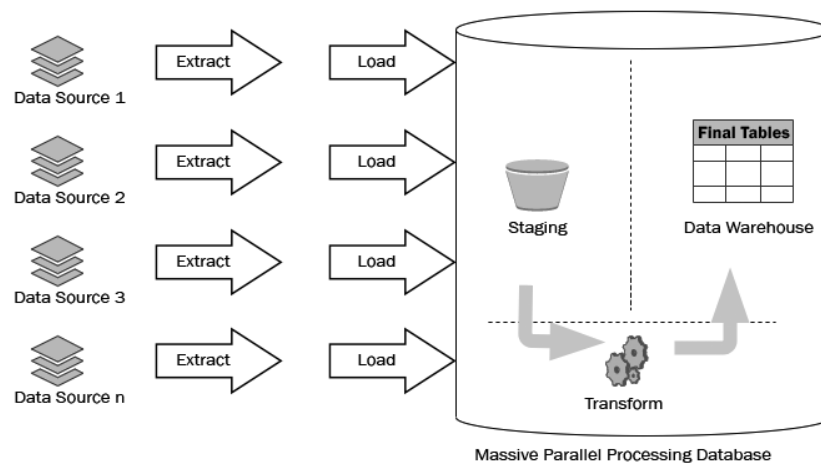
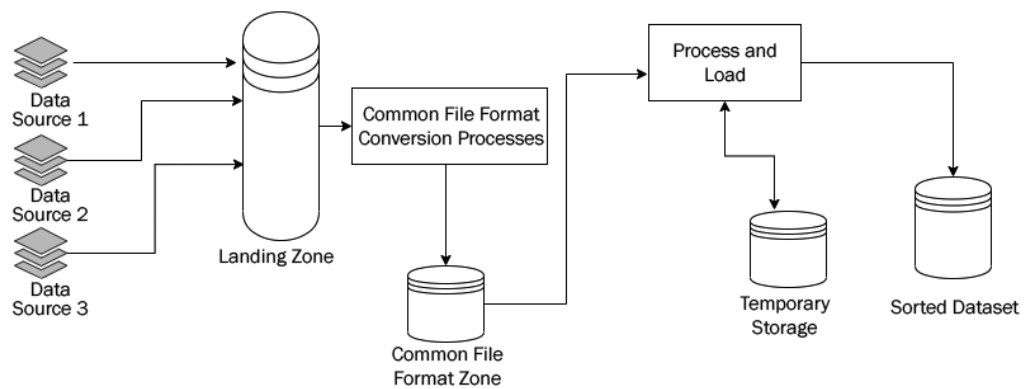
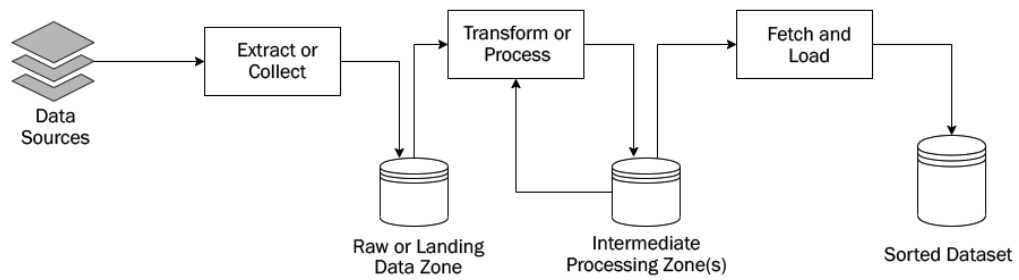
FILTER { field: 'value' } OPTIONS App

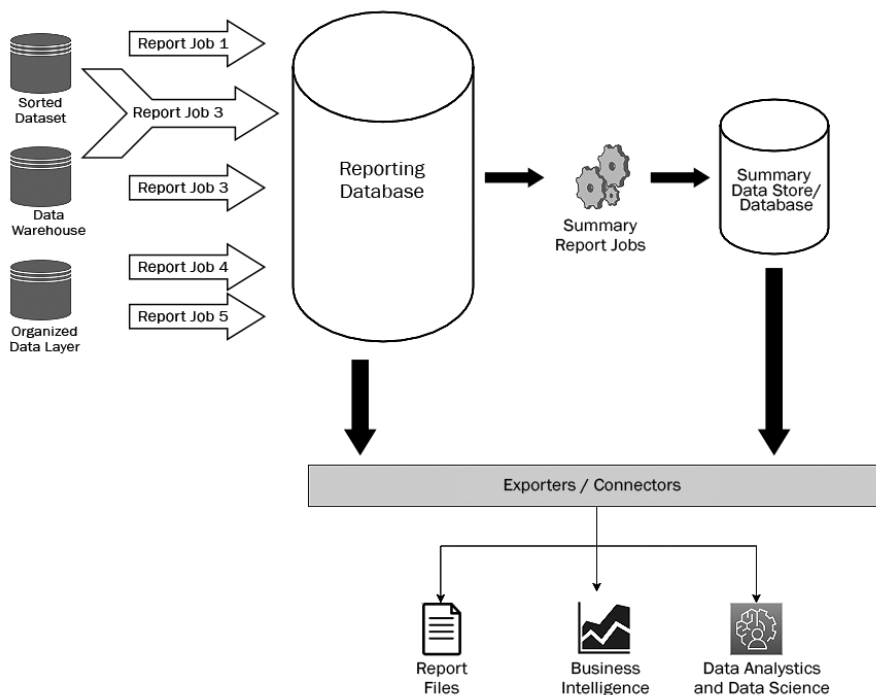
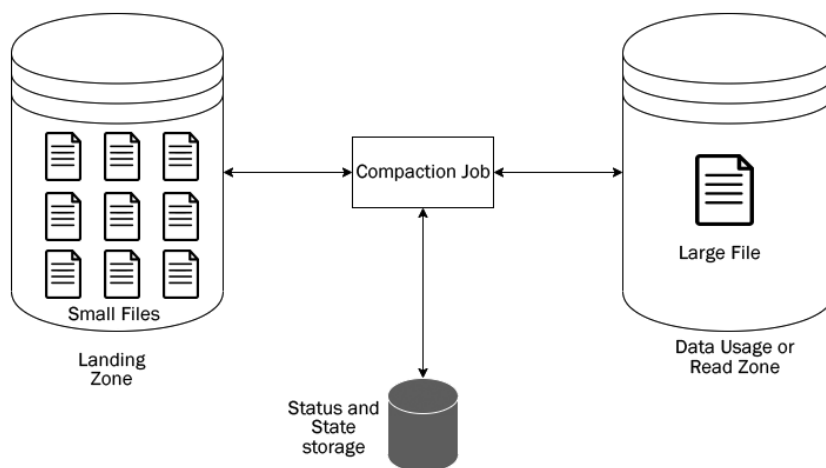
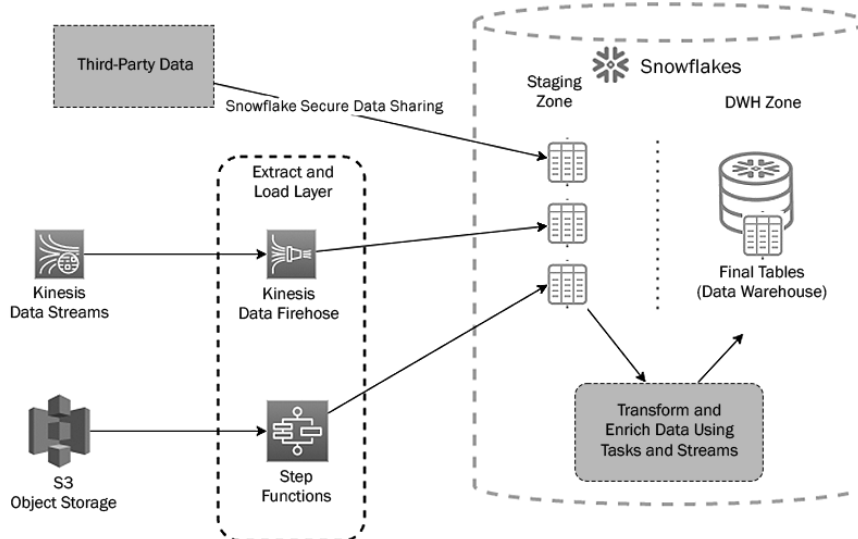
QUERY RESULTS 1-1 OF 1

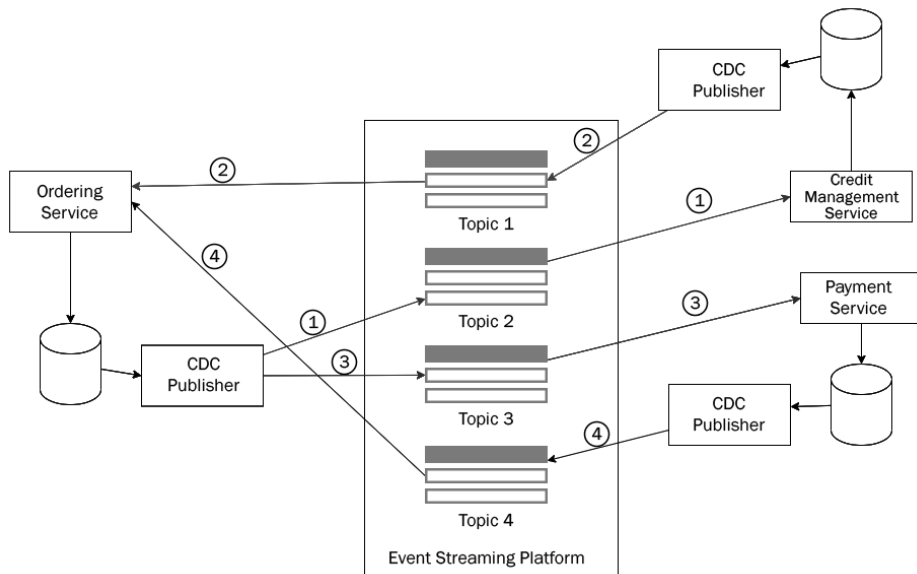
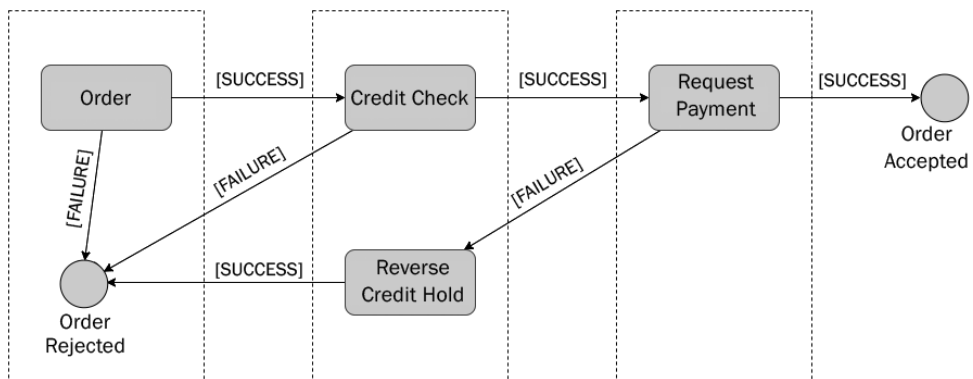
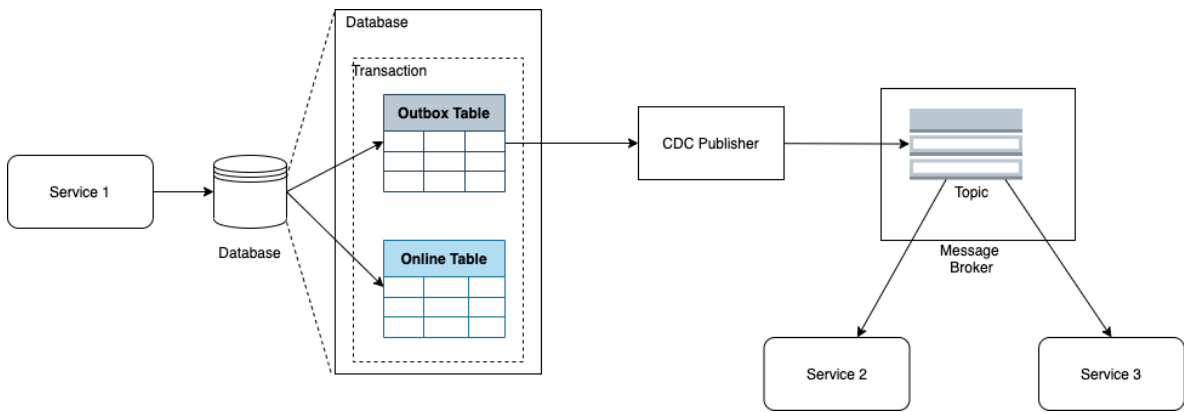
```

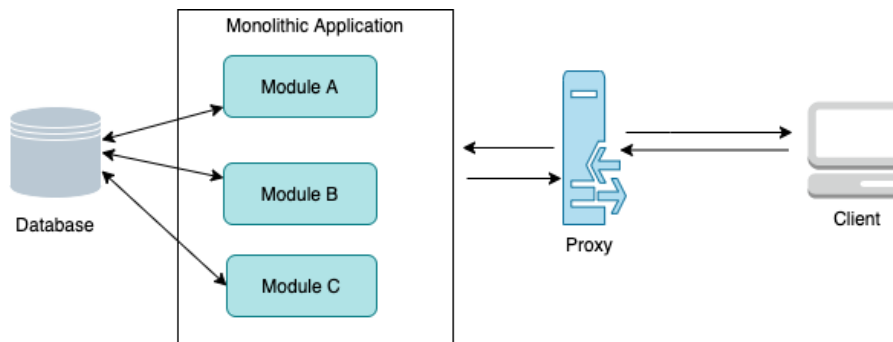
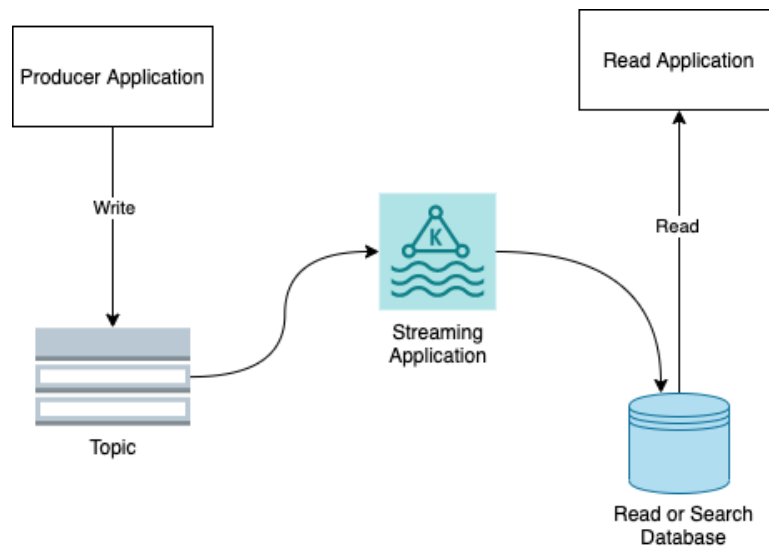
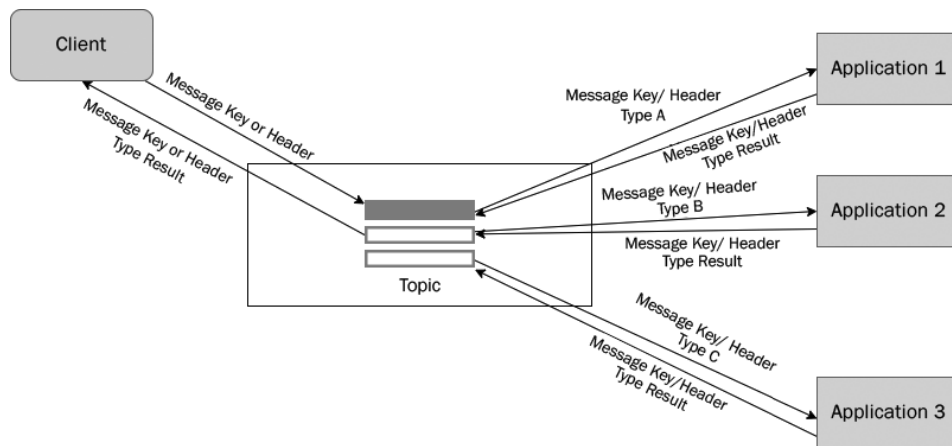
_id: ObjectId("620079e2ff213600e29c930a")
flagOwnRealty: "Y"
nameEducationType: "Higher education"
flagMobil: 1
flagWorkPhone: 1
nameIncomeType: "Working"
nameFamilyStatus: "Civil marriage"
nameHousingType: "Rented apartment"
cntChildren: 0
occupationType: ""
genderCode: "M"
flagEmail: 0
amtIncomeTotal: 427500
daysBirth: -12005
daysEmployed: -4542
flagPhone: 0
id: "5008804"
riskScore: 2
cntFamMembers: 2
flagOwnCar: "Y"
  
```

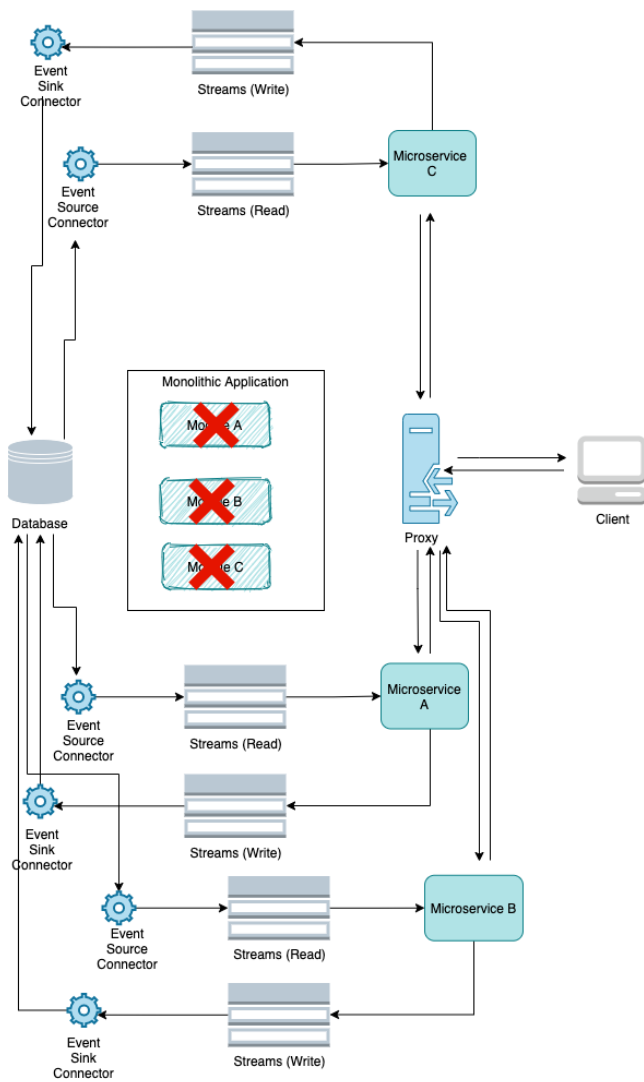
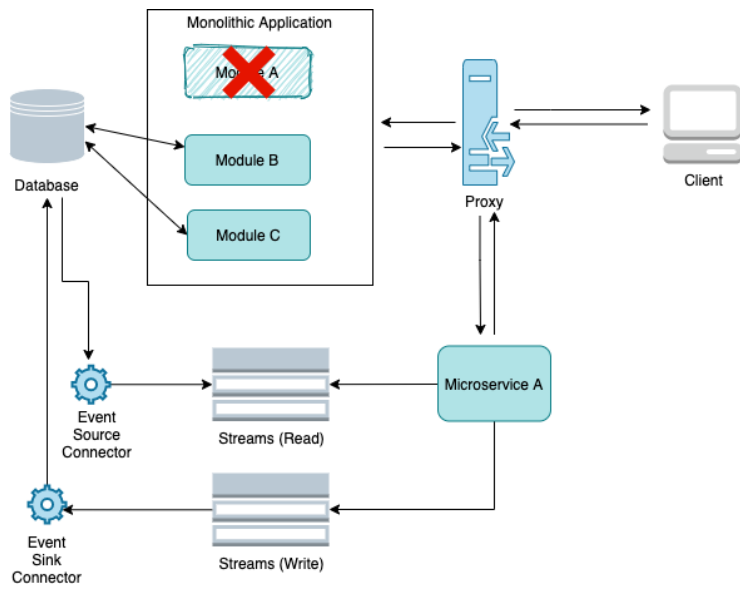
## Chapter 7: Core Architectural Design Patterns

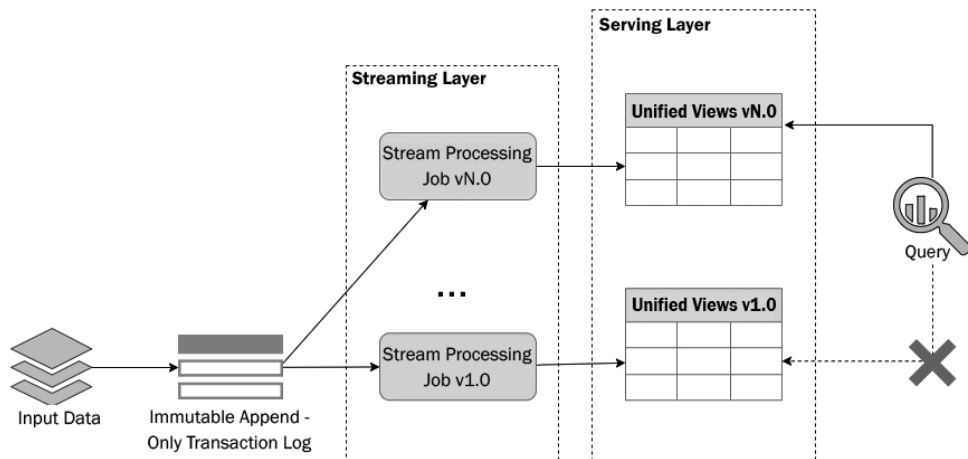
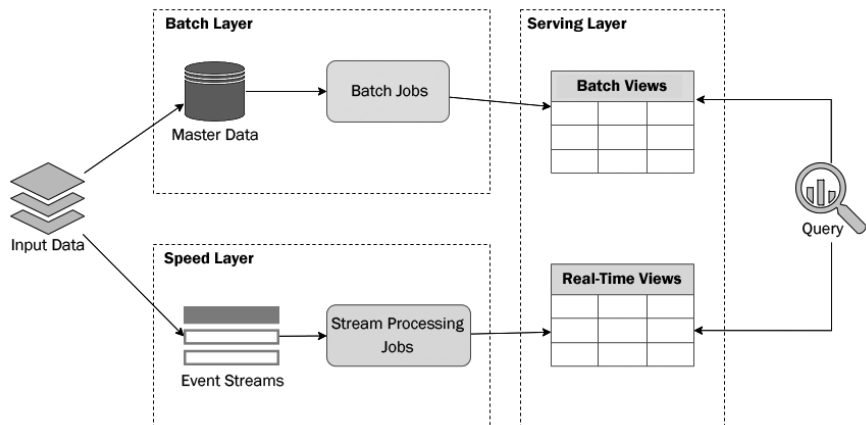
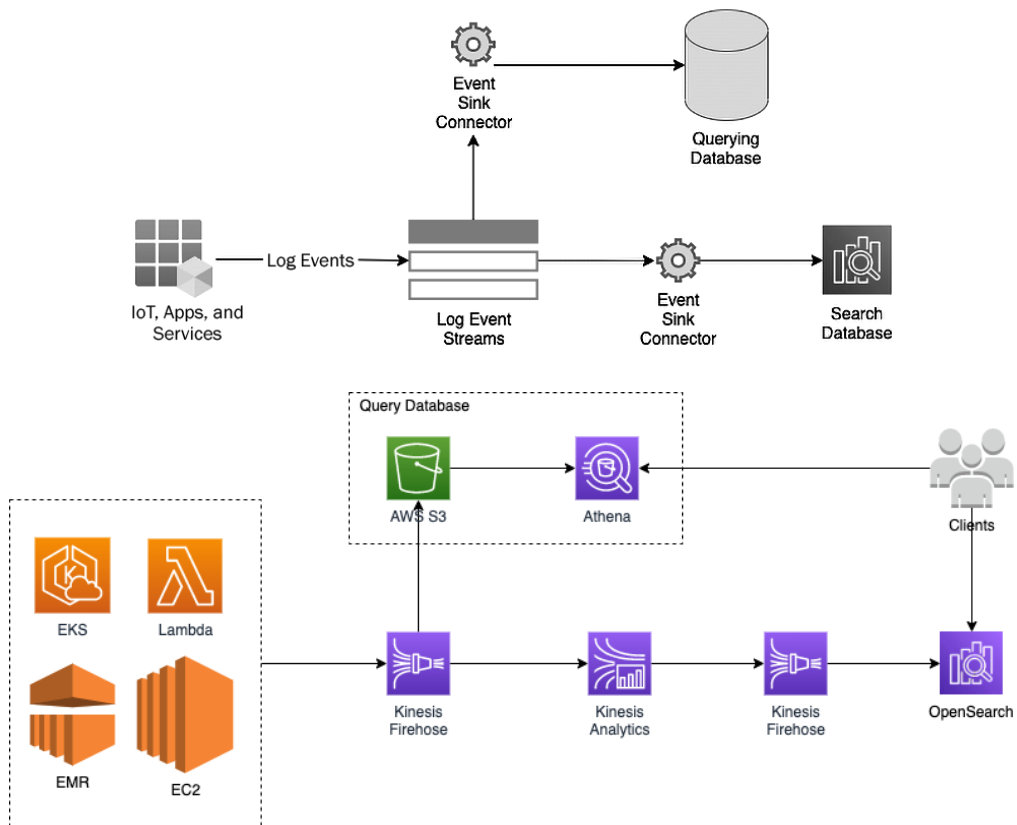




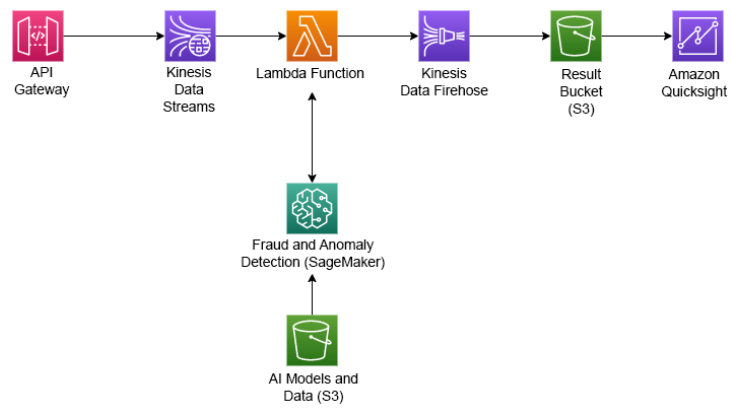
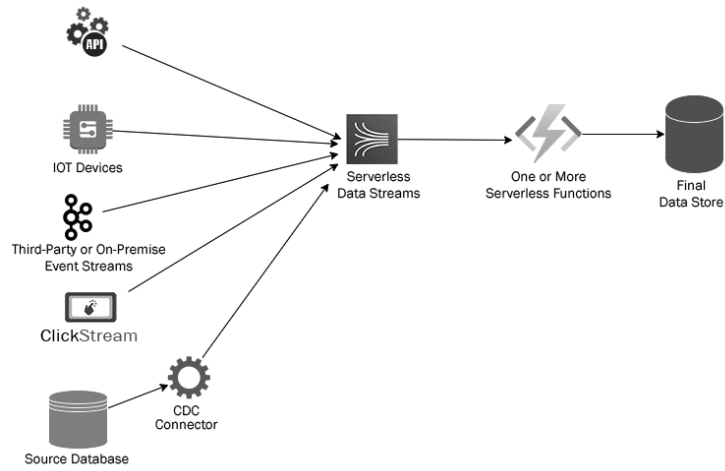
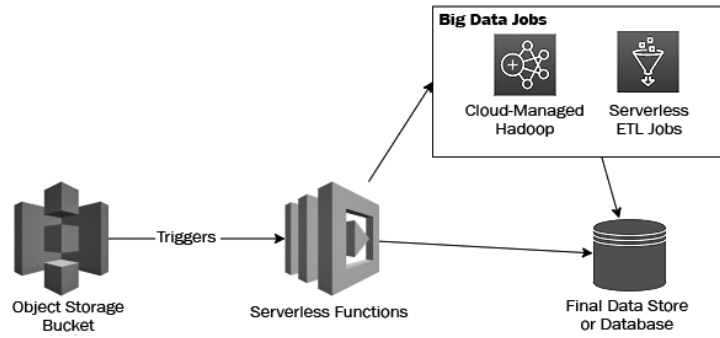




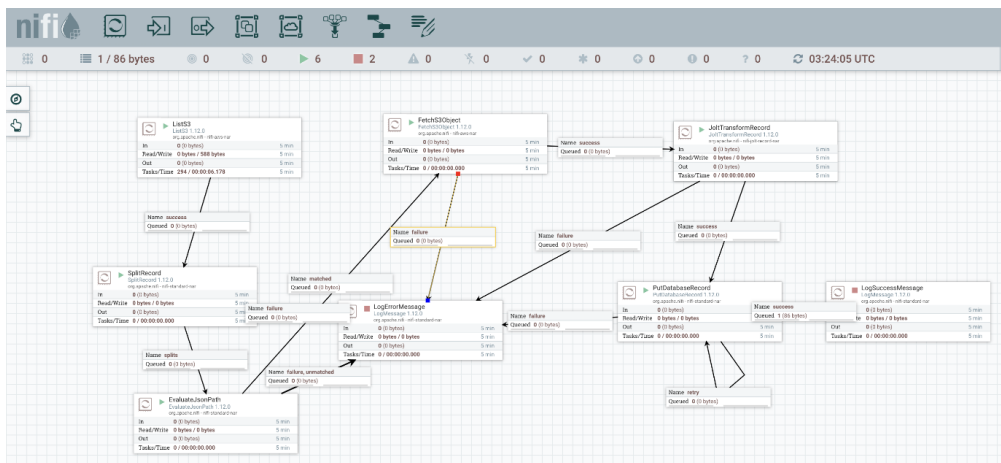
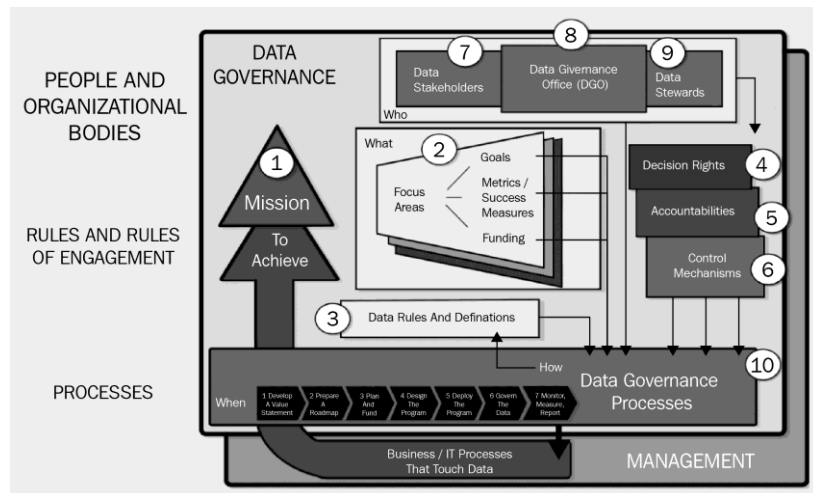








# Chapter 8: Enabling Data Security and Governance



## Processor Details

Running

STOP & CONFIGURE

SETTINGS

SCHEDULING

PROPERTIES

COMMENTS

Required field

Property	Value
Bucket	chapter8input
Region	US East (Ohio)
Access Key ID	Sensitive value set
Secret Access Key	Sensitive value set
Record Writer	JsonRecordSetWriter
Minimum Object Age	0 sec
Write Object Tags	False
Write User Metadata	False
Credentials File	No value set
AWS Credentials Provider service	No value set
Communications Timeout	30 secs
SSL Context Service	No value set

OK

## Processor Details

▶ Running

⚙ STOP & CONFIGURE

SETTINGS

SCHEDULING

PROPERTIES

COMMENTS

Required field

Property		Value	
Record Reader	?	JsonTreeReader	→
Record Writer	?	JsonRecordSetWriter	→
Records Per Split	?	1	

OK

## Processor Details

▶ Running

⚙ STOP & CONFIGURE

SETTINGS

SCHEDULING

PROPERTIES

COMMENTS

Required field

Property		Value	
Destination	?	flowfile-attribute	
Return Type	?	auto-detect	
Path Not Found Behavior	?	ignore	
Null Value Representation	?	empty string	
sourcekey	?	\$.key	

OK

## Processor Details

▶ Running

⚙ STOP & CONFIGURE

SETTINGS

SCHEDULING

PROPERTIES

COMMENTS

Required field

Property		Value	
Bucket	?	chapter8input	
Object Key	?	\${sourcekey}	
Region	?	US East (Ohio)	
Access Key ID	?	Sensitive value set	
Secret Access Key	?	Sensitive value set	
Credentials File	?	No value set	
AWS Credentials Provider service	?	No value set	
Communications Timeout	?	30 secs	
Version	?	No value set	
SSL Context Service	?	No value set	
Endpoint Override URL	?	No value set	
Signer Override	?	Default Signature	

OK

SETTINGS

SCHEDULING

PROPERTIES

COMMENTS

Required field

Property	Value
Record Reader	JsonTreeReader
Record Writer	JsonRecordSetWriter
Jolt Transformation DSL	Chain
Custom Transformation Class Name	No value set
Custom Module Directory	No value set
Jolt Specification	[[{"operation": "default", "spec": {"source": "\${sourcekey:s...}}
Transform Cache Size	1

[[{"operation": "default", "spec": {"source": "\${sourcekey:substringBefore('\_','\_')}"}}]]

1

2

3

4

5

6

```
1 [{
2   "operation": "default",
3   "spec": {
4     "source": "${sourcekey:substringBefore('_','_')}"
5   }
6 }]
```

OK

Running

STOP & CONFIGURE

SETTINGS

SCHEDULING

PROPERTIES

COMMENTS

Required field

Property	Value
Record Reader	JsonTreeReader
Database Type	MySQL
Statement Type	INSERT
Database Connection Pooling Service	MysqlDBCPConnectionPool
Catalog Name	No value set
Schema Name	No value set
Table Name	bundle_events
Translate Field Names	true
Unmatched Field Behavior	Ignore Unmatched Fields
Unmatched Column Behavior	Fail on Unmatched Columns
Update Keys	No value set
Field Containing SQL	No value set

OK

Controller Service Details

SETTINGSPROPERTIESCOMMENTS

Required field

Property	Value
Database Connection URL	<div>jdbc:mysql://ls-55ea092394d371f4971f55dbb...</div>
Database Driver Class Name	<div>com.mysql.jdbc.Driver</div>
Database Driver Location(s)	<div>/home/ec2-user/mysql-connector-java-8.0.28.jar</div>
Kerberos Credentials Service	<div>No value set</div>
Kerberos Principal	<div>No value set</div>
Kerberos Password	<div>No value set</div>
Database User	<div>dbmasteruser</div>
Password	<div>Sensitive value set</div>
Max Wait Time	<div>500 millis</div>

OK

Search Datasets, People, & more...

AnalyticsIngestionGovern

Manage Ingestion

Create, schedule, and run DataHub ingestion pipelines.

SourcesSecrets

Create new sourceRefresh

Search sources...

Type	Name	Schedule	Execution Count	Last Execution	Last Status	
<div></div>	test kafka	None	1	6/10/2022 at 4:54:59 PM	<div>Failed</div>	<div></div> <div>EDIT</div> <div>EXECUTE</div> <div></div>
<div></div>	af	None	4	6/10/2022 at 4:53:33 AM	<div>Failed</div>	<div></div> <div>EDIT</div> <div>EXECUTE</div> <div></div>

<1>

1 Choose Type2 Configure Recipe3 Schedule Execution4 Finish up

BigQuery

Redshift

Snowflake

Kafka

Looker

MySQL

Postgres

MongoDB

Azure AD

Okta

Glue

Oracle

Hive

Custom

## New Ingestion Source



- 1 Choose Type — 2 Configure Recipe — 3 Schedule Execution — 4 Finish up

### Configure Custom Recipe

For more information about how to configure a recipe, see the [Custom source docs](#).

```
1 source:
2   type: "nifi"
3   config:
4     # Coordinates
5     site_url: "https://localhost:8443/nifi/"
6
7     # Credentials
8     auth: SINGLE_USER
9     username: sinchan
10    password: [REDACTED]
11
12 sink:
13   type: datahub-rest
14   config:
15     server: "http://localhost:9002/api/gms"
```

AnalyticsDomainsUsers & Groups

New Ingestion Source

1 Choose Type — 2 Configure Recipe — 3 Schedule Execution — 4 Finish up

Create an Execution Schedule

Configure your ingestion source to run on a schedule.

\* Schedule

Provide a custom cron schedule.

0,30 \* \* \* \*

☒ Runs at 0 and 30 minutes past the hour.

\* Timezone

Select the timezone to run the cron schedule in.

America/New\_York

Previous

SkipNext

## New Ingestion Source



- 1 Choose Type — 2 Configure Recipe — 3 Schedule Execution — 4 Finish up

### \* Name

Give this ingestion source a name.

NIFI source

> Advanced

Previous





Done

## Manage Ingestion

Create, schedule, and run DataHub ingestion pipelines.

[Sources](#) [Secrets](#)

+ Create new source [Refresh](#)

	Type	Name	Schedule	Execution Count	Last Execution	Last Status
+		mysql	0,30 * * * *	42	6/5/2022 at 5:30:04 PM	 Running
+	Custom	NIFI source	0,30 * * * *	62	6/5/2022 at 5:30:04 PM	 Failed
+	Custom	S3Input source	0,30 * * * *	36	6/5/2022 at 5:30:04 PM	 Running

### Create new Domain



#### Name

Give your new Domain a name.



#### Description

An optional description for your new domain. You can change this later.



> Advanced

Cancel

Create

Create new group



Name

Give your new group a name.

offer\_analytics



Description

An optional description for your new group.

Offer Analytics team deliver technology to generate new offer based on data a



> Advanced

Cancel

Create



MySQL

347



AWS S3

7



Nifi

3

Domains



Sales

3



DataServices

2

Recently Viewed

bundle\_events

Schema [Documentation](#) Properties Lineage Queries Stats Validation

Back

✓ Save

**B** *I*

This table is a telecom provider's event collection table. It stores the customer events whenever a customer buys a phone recharge bundle, flowing in through various sources like MOB or WEB in realtime

This table is a telecom provider's event collection table. It stores the customer events whenever a customer buys a phone recharge bundle flowing in through various sources like MOB or WEB in



## bundle\_events

[Schema](#) [Documentation](#) [Properties](#) [Lineage](#) [Queries](#) [Stats](#) [Validation](#)

Reported at 5/30/2022, 10:00 AM EDT

[Normal](#)

[Blame](#)

0.0.0 - 2 days ago



Field	Description	Tags	Terms
customerid <small>(Number)</small> <a href="#">Primary Key</a>	Customer ID , unique to each customer	<a href="#">(edited)</a>	
bundleid <small>(Number)</small> <a href="#">Primary Key</a>	Bundle or the recharge Plan Id	<a href="#">(edited)</a>	
timestamp <small>(String)</small> <a href="#">Primary Key</a>	time when the bundle was purchased	<a href="#">(edited)</a>	
source <small>(String)</small>	stores the source type of the record /event ingested whether it is MOB /WEB/ PAPER/ PROMO/ REFERENCE. <a href="#">Read Less</a>	<a href="#">(edited)</a>	

Datasets > prod > s3 > chapter8input

[Dataset](#) [AWS S3](#)

### chapter8input

[Schema](#) [Documentation](#) [Properties](#) [Lineage](#) [Queries](#) [Stats](#) [Validation](#)

[Visualize Lineage](#)

[Impact Analysis](#)

#### 0 Upstream

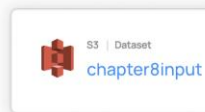
No Data

#### 1 Downstream

Data Task [Nifi](#)

**FetchS3Object**

Datasets > prod > s3 > chapter8input



[Pipeline](#) [Nifi](#)

### NiFi Flow

[Documentation](#) [Properties](#) [Tasks](#)

Contains 2 Tasks

Data Task [Nifi](#)

**FetchS3Object**

Data Task [Nifi](#)

**ListS3**

[View in Nifi](#)

[Tags](#)

No tags added yet. Tag entities to help make them more discoverable and call out their most important attributes.

[+ Add Tags](#)

[Glossary Terms](#)

No terms added yet. Apply glossary terms to entities to help discover their data.

[+ Add Terms](#)

[Owners](#)

[offer\\_analytics](#) [+ Add Owners](#)

[Domain](#)


[Sales](#) [X](#)


## Add Owners



### Owner

Find a user or group

 offer\_analytics X

 offer\_analytics ✓

### Type

Choose an owner type

Technical Owner

Cancel

Add

## Edit a Policy



- 1 Choose Policy Type
- 2 Configure Privileges
- 3 Assign Users & Groups

### Name

A name for your new policy.

View\_Analytics

### Type

The type of policy you would like to create.

Platform

The Platform policy type allows you to assign top-level DataHub Platform privileges to users. These include managing users and groups, creating policies, viewing analytics dashboards and more.

The Metadata policy type allows you to assign metadata privileges to users. These include the ability to manipulate metadata like ownership, tags, documentation associated with Datasets, Charts, Dashboards, & more.

### Description

An optional description for your new policy.

This policy is for viewing analytics only

Next

## Edit a Policy



- ✓ Choose Policy Type
- 2 Configure Privileges
- 3 Assign Users & Groups

### Privileges

Select a set of privileges to permit.

View Analytics X

Previous

Next

## Edit a Policy



- ✓ Choose Policy Type — ✓ Configure Privileges — 3 Assign Users & Groups

### Applies to

Select the users & groups that this policy should apply to.

#### Users

Select for specific users that this policy should apply to, or select 'All Users' to apply it to all users.

Search for users...

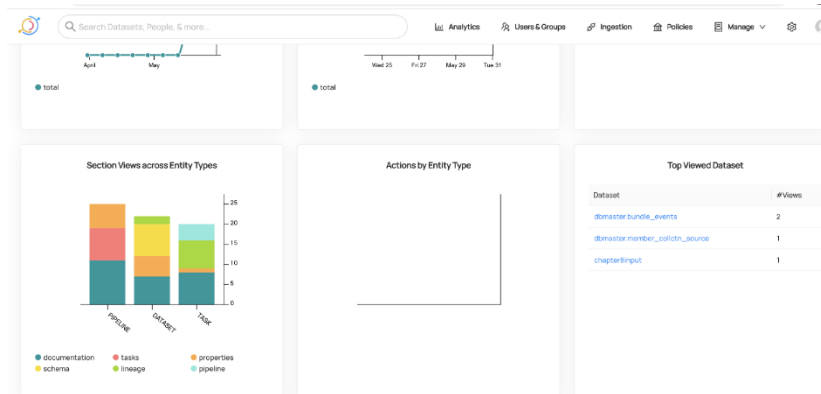
#### Groups

Select for specific groups that this policy should apply to, or select 'All Groups' to apply it to all groups.

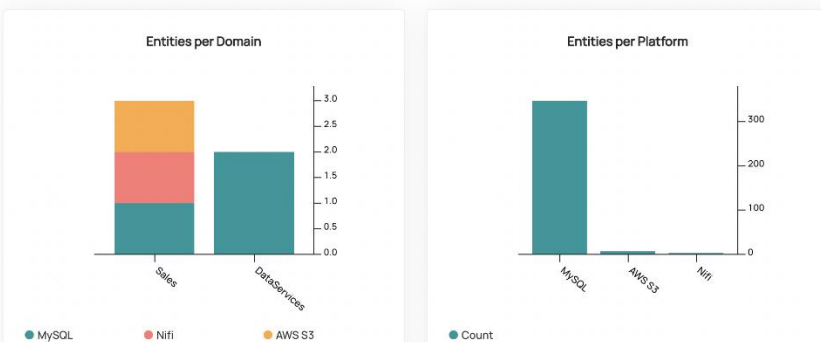
executives X

Previous

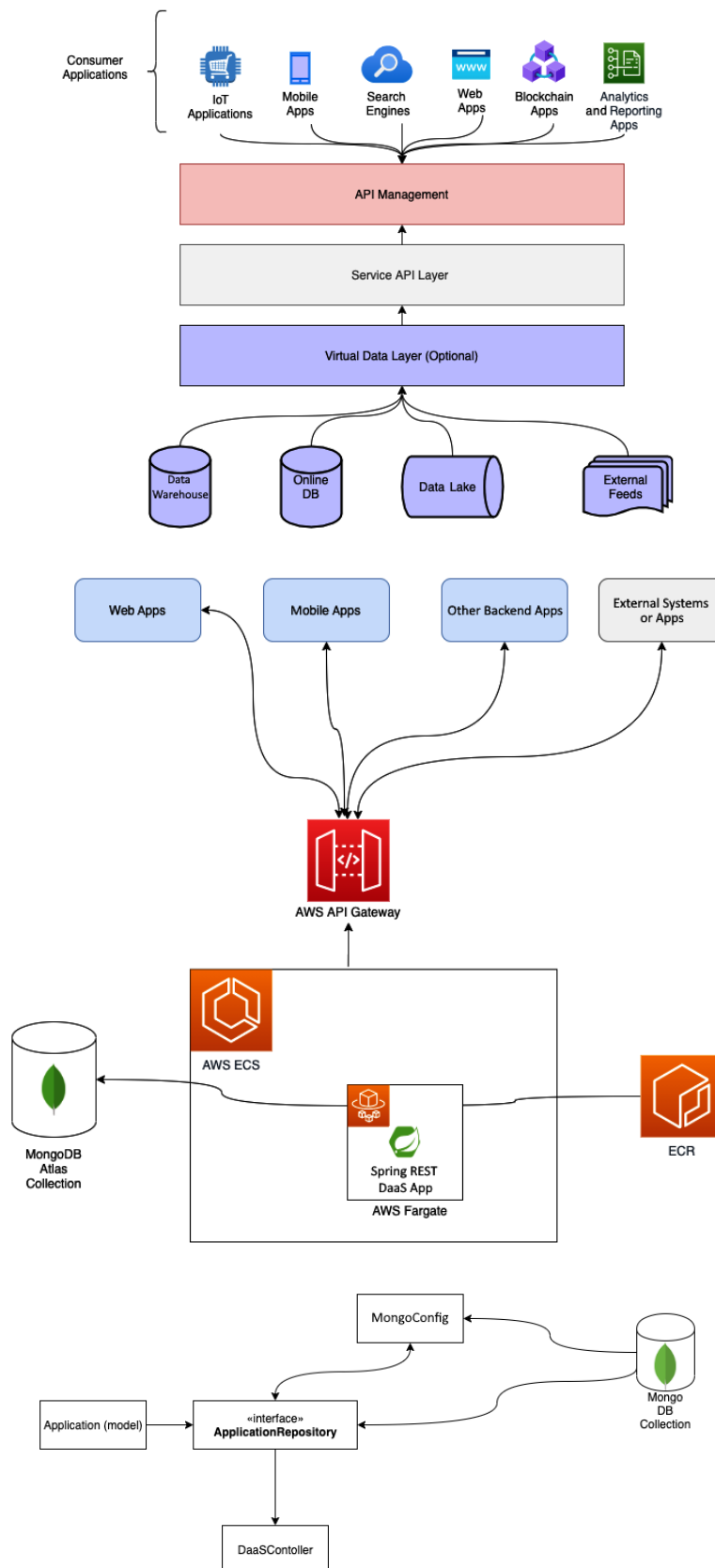
Save

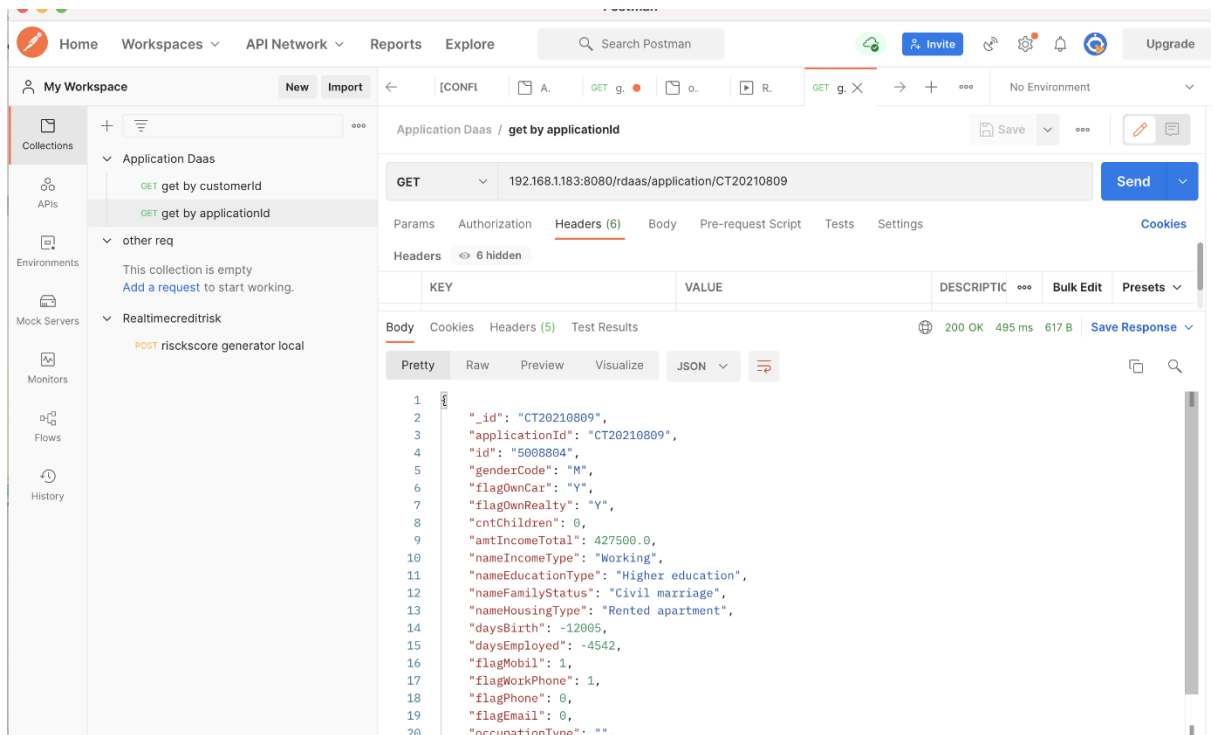
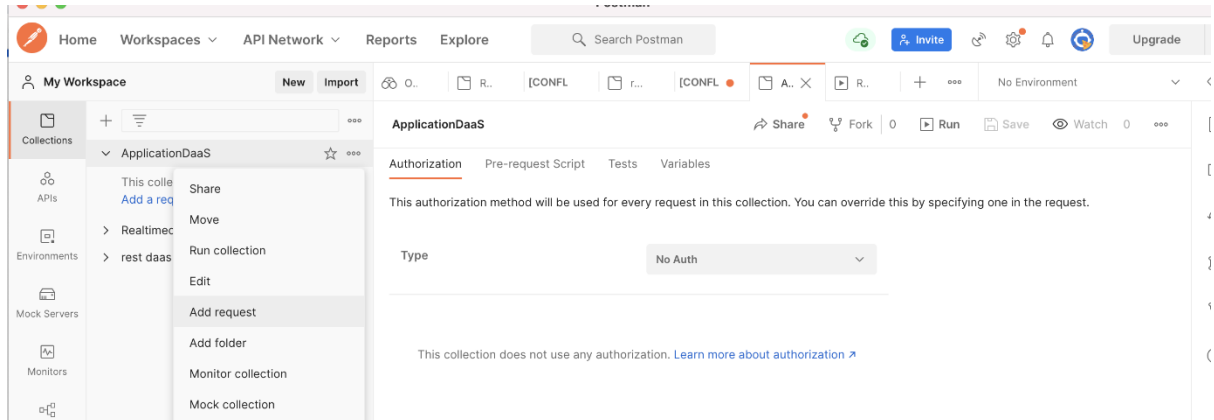
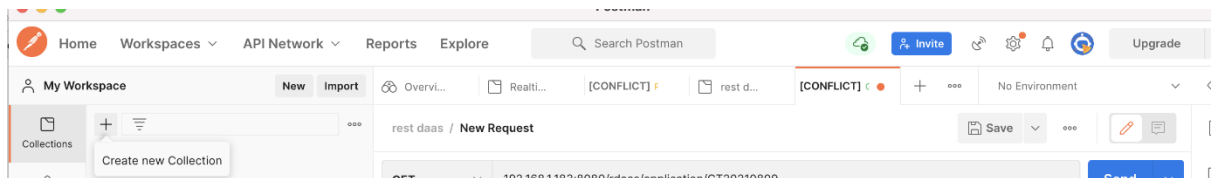


### Data Landscape Summary



## Chapter 9: Exposing MongoDB Data as a Service





# Create repository

## General settings

**Visibility settings** [Info](#)  
Choose the visibility setting for the repository.

☒ **Private**  
Access is managed by IAM and repository policy permissions.

☐ **Public**  
Publicly visible and accessible for image pulls.

**Repository name**  
Provide a concise name. A developer should be able to identify the repository contents by the name.

627443126298.dkr.ecr.us-east-2.amazonaws.com/

11 out of 256 characters maximum (2 minimum). The name must start with a letter and can only contain lowercase letters, numbers, hyphens, underscores, and forward slashes.

**Tag immutability** [Info](#)  
Enable tag immutability to prevent image tags from being overwritten by subsequent image pushes using the same tag. Disable tag immutability to allow image tags to be overwritten.

☒ **Disabled**

Once a repository is created, the visibility setting of the repository can't be changed.

Amazon ECR > Repositories

PrivatePublic

Private repositories (1)

Find repositories

View push commands

Delete

Edit

Create repository

Repository name	URI	Created at	Tag immutability	Scan frequency	Encryption type	Pull through cache
<input type="radio"/> restdaas	627443126298.dkr.ecr.us-east-2.amazonaws.com/restdaas	March 06, 2022, 12:11:52 (UTC-05)	Disabled	Manual	AES-256	Inactive

us-east-2.console.aws.amazon.com/ecs/home?region=us-east-2#/clusters

New ECS Experience

Amazon ECS

Clusters

Task Definitions

Account Settings

Amazon EKS

Amazon ECR

Repositories

AWS Marketplace

## Clusters

An Amazon ECS cluster is a regional grouping of one or more container instances on which you can run task requests. Each account receives one Amazon ECS service. Clusters may contain more than one Amazon EC2 instance type.

For more information, see the [ECS documentation](#).

Create Cluster

Get Started

View

list

card

No clusters found

## Create Cluster

### Step 1: Select cluster template

Step 2: Configure cluster

### Select cluster template

The following cluster templates are available to simplify cluster creation. Additional configuration and integrations can be added later.

**Networking only**

**Resources to be created:**

Cluster

VPC (optional)

Subnets (optional)

**For use with either AWS Fargate (Windows/Linux) or with External Instance capacity.**

**EC2 Linux + Networking**

**Resources to be created:**

Cluster

VPC

Subnets

Auto Scaling group with Linux AMI

**EC2 Windows + Networking**

**Resources to be created:**

Step 1: Select cluster template

### Step 2: Configure cluster

### Configure cluster

Cluster name\*

### Networking

Create a new VPC for your cluster to use. A VPC is an isolated portion of the AWS Cloud populated by AWS objects, such as Fargate tasks.

Create VPC ☐ Create a new VPC for this cluster

### Tags

Key	Value
<input type="text" value="Add key"/>	<input type="text" value="Add value"/>

### CloudWatch Container Insights

CloudWatch Container Insights is a monitoring and troubleshooting solution for containerized applications and microservices. It collects, aggregates, and summarizes compute utilization such as CPU, memory, disk, and network; and diagnostic information such as container restart failures to help you isolate issues with your clusters and resolve them quickly. [Learn more](#)

CloudWatch Container Insights ☐ Enable Container Insights

\*Required

Cancel

Previous

Create

☒ New ECS Experience  
[Tell us what you think](#)

Amazon ECS

Clusters

**Task Definitions**

Account Settings

Amazon EKS

Clusters

Amazon ECR

Repositories

AWS Marketplace

Discover software

Subscriptions

## Task Definitions

Task definitions specify the container information for your application, such as how many containers are part of your task, what resources they use, and which host ports they will use. [Learn more](#)

Create new Task Definition

Create new revision

Actions

Last updated on March 1

**Status:** **ACTIVE** INACTIVE

<input type="checkbox"/>	Task Definition	Latest revision status
		No results


## Create new Task Definition


### Step 1: Select launch type compatibility


Step 2: Configure task and container definitions

### Select launch type compatibility

Select which launch type you want your task definition to be compatible with based on where you want to launch your task.

**FARGATE**  
  
Price based on task size  
Requires network mode awsvpc  
AWS-managed infrastructure, no Amazon EC2 instances to manage

**EC2**  
  
Price based on resource usage  
Multiple network modes available  
Self-managed infrastructure using Amazon EC2 instances

**EXTERNAL**  
  
Price based on instance-hours and additional charges for other AWS services used  
Self-managed on-premise infrastructure with ECS Anywhere

\*Required

Cancel

Next step

## Create new Task Definition

Step 1: Select launch type compatibility

### Step 2: Configure task and container definitions

### Configure task and container definitions

A task definition specifies which containers are included in your task and how they interact with each other. You can also specify data volumes for your containers to use. [Learn more](#)

Task definition name\*  ⓘ

Requires compatibilities\* FARGATE

Task role  ⓘ

Optional IAM role that tasks can use to make API requests to authorized AWS services. Create an Amazon Elastic Container Service Task Role in the [IAM Console](#)

Network mode  ⓘ

If you choose <default>, ECS will start your container using Docker's default networking mode, which is Bridge on Linux and NAT on Windows. Windows tasks support the <default> and awsvpc network modes.

Operating system family  ⓘ

### Task execution IAM role

This role is required by tasks to pull container images and publish container logs to Amazon CloudWatch on your behalf. If you do not have the ecsTaskExecutionRole already, we can create one for you.



### Task execution IAM role

This role is required by tasks to pull container images and publish container logs to Amazon CloudWatch on your behalf. If you do not have the `ecsTaskExecutionRole` already, we can create one for you.

**Task execution role** You are giving permission to Elastic Container Service to create and use `ecsTaskExecutionRole`.



### Task size



The task size allows you to specify a fixed size for your task. Task size is required for tasks using the Fargate launch type and is optional for the EC2 or External launch type. Container level memory settings are optional when task size is set. Task size is not supported for Windows containers.

**Task memory (GB)** 1GB

The valid memory range for 0.5 vCPU is: 1GB - 4GB.

**Task CPU (vCPU)** 0.5 vCPU

The valid CPU range for 1GB memory is: 0.25 vCPU - 0.5 vCPU.

#### Task memory maximum allocation for container memory reservation



#### Task CPU maximum allocation for containers



### Container definitions



Add container

Container ...	Image	Hard/Soft ...	CPU Unit...	GPU	Inference A...	Essential ...
---------------	-------	---------------	-------------	-----	----------------	---------------

#### Add container



##### Standard

**Container name\*** restdaas



**Image\*** 627443126298.dkr.ecr.us-east-2.amazonaws.com/restdaas:v1



**Private repository authentication\*** ☐



**Memory Limits (MiB)** Soft limit 256



[Add Hard limit](#)

Define hard and/or soft memory limits in MiB for your container. Hard and soft limits correspond to the 'memory' and 'memoryReservation' parameters, respectively, in task definitions. ECS recommends 300-500 MiB as a starting point for web applications.

**Port mappings** Container port 8080 Protocol tcp



[Add port mapping](#)

Host port mappings are not valid when the network mode for a task definition is host or awsvpc. To specify different host and container port mappings, choose the Bridge network mode.

##### Advanced container configuration

\* Required

Cancel

Add

Services

Search for services, features, blogs, docs, and more

[Option+S]

Ohio

Sanchar Yadoo

New ECS Experience

Tell us what you think

Amazon ECS

Clusters

Task Definitions

Account Settings

Amazon EKS

Clusters

Amazon ECR

Repositories

AWS Marketplace

Discover software

Subscriptions

Task Definitions

Task definitions specify the container information for your application, such as how many containers are part of your task, what resources they will use, how they are linked together, and which host ports they will use.

Create new Task Definition

Create new revision

Actions

Run Task

Create Service

Update Service

Status: ACTIVE INACTIVE 1 selected

Filter in this page

Task Definition

restdaas

Latest revision status

ACTIVE

Last updated on March 13, 2022 1:38:15 PM (4m ago)

1-1

Page size

50

Services

Search for services, features, blogs, docs, and more

[Option+S]

Ohio

Sanchar Yadoo

New ECS Experience

Tell us what you think

Amazon ECS

Clusters

Task Definitions

Account Settings

Amazon EKS

Clusters

Amazon ECR

Repositories

AWS Marketplace

Discover software

Subscriptions

Run Task

Select the cluster to run your task definition on and the number of copies of that task to run. To apply container overrides or target particular container in

Launch type

FARGATE

EC2

EXTERNAL

Switch to capacity provider strategy

Operating system family

Linux

Task Definition

Family

restdaas

Revision

1 (latest)

Platform version

LATEST

Cluster

daas-cluster

Number of tasks

1

Task Group

Services

Search for services, features, blogs, docs, and more

[Option+S]

Ohio

Sanchar Yadoo

New ECS Experience

Tell us what you think

Amazon ECS

Clusters

Task Definitions

Account Settings

Amazon EKS

Clusters

Amazon ECR

Repositories

AWS Marketplace

Discover software

Subscriptions

VPC and security groups

VPC and security groups are configurable when your task definition uses the awsvpc network mode.

Cluster VPC\*

vpc-68e52d03 (172.31.0.0/16)

Subnets\*

subnet-e927d782 (172.31.0.0/20) - us-east-2a assign ipv6 on creation: Disabled

Security groups\*

restda-6983

Edit

Auto-assign public IP

ENABLED

Advanced Options

Task tagging configuration

Enable ECS managed tags

Propagate tags from

Do not propagate

Tags

Key

Value

✕

Assigned security groups

- ☒ Create new security group
- ☐ Select existing security group



[+ Add rule](#)

Type	Protocol	Port range	Source		
HTTP	TCP	80	Anywhere	0.0.0.0/0, ::/0	X
Custom TCP	TCP	8080	Anywhere	0.0.0.0/0, ::/0	X

- Amazon ECS
  - Clusters**
  - Task Definitions
  - Account Settings
- Amazon EKS
  - Clusters
- Amazon ECR
  - Repositories
- AWS Marketplace
  - Discover software
  - Subscriptions [↗](#)

Cluster : daas-cluster

Update

Status **ACTIVE**

Registered container instances	0
Pending tasks count	0 Fargate, 0 EC2, 0 External
Running tasks count	1 Fargate, 0 EC2, 0 External
Active service count	0 Fargate, 0 EC2, 0 External
Draining service count	0 Fargate, 0 EC2, 0 External


Run new Task Stop Stop All Actions ▾

Last updated on March 14, 2022 8:36

▼ Filter in this page

Launch type ALL

Task	Task def...	Container instanc...	Last status	Desired status	Started at	Started By	Group	Launch type
	<a href="#">d48e0fd46ec94f08...</a>	restdaasm...	--	RUNNING	2022-03-14 20:15:...		family:restdaasm1	FARGATE
	d48e0fd46ec94f0883fa32d048321db7							

- Account Settings
- Amazon EKS Clusters
- Amazon ECR Repositories
- AWS Marketplace Discover software
- Subscriptions 

Launch type FARGATE

Platform version 1.4.0

Task definition [restdaasm1:1](#)

Group family:restdaasm1

Task role	None
-----------	------

Last status **RUNNING**

Desired status RUNNING

Created at 2022-03-14 20:15:27 -0400

Started at 2022-03-14 20:15:58 -0400

Network mode `awsipc`

ENI Id [eni-0e3b9073f4e7dfe9b](#)

Subnet Id subnet-e927d782

Private IP 172.31.12.193

Public IP 18.220.175.8

Mac address 02:c7:97:38:fb:ec

## Containers

Last updated on March 14, 2022 8:37

Name	Container Runtime I...	Status	Image	Image Digest	CPU Units	Hard/Soft mem...	Essential
------	------------------------	--------	-------	--------------	-----------	------------------	-----------

Application Daas / **get by applicationId From ECS**

GET 18.220.175.8:8080/rdaas/application/CT20210809

Params Authorization Headers (7) Body Pre-request Script Tests Settings Cookies

none form-data x-www-form-urlencoded raw binary GraphQL

This request does not have a body

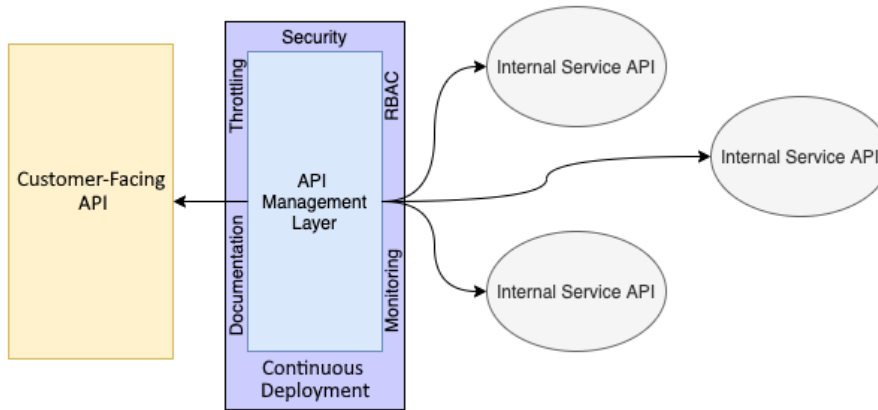
Body Cookies Headers (5) Test Results 200 OK 59 ms 617 B Save Response

Pretty Raw Preview Visualize JSON

```

1 {
2   "_id": "CT20210809",
3   "applicationId": "CT20210809",
4   "id": "5008804",
5   "genderCode": "M",
6   "flagOwnCar": "Y",
7   "flagOwnRealty": "Y",
8   "cntChildren": 0,
9   "amtIncomeTotal": 427500.0,
10  "nameIncomeType": "Working",
11  "nameEducationType": "Higher education",
12  "nameFamilyStatus": "Civil marriage",
13  "nameHousingType": "Rented apartment",
14  "daysBirth": -12005,
15  "daysEmployed": -4542,
16  "flagMobil": 1,

```



**API Gateway** ×

create, maintain, and secure APIs at any scale

Amazon API Gateway helps developers to create and manage APIs to back-end systems running on Amazon EC2, AWS Lambda, or any publicly addressable web service. With Amazon API Gateway, you can generate custom client SDKs for your APIs to connect your back-end systems to mobile, web, and server applications or services.

Choose an API type

**HTTP API**

Build low-latency and cost-effective REST APIs with built-in features such as OAuth2 and CORS support.

Works with the following: Lambda, HTTP backends

Import Build

**WebSocket API**

Build a WebSocket API using persistent connections for real-time use cases such as chat applications or dashboards.

Works with the following: Lambda, HTTP, AWS Services

Build

**REST API**

Develop a REST API where you gain complete control over the request and response along with API management capabilities.

Works with the following: Lambda, HTTP, AWS Services

Import Build

**REST API Private**

Create a REST API that is only accessible from within a VPC.

Works with the following: Lambda, HTTP, AWS Services

Import Build

Amazon API Gateway

APIs > Create

Show all hints?

APIs

Custom Domain Names

VPC Links

### Choose the protocol

Select whether you would like to create a REST API or a WebSocket API.

☒ REST ☐ WebSocket

### Create new API

In Amazon API Gateway, a REST API refers to a collection of resources and methods that can be invoked through HTTPS endpoints.

☒ New API ☐ Clone from existing API ☐ Import from Swagger or Open API 3 ☐ Example API

### Settings

Choose a friendly name and description for your API.

API name\*

RestDaasAPI

Description

test description

Endpoint Type

Regional

\* Required

Create API

Amazon API Gateway

APIs > RestDaasAPI (u9xrblum43) > Resources > /

APIs

Custom Domain Names

VPC Links

API: RestDaasAPI

Resources

Stages

Authorizers

Gateway Responses

Models

Resource Policy

Documentation

Dashboard

Settings

Usage Plans

Resources

Actions

/ Methods

/

RESOURCE ACTIONS

Create Method

Create Resource

Enable CORS

Edit Resource Documentation

API ACTIONS

Deploy API

Import API

Edit API Documentation

Delete API

Amazon API Gateway

APIs > RestDaasAPI (u9xrblum43) > Resources > /loanapplications/{appld} (41xokn) > GET

APIs

Custom Domain Names

VPC Links

API: RestDaasAPI

Resources

Stages

Authorizers

Gateway Responses

Models

Resource Policy

Documentation

Dashboard

Settings

Usage Plans

Resources

Actions

/loanapplications/{appld} - GET - Setup

/

/loanapplications

/[appld]

GET

### Choose the integration point for your new method.

Integration type

☐ Lambda Function

☒ HTTP

☐ Mock

☐ AWS Service

☐ VPC Link

Use HTTP Proxy integration

☐

HTTP method

GET

Endpoint URL

10.175.8:8080/daas/application/{appld}

Content Handling

Passthrough

Use Default Timeout

☒

Amazon API Gateway

APIs > RestDaasAPI (u9xrblum43) > Resources > /loanapplications/{appld} (41xokn) > GET

APIs

Custom Domain Names

VPC Links

API: RestDaasAPI

Resources

Stages

Authorizers

Gateway Responses

Models

Resource Policy

Documentation

Dashboard

Settings

Usage Plans

API Keys

Resources

Actions

Method Actions

Integration Request

Method Request

Method Response

Integration Response

Auth: NONE

ARN: arn:aws:execute-api:us-east-2:627443126298:u9xrblum43:/GET/loanapplications/\*

Type: HTTP

Paths: appld

Input passthrough: Yes

HTTP Status: 200

Models: application/json => Empty

HTTP status pattern: -

Output passthrough: Yes

Amazon API Gateway

APIs > RestDaasAPI (u9xrblum43) > Resources > /loanapplications/{appld} (41xokn) > GET

APIs

Custom Domain Names

VPC Links

API: RestDaasAPI

Resources

Stages

Authorizers

Gateway Responses

Models

Resource Policy

Documentation

Dashboard

Settings

Usage Plans

Resources

Actions

Method Actions

Integration Request

Method Request

Method Response

Integration Response

Auth: NONE

ARN: arn:aws:execute-api:us-east-2:627443126298:u9xrblum43:/GET/loanapplications/\*

Type: HTTP

Paths: appld

Input passthrough: Yes

HTTP Status: 200

Models: application/json => Empty

HTTP status pattern: -

Output passthrough: Yes

[Option+S]

Deploy API

Choose a stage where your API will be deployed. For example, a test version of your API could be deployed to a stage named beta.

Deployment stage

Stage name\*

Stage description

Deployment description

Cancel Deploy

ures, blogs, docs, and more [Option+S]

isAPI (u9xrblum43) > Stages > dev Show all hints ?

Create dev Stage Editor Delete Stage Configure Tags

Invoke URL: <https://u9xrblum43.execute-api.us-east-2.amazonaws.com/dev>

Settings Logs/Tracing Stage Variables SDK Generation Export Deployment History Documentation History Canary

Cache Settings

Enable API cache ☐

Default Method Throttling

Choose the default throttling level for the methods in this stage. Each method in this stage will respect these rate and burst settings. Your current account level throttling rate is **10000** requests per second with a burst of **5000** requests. [Read more about API Gateway throttles.](#)

New Import

Application Daas / get by applicationId From API Gateway Save

GET <https://u9xrblum43.execute-api.us-east-2.amazonaws.com/dev/loanapplications/CT20210809> Send

Params Authorization Headers (7) Body Pre-request Script Tests Settings Cookies

none form-data x-www-form-urlencoded raw binary GraphQL

This request does not have a body

Body Cookies Headers (7) Test Results 200 OK 69 ms 735 B Save Response

Pretty Raw Preview Visualize JSON

```
1
2  "_id": "CT20210809",
3    "applicationId": "CT20210809",
4    "id": "5008884",
5    "genderCode": "M",
6    "flagOwnCar": "Y",
7    "flagOwnRealty": "Y",
8    "cntChildren": 0,
9    "amtIncomeTotal": 427500.0,
10   "nameIncomeType": "Working",
11   "nameEducationType": "Higher education",
12   "nameFamilyStatus": "Civil marriage",
13   "nameHousingType": "Rented apartment",
14   "daysBirth": -12005,
15   "daysEmployed": -4542,
16   "flagMobil": 1,
```

Services Search for services, features, blogs, docs, and more [Option+S]

Amazon API Gateway APIs > RestDaasAPI (u9xrblum43) > Dashboard

APIs Custom Domain Names VPC Links

API: **RestDaasAPI**

Resources Stages Authorizers Gateway Responses Models Resource Policy Documentation **Dashboard** Settings Usage Plans API Keys Client Certificates

Invoke this API at: <https://u9xrblum43.execute-api.us-east-2.amazonaws.com/dev/>

Stage dev From 2/28/22 To 3/14/22

API Calls Latency Integration Latency

5xx Error

APIs

Custom Domain Names

VPC Links

API: RestDaasAPI

Resources

Stages

Authorizers

Gateway Responses

Models

Resource Policy

Documentation

Dashboard

Settings

Usage Plans

API Keys

Resources

Actions

Method Execution

/loanapplications/{appld} - GET - Method Request

Provide information about this method's authorization settings and the parameters it can receive.

Settings

Authorization NONE

Request Validator NONE

API Key Required false

Request Paths

URL Query String Parameters

HTTP Request Headers

Request Body

SDK Settings

Usage Plans

Create

Show all hints

Usage Plans

Create

Create Usage Plan

Usage Plans help you meter API usage. With Usage Plans, you can enforce a throttling and quota limit on each API key. Throttling limits define the maximum number of requests per second available to each key. Quota limits define the number of requests each API key is allowed to make over a period.

Name\* Basic

Description

Throttling

Enable throttling

Rate\* 2 requests per second

Burst\* 2 requests

Quota

Enable quota

100 requests per Month

\* Required

Next

### Configure Method Throttling

Resource	Method	Rate (requests per second)	Burst	
/loanapplications/{appld}	GET	2	2	

Close

Usage Plans

Create

Usage Plan API Keys

Subscribe an API key to this usage plan. Choose "Add API Key" below to search through your existing API keys. Once a key is associated with a plan, API Gateway will meter all requests from the key and apply the plan's throttling and quota limits.

Add API Key to Usage Plan

Create API Key and add to Usage Plan

Results per page: 100

Name

No associated API keys

<< < Page 1 > >



API Key

Create an API Key and add it to the Usage Plan

Name\*

test\_customer

API key\*

Auto Generate

Custom

Description

needs for testing

\* Required

Cancel

Save

APIs

Custom Domain Names

VPC Links

APIs RestDaasAPI

Resources

Stages

Authorizers

Gateway Responses

Models

Resource Policy

Documentation

Dashboard

Settings

Usage Plans

API Keys

Client Certificates

API Keys

test\_customer

Actions

Delete API Key

Configure

ID

qpv8aum3

Name

test\_customer

API key

Show

Description

needs for testing

Enabled

Enabled

Associated Usage Plans

Add to Usage Plan

Usage Plan	API	Stage
Basic	RestDaasAPI	dev

Application Daas / get by applicationId From API Gateway

Save

Send

GET

https://u9xrblum43.execute-api.us-east-2.amazonaws.com/dev/loanapplications/CT20210809

Params

Authorization

Headers (8)

Body

Pre-request Script

Tests

Settings

Cookies

Type

API Key

The authorization header will be automatically generated when you send the request.  
[Learn more about authorization](#)

Key

x-api-key

Value

UV3pQxt6BN9qqteaPjIYX6Lc2Y14tV5taG7F...

Add to

Header

Body

Cookies

Headers (7)

Test Results

200 OK 77 ms 735 B

Save Response

Pretty

Raw

Preview

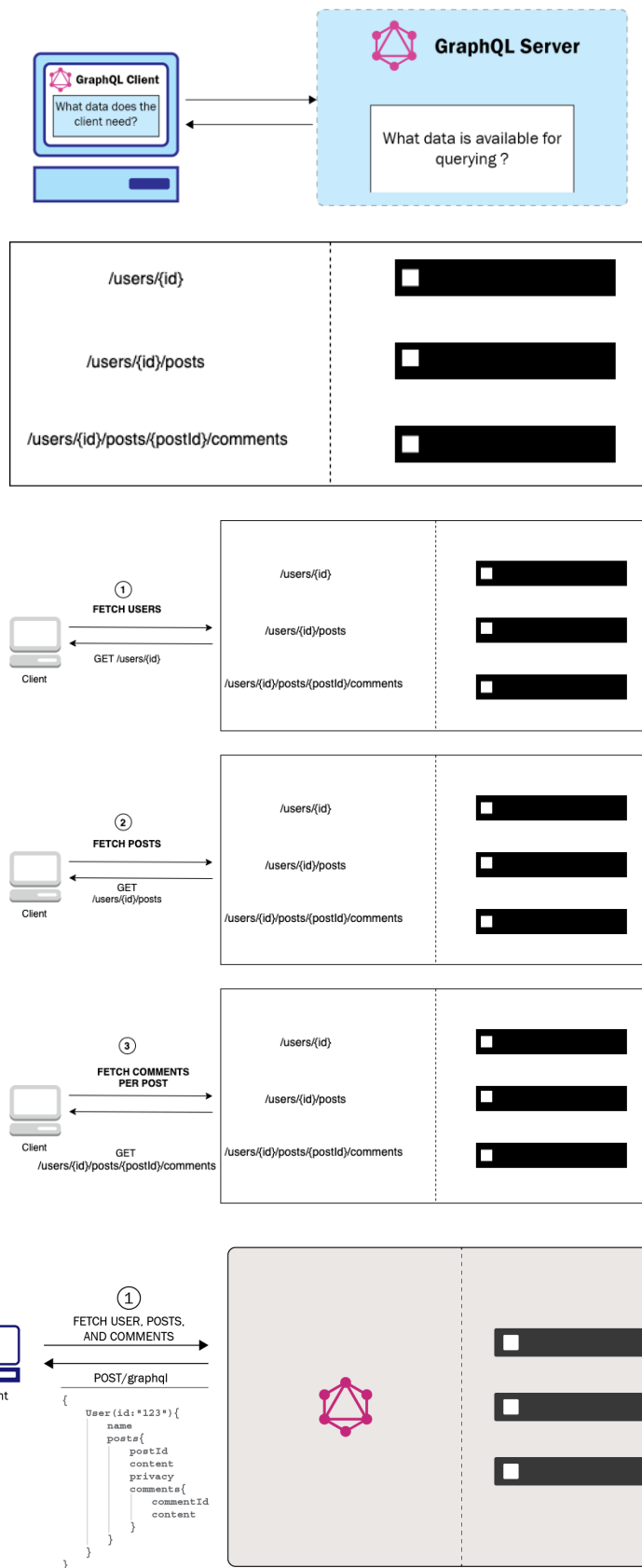
Visualize

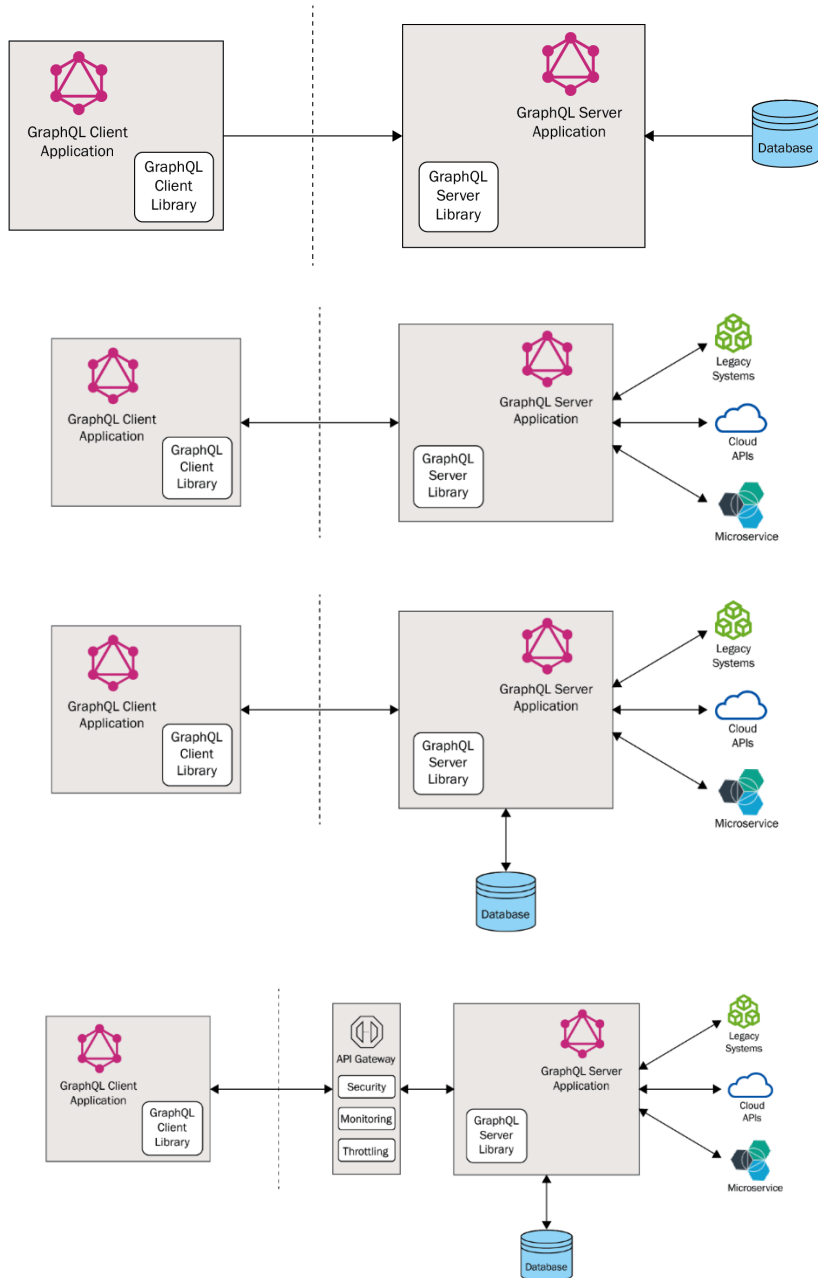
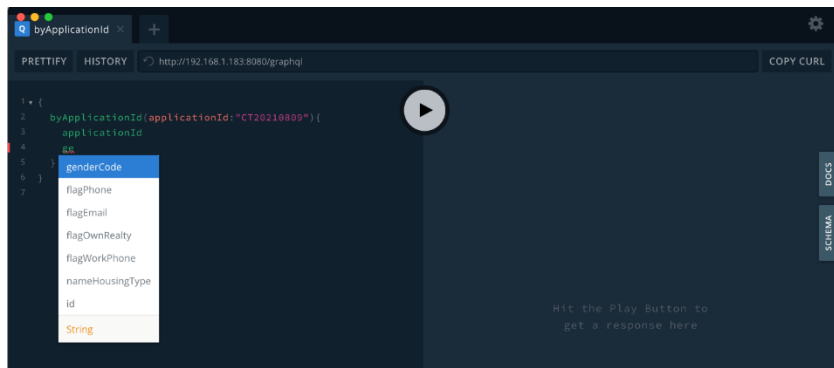
JSON

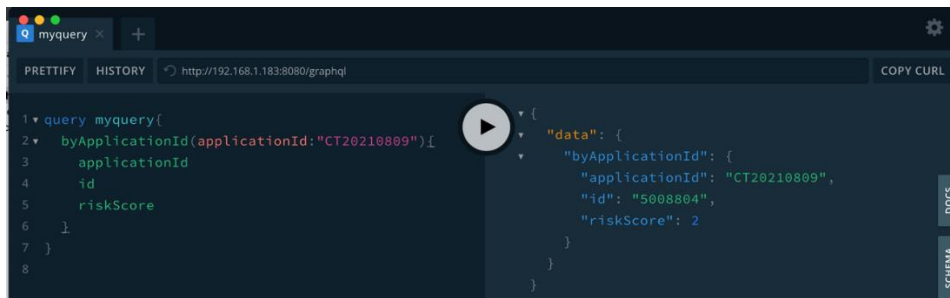
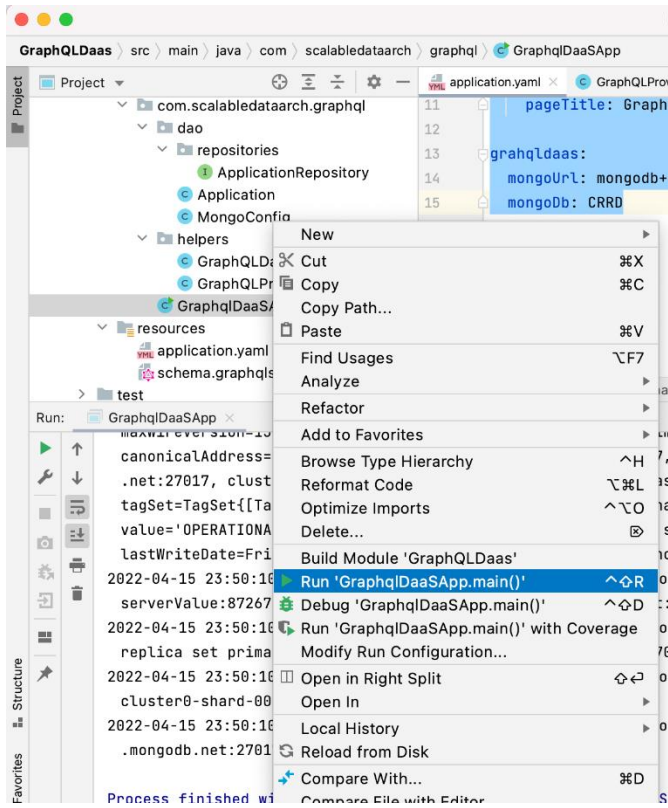
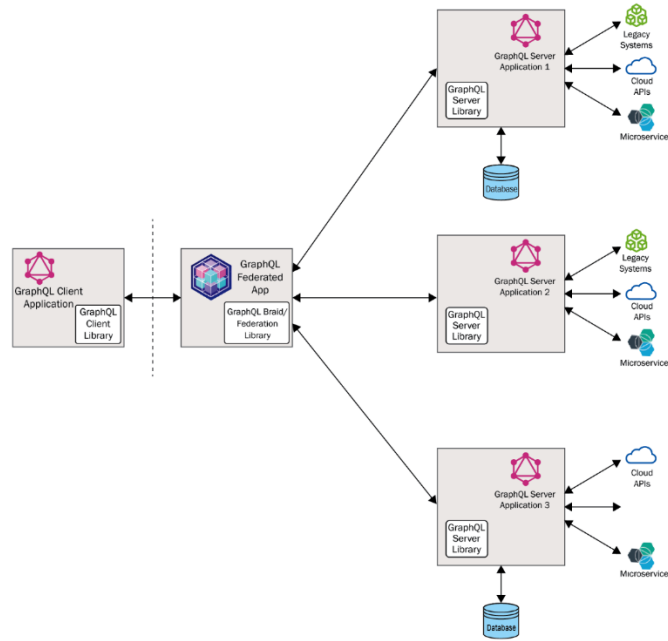
Copy

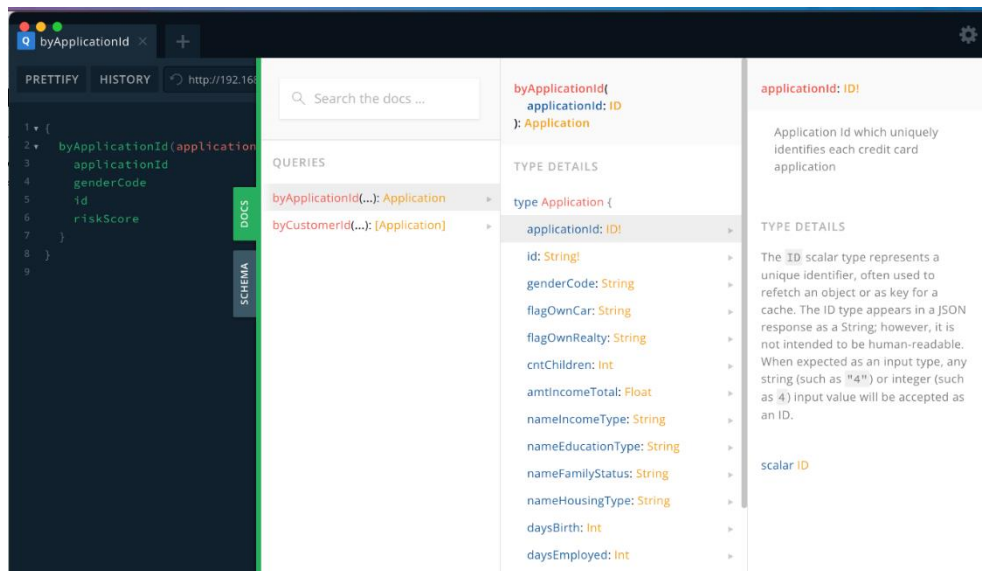
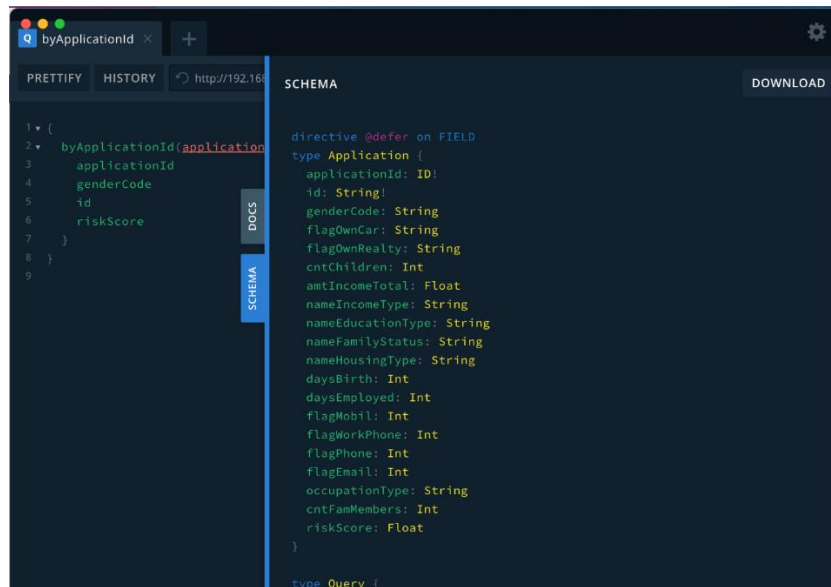
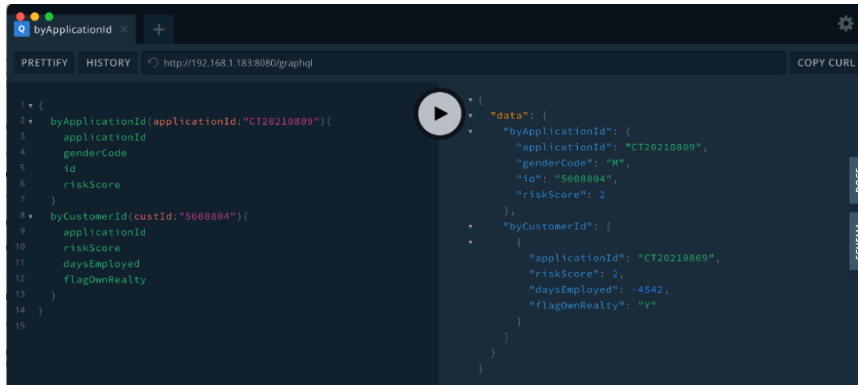
```
1 {
2   "_id": "CT20210809",
3   "applicationId": "CT20210809",
4   "id": "5008804",
5   "genderCode": "M",
6   "firstName": "V"
```

## Chapter 10: Federated and Scalable DaaS with GraphQL

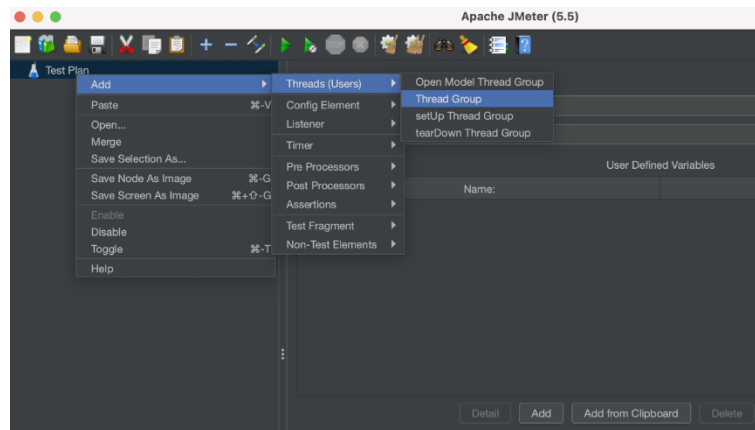
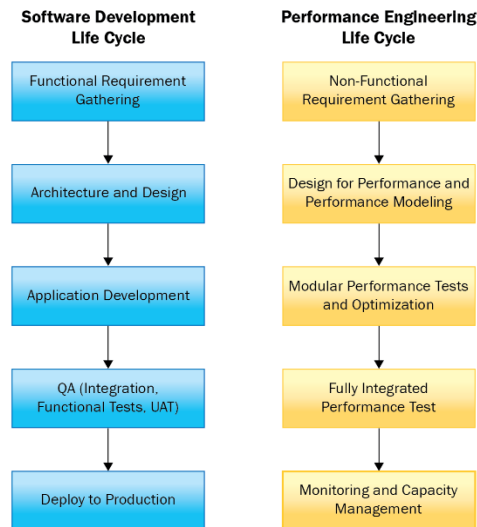








# Chapter 11: Measuring Performance and Benchmarking Your Applications



Apache JMeter (5)

Thread Group

Name: REST Thread Group

Comments:

Action to be taken after a Sampler error

☒ Continue ☐ Start Next Thread Loop ☐ Stop

Thread Properties

Number of Threads (users): 50

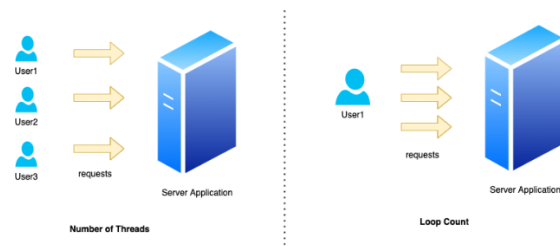
Ramp-up period (seconds): 50

Loop Count: ☐ Infinite ☒ 5

☒ Same user on each iteration

☐ Delay Thread creation until needed

☐ Specify Thread lifetime



Test Plan

REST Thread Group

HTTP

Add

Add Think Times to children

Start

Start no pauses

Validate

Cut

Copy

Paste

Duplicate

Remove

Open...

Merge

Save Selection As...

Save Node As Image

Save Screen As Image

Enable

Disable

Toggle

Thread Group

Sampler

Logic Controller

Pre Processors

Post Processors

Assertions

Timer

Test Fragment

Config Element

Listener

Flow Control Action

HTTP Request

Debug Sampler

JSR223 Sampler

AJRP/1.3 Sampler

Access Log Sampler

BeanShell Sampler

Bolt Request

FTP Request

GraphQL HTTP Request

JDBC Request

JMS Point-to-Point

JMS Publisher

JMS Subscriber

JUnit Request

Java Request

LDAP Extended Request

LDAP Request

Mail Reader Sampler

Number of Threads (users): 50

Ramp-up period (seconds): 50

Loop Count: ☐ Infinite ☒ 5

☒ Same user on each iteration

☐ Delay Thread creation until needed

Test Plan

REST Thread Group

HTTP Request

Name: HTTP Request

Comments:

Basic

Web Server

Protocol (http): http Server Name or IP: 192.168.1.183 Port Number: 8080

HTTP Request

GET Path: /daas/application/CT20210809 Content encoding:

☐ Redirect Automatically ☒ Follow Redirects ☒ Use KeepAlive ☐ Use multipart/form-data ☐ Browser-compatible headers

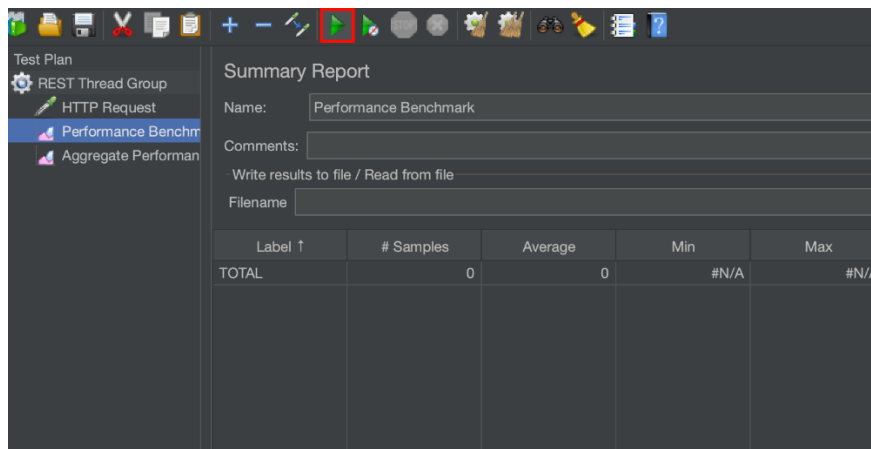
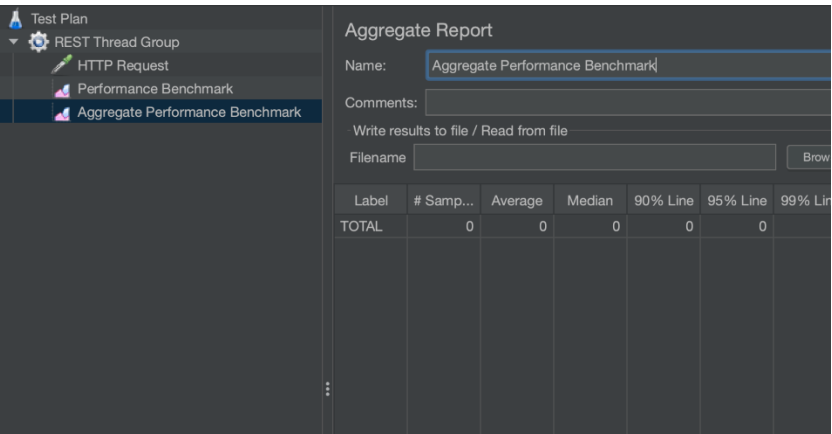
Parameters

Body Data

Files Upload

Send Parameters With the Request:

Name: Value URL Encode? Content-Type Include



Summary Report

Name: Performance Benchmark

Comments:

Write results to file / Read from file

Filename

Browse...

Log/Display Only:

Errors

Successes

Configure

Label	T	# Samples	Average	Min	Max	Std. Dev.	Error %	Throughput	Received KB/sec	Sent KB/sec	Avg. Bytes
HTTP Request		250	47	40	62	4.14	0.00%	5.1/sec	3.11	0.74	627.0
TOTAL		250	47	40	62	4.14	0.00%	5.1/sec	3.11	0.74	627.0



Aggregate Report

Name:Aggregate Performance Benchmark

Comments:

Write results to file / Read from file

Filename

Browse...

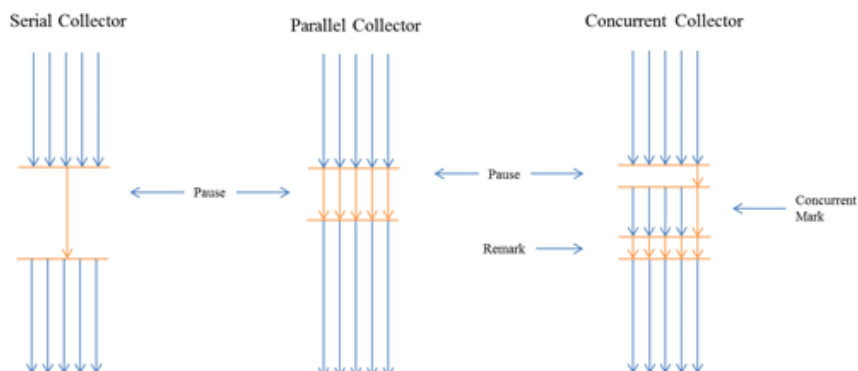
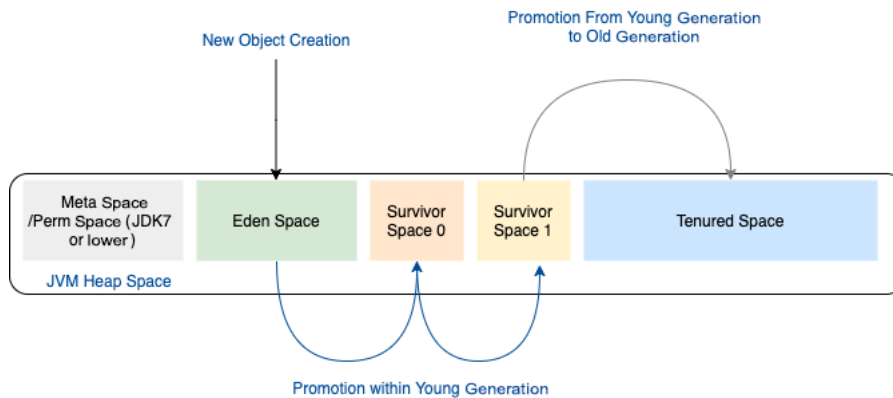
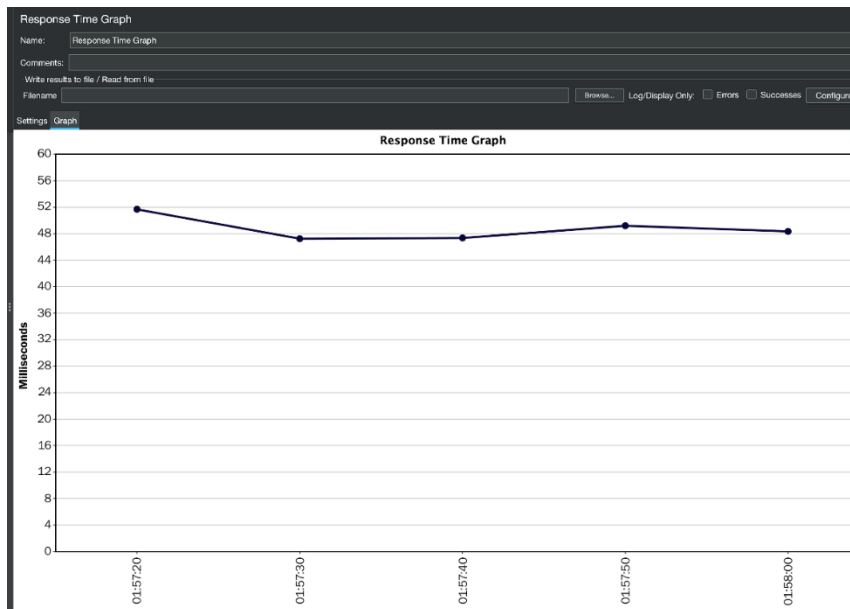
Log/Display Only:

☐ Errors

☐ Successes

Configure

Label	# Samples	Average	Median	90% Line	95% Line	99% Line	Min	Maximum	Error %	Throughput	Received K...	Sent KB/...
HTTP Request	250	47	47	54	58	61	40	62	0.00%	5.1/sec	3.11	0.74
TOTAL	250	47	47	54	58	61	40	62	0.00%	5.1/sec	3.11	0.74
						</						



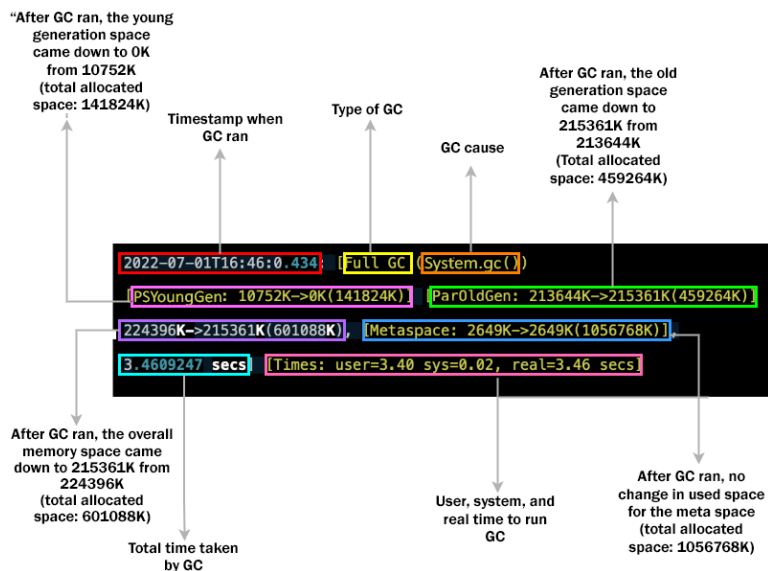
	T	T	T		
	E	E	T	E	
			E	S	S
	T	E	E	S	
	T	T	T	T	



```

2022-07-01T16:45:0.193: [GC (Allocation Failure) [PSYoungGen: 71912K->10752K(141824K)] 101680K->101012K(316928K), 0.3575121 secs] [Times: user=0.22 sys=0.06, real=0.38 secs]
2022-07-01T16:46:0.357: [GC (Allocation Failure) [PSYoungGen: 141832K->10752K(141832K)] 232084K->224396K(359424K), 0.569666 secs] [Times: user=0.45 sys=0.02, real=0.56 secs]
2022-07-01T16:46:0.434: [Full GC (System.gc()) [PSYoungGen: 10752K->0K(141824K)] [ParOldGen: 213644K->215361K(459264K)] 224396K->215361K(601088K), [Metaspace: 2649K->2649K(1056768K)], 3.4609247 secs] [Times: user=3.40 sys=0.02, real=3.46 secs]
2022-07-01T16:49:0.984: [GC (Allocation Failure) [PSYoungGen: 131072K->10752K(190464K)] 346433K->321225K(649728K), 0.1407158 secs] [Times: user=0.07 sys=0.08, real=0.14 secs]

```





Applications

NEW  
NEW SAVING  
SUBMITTED  
ACCEPTED  
RUNNING  
FINISHED  
FAILED  
KILLED

Scheduler

Tools

Applications Metrics

Apps Submitted		Apps Pending		Apps Running		Apps Completed		Apps Failed			
2033		0		6		1964		46			
Show 20 entries											
Node State	Node HTTP Address	Last health-update	Containers	Total Mem	Allocated Mem	Actual Mem Used	Total VCores	Allocated VCores	CPU consumed percent	Is Volatile	Shuffle Running
RUNNING	10.7.197.236:8042	4-Jun-2020 19:35:04	0	112 GB	0 B	0 B	32	0	1.1662779	false	true
RUNNING	10.7.197.127:8042	4-Jun-2020 19:36:49	2	112 GB	14.88 GB	2.33 GB	32	3	5.0208335	false	true
RUNNING	10.7.192.42:8042	4-Jun-2020 19:35:25	8	112 GB	72.63 GB	4.68 GB	32	13	7.3933687	false	true

IMENTS > DEFAULT > CLUSTER\_0 >

iew

Clients

+ New

Producers Consumers Consumer lag

Search consumer groups

Consumer group ID	Messages behind	Number of consumers	Number of topics
analyticsTeam	1,654	1	1

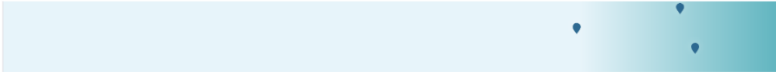
analyticsTeam

5,470  
Total Messages behind

+2 message  
5 second in

Current progress in processing

Maximum lag per consumer  
2,757



10,000

5,000

Client ID	Consumer ID	Partition	Messages behind	Current offset	End offset
--	--	1	1454	4243	5697
--	--	0	2757	7723	10480
--	--	2	1259	3337	4596

## Chapter 12: Evaluating, Recommending, and Presenting Your Solutions

Task no.	Task name	Dependency
1	Creating Git user registration and a master repository	
2	Creating local repositories on a PC	Task 1
3	Creating a simple Hello World whose output will be shown in Hindi in Java	
4	Creating R2	
5	Reviewing the code of R2	Task 4
6	Pushing the code of R2	Task 2 and Task 5
7	Creating a data model of R2	
8	Creating a data model of R3-b	
9	Reviewing the data model of R3-b	Task 8
10	Pushing the data model of R3-b	Task 1 and Task 8

Architectural Decision Matrix

	Architecture 1	Architecture 2	Architecture 3	Architecture 4
Criteria 1				
Criteria 2				
Criteria 3				
Criteria 4				
Criteria 5				

Architectural Decision Matrix

	Architecture 1	Architecture 2	Architecture 3	Architecture 4
	Scale	Scale	Scale	Scale
Criteria 1	3	2	4	4
Criteria 2	3.5	3.5	3	2.5
Criteria 3	4.5	2.5	3	5
Criteria 4	2	4	4	3.5
Criteria 5	2.5	3.5	3	4.5

Architectural Decision Matrix

	Architecture 1	Architecture 2	Architecture 3	Architecture 4
	Scale	Weight	Scale	Weight
Criteria 1	3	2	4	2
Criteria 2	3.5	3	3	3
Criteria 3	4.5	2	3	3
Criteria 4	2	4	4	4
Criteria 5	2.5	4	3	4

Architectural Decision Matrix

	Architecture 1			Architecture 2			Architecture 3			Architecture 4		
	Scale	Weight	Score	Scale	Weight		Scale	Weight	Score	Scale	Weight	Score
Criteria 1	3	2	6	2	2	4	4	2	8	4	2	8
Criteria 2	3.5	3	10.5	3.5	3	10.5	3	3	9	2.5	3	7.5
Criteria 3	4.5	2	9	2.5	2	5	3	3	9	5	2	10
Criteria 4	2	4	8	4	4	16	4	4	16	3.5	4	14
Criteria 5	2.5	4	10	3.5	4	14	3	4	12	4.5	4	18
Total Desirability score	43.5			49.5			54			57.5		

