Big data driving behavior analysis of IoV (Internet of Vehicles)

1. Scenario introduction

IoV system will record the vehicle status at all times, such as Driver ID, License plate number, Latitude, Longitude, Speed, Direction, Rapidly speedup, Rapidly slowdown, Neutral slide, Over speed, Oil leak .etc. of every single time.

For example the data shown in the figure below. Each row of data represents a record. If the value of is 1, it means a violation, and if it is empty, it means that there is no violation in this record (Those data is anonymized and does not contain personal sensitive information).

Driver ID	License plate number	Latitude	Longitude	Speed	Direction	Location	Record report time	Rapidly speedup	Rapidly slowdowr	Neutral slide	Neutral slide finished	Neutral slide time	Overspeed	Overspeed finished	Overspeed time	Fatigue driving	Hthrottle stop	Oil leak	
haowei1000008	A709GB	38.504	106.2094	102	66		1/2/2017 16:00)											
hanhui1000002	AZI419	31.249	121,4862	65	250		1/2/2017 16:00)											
haowei1000008	A709GB	38,505	106.2118	86	66		1/2/2017 16:00)											
hanhui1000002	AZI419	31.248	121.4841	76	250		1/2/2017 16:00)											
haowei1000008	A709GB	38.506	106.2137	65	66		1/2/2017 16:00)											
hanhui1000002	AZI419	31.247	121.4822	73	250		1/2/2017 16:00)											
haowei1000008	A709GB	38.507	106.2169	114	66		1/2/2017 16:00)											
hanhui1000002	AZI419	31.246	121.4804	66	250		1/2/2017 16:00)											
hanhui1000002	AZI419	31.246	121.48	143	250		1/2/2017 16:00)					1						
haowei1000008	A709GB	38.508	106.2199	104	66		1/2/2017 16:00)											
hanhui1000002	AZI419	31.245	121.4769	129	250		1/2/2017 16:00)											
hanhui1000002	AZI419	31.245	121.4766	120	250		1/2/2017 16:00)						1	10)			
haowei1000008	A709GB	38.51	106.2226	96	66		1/2/2017 16:01												1
hanhui1000002	AZI419	31.244	121.4746	83	250		1/2/2017 16:01												Γ
hanhui1000002	AZI419	31.244	121.4724	115	250		1/2/2017 16:01		1	1									I
haowei1000008	A709GB	38.511	106.2252	90	66		1/2/2017 16:01												
hanhui1000002	AZI419	31.243	121.4717	87	250		1/2/2017 16:01												
haowei1000008	A709GB	38.512	106.2279	96	66		1/2/2017 16:01												
hanhui1000002	AZI419	31.242	121.4691	99	250		1/2/2017 16:01												
haowei1000008	A709GB	38.513	106.2299	69	66		1/2/2017 16:01												
hanhui1000002	AZI419	31.242	121.4671	74	250		1/2/2017 16:01												
haowei1000008	A709GB	38.514	106.232	75	66		1/2/2017 16:01												
hanhui1000002	AZI419	31.241	121.465	79	250		1/2/2017 16:01												
haowei1000008	A709GB	38.515	106.2344	82	66		1/2/2017 16:01												
hanhui1000002	AZI419	31.24	121.463	74	250		1/2/2017 16:01												
haowei1000008	A709GB	38.516	106.2374	107	66		1/2/2017 16:02	2											
hanhui1000002	AZI419	31.239	121.4605	94	250		1/2/2017 16:02	2											
haowei1000008	A709GB	38.516	106.2379	94	66		1/2/2017 16:02	2								1	1		
haowei1000008	A709GB	38.517	106.2394	62	66		1/2/2017 16:02	2											
hanhui1000002	AZI419	31.239	121.4582	86	250		1/2/2017 16:02	2											
hoowoi100008	A700CR	29 519	106 2416	74	66		1/2/2017 16:02												

The data definition and description are as follows:

Record	Remark
information	
Driver ID	
License plate	
number	
Latitude	
Longitude	
Speed	
Direction	
Location Name	
Time	
Rapidly speedup	If the value is 1, rapidly speedup occurs.
Rapidly slowdown	If the value is 1, rapidly slowdown occurs.
Neutral slide	If the value is 1, neutral slide occurs.
Neutral slide	If the value is 1, it means neutral slide is finished
finished	
Neutral slide	Neutral slide duration
duration	
Over speed	If the value is 1, over speeding occurs.

Over speed finished	If the value is 1, it means over speeding is				
	finished				
Over speed	Over speeding duration.				
duration					
Fatigue driving	If the value is 1, fatigue driving occurs.				
Idle speed	If the value is 1, idle speed stopping occurs.				
Oil leak	If the value is 1, oil leakage occurs.				

Big data technology can dig out more value from those mass data, for example, for a single driver, how many times rapidly speed, how long time over speed, how many times oil leak, to further conclude the driver behavior, those conclusion can be further referred in security, loan, and other fields as input information again.

This lab will show that, with the IoV data (shared by IoV system by http link) and prepared programs, how to use HUAWEI CLOUD Big Data Service to do the one-stop analysis from end to end with a unified data governance platform.

The final objective of the analysis is to obtain the following data.

Driver ID	License plate number	Rapidly speedup times	Rapidly slowdown times	Neutral slide times	Neutral coasting duration	Over speed times	Over speed duration	Fatigue driving times	ldle speed times	Oil leak times
panxian1000005	AX542C	395	434	330	2930	3531	33946	4307	417	441
shenxian1000004	ADJ750	374	356	297	2810	3126	31494	3767	383	366
zouan1000007	A58M83	360	385	315	2997	3181	31248	3594	389	385
zengpeng1000000	AZQ110	340	344	272	2894	2763	25479	3274	284	337
xiexiao1000001	AEB132	264	261	248	2525	2324	23434	2720	314	253
duxu1000009	AT75H8	238	284	247	2632	2301	22338	2814	264	248
hanhui1000002	AZI419	401	444	327	2844	3349	31813	3997	433	371
likun1000003	AVM936	341	354	291	3043	3044	28728	3552	347	376
haowei1000008	A709GB	321	314	255	2659	2639	25522	3204	312	318
xiezhi1000006	A6CU11	255	310	254	2074	2535	23942	2931	312	279

2. Architecture Overview:



Data Lake Factory (DLF) is a big data platform designed specifically for the HUAWEI CLOUD. It manages diverse big data services and provides a one-stop big data development environment and fully-managed big data scheduling capabilities.

Cloud Data Migration (CDM) implements data mobility by enabling batch data migration among homogeneous and heterogeneous data sources.

Object Storage Service (OBS) is a cloud storage service optimized for storing massive amounts of data. It provides unlimited, secure, and highly reliable storage capabilities at a relatively low cost.

MapReduce Service (MRS) provides enterprise-level big data clusters on the cloud. Tenants can fully control clusters and easily run big data components such as Hadoop, Spark, HBase, Kafka, and Storm.

3. Pre-conditions:

Open the Chrome/Firefox explorer to access the HUAWEI CLOUD login website (www.huaweicloud.com):

/ looodine Login	Account I	Login
------------------	-----------	-------

Account name or email				
Password	Ø			
Mobile Number Login	Remember me			
Log In				
Free Registration	Forgot Password			
IAM User Login	HUAWEI ID Login			
Use Another Account ~				

4. Setup the single services including network, Security Group, OBS, CDM and MapReduce

4.1 Setup the internal network

Log in to the management console, choose "Service List" > "Network" > "Virtual Private Cloud" as follows:

HUAWEI	HUAV	VEI CLOUD	Console	♦ Bangkok	▼
		My Resource	s [Bangkok]		
HUAWEI	CLOUD Conso	le 🛛 Bangkok 🗸 🗸		Search	Q More English APClouddemoTH English
Networ	rk Console	Virtual Private Cloud ⑦			Quick Links Create VPC
Dashboa	ard Private Cloud		All projects	Name 🔻	Q Search by Tag ⊗ C □

Click "Create VPC" in the upper right corner, configure as follows:

[Region]: Bangkok

[Name]: Enter a user-defined name. In this lab, "vpc-driver-behavior" is used as an example.

[CIDR Block]: keep default

Basic Information	
Region	AP-Bangkok 💌
	Regions are geographic areas isolated from each other. Re latency and quick resource access, select the nearest regio
Name	vpc-driver-behavior
CIDR Block	192 · 168 · 0 · 0 / 16
	Recommended: 10.0.0.0/8-24 (Select) 172.16.0.0/12-24 (Select)

Configure Default Subnet as follows:

[Name]: Enter a user-defined name. In this lab, "subnet-driver-behavior" is used as an example.

[Other configuration]: keep default

Click Create Now at lower right corner.

Default Subnet	
Name	subnet-driver-behavior
CIDR Block	192 • 168 • 0 / 24 ? Available IP Addresses: 251 The CIDR block cannot be modified after the subnet has been created.
Associated Route Table	Default ⑦
Advanced Settings 🔻	Gateway DNS Server Address DHCP Lease Time Tag

4.2 Setup Security Groups

In the Virtual Private Cloud console, click "Security Groups" as follows:

≡	Network Console
Ô	
&	Dashboard
,000,	Virtual Private Cloud
Ô	Subnets
	Route Tables
	Access Control
	Security Groups
Ø	Network ACLs

HUAWEI	HUAWEI CLOUD Conso	e o Bangkok 💌	Search Q	More [®] English	APClouddemoTH
≡	Network Console	Security Groups ⑦		📝 Quick Links	Create Security Group
\$	Dashboard		All projects 🔹 Name	Ŧ	QC

Click "Create Security Groups" at upper right corner, configure as follows:

[Name]: Enter a user-defined name. In this lab, "sg-driver-behavior" is used as an example.

[Template]: General-purpose web server

[Other configuration]: keep default

Click OK

Create Security Group

* Name	sg-driver-behavior				
* Template	General-purpose web server 🔹				
Description	The security group is for general-purpose web servers and includes default rules that allow all inbound ICMP traffic and inbound traffic on ports 22, 80, 443, and 3389.The security group is used for remote login, ping, and hosting a website on ECSs.				
0/25 Show Default Rule 🔻					
	OK Cancel				

4.3 Upload the analysis jar file to OBS

Use your test PC, open the Chrome/Firefox explorer, copy and paste the URL: https://sandbox-experiment-resource-north-4.obs.cn-north-4.myhwclouds.com/mrsspark/driver_behavior.jar

Download the file to "D:\Home\user\" folder.

Choose "Service List" > "Storage" > "Object Storage Service" as follows:

HUAWE	HUAWEI CLOUD	nsole			Search	Q More [®] En	nglish A	PClouddemoTH ro_test	🖻
Ξ	Object Storage Service		Object Storage Service ⑦	Open Source Software Notice		Create	e Bucket	Buy OBS Pa	ickage
@ &	Object Storage		OBS Browser+	obsutil 🕹 Download	obsfs 🕹 Download	Get SDK			
AXA.	Parallel File System		Download	A CLI tool. It supports basic operations on buckets and	A tool for mounting parallel file system. It enables you to	Obtain access keys (AK and SK)			
	My Packages Data Express Service	e e	It supports batch upload of large files and folders.	objects. Learn more	operate objects in your local file system.	Visit OBS growth map			
0	CDN	ø	Learn more		Learn more				
	Object Storage Migration								

Click "Create Bucket" in the upper right corner, configure as follows: [Region]: Bangkok [Bucket Name]: Enter a user-defined name. In this lab, "obs-driver-behavior2" is used as an example.

[Other Configurations]: keep default

Click "Create Now" at lower right corner.

< Create Bucket	
Region	AP-Bangkok 👻
(10)	Regions are geographic areas isolated from each other. Resources are region-specific and cannot be used across regions through internal network connections. For low network
	latency and quick resource access, select the nearest region. Once a bucket is created, the region cannot be changed.
Bucket Name	obs-driver-behavior2
	Naming conventions: - The name must be olobally unique in OBS.
	The name of a bucket or parallel file system can be reused 30 minutes after the bucket or parallel file system is deleted. The name must contain 3 to 63 characters. Only lowercase letters, dioits, hvohens (-), and periods (-) are allowed.
	- The name cannot start or end with a period (.) or hyphen (-), and cannot contain two consecutive periods (.) or contain a period (.) and a hyphen (-) adjacent to each other. - The name cannot be an IP address.
	- If the name contains any periods, a security certificate verification message may appear when you access the bucket or its objects by entering a domain name.
Storage Class	Standard Infrequent Access Archive
	Optimized for frequently accessed (multiple times per month) data such as small and essential files that require low latency.
	The storage class of a bucket is inherited by objects uploaded to the bucket by default. You can also change the storage class of an object when uploading it to the bucket. Lear more
Default Encryptio	Enabling default encryption upon bucket creation is supported only in AP-Hong-Kong, AP-Singapore, CN East-Shanghai2, CN North-Beijing1, CN South-Guangzhou, AF-Johannesburg.
Direct Reading	Direct reading of Archive data is supported in the following region: CN North-Beijing1, CN East-Shanghai1, CN East-Shanghai2, CN North-Beijing4, CN South-Guangzhou, .
Buckets are billed	
Bucket creation is free	of charge. You will be billed for the service only after using billable items. Pricing details
lick the nar	ne of the bucket you just created and go to Objects part to upload the jar file as follows:
bject Storage Service	Object Storage / obs-driver-behavior2
Namiou	Objects Deleted Objects Fragments
blacts	
armissions	Objects are basic units of data storage. In OBS, files and folders are treated as objects. Any file type can be uploaded and managed in a bucket. Learn more
asic Configurations	Vinoad Orject Vinoad Orject Vinoad Orject Vinoad Orject Vinoad Orject Vinoad Vino
Jamain Name Mamt	
Optoad Ob	Ject How to upload a File Greater than 5 GB?
Storage Class	Standard Infrequent Access Archive
	Ontimized for frequently accessed (multiple times per month) data such as small and essential files that require low latency.
	The default starses class is the same as that of the hucket. You can change the starses class according to your actual needs.
	more
Upload Object	Note: If the bucket is not versioning-enabled, uploading a file/folder with the name that already exists in the bucket will replace th existing file/folder.
	OBS
	Drag files or folders here to upload. Or add file
	(A maximum of 100 files can be uploaded at a time. The total size cannot exceed 5 GB.)
Encryption	Encrypts the file for secure storage. The encryption status of the encrypted file cannot be changed
Enci y priori	KMS encryption
	Upload Cancel

Create a folder name "input", and click "OK":

Object Storag	ge Service	Object Stora	ge / obs-driver-behavior2	đ								
Overview		Objec	ts Deleted Objects	s Fragments	5							
Objects		Objects a	are basic units of data storage	e. In OBS, files and fo	lders are treated a	s objects. Any f	ile type can be uplo	paded and mar	aged in a b	ucket. Learn more		
Permissions		Uplo	ad Object Create Fol	der Restore	Delete	Change St	orage Class		Enter an o	bject name prefix.	Q	С
Basic Configura	ations 💌		Name ↓Ξ	Storage ↓Ξ	Size ↓Ξ	Encrypt	Restoration St	Last Modifie	d 1∃	Operation		
Domain Name	Mgmt		driver_behavior.jar	Standard	25.74 KB	No		Oct 21, 2020) 22:26:1	Download Share More •		
Folder Name	input Naming rules: - You can creat	e folders witi	h a single level or multiple l	levels.								
	 The name of 	a single-leve	l folder cannot contain the	following characters:	:\:*?"<							
	- The name ca	nnot start or	end with a period (.).									
	- Use single sla	ishes (/) to se	eparate levels of a folder.									
	- The absolute	path of the f	older cannot exceed 1023 c	haracters.								
	- Cannot conta	in two or mo	re consecutive slashes (/).									
		OK	Cancel									

Prepare the AK (Access Key Id) and SK (Secret Access Key) of OBS for later use:

Switch over to Console, click "My Credentials" as follows:

HUAWEI	HUA	WEI CLOUD	Console	Bangkok	v		Sea	arch	Q	More	English	APClouddemoTH Leo_test
		My Resource	es [Bangkok]				View Resources	s in All Regions 🔞	Basic I My Cre	nformation edentials		
යි		Elastic Cloud	l Server	91	Relational Database Service	17	Auto Scaling	5	Identit	y and Acce	ess Manager	ment
00		Bare Metal S	erver	0	Elastic Volume Service	151	Volume Backup Service	0	Tag M	anagemen		
		Virtual Privat	e Cloud	61	Elastic Load Balance	18	Elastic IP	59	Opera	tion Log		
Pro	jects ss key:	Access	Keys	after they are gener	ated. Keep them secure, change t	hem periodically,	and do not share them with a	nyone.				
e) Creat	e Access Key	You can cr	eate 1 more acces	s keys.				En	ter an acci	ess key ID.	Q

If the "Create Access Key" button is grey color, and appear the error message "you can create 0 more access keys". Please delete the existing access key, and create new one.

Input password of HUAWEI CLOUD Account and click "confirm", choose "save file to" then click "confirm".

Remark: Please save the AK/SK file to "D:\home\user" directory, keep "credentials.csv" file safely, it will been used at follow operation.

You can open the credentials.csv file and see:

11	D	0
User Name	Access Key Id	Secret Access Key

4.4 Purchase CDM service

4.4.1 Create CDM cluster

Choose "Service List" > "Migration" > "Cloud Data Migration", as follows:

HUAWEI CLOUD	Bangkok Console Service List Favorites	Q More APClouddem. 🖻 🕮 👁
	Cluster Management @	부 Виу СОМ
CDM	You can create 29 more clusters. Start Stop Restart Delete	Authorize EIP Check Enter a cluster name. Q C

Click Buy CDM in the upper right corner.

Configure basic information as follows:

[Current Region]: AP-Bangkok

[Cluster Name]: Enter a user-defined name. In this lab, "cdm-driver-behavior" is used as an example. [Instance Type]: cdm.large

[VPC]: The one you just created, in this lab, "vpc-driver-behavior" is used as an example.

[Subnet]: The one you just created, in this lab, "subnet-driver-behavior" is used as an example.

[Security Group]: The one you just created, in this lab, "sg-driver-behavior" is used as an example. [Other Configurations]: Keep default

* Current Region	AP-Bangkok 👻				
	Regions are geographic areas isolated connections. For low network latency	d from each other. Resources ar and quick resource access, sel	e region-specific and cannot b ect the nearest region.	e used across regions through internal ne	twork
* AZ ③	ap-southeast-2a ap-sou	theast-2b			
* Cluster Name	cdm-driver-behavior	0			
* Version	2.8.2				
* Instance Type	Flavor	vCPUs/Memory	Assured/Maximum Ba	Concurrent Extractors	
	o cdm.large	8 vCPUs 16 GB	0.8/3 Gbps	20	
	Cdm.xlarge	16 vCPUs 32 GB	4/10 Gbps	100	
	Selected Instance Specifications: cdn	n.large 8 vCPUs 16 GB			
* Quantity	1				
* VPC ③	vpc-driver-behavior	✓ View VPC C			
* Subnet 👩	subnet-driver-behavior(192.168.0.0/	24) 👻			
* Security Group	sg-driver-behavior	✓ View Security Grou	P C		
* Auto Shutdown (?)	Auto Shutdown and Scheduler	d Startup/Shutdown cannot be	enabled at the same time.		
* Scheduled Startup	—				
Running Price: \$0.56 USD	/hour + Hangup Price: \$0.01 USE)/hour			
CDM billing is based on cluster runni when no data is migrated. This price	ng duration. You are advised to shut down or de is an estimate and may differ from the final pric	elete the cluster te in your bill.			Buy Now

After confirming the configuration parameters, click Buy Now in the lower right corner. On the displayed page, confirm the configuration and click Submit to apply for the CMD Service. It takes around 5-10 mins.

4.4.2 Configure CDM to collect data from HTTP (data source) link to OBS(data

destination)

Purchase an internet access (EIP) for CDM first

Choose "Service List" > "Network" > "Elastic IP", as bellowing:

HJAWEI	HUAWEI CLOUD Consol	e Bangkok v Search Q More English APClouddemoTh	
≡	Network Console	EIPs ⑦	Buy EIP
යි	Dashboard	Unbind Modify Bandwidth Release	
.00.	Virtual Private Cloud	All projects v EIP v Q Search by Tag & C	6
6	Route Tables	EIP Moni Status EIP Type 🖓 Bandwidth Bandwidth D Associated Ins Billing Mode Enter Operation	

Click "Buy EIP" at upper right corner.

Configure the basic information as follows

[Billing Mode]: Pay-per-use

[Region]: AP-Bangkok

[Billed by]: Traffic

[Bandwidth]: 100

[Bandwidth Name]: Enter a user-defined name. In this lab, "bandwidth-driver-behavior" is used as an example.

[Other configuration]: Keep default

Click "Next" and "Submit"

per-use oud resources in the same region. The region cannot be changed after the EIP is purchased.
oud resources in the same region. The region cannot be changed after the EIP is purchased.
oud resources in the same region. The region cannot be changed after the EIP is purchased.
service availability rate
Service or unability rate
Traffic For light/sharply fluctuating traffic
ve of usage duration; configurable maximum bandwidth size.
50 100 Custom - 100 + The value ranges from 1 to 300 Mbit/s.
ve

Switch to CDM console and click "Bind EIP" to bind the public IP to CDM to get the internet access.

	Cluster Management [®] Buy CDM
CDM	You can create 28 more clusters. Start Stop Restart Delete Authorize EIP Check Enter a cluster name. Q C
Cluster Management	Cluster Name 🔶 Cluster Status 💠 Internal Network A 💠 Public Network Ad 💠 Operation
Help Center	□ ∨ cdm-driver-behavior ③ Running 192.168.0.104 - Job Management Bind EIP More ▼

Bind	EIP								
Ð	Are you sure you want to bind this EIP?								
				C					
	EIP 🌲	Status 🌻	Bandwidth (Mbit/s)	\$					
0	159.138.243.101	🖉 Unbound	100						
		OK Cance	I						
Click th	e CDM you just	created and c	lick "Job Manage	ment"					

🗓 Delete 📑 Driver Management

🕑 Create Link

	Cluster Management > cdm-driver-behavior							
CDM	Basic Information	Cluster Configuration						
Cluster Management	Cluster Information							
Help Center	Cluster Name:	cdm-driver-behavior		Cluster Management:	Job Management			
	Cluster Status:	Running		Auto Shutdown:	Disabled			
	Number of Nodes:	1		Scheduled Startup/Shutdown:	Disabled			
	Version:	2.8.2		Notification:	Disabled			
Click "Links" and "Cro	eate Link"							
Cluster Management > cdm	-driver-behavior > Links							
Table/File Migration	Entire DB Migration	Scenario Migration	Historical J	lobs Links Agent	ts Settings			

Choose HTTP to connect with data source, and click "Next" at bottom to name it as "cdm-connectorsource", and "Save"

Select Connector						2 Configure
Data Warehouse	Data Warehouse Service	FusionInsight LibrA		Data Lake Insight		
Hadoop	MRS HDFS	MRS HBase		MRS Hive	FusionInsight HDFS	FusionInsight HBase
	FusionInsight Hive	Apache HDFS	The H	ITTP link is used to download a	Apache Hive	Hadoop release version
Object Storage	HUAWEI CLOUD OBS	Alibaba Cloud OSS	single file from the Internet. You are advised to create jobs using APIs.		Amazon S3	Tencent Cloud COS
File System	FTP	SFTP		HTTP	Network Attached Storage	Scalable File Service (SFS Turbo)
Select Connector						Configure
* Name	cdm-connector-source					
* Connector	HTTP	-				
× Cancel	< Previous 🗣 Test 🖪 Sa	ive				

Click "Links" and "Create Link" for "cdm-connector-destination" again:

Table/File M	ligration Entire DB	Migration Scenar	io Migration Histori	cal Jobs Links	Agents Settings	
O Create Link ☐ Delete						
Select Connector					2 Configure	
Data Warehouse	Data Warehouse Service	FusionInsight LibrA	Data Lake Insight			
Hadoop	MRS HDFS	MRS HBase	MRS Hive	FusionInsight HDFS	FusionInsight HBase	
	FusionInsight Hive	Apache HDFS	Apache HBase	Apache Hive	Hadoop release version	
Object Storage	HUAWEI CLOUD OBS	Alibaba Cloud OSS	Qiniu Cloud Object Storage (KODO)	Amazon S3	Tencent Cloud COS	

Put the name, AK SK, and keep other configuration as default. And Click "Save".

Select Connector			2	Conf
* Nama	odm connector destination			
* Connector				
Object Storage Type	HUAWEI CLOUD OBS			
* OBS Endpoint ⑦	obs.ap-southeast-2.myhuaw			
* Port ③	443			
* OBS Bucket Type	Object Storage 👻			
* AK (?)				
* SK ⑦				

Create "Table/File Migration" from "cdm-connector-source" to "cdm-connector-destination" you just created.

Cluster Management + cdm-driver-behavior + Table/File Migration						
Table/File Migration Entire DB Migration Scenario Migration Historical Jobs Links Agents Settings						
O Create Job [™] Delete [™] Export [™] Import Schedule ✓ Job name	QC					
⊙ 🖋 ▷ 📅 < Name ≎ Link Details Cr ≎ Last Execution ≎ Dura ≎ Write S Status Grou Operation						
Enter a group name. Q						

Configure the information as follows:

[Job Name]: Enter a user-defined name. In this lab, "cdm-data-migration" is used as an example. [Source Link Name]: Choose "cdm-connector-source".

[Source URL]:https://driver-behavior-analysis-source.obs.ap-southeast-

3.myhuaweicloud.com/detail_record_2017_01_02_08_00_00|https://driver-behavior-analysis-

source.obs.ap-southeast-3.myhuaweicloud.com/detail_record_2017_01_03_08_00_00|https://driverbehavior-analysis-source.obs.ap-southeast-

3.myhuaweicloud.com/detail_record_2017_01_04_08_00_00|https://driver-behavior-analysissource.obs.ap-southeast-3.myhuaweicloud.com/detail_record_2017_01_05_08_00_00|https://driverbehavior-analysis-source.obs.ap-southeast3.myhuaweicloud.com/detail_record_2017_01_06_08_00_00|https://driver-behavior-analysissource.obs.ap-southeast-3.myhuaweicloud.com/detail_record_2017_01_07_08_00_00|https://driverbehavior-analysis-source.obs.ap-southeast-

3.myhuaweicloud.com/detail_record_2017_01_08_08_00_00|https://driver-behavior-analysissource.obs.ap-southeast-3.myhuaweicloud.com/detail_record_2017_01_09_08_00_00|https://driverbehavior-analysis-source.obs.ap-southeast-

3.myhuaweicloud.com/detail_record_2017_01_10_08_00_00|https://driver-behavior-analysis-

source.obs.ap-southeast-3.myhuaweicloud.com/detail_record_2017_01_11_08_00_00

[Destination Link Name]: Choose "cdm-connector-destination"

[Bucket Name]: Choose the OBS bucket you just created.

[Write Directory]: Choose the input.

[File Format]: Choose Binary

[Other Configurations]: Keep default

Click "Next"



* Job Name	cdm-data-migration			
Source Job Configuration		Destination Job Configuration		
* Source Link Name	cdm-connector-source +	* Destination Link Name	cdm-connector-destin	• +
* Source URL ③	https://driver-behavior-analy	* Bucket Name	obs-driver-behavior2	
Entries Files	Yes No	* Write Directory ③	/input/	
* File Format ③	Binary	* File Format ③	Binary	•
Show Advanced Attributes		Duplicate File Processing Method	Replace	¥
		0		
		Show Advanced Attributes		



Keep the default configuration and Click "Save" as follows:

Configure Basic		Map Field	3 Configure Task
Configure Task	Never		
Group ⑦	DEFAULT - O Add	🖋 Edit 🚡 Delete	
Schedule Execution	Yes No		
× Cancel < Previous	Save de Save and Run		

4.5 Purchase and configure MRS Service

Choose "Service List" > "El Enterprise Intelligence" > "MapReduce Service", as bellowing:

HUAWEI CLOUD Console	e o Bangkok 👻		Search	Q More	English APClouddemoRegion
MRS	Active Clusters ⑦				🔁 Buy Cluster
Clusters		All projects	▼ All statuses	Enter a cluster nam	e/ID. Q Search by Tag 😵 C
Active Clusters	Name/ID	Cluster V Nodes St	atus Billing Mode	AZ	Enterpris Operation
Cluster History		No	o data available.		
Data Connections					
Operation Logs					
Help					

Click Buy Cluster in the upper right corner.

Configure basic cluster information as follows:

[Region]: Bangkok

[Cluster Name]: Enter a user-defined name. In this lab, "mrs-driving-behavior" is used as an example. [Cluster version]: MRS 2.1.0

[Cluster type]: Analysis cluster

K MapReduce Service	Quick Config	Custom Config		
1 Software Configuration	— (2) Configure Hardware ——	3 Set Advanced	Options	
Region	AP-Bangkok Regions are geographic areas isola regions through internal network nearest region. Learn how to selec	ated from each other. Res connections. For low network of a region.	ources are region-spec work latency and quick	ific and cannot be used across resource access, select the
Cluster Name	mrs-driving-behavior)		
Cluster Version	MRS 2.1.0	•		
Cluster Type	Analysis cluster St	treaming cluster	Hybrid cluster	

[Analysis Components] : Select all

Analysis Components

	Name	Version	Description
	Presto	308	An open source distributed SQL query engine.
	Hadoop	3.1.1	A distributed data storage and processing framework for large da
	Spark	2.3.2	A fast and general engine for large-scale data processing.
~	HBase	2.1.1	A scalable, distributed database that supports structured data sto
~	Hive	3.1.0	A data warehouse infrastructure that provides data summarizatio
~	Hue	3.11.0	A component that provides the Hadoop UI capability, which enab
~	Loader	2.0.0	A tool developed based on open source Sqoop 1.99.7, designed fo
~	Tez	0.9.1	An application framework which allows for a complex directed-ac
~	Flink	1.7.0	A framework and distributed processing engine for stateful comp
~	Impala	3.2.0	An SQL query engine for processing huge volumes of data.
~	Kudu	1.9.0	An column-oriented data store.

[Kerberos Authentication] : disable [Username] : admin [Password] : fdtDWS%!\$709 [Confirm Password] : fdtDWS%!\$709 Click Next.

Kerberos Authentication	?	
Username	admin	
Password		The password will be required to log in to the MRS Manager.
Confirm Password	••••••	

The hardware configuration is as follows:

[Billing mode]: pay-per-use

[AZ]: default value

[VPC]: The one you just created, in this lab, "vpc-driver-behavior" is used as an example.

[Subnet]: The one you just created, in this lab, "subnet-driver-behavior" is used as an example.

[Security Group]: The one you just created, in this lab, "sg-driver-behavior" is used as an example. [EIP]: Bind later.

1 Software Configuration	- 2 Configure Hardware (3) Set Advanced Options
Billing Mode	Pay-per-use Yearly/Monthly
AZ	AZ1 AZ2 ⑦
VPC	vpc-driver-behavior View VPC (2)
Subnet Security Group	Auto crea C Manage Security Group () sg-driver-behavior(()
EIP	Bind later

[Cluster node]:

Master: keep default Instance Specifications; Cluster HA: Disabled;

Analysis Core: keep default Instance Specifications; Set the Instance Count to 1, retain the default values for other parameters.

CPU Architecture	x86 Kunpeng		
Cluster Node	Node Type	Instance Specifications	Instance Count
	Master	General computing-plus 8 vCPUs 32 GB c3.2xlarge.4 System Disk High I/O 100 GB x 1 Data Disk High I/O 200 GB x 1	1 Cluster HA
	Analysis Core 🕐	General computing-plus 8 vCPUs 32 GB c3.2xlarge.4 System Disk High I/O 100 GB x 1 Data Disk High I/O 100 GB x 1	- 1 +
	Analysis Task 🕐		\oplus

[Login Mode]: password [Username]: root [Password]: fdtDWS%! \$709 [Confirm password]: fdtDWS%! \$709 Click Next.

Login Mode	Password	Key Pair			
Username	root				
Password		The passw	ord will be require	d during remote l	ogin to the ECS.
Confirm Password	•••••				

The advanced Options is as follows:

[Alarm]: disabled

Retain the default values for other parameters.

1 Software Configuration ——	— 2 Configure Hardware		Set Advanced Op	tions			
Tag	Tag key You can add 10 more tags	to the cluster.	Tag value				
Auto Scaling Back to the previous step to set Task node specifications before configuring an auto scaling policy.							
Bootstrap Action	Name	Exe	Execution Time		Operation		
	Add ⑦						
	Up to 18 scripts can be add	ded for bootstrap	actions.				
Agency	Bind later	MRS_ECS_DEFAU	LT_AGENCY	Available agencies	0		
Alarm	Disable	able					
	Enable this function so that	t O&M personnel	l can use alarm inf	ormation to locate faults.			

After confirming the configuration parameters, click Buy Now in the lower right corner. On the displayed page, confirm the configuration and click Submit to apply for the MRS Service. It takes about 10 to 20 minutes to start the service.

5. Use DLF to schedule the data analysis process

Choose Service List > "El Enterprise Intelligence" > "Data Lake Factory", and load the jar file from OBS as bellowing:

HUAWEI CLOUD	Data Development 🔹 📀 Bangkok 🔹 Console Service List 🔹 Favorites Workspace Leo_test 🔹 🛇 Q. More [*] APClouddem 🔊 💬
=0	
Se Dashboard	Manage Resource
Connection ^	
Manage Connection	
Development ^	Enter a keyword. Q 💿 🗇 Create Resource 🔮
Develop Script	Name Time & Storage Bo & Depend Create Created On & Depending Operation
🗞 Develop Job	Resource Directory Resource Directory Intell Type + Storage Fa+ Dependum Create Create Operation Operation
Monitoring ^	
Overview	
🖅 Monitor Job	
E Monitor Instance	No records found.
Monitor PatchData	
Manage Notification	
Manage Backup	
Configuration ^	
Manage Host Connection	
📰 Manage Resource 🔮	
Specifications	

Configure the information as follows:

[Name]: Enter a user-defined name. In this lab, "jar-driver-behavior" is used as an example.

[Type]: jar

[Resource Location]: obs

[Main JAR Package]: choose the jar file you uploaded previous.

[Other Configurations]: Keep default

Click "OK".

Create Resource



Configure the basic information as follows:

[Name]: Enter a user-defined name. In this lab, "job-driver-behavior" is used as an example.

[Other Configurations]: Keep default

Click "OK".

Create Job

A maximum of 10000 jobs can be created. You can create 10000 more jobs.

* Job Name	job-driver-behavior
* Processing Mode	• Batch processing
* Creation Method	Create Empty Job Create Based on Template
* Select Directory	/Job Directory/ ····
Owner	Owner
Priority	O High ○ Medium ○ Low
Agency 🕜	Select an agency
* Log Path	obs://dlf-log-0581b95a0b8010e32f81c015009f6587/
	 I agree to create OBS bucket obs://dlf-log-0581b95a0b8010e32f81c015009f658 7/. This bucket is used only for storing run logs of DLF jobs. To change the log path, go to the OBS bucket configuration page.

5.1 Draw your data flow from CDM to MRS Spark by dragging the icon on the graphical console



Click the icon of CDM and load the information to DLF as follows:

 $\label{eq:constraint} \ensuremath{\mathsf{[CDM Cluster Name]: Choose the CDM cluster you created previous.} \ensuremath{\mathsf{a}}$

[CDM job name]: Choose the CDM "Table/File Migration" job you created previous

[Other Configurations]: Keep default



Click the icon of MRS to load the information to DLF and configure spark job as follows:

[MRS Cluster]: Choose the MRS cluster you created previous.

[Spark job name]: Enter a user-defined name. In this lab, "spark-driver-behavior" is used as an example. [JAR Package]: Click ... and choose the jar file you configured on DLF

[Program Parameter]: key is "--class", value is "com.huawei.bigdata.spark.examples.DriverBehavior"



[Jar File Parameters]:AK SK 1(Attention: every parameter need use "space" to separate) [Input data path]: Click OBS to choose the "input" directory on the OBS bucket you created previous. [Output data path]: obs://obs-driver-behavior2/output/ (Type manually)

[Node Name]: Enter a user-defined name. In this lab, "mrs-spark-driver-behavior" is used as an example.

> 💦 * job-driver-beha	× +	
Node Library 💙	🖾 Save ▷ Test 🖸 Submit 🖉 Clear 🎞 Full Screen 🛛 🖨 Export 😳 Refr	esh 🔄 Monitor
Data Integration 🔻		
	Enter a node name. Q 😔 🔾 🖄 🤻 🖧 🗍	Program Parameter ③
CDM Job DIS Stream		class com.huawei.bigdata.spark.exa
6 🗖		lar File Deremetere
DIS Dump DIS Client	\sim	Jai File Parameters
	spark	
Rest Client Import GES	odm data migration	
& 	ior	
MRS Kafka Kafka Client		63/2047
Compute & Analytics 🔻		Input Data Path Click OBS to choose the "input" directory
<u>a</u>		obs://obs-driver-behavior2/input/ × HDFS OBS
DLI SQL DLI Spark		Output Data Path Type as below manually
Sport		
DWS SQL MRS Spark SQL		obs://obs-driver-behavior2/output涨 HDFS OBS
soort"		_
MRS Hive MRS Spark		Node Name *
C DA		mrs-spark-driver-behavior
MRS Spark MRS Flink		

Connect the CDM and MRS as follows:



5.2 Click Save and Submit to run the process automatically.



> 🛃 job	driver-behavi	× +				
Node	Node Library 🗸 💾 Save Þ Test 🙆 Stop 🖉 Clear 🎞 Full Screen 🕒 Export 😳 Refresh					
Data Integra	ition 💌					
CDM Job	X DIS Stream	Information	×	iame.	Q	
DIS Dump	DIS Client	The job is scheduled successfully. You can go to the O&M center to view the running progress.)			
Kest Client	Kafka Client	Go to O&M Center Cancel				
Compute &	Analytics 🔻					
Cau DLI SQL	DLI Spark					

You could find the task is running following the process.

=	Job Monitoring > job-driver-behavior O Normal > Submit (a) Stop	🔃 Pause 🄄 Restore 😹 Edit		Auto Refresh 5 s 🔹 C
	cdm-data-migration mrs-spark-driver-behavior			
Ē	Job Instances			- 8 8
10	Stop Rerun Continue Succeed Manual Run		2020.10.16 - 2020.10.22	All statuses 👻 C
	Status Running T 🖓 Planned Start Time	e Actual Start Time E	ind Time Running Dura	Operation
~	A S Running Normal Sche Oct 22, 2020 00:37	7:24 Oct 22, 2020 00:37:25	1.0	Stop Rerun View Wai More 🔻
	Name Type Status Ri	unning Dura Start Time	Retry Count Error Message	Operation
	mrs-spark-driver-beh Spark 🧿 Running	Oct 22, 2020 00:38:1	. 0	View Log Manual Retry More -
	cdm-data-migration CDM Job 🥑 Run successfully 0.	.73 Oct 22, 2020 00:37:2	. 0	View Log Manual Retry More -
	Job Monitoring > job-driver-behavior I Stopped > Submit I Stop	0 🖲 Pause 🗿 Restore 🛃 Edit		Auto Refresh 5s 🔹 🤇
	cdm-data-migration mrs-spark-driver-behavior			
0	Job Instances			- 8 E
	Stop Rerun Continue Succeed Manual Run		2020.10.16 - 2020.10.22	🛱 All statuses 👻 C
	Status Running T 🖓 Planned Start Time	e Actual Start Time	End Time Running Dura	Operation
	Run successfully Normal Sche Oct 22, 2020 00:37	7:24 Oct 22, 2020 00:37:25	Oct 22, 2020 00:39:37 2.2	Stop Rerun View Wai More 🔻
	Name Type Status	Running Dura Start Time	Retry Count Error Messag	e Operation
	cdm-data-migration CDM Job 🥑 Run successfully	0.73 Oct 22, 2020 00:3	7:2 0	View Log Manual Retry More
	mre-enark-drivar-heb Snark 🔗 Run successfully	1.45 Oct 22.2020.00:3	8:1 0	View Log Manual Betry More

6. Check result on OBS bucket

≡	Object Storage Service	Object Storage / obs-driver-behavior2
ے &	Overview	Objects Deleted Objects Fragments
~~~	Objects	Objects are basic units of data storage. In OBS, files and folders are treated as objects. Any file type can be uploaded and managed in a bucket. Learn more
7.0.1	Permissions	Upload Object         Create Folder         Restore         Delete         Change Storage Class         Enter an object name prefix.         Q         C
	Basic Configurations	Name JΞ Storage JΞ Size JΞ Encrypt Restoration St Last Modified J.F Operation
0	Domain Name Mgmt	□ 🗗 output Share   Copy Path   More ▼
Ð	Cross-Region Replication	🗌 🖻 input Share   Copy Path   More 🕶
Ø	Image Processing	driver_behaviorjar Standard 25.74 KB No Oct 21, 2020 22:26:1 Download Share More ▼
$\Phi$	Inventories	
$\equiv$	Object Storage Service	Object Storage / obs-driver-behavior2 / output
٢		
ß	Overview	Objects Deleted Objects Fragments
0.0	Objects	Objects are basic units of data storage. In OBS, files and folders are treated as objects. Any file type can be uploaded and managed in a bucket. Learn more
7.07	Permissions	Upload Object         Create Folder         Restore         Delete         Change Storage Class         Enter an object name prefix.         Q         C
	Basic Configurations	Name ↓Ξ         Storage ↓Ξ         Size ↓Ξ         Encrypt         Restoration St         Last Modified ↓F         Operation
0	Domain Name Mgmt	← Back
6	Cross-Region Replication	
Ð	Image Processing	part-00000-7e40ec1e-bcf2-4dbd-8950-e Standard 622 byte No Oct 22, 2020 00.39.3 Download Share More 🕶
$\Delta$	Inventories	

## Click Download to download result to your test PC "/Home/user"

panxian1000005	AX542C	395	434	330	2930	3531	33946	4307	417	441
shenxian1000004	ADJ750	374	356	297	2810	3126	31494	3767	383	366
zouan1000007	A58M83	360	385	315	2997	3181	31248	3594	389	385
zengpeng1000000	AZQ110	340	344	272	2894	2763	25479	3274	284	337
xiexiao1000001	AEB132	264	261	248	2525	2324	23434	2720	314	253
duxu1000009	AT75H8	238	284	247	2632	2301	22338	2814	264	248
hanhui1000002	AZI419	401	444	327	2844	3349	31813	3997	433	371
likun1000003	AVM936	341	354	291	3043	3044	28728	3552	347	376
haowei1000008	A709GB	321	314	255	2659	2639	25522	3204	312	318
xiezhi1000006	A6CU11	255	310	254	2074	2535	23942	2931	312	279

Note: The above data are in order License plate number, Rapidly speedup times, Rapidly slowdown times, Neutral slide times, Neutral coasting duration, Over speed times, Over speed duration, Fatigue driving times, Idle speed times, Oil leak times