

Qu Yuan on the cloud construction document

First, architecture diagram

Data entry

Kafka: Log behavior data

Mysql: User data

Processing engine

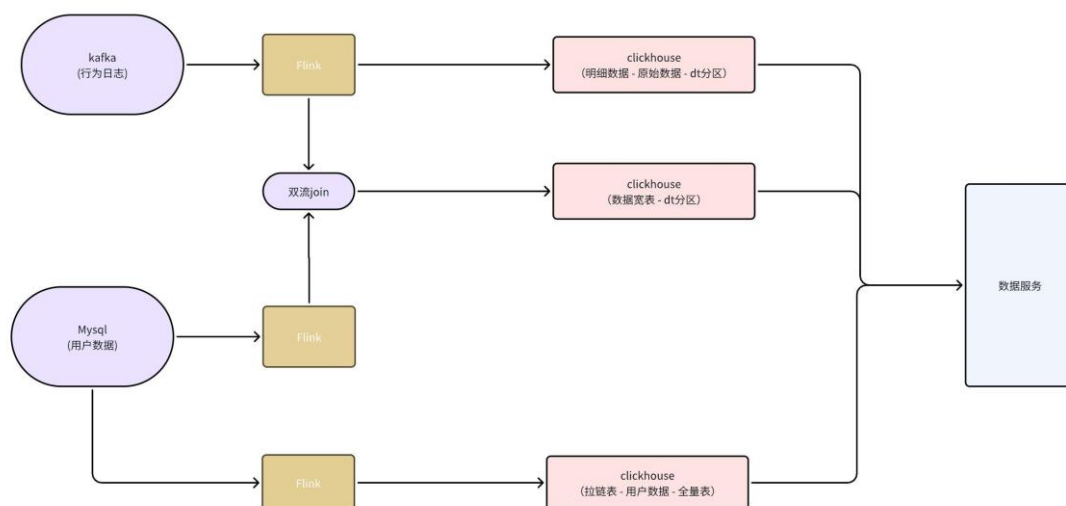
Flink: Supports stream batch computing

Storage database

Clickhouse: Offline real-time, multiple library table engine, good at indicator statistics

Data Output

Springboot: Microservices



2. Modules

1. Number of tourists

Current Online: Number of users + visitors currently online, in real time

Number of **visitors Today**: **Number** of visits that day, pv, updated every hour

Views of the week: Views of the week, updated every hour

Number of visitors this month: Number of visits in the month pv, updated every hour

Custom cycle pv: dt partition, daily data statistics (at least in days)

2. User portrait

Age: Ratio by age group

Sex: Male to female sex ratio

Region: Number of users in provinces and cities nationwide (the same province that registered the most recent 10 visited IP)

Occupation: Proportion by occupation, general users, teachers, students, scholars, others

3. Core indicators

Number of visitors Trend: The number of visits in a customized statistical period is PV, at least in days

Registered User Trend: Indicates the number of registered users in a user-defined statistical period, expressed in at least one day

4. Source Trend

Search engine: user-defined statistical cycle through the search engine to jump to the number of Qu Yuan on the cloud, at least in days

Direct access: The number of times of direct access to Qu Yuan on the cloud through the domain name of the custom statistical period, at least in days

External links: The number of visits to Qu Yuan on the cloud through other external links of the custom statistical period, at least in days

Custom source: Custom statistical cycle does not belong to the above 3 kinds of access to the number of Qu Yuan on the cloud, at least in days

5. Source proportion

Source proportion: Indicates the proportion of search engine, external link, direct access, and source in the customized statistical period

6. Ranking

Location ranking: User-defined statistical period, the name and number of the area with the top number of location visits, at least in days

Interior ranking: Custom statistical period, the name and number of areas with the top number of interior visits, at least in days

Digital ranking: User-defined statistical period: The name and number of the area with the highest number of visits by all digital people, at least in days

Exhibit ranking: Custom statistical period, the name and number of areas with the top number of visits of all exhibits, at least in days

lii. Table

1. User table

Table name: MYSQL: `test.cloud_quyuan_user_test`

Update period: real time

Processing logic: Modify, delete, add

Field name	Field description	Data type	Non-null	Sample data
user_id	User ID (primary key)	INT	Y	1
user_name	Username	VARCHAR	Y	Zhang Fei
gender	Gender	INT	Y	male
age	age	INT	N	48
address	District	TEXT	N	Szechuan
occupation	Occupation	VARCHAR	N	Students
create_time	CREATE_time	DATE	Y	2023-10-13 09:54:02
upadte_tim	Update time	DATE	Y	2023-10-13 09:54:02

e				
---	--	--	--	--

2. Behavior Journal Form (ods)

Table name: CLICKHOUSE: `quyuan.quyuan_ods_src_behavior_log`

Data source: KAFKA: `cCloud_quyuan_test`

Partition: `dt`

Sort key: `Module, name`

Table Engine: Distributed Table Engine + ReplicatedMergetree Table Engine (2 copies)

Update cycle: Real time

Processing logic: original data into the database

Field name	Field description	Data type	Non-null	Sample data
location	Location (front/back)	String	Y	Front desk
model	Modules	String	Y	Landing page
name	name	String	Y	User logged in successfully
describe	Describe	String	N	Click the login button. After verification, login is successful
action	Type of action (click/stay)	String	N	Click
content	Data content	String	N	{ "info":" Login successful ", "ID":12, "time":"202310130954" } json format, different module names correspond to different content on the cloud Qu yuan management background field table

timestamp	Timestamp	DateTime	Y	2023-10-13 09:54:02
dt	Date (partition)	Date	Y	2023-10-13

3 Behavior Journal Table Wide Table (dws)

Table name: CLICKHOUSE: `quyuan.quyuan_dws_behavior_log`

Data source: KAFKA: `cloud_quyuan_test`

MYSQL: `test.cloud_quyuan_user_test`

Partition: `dt`

Sort key: `Module, name`

Table Engine: Distributed Table Engine + ReplicatedMergeTree Table Engine (2 copies)

Update cycle: Real time

Processing logic: data cleaning, dimension table association

Field name	Field description	Data type	Non-null	Sample data
location	Location (front/Back)	String	Y	Front desk
model	Modules	String	Y	Landing page
name	Name	String	Y	User logged in successfully
describe	Describe	String	N	Click the login button. After verification, login is successful
action	Type of action (click/stay)	String	N	Click
content	Data content	String	N	{ "info": "Login successful", "ID": 12, "time": "202310130954" }

				json format, different module names correspond to different content on the cloud Qu yuan management background field table
user_name	Username	VARCHAR	Y	Zhang Fei
gender	Gender	INT	Y	male
age	Age	INT	N	48
address	District	TEXT	N	Szechuan
occupation	Occupation	VARCHAR	N	Students
timestamp	Timestamp	DateTime	Y	2023-10-13 09:54:02
dt	Date (partition)	Date	Y	2023-10-13

4. User Information Form (ods)

Table name: CLICKHOUSE: `quyuan.quyuan_ods_src_user_info`

Source table: MYSQL: `test.cloud_quyuan_user_test`

Partition: Unpartitioned

Sort key: User ID

Table engine: Distributed table engine + ReplicatedVersionedCollapsingMergetree table engine (2 copies)

Update cycle: Real time

Processing logic: Modify, add

Field name	Field description	Data type	Non-null	Sample data
user_id	User ID (primary key)	INT	Y	1
user_name	Username	VARCHAR	Y	Zhang Fei

gender	Gender	INT	Y	male
age	Age	INT	N	48
address	District	TEXT	N	Szechuan
occupation	Occupation	VARCHAR	N	Students
create_time	Creation time	DATE	Y	2023-10-13 09:54:02
end_time	Effective time	DATE	Y	2099-12-31 59:59:59

Four, API interface

<http://192.168.99.75:51101/doc.html>