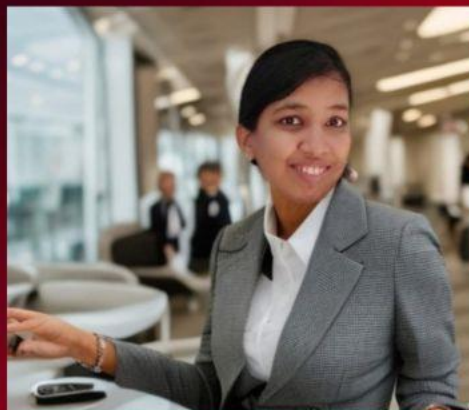


Lightning Talk

Data Experience & Tools

How Agoda Empowered Employees to
Build Efficient Data Pipelines



**SWATI
JAIN**

Engineering Manager
Agoda

BANGKOK
30 OCTOBER 2024
HOTEL NIKKO BANGKOK
#DDSUMMIT



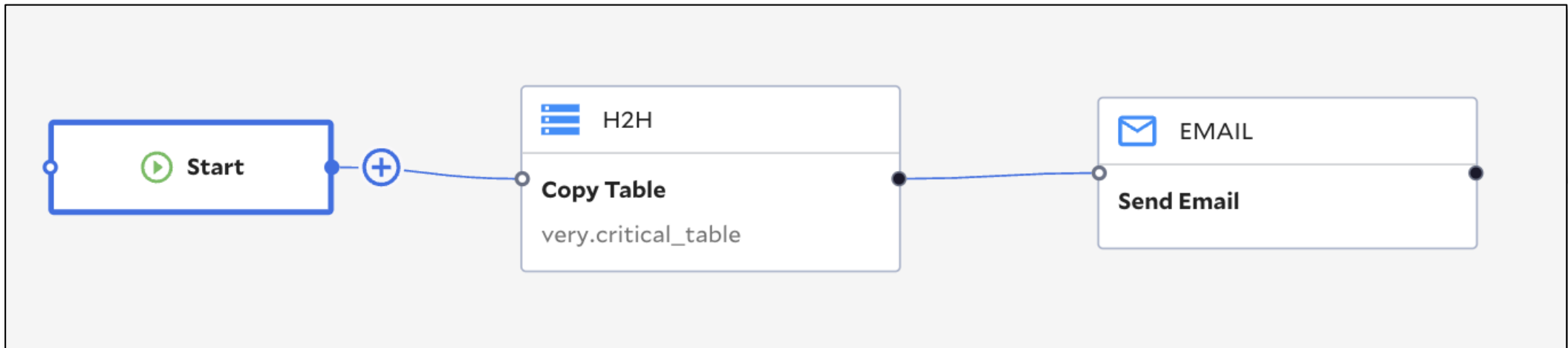
Agoda Workflow Editor

30 Oct 2024

Swati Jain

What is Workflow Editor?

1. **Workflow** : Sequence of tasks performed in particular order
2. **Action** : One Task eg: Copy Data , Send Email etc
3. **Scheduler** : Executes workflow at defined times



Why a Workflow Editor?

- No Code / Low Code Platform
- ANYONE (including non-Tech) can easily build & manage complex data pipelines
- Powerful and Flexible enough for Developers who need more advance features



3 Years Ago

Old Workflow Editor

The screenshot displays the Oozie Editor interface. At the top, the title bar reads "Oozie Editor" and "Workflows". Below this, a navigation bar includes "Agoda Tools" and several icons. A dropdown menu titled "Editors" is open, showing options for "Workflows", "Coordinators", and "Bundles". The main workspace is currently empty. A configuration panel for an "ETL SQL to Hive" job is visible, featuring an "Input" section with a "Connection" dropdown and a "Query" section with a text area for the query string and a "Query file" input field. The query string area contains an example query: "Example Query String -- SQL to Hive query need \$CONDITION SELECT client_ip, rec_created_when F".

Rethinking Workflows in Agoda.. Objectives

- Developer happiness
- Change Management
 - Version Control and Rollback
 - Approvals and Change Management
- Custom IDE
 - Query Unit Testing
 - Test runs
 - Easy to transition from Development to Production

Resolving Core Formats and Interfaces

Workflows Definition: Document or Code ?



- Easy to transform
- Easy to learn
- Validatable
- Code is environment dependent

```
workflow.yml 1.15 KiB
1 workflow:
2   id: asia-test-1
3   name: asia-test-1
4   description: ''
5   version: 2
6   actions:
7     - id: test-1-hjif
8       name: test-1
9       type: h2h
10      spec:
11        query: test-1-hjif.query.sql
12        targetTable: '{{resolve_table_name("asia_test_test_3')}}'
13        targetDatabase: etl_test
14        partitionBy:
15          staticValue:
16            columns:
17              - name: master
18                type: string
19                value: master
20        enableValidationRule: true
21        trackingColumns: []
22        mergeBy:
23          overwritePartition: {}
24        advancedETLParameters:
25          - key: etl.prop.path
26            value: /user/hk-hdp-apps-svc/etl/qa/config/latest/etl-
27        additionalSparkConfigurations:
28          - key: spark.executor.extraJavaOptions
29            value: >-
30              -Detl.prop.path=/user/hk-hdp-apps-svc/etl/qa/config/
31        dependsOn: []
```



```
if is_debug_workflow:
    debug_workflow_path = os.path.join(temp_dir, 'deb

try:
    str_xml = open(debug_workflow_path, "r", encod
except Exception as e:
    print('Cannot parse debug workflow: ' + str(e)
    str_xml = None

if str_xml:
    workflow_xml = ET.fromstring(str_xml)
    if workflow_xml:
        for element in workflow_xml.iter('Node'):
            try:
                if element.find('GuiSettings').at
                    debug_text_box = element.find
                    original_workflow = debug_tex
                    debug_tex

                    workflow_dir = re.search(r'(.

            except Exception:
                pass
print('workflow_dir: ' + str(workflow_dir))

code_path = os.path.join(workflow_dir, folder, script
exec(open(code_path).read(), globals())
```

Workflows Development : Web IDE vs Local IDE?

- No learning git commands
- No Laptop Instructions
- Customizable experience
- Validations and immediate feedback
- Placeholders/hints



UNIFIED DATA PORTAL | etl-tester | Build | Query & Visualize | Discover

Workflow: test

Select a workflow: description_test

Branch: read-only | main

Review and Deploy | Save Changes

ACTIONS | PARAMETERS | JOBS

Your Actions: H2H test

Destination table: etl_test

Table name: col_desc_test

Suggested table

Branch	Resolved Table Name
main	col_desc_test

DIAGRAM | PROBLEMS (2) | TESTS

Documentation | Help?



```
workflow.yml 1.15 KiB
1 workflow:
2   id: asia-test-1
3   name: asia-test-1
4   description: ''
5   version: 2
6   actions:
7     - id: test-1-hjif
8       name: test-1
9       type: h2h
10      spec:
11        query: test-1-hjif.query.sql
12        targetTable: '{{resolve_table_name("asia_test_test_3')}}'
13        targetDatabase: etl_test
14        partitionBy:
15          staticValue:
16            columns:
17              - name: master
18                type: string
19                value: master
20        enableValidationRule: true
21        trackingColumns: []
22        mergeBy:
23          overwritePartition: {}
24        advancedETLParameters:
25          - key: etl.prop.path
26            value: /user/hk-hdp-apps-svc/etl/qa/config/latest/etl-
27        additionalSparkConfigurations:
28          - key: spark.executor.extraJavaOptions
29            value: >-
30              -Detl.prop.path=/user/hk-hdp-apps-svc/etl/qa/config/
31        dependsOn: []
```

Workflows Storage: Gitlab vs Database?

- Universal source of truth
- Change Management
- Approval Flow
- CI/CD tooling



main v etl-tester / + v

Merge branch 'new-feature' into 'main' ⋮
 Good-Dev authored 1 hour ago

Name	Last commit
workflow 1	fix: mismatch wf ids
workflow 2	fix: mismatch wf ids
workflow 3	Merge feature/asia-test

uuid	type	data
e47b86ff-763d-4903-af1a-7bac64d132de	oozie-workflow2	{ "history": { "oozie_id": "0301671-220210181419941-oozie-oozi-W", "properties":
a2104577-cc50-4e43-8a18-b30c29a04e7	oozie-workflow2	{ "history": { "oozie_id": "0072171-220223133615320-oozie-oozi-W", "properties":
bf0f9606-5deb-4d7e-8009-c69487a9d8cb	oozie-workflow2	{ "history": { "oozie_id": "0312177-220223133615320-oozie-oozi-W", "properties":
af9c1743-d48f-4bcc-a7b9-ddb2232492c5	oozie-workflow2	{ "history": { "oozie_id": "0000012-190710180518916-oozie-oozi-W", "properties":
60d1762c-beff-4f76-bbfe-be8bc88afeb2	oozie-workflow2	{ "history": { "oozie_id": "0041345-200114134659416-oozie-oozi-W", "properties":
e705b9df-f988-490f-b5fc-d88d5014d05e	oozie-workflow2	{ "history": { "oozie_id": "0000262-200129143240348-oozie-oozi-W", "properties":
9a7bd718-6b57-4b22-9d8c-72e0115c6de6	oozie-workflow2	{ "history": { "oozie_id": "0091332-201208154137433-oozie-oozi-W", "properties":
d42a1945-b1da-4cc8-a153-7f97ab14f0c7	oozie-workflow2	{ "history": { "oozie_id": "0000013-190710180518916-oozie-oozi-W", "properties":
2d58b66e-73a1-47a7-9f85-b687b4f07238	oozie-workflow2	{ "history": { "oozie_id": "00454074-220210181419941-oozie-oozi-W", "properties":
27777777-7777-7777-7777-777777777777	oozie-workflow2	{ "history": { "oozie_id": "0072987-220223133615320-oozie-oozi-W", "properties":
2-72727-4e62-8e8c-99c73afea725	oozie-workflow2	{ "history": { "oozie_id": "0000499-190711134246693-oozie-oozi-W", "properties":
a0a0f17-8526-4810-999f-aae7271f3d5d	oozie-workflow2	{ "history": { "oozie_id": "0000010-190627113630766-oozie-oozi-W", "properties":
e0060e35-f182-4970-b57d-95f48bad38e3	oozie-workflow2	{ "history": { "oozie_id": "0390581-201208154137433-oozie-oozi-W", "properties":
dce8e203-e527-467f-9edc-639ed9f6e7ef	oozie-workflow2	{ "history": { "oozie_id": "0000011-190627113630766-oozie-oozi-W", "properties":
76c90f9d-2fbc-415f-8918-232e7b60397f	oozie-workflow2	{ "history": { "oozie_id": "0394622-210208105406247-oozie-oozi-W", "properties":
e73bff29-c288-4d38-af31-4cadeb66b413	oozie-workflow2	{ "history": { "oozie_id": "0205718-220210181419941-oozie-oozi-W", "properties":
30cc1991-dbee-45e0-9d7b-63d7144afa29	oozie-workflow2	{ "history": { "oozie_id": "0000036-190617161548293-oozie-oozi-W", "properties":
fbfb6269-f818-4586-9c4e-e5a303731842	oozie-workflow2	{ "history": { "oozie_id": "0000012-190627113630766-oozie-oozi-W", "properties":
7e35d98d-bba3-4b3a-9732-8ffdf70b2f9d3	oozie-workflow2	{ "history": { "oozie_id": "0061242-210208105406247-oozie-oozi-W", "properties":
532422fe-d978-4a26-8f73-750bf624d189	oozie-workflow2	{ "history": { "oozie_id": "0301783-220210181419941-oozie-oozi-W", "properties":
00080e9b2-5849-404c-bd99-5c99fffe08ce	oozie-workflow2	{ "history": { "oozie_id": "0000037-190617161548293-oozie-oozi-W", "properties":
4f95e66c-2710-4735-8342-2c0c4bce74dd	oozie-workflow2	{ "history": { "oozie_id": "0000013-190627113630766-oozie-oozi-W", "properties":
4bc6213f-d252-4693-ab2d-7a81279c3ed3	oozie-workflow2	{ "history": { "oozie_id": "00082678-210421152957392-oozie-oozi-W", "properties":
85bb9559-cbcd-4844-b905-2bca2315d13f	oozie-workflow2	{ "history": { "oozie_id": "0000038-190617161548293-oozie-oozi-W", "properties":
562dac62-95fd-4d37-87bb-c1b43adf9809	oozie-workflow2	{ "history": { "oozie_id": "0000014-190627113630766-oozie-oozi-W", "properties":
f361605c-f50a-40f2-b29a-83e6e4dc62c6	oozie-workflow2	{ "history": { "oozie_id": "3147086-210625140657362-oozie-oozi-W", "properties":
b3779423-3184-4e78-8881-918d6ecec13	oozie-workflow2	{ "history": { "oozie_id": "0000039-190617161548293-oozie-oozi-W", "properties":

Development and Production Environments

- Environment selection based on branch
- Testing Enforced before Deployment to Production

The screenshot displays a workflow management interface. On the left, a sidebar shows workflow selection (atom-how-to-test-your-workflows) and branch selection (feature/atom-how-to-test-your-workflows..). Below this are 'Review and merge' and 'Commit change' buttons. The main area shows a 'quick' workflow with a 'Target table' section. The 'Destination table' configuration includes a database name 'etl_test' and a table name template: `{{resolve_table_name("very_important_table")}}`. A table below shows the resolved table names for different branches: 'current' maps to `very_important_table__feature_atom_how_to_test_your_workflows_initial_dev` and 'main' maps to `very_important_table`. An inset window shows test results for 'my-test-1', indicating that all tests passed.

Workflow

Select a workflow

atom-how-to-test-your-workflows

Branch

Select a branch

feature/atom-how-to-test-your-workflows..

Review and merge

Commit change

ACTIONS PARAMETERS JOBS

Your Actions

H2H quick

quick

Target table

Destination table

Database name

etl_test

Table name

`{{resolve_table_name("very_important_table")}}`

Branch	Resolved Table Name
current	<code>very_important_table__feature_atom_how_to_test_your_workflows_initial_dev</code>
main	<code>very_important_table</code>

DIAGRAM PROBLEMS (1) TESTS

(0/1) tests running in progress...

(1/1) tests running in progress...

PASSED: [h2h-sql-ut] my-test-1

Tests Summary

ALL TESTS PASSED (1/1)

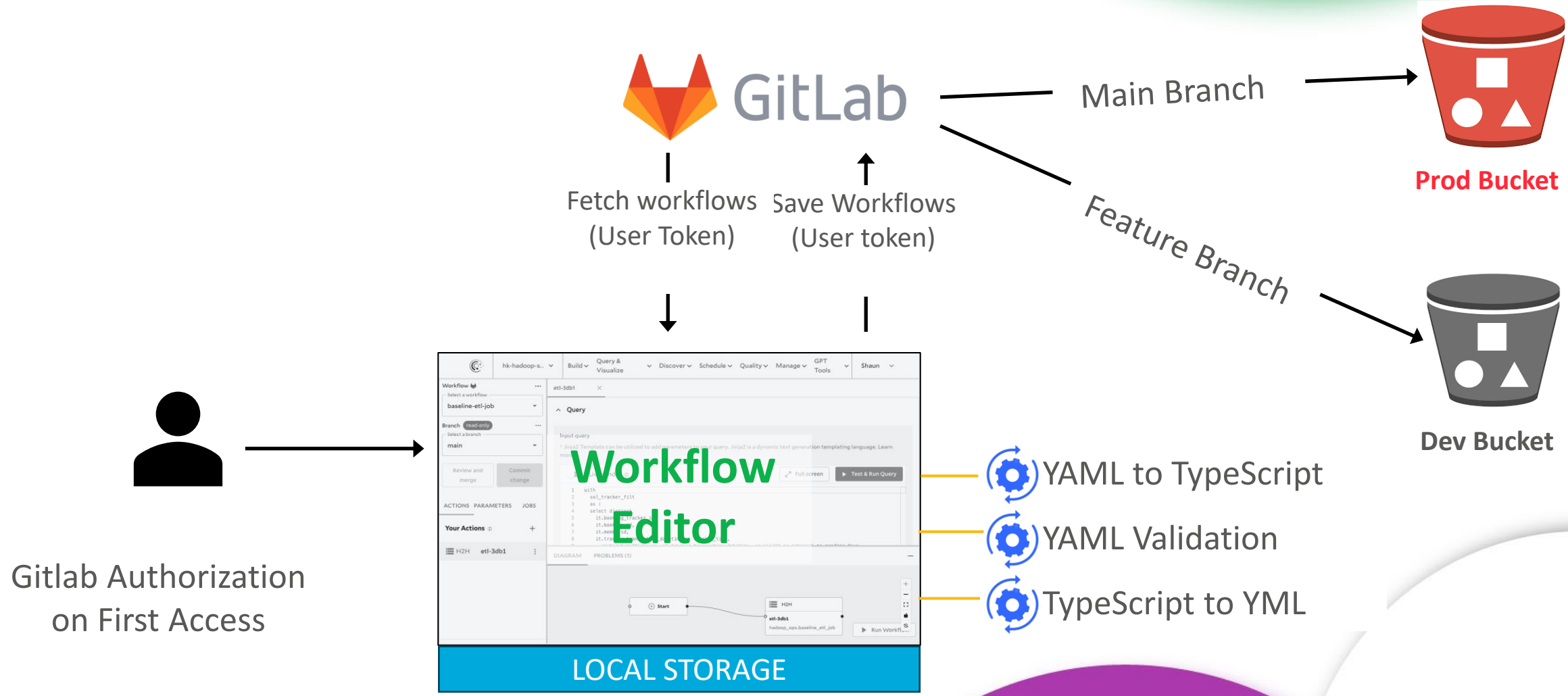
PASSED: [h2h-sql-ut] my-test-1

Final Decisions :

- ✓ Document
- ✓ Web IDE and Local Both
- ✓ Gitlab
- ✓ Environments : Branch based

User Flows and Tech Stack

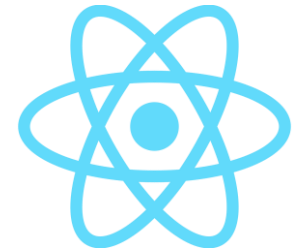
User Flows : Simplified



Tech Stack of Workflow Editor

1. NextJS and React (Framework)
2. Valtio (State Management)
3. Material UI
4. Python + FastAPI (CI/CD)
5. Zod Schema
6. ReactFlow (Workflow Diagram)
7. No Backend (Preferably)

NEXT.js



valtio

Agoda's Workflow Editor Today

Agoda's Workflow Editor

The screenshot displays the Agoda Unified Data Portal interface for editing a workflow. The top navigation bar includes the 'UNIFIED DATA PORTAL' logo, the user 'etl-tester', and menu items for 'Build', 'Query & Visualize', 'Discover', 'Schedule', 'Quality', and 'Manage'. The user's name 'Swati' is visible in the top right.

The left sidebar contains the following sections:

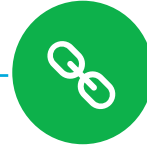
- Workflow:** A dropdown menu showing 'Count Booking'.
- Branch:** A dropdown menu showing 'main' with a 'read-only' status.
- Buttons:** 'Review and Deploy' and 'Save Changes'.
- Navigation:** 'ACTIONS', 'PARAMETERS', and 'JOBS' tabs.
- Your Actions:** A list of actions including 'H2H count booking', 'EMAIL send email', and 'H2H count booking-copy'.
- Footer:** 'Documentation' and 'Help?' links.

The main workspace shows the configuration for the 'count booking' workflow. It includes a configuration panel with sections for 'Action', 'Add Test', 'Query', 'Target table', 'Incremental mode (optional)', 'Partition', 'Merge', 'Quality and tracking', and 'Advanced setting'. Below this is a diagram view showing a sequence of steps: 'Start' (green play button), 'EMAIL send email', 'H2H count booking-copy' (with a table icon), and 'H2H count booking' (with a table icon). The 'H2H' steps include a table icon and a configuration snippet: `etl_test. {{resolve_table_name("book..`. A 'Run Workflow' button is located at the bottom right of the diagram area.

Agoda's Workflow Editor at a Glance

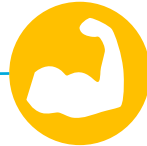


Workflow Editor



Decouple Workflow Definition from Execution

Allow changing underlying execution engine without touching user definitions



Empowered Employees

No Code / Low Code empower **ALL** employees to build complex data pipelines without technical expertise



Best Practices Out of the Box

Safely transition from Development to Production, Version Controlled, Ease of Rollbacks

agoda



Workflow Editor

agoda



Workflow Editor