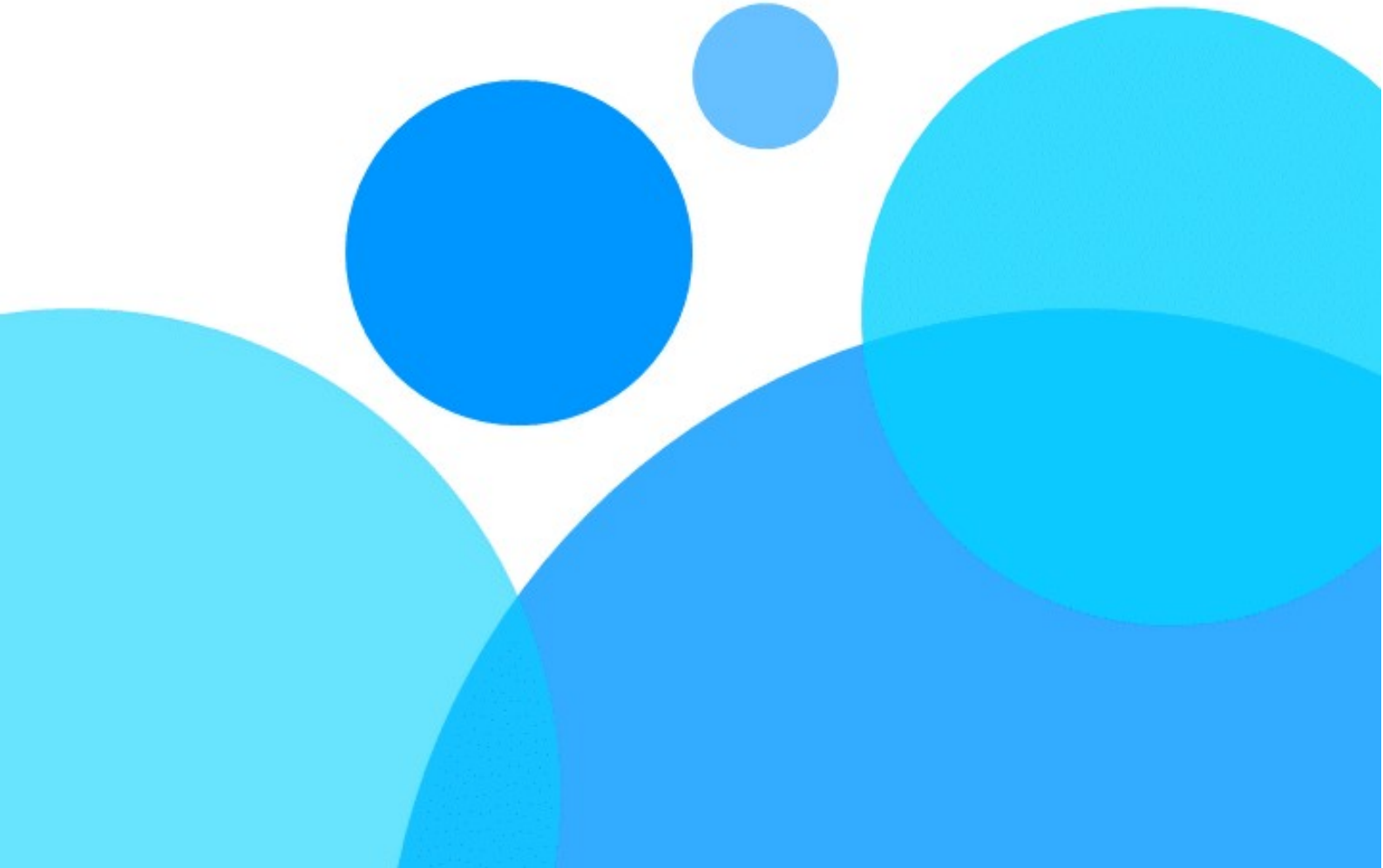
FineBI
帆软商业智能

Freely unleash the potential of data

FineBI V5.1.5 Direct Connect Performance Test Report



Document version.	V1.0
Last updated on.	2020-09-08
Author.	Performance Testing Group
Creation Date	2020-09-01

Revision record.

Versions	Modify date	Change Description
V1.0	2020.09.08	Output first draft

Cata
log

1 Overview

1.1 Purpose

Summarize and report on the performance of FineBI 5.1.5 based on clickhouse, vertica direct connection.

1.2 Background

FineBI version 5.1.5 is released with a companion direct connect performance test report.

1.3 Noun Description

Noun	Description
①Original table	The base tables added to the BI engine (added db database tables, sql datasets, or Excel datasets).
②Results Table	The result set of a self-help dataset, or a table made in a Dashboard (Dashboard).
③samples	The total number of iterations completed for the transaction.
④90%Line	90% of the sample time did not exceed this value.

1.4 Scope

This test is based on clickhouse, vertica, and focuses on 5.1.5 self-service datasets, dashboards
(Dashboard) preview, concurrency, and other performance tests.

1.5 Conclusion

The dashboard preview is mainly affected by database queries, and the million grouped database queries take longer.

Excel export length and export file size are positively correlated.

Self-help dataset 23.97% of scenario operation time is longer than 3

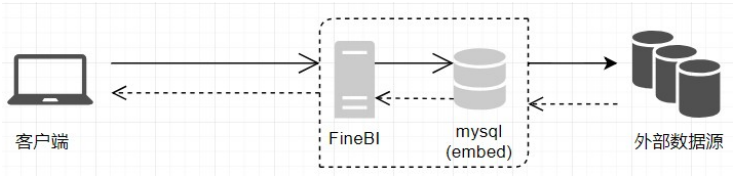
seconds; performance of simulated scenarios is 40% better than the previous version.

The performance of Excel dataset is not much different than spider, and the upload speed of csv file is 7~8 times faster than the upload speed of excel file.

The concurrent findings are detailed in section 4.4 below.

2 Test Environment

2.1 Test network



Note: All machines are on the same LAN, which excludes the influence of network factors on the system.

2.2 Server Configuration

Application Deployment	Server Address	Operating System	CPU	Number of physical cores	Logic cpu number	Memory	Disk Type
FineBI	188.168*80	CentOS 7	2 Intel(R) Xeon(R) CPU E5-2620 v4 @ 2.10GHz	2*8	32	64G-96G	2T (mechanical)
Mysql Configuration Library	188.168*215	CentOS 7	2 Intel(R) Xeon(R) CPU E5-2620 v4 @ 2.10GHz	2*8	32	64G	2T (mechanical)
clickhouse	188.168*72	CentOS 7	1 block Intel(R) Xeon(R) CPU E5-2620 v4 @ 2.10GHz	1*8	16	64G	2T (mechanical)
vertica	188.168*72	CentOS 7	1 block Intel(R) Xeon(R) CPU E5-2620 v4 @ 2.10GHz	1*8	16	64G	2T (mechanical)
Jmeter Negative Carrier	188.168*13	Windows 7	1 block Intel(R) Core(TM) i7-3770 CPU @ 3.40GHz	1*6	6	20G	300G(Machine machinery)

2.3 Database Configuration

Database Type	Test scenarios	Versions and parameters

Mysql Version	Configur ation Library	5.7 innodb_buffer_pool_size=1G innodb_log_file_size=48M max_connections=1000
clickhouse	Data source	20.3.8.53 mark_cache_size=5368709120 max_concurrent_queries=100 max_connections=4096 max_memory_usage=16G
vertica	Data source	Vertica Analytic Database 9.3.0

2.4 FineBI Configuration

FineBI Versions	Version 5.1.5 (0902)
jvm memory	16G
Tomcat thread count settings (server.xml)	maxThreads="800" minSpareThreads="100" maxSpareThreads="200" acceptCount="500"
Configuring library connection pools (db.properties)	hibernate.initialSize=50 hibernate.maxActive=500

3 Test content and method

3.1 Test Range

Single-user request time, core order of magnitude scenario response time.
Self-service dataset, dashboard (Dashboard) preview, editing concurrency.

3.2 Test content

Core order of magnitude description.

Number of rows of original table data	1 million		100 million
Number of groups (number of rows of grouped table results)	100	10,000	1 million
Number of data columns (results table_summary)	20(5 dimensions + 15 indicators)		
Number of data columns (result table_detail)	20		100

Raw table data, 10 million rows, 100 million rows.

The breakdown contains 20 columns and 100 columns.

Grouping table with 20 columns, all 5 dimensions and 15 indicators. The number of result rows in the grouping table is 100, 10,000, 1,000,000.

3.2.1 Dashboar

d scenarios 1.

Characteristics	Description
Subgroup Summary	Median, variance, standard deviation, number of records, averaging, ring period value, ring period ratio, maximum value
Totaling Method	Sum total, average total, minimum total, median total, and median total

Basic features

Year Month Interval	Year Month Interval Control
	Year Month Interval Control Breakdown
Controls and dimensions sorted by other fields	Dropdown tree dropdown operation
	Drop down tree grouping table display
	Control sorts by other fields (text drop-down, ascending order)
	Text drop-down operation
	Dropdown tree dropdown operation
	Drop down tree grouping table display
	Custom sorting of control bound fields
	Text drop-down operation
	Row header fields a Ascending order by associated field b
Control option values support filtering	Dropdown tree dropdown operation
	Drop down tree grouping table display
	The control option values support filtering (the text dropdown ends with
	Drop down tree grouping table display
	Belongs to 100 Dropdown tree dropdown operations
	Belongs to 50 Dropdown tree dropdown operations
	Control option values support filtering (text dropdown belongs to 100 items)
	Control option values support filtering (text dropdown belongs to 50 items)
	Control option values support filtering (text dropdown belongs to 50 items)
	Control Binding Field Filtering Top 100
	Control Binding Field Filtering Top 100 Dropdown Tree Dropdown Operations
	Control Binding Field Filtering Top 100 Dropdown Tree Display
	Control Binding Field Filtering Top 10
	Control Binding Field Filtering Top 10 Dropdown Tree Dropdown Operations
	Control Binding Field Filter Top 10 Dropdown Tree Display
	Control Binding Field Filtering Top 50
	Control Binding Field Filtering Top 50 Dropdown Tree Dropdown Operations
	Control Binding Field Filtering Top 50 Dropdown Tree Display
Schedule filtering support TopN	First 1000 items
	First 10w entries
Component Filtering	The detailed components belong to 100 items
	The detailed components belong to 50 items
	Aggregate components belong to 100 items
	Aggregate components belong to 50 items
	The grouping indicator filters the endings based on the text field
	Set the end of the itemized text field to
	The detail date field is set to not be between two dates
	Aggregate indicator filtered by date field not between two dates
	Breakdown 1 numeric field set between two numbers
	Aggregate 1 numeric field set between two numbers
	Dimension field filtering belongs to 100 items
	Dimension field filtering belongs to 50 items
	The grouping dimension filters the endings based on a dimension field that is
	Grouping table dimension fields to filter the first 100
	Grouping table dimension fields to filter the first 10
	Grouping table dimension fields to filter the first 50
	The dimension filters the largest 100 based on a metric
	The dimension filters the largest 10 based on a metric
	The dimension filters the maximum of 50 based on a metric
	Analysis of regional indicators set to filter the maximum of 100
	Analysis of regional indicators set to filter the largest 10
	Analysis of the regional indicators set to filter the largest 50
Linkage	Indicator linkage operation

	Linkage passing filter conditions, indicator detail filtering
	Dimensional linkage operation
	Linkage delivery table header dimension filtering, table header indicator filtering
	Indicator dimension linkage
	Indicator dimension linkage operation
Filtering to calculate sorting priority	The cumulative value is calculated after sorting
	All values are not computed twice after dimensional filtering
	All values are calculated twice after dimensional filtering
	All values in the group are not calculated twice after dimensional filtering
	All values in the group are calculated twice after dimensional filtering
	Cumulative values are calculated twice after dimensional filtering
	Cumulative values within a group are calculated twice after dimensional filtering
	Ranking is calculated twice after dimensional filtering
	Rankings are not calculated twice after dimensional filtering
	Intra-group ranking is calculated twice after dimensional filtering
	Intra-group ranking is not calculated twice after dimensional filtering

2.Function Features

Characteristics	Description
Basic Functions	Addition, subtraction, multiplication, division, numeric functions, text functions, date functions, logic functions Number, other functions
Aggregation functions	Sum_agg, Avg_agg, Max_agg, Count_agg, Countd_agg, Median_agg, Stdev_agg, Var_agg

3.Chart

Major Categories	Chart Subcategories	Display adaptation mode
Polar coordinate system graphs	Radar map	Standard adaptation
Right angle axis system graph	Heat points, funnel charts, dashboards, filled maps, text, pie charts, rectangular blocks, points, bar charts, lines, areas	Standard adaptation
		Standard adaptation, holistic adaptation
Non-rectangular coordinate axis system graph	Pie charts, funnel charts, rectangular tree blocks, aggregated bubbles Charts, multi-layer pie charts, word clouds	Standard adaptation
Geographical coordinate system graphics	Point map, heat map, flow map	Standard adaptation

4. Excel Export

Characteristics	Description
Schedule Export	Same columns (20 columns): Number of rows: 10,000, 500,000, 1,000,000, 5,000,000 Same number of rows (1 million): Number of columns: 10, 20, 50, 100
Grouped Table Export	Without calculation: 20 columns of 100 subgroups, 10,000 subgroups, 350,000 subgroups
	With calculations: number of de-duplicated records, median, variance, quick calculation - cumulative values, sorting, custom grouping, text controls
Cross Table Export	Without computation: 300 row dimensions 1200 column dimensions, original table 1 million rows, 5 million rows, 10 million rows
	With calculations: number of de-duplicated records, median, variance, quick calculation - cumulative values, sorting, custom grouping, text controls

3.2.2 Self-service data sets

1.Single step - based on db/sql table

Characte ristics	Desc ript ion
Subgroup Summary	Median, variance, standard deviation, averaging, number of records, de-duplicated counting, maximum value, 1:1 custom grouping Summation
New column	Time Difference, Get Time, Text Grouping Assignment, Value Grouping Assignment, Formula
Filterin g	Text belongs to, text ends with, value between, non-empty, formula
join	Intersection merge, left merge, right merge, merge set merge
union	Top and bottom merge
Field Settings	Field renaming, field undisplay

2.Single step - based on excel table

Characte ristics	Desc ript ion
Subgroup Summary	Sum, variance, average, record count, de-duplicate count
New column	Get time, text grouping assignment, value grouping assignment, formula
Filterin g	Text belongs to, text ends with, value between, non-empty, formula
join	Intersection merge, left merge, right merge, and merge set merge
union	Top and bottom merge
Field Settings	Change name, cancel field

3.Simulation of combination scenarios

Simulation of the combination of scenarios designed to add columns + filtering random combination of scenarios, involving 5 steps, 10 steps, 20 steps a total of 8 use cases. The new columns are all functionally added columns, and the operations involved in filtering and adding columns are shown below.

Characte	Desc

ristics	ript ion
Filtering	text belongs to, date belongs to, text ending yes/no, value between, non-empty, text contains/does not contain, value greater than, Before/after a date, value less than or equal to, date not equal to, date between, value less than, text beginning with is, four formulae, function formula
New column	POWER, PROMOTION, DATETONUMBER, LOWER, SWITCH, RADIANS, RAND, NVL, ACOS, SWITCH. SECOND, TODOUBLE, MID, ABS, YEAR, UPPER, IFP, LN, REGEXP, ASIN, CHAR, WEEKDAY, DATEDIF, AND, INT, DATEDELTA, OR, DATE, four formulae

3.2.3 Concurrent

Dashboard concurrency: 10 dashboard previews of a single scene; 50, 100, 150, 200 concurrent tests, concurrency lasts 10 minutes

Self-service data set concurrency: select 5 self-service data sets from the self-service data set single scenario: Edit-Preview Last Step - Exit Data Prep Preview; select 3 self-service data sets from the combined scenario: Edit-Preview Last Step
Step - Click on Step 2 Preview - Exit Data Prep Preview; Pick 2 self-service datasets from the Excel dataset: Edit - Preview Last Step - Exit Data Prep Preview; Press test 10, 20, 30 concurrently, concurrently for 10 minutes

Hybrid scenario: 5 dashboards selected in dashboard concurrency, 2 self-service datasets in single scenario, 2 self-service datasets in hybrid scenario, and 1 Excel dataset in self-service dataset concurrency; 10, 20, 30 concurrency tested, concurrency lasts 10 minutes

3.3 Testing Tools

Browser chrome, compression test jmeter.

4 Test results and analysis

4.1 Dashboard

The test results are using the vertica database
The page preview length is affected by the number of rows of the original table, the number of rows of the result table (grouped tables correspond to the number of groups), the number of columns of the result table, etc. The test results are all single-user operations with no other query tasks in the database (the response time will be slower when there are other query tasks or concurrency in the database).

4.1.1 Base feature preview time distribution

Order of magnitude description.

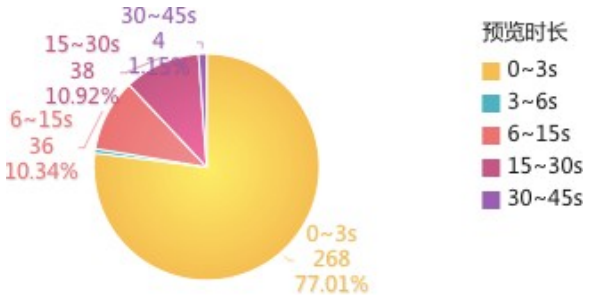
The original table uses DB/sql tables with 10 million rows and 100 million rows of data

Detailed table with 20 and 100 columns

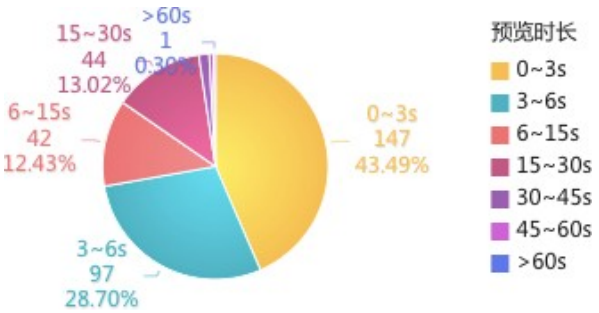
The base scenario produces a grouping table with 20 columns, all with 5 dimensions and 15 indicators. The number of rows in the grouping table is 100, 10,000, 1,000,000

Test results.

10 million orders of magnitude: a total of 348 test results. Among them, 77.01% of scenes are displayed within 3s; 0.57% of scenes 3~6s; 10.34% scenarios 6~15s; 10.92% scenarios 15~30s; 1.15% scenarios 30~45s; database query time accounts for 90%~95% of the total response time; scenarios greater than 6s are mainly those with a grouping of 1 million.



Under the 100 million order of magnitude: there are 338 test results. Among them, 43.49% of scenes are displayed within 3s; 28.7% of scenes 3~6s; 12.43% of scenes 6~15s; 13.02% of scenes 15~30s; 1.48% of scenes 30~45s; 0.59% of scenes 45~60s; 0.3% scenario is greater than 60s; database query time accounts for 90%~95% of the total response time.



Scenes	Order of magnitude	Response time	Number of records	Percentage of
Basic scenes	10 million	Within 3s	268	77.01%
		3~6s	2	0.57%
		6~15s	36	10.34%
		15~30s	38	10.92%
		30~45s	4	1.15%
	100 million	Within 3s	147	43.49%
		3~6s	97	28.7%
		6~15s	42	12.43%
		15~30s	44	13.02%
		30~45s	5	1.48%
		45~60s	2	0.59%
		Greater than 60s	1	0.3%

Weak performance scenarios

1 million grouping of all scenarios

Rows 32k/columns 1200 Cross-tab preview

4.1.2 Function preview time distribution

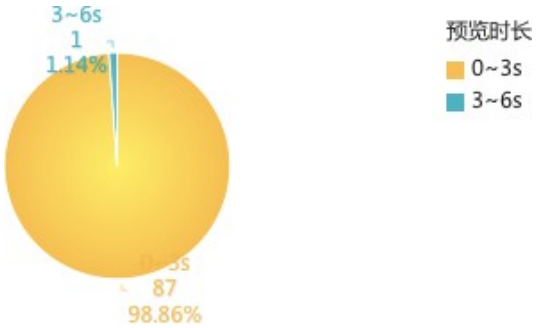
Order of magnitude description.

The original table uses DB/sql table with 10 million rows and 100 million data.

The base scenario was created with a grouping table of 10 columns, all with 5 dimensions and 5 indicators. The number of rows in the grouping table is 10,000.

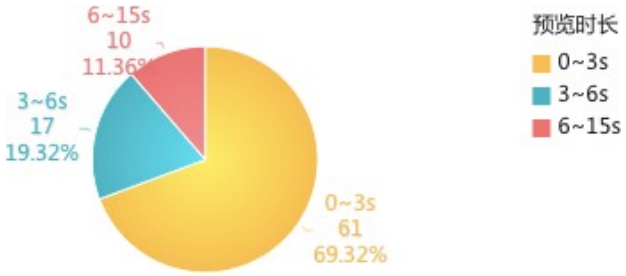
Test results.

Under 10 million orders of magnitude: there are 88 test results. Among them, 98.86% of scenarios are displayed within 3s; 1.14% of scenarios are 3~6s; database query time accounts for 90%~95% of the total response time.



100 million orders of magnitude: 82 test results in total. 69.32% of them previewed within 3s; 19.32% previewed 3~6s.

11.36% preview 6~15s; database query time accounts for 90%~95% of the total response time.



Scenes	Order of magnitude	Response time	Number of records	Percentage of
Function	10 million	Within 3s	87	98.86%
		3~6s	1	1.14%
	100 million	Within 3s	61	69.32%
		3~6s	17	19.32%
		6~15s	10	11.36%

Weak performance scenarios

Scenes	Sub-characteristics	Description	Number of original table rows	Number of groups
Function	Date Functions	DATEDIF	100 million	10,000
		DAYSOFMONTH	100 million	10,000
		DAYVALUE	100 million	10,000
	Aggregation functions	countd_agg	10 million / 100 million	10,000
		median_agg	10 million / 100 million	10,000

4.1.3 Combined scene preview length distribution

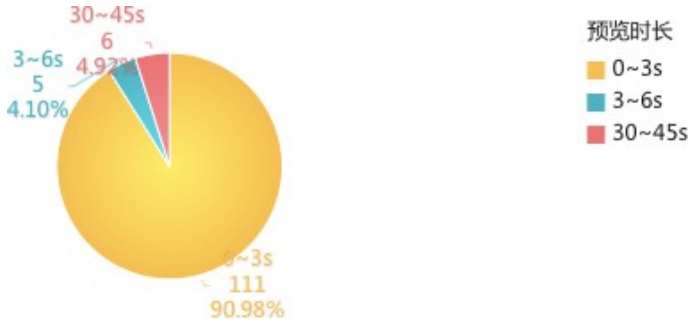
Order of magnitude description.

The original table uses a self-service dataset with 10 million rows and 100 million data.

The grouping table created by the combined scenario has 25 columns, all 10 dimensions and 15 indicators. The number of rows in the grouping table is 10,000.

Test results.

Under 10 million orders of magnitude: there are 122 test results. Among them, 90.98% scenarios are displayed within 3s; 4.1% scenarios are 3~6s; 4.92% scenarios are 30~45s; database query time accounts for 90%~95% of the total response time.



Scenes	Order of magnitude	Response time	Number of records	Percentage of
Combination scenes	10 million	Within 3s	111	90.98%
		3~6s	5	4.1%
		30~45s	6	4.92%

Weak performance scenarios

Scenes	Sub-characteristics	Description	Number of original table rows	Number of groups
Combination scenes	Subgroup Summary	Median	1 million	10,000/30 0,000
	Aggregation functions	countd_agg	1 million	10,000/30 0,000
		median_agg	1 million	10,000/30 0,000

4.1.4 Chart Preview Duration Distribution

Test scenario description.

The test template involves function points containing various diagrams, as detailed in the scenarios in section 3.2.1.

The page preview length is affected by the number of original table rows, the number of result table rows (grouping), and the number of result table columns. The test results are for single-user operation.

Grouping logic: All dimension fields in the analysis area are involved in the fine-grained (grouping) division, and the division logic is the

same as the grouping table logic, that is, the effect of dragging in dimension fields in the attribute area is equivalent to dragging in dimension fields in the grouping table; the grouping order is from left to right, from top to bottom, and the horizontal and vertical axes take precedence over the attribute panel.

Order of magnitude description.

The original table uses DB/sql tables with 10 million rows of data.

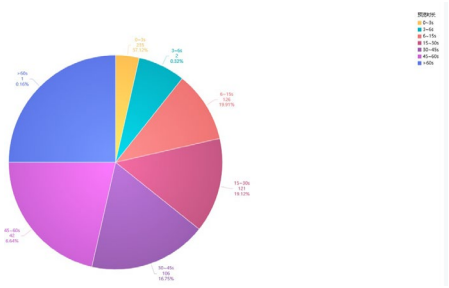
All scenes are produced in groups of 500, 5000, 30,000, and 1,000,000.

All scenarios use dimension fields including: 2 dimension fields, 3 dimension fields, and 5 dimension fields.

All scenarios use 1 indicator field.

Test results.

Under the 10 million order of magnitude, there are 128 test scenes, 37.12% of which are within 3s, 0.32% are 3~6s, 19.91% are 6~15s, 19.12% are 15~30s, and 23.55% are more than 30s.



Weak performance scenarios

Basic scenes

Chart Broad Category	Chart Subcategories	Computing Scenarios	Response time s
Coordinate axis system graphics	Dashboard, pie chart. Line graphs, bar graphs	Number of de-duplicated records + table header filtering	48-52
		Quick Calculation	40-42
		TOPN Filtering	36-38
		Aggregation method (mean, maximum, minimum, standard deviation, variance)	17-19

Big Data Scenarios

Major Categories	Chart Subcategories	Display adaptation mode	Number of groups	Response time s
Polar coordinate system graphs	Radar diagram	All modes	1 million	7
Coordinate axis system graphics	Heat points, funnel charts, dashboards, filled maps, text, pie charts, rectangular blocks, points, bars, lines, areas	Standard adaptation	1 million	6~10
		Width-adapted, height-adapted Response, overall adaptation	1 million	6~10

Non-coordinate axis system graphics	Pie charts, funnel charts, rectangular tree blocks, aggregated bubble charts, multi-layer pie charts, word clouds	All modes	1 million	6~8
Geographical coordinate system graphics	Point map, heat map, flow map	All modes	1 million	10~11

4.1.4 Excel Export

1.Schedule Export

Order of magnitude description.

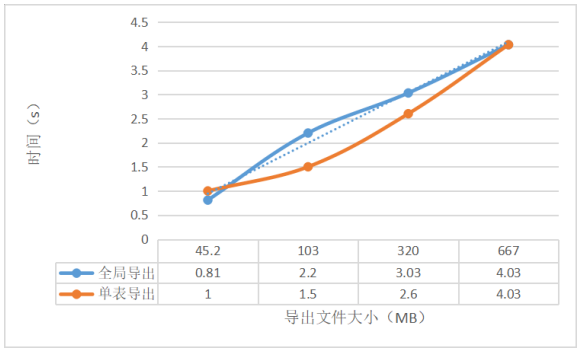
Keep 20 columns constant and test 10,000 rows, 500,000 rows, 1,000,000 rows, 5,000,000 rows

Keep 1 million rows constant and test 10, 20, 50, 100 columns

Test results.

When the number of rows is the same, the export time is positively correlated with the number of columns

When the number of columns is the same, the length of the export is positively correlated with the number of rows.



2.Grouped Table Export

Order of magnitude description.

The original table has 10 million rows and 20 columns, and the result table has 100 rows, 10,000 rows, and 350,000 rows.

Test results.

Grouped table export without calculation: export time is proportional to file size, within 2 seconds for 10,000 groups and 29 seconds for 350,000 groups.

Export length with calculations: Similar to the export length without calculations, except for the custom group export length of 4-5 seconds and the six component export length of 13 seconds.

Scenes	Number of groups	Export Method	File Size	Export time(s)
	100	Global Export	59.5K	1.05
		Single Table	20.4k	0.52

Grouping table (without calculation)		Export		
	10,000	Global Export	1.54M	1.81
		Single Table Export	1.51M	1.24
	350,000	Global Export	45.1M	29.54
		Single Table Export	45.1M	27.53
counted_agg	10,000	Global Export	1.52M	0.59
	10,000	Single Table Export	1.49M	0.29
Variance	10,000	Global Export	1.77M	2.17
	10,000	Single Table Export	1.74M	1.66
Quick Calculation (Cumulative value)	10,000	Global Export	1.64M	1 .89
	10,000	Single Table Export	1.57M	1 .40
Sort by	10,000	Global Export	1.60M	1 .87
	10,000	Single Table Export	1.57M	1 .47
Customized grouping	10,000	Global Export	7.4M	4 .98
	10,000	Single Table Export	7.38M	4 .74
Text control + grouping table	10,000	Global Export	1.53M	0 .59
	10,000	Single Table Export	1.51M	0 .29
Six components	10,000	Global Export	14.2M	13.65

3.Cross Table Export

Order of magnitude description.

Cross-tabulation without calculation: column grouping 1300, row grouping 300, original table 1 million rows, 5 million rows, 10 million rows

Cross-tabulation: column grouping 1300, row grouping 300, original table 10

million rows

Test results.

Export time with calculation: similar to export time without calculation except 13-14 seconds for de-duplicated records and 18 seconds for six components

Export length and file size are positively correlated

Sub-characteristics	Number of original table rows	Column grouping	Row grouping	Description	File Size	Export duration (s)
Cross-tabulation (without calculation)	100w	1300	300	Global Export	15.7M	0.95
				Single Table Export	15.7M	0 .63
	500w	1300	300	Global Export	15.7M	2 .67
				Single Table Export	15.7M	2 .29
	1kw	1300	300	Global Export	15.7M	1 .18
				Single Table Export	15.7M	0 .87
counted_agg	1kw	1300	300	Global Export	15.7M	14.18
			300	Single Table Export	15.7M	13.76
Variance	1kw	1300	300	Global Export	15.9M	1 .39
			300	Single Table Export	15.9M	0 .99
Quick calculation (cumulative) (Value)	1kw	1300	300	Global Export	22.5M	1 .31
			300	Single Table Export	22.5M	0 .84
Sort by	1kw	1300	300	Global Export	15.8M	1 .26
			300	Single Table Export	15.8M	0 .99
Customized grouping	1kw	1300	300	Global Export	17.3M	1 .69
			300	Single Table Export	17.3M	1 .16

Text control + groupin g table	1kw	1300	300	Global Export	15.7M	0 .59
			300	Single Table Export	15.7M	0 .29
Six components	1kw	1300	300	Global Export	84.7M	18.08

4.3 Data Preparation

Note: The test results are all single-user operations (concurrent operations, the length of time will slow down); directly connected data processing users and data analysis users in the operation of no difference. The following are the results of the operation with data processing users.

4.3.1 Self-

service dataset

1. single scene

Test scenario description.

The self-service data sets are all single-step and involve only one function point, including grouping summary, adding new columns, filtering, merging, and field setting, see Section 3.2.2 for detailed scenarios.

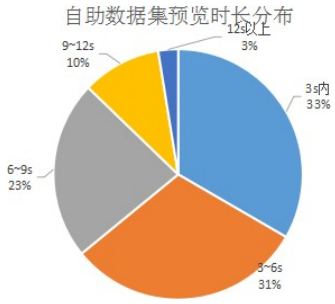
Use the clickhouse database.

Scenes	Number of rows	Number of columns	Number of groups
Grouping Summary Scenarios	1 million	20~30	10 dimensions: 10,000 groups
Other Scenes	Small table 1 million, 5 million Large table 10 million	20~30	-

Self-service dataset editing test results.

A total of 39 preview lengths were measured, 6which about 33% of the scenes

were within 3s, 31% were 3-6s, 23% were 6-9s, 10% were 9-12s, and 3% were more than 12s.



Scenes	Preview Length	Number of records	Percentage of
Self-service data sets	0~3 seconds	21	33%
	3~6 seconds	7	31%
	6~9 seconds	11	23%
	9~12 seconds	3	10%
	12 seconds or more	3	3%

Scenes longer than 9 seconds

Scenes	Operation steps
Grouping summary (10,000 groups) - summing, de-duplicating counts	Click to preview
Group summary (300,000 groups) - de-duplicate counts, standard deviation, summation	Click to preview

2. Simulation of combination scenarios

Test scenario description.

The new column+filter random combination involves 5 steps, 10 steps, and 20 steps. topN and average related functions are not supported by direct connection, so they are not involved in the simulation scenario testing. See Section 3.2.2 for detailed scenario descriptions.

Use the clickhouse database.

Order of magnitude description.

The original table has 10 million rows and 25 columns before adding new columns

Test results.

The preview time for each step of each scene is less than 3s, which is 1~2s performance improvement over the 0803 version.

Complex dataset combination scenarios that are more influenced by db

A.5 Step test results

In each step of each scene in the 5-step process, 80% of the previews were less than 2s long, and 20% of the previews were longer than 2s, but not more than 3s.

Scenes	Step 1 Preview	Step 2 Preview	Step 3 Preview	Step 4 Preview	Step 5 Preview
2 Add +3 Filtering	2.05s	2.05s	2.05s	2.23s	1.67s
3 Add +2 Filtering	1.81s	1.92s	1.61s	1.56s	2.02s
4 Add +1 Filter	1.77s	1.92s	1.79s	1.88s	1.76s
1 Filter +2 Add +2 Filter	1.56s	1.68s	1.99s	1.82s	1.81s
2 Add+2 Filter+1 Add	1.87s	2.08s	1.60s	1.39s	1.92s
3 Add +1 Filter +1 Add	1.75s	1.51s	1.71s	1.75s	1.90s

B.10 Step Test Results

In each of the 10 steps, 90% of the previews were within 2s and 10% were longer than 2s but less than 3s.

2 Add +5 Filter +3 Add			
Step 1 Preview	1.77s	Step 6 Preview	1.80s
Step 2 Preview	1.80s	Step 7 Preview	1.76s
Step 3 Preview	1.82s	Step 8 Preview	1.82s
Step 4 Preview	1.76s	Step 9 Preview	1.99s
Step 5 Preview	1.91s	Step 10 Preview	2.22s

C.20 Step Test Results

Of the individual steps in the 20-step process, 40% of the step previews were within 2s and 60% of the step previews were within 2~3s.

5 Add +5 Filtering +5 Add +5 Filtering			
Step 1 Preview	1.98s	Step 11 Preview	2.06s
Step 2 Preview	1.67s	Step 12 Preview	2.05s
Step 3 Preview	1.95s	Step 13 Preview	1.83s
Step 4 Preview	1.97s	Step 14 Preview	2.16s
Step 5	1.66s	Step 15	1.75s

Preview		Preview	
Step 6 Preview	1.89s	Step 16 Preview	2.16s
Step 7 Preview	2.14s	Step 17 Preview	2.00s
Step 8 Preview	2.16s	Step 18 Preview	2.15s
Step 9 Preview	2.19s	Step 19 Preview	2.06s
Step 10 Preview	2.08s	Step 20 Preview	2.22s

4.3.2 Excel dataset

Test scenario description.

Upload Excel/csv files, pay attention to the upload time (upload, default), save time (new_table_id).

Excel-based self-service datasets are all single-step, involving only one function point, including grouping summary, adding new columns, filtering, merging, and field settings, as described in Section 3.2.2 for detailed scenarios.

The test results are all single-user operations for the data processing user.

Order of magnitude description.

The amount of data to be uploaded in Excel/csv files is 10,000 20/100 columns, 100,000 20/100 columns, 300,000 20E columns, 500,000 20 columns, and the size should not exceed 100M.

E Excel-based self-help dataset with 200, 10w, 30w data volumes.

Test results.

A. Excel/csv file upload

Excel files import about 95.2w cells per second.

The csv file imports about 555.56w cells per second.

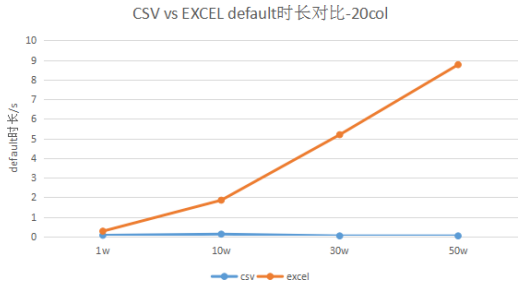
The speed of re-uploading is almost indistinguishable from uploading an Excel dataset directly.

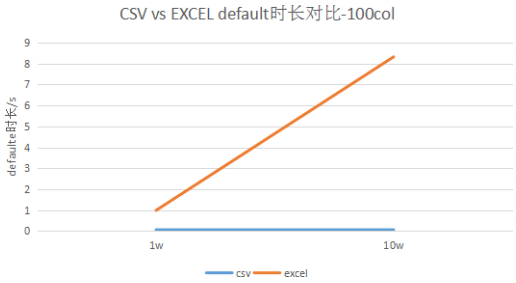
Excel files take about 1s for every 100w cells saved.

The upload time of csv file is about 4~5 times better than the upload time of Excel file (upload+default), and the main difference is in the default time.

Compared to the 0803 version, the 0901 directly connected Excel/csv dataset upload speed is 7 to 8 times faster.

Document Category	Number of rows	Number of columns	upload	default	new_table_id
csv	1w	20	0.04s	0.06s	0.16s
	10w		0.39s	0.12s	0.19s
	30w		1.03s	0.03s	0.29s
	50w		1.79s	0.03s	0.43s
	1w	100	0.17s	0.08s	0.17s
	10w		1.68s	0.08s	0.52s
excel	1w	20	0.04s	0.27s	0.18s
	10w		0.32s	1.85s	0.23s
	30w		1.02s	5.19s	0.35s
	50w		1.59s	8.76s	0.46s
	1w	100	0.17s	0.99s	0.21s
	10w		1.73s	8.32s	0.46s





B. Excel-based self-service dataset preview

A total of 22 preview lengths were measured, of which 86.56% of the scenes were

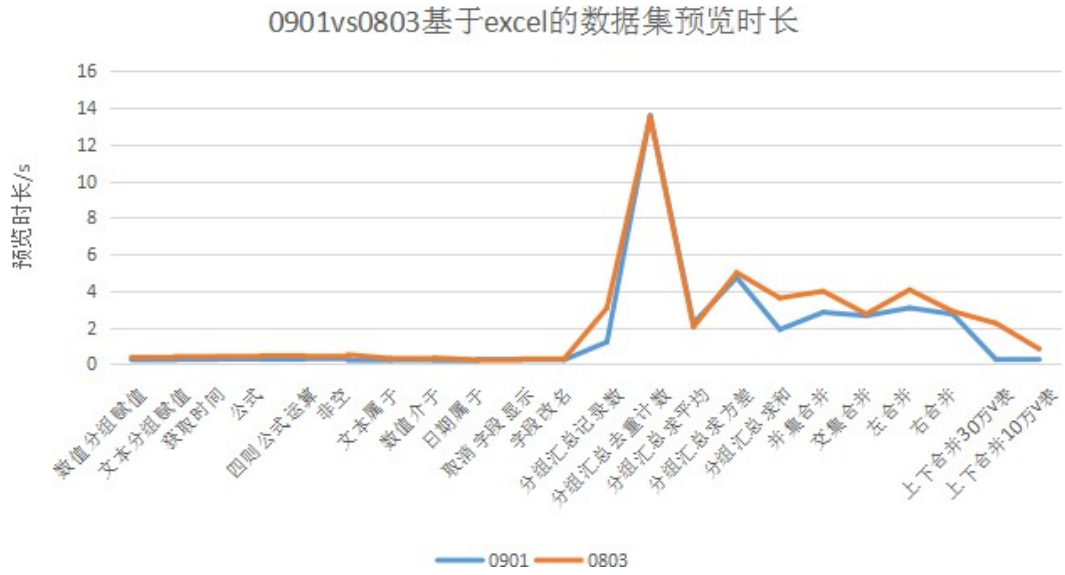
within 3s and 13.64% were greater than 3s.

Scenarios beyond the 3s are.

Scenes	Operation steps
Group summary-variance, de-weighted counts	Click to preview

Left-right merge - Left merge	Click to preview

The Excel-based self-service dataset shows little difference in performance compared to the previous version.



4.4 Edit Preview Concurrent Tests

4.4.1 Test scenario description

Dashboard concurrency: 10 dashboard previews of a single scene; 50, 100, 150, 200 concurrent tests, concurrency lasts 10 minutes

Self-service data set concurrency: select 5 self-service data sets from the self-service data set single scenario: Edit-Preview Last Step - Exit Data Prep Preview; select 3 self-service data sets from the combined scenario: Edit-Preview Last Step
Step - Click on Step 2 Preview - Exit Data Prep Preview; Pick 2 self-service datasets from the Excel dataset: Edit - Preview Last Step - Exit Data Prep Preview; Press test 10, 20, 30 concurrently, concurrently for 10 minutes

Hybrid scenario: 5 dashboards selected in dashboard concurrency, 2 self-service datasets in single scenario, 2 self-service datasets in hybrid scenario, and 1 Excel dataset in self-service dataset concurrency; 10, 20, 30 concurrency tested, concurrency lasts 10 minutes

Script scenarios	
Dashboard concurrency	Log in only once and loop through the top 10 time consuming dashboards

Self-service data set concurrency	Log in only once and edit 10 self-help data sets in a loop
Mixed scenes	Log in only once, cycle through 5 self-service datasets + cycle through 5 dashboards

Concurrency】Dashboard concurrency setting 50, 100, 150, 200 concurrency test; Self-service data set concurrency design 10, 20, 30 concurrency test; Combined scenario concurrency design 10, 20, 30 concurrency test.

Database usage] Concurrently, we use vertica database.

[Concurrency duration] The concurrency lasts 10 minutes.
[Concurrent start preparation] Restart the database and BI project after each concurrency.
[Concurrent Cache Settings] Turn off cache for separate operations and turn on cache for concurrent operations.

4.4.2 Test

result 1.

Dashboard

Template Name	Table 1	Table 2	Table 3	Table 4	Table 5	Table 6	Table 7	Table 8	Table 9	Table 10	Overall
concurrent /ms	1186	1186	2315	1560	1460	1006	300	905	400	455	1331

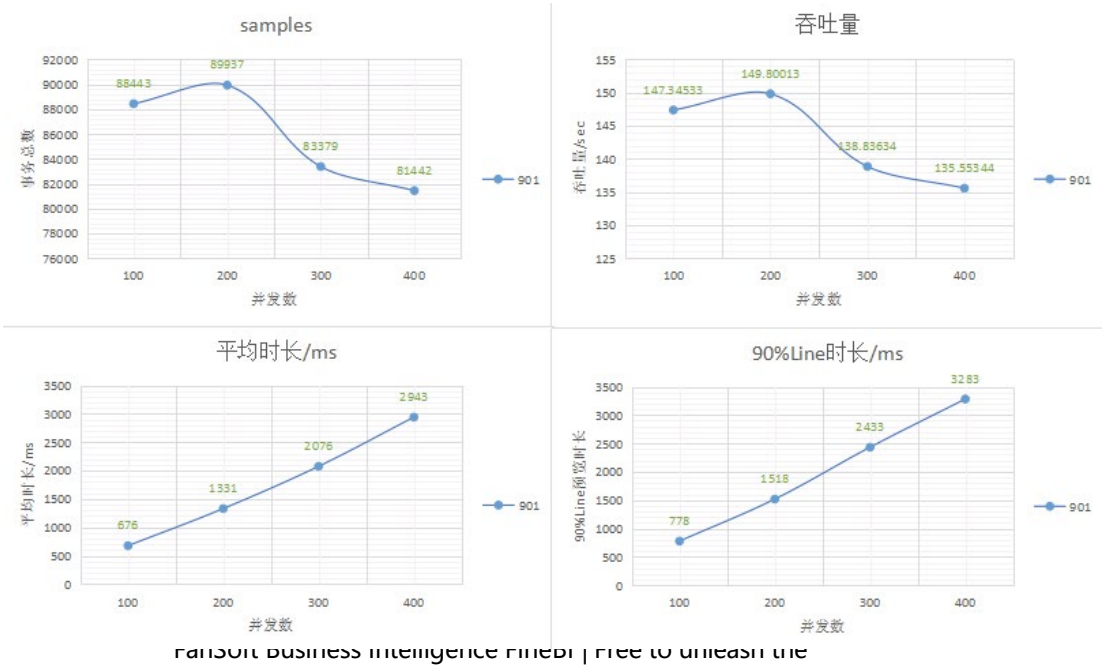
single

operation time

In terms of total number of transactions and throughput, the best performance in terms of throughput and total number of transactions is seen with 200 concurrency.

In terms of average time and 90%Line time, the time tends to increase significantly after 200 concurrency; the average time after 200 concurrency is significantly higher than the average time of separate operations.

Maximum 200 concurrency for the current scenario.



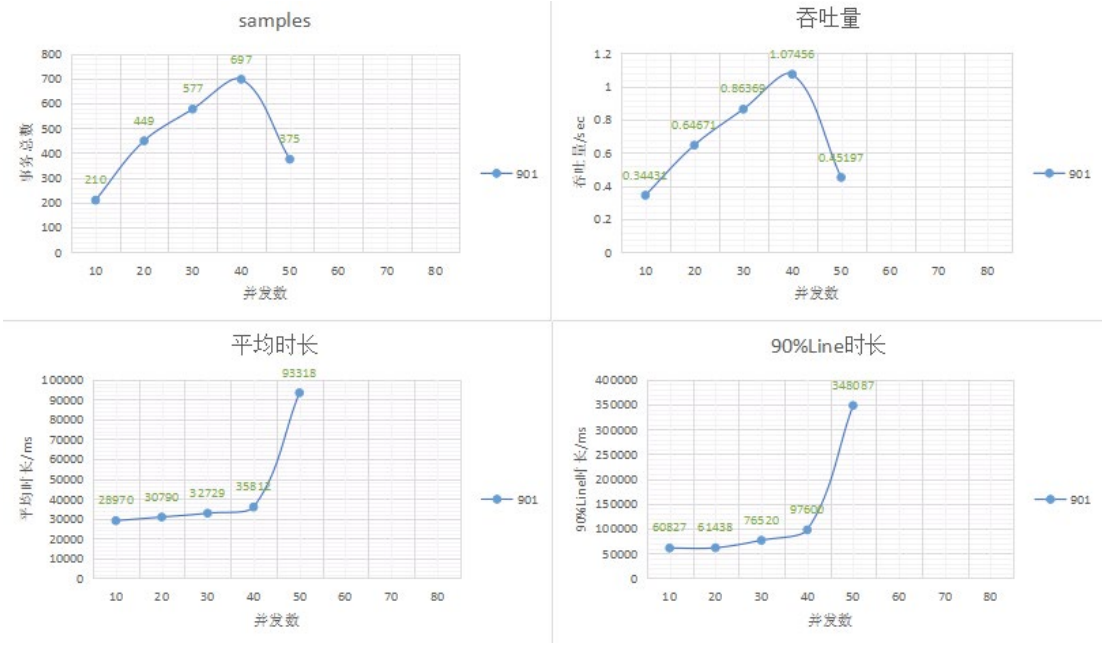
2.Self-service data set concurrency

Duration of individual operation

Operation	Access to data preparation	Go to Business Package 1	Table 1	Table 2	Table 3	Table 4	Table 5	Go to Business Package 2	Table 6	Table 7	Table 8	Go to Business Package 3	Table 9	Table 10	Overall
Duration /ms	127	1442	7460	12601	8078	9628	1449	307	8480	1253	1467	280	1751	4359	3918

In terms of total number of transactions and throughput, throughput and things always perform best when 40 concurrent.

In terms of average response time and 90%Line time, 40 the average time after concurrency is already significantly higher than the time of a single operation.



3.Hybrid scenario concurrency

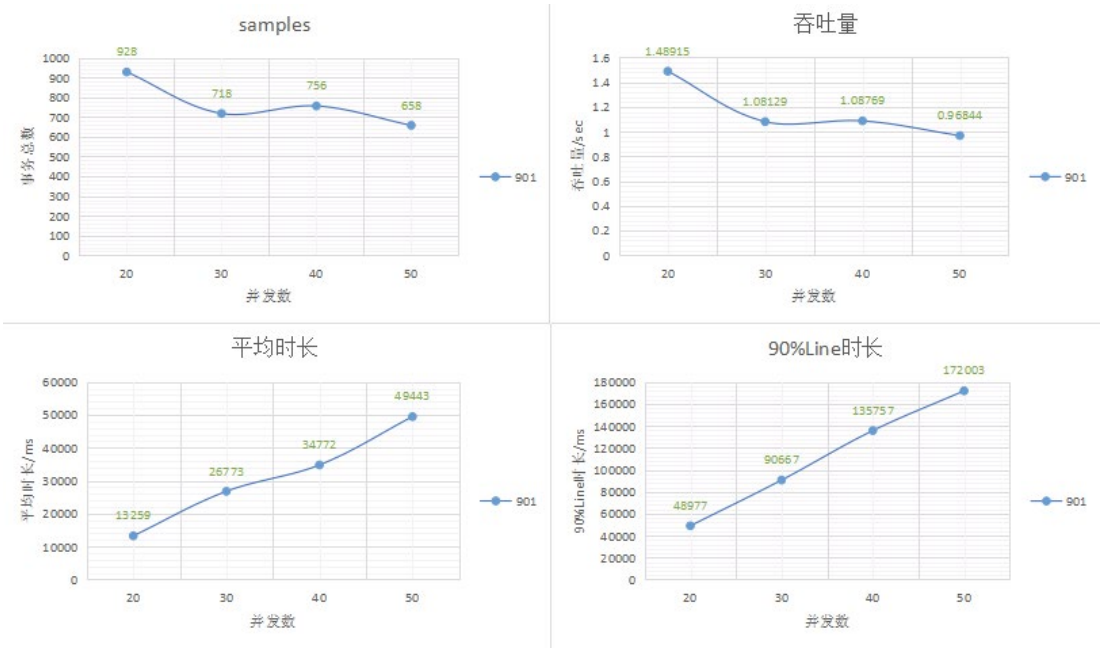
Duration of individual operation

Name	Table 1	Table 2	Table 3	Table 4	Access to data preparation	Access to the industry Service Package 1	Table 5	Table 6	Go to Business Package 2	Table 7	Table 8	Access to the industry Service Package 3	Table 9	Overall
Duration /ms	7279	3494	1323	1267	337	1365	14168	8172	370	10517	1752	278	8685	4241

In terms of throughput and total number of transactions, there has been a decreasing trend, with the best performance in terms of throughput and total number of transactions when 20 concurrent.

In terms of average time and 90% LIne time, 20 the response time increases significantly after concurrency and is significantly higher than the response time of a single operation.

Overall, it seems that the optimal 20 concurrent.



4.Summary

Scenes	Maximum number of concurrency
Dashboard Preview	200
Mixed scenes	20
Self-service dataset editing & preview	40