Using the IBM Cloud Application Performance Management Database Load Projections Spreadsheet September 2018



Using the IBM Cloud Application Performance Management Database Load Projections Spreadsheet

Document version 1.4

IBM Cloud Application Performance Management

Contract with IBM Corp.

CONTENTS

List of	Figures	8		iii		
Revisi	on Histo	ory		iv		
1	Introdu	luction1				
2	adsheet works	1				
	2.1	Typica	I Usage Scenario	2		
	2.2	Explan	ation of Prefetch DB Details worksheet	3		
	2.3	Explan	ation of PrefetchDBSummary worksheet	4		
		2.3.1	Summary results by agent type	4		
		2.3.2	Summary results – estimated data volume	6		
	2.4	Explan	ation of MongoDB&DatamartDB worksheet	6		
		2.4.1	Input parameters	7		
		2.4.2	Calculated values	7		
		2.4.3	Summary results – estimated data volume for MongoDB database	7		
		2.4.4	Summary results – estimated data volume for Datamart database	8		

LIST OF FIGURES

Figure 1 Worksheet tabs appear at the bottom of the spreadsheet	2
Figure 2 Sample calculations for Linux OS agent tables	3
Figure 3 Sample summary results by agent type	
Figure 4 Sample summary results – estimated data volume	
Figure 5 MongoDB&DatamartDB worksheet	

REVISION HISTORY

Date	Version	Revised By	Comments	
2015/03/02	1.0	JTJ	Original document for IBM Monitoring 8.1	
2015/06/18	1.1	JTJ	Updated for IBM Performance Management 8.1.1	
2015/10/23	1.2	JTJ	Updated for IBM Performance Management 8.1.2	
2016/04/19	1.3	JTJ	Updated for IBM Performance Management 8.1.3	
2017/07/31	1.4	JTJ	Updated for IBM Cloud Application Performance Management 8.1.4	
2018/09/24	1.4	JTJ	Updated for IBM Cloud Application Performance Management 8.1.4 IF0007 (up to 24 hour history in mongodb)	

1 Introduction

This document describes how to use the IBM Cloud Application Performance Management Database Load Projections spreadsheet.

The Cloud APM Database Load Projections spreadsheet was created to simplify the task of producing a disk space estimate for the Prefetch, MongoDB and Datamart databases. This spreadsheet includes the data set (also referred to as *attribute group*) information for more than 50 different agent 1 types, and allows the user to perform "what-if" exercises to see the database load for different agent environments. The spreadsheet includes two predefined charts showing the contribution of each agent type to the total database insert load and disk space usage. Because it is implemented in a standard spreadsheet format, other charts can be generated easily.

The total data size estimate given in the *PrefetchDBSummary* worksheet ("Total GB of disk usage") includes an extra 50% overhead for indexes, log space, and so on.

Projections produced by this spreadsheet should be viewed as rough estimates, but should be useful in making configuration planning decisions and in performing sensitivity analysis and what-if exercises. The actual disk storage required for a given monitoring configuration will depend on complex interrelationships among many variables, not all of which have been, or could be, modeled. It is the responsibility of the user to validate the spreadsheet inputs and outputs.

THE TOOL IS PROVIDED ON AN "AS IS" BASIS. IBM CORP. DOES NOT GUARANTEE THE PERFORMANCE OF THE TOOL OR THE RESULTS CALCULATED BY THE TOOL.

2 How the spreadsheet works

The spreadsheet is made up of six worksheets. For all of the worksheets within the spreadsheet, input cells are shown with a green background. Cells showing calculations based on the input parameters are shown with a yellow background.

The spreadsheet consists of the following worksheets:

- The ReadMe worksheet describes the spreadsheet and the limitations of its use.
- The MongoDB&DatamartDB worksheet allows you to estimate the disk space usage for the two databases that are used for transaction tracking data and alarms. The user enters the number of monitored applications, the average user transaction rate, the average monitored components involved per transaction, and the maximum number of alarms expected in 24 hours. Yellow cells show the estimated load for the two databases based on these input parameters.

© Copyright IBM Corp. 2018

¹ In this document, the term *agent* also refers to *data collector* (such as the J2SE data collector).

- The PrefetchDBSummary worksheet is the main worksheet, showing a list of agent types. The user enters the number of managed resources in their environment for each agent type. For certain agent types where the number of rows written is highly variable, additional Size parameters can be specified to reflect the size of the monitored environment. For example, for the VMware VI agent, the Size parameter is used to specify the expected number of VMs monitored by each VMware VI agent. Yellow cells show summary calculations based on input parameters on the Summary and Details worksheets.
- The PrefetchDBDetails worksheet lists all of the prefetch database tables in alphabetical order by agent product code (which is shown in the first column on the PrefetchDBSummary worksheet). The "Rows per Interval per Agent" column can be used to specify the expected number of rows per collection sample for each agent table. Default values are provided for most data sets, based on values observed in test environments or expected values. Yellow cells show Size parameter values referenced from the PrefetchDBSummary worksheet, and calculated values for each agent table in the prefetch database.
- The *InsertsByAgentType* worksheet shows a column chart of the total row insert rate projected for each agent type.
- The DiskSpaceByAgentType worksheet shows a column chart of the expected disk space usage by agent type.

To move from one worksheet to another, click one of the tabs along the bottom of the spreadsheet, which are shown in Figure 1 below.



Figure 1 Worksheet tabs appear at the bottom of the spreadsheet

2.1 Typical Usage Scenario

In a typical usage scenario, the user brings up the *PrefetchDBSummary* worksheet. The user specifies the expected number of agents for each type. For agent types that have Size parameters, the user specifies a value that reflects the size of their monitored environment. The user can also specify whether the number days of detailed data retained for the agent tables is the default of 8 days or the best practice values provided by agent developers.

No further input is required. If desired, the user can use the *Prefetch DB Details* worksheet to specify an expected number of rows written per collection sample for each agent table. The user can also specify a custom value for the number of days of detailed data retained for each table.

The yellow cells show calculated values based on the input values specifying the number of agents and the Size parameters. The following important estimated values are highlighted in red:

- Total database row inserts per minute
- Total disk space usage in GB

These values are important to consider in planning the hardware for running the Cloud APM server.

2.2 Explanation of Prefetch DB Details worksheet

The *PrefetchDBDetails* worksheet has calculated values for each agent table in the prefetch database. Figure 2 shows sample calculated values for Linux OS agent tables (product code LZ).



Figure 2 Sample calculations for Linux OS agent tables

- Product Code (column A) is the two digit code for the agent type
- Agent type (column B)
- DB Table Name (column C) is the table name as it appears in the prefetch database.
- Rows per Interval per Agent (column D) is an input column. Default values are
 provided for most agent tables based on observed values on test systems or
 expected values. These values can be overridden to better reflect the monitored
 environment. For some agent tables, the Size parameter from the
 PrefetchDBSummary worksheet is referenced (for example, the WebSphere MQ
 agent Queue Status table), and in those cases, the cell has a yellow background.
- Description of expected rows per interval (column E) provides information to help in setting an appropriate value for Rows per Interval per Agent (column D).
- Number of V8 Agents (column F) is referenced from the *PrefetchDBSummary* worksheet.
- Number of V6/V7 Agents (column G) is referenced from the PrefetchDBSummary worksheet. Certain types of V6/V7 agents can be monitored using the hybrid gateway.
- Retained days, custom value (column H) is an input column. By default, detailed
 data in the agent tables is retained for 8 days. The set_metric_retention.sh script
 can be used to customize the retention values, and this column in the spreadsheet
 can be used to reflect the custom values.
- Retained days (column I) shows the retention value in effect for the table, and
 used in the spreadsheet calculations. If a custom value is specified in column H,
 the custom value is shown. If no custom value is specified and if the best practice
 retention values have been selected on the *PrefetchDBSummary* worksheet, the
 best practice value is shown. Otherwise, the default retention value of 8 days is
 shown.

- Upload bytes per second per agent (column J) is a calculated value based on the number of columns and rows of data uploaded to the agent table and the upload interval for the agent table (1 minute for most tables).
- DB rows inserted per minute per agent (column K) is a calculated value based on the Rows per Interval per Agent (column D) and the upload interval for the agent table (1 minute for most tables).
- DB rows inserted per minute total (column L) is a calculated value based on the Rows per Interval per Agent (column D), the Number of V8 Agents (column F), the Number of V6/V7 Agents (column G), and the upload interval for the agent table (1 minute for most tables).
- DB KB inserted per minute total (column M) is a calculated value based on DB rows inserted per minute total (column L) and the expected size of database rows for the table.
- Total table rows (column N) is a calculated value based on DB rows inserted per minute total (column L) and the number of days of data to be retained in the database (column G).
- Total table size (MB) (column O) is a calculated value based on Total table rows (column N) and the expected size of database rows for the table.

2.3 Explanation of PrefetchDBSummary worksheet

The *PrefetchDBSummary* worksheet has input values and calculated values for each agent type supported in the prefetch database, and summary calculations for the estimated data volume.

2.3.1 Summary results by agent type

Figure 3 shows the table of agent types in the *PrefetchDBSummary* worksheet.

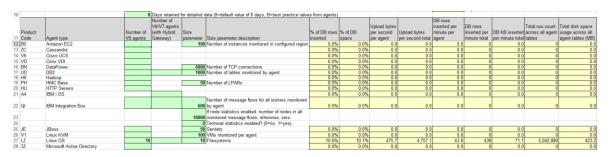


Figure 3 Sample summary results by agent type

- Days retained for detailed data (column D, row 10) specifies whether the detailed data for the agents is kept for 8 days (the default) or kept according to the best practice retention values provided by the agent developers (by specifying "B"). If desired, custom retention values for each agent table can be specified on the PrefetchDBDetails worksheet.
- Product Code (column A) is the two digit code for the agent type.

- Agent type (column B).
- Number of V8 Agents (column C) is an input value that specifies the expected number of agents of that type in the monitored environment.
- Number of V6/V7 Agents (column D) is an input value that specifies the expected number of V6/V7 agents of that type in the monitored environment. V6/V7 agent types with green input boxes for column D should be eligible for monitoring using the hybrid gateway.
- Size parameter (column E) is an input value that is used by certain agent types
 that can generate a potentially large number of agent table rows for each collection
 sample. For example, the MQ agent generates a row for every queue managed by
 the monitored queue manager. Sample values are given for the agent types that
 have a Size parameter specified. The user should modify these values to match
 the expected values for their monitored environment.
- Size parameter description (column F) applies to the Size parameter (if used) for the agent type.
- % of DB rows inserted (column G) is a calculated value showing the relative percentage of DB rows inserted per minute total (column L) for this agent type vs. the total for all agent types.
- % of DB space (column H) is a calculated value showing the relative percentage of Total disk space usage across all agent tables (column O) for this agent type vs. the total for all agent types.
- Upload bytes per second per agent (column I) is a calculated value based on the sum across all tables for the agent type of Upload bytes per second per agent (column H) from the PrefetchDBDetails worksheet. Additional network activity for heartbeat and protocol overhead is also included. This value represents the estimated average network overhead due to monitoring on the agent machine.
- Upload bytes per second total (column J) is a calculated value for the total network overhead across all agents of the agent type, and is the product of Upload bytes per second per agent (column I), Number of V8 Agents (column C) and Number of V6/V7 Agents (column D).
- DB rows inserted per minute per agent (column K) is a calculated value based on the sum across all tables for the agent type of DB rows inserted per minute per agent (column I) from the PrefetchDBDetails worksheet.
- DB rows inserted per minute total (column L) is a calculated value based on the sum across all tables for the agent type of DB rows inserted per minute total (column J) from the PrefetchDBDetails worksheet.
- DB KB inserted per minute total (column M) is a calculated value based on the sum across all tables for the agent type of DB KB inserted per minute total (column K) from the PrefetchDBDetails worksheet.
- Total row count across all agent tables (column N) is a calculated value based on the sum across all tables for the agent type of Total table rows (column L) from the PrefetchDBDetails worksheet.

 Total disk space usage across all agent tables (MB) (column O) is a calculated value based on the sum across all tables for the agent type of Total table size (MB) (column M) from the PrefetchDBDetails worksheet.

2.3.2 Summary results – estimated data volume

Figure 4 shows sample results of the estimated data volume:

3 Estimated data volume for Prefetch database			
4 Total database row inserts	Total disk space usage	Total data inserted into DB	Network usage - total uploaded data
5 69.1 rows per second	4.1 Total data in DB (GB)	15.6 KB per second	43.5 KB per second
6 4,148 rows per minute	6.1 Total GB of disk usage	934.5 KB per minute	2.6 MB per minute
7 37,067 rows per hour		54.8 MB per hour	153.0 MB per hour
8 0.89 million rows per day		1.3 GB per day	3.6 GB per day
9			0.357 Mbps received + transmitted

Figure 4 Sample summary results – estimated data volume

- Total database row inserts (column A) shows the estimated total row insert rate
 across all agent types. Results are expressed in rows per second, rows per
 minute, rows per hour and million rows per day. The rows per minute value is
 highlighted in red because it is useful in determining the hardware requirements for
 the Cloud APM server.
- Total disk space usage (column C) shows the estimated total disk space usage across all agent types.
 - The Total data in DB (GB) value is the sum of the Total disk space usage across all agent tables (MB) (column O) values.
 - The Total GB of disk usage value is an estimate of the total disk space requirement, which includes 50% above the Total data in DB (GB) value to account for indexes, transaction logs, etc.
- Total data inserted into DB (column G) shows the estimated total amount of data inserted into the database, and is the sum of the DB KB inserted per minute total values (column M), expressed in KB per second, KB per minute, MB per hour and GB per day.
- Network usage total uploaded data (column K) shows the estimated network bandwidth usage (received and transmitted) by the Cloud APM server. This is calculated by summing the *Upload bytes per second total* (column J) values, and expressed as KB per second, MB per minute, MB per hour, GB per day and Mbps received + transmitted.

2.4 Explanation of MongoDB&DatamartDB worksheet

The *MongoDB&DatamartDB* worksheet has input values and calculated values related to the two databases that are used for transaction tracking data. Figure 5 shows the contents of this worksheet.

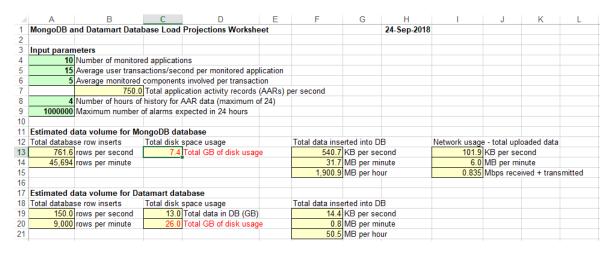


Figure 5 MongoDB&DatamartDB worksheet

2.4.1 Input parameters

- Number of monitored applications is the number of applications for which transaction tracking is enabled in the monitored environment.
- Average user transactions/second per monitored application should represent the average expected user transaction rate per application across all of the monitored applications.
- Average monitored components involved per transaction is a measure of the
 application complexity. A user interaction with a web page could involve several
 monitored components, including a web server, application server, database
 server, messaging middleware, and even the browser if JavaScript injection is
 configured. For each monitored component in the user transaction flow, an
 application activity record (AAR) is generated and tracked in the MongoDB and
 Datamart databases.
- Number of hours of history for AAR data can be set to see the effect of increasing the amount of AAR history kept in mongodb. The default value is 4 hours and the maximum value is 24 hours.
- Maximum number of alarms expected in 24 hours is used in estimating the size of the alarm mongo database.

2.4.2 Calculated values

• Total application activity records (AARs) per second is calculated by multiplying the three input parameters together.

2.4.3 Summary results – estimated data volume for MongoDB database

- Total database row inserts (column A) shows the estimated total row insert rate. There is one insert per AAR and one insert per alarm.
- Total GB of disk usage (column C) shows the estimated total disk space usage based on the number of hours of history for AAR data kept in mongodb.

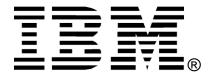
- Total data inserted into DB (column F) shows the estimated total amount of data inserted into the database. Values are expressed in KB per second, MB per minute and MB per hour.
- Network usage total uploaded data (column J) shows the estimated network bandwidth usage (received and transmitted) by the Cloud APM server. Results are expressed as KB per second, MB per minute and Mbps received + transmitted.

2.4.4 Summary results – estimated data volume for Datamart database

- Total database row inserts (column A) shows the estimated total row insert rate. There is one insert per monitored user transaction.
- Total disk space usage (column C) shows the estimated total disk space usage.
 These calculations reflect eleven days of data. The Total GB of disk usage is
 inflated by a factor of two to reflect the additional storage required for indexes, log
 files and so on.
- Total data inserted into DB (column F) shows the estimated total amount of data inserted into the database. Values are expressed in KB per second, MB per minute and MB per hour.

Trademarks

- DB2, IBM, Lotus, Tivoli, Rational, and WebSphere are trademarks or registered trademarks of IBM Corporation in the United States, other countries, or both.
- Excel, Windows and Windows NT are registered trademarks of Microsoft Corporation in the United States, other countries, or both.
- Java and all Java-based trademarks and logos are trademarks or registered trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.
- Other company, product, and service names may be trademarks or service marks of others.



© Copyright IBM Corporation 2018

IBM United States of America

Produced in the United States of America

All Rights Reserved

The e-business logo, the eServer logo, IBM, the IBM logo, OS/390, zSeries, SecureWay, S/390, Tivoli, DB2, Lotus and WebSphere are trademarks of International Business Machines Corporation in the United States, other countries or both

Lotus, Lotus Discovery Server, Lotus QuickPlace, Lotus Notes, Domino, and Sametime are trademarks of Lotus Development Corporation and/or IBM Corporation.

Java and all Java-based trademarks and logos are trademarks of Sun Microsystems, Inc. in the United States, other countries or both.

Other company, product and service names may be trademarks or service marks of others.

INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PAPER "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

Information in this paper as to the availability of products (including portlets) was believed accurate as of the time of publication. IBM cannot guarantee that identified products (including portlets) will continue to be made available by their suppliers.

This information could include technical inaccuracies or typographical errors. Changes may be made periodically to the information herein; these changes may be incorporated in subsequent versions of the paper. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this paper at any time without notice.

Any references in this document to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to:

IBM Director of Licensing IBM Corporation 4205 South Miami Boulevard Research Triangle Park, NC 27709 U.S.A.

