IDUG

Now you see it, Db2 AI for z/OS 1.5

Akiko Hoshikawa, IBM

September 13, 2022
Midwestern USA RUG event
Db2 for z/OS track

Disclaimer

The information contained in this publication is provided for informational purposes only. While efforts were made to verify the completeness and accuracy of the information contained in this publication, it is provided AS IS without warranty of any kind, express or implied. In addition, this information is based on IBM's current product plans and strategy, which are subject to change by IBM without notice. IBM shall not be responsible for any damages arising out of the use of, or otherwise related to, this publication or any other materials. Nothing contained in this publication is intended to, nor shall have the effect of, creating any warranties or representations from IBM or its suppliers or licensors, or altering the terms and conditions of the applicable license agreement governing the use of IBM software.

References in this presentation to IBM products, programs, or services do not imply that they will be available in all countries in which IBM operates. Product release dates and/or capabilities referenced in this presentation may change at any time at IBM's sole discretion based on market opportunities or other factors, and are not intended to be a commitment to future product or feature availability in any way. Nothing contained in these materials is intended to, nor shall have the effect of, stating or implying that any activities undertaken by you will result in any specific sales, revenue growth or other results.

Agenda

- Overview
- Features of Db2 AI for z/OS
  - SQL Optimization
  - Distributed Connection Control
  - System Assessment and Performance Insights
- Requirements
  - Requirement to run Db2ZAI
  - Capacity planning
  - Up and running experience
- Beyond 1.5
- Q&A

Overview

Product Releases and Feature Updates

2H 2018 Db2ZAI V1.1 SQL Optimization –Learning SQL patterns

1H 2019 Db2ZAI V1.2 System Assessment – out of box performance tuning recommendation

1H 2020 Db2ZAI V1.3 Distributed Connection Control – manage remote connections

2H 2020 Db2ZAI V1.3.1

- Self healing access path
- Dynamic IP support – IP address that are not yet learned
- RACF/MFA Security support

1H 2021 Db2ZAI V1.4

- High availability support
- SQL Optimization dash board update
- Integration with vendor data (Query Monitor, etc) to drill down the exception analysis

1H 2022 Db2ZAI V1.5

- Application/user-id based filtering in DCC
- High level application view in system assessment
- Granular control of features
- Db2 13 support

Db2 AI for z/OS

Enhance performance, reduce resource consumption and maintain Db2 for z/OS health

SQL Optimization

- Create better access path by predicting likely SQL behavior (value of host variable, rows fetched, etc.)
- Apply additional runtime optimization (better sort method or parallelism)
- Automatically stabilized the dynamic queries and provide a capability to lock/unlock access path
- Automatically self tune the access path

Cost reduction : Automatic Access Path Tuning, Statement level access path lock

System Assessment

- Leverage AI to learn what is normal and performs exception-based analysis based on the learned thresholds
- Provide recommended actions to take for each exception based on Db2 SME knowledge
- Provide the additional correlation information for each exception to speed up analysis
- Automatically collect key performance information and visualize them to ease analysis in data sharing group level

Performance & Resiliency: Automate performance analysis.

Performance Insights (*new in v1.5)

- Interactive (real-time) Db2 performance analysis
- Integrates new application-level data
- Graphically driven with launch capabilities for deeper dives

Productivity : Years of experience embedded within

Distributed Connection Control

- Leverage AI to learn the distributed connection behavior to control the flood of connections and threads using Db2 profile function
- Automatically learn and visualize connection and thread behavior
- Create warning and exception alerts with additional connection/thread information

Enhance resiliency in Hybrid cloud environment

What have other users seen ?

Large European Bank

Average transaction response time

18% ↓

Average transaction CPU utilization

19% ↓

40M online transactions executed in 3 hours utilizing System Assessment recommendation

Large European Bank

Overall elapsed time improvement

66% ↓

CPU consumption

48% ↓

“Just amazing to be able to choose a better access path by capturing host variables without using the reopt parameter”

Utilizing SQL Optimization

Large US Insurer

SQL sorts improved

39M

SQL statements ‘learned’ into a more optimal access path

2000+

CPU savings

30%

All within 48 hours of up and running using SQL Optimization

SQL Optimization

Leverage AI to Improve The Access Path Selection

Predict likely value of Host variable/parameter markers

- Db2ZAI predicts likely literal values using models based on historical usage
- Random Forest Algorithm to predict program variable value
- Helps optimizer choose different table join sequence and join methods

```
WHERE C_CUSTKEY = O_CUSTKEY
AND L_ORDERKEY = O_ORDERKEY
AND C_NATIONKEY = N_NATIONKEY
AND O_ORDERDATE >= ?
AND O_ORDERDATE < ?
AND C_STATE = ?
GROUP BY … ORDER BY …
```

Actual number of rows fetched by the application

- SQL might qualify hundreds or thousands of rows
- Application only fetches a subset of the qualifying rows (one screen)
- Db2ZAI learns application behavior and adjusts access path selection appropriately
- Liner Regression algorithm to predict number of rows fetched by Application

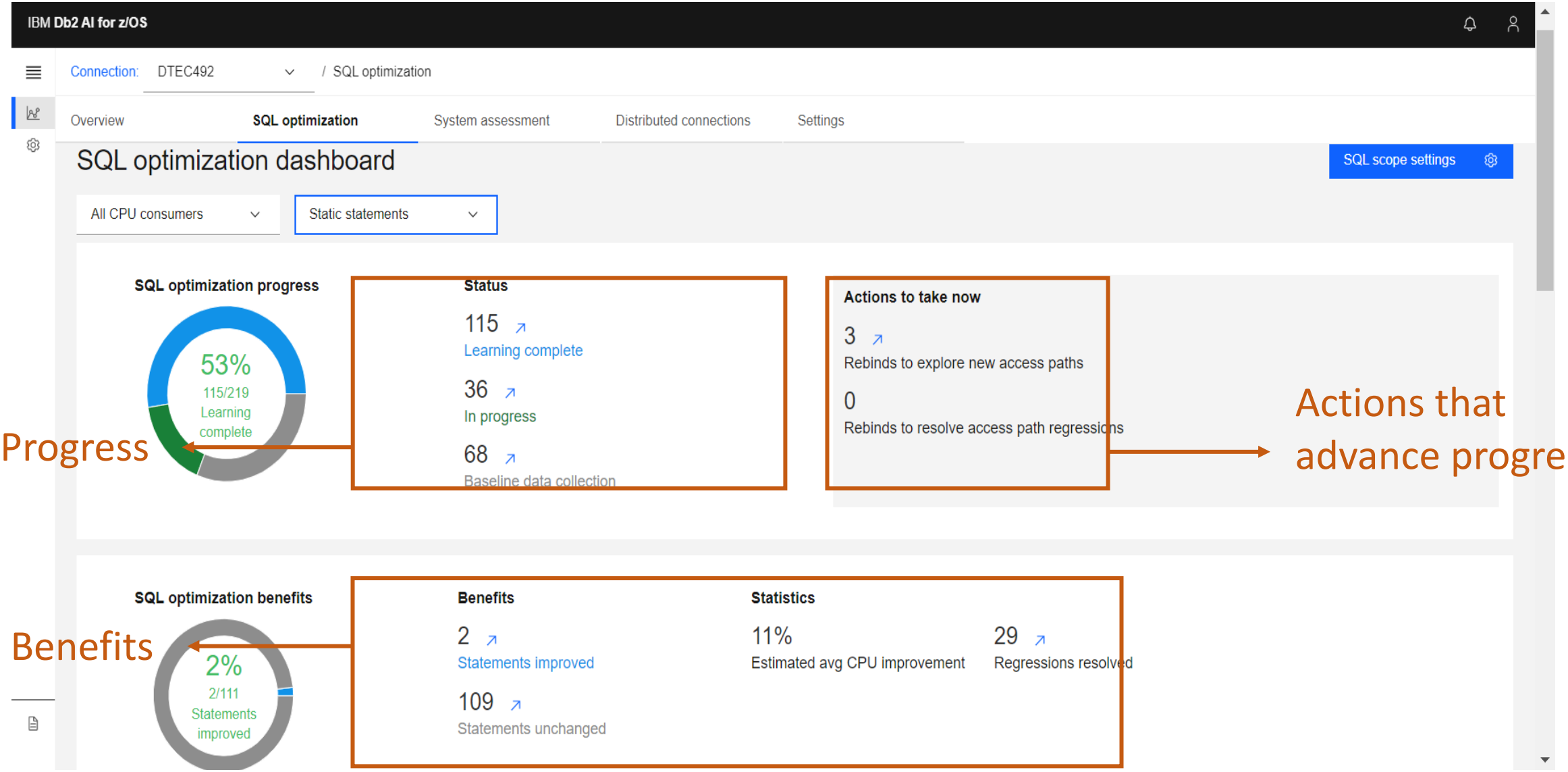
Parallelism exploitation

- Learns from prior execution cost and turn on the parallelism safely
- Liner regression algorithm to predict CPU time

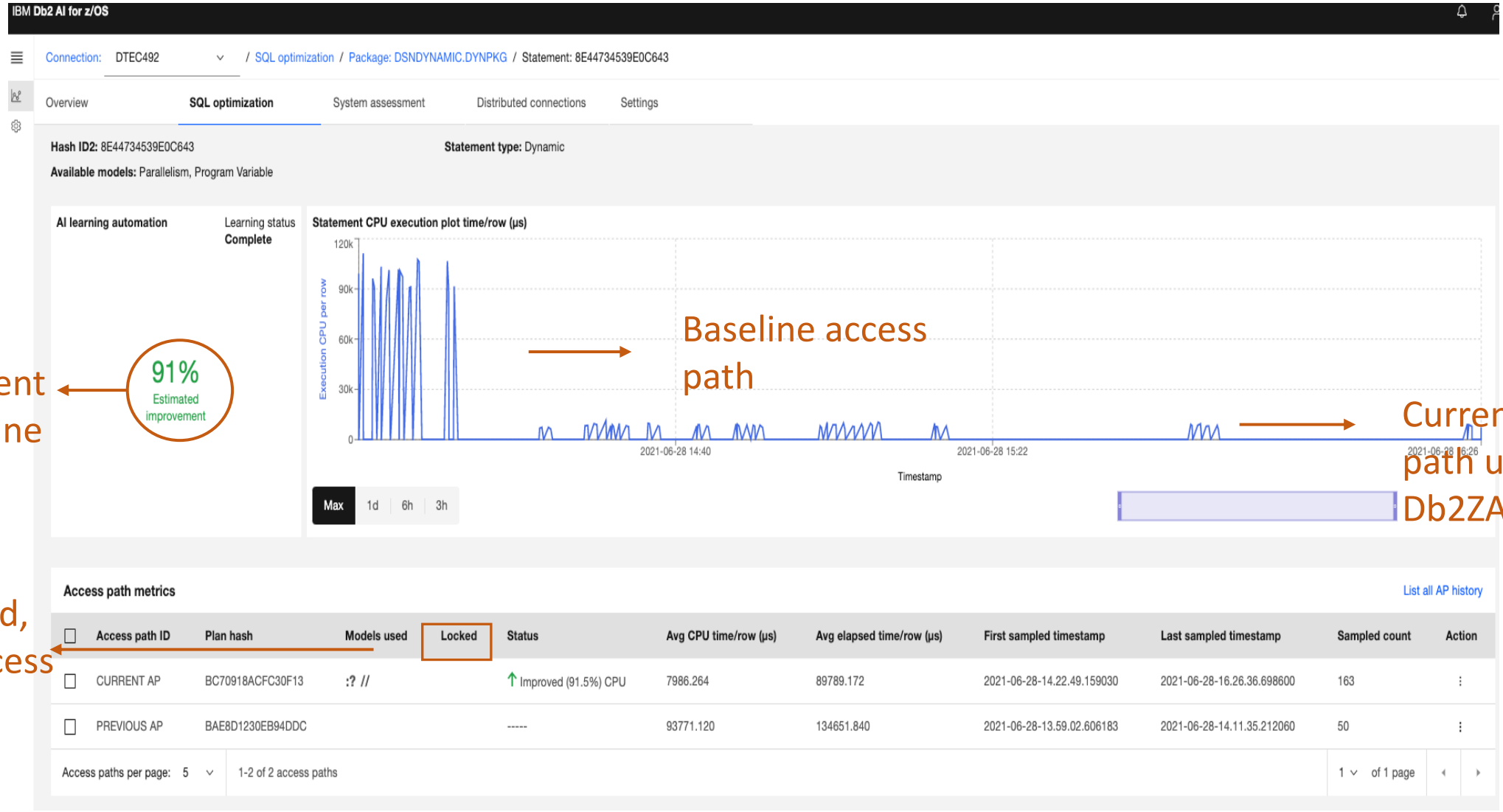
RDS SORT algorithm and memory allocation

- Learns from prior SQL statement execution and adjusts sort algorithm
  - DISTINCT, ORDER BY
  - Possibly more Z SORT exploitation if running on z15

Db2ZAI SQL Optimization Dashboard



Db2ZAI SQL Optimization Detail



Estimated improvement over baseline

Baseline access path

Current access path using Db2ZAI model

Lock a good, proven access path

(Customer POC #1) Access Path : Before AI / After AI

The query that selects the full parameter index instead of the empty parameter, thanks to AI

BEFORE

```
DB2 SSID ==> TDA1              Type ==> PACKAGE          Program ==> DFGMN252
Collid   ==> ATCOR
Version  ==> 2019-05-12-15.24.50.663342
Total/Avg => A

Interval Date => 30-08-21   Interval Time => 00:00:00    Elapsed Time => 01-17
--------------------------------------------------------------------------------

Q -View SQL text, T -View Tables/Indexes, E -Explain, D -View Detail

   SQL_CALL  SQL         TIMEPCT CPUPCT  INDB2_TIME   INDB2_CPU    GETPAGE
   --------  ----------  ------- ------- ------------ ------------ ----------
_  OPEN              12  99.99%  99.97% 00:00.257064 00:00.074908   38679.08
_  CLOSE             12    .00%    .01% 00:00.000009 00:00.000008       0.00
_  FETCH             12    .00%    .01% 00:00.000010 00:00.000008       0.00
******************************** BOTTOM OF DATA ********************************

.XB XFGINVI2 HYS(28012364)
.CUSTOMER_NUM
.INVOICE_ID
.INVOICE_YEAR_NUM
.CURRENCY_CODE
.XE



Total/Avg => T                DB2 SSID ==> TDA1
 Program  ==> DFGMN252         Type     ==> PKGE      Statement     ==> 0001490
 Ctoken   ==> 1AC3855E165CEB90                        Section       ==> 00001
 Collid   ==> ATCOR                                   SQL Call Type ==> OPEN
 Version  ==> 2019-05-12-15.24.50.663342

 Interval Date => 30-08-21  Time => 00:00:00  Elaps Time => 01-17   Samp => 100%
--------------------------------------------------------------------------------

 TABLE               INDEX    TB_SEQ_GP TB_IDX_GP TB_LNK_GP IS_GETP   IS_TBGETP
 ------------------  -------- --------- --------- --------- --------- ---------
 T_IEF_INVOICE_INFO  XFGINVI2                                   5436    458652
 T_IEF_INVOICE_INFO                   0    458652         0
 SYSSPTSEC_DATA      DSNSPDXA                                     12         0
 SYSSPTSEC_DATA                      36         0         0
 SPTR                DSNSPT01                                     12         6
 SPTR                                 0         6         0
******************************** BOTTOM OF DATA ********************************
```

AFTER

```
DB2 SSID ==> TDA1              Type ==> PACKAGE          Program ==> DFGMN252
Collid   ==> ATCOR
Version  ==> 2019-05-12-15.24.50.663342
Total/Avg => A

Interval Date => 01-09-21   Interval Time => 00:00:00    Elapsed Time => 01-17
--------------------------------------------------------------------------------

Q -View SQL text, T -View Tables/Indexes, E -Explain, D -View Detail

   SQL_CALL  SQL         TIMEPCT CPUPCT  INDB2_TIME   INDB2_CPU    GETPAGE
   --------  ----------  ------- ------- ------------ ------------ ----------
_  OPEN              18  99.71%  97.59% 00:00.002427 00:00.000243       8.22
_  CLOSE             18    .16%   1.60% 00:00.000004 00:00.000004       0.00
_  FETCH             18    .12%   1.20% 00:00.000003 00:00.000003       0.00
******************************** BOTTOM OF DATA ********************************

.XB XFGINVI8 HYS(64496311) BP(BP2)
.BUYER_CUST_NUM
.XE



Total/Avg => T                DB2 SSID ==> TDA1
Program  ==> DFGMN252         Type     ==> PKGE      Statement     ==> 0001490
Ctoken   ==> 1AC3855E165CEB90                        Section       ==> 00001
Collid   ==> ATCOR                                   SQL Call Type ==> OPEN
Version  ==> 2019-05-12-15.24.50.663342

Interval Date => 01-09-21  Time => 00:00:00  Elaps Time => 01-17   Samp => 100%
--------------------------------------------------------------------------------

TABLE               INDEX    TB_SEQ_GP TB_IDX_GP TB_LNK_GP IS_GETP   IS_TBGETP
------------------  -------- --------- --------- --------- --------- ---------
T_IEF_INVOICE_INFO  XFGINVI8                                    30         0
SYSSPTSEC_DATA      DSNSPDXA                                    24         0
SYSSPTSEC_DATA                      72         0         0
SPTR                DSNSPT01                                    24         6
SPTR                                 0         6         0
******************************** BOTTOM OF DATA ********************************
```

Distributed Connection Control

Managing Inbound Network Traffic

- Db2 administrators lack the knowledge of remote application behavior
  - No protection from connection flooding and system level impact
- Db2 administrators lack the knowledge to set up Db2 mechanisms used to control connections
  - Setting connection controls too low can cause an application outage
  - Setting connection controls too high reduces their effectiveness

"IBM must give me a better way to protect Db2."

"A misconfigured application property caused us a multi million-dollar outage"

"If IBM doesn't build the controls, then we need to build it."

Distributed Connection Control

- Automatically learns about connection and thread behavior
- Identifies and resolves problems before they cause an impact
- Improve the visibility of your Db2 environment
- Uses learned behavior to identify why threads are getting blocked
- Profile recommendations to prevent connection flooding from impacting other applications

DCC – Solution Overview

Analyze and Enable Protection:

IFCID 365 (location stats) and 411 (client application) or 412 (client userid) collected in Db2ZAI tables

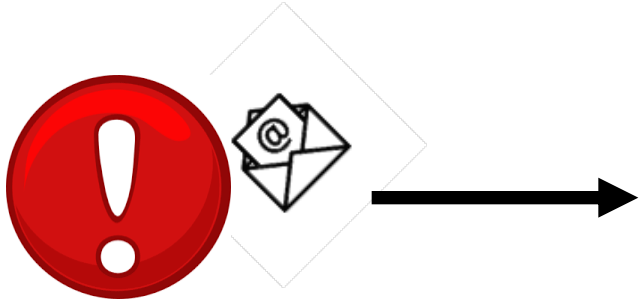
User initiates the “training” process to analyze 365, 411/412 data, generate recommended profiles.

User reviews training data.
Discovers insight about applications.
Reviews recommended profiles, and edits if necessary.

User activates profiles
Protection is enabled.

Monitor and Troubleshoot:

IFCID 402 (Profile threshold) data collected in Db2 table when profile thresholds exceeded



“Alerts” are raised in the Db2ZAI UI.
User receives an email notification.

User can view the “scorecard” to determine the appropriate response.

Generating profile recommendations

1

Prioritize remote applications

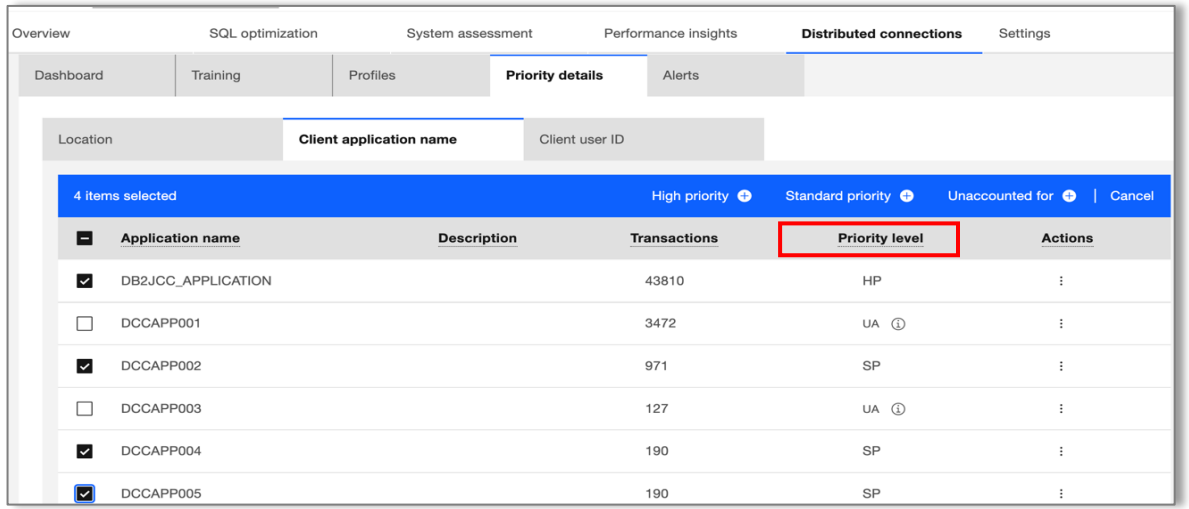
- HP, SP, and UA

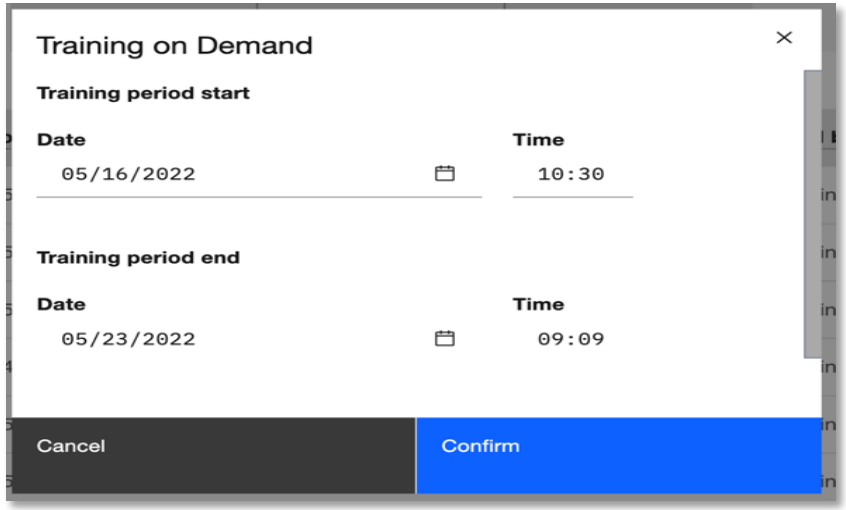
2

Select time range that captures peak workloads

3

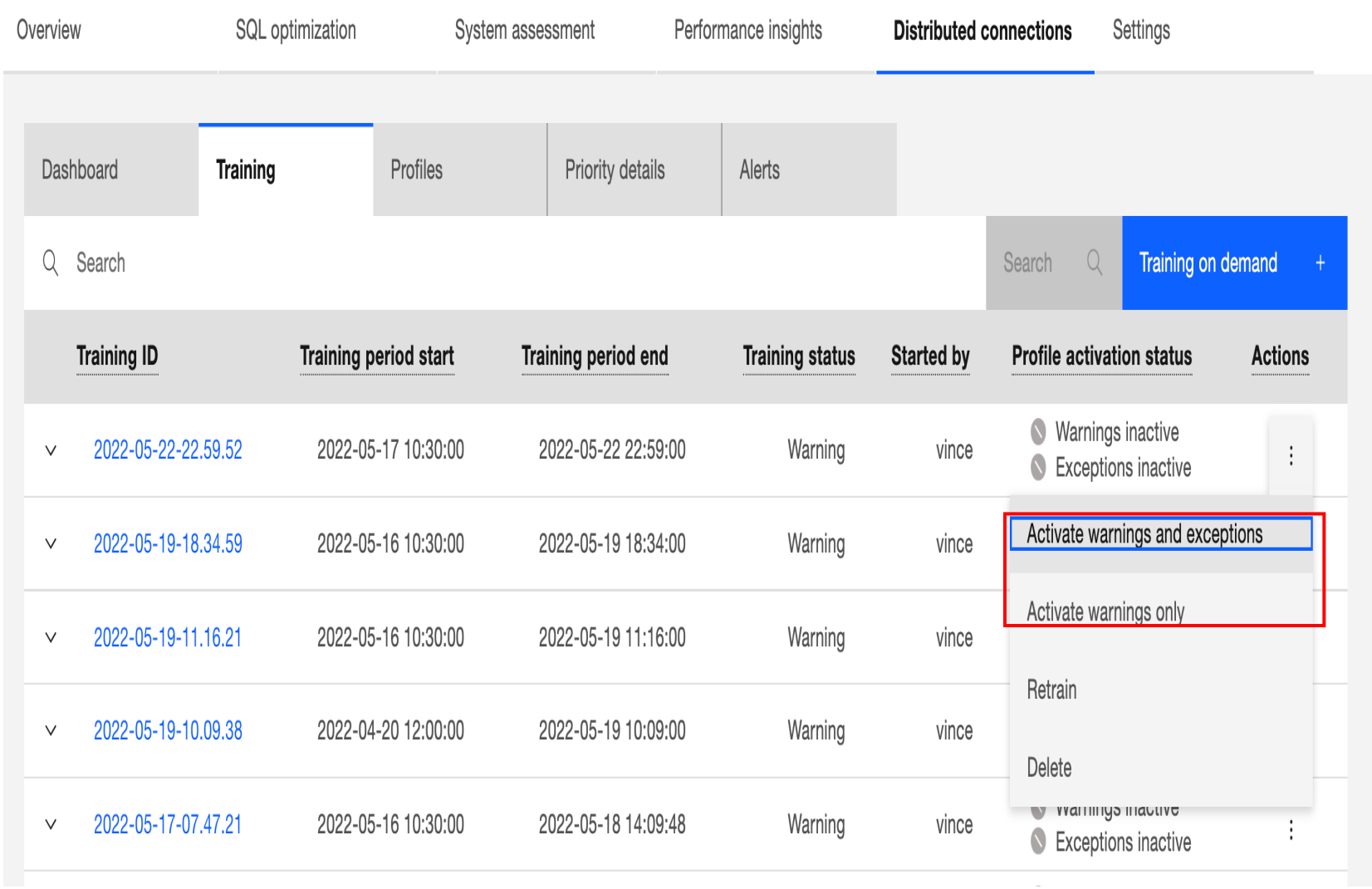
Review and edit recommendations





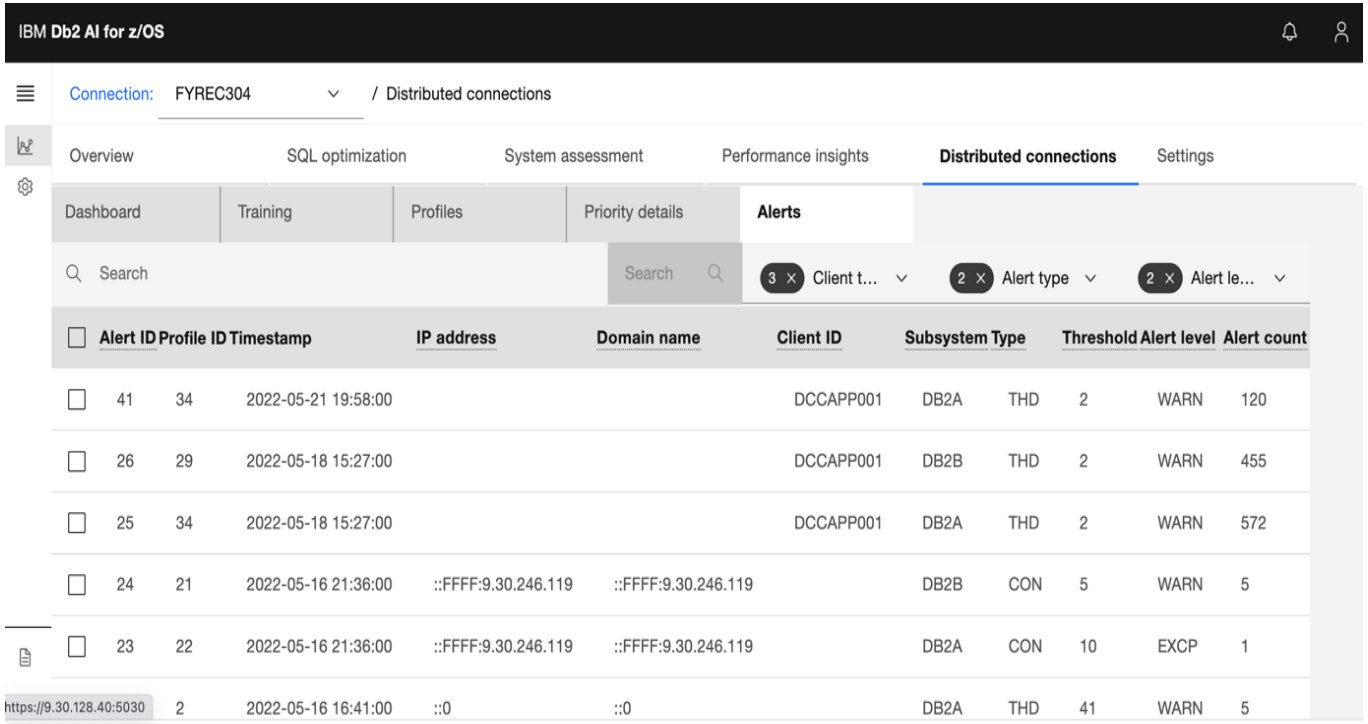
Enabling profile recommendations

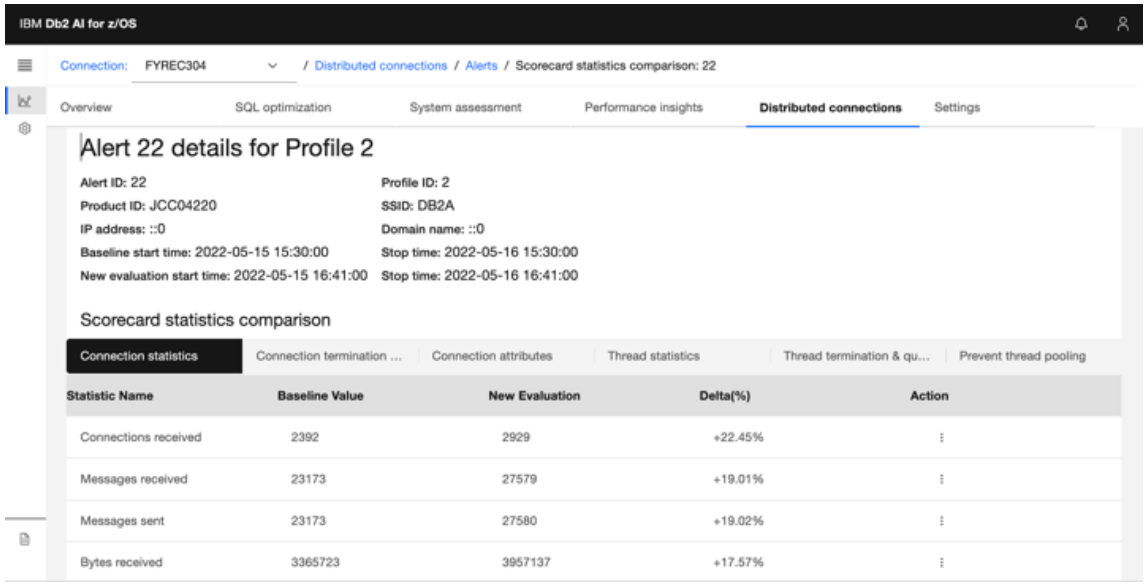
- Two enablement options
  - Warning only
  - Warning & Exceptions

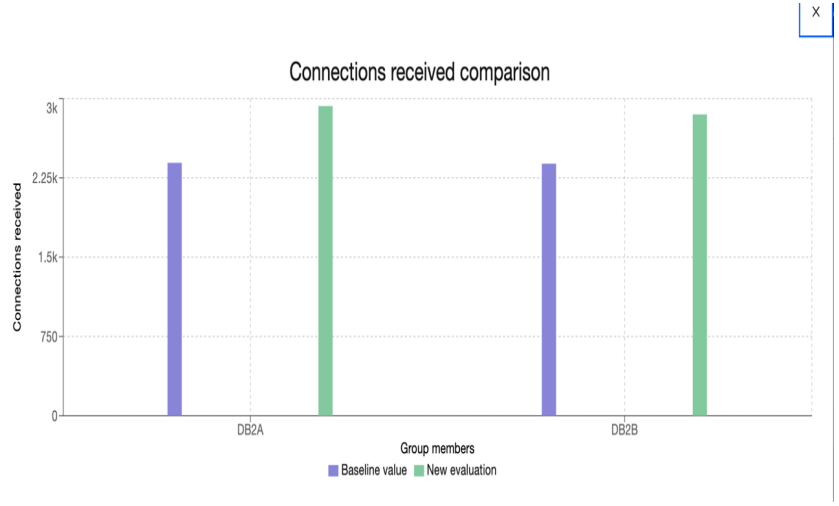


Monitoring Alerts

- Consolidated view of Profile alerts
- Ability to compare data from training with data from alert time





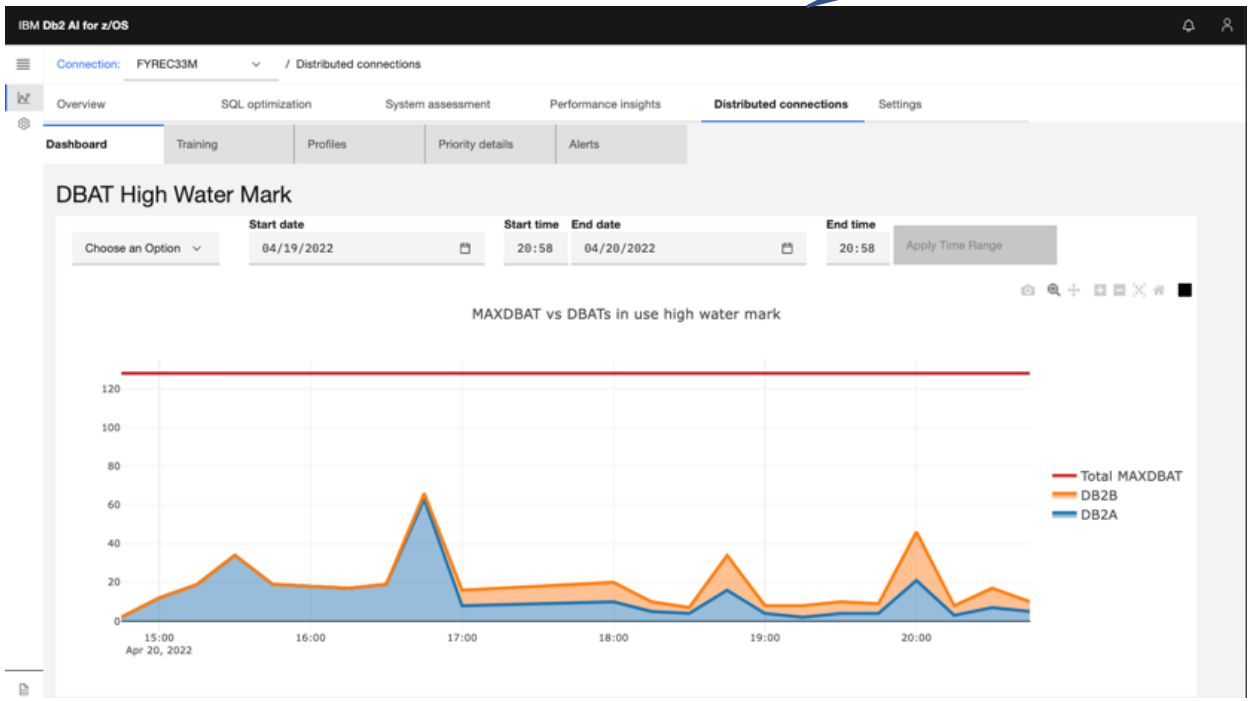


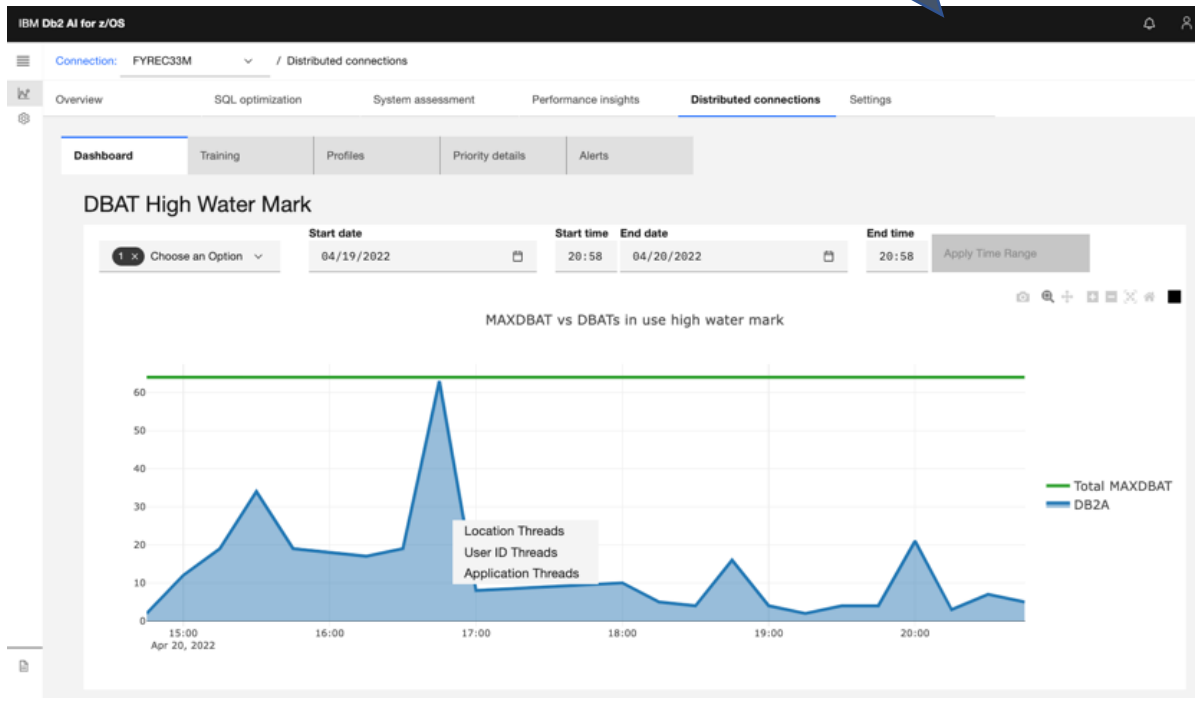
DCC– Intelligent Dashboard for Monitoring Health and Best Practices

- Quick view to monitor stability of inbound traffic to Db2
- High level view of DBATs across Db2 group members

View from a group perspective

View each member

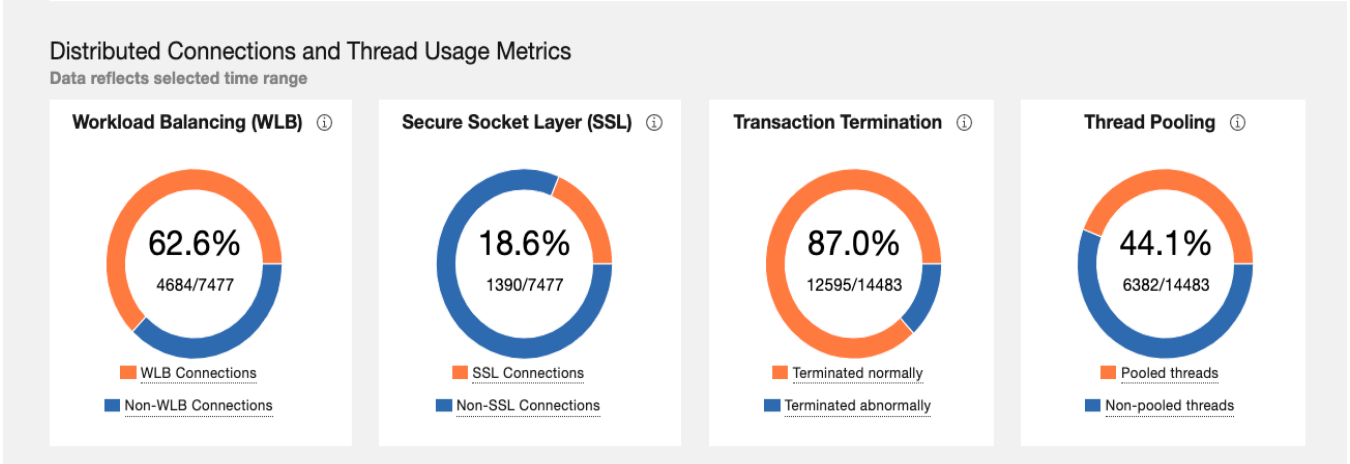


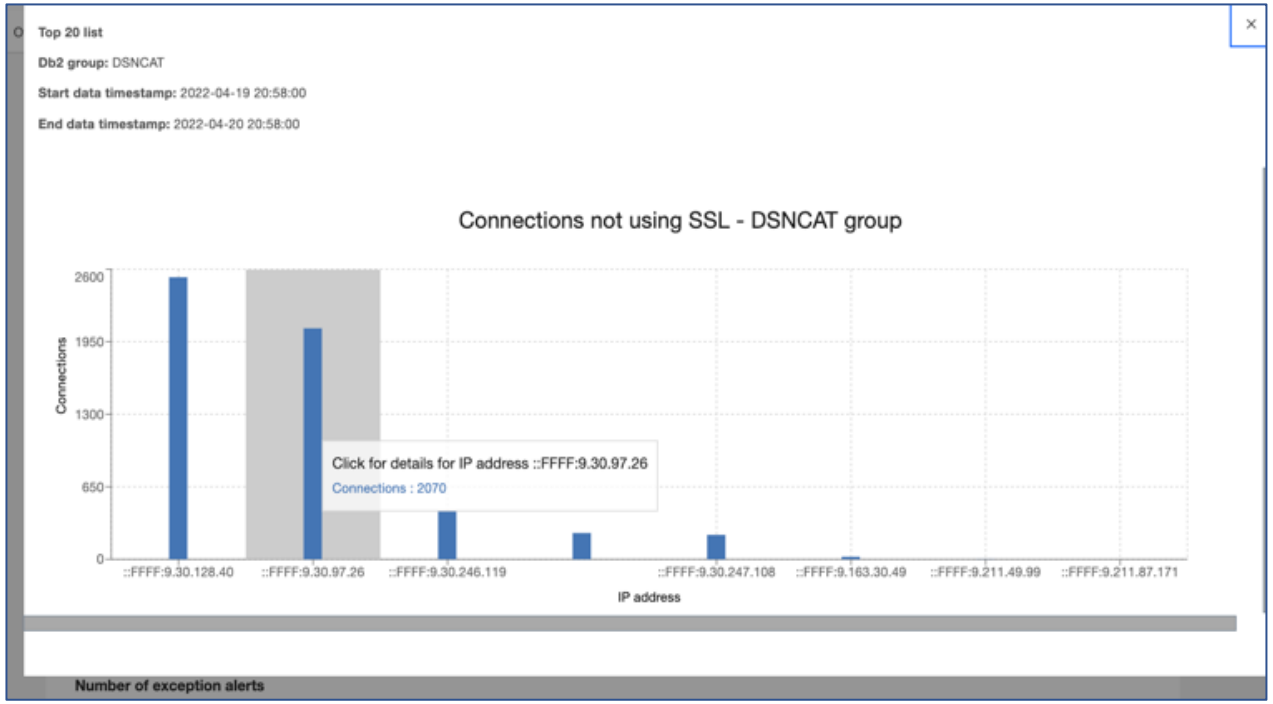


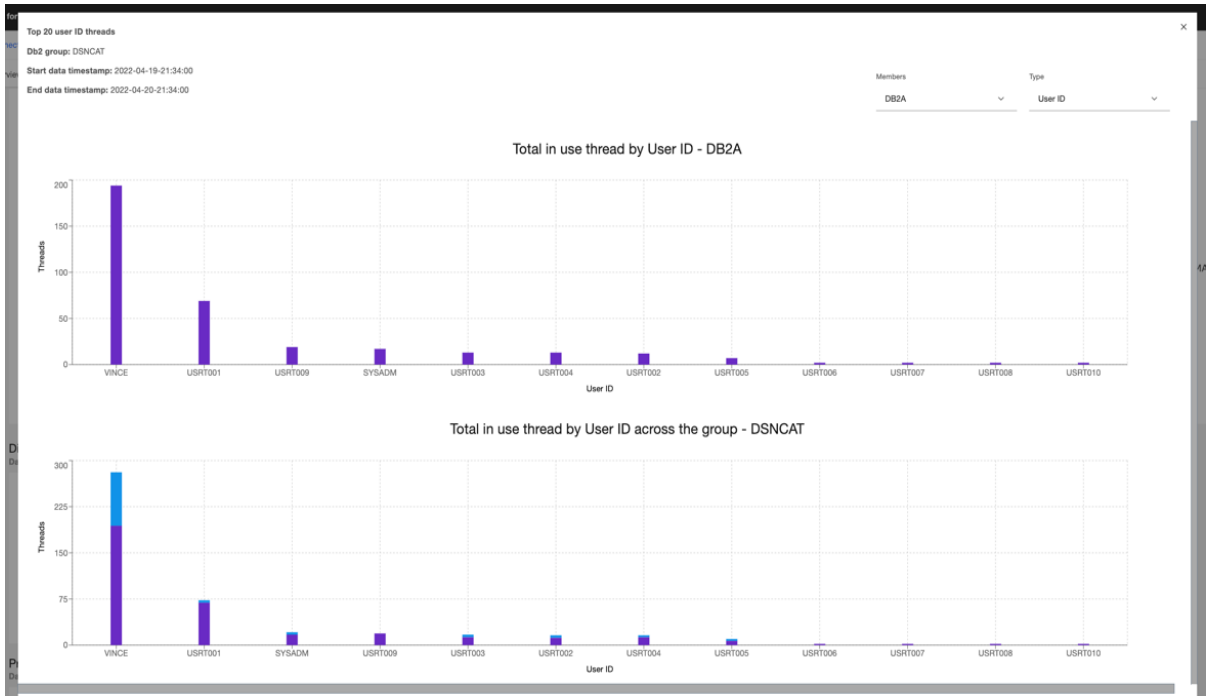
DCC– Intelligent Dashboard for Monitoring Health and Best Practices

- Drill down to identify problematic applications
  - Find top 20 servers not Sysplex enabled
  - Find top 20 servers not using secure connections
  - Find top 20 threads causing transactions to be terminated abnormally
  - Find top 20 threads not allowing pooling

- New visibility into application-level controls for DBAs
  - Improved visibility into on how applications are performing
  - New application-based statistics and connection controls







System Assessment & Performance Insights

Too Many Metrics !

Vast amount of data

Diminishing Skills

System Assessment - Recommended Workflow

Define work periods *(typically one time, or change if workload windows change)*

Schedule training *(typically one time; retrain only when workload changes significantly)*

Schedule assessment or **on-demand** *(typically scheduled to run once daily to process last 24 hour's data)*

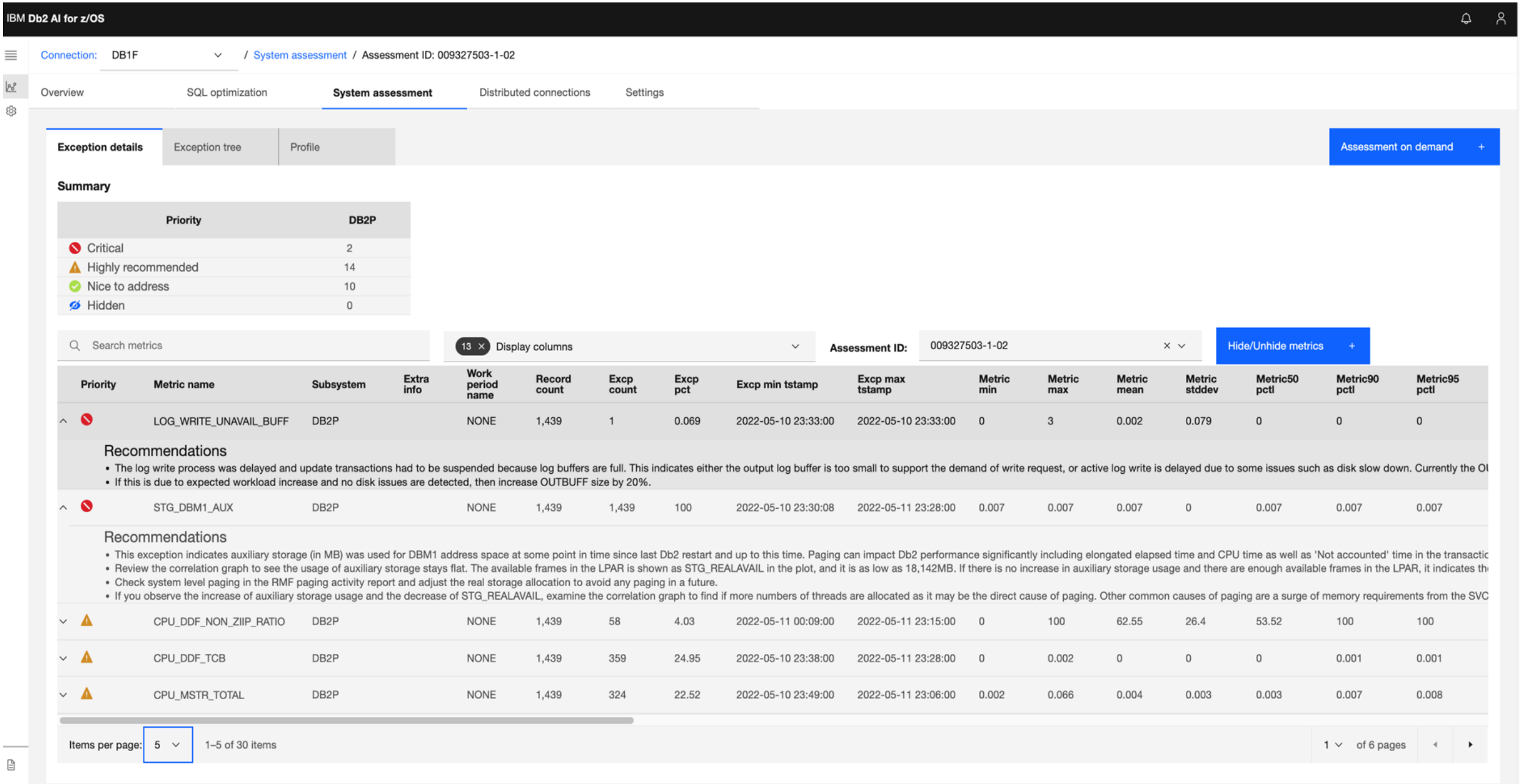
Review assessment results *(interactive visualizations, statistical summaries and recommendations)*

Take recommended actions *(e.g., configuration change - such as zparm updates, buffer pool updates, or hide exceptions)*

Continue to review assessment results *(use daily scheduled assessment results to refine and improve Db2 systems)*

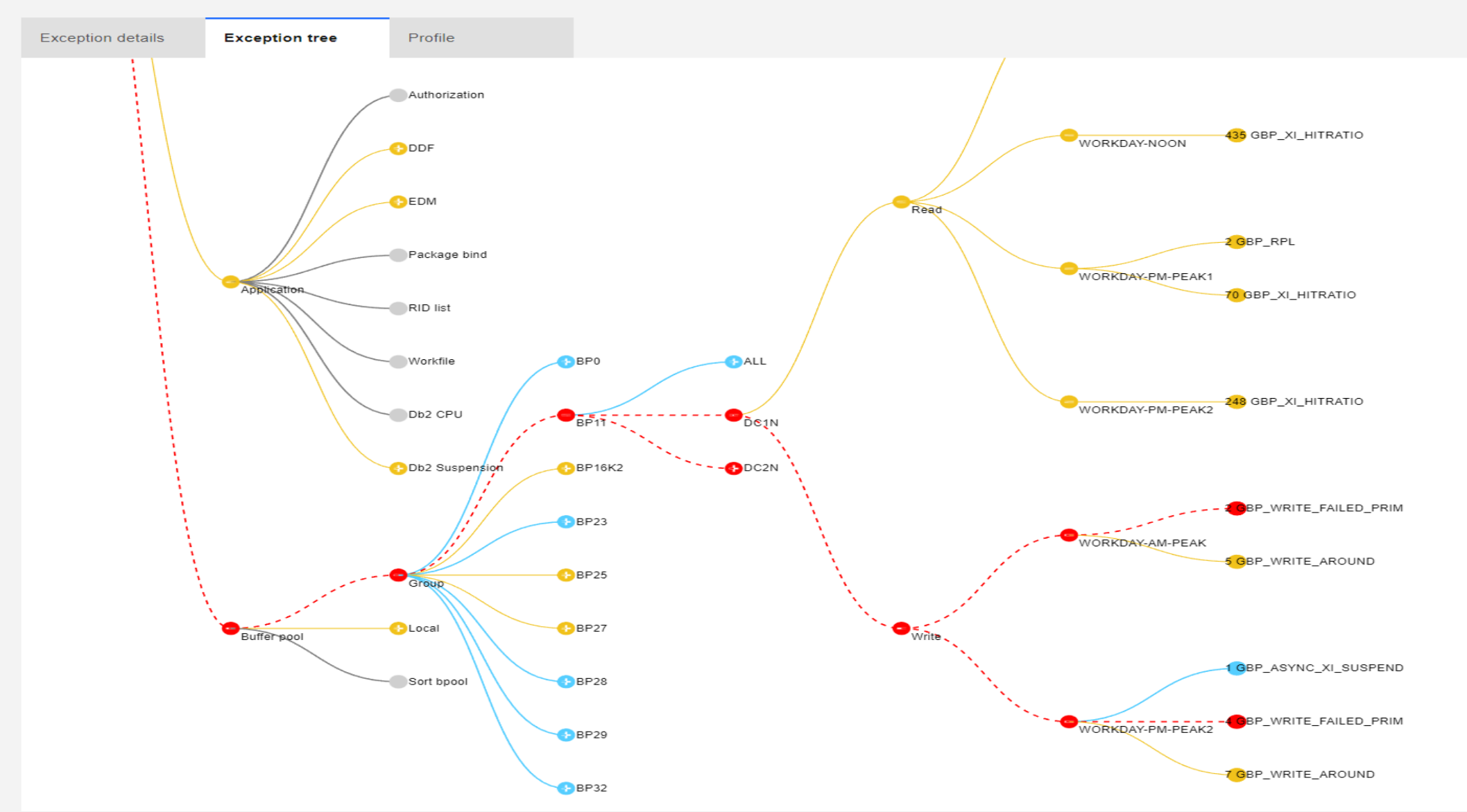
Re-Train as needed (*kick off another training when the system configuration and/or workload changes*)

Exception Summary



System Assessment Exception Tree

Interactive tree view for exception summary and deep dive



System Assessment – Exception Recommendations

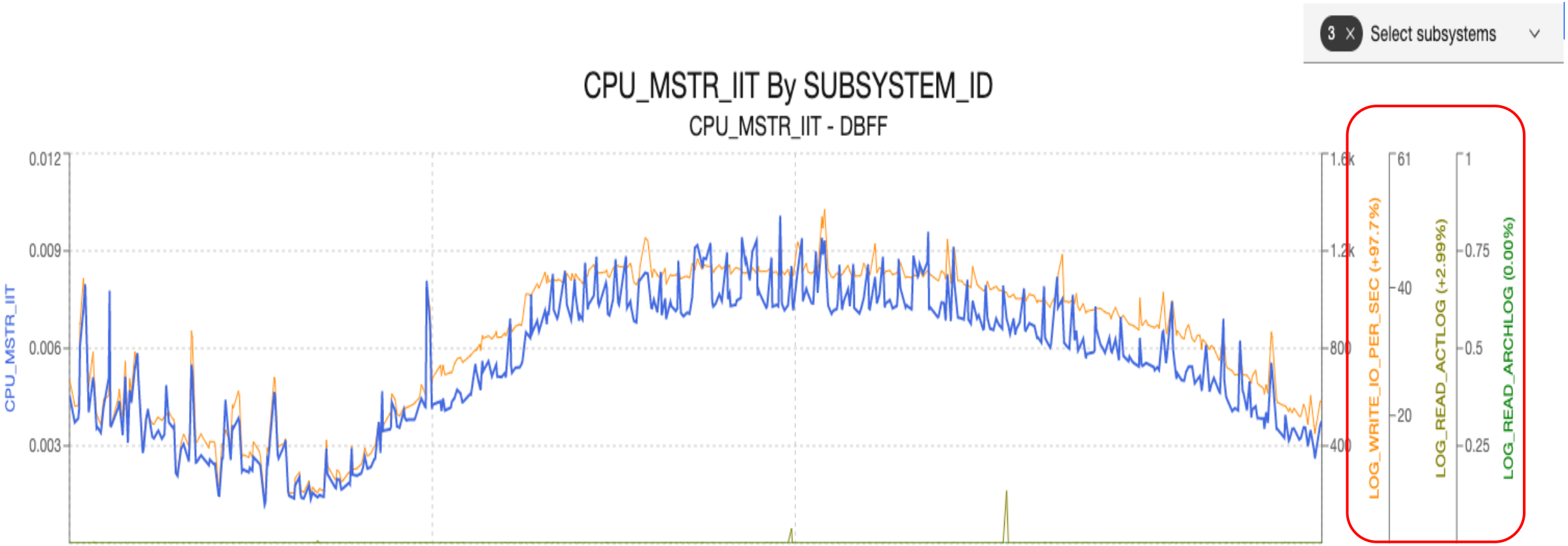
- Recommendations specific to Db2 environment configuration (zparms, bufferpool, etc)
- Guidelines for specific changes, as well as pointers about other related metrics to investigate

Recommendation for exception GBP_WRITE_FAILED_PRIM

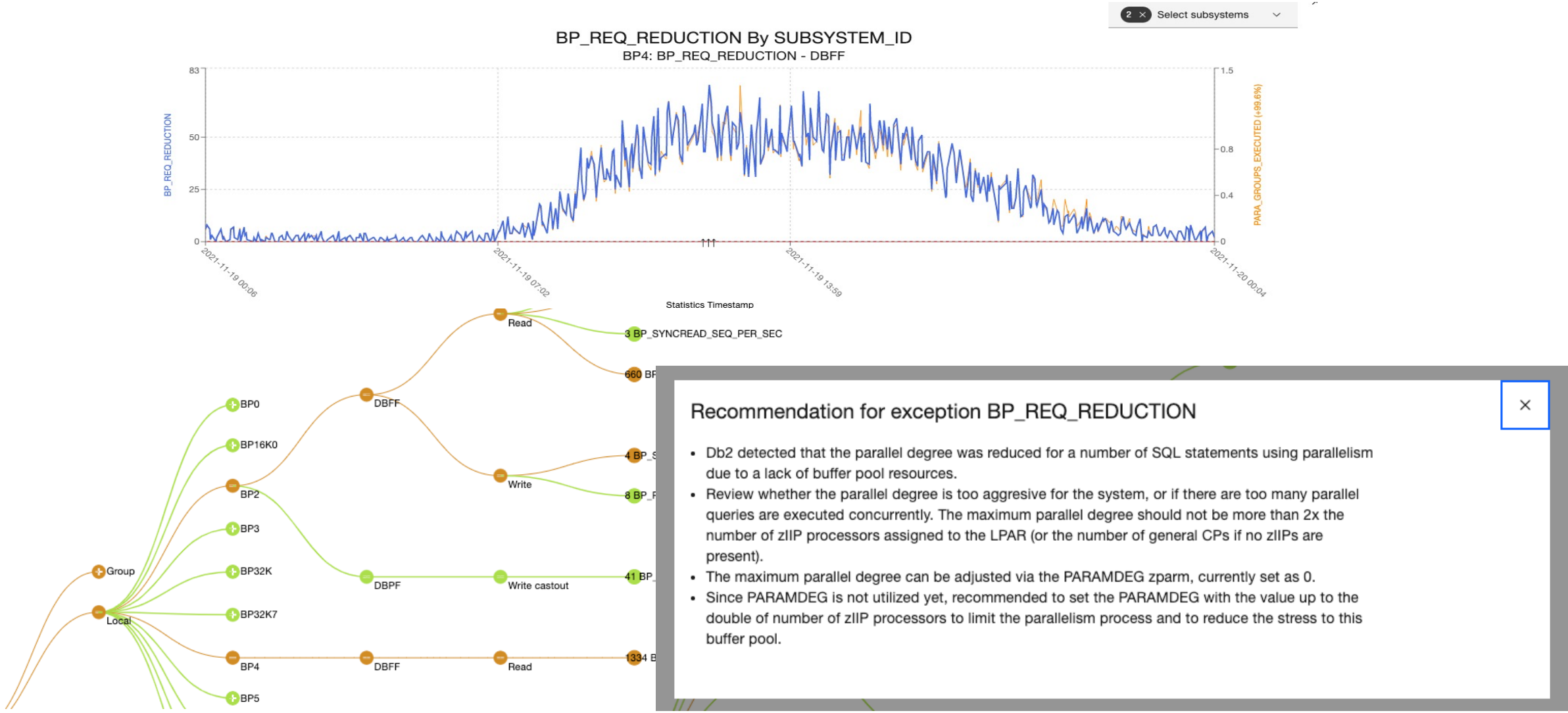
- A number of write failures to the primary GBP have occurred. This indicates that the data pages in the coupling facility are being consumed faster by changed pages than the Db2 castout process can write back to the disk. Db2 will retry the affected write after the castout, but this will impact the application commit time. If retry continues to fail, the pages will be added to the logical page list (LPL) to prevent further access.
- Review the graph for GBP write, castout I/O, and write failures to see whether this is caused by a surge in update (I/U/D) activity from batch or utility. If this is caused by the surge in update activity but an infrequent occurrence, continue to monitor the exception and make sure castout is catching up.
- If this problem occurs frequently, consider reducing or spreading the overall system activity, or decrease the GBP castout thresholds (CLASST, GBPOOLT, GBPCHKPT). Currently castout class threshold (CLASST) is set as (5, 0), group buffer pool threshold (GBPOOLT) is set as 30, group checkpoint interval (GBPCHKPT) is set as 4 minutes.
- If castout activity is happening frequently, then it is possible that disk performance is not catching up with castout activities. Consider increasing the number of data entries in the (primary) group buffer pool by adjusting the directory/data ratio, and/or increase the total size of the group bufferpool.

System Assessment – Exception Plots with Correlations

- Interactive visualizations to understand and investigate system exceptions
- Identify exception thresholds, timeframes for exceptions, correlations

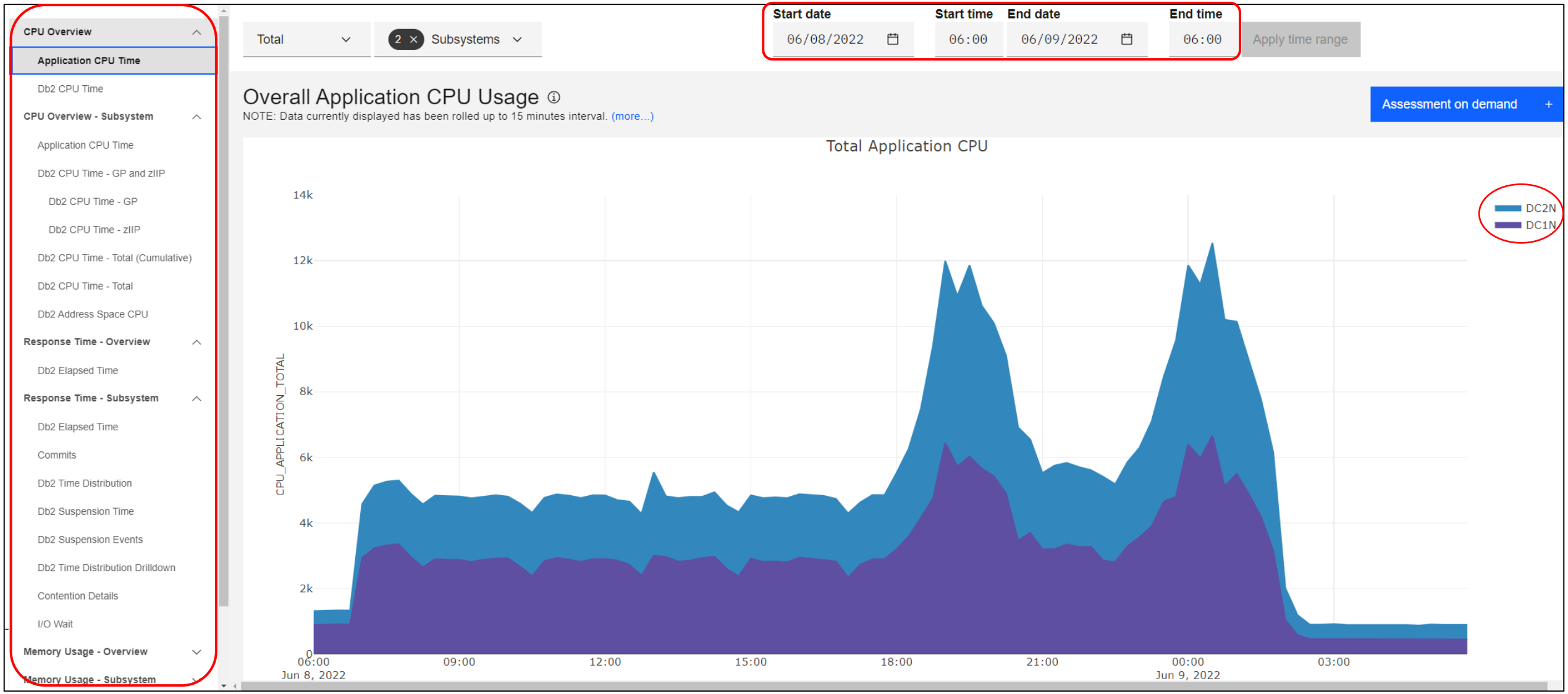


System Assessment Exception and Recommendation



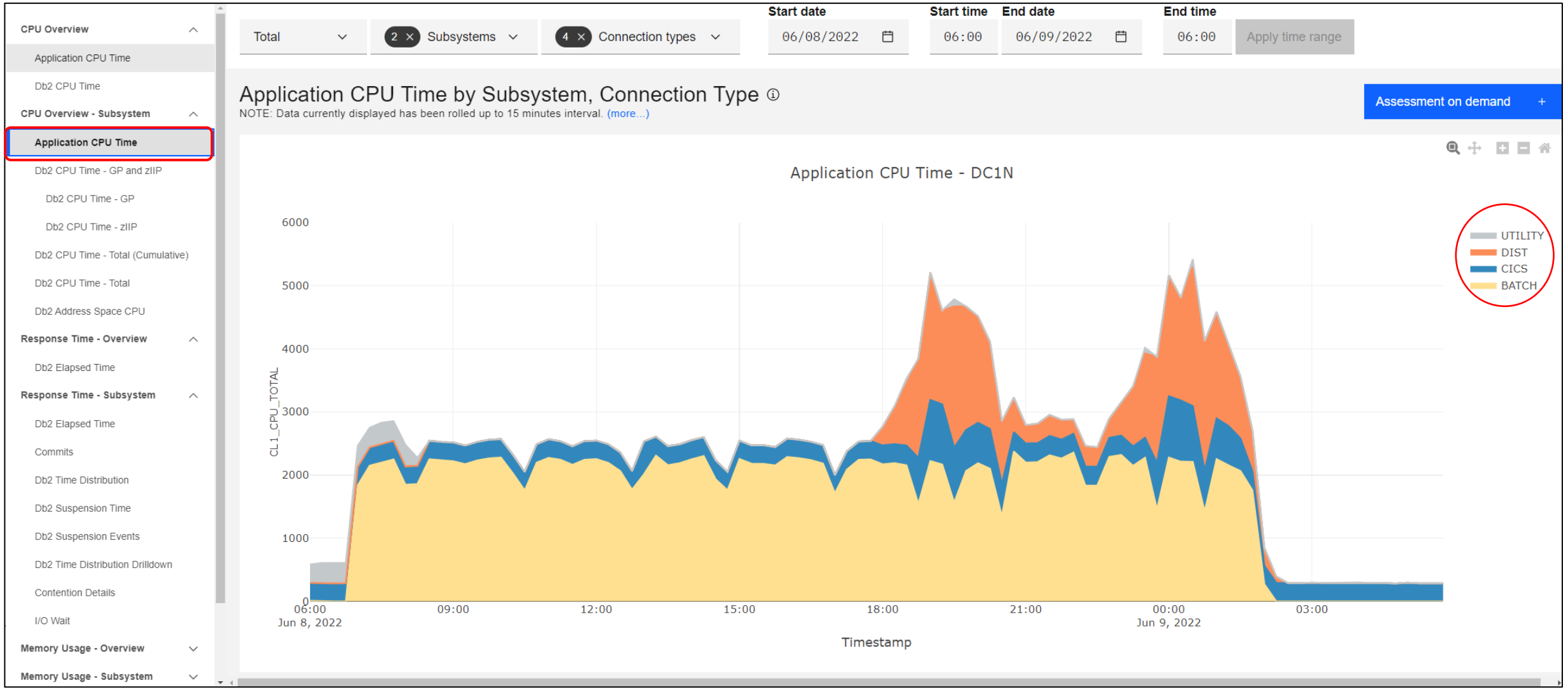
Performance Insights – Top-Down Guided Analysis

- Reports to help in top-down performance analysis



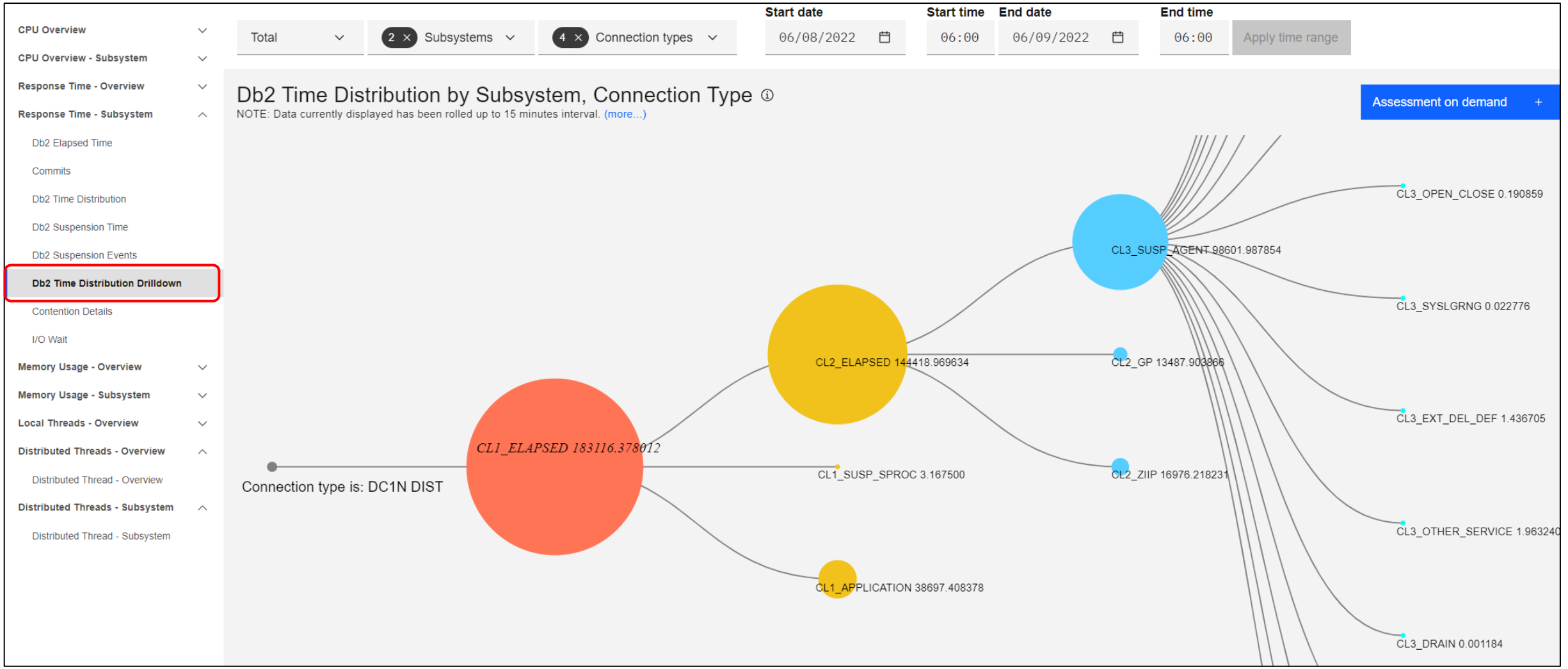
Performance Insights – Top-Down Guided Analysis

- Drill-downs from data sharing to subsystem and connection type metrics



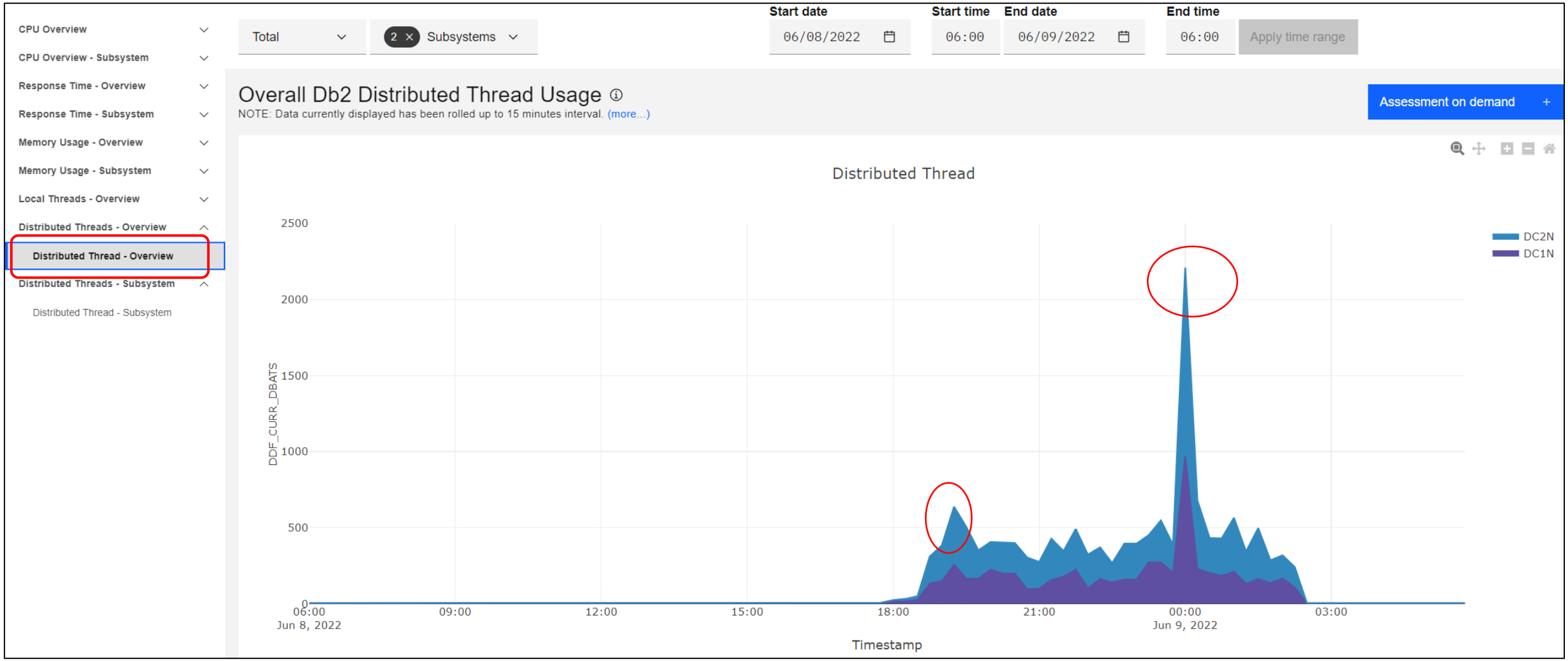
Performance Insights – Top-Down Guided Analysis

- Db2 suspension analysis for response time issues

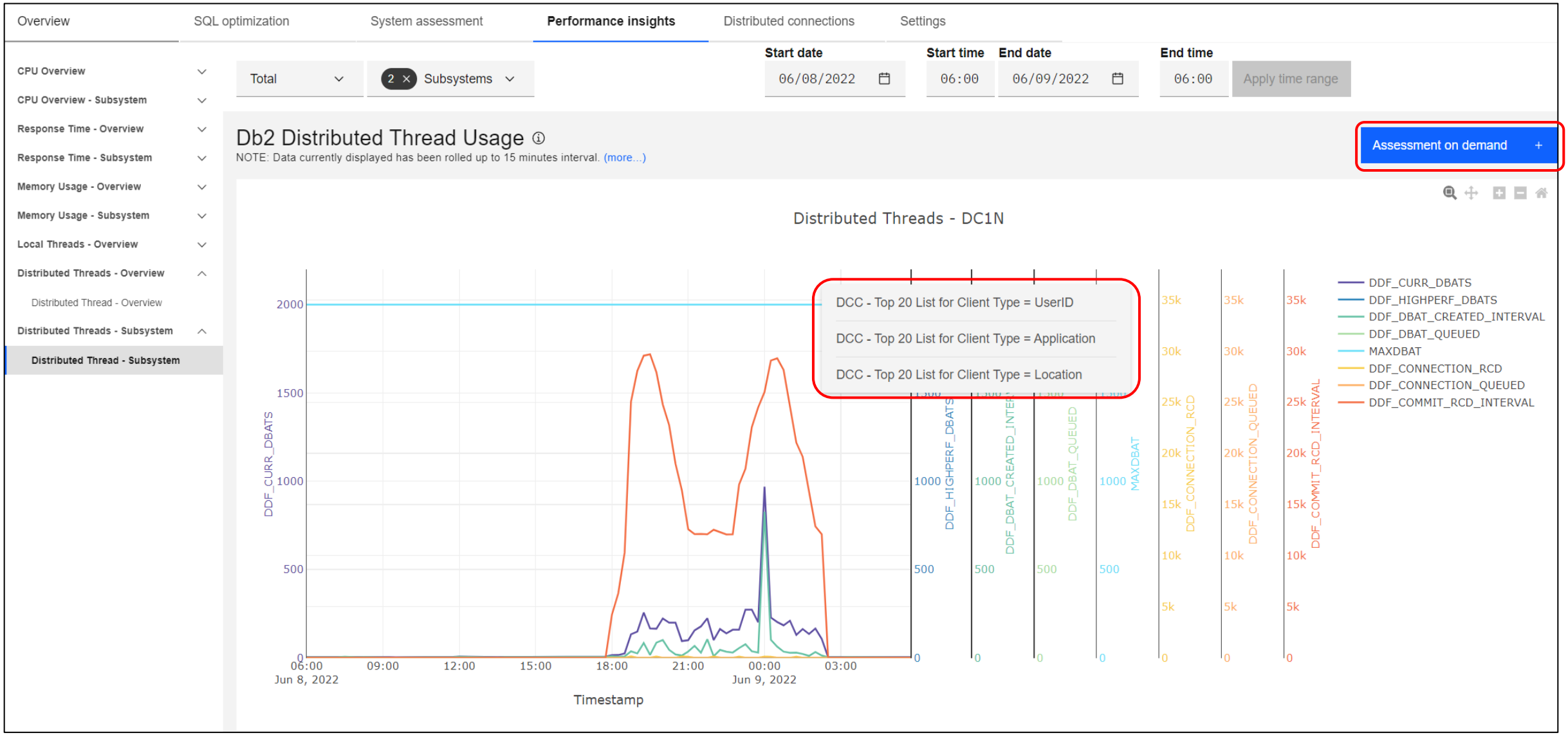


Performance Insights – Top-Down Guided Analysis

- Multiple reports for a given timeframe for a thorough performance analysis

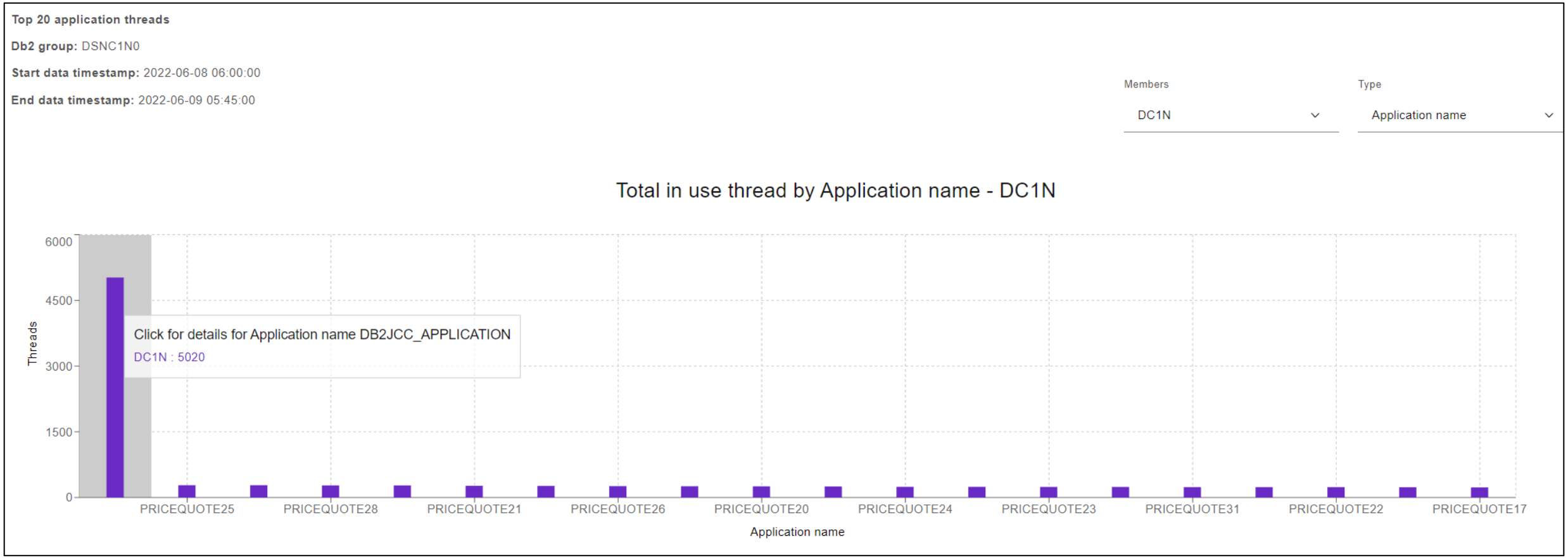


Performance Insights – Drill Down to System Assessment and Distributed Connection Control



Performance Insights – Drill Down to Distributed Connection Control

- Identify top applications contributing to distributed thread usage



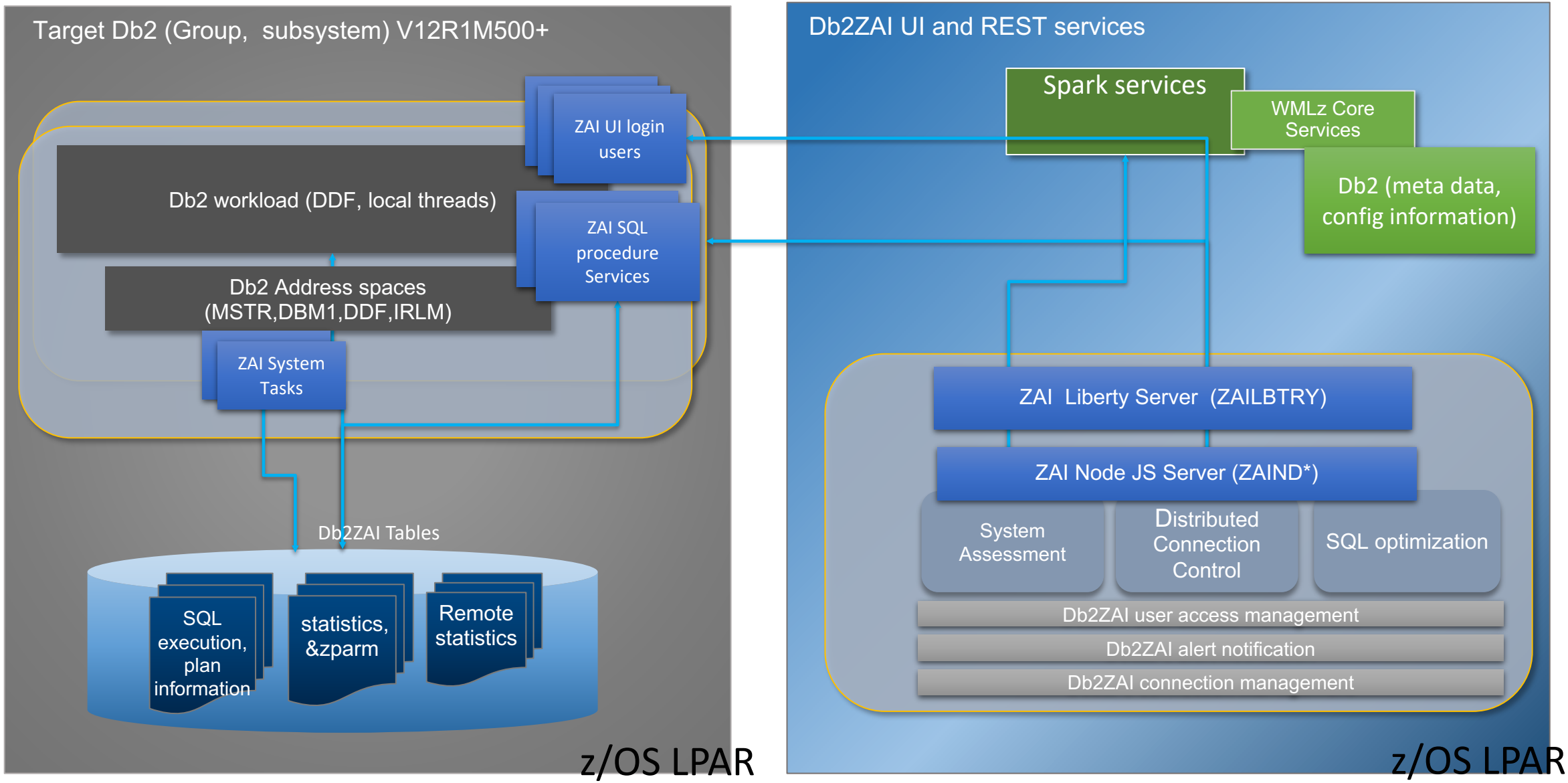
System Assessment Trial (Technology Preview) using your SMF

No charge service to experience one-day System Assessment result using your production SMF data without installing Db2ZAI

1) Determine the date & subsystem or Db2 group to be assessed
2) Send the SMF 100/102 to IBM Db2 using standard case and notify Db2 zAI team
3) IBM will run the in-house assessment
4) Attend the Webex session to review the assessment results

S/W and H/W Requirement

Db2 AI for z/OS – Logical View



S/W Requirement

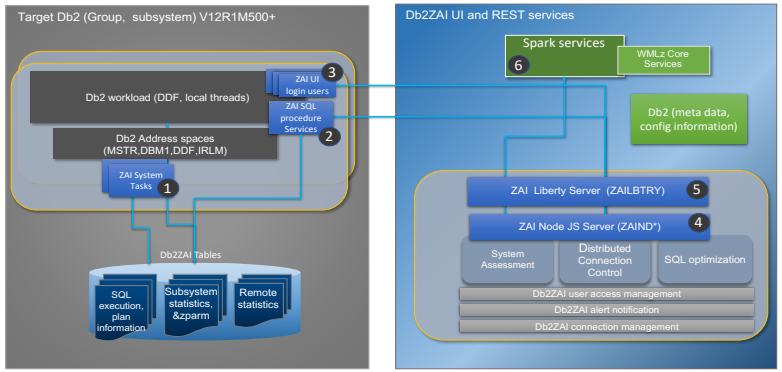
- Db2 AI for z/OS 1.5 (HCOY150)
  - IBM WMLz 2.4 (HAQN240)
    - Embedded Node.js
    - Embedded bash 4.3.48
  - Spark (HSPK120)
    - Anaconda(HANA110) and MDS(HMDS12) are also included in the package (bundled) but no need to install
- IBM 64-bit SDK z/OS Java (HJV800)
- z/OS maintenance
- Db2 for z/OS maintenance

Base product Maintenance Requirement

https://www.ibm.com/docs/en/db2-ai-for-zos/1.5.0?topic=installation-software-needed

Prerequisite	Version	APAR or PTF
z/OS	– 2.4 or higher, plus Product Registration Services or – 2.5 or higher, plus Product Registration Services	For z/OS 2.4 apply the following PTFs: – UI64830, UI64837, UI64839, UI64940, and UI65567 For z/OS 2.5 apply the following PTFs: – UI64830, UI64837, and UI64940
Db2	– 12 (either 5650-DB2 or 5770-AF3), function level 500 activated plus the APARs listed or – 13 (either 5698-DB2 or 5698-DBV), function level V13R1M100 or later activated plus the APARs listed	Db2 12: – PH40243, PH43479, PH44983, PH36105, PH43916, PH41978, PH40041, PH45989, PH46564, PH46960, PH47640, PH47641, and PH46962 Db2 13: – PH44814, PH46031, PH46028, PH46403, PH46603, PH46960, PH47640, PH47641, and PH46962

Software and Hardware Requirement : Summary



Factors influencing the capacity

- SQL Optimization
  - Number of statements in scope (both static and dynamic statements)
  - Number of packages in scope (static only)
  - Complexity of the statements
  - For dynamic statement – frequency of full prepare occurrence (i.e. is the cache large enough to contain the statements that are in scope)
- SA/DCC
  - Numbers of members in a group
  - STATIME_MAIN subsystem parameter
  - Numbers of buffer pools and group buffer pools
  - Work periods
  - Number of IP addresses

Disclaimer:
Capacity projections are based on the performance measurements using standard IBM benchmarks in a controlled environment. The actual capacity requirement or performance that any user will experience will vary depending upon many factors, including hardware and software configuration, workload characteristics, types of UI operations, concurrency, numbers of SQL statements in scope, number of IP addresses connecting to Db2 members. Therefore, no assurance can be given that an individual user will observe results similar to those stated here.

Minimum requirement

Target Db2 (server)

- Db2 12 for z/OS V12R1M500 or later
- zEC12 or later
- Additional 200GB disk space (Db2zAI tables)
- Additional memory to support buffer pools
- Each Db2 member (subsystem) requires 1-2 % of one GP, 4 to 8 % of one zIIP for Data collection

Db2ZAI UI and Rest Service (This does not need to be separate LPAR from above)

- WML for z/OS V2.3 or above
- zEC12 or later
- 10 GB memory (for 20way SA assessment)
- 10-100 GB disk space
- When system assessment executes assessment, it requires up to 2 zIIPs.
- General requirement (without SA execution) is 1-2% of one GP. Each UI operation may add up to additional 1.5% of one GP.

Up and Running Experience

- Unclear documentation
  - Requirement for pre-req products
    - Download & install anaconda/Python
  - Installation & configuration in USS / outside of Db2z domain skills
- Security set up

1. Improved Db2zAI documentation
   Requirement :
   https://www.ibm.com/docs/en/db2-ai-for-zos/1.5.0?topic=installation-software-needed

2. Step by step :
   https://www.ibm.com/docs/en/db2-ai-for-zos/1.5.0?topic=installing-configuring-db2-ai-zos-solution

3. Redpaper: WMLz and Db2ZAI combined
   https://www.redbooks.ibm.com/redpapers/pdfs/redp5643.pdf
   Update for 1.5 in progress

4. Security setup example included ..

5. Assign development focal point to assist all the way

Installing and Configuring IBM Db2 AI for IBM z/OS v1.4.0

An IBM Redpaper publication

Published on 21 July 2021, updated 04 January 2022

Up and Running Experience (Security setup)

https://www.ibm.com/docs/en/db2-ai-for-zos/1.5.0?topic=db2zai-enabling-passticket-support

Enabling PassTicket support for Db2ZAI

Last Updated: 2022-07-27

You need to enable RACF PassTicket support for Db2ZAI. Db2ZAI uses PassTickets for network security.

Overview

The RACF PassTicket is a one-time-only password that is generated by a requesting product or function. It is an alternative to the RACF password and password phrase that removes the need to send RACF passwords and password phrases across the network in clear text.

Environmental scenarios

The RACF commands to be issued depend upon your particular Db2ZAI setup.

The three scenarios covered in this topic are:

- Scenario 1 - the metadata Db2® and the target Db2 are on the same subsystem.
- Scenario 2 - the metadata Db2 and the target Db2 are on different subsystems, but in the same RACF database.
- Scenario 3 - the metadata Db2 and the target Db2 are on different subsystems, and also in a separate RACF database.

Scenario 1 - the metadata Db2 and the target Db2 are on the same subsystem

Issue the following commands. These commands are available in sample job `hlq.SCOYBASE(COYPTK1)`. This job must be customized and run by the security administrator.

Note: **`ACCESS(UPDATE)`** allows the generation of PassTickets. **`ACCESS(READ)`** allows the evaluation of PassTickets.

```
SETROPTS CLASSACT(PTKTDATA)
SETROPTS RACLIST(PTKTDATA)
SETROPTS GENERIC(PTKTDATA)

RDEFINE PTKTDATA <APPLNAME> SSIGNON(KEYMASKED(1234567890ABCDEF)) +
 APPLDATA('NO REPLAY PROTECTION') UACC(NONE)
RDEFINE PTKTDATA IRRPTAUTH.<APPLNAME>.* UACC(NONE)
PERMIT IRRPTAUTH.<APPLNAME>.* CLASS(PTKTDATA) ID(<LIBERTY_OWNER>) ACCESS(UPDATE)
PERMIT IRRPTAUTH.<APPLNAME>.* CLASS(PTKTDATA) ID(<DIST_OWNER>) ACCESS(READ)

SETROPTS RACLIST(PTKTDATA) REFRESH
```

2023 and Beyond

2H 2018 Db2ZAI V1.1 SQL Optimization –Learning SQL patterns

1H 2019 Db2ZAI V1.2 System Assessment – out of box performance tuning recommendation

1H 2020 Db2ZAI V1.3 Distributed Connection Control – manage remote connections

2H 2020 Db2ZAI V1.3.1 Self healing, Dynamic IP and RACF/MFA

1H 2021 Db2ZAI V1.4 HA, SQL OPT Dash board, Integration with vendor data

1H 2022 Db2ZAI V1.5 Application/userid filtering in DCC, Performance Insights, V13

2023 & beyond Db2zAI as AI engine for Db2 engine and Db2 tools

- Streamline SQL optimization to minimize the operational risk of access path updates
  - Simplify access path management and proactively protect regression
- Integration with Query tuning services
- Exploitation of Db2 13 features and metrics
- Adaptive Db2
  - Dynamic workfile management, Adaptive MAXDBATs, etc .

Self healing and tuning - SQL workload and Db2 systems

Questions?

Akiko@us.ibm.com