Ed Woods, IBM Corporation
woodse@us.ibm.com

Session#15839
Friday, August 8, 2014: 10:00 AM-11:00 AM

Predictive Analytics And IT Service Management

Agenda

- What is Predictive Analytics?
- Examples
- How is predictive analytics relevant to IT Service Management?
- Typical monitoring and management paradigms
- Real time information versus historical data collection
- Univariate versus multivariate analysis
- Examples of relevant metrics
- Components of a solution
- Roadmap

What Is Predictive Analysis?

- An area of analysis that deals with extracting information from data and using it to predict future trends and behavior patterns
- Relies on capturing relationships between explanatory variables and the predicted variables from past occurrences
  - Exploit the information to predict future outcomes
- Accuracy and usability of results will depend greatly on the quality of data analysis and the quality of assumptions
- Predictive analysis is used in many facets of business
  - Common example would be credit score
    - Function of many data items
    - Income, payment history, amount of outstanding debt, etc…

Analytics Has Become Mission Critical
Impacts Bottom Line Results Across All Industries And IT

Industries

Banking

- Increase account profitability

Insurance

- Retain policy holders with better service & marketing

Retail

- Understand sales patterns

Telecommunications

- Reduce churn with custom retention offers

Operations

Industrial

- Predict maintenance issues before occur

Retail

- Improve store performance with P&L reports

Telecommunications

- Understand & manage network traffic

Insurance

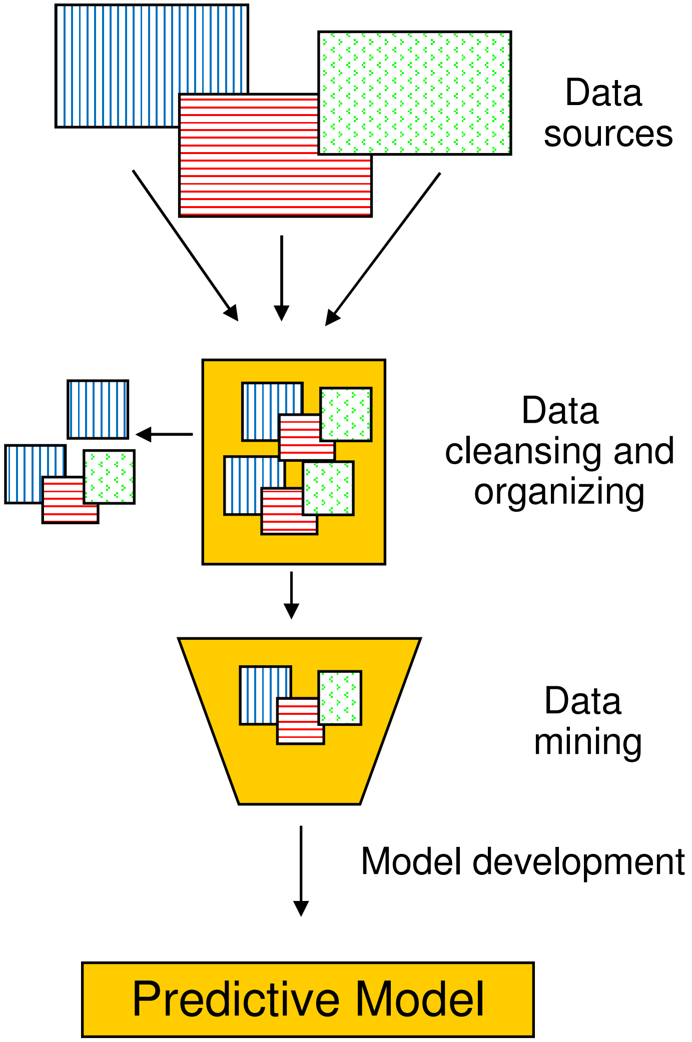
- Streamline claims process

Government

- Reduce fraud and waste

Steps In The Predictive Analytics Process

- Data organization and cleansing
  - Identify data sources
- Data Mining
  - Analysis of data to identify underlying trends, patterns, or relationships
  - Identify data to be used to develop the predictive model
- Model Development - Regression models
  - Regression modeling describes the relationship between dependent variable (the variable to be predicted) and independent explanatory variables
  - Regression models imply some level of causation (versus correlation)

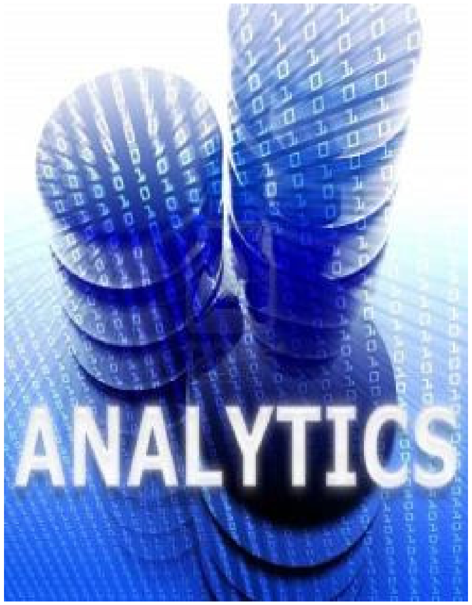


Predictive Analytics
About Regression Models And Types Of Models

- Regression models are the core of predictive analytics
- A wide variety of models can be applied
  - Linear regression model
    - Analyzes the relationship between the response or dependent variable and a set of independent or predictor variables
  - Partial or Stepwise regression
    - Modeler does not specify all the explanatory variables
    - Variables are added iteratively
  - Logit or Probit regressions
    - Allow one to predict a discrete outcome (yes/no) from a set of variables
  - Time series models
    - Used for predicting or forecasting the future behavior of variables
    - Data points taken over time may have an inherent time relation
    - Developed to decompose the trend, seasonal and cyclical component of the data
  - Many more models……

Analytics for System z –
Enhanced Search, Optimize, and Predict technology

- **Huge amount of critical IT operational data** (SMF, log, journal) .. More than distributed-only environments.
  - Focus on problem determination and time to resolution while placing premium on availability of services and applications**.**



- **90% of the Fortune 1000 companies are running z** and have 'Systems of Record' dependencies for transactional processing and data serving applications .

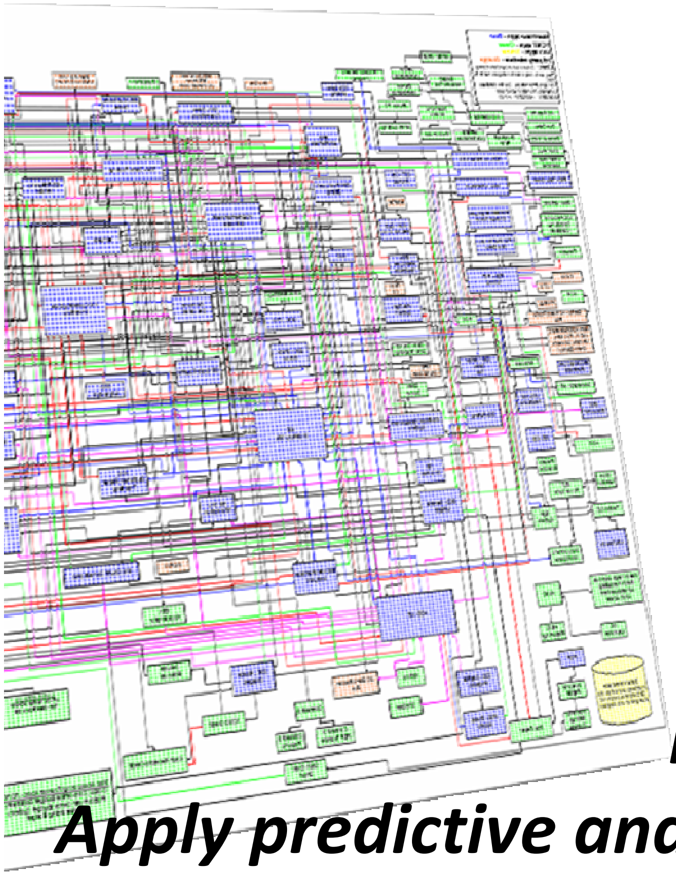
- By 2016, **20% of Global 2000 enterprises will have IT operations analytics** architecture in place, up from < 1% today, looking to integrate across their enterprise to reduce outages (Gartner).

Examples Of Predictive Analytics Commonly Applied to IT

- Performance modeling
  - z/OS workload right sizing and load balancing
    - Model workload placement using SMF data as input
- Trending and forecasting of workload/resource utilization
  - Workload performance trends
    - Discern patterns in resource utilization
    - Capacity planning
  - The common question >> When will a critical resource reach breaking point?
- ‘What If’ Analysis examples
  - DB2 buffer pool analysis
    - DB2 performance trace data to determine optimal pool sizing and object placement
  - DB2 SQL and object tuning
    - DB2 Explain analysis based on DB2 Catalog statistics and SQL call changes

What Is IT Service Management?

Customers want to improve this….



Why?

Integration

Manage Exponential Change

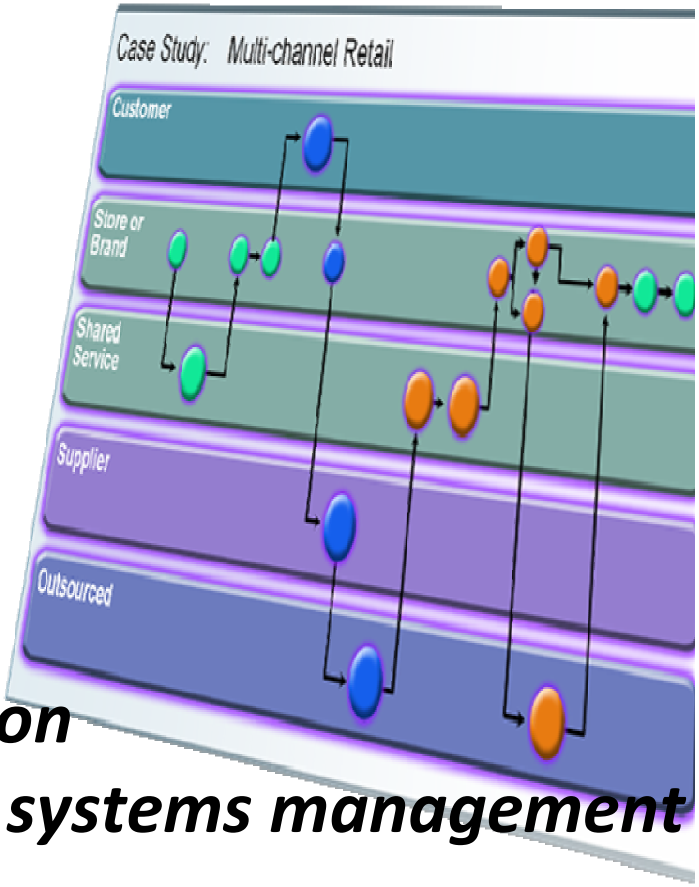
Reduce Complexity

Ensure Compliance

Reduce Cost

Improve Control

… to run their business like this.



Focus on integration

Apply predictive analytics methods to systems management

A Goal For Many Shops
Make Systems Management More ‘Proactive’

- In many shops systems management tends to be done ‘ad hoc’
  - Some alert generation – varies by shop
    - Some shops very alert driven – many are not
  - Often notification consists of ‘call the help desk’
- Many customers want to be more ‘proactive’
  - Definition of proactive may vary
    - Proactive for some installations may mean more rapid alert and notification of technical and/or business application issues
    - Proactive for some installations may mean notification ***prior*** to the problem
      - Alert when utilization indicates a potential issue in the future
      - Alert when I’m within 90% of the wall

The Typical Monitoring Paradigm

- Traditional monitoring strategy
  - Monitor key resources based upon established ‘best practices’
    - Resource utilization and resource bottlenecks
  - Monitor performance and availability
    - Key Performance Indicators (KPIs)
      - Examples – Response time, transaction rate, technical component, software subsystem, or business application availability
    - Monitor based on established SLA’s
  - Alert notification about performance bottlenecks and outages
    - Notification via monitoring UIs, paging, emails
- Real time monitoring versus historical
  - Real time monitoring for current utilization and status
  - Historical data collection for trending and after the fact analysis

Most shops monitor – but how predictive is it?

Problem Analysis And Resolution In Many IT Environments

- Problem identification and notification may be ad hoc
  - Alert notification via phone calls, emails, or paging
- Problem analysis is often after the fact
- Problem analysis and resolution often involves rounding up the usual suspects (and getting them to confess)
- Issue resolution relies heavily on the knowledge and intuition of the technical staff
  - Knowledge of the systems and business applications
  - Understanding ***complex problems*** will be ***multivariate*** in nature

The Problem:
Traditional Monitoring Approaches Have Limitations

- Many tools, data sources and metrics available
  - Many are Resource/Single Metric Focused (Univariate)
- Often many missed, or misinterpreted events
- In many shops not enough time, and/or resources to correlate completely
  - May require many people and groups to collaborate effectively
  - Many resources and no obvious resource inter-relationships

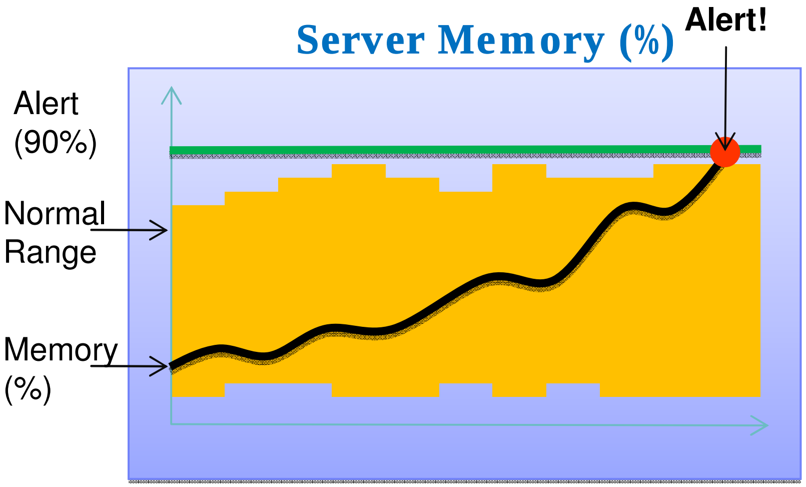
Univariate - refers to an expression, equation, function or polynomial of only one variable

Multivariate - encompasses the simultaneous observation and analysis of more than one statistical variable

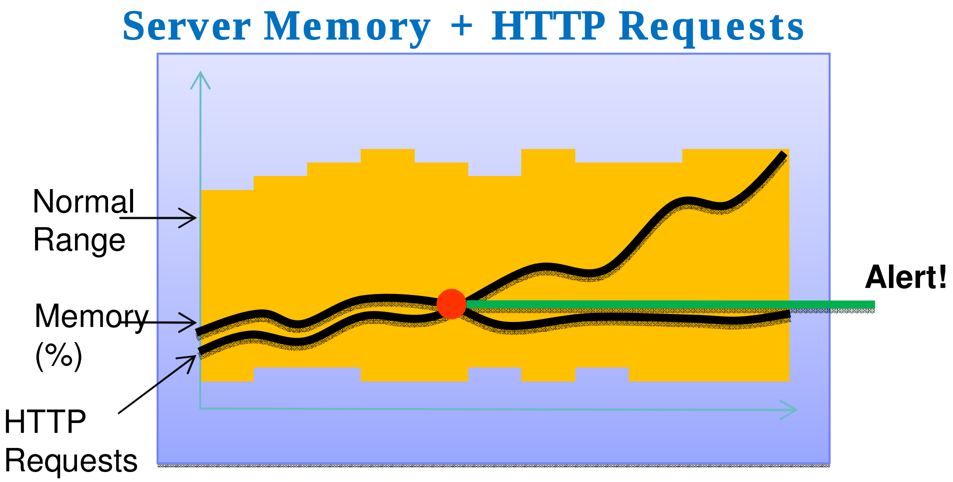
Why Multivariate Analysis?

- Multivariate analysis expands the relevance of the predictive analytic approach
  - Provides context through correlation
- Example – credit rating metrics
  - Payment history – how relevant if I do not consider other metrics?
  - Income – again how relevant if I do not consider other metrics?
- Multivariate is important for IT Service Management
  - Many business applications are composite in nature
    - Many components, platforms, core technologies
  - Many critical resources are shared and inter-related
    - Mainframes support many applications
    - Networks may support a wide array of workloads

Multivariate Analysis In An IT Context – An Example



Static Threshold = Short Warning



Multivariate = Alerts earlier on Deviation

Multivariate analytics detects problems sooner by detecting the deviation of metrics that normally move together.

For example:

- **Memory consumption is normally correlated to HTTP requests**

- **But when memory deviates from HTTP Requests, as would happen with a memory leak, this indicates a problem and an alert is generated.**

- **The alert is generated much sooner than waiting for a static threshold violation.**

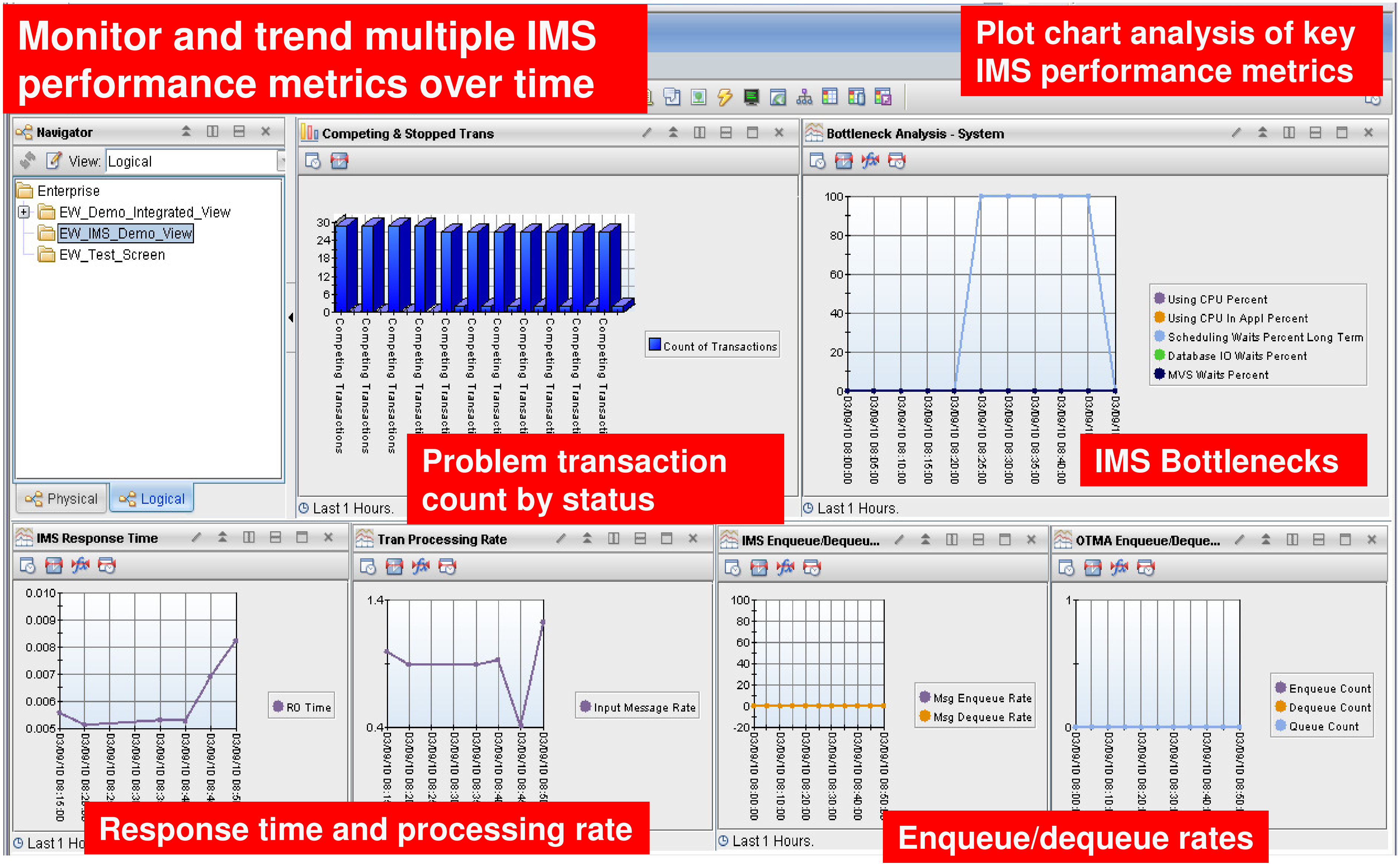
This advanced warning time helps you become proactive and mitigate damage before customer service is impacted.

It also help reduce threshold alerts due to normal threshold violation correlated with HTTP Requests.

Examples Of IT- related Multivariate Metrics

- DB2 example
  - DB2 object lock conflict >>
    - long running SQL call >> high In-DB2 time >> longer thread elapsed time >> longer DB2 query time
- IMS example
  - High IMS message region occupancy time >>
    - IMS transactions queued >> longer IMS transaction scheduling time >> longer IMS response time >> lower IMS transaction processing rate
- MQ example
  - Lower MQ message input rate >>
    - Higher MQ message queue depth >> lower transaction processing rate >> longer CICS/IMS transaction response time

An Example Of Multivariate Analysis For IMS Performance



Monitor and trend multiple IMS performance metrics over time

Plot chart analysis of key IMS performance metrics

Problem transaction count by status

IMS Bottlenecks

Response time and processing rate

Enqueue/dequeue rates

Identifying The Critical Metrics
Defining The Data Sources

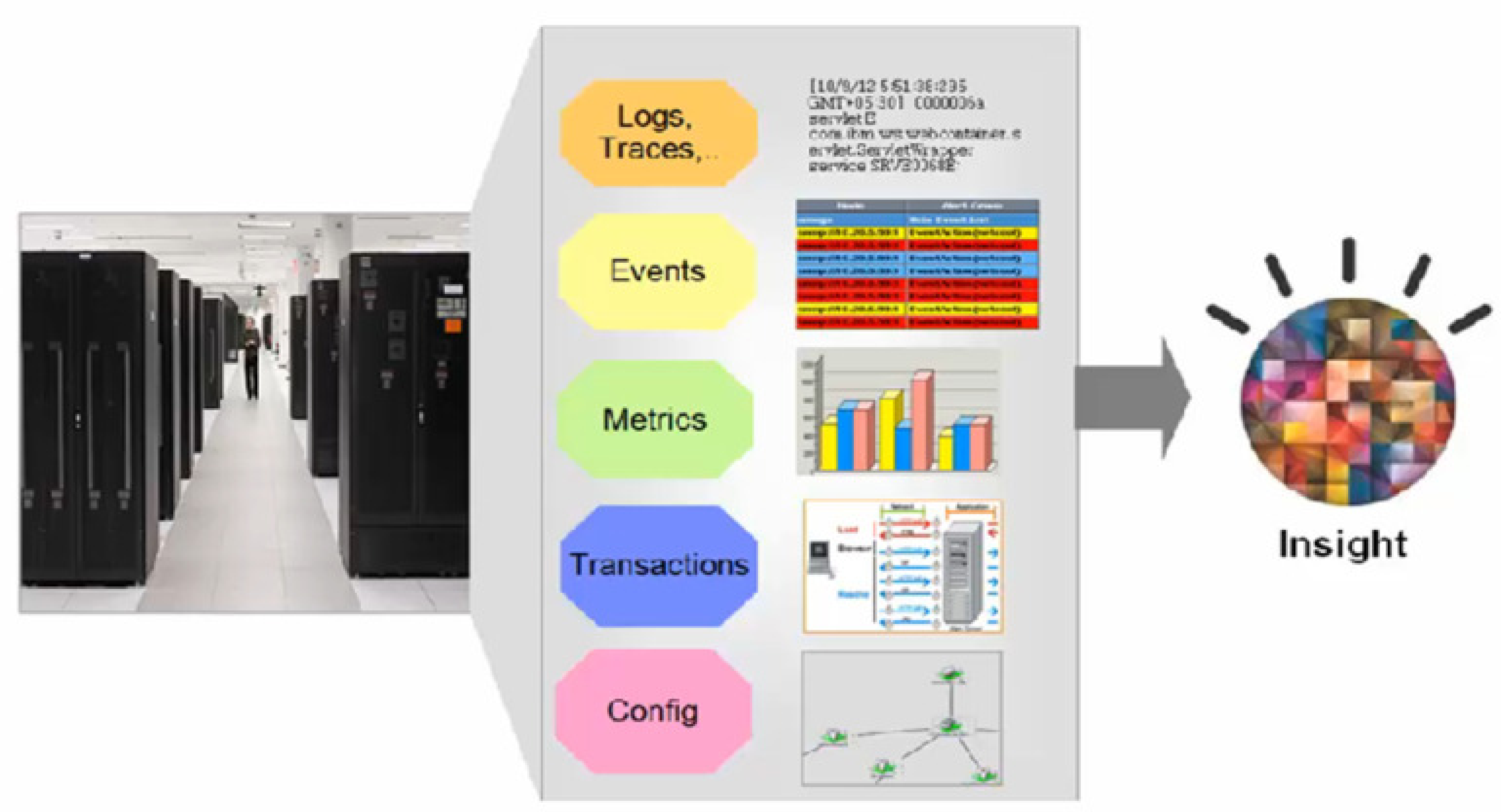
- Knowledge of business applications
  - Internal operational processes
  - Known issues based upon prior operational experience
  - Maintaining a history of common alerts/events
- Identify critical performance metrics as established by ‘best practices’ documented in commonly available sources
  - IBM documentation and IBM Red Books
  - Share, CMG, IDUG, Pulse, IOD and other user group presentations
- Define a list of the most critical metrics to track
  - Consider each component/platform for the application(s)
  - Consider various data sources
    - Monitoring, automation, console logs, application data sources

Predictive Analytics
Categories And Sources Of Information For Analysis

- Messages and events
  - System console messages
    - z/OS console messages, CICS, IMS, MQ messages
    - System message logs from open systems sources
  - Application message logs (including error messages)
  - Various abend and error messages
- Alerts
  - Alerts from various monitoring sources
- Monitored metrics
  - Real time monitoring – critical system and resource metrics
  - Historical monitoring and collection
    - Critical system and resource metrics collected for historical analysis
    - Detail and summary historical data

The Challenges Of Message Management

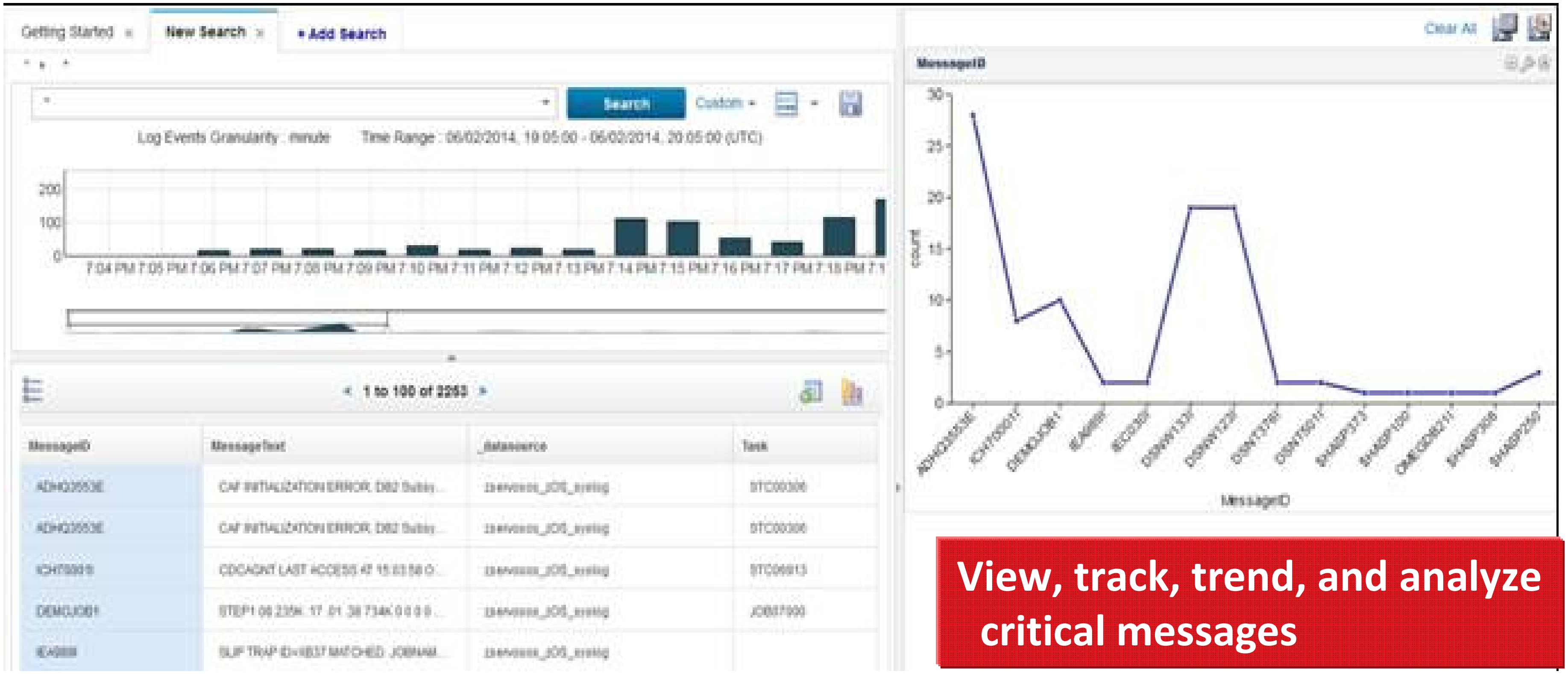
Operators and subject matter experts are overwhelmed with **volumes of data** that they **manually process** to determine the cause, location and scope of a problem.



```
[10/9/12 5:51:36:235
GMT+05:30] 0000036a
servlet E
com.ibm.ws.webcontainer.s
ervlet.ServletWrapper
service SRVE0068E
```

- Only 3% of the data generated is operations-oriented metric data
- 97% is unstructured/semi-structured data
- An enterprise with 5000 servers generates over 1.3 TB of data per day

Messages Provide Important Input To The Analytics Process



1 to 100 of 2253

MessageID	MessageText	_datasource	Task
ADHQ2553E	CAF INITIALIZATION ERROR, DB2 Subsy...	zservices_zOS_syslog	STC00306
ADHQ2553E	CAF INITIALIZATION ERROR, DB2 Subsy...	zservices_zOS_syslog	STC00306
ICH70001I	CDCAGNT LAST ACCESS AT 15:03:58 O...	zservices_zOS_syslog	STC06913
DEMOJOB1	STEP1 00 235K 17 .01 38 734K 0 0 0 0 ...	zservices_zOS_syslog	JOB07000
IEA989I	SLIP TRAP ID=X0B37 MATCHED. JOBNAM...	zservices_zOS_syslog	

View, track, trend, and analyze critical messages

- Messages highlight issues and events IT platforms
  - z/OS subsystem messages, application errors, abends, notifications, alerts

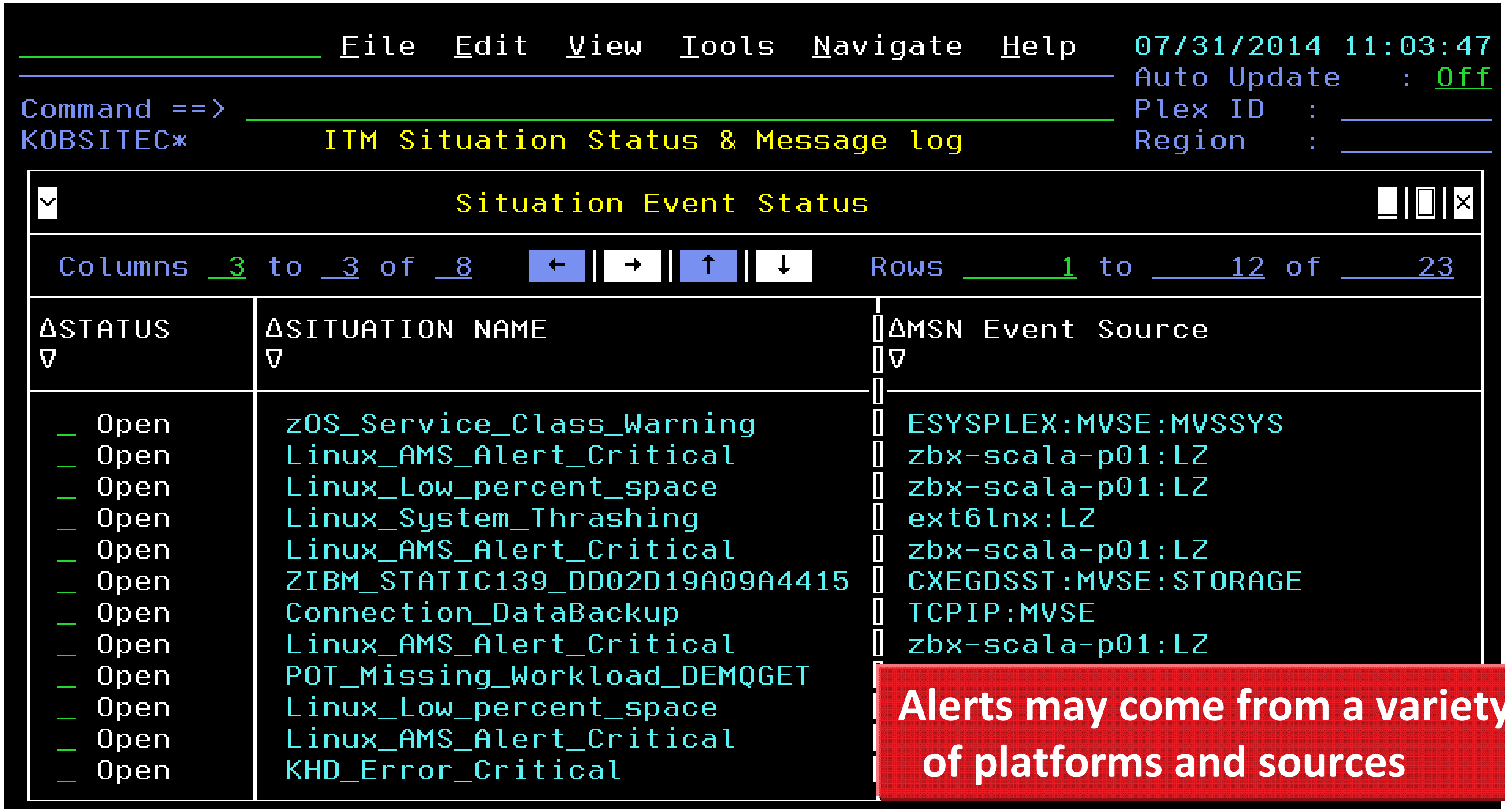
The Importance Of Messages For Analytics

- What is one of the most common causes of DB2 z/OS outages?
  - “What Happened to My DB2? The Top Missteps in High Availability” https://share.confex.com/share/121/webprogram/Session13729.html
    - Share Conference presentation, John Tobler & Nigel Slinger

The Importance Of Messages For Analytics

- What is one of the most common causes of DB2 z/OS outages?
  - “What Happened to My DB2? The Top Missteps in High Availability” https://share.confex.com/share/121/webprogram/Session13729.html
    - Share Conference presentation, John Tobler & Nigel Slinger
- ***Missing critical messages*** - a common source of DB2 outages
  - Are you monitoring all the most critical messages?
  - There are many critical messages that indicate potential issues that may impact availability

The Importance Of Alerts For Analytics

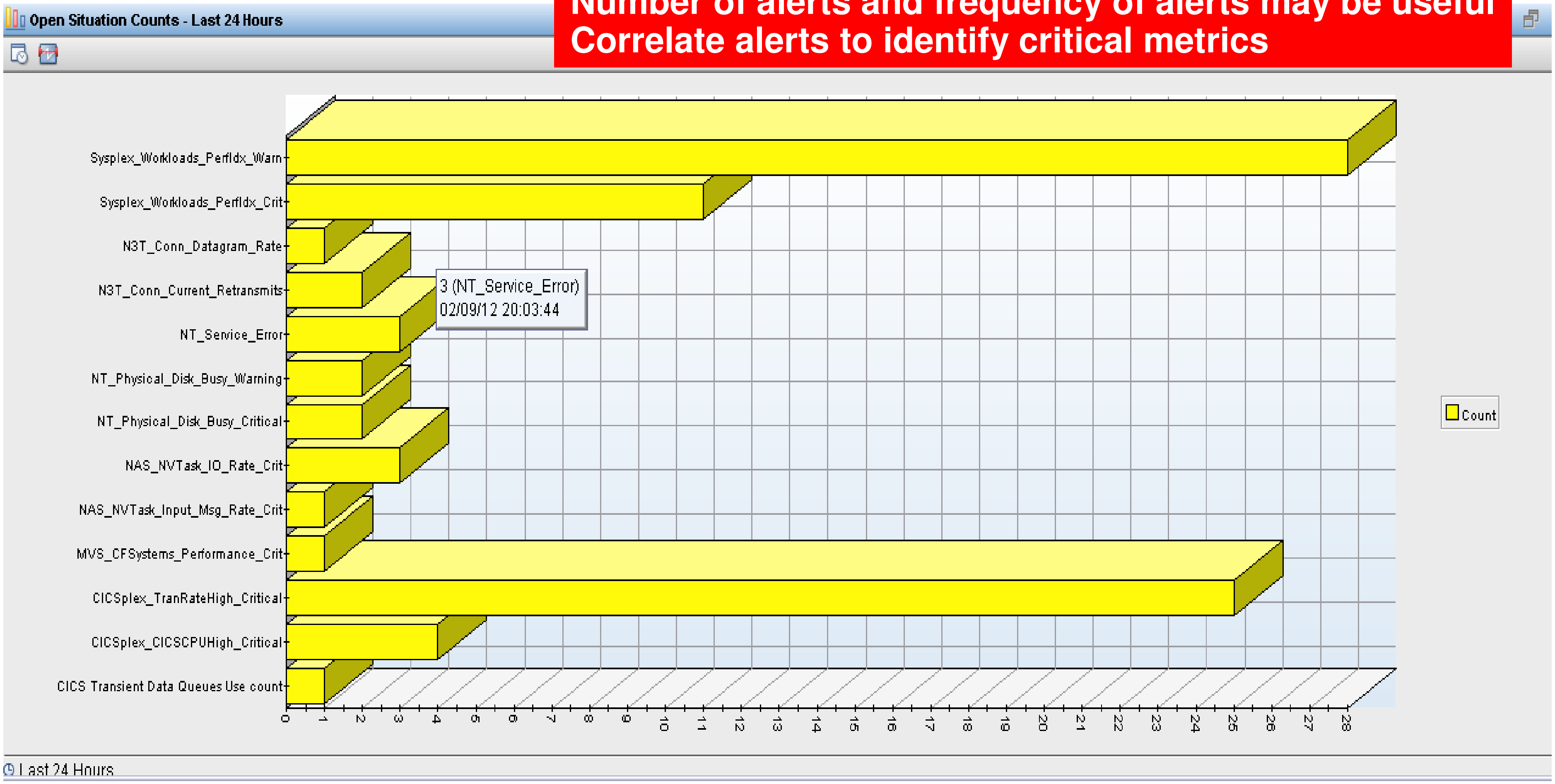


```
___________________  File  Edit  View  Tools  Navigate  Help    07/31/2014 11:03:47
                                                                Auto Update    : Off
Command ==> ______________________________________________      Plex ID   : ________
KOBSITEC*        ITM Situation Status & Message log             Region    : ________

                       Situation Event Status

 Columns _3 to _3 of _8     ← | → | ↑ | ↓     Rows _____1 to ____12 of ____23

 ΔSTATUS    ΔSITUATION NAME                    ΔMSN Event Source
 ∇          ∇                                  ∇

 _ Open      zOS_Service_Class_Warning          ESYSPLEX:MVSE:MVSSYS
 _ Open      Linux_AMS_Alert_Critical           zbx-scala-p01:LZ
 _ Open      Linux_Low_percent_space            zbx-scala-p01:LZ
 _ Open      Linux_System_Thrashing             ext6lnx:LZ
 _ Open      Linux_AMS_Alert_Critical           zbx-scala-p01:LZ
 _ Open      ZIBM_STATIC139_DD02D19A09A4415     CXEGDSST:MVSE:STORAGE
 _ Open      Connection_DataBackup              TCPIP:MVSE
 _ Open      Linux_AMS_Alert_Critical           zbx-scala-p01:LZ
 _ Open      POT_Missing_Workload_DEMQGET
 _ Open      Linux_Low_percent_space
 _ Open      Linux_AMS_Alert_Critical
 _ Open      KHD_Error_Critical
```

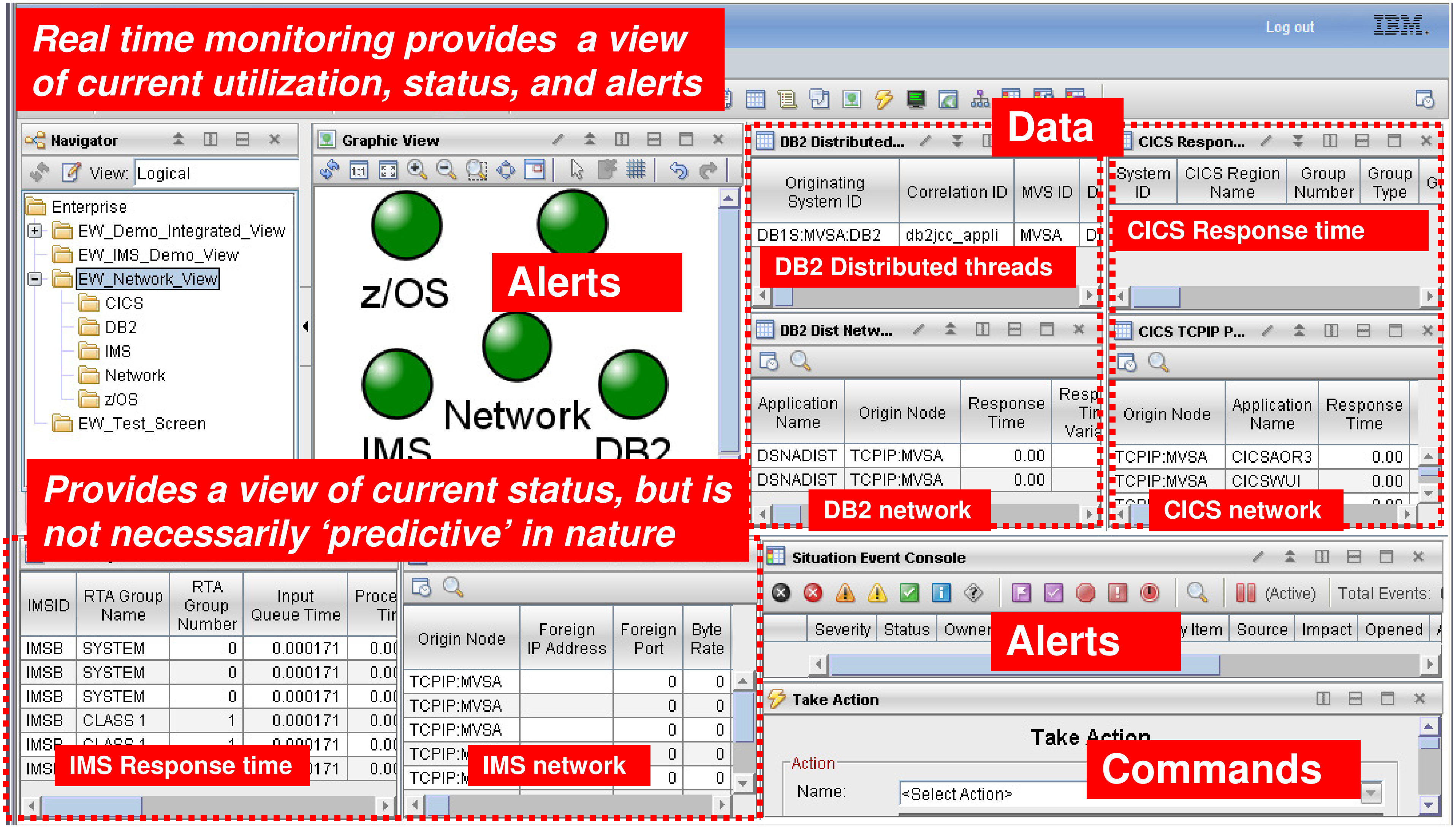
Alerts may come from a variety of platforms and sources

Don’t Overlook Alerts
Alerts Can Provide Valuable Metrics

Alerts may be a useful source of metrics for analysis
Number of alerts and frequency of alerts may be useful
Correlate alerts to identify critical metrics


Open Situation Counts - Last 24 Hours
Sysplex_Workloads_PerfIdx_Warn
Sysplex_Workloads_PerfIdx_Crit
N3T_Conn_Datagram_Rate
N3T_Conn_Current_Retransmits
NT_Service_Error
NT_Physical_Disk_Busy_Warning
NT_Physical_Disk_Busy_Critical
NAS_NVTask_IO_Rate_Crit
NAS_NVTask_Input_Msg_Rate_Crit
MVS_CFSystems_Performance_Crit
CICSplex_TranRateHigh_Critical
CICSplex_CICSCPUHigh_Critical
CICS Transient Data Queues Use count
3 (NT_Service_Error)
02/09/12 20:03:44
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28
Count
Last 24 Hours

Real Time Monitoring Provides A Starting Point For Analysis



Real time monitoring provides a view of current utilization, status, and alerts

Data

Alerts

Provides a view of current status, but is not necessarily 'predictive' in nature

View: Logical

- Enterprise
  - EW_Demo_Integrated_View
  - EW_IMS_Demo_View
  - EW_Network_View
    - CICS
    - DB2
    - IMS
    - Network
    - z/OS
  - EW_Test_Screen

DB2 Distributed threads

Originating System ID	Correlation ID	MVS ID	D
DB1S:MVSA:DB2	db2jcc_appli	MVSA	D

CICS Response time

System ID	CICS Region Name	Group Number	Group Type	G

DB2 network

Application Name	Origin Node	Response Time	Resp Tim Varia
DSNADIST	TCPIP:MVSA	0.00	
DSNADIST	TCPIP:MVSA	0.00	

CICS network

Origin Node	Application Name	Response Time
TCPIP:MVSA	CICSAOR3	0.00
TCPIP:MVSA	CICSWUI	0.00
TCP		0.00

IMS Response time

IMSID	RTA Group Name	RTA Group Number	Input Queue Time	Proce Ti
IMSB	SYSTEM	0	0.000171	0.00
IMSB	SYSTEM	0	0.000171	0.00
IMSB	SYSTEM	0	0.000171	0.00
IMSB	CLASS 1	1	0.000171	0.00
IMSB	CLASS 1	1	0.000171	0.00
IMS			171	0.00

IMS network

Origin Node	Foreign IP Address	Foreign Port	Byte Rate
TCPIP:MVSA		0	0
TCPIP:MVSA		0	0
TCPIP:MVSA		0	0
TCPIP:M		0	0
TCPIP:M		0	0

(Active) Total Events:

Severity	Status	Owner	y Item	Source	Impact	Opened

Alerts

Take Action

Action

Name: <Select Action>

Commands

Other Examples Of Common z/OS Critical Performance Metrics

WebSphere MQ

Queue depth
Message send/receive rate
DLQ depth
Channel status and performance

CICS

Transaction response time
Transaction rate
Region CPU rate
File I/O count
String waits
Abend messages

z/OS

System CPU rate
Paging rate
WLM Performance Index
DASD I/O MSR time and rate
Critical console messages

WebSphere

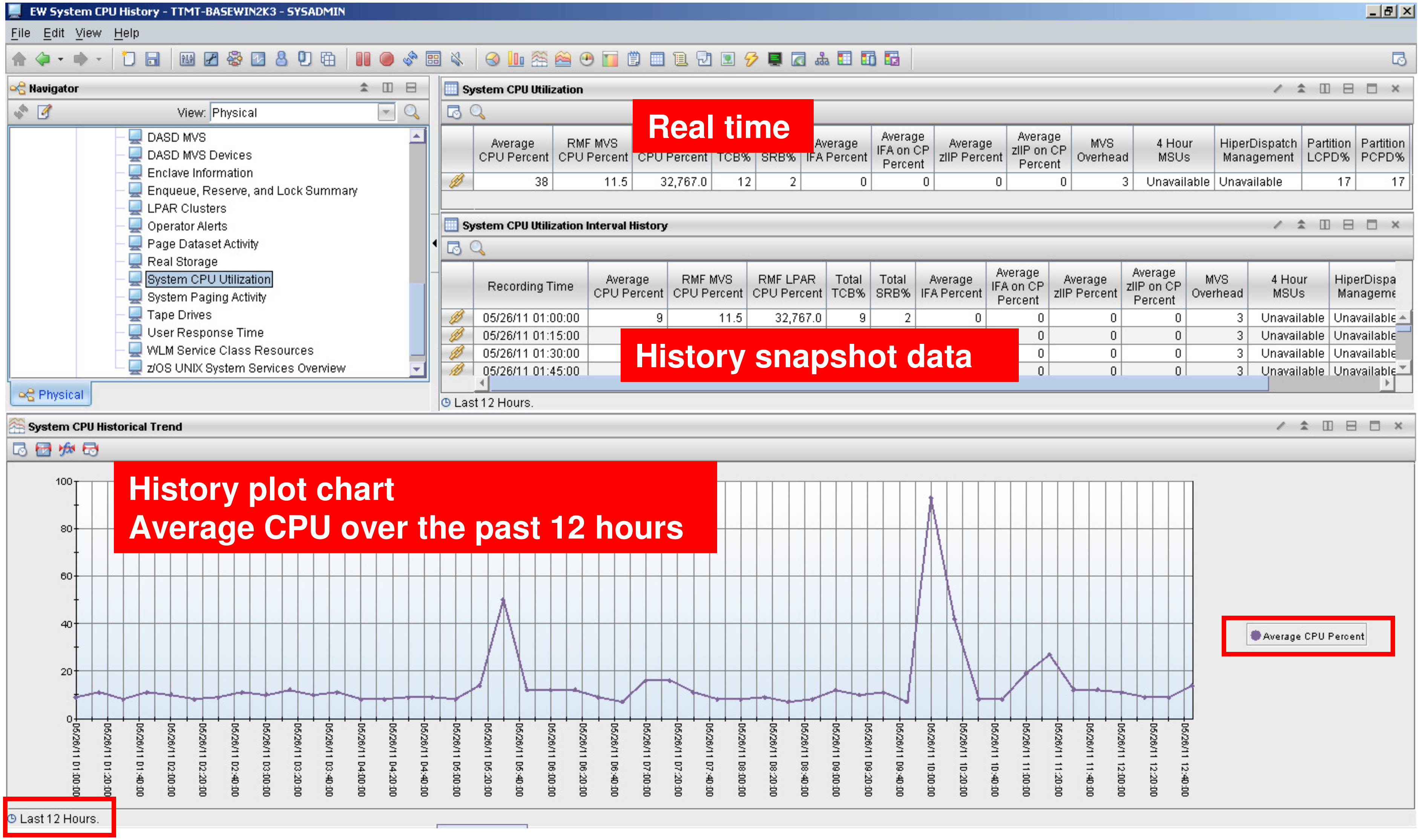
Method call count and elapsed time
Heap size
Garbage collection
Connection pool utilization

Network

Network Connection status and performance
Network interface utilization

Historical Data Analysis

Helps Identify Critical Metrics, Trends, Usage Patterns And Potential Issues



System CPU Utilization

Real time

	Average CPU Percent	RMF MVS CPU Percent	CPU Percent	TCB%	SRB%	Average IFA Percent	Average IFA on CP Percent	Average zIIP Percent	Average zIIP on CP Percent	MVS Overhead	4 Hour MSUs	HiperDispatch Management	Partition LCPD%	Partition PCPD%
	38	11.5	32,767.0	12	2	0	0	0	0	3	Unavailable	Unavailable	17	17

System CPU Utilization Interval History

History snapshot data

	Recording Time	Average CPU Percent	RMF MVS CPU Percent	RMF LPAR CPU Percent	Total TCB%	Total SRB%	Average IFA Percent	Average IFA on CP Percent	Average zIIP Percent	Average zIIP on CP Percent	MVS Overhead	4 Hour MSUs	HiperDispa Manageme
	05/26/11 01:00:00	9	11.5	32,767.0	9	2	0	0	0	0	3	Unavailable	Unavailable
	05/26/11 01:15:00							0	0	0	3	Unavailable	Unavailable
	05/26/11 01:30:00							0	0	0	3	Unavailable	Unavailable
	05/26/11 01:45:00							0	0	0	3	Unavailable	Unavailable

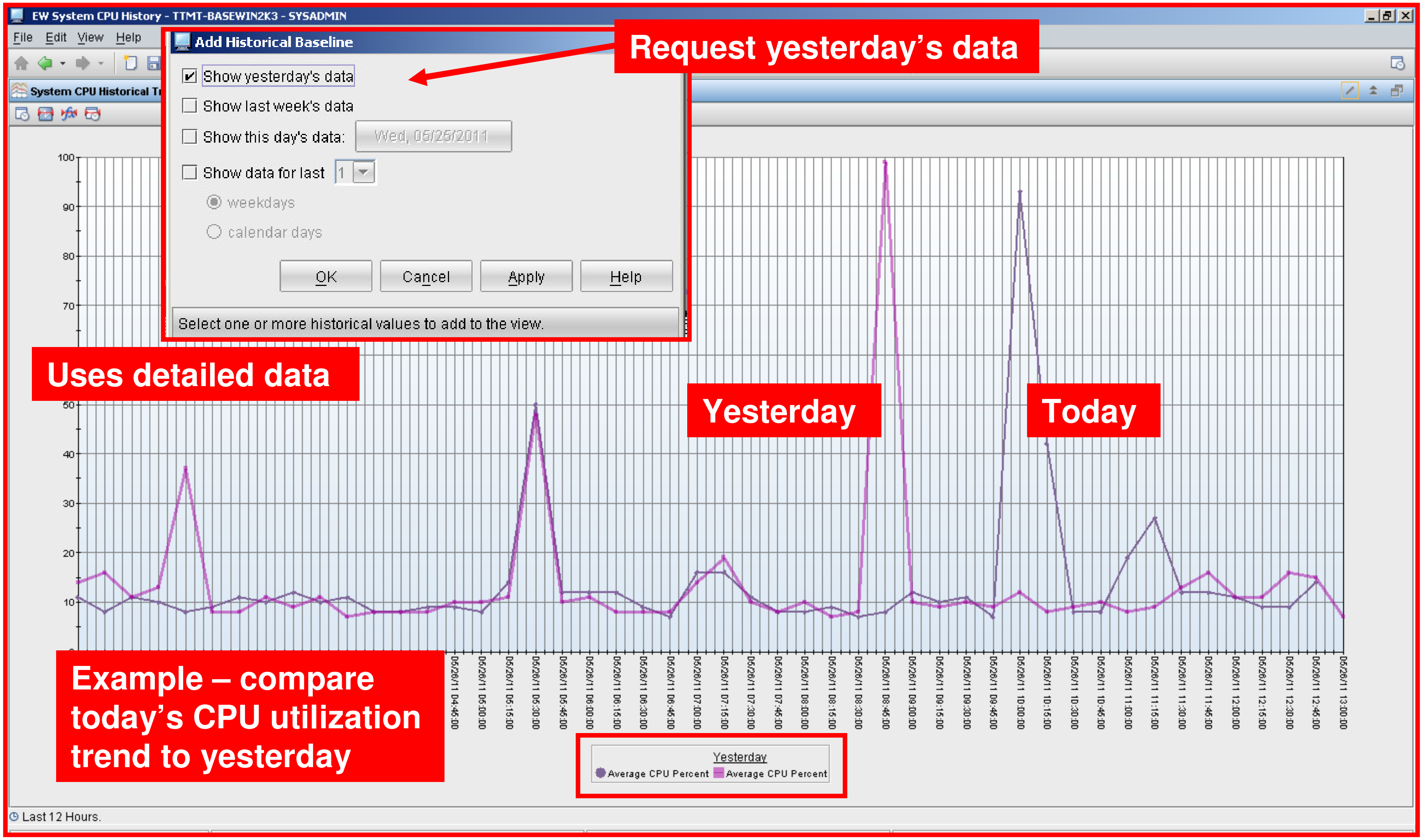
Last 12 Hours.

System CPU Historical Trend

History plot chart
Average CPU over the past 12 hours

Last 12 Hours.

Another Example - Historical Baseline Data To Compare Past Trends To Current Trends



EW System CPU History - TTMT-BASEWIN2K3 - SYSADMIN

Add Historical Baseline

☑ Show yesterday's data

□ Show last week's data

□ Show this day's data: Wed, 05/25/2011

□ Show data for last 1

○ weekdays

○ calendar days

Select one or more historical values to add to the view.

Last 12 Hours.

Request yesterday's data

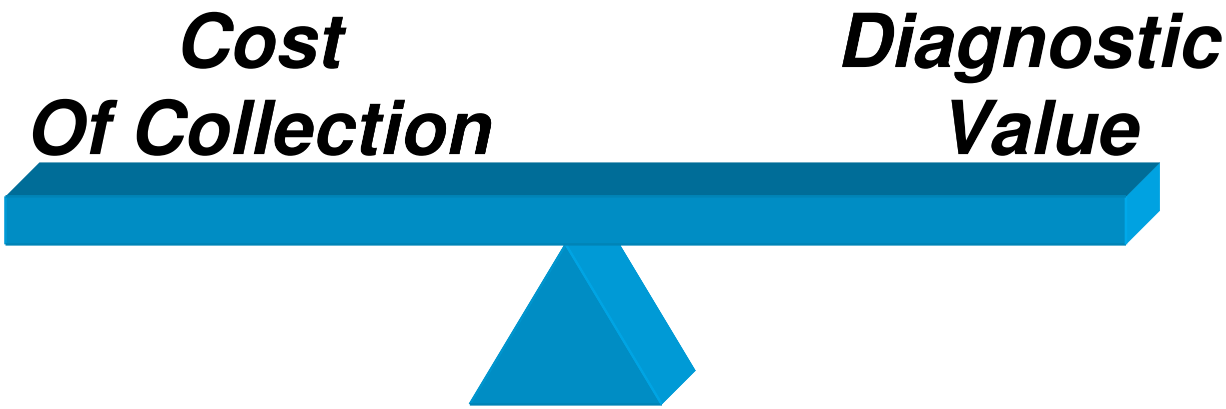
Uses detailed data

Yesterday

Today

Example – compare today's CPU utilization trend to yesterday

Predictive Analytics Often Begins With History
Historical Data Collection Considerations



Monitoring And History Collection Trade-off

- Historical data collection varies in cost and quantity
  - CPU, memory, and software process cost of collection
  - Cost of data storage and retention
  - Cost of retrieval and post processing
  - Ease of review and analysis
- Some historical data will be more relevant and useful than other data
  - Consider the context, nature, and meaningfulness of the data

Types Of Historical Monitoring Data

- Know the nature and characteristics of the history data being collected

- Detail data
  - Data that documents/measures detail of a specific event
  - Often high quantity data and the most detailed for analysis
  - May pose the greatest challenge in terms of cost, retention, post processing
  - Examples – DB2 Accounting records, CICS SMF 110 records, IMS log records

- Summary data
  - Data that summarizes underlying detail data
  - Either an aggregation or an averaging of underlying detail records
  - May be useful for longer term trending and analysis
  - Reduces quantity of data and reduces cost of retention, post processing
  - Less detail may mean less diagnostic value

Types Of Historical Monitoring Data - continued

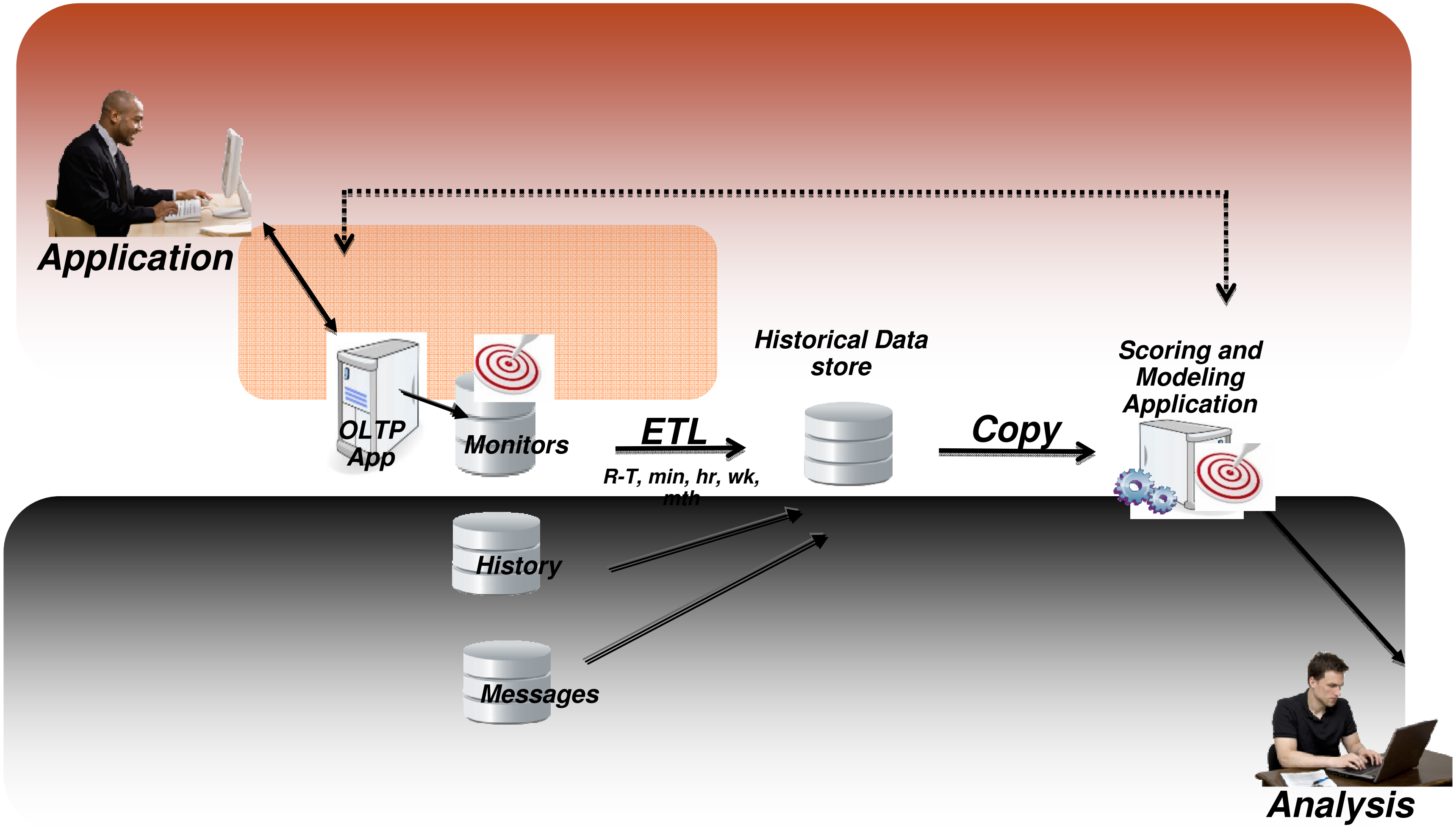
- Interval data
  - History data that includes an encapsulation of one or multiple events to a specified time interval
  - The data will include all activity within that given time interval
  - Useful for problem analysis and trending analysis
  - Examples – DB2 statistics records

- Snapshot data
  - Typically a point in time snapshot of activity
  - Snapshots are usually based on a specified time interval
  - Snapshots may be taken of types of history (detail, summary, or interval)
  - Snapshots will show activity at time of the snapshot, but may/may not reflect activity between snapshots
  - Useful for problem analysis and trending analysis
  - Useful as an aid in setting alert thresholds
  - Examples –snapshot history captured by performance monitoring,

The Components Of An IT Service Management Solution Built On Predictive Analytics

- ***Methodology consists of 3 core components***
  - Analytic data sources
    - Events, message logs, real time monitoring collection and analysis, historical performance metrics
  - Modeling and analysis component
    - Analytic engine to analyze, correlate and ‘score’ information
  - Reporting and visualization
    - View and display output of the analytic process
    - View actual versus ‘predicted’ outcomes

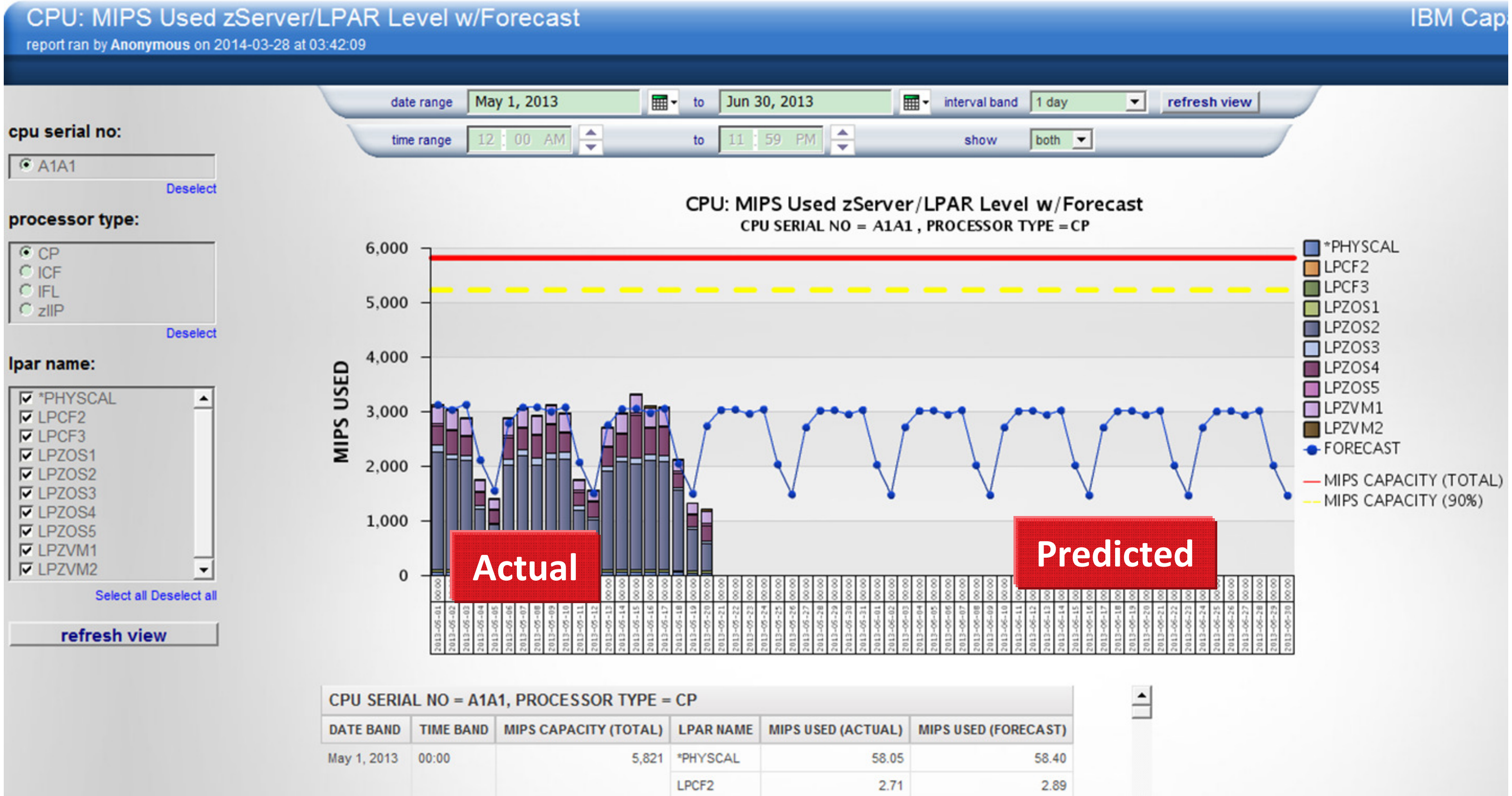
Understanding the Process Of Collection And Scoring



Reporting And Visualization

Answers the questions

- Will current capacity support usage demands 6 months from now? One year from now?
- Are there any cyclical or seasonal trends that indicate a need for additional capacity?



CPU: MIPS Used zServer/LPAR Level w/Forecast

report ran by **Anonymous** on 2014-03-28 at 03:42:09

IBM Cap

date range May 1, 2013 to Jun 30, 2013 interval band 1 day refresh view

time range 12 : 00 AM to 11 : 59 PM show both

cpu serial no:

● A1A1

Deselect

processor type:

● CP
○ ICF
○ IFL
○ zIIP

Deselect

lpar name:

☑ *PHYSCAL
☑ LPCF2
☑ LPCF3
☑ LPZOS1
☑ LPZOS2
☑ LPZOS3
☑ LPZOS4
☑ LPZOS5
☑ LPZVM1
☑ LPZVM2

Select all Deselect all

refresh view

CPU SERIAL NO = A1A1, PROCESSOR TYPE = CP

DATE BAND	TIME BAND	MIPS CAPACITY (TOTAL)	LPAR NAME	MIPS USED (ACTUAL)	MIPS USED (FORECAST)
May 1, 2013	00:00	5,821	*PHYSCAL	58.05	58.40
			LPCF2	2.71	2.89

An Example IBM Solution
Capacity Management Analytics - zCMA

- **Capacity Management Analytics**
  - Combines the following
    - TDSz for historical collection
    - SPSS for modeling and scoring
    - Cognos for visualization
- **System/Workload Characteristics, Performance and Trending**
  - What's driving demand?
  - Capacity constraints causing bottlenecks and what's being impacted
  - Anomalies occurred that impacted resource usage and/or performance
- **System/Workload Optimization, Prediction and Forecasting**
  - Available capacity to move workloads / applications to alleviate bottlenecks
  - Balance resource usage across servers/LPARs/VMs and defer capacity upgrade

Thank You!!

Check Out My Blog

http://tivoliwithaz.blogspot.com

Tivoli With A z - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites
Address http://tivoliwithaz.blogspot.com/ Go Snagit
Share Report Abuse Next Blog»
Create Blog Sign In
Internet

Tivoli With A z

This is a blog to discuss what is happening in the area of IBM z/Series, Tivoli, OMEGAMON monitoring, System Automation, and other relevant IBM Tivoli technology for z/OS performance and availability management.

Ed Woods
IBM Corporation

Friday, February 5, 2010

OMEGAMON DB2 Near Term History

OMEGAMON DB2 has a very useful Near Term History (NTH) function. NTH provides an easy way to be able to retrieve and review DB2 Accounting and Statistics records from the past few hours of DB2 processing. The data is stored in a set of VSAM files allocated to the OMEGAMON collection task. How far back the history goes depends upon the size of the files and the amount of data being written to these files. Now some of the data volume is driven by the DB2 workload activity. Accounting records are typically written when a DB2 thread terminates processing, and it is the Accounting data that is often looked at by the analyst when studying what DB2 applications have been doing. Statistics records are created on a time interval basis. Usually, you will have much more accounting data than statistics data. Also, OMEGAMON has the ability to pull in additional trace IFCIDs to get information on things such as dynamic SQL activity.

To understand the amount of data being gathered by NTH, there are displays that show the number of records written to the NTH files, by type. In the example I show, you see an example of common NTH settings/options, and then you see the record count in the NTH record information display. If you look carefully you see that 'Perf-Dyn SQL' has a lot of records written relative to the other record types. This is a good way to understand the impact of enabling certain collection options, such as dynamic SQL collection, and see how many trace records are being gathered, as a result.

Posted by Ed Woods at 3:13 PM 0 comments

ED WOODS

I'm an IT Specialist with IBM Corporation supporting Tivoli Performance solutions on z/OS. Please note that comments made on this blog are my own, and do not necessarily reflect the position of IBM Corporation.

View my complete profile

Links To My Articles

- DB2 Thread Situations
- OM XE For Mainframe Networks
- Situation usage and best practices
- Situation best practices - part 2
- Article on policy automation
- Article on monitoring DB2 dynamic SQL
- IMS historical performance analysis

Useful Links

- Link to IBM Tivoli product information
- Link To Tivoli User Group
- Link to OPAL
- Tivoli System z Blog