



IBM 2019
Cloud
& **AI** 高峰會

資料品質對IT AIOps的影響

—

Fission Lin
Advisory IT Architect

*If Your Data Is Bad, Your Machine Learning
Tools Are Useless*
~ Thomas C. Redman



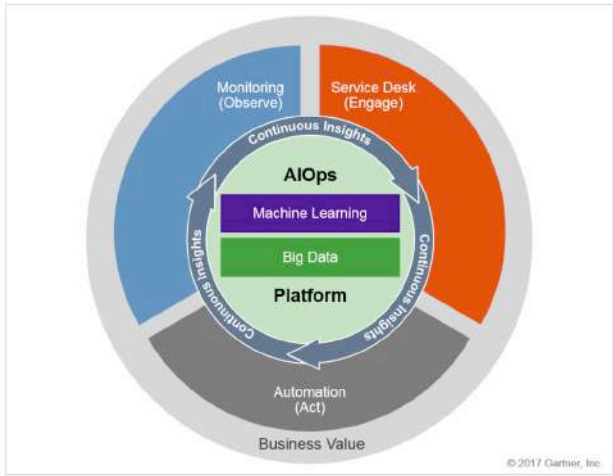
01

IT AIOps簡介

IBM 2019
Cloud
& **AI** 高峰會

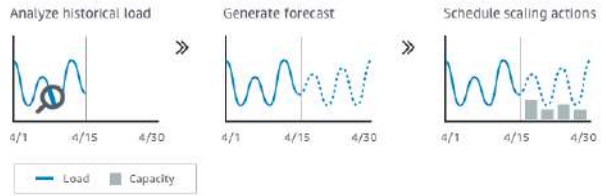
IT AIOps簡述

IT AIOps的定義



Source: Gartner (August 2017)

IT AIOps的範例



IT AIOps的效益

- 穩定
 - 提升系統MTBF (Mean Time Between Failure)
 - 降低系統MTTR (Mean Time To Repair)
 - 降低系統事故發生的機會
 - 提升系統Availability
- 高效
 - 提升自動擴縮容效率
 - 提升維運人員效率(維運設備數/人)
 - 提升使用者體驗及滿意度
- 節省
 - 節省資源投入
 - 節省人力投入
 - 節省時間的浪費



02

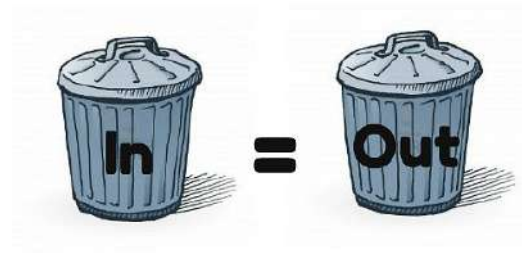
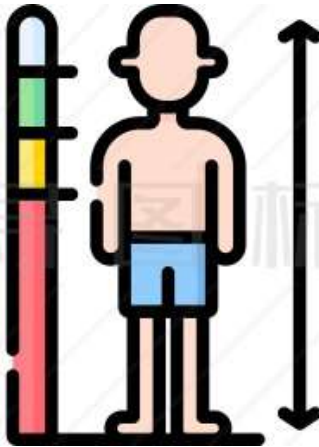
IT AIOps的資料來源與 分析質量建議

IT AIOps的主要資料來源與分析質量建議

座右銘：

~資料的品質、特徵的選取決定了機器學習的上限，模型(Model)和算法只是逼近這個上限而已。

~Garbage In, Garbage Out



IT AIOps的主要資料來源與分析質量建議(續)

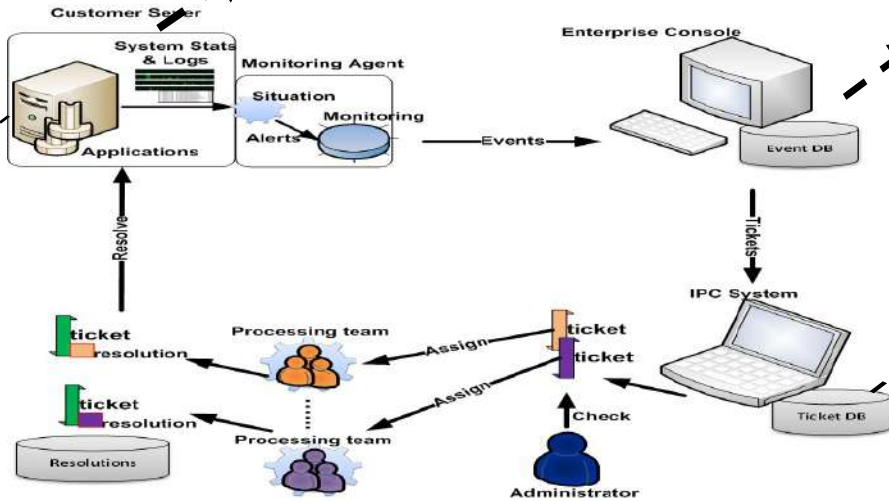
```

2010-05-02 00:21:39 Command: put "E:/Tomcat/apps/index.html" "/disk/...
2010-05-02 00:21:40 Status: File transfer successful, transferred 823 bytes...
2010-05-02 00:21:41 Command: cd "/disk/storage006/users/lt...
2010-05-02 00:21:42 Command: cd "/disk/storage006/users/lt...
2010-05-02 00:21:42 Command: put "E:/Tomcat/apps/record1.html" "/disk/...
2010-05-02 00:21:42 Status: Listing directory /disk/storage006/users/lt...
2010-05-02 00:21:42 Status: File transfer successful, transferred 1,232 bytes...
2010-05-02 00:21:42 Command: put "E:/Tomcat/apps/record2.html" "/disk/...
2010-05-02 00:21:42 Response: New directory is: "/disk/storage006/users/lt...
2010-05-02 00:21:42 Command: mkdir "libraries"
2010-05-02 00:21:42 Error: Directory /disk/storage006/users/lt...
2010-05-02 00:21:44 Status: Retrieving directory listing...
2010-05-02 00:21:44 Command: ls
2010-05-02 00:21:45 Command: cd "/disk/storage006/users/lt...
    
```

日誌

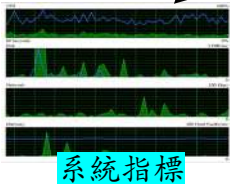
Event Type	Time Stamp	Description
DB_Down	[16/Sep/2015 14:01:38 +1000]	xxxxxxxxxxxx
Service_Unavailable	[16/Sep/2015 14:01:42 +1000]	xxxxxxxxxxxx
Server_Restart	[16/Sep/2015 14:31:02 +1000]	xxxxxxxxxxxx
SVC_TEC_HEATBEAT	[16/Sep/2015 14:32:00 +1000]	xxxxxxxxxxxx
...

事件



Description of ticket
 Time: 20xx-xx-01 hh:mm:ss
 Message:
 Failed to write a record to destination file xxx
 Warning:NAS Mount is failing on nas03a.host1.com:/zzz

Ticket



系統指標

IT AIOps的主要資料來源與分析質量建議(續)

類型： 系統指標

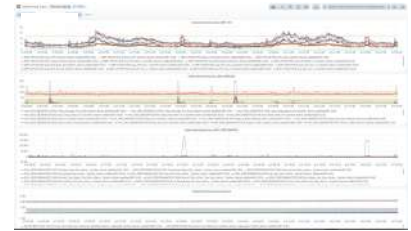
資料性質： 結構化數據，固定時間區間採樣，會伴隨採樣的時間記錄，專業術語稱之為時間序列

範例： 使用率、流量、數量(ex. IOPS、線上使用者人數、傳送位元數、吞吐量)、回應時間

應用場景： 基於系統行為的異常檢測、容量預測

質 KPI： 維度、資料收集區間、缺值率

量 KPI： 連續7~14天不間斷收集的效能指標數據



IT AIOps的主要資料來源與分析質量建議(續)

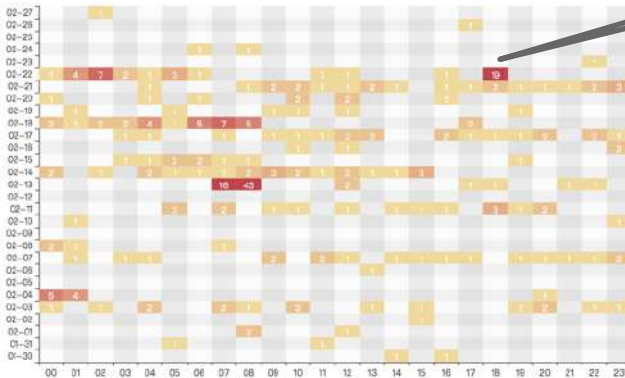
#	資料描述	指標	序列	最大缺值率	連續5分鐘缺值	連續5分鐘缺值
1	2019年2月資料	KLZ_CPU	26個序列[System_Name,CPU_ID]	1.10%	35次	3次
2		KLZ_VM_STATS	26個序列[System_Name]	0.56%	47次	2次
3		KLZ_NETWORK	151個序列[System_Name,Network_Interface_Name]	0.47%	245次	12次
4		KLZ_DISK_ID	200個序列[System_Name,Dev_Name]	0.42%	337次	34次
5	2019年3月資料	SMP_CPU	4個序列[System_Name]	11.58%	3次	無
6		UNIX_MEMORY	4個序列[System_Name]	17.58%	7次	無
7		NETWORK	18個序列[System_Name,Network_Interface_Name]	36.67%	310次	1次
8		DISK_PERFORMANCE	350個序列[System_Name,Disk_Name]	76.03%	166684次	627次
9		KLZ_CPU	26個序列[System_Name,CPU_ID]	0.73%	35次	無
10		KLZ_VM_STATS	26個序列[System_Name]	0.64%	48次	無
11		KLZ_NETWORK	151個序列[System_Name,Network_Interface_Name]	0.53%	167次	無
12		KLZ_DISK_ID	200個序列[System_Name,Dev_Name]	0.62%	225次	無
13		SMP_CPU	4個序列[System_Name]	11.15%	6次	無
14		UNIX_MEMORY	4個序列[System_Name]	12.34%	6次	無
15		NETWORK	18個序列[System_Name,Network_Interface_Name]	32.34%	512次	無
16		DISK_PERFORMANCE	350個序列[System_Name,Disk_Name]	74.30%	100798次	438次

指標角度

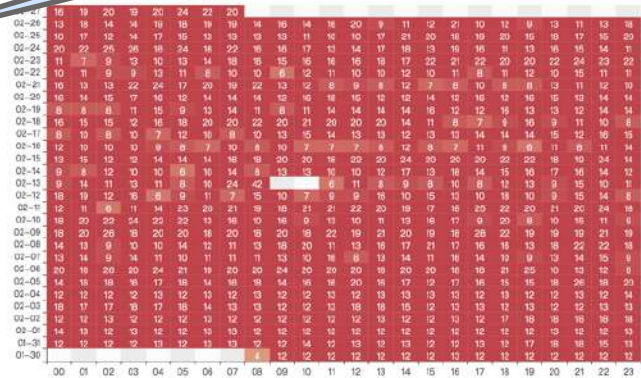
每小時資料缺失數
(如：2月22日18點缺失19
個採樣資料點)

伺服器角度

2019年2月資料缺失統計

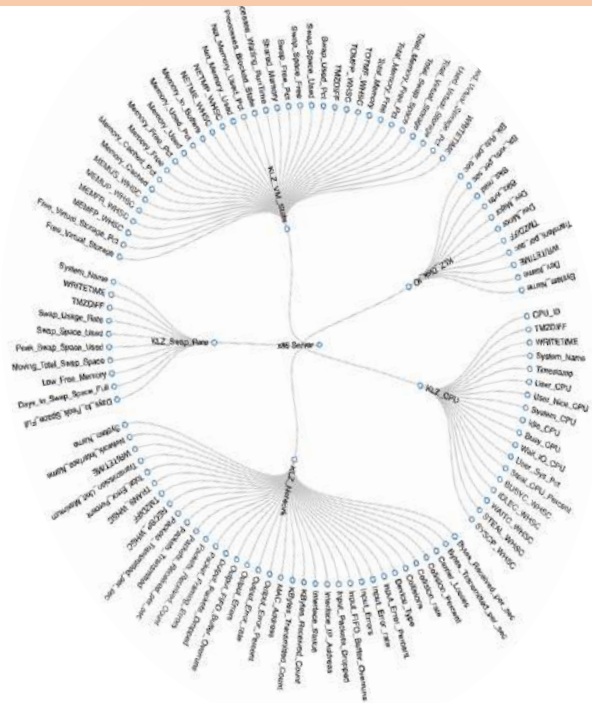


2019年2月資料缺失統計

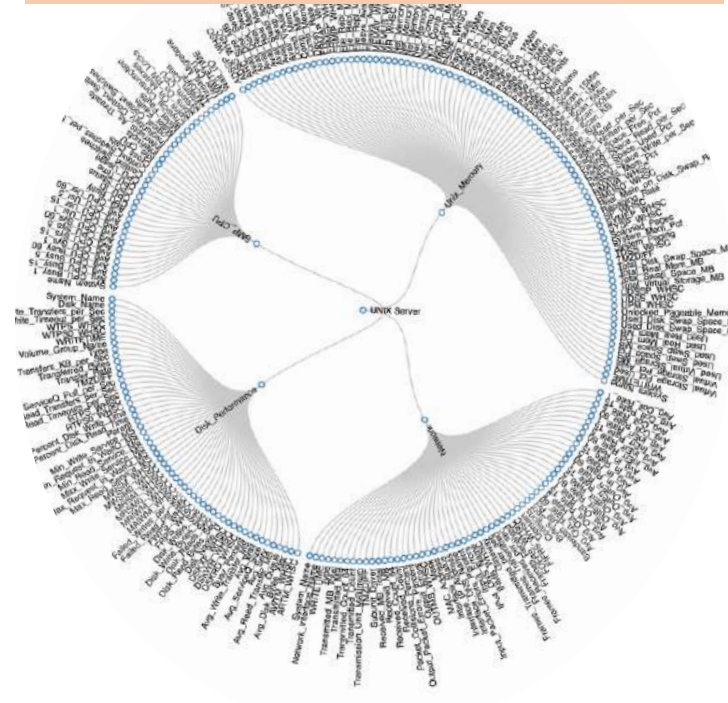


IT AIOps的主要資料來源與分析質量建議(續)

x86伺服器監控指標：481維(最高)



Unix伺服器監控指標：9409維(最高)



IT AIOps的主要資料來源與分析質量建議(續)

時間區間:

- 2019年2月20日 ~ 2019年2月22日

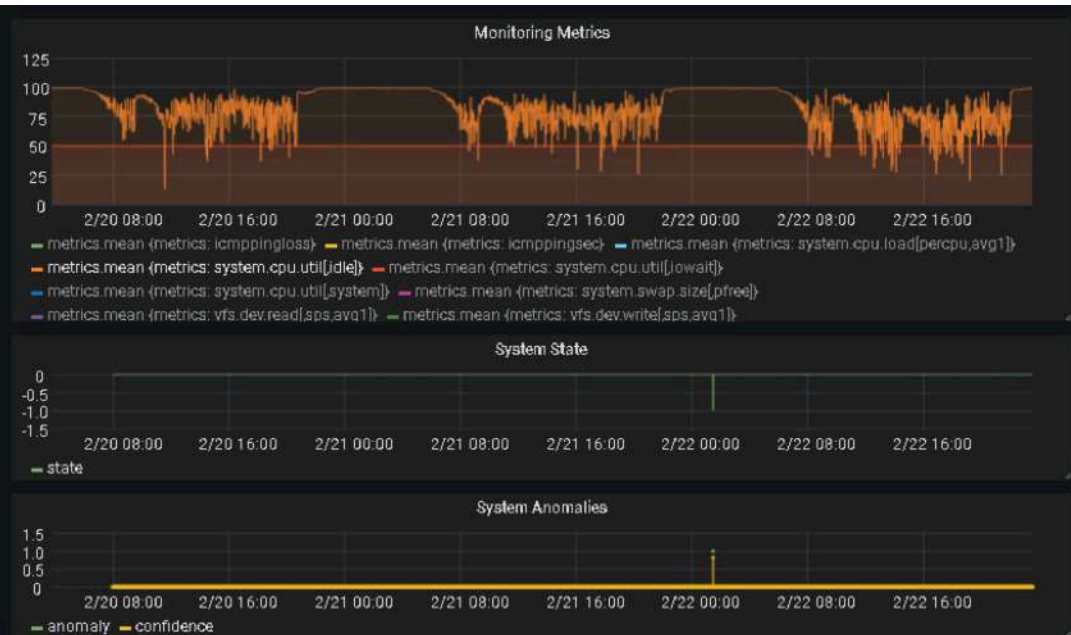
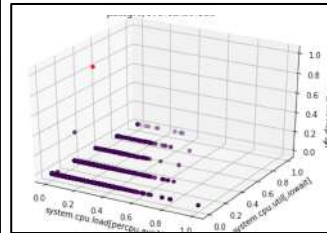
監控告警數量:

- 32

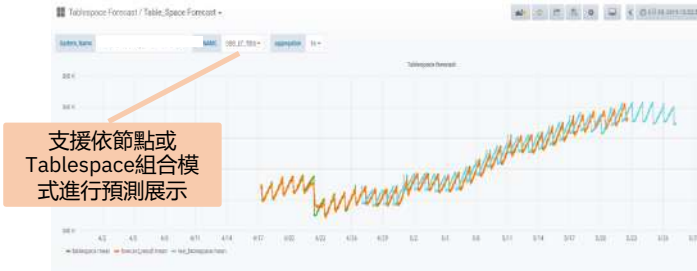
告警設定:

- CPU閒置值低於50%

系統行為模型



IT AIOps的主要資料來源與分析質量建議(續)



IT AIOps的主要資料來源與分析質量建議(續)

類型： 日誌

資料性質： 半結構化數據

範例： 應用程式日誌、中介軟體日誌

應用場景： 基於日誌的異常檢測與根因分析

質 KPI： 平均日誌行數/交易、平均可自動跨接系統日誌數/交易、異質日誌數量

量 KPI： -樣本數達到一定規模，典型值：樣本數>模版數量x500

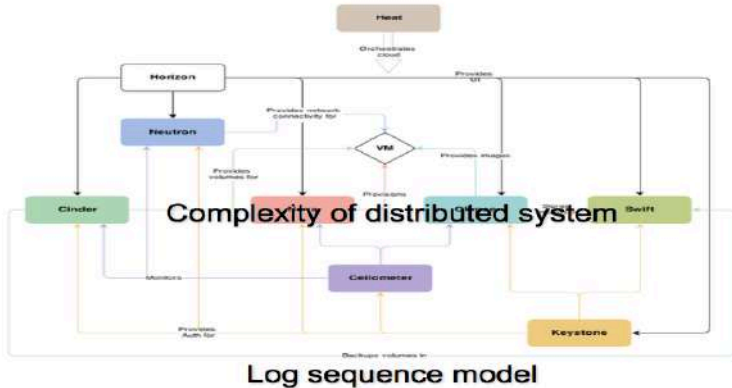
-日誌交易序列樣本達到一定規模，典型值：交易總數>200，
且成功交易佔>90%

-日誌來自於持續運行的系統，典型值：連續一~四星期的日誌

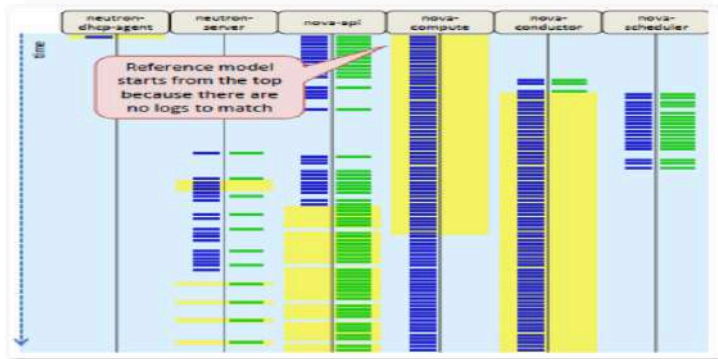
其它： 業務容器若採用多執行緒執行，日誌可以透過標示符號解決交錯(interleave)的問題，如ThreadID、ProcessID

```
2010-05-02 00:21:39 Command: put "E:/Tomcat/apps/index.html" "/disk/...
2010-05-02 00:21:40 Status: File transfer successful, transferred 823 bytes...
2010-05-02 00:21:41 Command: cd "/disk/storage006/users/lt...
2010-05-02 00:21:42 Command: cd "/disk/storage006/users/lt...
2010-05-02 00:21:42 Command: put "E:/Tomcat/apps/record1.html" "/disk/...
2010-05-02 00:21:42 Status: Listing directory /disk/storage006/users/lt...
2010-05-02 00:21:42 Status: File transfer successful, transferred 1,232 bytes...
2010-05-02 00:21:42 Command: put "E:/Tomcat/apps/record2.html" "/disk/...
2010-05-02 00:21:42 Response: New directory is: "/disk/storage006/users/lt...
2010-05-02 00:21:42 Command: mkdir "libraries"
2010-05-02 00:21:42 Error: Directory /disk/storage006/users/lt...
2010-05-02 00:21:44 Status: Retrieving directory listing...
2010-05-02 00:21:44 Command: ls
2010-05-02 00:21:45 Command: cd "/disk/storage006/users/lt...
```

IT AIOps的主要資料來源與分析質量建議(續)



- 正常行為的日誌參考模型
- 即時由系統收集的日誌
- 前兩者產生落差的日誌區塊



IT AIOps的主要資料來源與分析質量建議(續)

類型: 事件

資料性質: 半結構化數據

應用場景: 基於因果關係的事件壓縮

質 KPI: 事件欄位數量、事件欄位缺值率、事件分類數量

量 KPI: -需要近三個月的或不少於100000條的連續事件
-同一類型的告警事件，最少需要的歷史數據訓練集的大小超過1000條

Event Type	Time Stamp	Description
DB_Down	[16/Sep/2015 14:01:39 +1000]	xxxxxxxxxxxxxxxx
Service_Unavailable	[16/Sep/2015 14:01:42 +1000]	xxxxxxxxxxxxxxxx
Server_Restart	[16/Sep/2015 14:31:02 +1000]	xxxxxxxxxxxxxxxx
SVC_TEC_HEATBEAT	[16/Sep/2015 14:32:00 +1000]	xxxxxxxxxxxxxxxx
⋮	⋮	⋮

IT AIOps的主要資料來源與分析質量建議(續)

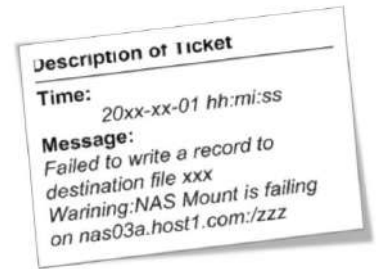
類型: Ticket

資料性質: 半結構化數據

應用場景: Ticket自動分類與處理

質 KPI: Ticket欄位數量、Ticket欄位缺值率、Ticket分類數量

量 KPI: 需要三個月或者不少於10000張的Ticket數據，每種故障原因的Ticket數量需要>100張







IBM 2019
Cloud
& AI 高峰會