



Insight2014 大數據高峰會

千思萬想 · 一瞬在握

產業轉型時機來臨，商機加值就在此刻！





Insight2014 大數據高峰會

即時洞悉：
以現有分析工具為基礎，打造即時海量分析力

財團法人資訊工業策進會
前瞻科技研究所

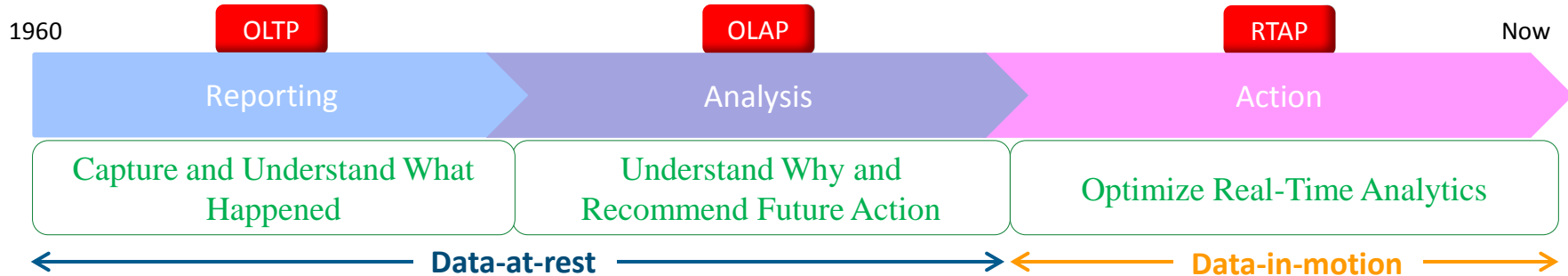
研發經理 余承叡
fishyu@iii.org.tw

簡報大綱

- 資訊分析的革命
- 巨量資料分析演化
- 巨量資料即時分析應用
- 下一波資訊革命-物聯網 Internet of Things



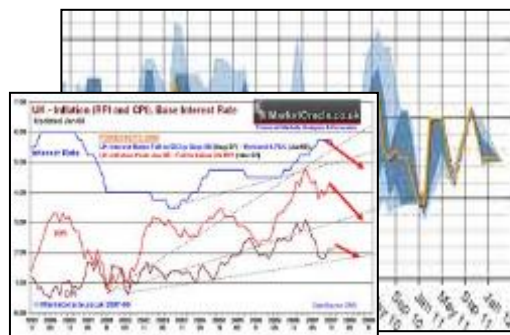
資訊分析的革命



- | | | |
|---|---|--|
| <ul style="list-style-type: none"> • 1960 年, Hierarchical Databases 問世 • 線上 OLTP 交易成為商業系統的一部分 • 全球每天完成進百萬兆的驚人交易量 | <ul style="list-style-type: none"> • Relational Database 演進到資料倉儲的概念產生 • 延伸系統從 OLTP 到 OLAP 的觀念 • OLAP 更注重於了解問題的原因以提供更好的建議 | <ul style="list-style-type: none"> • OLTP 與 OLAP 屬於資料儲存技術, 無法滿足即時應用需求 • Real-time Analytics Processing (RTAP) 改變系統設計觀念 • 從資料儲存後分析, 走向資料在 in-motion 狀態產生結果 |
|---|---|--|



Dashboards/Daily Reports

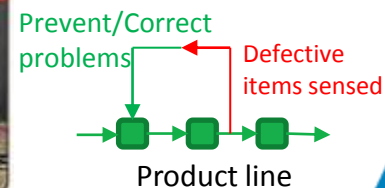


Trends/Patterns/Forecasts



Alerting/Feedback/Detecting failures

Stream Computing



Ref: IBM, InfoSphere Streams Red Book, 2013

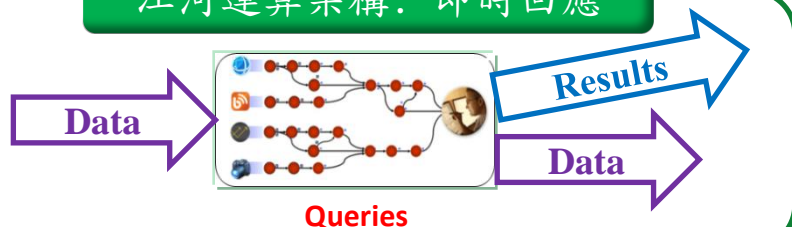
從傳統分析走向大數據即時分析

傳統運算架構：批量處理



In traditional computing, analytics move through data

江河運算架構：即時回應



In stream computing, data move through analytics

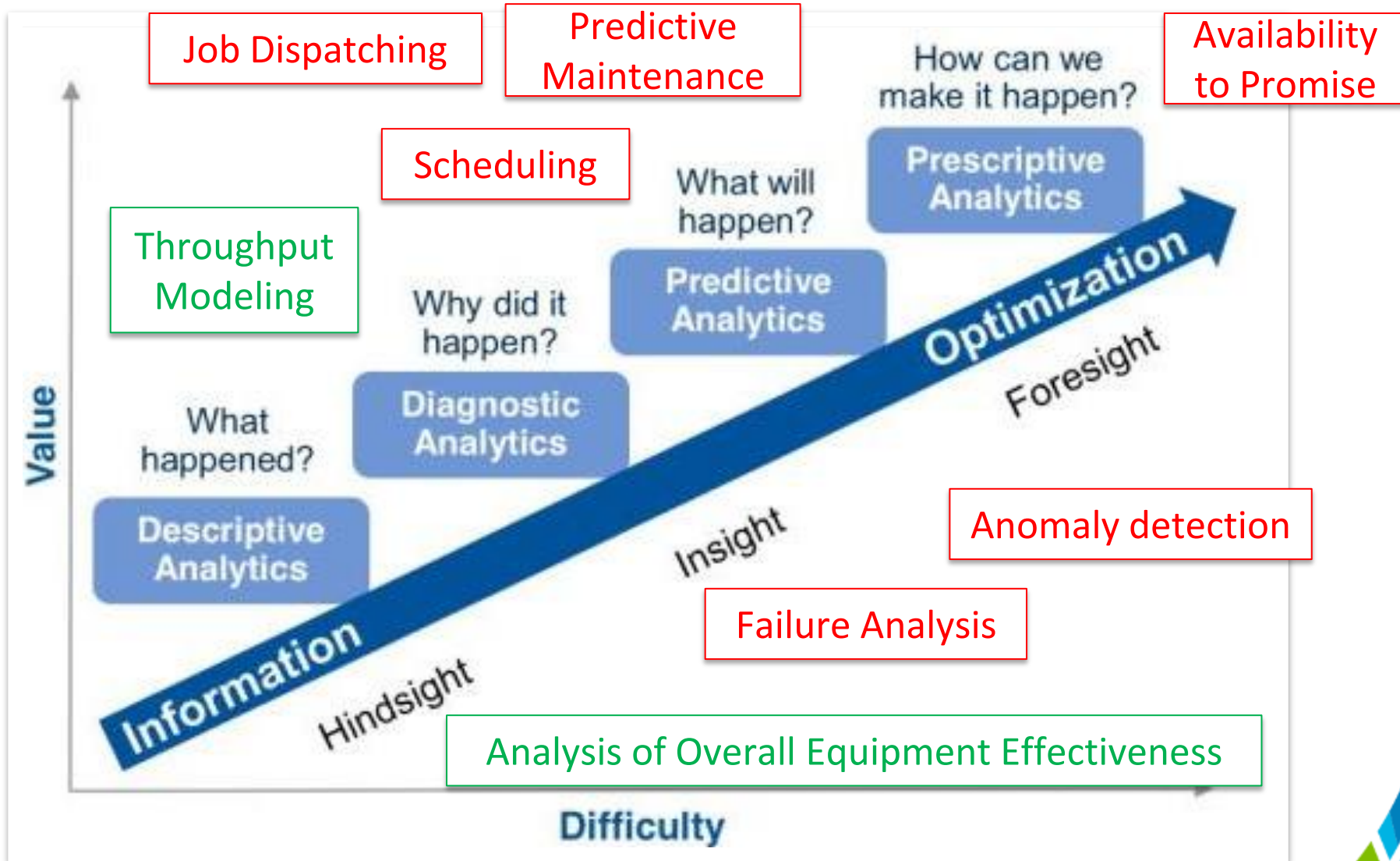
<Historical fact finding with data-at-rest>

- Batch Paradigm, pull model
- Query-driven: submits queries to static data
- Relies on Databases, Data Warehouses

<Real time analysis of data-in-motion streaming data>

- Streaming data
 - ❑ A stream of structured or unstructured data-in-motion
- Stream Computing
 - ❑ Analytic operations on streaming data in real-time
 - ❑ Operations “fire” as data passes, unneeded data can be immediately discarded

巨量資料分析演進: From information to optimization



巨量資料即時分析應用

案例: 製造業- 以OEE分析及生產效能建模為例

“Manufacturing stores more data than any other sector — close to 2 exabytes of new data stored in 2010.” - McKinsey & Company

- **總體設備效率 (Overall Equipment Effectiveness, OEE) 分析**
 - OEE 能被當作製造業機台設備初始表現的 “標竿” 衡量。
 - OEE 能夠針對一條生產線來計算,並可以用來比較整個工廠之不同製造生產線的績效,藉此找出差勁表現的生產線。
 - 若機台是個別在運作, OEE 的衡量能確認哪一機台製造績效表現最差,以指出該在哪裡集中做全面生產管理(Total Productive Management, TPM) 的資源改善。
- **生產效能模型建立**
 - 不同機型生產不同規格之產品,會有不同的生產效率;如何藉由大量之生產資料,鑑別重要的效率影響因子、建立生產效率模型,以預測新品生產時間、建立標準生產效能。

製造業於巨量資料分析之技術缺口

• 資料收集與儲存方面

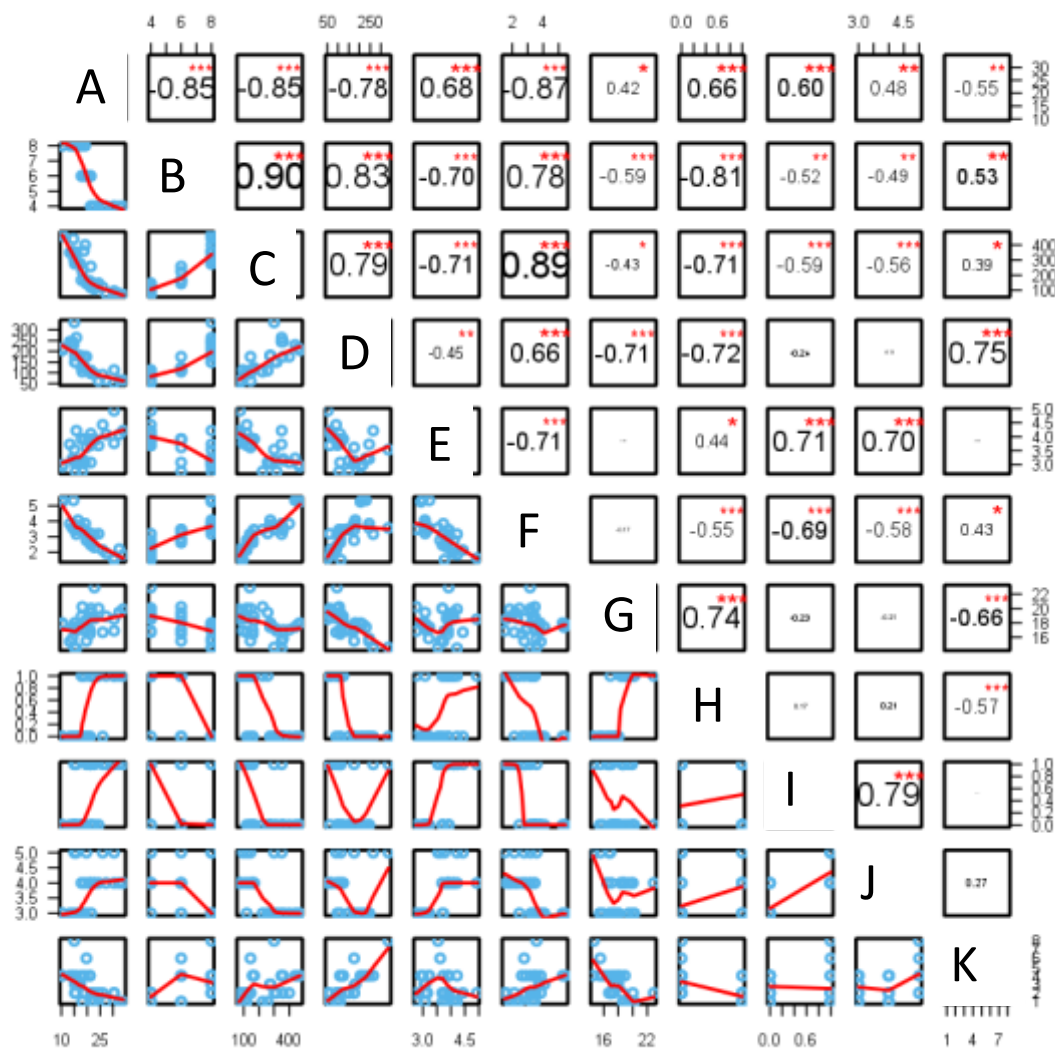
- 製程依產品需求不同往往需經歷數百個不等的步驟，每個步驟則需耗費數秒至數小時高變異時間差，同時需高頻率地監控各種製程指標。
- 針對製程中所產生的即時且大量資料，如何有效確保其完整性並減少遺漏，同時設計具有延展性且低價格的資料收集方式及儲存架構，為目前製造業所面臨之技術缺口。

• 資料處理與分析方面

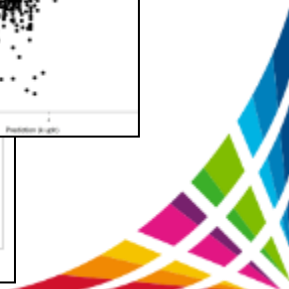
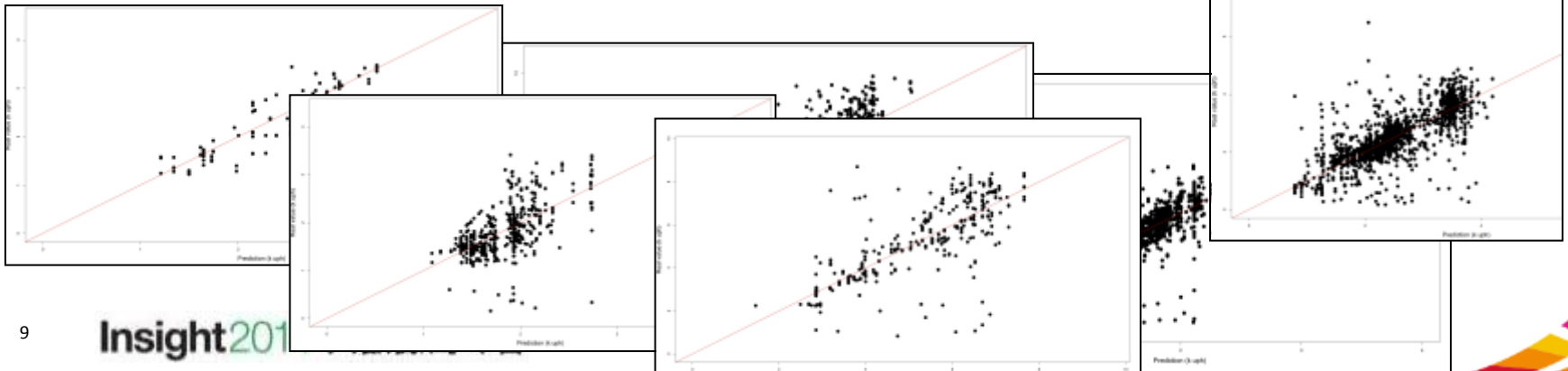
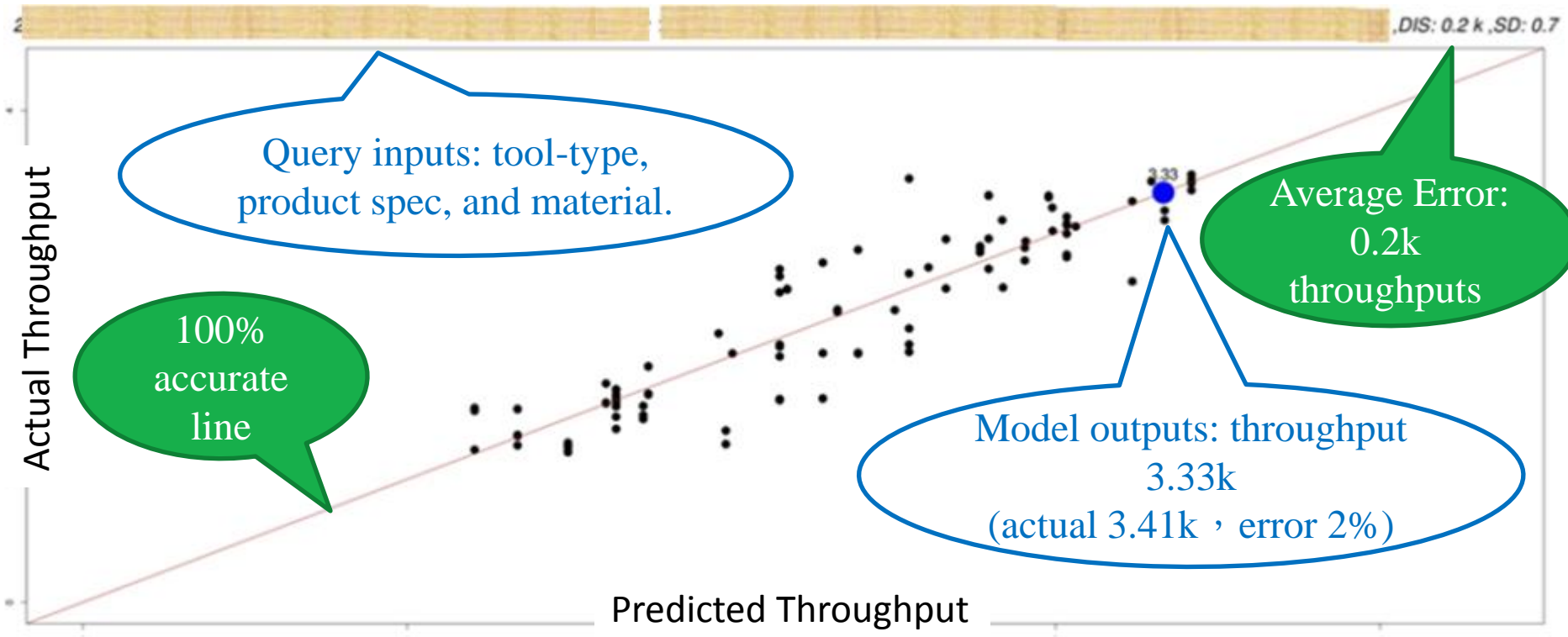
- 由於製程資料同時具備多元、異質、分屬不同IT系統、由不同製程站點產生等特性，使得製程資料具有不同步及不正確兩項課題
- 資料的不同步性提高了即時分析如資料整合、關聯查詢、交叉比對等步驟之困難度
- 資料的不正確性則降低了分析結果的可信度
- 因此如何在分析的過程中，即時識別不正確的資料、追蹤不正確的原因並試圖修正，同時能藉由有限資料進行分析推論，並採用漸進式方式以新資料更新分析結果，皆為目前製造業所面臨之技術缺口。



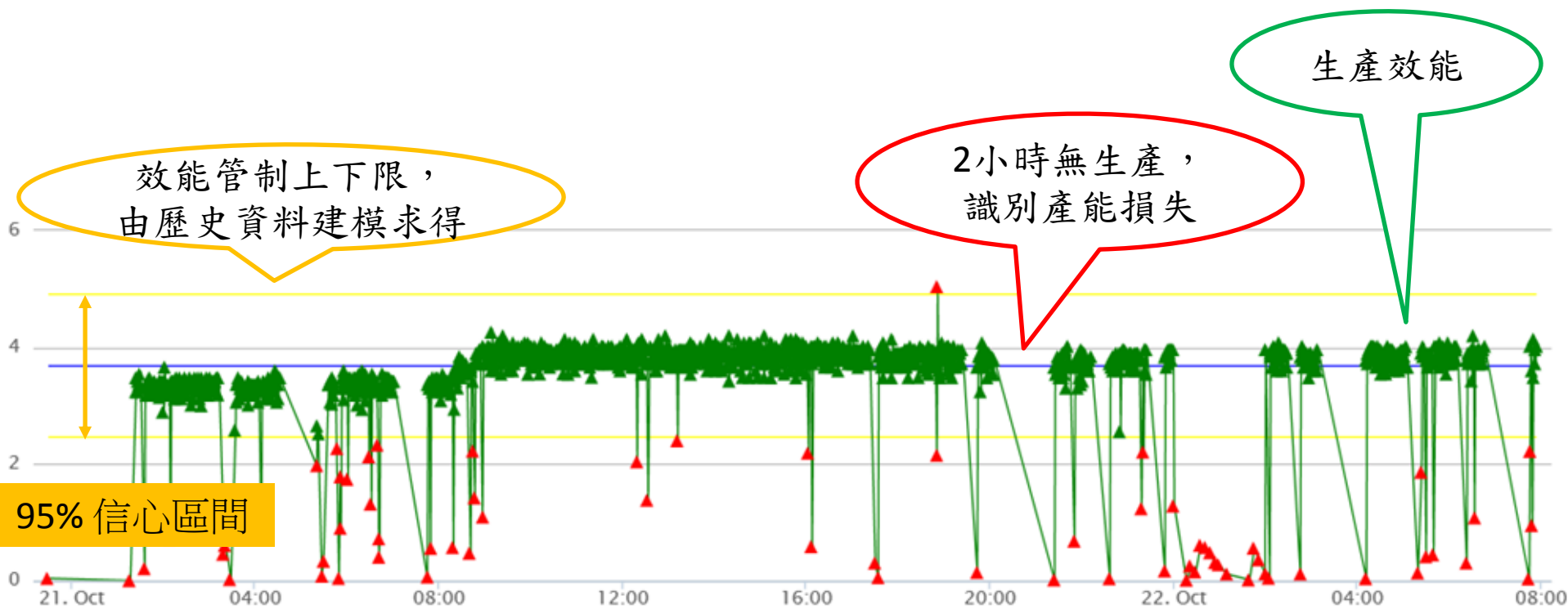
以現有的統計工具觀察資料間的相關性



建立生產效能的預測模型



整合運用IBM Streams及R語言之預測建模功能： 實現生產效能即時監控



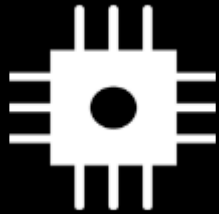
即時生產效能圖，可細緻控管生產效能

運用Stream Computing運算的優勢

- IBM InfoSphere Streams 的好處
 - 方便佈署於叢集伺服器，加速巨量資料運算
 - 配備豐富的資料處理與分析元件庫，縮短開發時間
 - 視覺化的開發及管理介面，縮短學習曲線
 - 支援平行化(parallelization)及管線化(pipeline)加速，可快速處理大量且即時的資料
- IBM InfoSphere Streams 於製造業之實證經驗
 - 快速處理半結構化的全廠機台日誌資訊，延遲<1分鐘
 - 即時統計廠區產能，資料處理量達 2TB/week、1MB/sec
 - 快速更新預測模型，產效預測準確度 <10%
- IBM InfoSphere Streams 還可快速導入於電信、金融、零售、健康照護、智慧建築等領域應用



下一波資訊革命- 物聯網 Internet of Things



Our world is becoming

INSTRUMENTED



Our world is becoming

INTERCONNECTED



All things are becoming

INTELLIGENT

影片分享: BIG DATA大未來





Thank You