



釋放雲端潛能 駕馭海量資料

大資料雲端計算分析應用

- 陳志昇 / 騰雲計算
- Vincent_Chen@tcloudcomputing.com

大資料雲端計算分析應用 -- 手機行動上網分析



雲端計算分析平臺具備融合多種來源和格式的非結構化資料的能力，並通過簡訊、上網分詞與定位技術、將非結構化的資料形成標籤化的使用者檔案，並最終實現人與內容、人與商品、人與人的智能配對。



核心價值

整合非結構化的資料



形成用戶檔案——User Profile



自動運營——Smartpush

AM9:00



手機購物用戶：習慣在早上9點登陸淘寶、凡客等用戶端查詢商品資訊

AM10:00



資訊需求用戶：習慣在早上10點流覽當日熱點新聞資訊

PM19:30

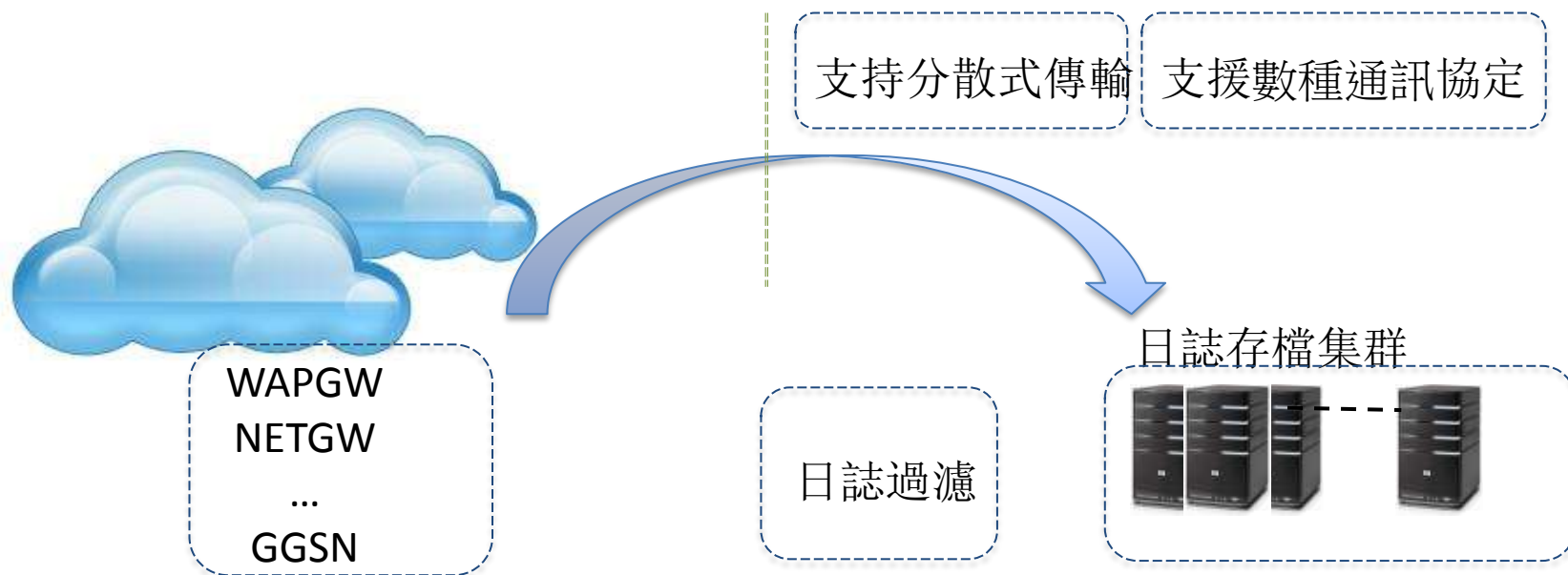


手機遊戲用戶：每天晚上頻繁使用手機玩遊戲，活躍時間集中在19點後

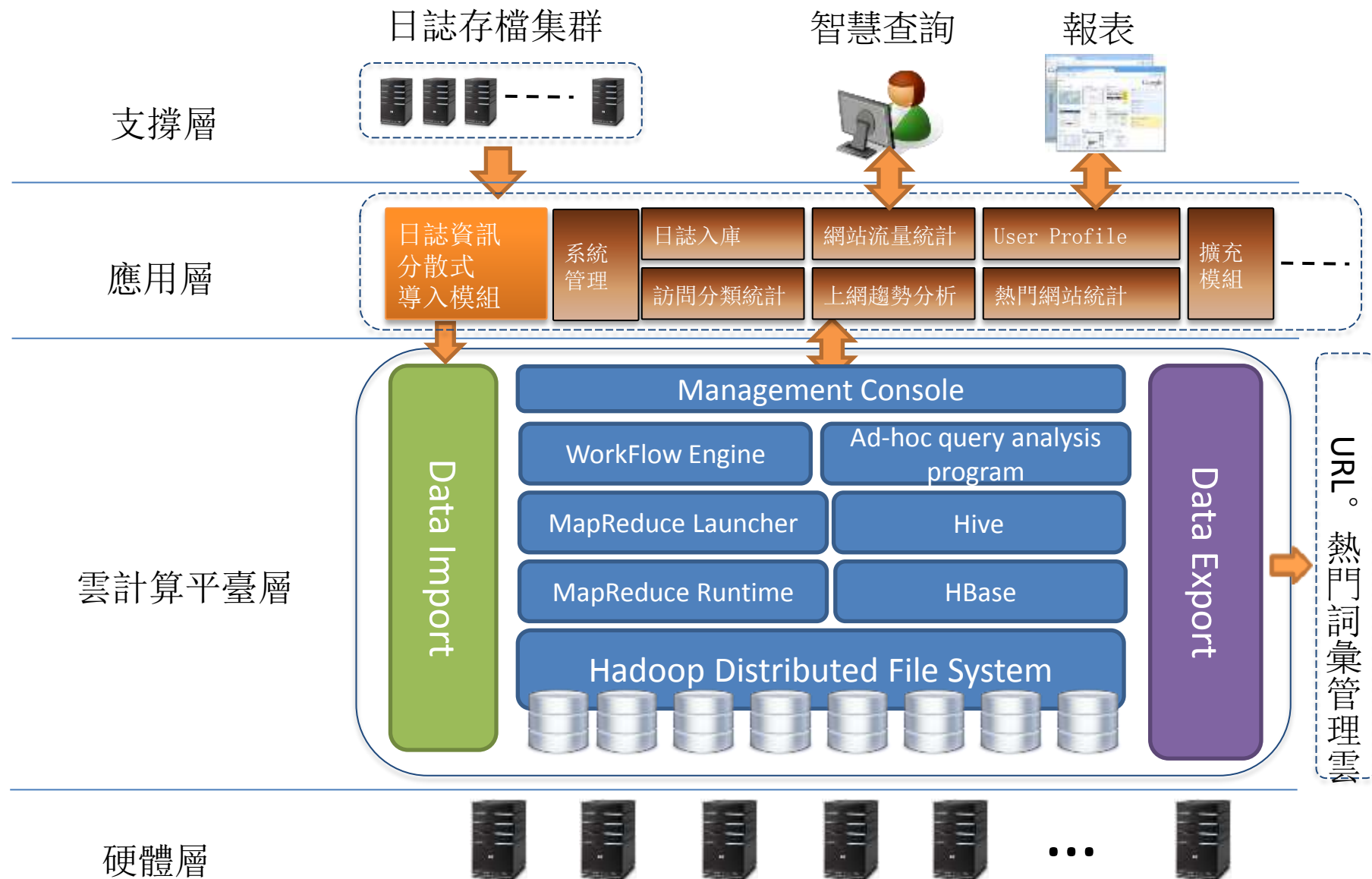
上網日誌資料需求

手機上網日誌分析系統所支援的資料類型有 .wap、.net數據；通過對資料來源按照過濾規則進行安全過濾，生成可被系統識別的資料分析來源；

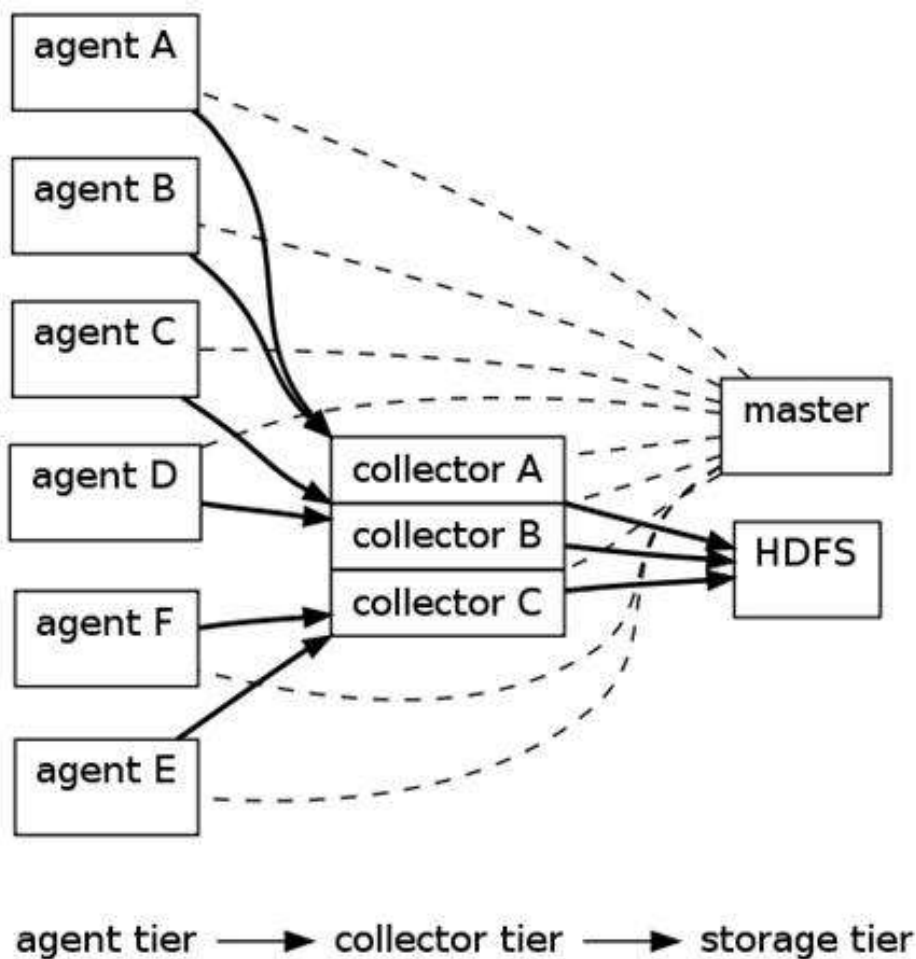
資料獲取方式：可通過傳統方式像FTP進行資料傳輸獲取、亦可支援私有協定傳輸；



整體平臺架構



日誌資訊 - 分散式導入

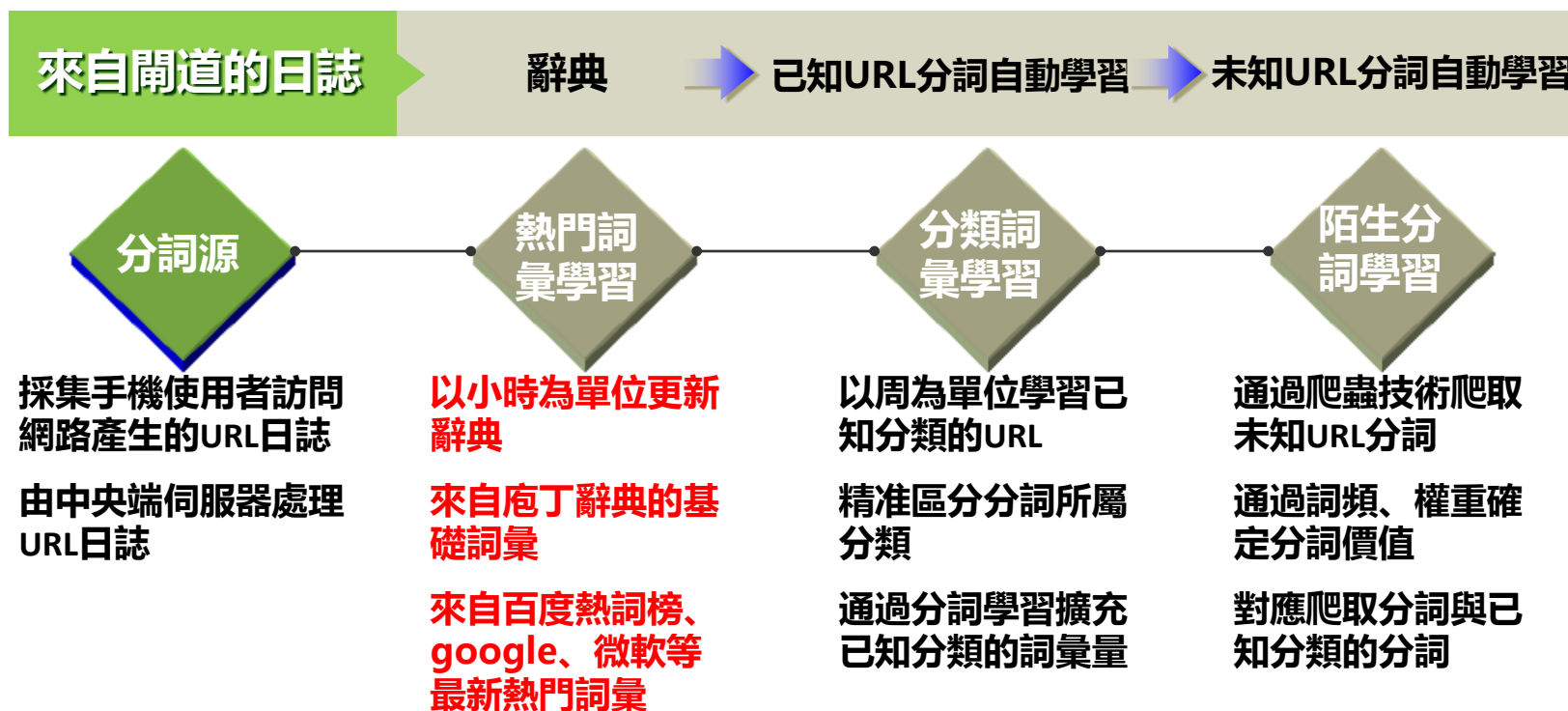


日誌資訊分散式導入模組：是分散式、可靠、和高可用的海量日誌採集、聚合和傳輸的模組，該模組可以同時向多個ftp伺服器發起讀取資料請求併發存入分散式檔案系統，大大提升入庫效率。

該模組提供了從console（控制台）、RPC（Thrift-RPC）、text（文件）、tail（UNIX tail）、syslog（syslog日誌系統，支援TCP和UDP等2種模式）、exec（命令執行）等資料來源上收集資料的能力。同時，分散式日誌收集模組可作數據接受方，可以是console（控制台）、text（文件）、dfs（HDFS文件）、RPC（Thrift-RPC）和syslogTCP（TCP syslog日誌系統）等

不斷成長的智慧分詞庫

智慧雲系統的分詞庫，一直在不斷成長和不斷完善。通過學習百度熱詞榜吸納新的詞彙；通過對已知分類URL的分詞學習擴充不同分類的分詞量；通過對未知分類URL的分詞學習確定URL的分類。



定義用戶特徵 – 多樣化的分析維度

上網活躍度：高
終端：HTC

上海 28歲
遊戲迷
每月80M流量
晚8點
男性
android
音樂下載



智慧雲將雜亂無章的用戶標籤細分成五個維度詮釋每個獨立用戶

基礎屬性特徵：
年齡、性別、上網時間等

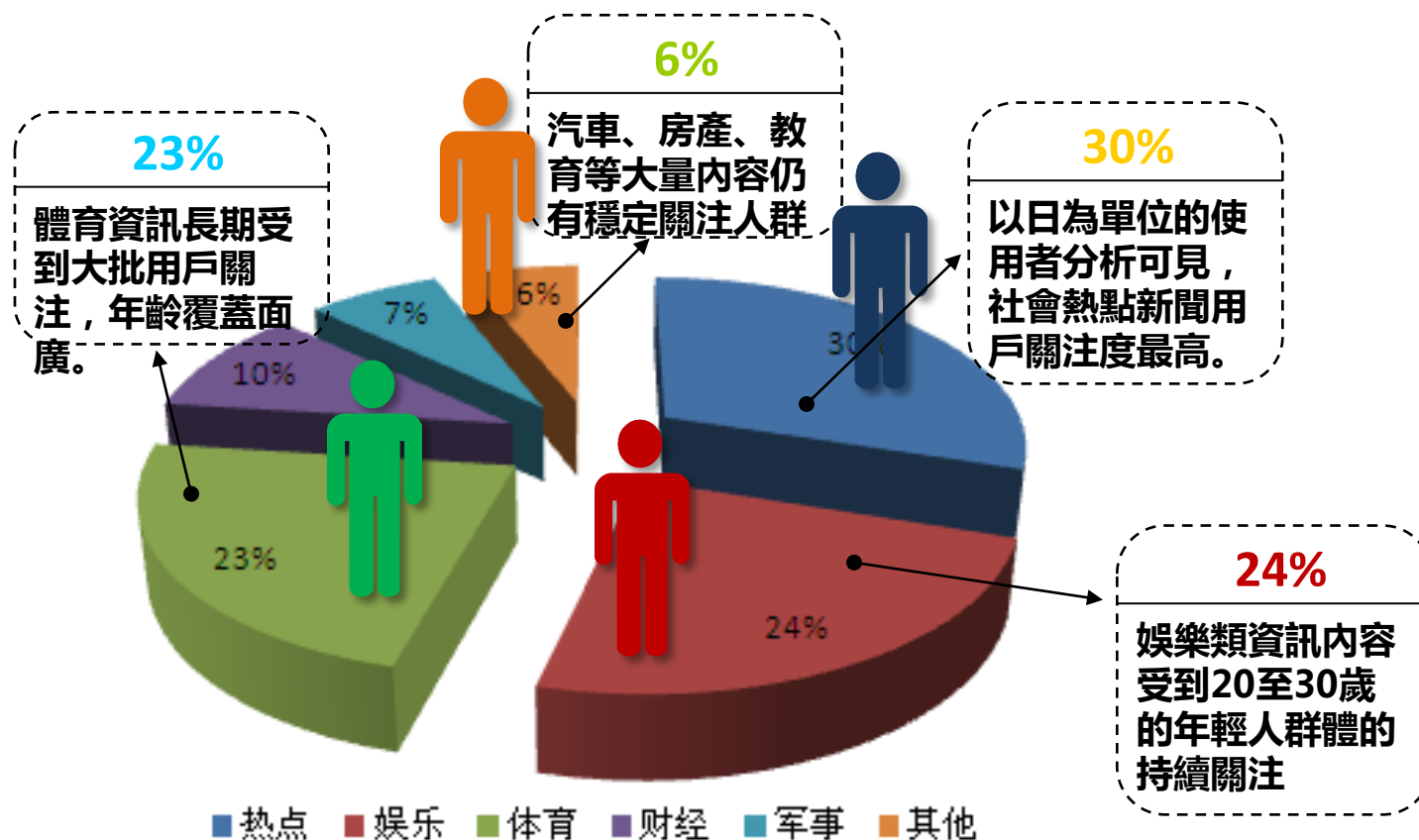
內容喜好特徵：
軍事、體育、音樂、遊戲等

行為特徵：
下載、流覽、搜索、郵件等

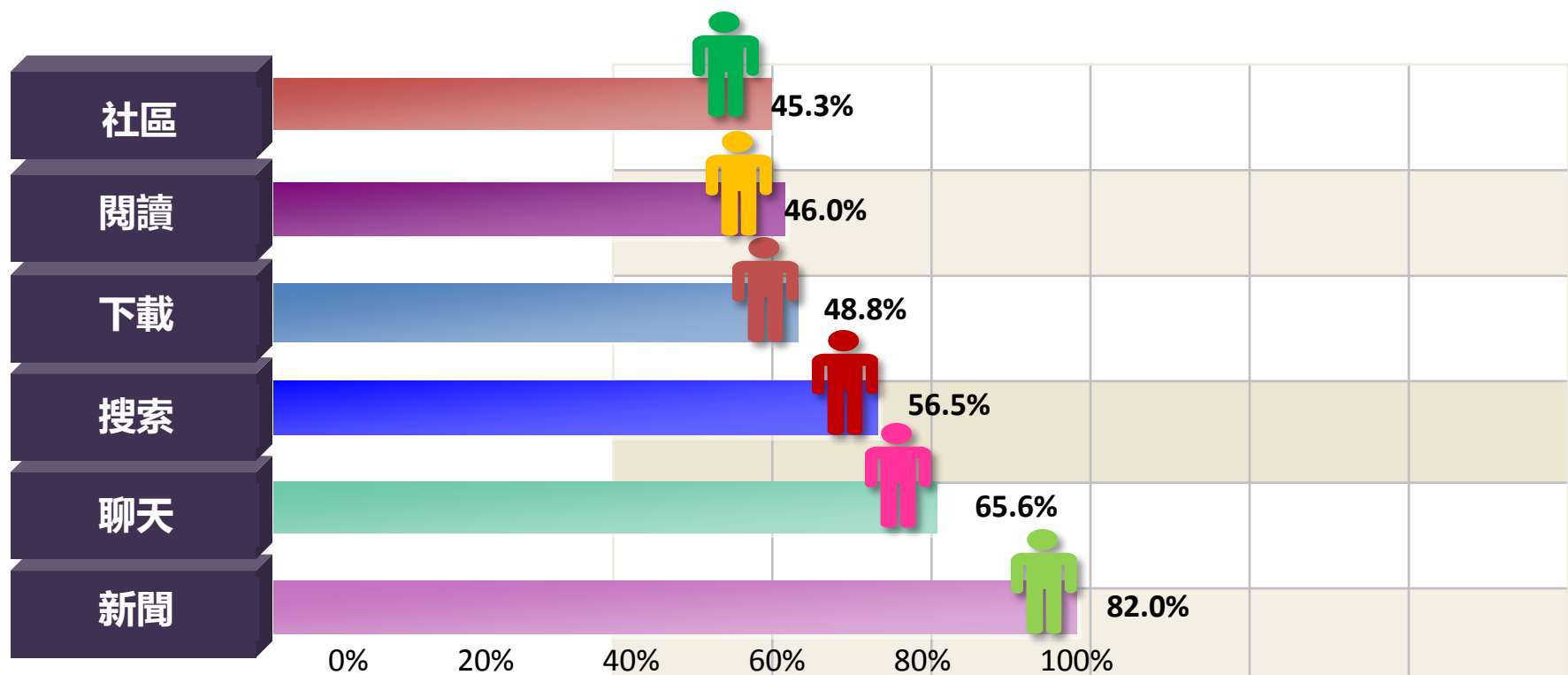
終端屬性特徵：
手機品牌、型號等

應用偏好特徵：
通訊、音樂、遊戲、閱讀等

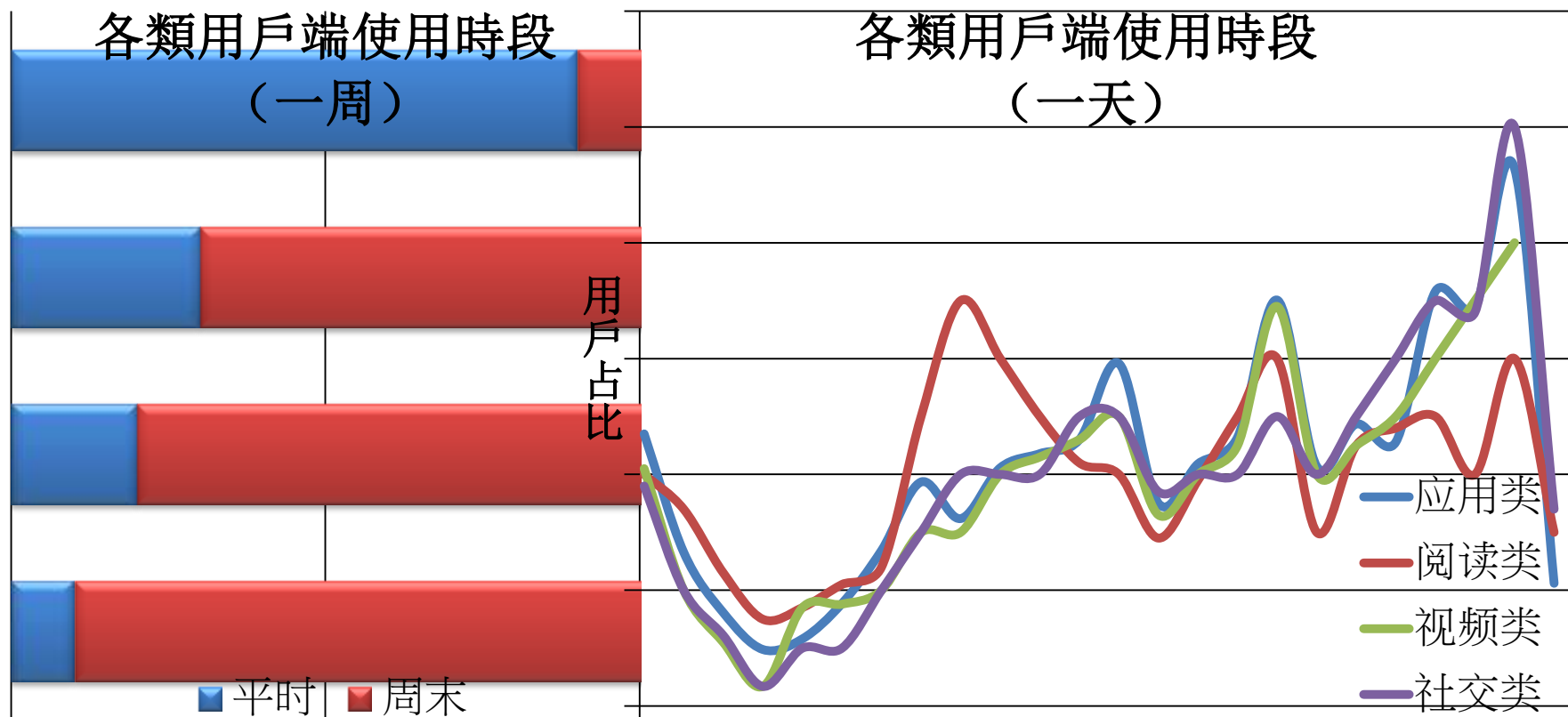
使用者喜好分析 – 瀏覽網站內容



手機上網用戶行為分析



手機用戶端分析



SmartPush自動配對



內容

國際
軍事
巴格達
伊拉克
爆炸
士兵



商品

購物
團購
購買
人字拖
凡客誠品



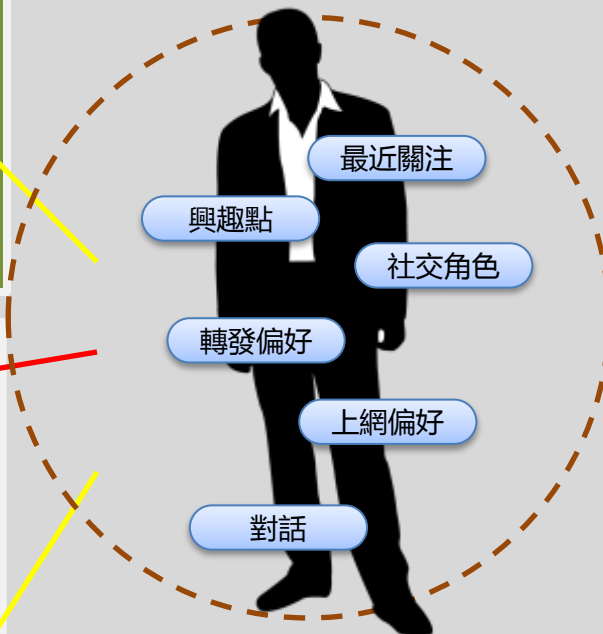
業務

動漫
漫畫
海賊王
蒙其D路飛
日本



業務

情感
小說
愛情
閱讀
女性



男性 30歲 流量80M
白領 國貿 iphone
新聞 軍事 漫畫 交友



女性 28歲 流量50M
歌手 上海 HTC
購物 小說 張小嫻 交友



SmartPush自動運營

自動運營主要體現在：選擇恰當的資訊推送管道將個性化內容送達使用者手中。

管道一：WAP PUSH

智慧可通過嵌套、門戶及手機用戶端等管道完成產品的自動運營，實現應用及服務的個性化推薦



頁面嵌套



衝浪門戶



手機用戶端

多管道
自動運營

管道三：API介面能力開放
開放API介面，擴展更多管道

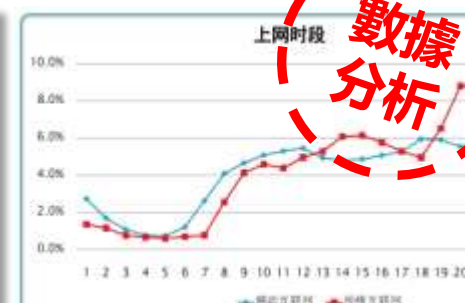
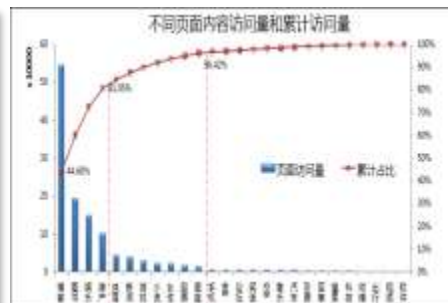
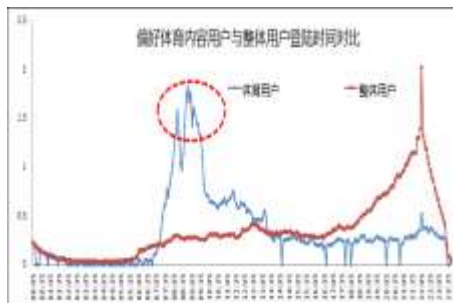


管道二：短訊&圖片訊息
通過傳統的SMS/MMS管道
完成內容的推送



案例分享 - 報表

目前已有7個省運營商使用智慧平臺對使用者屬性分析和資料採擷，並自動運營



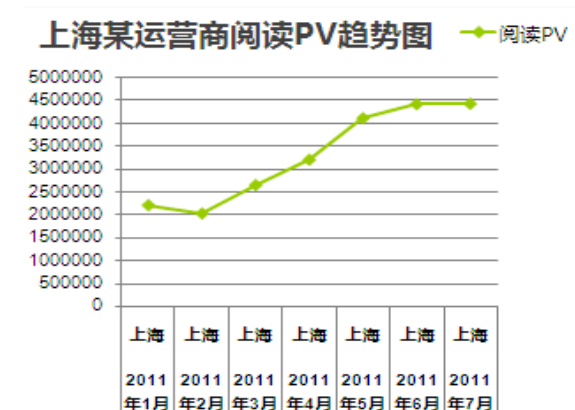
數據分析

特定内容使用者訪問時段分析

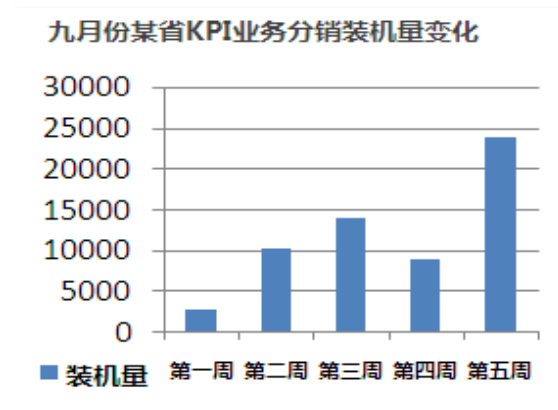
財經類熱門網站分析

使用者内容喜好訪問量分析

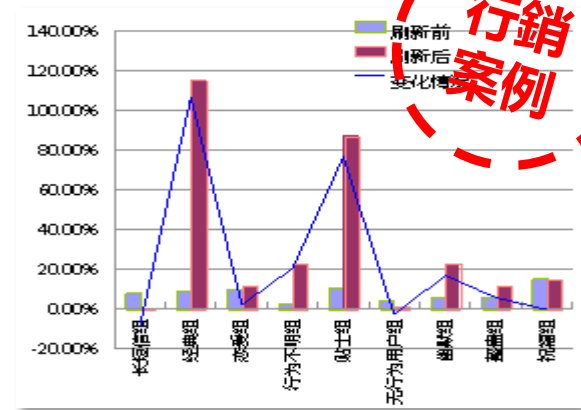
用戶上網時段趨勢分析



閱讀業務PV提升運營案例



KPI業務提升運營案例



行銷案例

短彩轉發業務運營案例

謝謝！

Vincent_Chen@tcloudcomputing.com

