

從「條數據」到「塊數據」

大數據時代真正到來的標誌

全面闡述大數據時代的顛覆性變革
一個新的時代到來了

塊數據1.0

藍天書
圖書

大數據戰略重點實驗室 著
連玉明 主編

數據是驅動商業向前發展的核心，更是人類社會的未來。
DT時代不僅僅是技術的升級，更是思想意識的巨大變革，
我可以肯定地說，這無疑是一個更波瀾壯闊的大時代。

——馬雲

拉開大數據序幕的真正推手
推動企業、政府巨變的引爆器
風口之上，得「塊」者得天下



塊數據

1.0

大數據時代
真正到來的標誌

大數據戰略重點實驗室◎著

連玉明◎主編

大數據對經濟發展的推動已經超出了所有最偉大的預言家的預測，而本書認為大數據對社會發展的推動更將是一個史詩般的劇目。這齣劇的序幕正準備拉開或者說剛剛拉開，而拉開序幕的推手就是塊數據。

——大數據戰略重點實驗室

Business 041

作者：大數據戰略重點實驗室

編輯：藍天圖書編輯組

出版：紅出版（藍天圖書）

地址：香港灣仔道 133 號卓凌中心 11 樓

出版計劃查詢電話：(852) 2540 7517

電郵：editor@red-publish.com

網址：http://www.red-publish.com

香港總經銷：香港聯合書刊物流有限公司

台灣總經銷：貿騰發賣股份有限公司

地址：新北市中和區中正路 880 號 14 樓

電話：(866) 2-8227-5988

網址：http://www.namode.com

出版日期：2020 年 2 月

圖書分類：資訊管理／商業策劃

ISBN：978-988-8568-72-7

定價：港幣 148 元正／新台幣 596 圓正

塊數據 1.0 by 大數據戰略重點實驗室

中文繁體字版 2019 由香港紅投資有限公司

本作品中文繁體版經中信出版集團股份有限公司授予香港紅投資有限公司獨家發行，非經書面同意，不得以任何形式，任意重製轉載。

編撰委員會

編委會主任：陳剛

編委會常務副主任：劉文新

編委會副主任：聶雪松 高衛東 徐昊 連玉明
康克岩 李瑤

主編：連玉明

主要撰稿人：王黔 唐振江 劉春輝 陳棟
王鵬飛 徐先文 李祥 王麒
趙剛 楊達 白延成 張清
諶貽建 張潔 龍榮遠 張龍翔
鄒濤 翟斌 楊官華 王倩茹
陳威

學術秘書：李瑞香 龍婉玲

大數據是什麼並不重要，重要的是大數據改變了人們對世界的認識。大數據不僅改變了貴州和貴陽對世界的認識，更重要的是也改變了世界對貴州和貴陽的認識。貴州和貴陽已經成為中國大數據發展的戰略策源地，並正在成為引領全球大數據發展的重要風向標。這個戰略策源地和重要風向標的主要標誌就是以塊數據為核心的理論創新，搶佔了大數據發展的理論制高點和國際話語權。

應該肯定，塊數據是貴州和貴陽發展大數據理論創新和實踐探索的產物。如果說，大數據第一次讓貴州和貴陽站在了世界面前，那麼，在這面前的正是標誌著大數據時代真正到來的塊數據。塊數據對未來文明的改變和推動已經超出所有偉大預言家的預測，這一史詩般劇碼的序幕正在拉開或者剛剛拉開。事實上，對塊數據的認識和研究從來沒有停歇過，我們建構了一個從 1.0 到 5.0 的塊數據理論體系，試圖探索和揭示塊數據的本質、規律和價值。正如我在《塊數據 2.0：大數據時代的範式革命》一書主編序中描述的那樣：「這是一場由科技引發的社會變革，它將改變我們的思維方式和生

活方式，改變世界上物質與意識的構成，改變我們的世界觀、價值觀和方法論。我們不得不更多地關注它、研究它和把握它，因為我們每個人都置身其中。這也是我們必須對塊數據肅然起敬，並持續探尋的根本動因。」

不僅如此，我們的研究發現，塊數據具有更加深遠的意義：

第一，塊數據是大數據時代真正到來的標誌。新一輪科技革命和產業變革正處於重要交匯期。隨著信息技術和人類生產生活交匯融合，我們已進入以大數據為標誌的信息化發展新階段。人類將以塊數據為標誌，真正步入大數據時代。塊數據就是把各個分散的點數據和分割的條數據匯聚在一個特定平台上並使之發生持續的聚合效應。這種聚合效應通過數據多維融合與關聯分析對事物作出更加快速、更加全面、更加精準和更加有效的研判和預測，從而揭示事物的本質和規律，推動秩序的進化和文明的增長。簡言之，塊數據就是研究數據運動規律的數據哲學。數據是運動的，數據運動是有規律的，數據運動所揭示的是數字文明時代秩序的增長。我們認為，塊數據是大數據發展的高級形態，是大數據融合的核心價值，是大數據時代的解決方案。

第二，塊數據是數據、算法、場景融合應用的價值體系。數據、算法、場景是治理科技的三大核心要素。其中，數據是基礎，算法是手段，場景是目的。塊數據價值鏈是實現超越資源稟賦的價值整合，是以全產業鏈、全服務鏈和全治理鏈為核心的價值體系。通過數據、算法、場景的疊加效應，在塊數據系統架構下減量化的數據採集、數據傳輸、數據存儲、數據分析和數據應用的數據觀和

方法論，為我們建構一個融合技術流、物質流、資金流、人才流、服務流的價值系統。以發現塊數據內海量複雜數據的潛在關聯和預測未來為目標，以複雜理論的系統思想為主要範式，實現對不確定性和不可預知性更加精準的預測。從條數據到塊數據的融合，人類社會的思維模式和行為範式將產生顛覆性變革。這種變革的文化力量就是以人為原點的數據哲學，它不僅革新了我們的世界觀、價值觀和方法論，而且開啟了我們的新時代、新生活和新未來。

第三，塊數據引領和催生新的組織模式，成為改變未來的新力量。塊數據既是一種經濟模式，也是一種技術革新，更是一種新的世界觀、價值觀和方法論，引領和催生新的組織模式。物物互聯、業業互聯、人人互聯，構成一個自組織、自激活的數據實時流動、共用、聚合的數據化網路生態圈。這個生態圈通過外部資源強化自身戰略地位，實現組織的再平衡。塊數據組織是一個資源分享、平台無限擴展的超穩定組織結構源，預示著組織發展的新方向。無邊界組織、自組織、雲組織等，都可以看作一種正在萌芽和生長著的塊數據組織。利他主義的數據文化是構成塊數據組織的理論基石，其出發點是數據人假設。塊數據組織中，數據力上升為組織的核心競爭力。數據力與數據關係影響著社會關係，這將引發整個社會發展模式前所未有的變革和重構。所有這一切，都預示著塊數據組織「平台化、關聯度和聚合力」三位一體所帶來的強大組織勢能，實現組織的自激活和自適應環境變化，是組織存續與發展的根本動力所在，並最終形成共用型組織新範式。

第四，啟動數據學成為人工智能時代大數據發展新的解決方

案。數據無序增長預示著超數據時代的來臨。小數據時代，數據越大，價值越大；大數據時代，數據越大，價值越小。人類對未來的把握和勝算，在於數據被有效存儲的同時又被迅速激活。我們沒有把大數據僅僅看作所謂的「大」的數據，而是把大數據看作一種「活」的數據，因為只有激活，大數據才有生命，才能成為未來世界人們賴以生存與發展的土壤和空氣。我們探索用生命科學的方法解決數據擁堵問題，以數據社會學的思維打開沉澱的數據寶藏。塊數據就是數據通過算法作用於場景，這種作用的動力就是激活數據學，它為我們尋找這小部分的有效數據提供了解決方案。作為一種理論假說，激活數據學就像一座朝向深邃的大數據宇宙的「天眼」。它是未來人類進入雲腦時代的預報，是關於混沌的數據世界的跳出決定論和概率論的非此即彼、亦此亦彼的複雜理論的大數據思維範式革命。

第五，塊數據提出的數據進化論、數據資本論、數據博弈論或將成為數字文明的「新三論」。美國學者塞薩爾·伊達爾戈的《增長的本質》一書被譽為「21世紀經濟增長理論的重要里程碑」。因為本書提出了一個重要觀點，經濟增長的本質是信息的增長，或者說秩序的增長。他認為，善於促進信息增長的國家會更昌盛。我們提出的數據進化論、數據資本論和數據博弈論，正是重構數位文明時代人與技術、人與經濟、人與社會的新秩序。我們的研究認為，數字文明時代增長的本質不是 GDP 的增長，而是文明的增長和秩序的增長。新「三論」的提出，對社會結構、經濟機能、組織形態、價值世界進行了再塑造，對以自然人、機器人、基因人為主

體的未來人類社會構成進行了再定義，對以數據為關鍵要素的新型權利範式和權力敘事進行了再分配。這既是研究未來生活的宏大構想，也是研究未來文明增長和秩序進化的重大發現。

我們對塊數據的研究才剛剛開始。數據驅動、算法驅動、場景驅動下的雲腦時代正在到來，激蕩著我們對於未來所有的憧憬和渴望。

我們永遠在路上。

連玉明

大數據戰略重點實驗室主任

2019年12月

大數據，作為一個影響深遠的名詞，2011年第一次出現在知名技術諮詢公司高德納公佈的年度新興技術成熟度曲線上。從那時起，各界對大數據的理論研究與應用實踐一直保持著前所未有的狂熱。大數據成為從少數人關注、少數人研究到整個社會高度關注的現象級詞匯。但在2016年公佈的年度曲線中，大數據作為一個獨立名詞已經悄然隱去，卻出現了一系列包括公民數據科學、微型數據中心、自我實現的高級分析等與其直接相關，可判斷人類意圖的技術、物聯網等與其間接相關的大數據衍生概念。這意味著，大數據已經進入從概念推廣到全面落地的關鍵轉折期，作為一種新的能源、權利和組織方式，全面滲透並深刻影響著人類政治、經濟、文化和社會生活的方方面面。

塊數據是貴陽發展大數據的理論創新和實踐探索的產物。應用場景是創新的唯一源泉，任何科學理論都不是憑空產生的，而是時代或實踐的產物。貴州省貴陽市在發展大數據實踐中的理論探索和創新正是塊數據這個新概念、新思想、新理論的源泉。塊數據推

動貴陽在大數據發展中搶佔了理論創新制高點、實踐創新制高點和規則創新制高點。

塊數據是大數據時代的引爆器。本書在「2015 貴陽國際大數據產業博覽會暨全球大數據時代貴陽峰會新聞發佈會」上首發以來，引起業界和社會強烈反響，並受到中央有關領導的高度評價，其銷量在亞馬遜（Amazon）網站曾一度排名前 10 位。馬雲認為，塊數據極具前瞻性，站在了時代的最前沿。如果說，大數據正在改變人對世界的看法，那麼，塊數據就是告訴你人是如何改變對世界看法的。從這個意義上講，塊數據是大數據的解決方案。

第一，塊數據是從數據到數聚的過程，這是塊數據的起點。數據是分散的、分割的、碎片化的，當這些分散的、分割的、碎片化的數據聚合在一起的時候，就開始產生「塊」，「塊」就是一個多維的無限的變量。這個多維是思維範式，無限是跨界，變量是一種不確定性和不可預知性，這是大數據時代我們認識世界的基礎，也是改造世界的方法。

第二，塊數據是一個從解構到重構的過程，這是塊數據的機制。一旦數據被集聚，就會形成「塊」，這種「塊」會對物質、能量、要素、權力、意識等進行解構。大數據時代人們獲取信息的方式、交往和交友的方式、生活方式、意識形態、社會組織模式都將發生深刻的變革，這種變革的本質就是解構。每一次解構的結果都會產生新的重構，比如權力被權利所替代，這就是解構中的重構。

第三，塊數據是從多維到共享的過程，這是塊數據的價值。大數據時代帶給我們最大的好處是什麼？如果概括起來解釋就是多

維和共享，就是每一個人在大數據時代能夠快速分享人類最先進的文明成果，這種多維和分享是可在任何時間、任何地點、任何人、任何事、任何方式下獲得任何信息，這就是共享的魅力。共享是大數據時代對人類最大的貢獻。我們過去不知道的事現在可以知道，我們過去不能獲得的信息現在可以獲得，過去少數人擁有的東西，現在大多數人都能擁有，這就是共享。共享正在成為一個新時代的標誌。所以，得「塊」者得天下，得「塊」者得未來。

塊數據是一個創新的課題。本書從三個部分通過 10 個章節加以論述。第一部分包括第一章、第二章、第三章，首次提出塊數據的定義，並闡釋了塊數據的基本屬性和價值。第二部分包括第四章至第七章，提出塊數據的形成、集聚及運營模式，塊數據公共平台和大數據市場交易機制，並指出開放、共享、連接是塊數據形成的基本機制。第三部分包括第八章至第十章，提出塊數據實踐應用所形成的全產業鏈、全服務鏈和全治理鏈，將更大程度地驅動產業升級、服務改善民生、推進城市轉型和倒逼政府改革。

塊數據的產生，將給我們的時代帶來跨界、跨代、顛覆式變革。人類將以「塊數據」為標誌，真正步入大數據時代。以塊數據為核心的數據驅動將解構和重構資源配置方式，從競爭逐利到合作共贏、從信息不對稱到交互零成本、從集中化到平台化、從流程化到模塊化、從重資產到輕資產、從點對點到多對多，試圖完全「建立一種新的生產函數」，把從來沒有的關於生產要素和生產條件的「新組合」引入生產體系，從根本上改變了傳統生產力與生產關係，整個社會生產關係都被打上數據關係的烙印。這種新的數據關

係，深刻地推動整個經濟結構的變革和價值鏈重構。這種重構將進一步變革數據力與數據關係，實現效率與公平的高度統一，實現共享經濟到共享社會的轉變。

連玉明

大數據戰略重點實驗室主任

2018年11月8日於北京

推薦序一

受摩爾定律驅動的信息技術的不斷廉價化、互聯網的普及及其延伸所帶來的信息技術無處不在的應用，催生了大數據時代的到來。進而使信息化進入以數據的廣泛關聯、跨域融合和深度應用為特徵的智慧化階段。在當前的大數據熱潮中，相關的書籍、文章可謂車載斗量，共識與各種觀點的爭鳴共存。本書卻從一個新穎的視角——塊數據，來看待大數據及其未來的發展，頗有創意，也發人思考。讀完本書，感受有三。

其一，「條數據」和「塊數據」的劃分，師法自然，抓住了數據的本質。從其定義和靜態的角度看，「條」是一個領域或行業內縱深數據的集合，可以反映本領域或行業的規律，無疑具有很大價值。「塊」是一個物理區域或行政區劃內眾多「條數據」的集合，更能反映現實世界和社會的極度複雜性，其綜合應用無疑會帶來數據價值的顯著提升。從動態的視角看，重視「塊數據」是為了避免僅僅關注「條數據」而可能帶來的新的數據孤島現象，更體現了一種對信息化建設的發展性思維。

其二，討論大數據，很多人都會談到當前價值密度高的數據大多掌握在政府手中，期待政府依法依規開放數據並推動數據共享。本書恰恰是從政府管理和社會治理的視角，使用了在管理和治理體系中常用的「條」、「塊」概念，著眼於當前「最有價值」的一部分數據的有效、高效應用，其意義不言而喻。

其三，本書源於貴陽市領導在推進大數據應用及其產業發展過程中的思考和實踐，既有理論的闡述，也有實踐的支撐，呈現了對諸多方面極有價值的參考。期待貴陽市在本書指引下，在其大數據驅動的智慧城市建設中，積極探索，穩步前行，為信息化建設的新階段——智慧化，樹立典範。

梅 宏

全國信標委大數據標準工作組組長
中國科學院院士、上海交通大學副校長
2015 年 2 月

推薦序二

雲計算、大數據現在已經成了科技發展的代名詞，我和不少政府官員討論過雲計算、大數據對社會經濟發展和資源配置的影響，深感貴州政府是對新技術最敏銳、最大膽的實踐者之一。

數據是驅動商業向前發展的核心，更是人類社會的未來。我們看到這樣一個趨勢：以控制為出發點的 IT 時代正在走向以激活生產力為目的的 DT（數據技術）時代。從組織內部角度來看，DT 會改變一個組織的溝通、生產、消費方式，驅動它的架構、文化的變革；從跨組織角度來看，由於 DT 時代的「利他」思維取代了 IT 時代的「利我」思維，組織與組織的合作將遠大於競爭，跨組織的協同會頻繁發生，而且將變得越來越敏捷，越來越高效。這不僅僅是技術的升級，更是思想意識的巨大變革，我可以肯定地說，這無疑是一個更波瀾壯闊的大時代。

阿里巴巴集團於 2014 年 4 月 17 日與貴州省政府簽訂了「全面戰略合作協議」，項目之一「雲上貴州」已經取得一些成績，成為政府運營雲計算和大數據的最佳實踐。政府作為一個組織，生來就

是一個極為重要的數據生產和交換平台。數據本身並不能創造價值，只有讓更多的人對其進行分析和運用，才能成倍地創造價值。貴州省政府、貴陽市政府、貴安新區就是本著這樣一個開放的心態，將自身打造為推動地區發展的數據引擎——而非數據倉庫，將政府數據公開，鼓勵那些有創造力的年輕人在這個平台上創業，挖掘新的商業機會，既提高了執政效率又促進了民生服務。這不僅是對雲計算領域的預判和佈局，也給本地區注入了新的活力，讓當地年輕人有機會站在時代的最前沿。

很多地方都在談數據經濟、雲計算產業，但貴州下如此大的決心，善於利用阿里巴巴集團業已搭建的雲計算和數據平台，跟我們一起幹、堅持幹、務實幹，一起探索未來，這樣的勇氣和魄力值得敬佩。

作為一位專研型官員，貴陽的陳剛書記既是上述實踐的推動者、實施者，也是大數據理論的研究者。我不懂技術，但深知一邊做，一邊想，每天努力，未來才會到來。

祝願我們理想成真。

馬雲

阿里巴巴集團董事局主席

2015年新春於杭州

序

塊數據是一個新名詞。到目前為止，人類生活、生產產出的大數據，無論是搜索引擎、電商、社交平台形成的企業大數據，還是天氣預報形成的科學大數據，還是各政府部門形成的行業大數據，都可以定義成「條數據」。而塊數據則是以一個物理空間或行政區域形成的涉及人、事、物的各類數據的總和。

在此書完稿後，不少人士甚至是專業人士都問我，你能用一個形象的比喻把塊數據說清楚嗎？我想舉個例子說明：一個百貨商場賣出許許多多的商品，每個商品有從原材料到加工成商品的生產過程的數據，也有品牌設計、廣告營銷和銷售數量的數據，以及產品售後服務、商戶反饋等服務數據，這些都是以產品為中心的「條數據」，而百貨商場銷售的商品種類、數量，男女老少在商場的購物、娛樂情況，天氣、公交和停車場對商場經營情況的影響等卻易被忽視。事實上，當商場的影院播映一部聚人氣的大片時，商場的銷售量也會隨之上升，這些數據可以稱為塊數據，這個「塊」是指這個商場，商場這個物理空間產生的數據總和就是商場的塊數據。

當塊擴大到社區和城市層面時，在這個塊上形成的數據總和就是本書所指的塊數據。

現任國家大劇院院長的陳平先生，在 21 世紀初時任北京市東城區區委書記期間，探索並構建了一整套的「網格化治理」城市的模式，時任微軟總裁比爾·蓋茨先生由衷評價：「東城區網格化是人類發明計算機以來在社會領域應用方面最偉大的發明。」國家建設部在全國推廣複製這一模式。本人在北京朝陽區工作期間，學習借鑒東城經驗，並將這個模式從對「物」的管理，轉向對社會治理中動態的「人」、「事」和「物」的管理，探索了「全模式」社會治理的路徑。由於受技術的局限和歷史因素限制，這些探索儘管已經足夠了不起了，仍屬於初步探索。例如，「萬米網格」給所有公共物件進行人工數字編碼，然後靠網格員手持掌上電腦（PDA），依靠人的巡查去控制各個公共物件的運轉情況，把這些公共物件的信息匯聚在一起，形成物聯網（IoT）的初級階段，也可以稱為手工物聯網。

過去的 10 年，人類在互聯網、雲計算、大數據、物聯網等領域突飛猛進，這些技術的進步成就了矽谷的蘋果、谷歌（Google）和中國 BAT（百度、阿里巴巴、騰訊）等這樣一批互聯網企業，人類目前已經認識到，「大數據」將在經濟領域徹底顛覆人類自工業革命以來積累形成的經濟模式和商業模式。

塊數據的貢獻，將會使人類在大數據領域初步探索並逐步形成的條數據基礎上，在一個物理空間形成匯聚，形成條數據和塊數據的融合式發展。這正是本書以「塊數據——大數據時代真正到來

的標誌」為題目的原因。只有條數據和塊數據融合式發展，人類所描述的大數據的時代特徵才真正開始實現。

本書對塊數據的基本概念、基本特徵進行了闡述，對塊數據的聚集路徑進行了設計，進而特別提出了塊數據有別於條數據的最大特徵——共享性和可交易性。與此同時，本書指出塊數據對經濟發展的推進，特別是數據共享和可交易性為時下推崇的「大眾創業、萬眾創新」的創新驅動發展模式，插上了騰飛的翅膀。更為重要的事情是，塊數據是提高社會治理能力、更好地服務民生的「共性需求」和「長尾需求」的一條現實的可實現路徑。此外，大數據讓「世界變成透明的」，塊數據在對監督政府依法行政行為和倒逼政府職能改革方面可實現其他任何外力手段不能達到的推動力。

需要說明的是，本書所講的塊數據不是過去這些年各地方政府的「智慧政府」、「智慧城市」的「集成版」。與後者有別，塊數據構建的模式是開放的、共享的和可自更新的，是一種可持續發展的模式。它解決的不僅僅是「數據孤島」和統一平台標準問題，更多的是一種「互聯網思維」模式，由政府、社會和個體共同以「眾包」的模式構建一個面向未來的數據平台。本書對這方面的內容從機制上和構架上進行了闡述。

本書適合互聯網的創業者閱讀，讀者會從本書內容中判斷互聯網的未來發展方向和商業趨勢。本書適合企業管理者閱讀，當下對大數據概念的炒作可能讓你對互聯網發展方向有點兒迷茫，本書可以告訴你互聯網大數據發展的方向，讓你從一個全新的視角重新認識大數據。本書也適合金融家、投資者閱讀，你會從中找到投資

的方向和發現成千上萬的新投資機會，或許你就是下一批「BAT」的早期股東。本書同樣適合政府公職人員閱讀，社會治理能力提升的根本出路在於信息化，而塊數據是實現社會信息化的必由之路。你會從本書中找到提升政府治理能力的方向。

最後還要說明的是，本書不是互聯網、大數據行業的入門讀物。當然，如果你有志於關注互聯網、了解大數據的話，這本書或多或少會對你有幫助。大數據是新生事物，現在不存在真正意義上的大數據專家。若干年可能甚至若干月以後，你會發現這本書的有些觀點已經過時、落伍，甚至是錯誤的，那麼本書出版的目的也許就實現了，因為這意味著大數據特別是塊數據又取得了新的進展和突破，而這也正是本人所希望看到的。

感謝在本書創作、完稿過程中各方的支持者。

陳剛

2015年2月於貴陽

總序	VII
主編序	XIII
推薦序一	XVII
推薦序二	XIX
序	XXI

引言	XXIX
----	------

➤ 上篇

塊數據的定義與特徵

第一章 塊數據的出現——時也，勢也

- (一) 條數據——越長大越孤單 004
- (二) 塊數據——坐擁天時和地利的寵兒 013
- (三) 「塊」思辨——塊數據哲學體系初步構建 016

第二章 塊數據素描——「金礦」是怎麼煉成的

- (一) 高度關聯性：「你」「我」「他」與「們」 021
- (二) 立體性：塊數據的維度結構 024
- (三) 活性：隨時隨地都在更新 027
- (四) 主體性：以人或物的活動為起源 030
- (五) 開放性：塊數據的基因表達 035

第三章 塊數據的價值——得「塊」者得天下

- (一) 商業模式發掘潛力巨大 041
- (二) 社會領域應用前景廣闊 046
- (三) 政府職能發生深刻轉變 049
- (四) 社會公共文化被重新定義 052

➤ 中篇

塊數據的形成、集聚及運行模式

第四章 塊數據的挑戰——雄關漫道真如鐵

- (一) 頂層機制設計時不我待 058
- (二) 數據結構化挑戰巨大 059
- (三) 數據安全問題日益凸顯 061
- (四) 保護公民隱私刻不容緩 062
- (五) 行業重大變革即將到來 065

第五章 塊數據的形成——使數據「聚變」

- (一) 「寬帶城市」為塊數據形成提供「反應堆」 069
- (二) 「無線城市」為塊數據形成注入「催化劑」 074
- (三) 「物聯城市」為塊數據形成安裝「引爆器」 080

第六章 塊數據的匯聚和處理——讓「富礦」名副其實

- (一) 打造公共平台，讓塊數據匯聚得更迅速、更全面、更智能 092
- (二) 打破技術瓶頸，將數據處理能力作為公共平台的核心支撐 099
- (三) 增益塊數據價值，使「雲上富礦」飄進千家萬戶 111

第七章 塊數據開放與市場交易機制構建——數據資產化

- (一) 數據開放機制構建——史無前例的深度開放 120
- (二) 大數據交易機制構建——把數據變成一門生意 125
- (三) 網絡及數據安全保障機制構建——守住國家安全的第四維 132

➤ 下篇

塊數據應用

第八章 商業領域：大數據，大產業

- (一) 服務業領域——告訴你誰會點擊、購買、差評 143
- (二) 工業領域——智能製造 151
- (三) 農業領域——精準性和精細化 157
- (四) 塊數據的全產業鏈——一片新的藍海 160

第九章 民生領域：大數據，大民生

- (一) 文化教育——數字化、在線化和個性化 167
- (二) 健康醫療——「你未來五年患癌的概率是 80%」 171
- (三) 就業創業——找到自己真正熱愛並適合的工作 174
- (四) 社會保障——全天候關懷與長尾化幫扶 178

- (五) 交通出行——尋找最短最暢通的那條路 180
- (六) 城市安全——由末端治理轉向源頭預防 184
- (七) 塊數據的全服務鏈——無處不在的精心呵護 187

第十章 政府治理領域：大數據，大治理

- (一) 透明政府——讓政府變得不再神秘 191
- (二) 智慧政府——打造塊數據時代的政府 2.0 199
- (三) 法治政府——把「牛欄」變成銅牆鐵壁 207
- (四) 責任政府——塊數據時代的政府使命 213
- (五) 塊數據的全治理鏈——共建共治的新格局 218

結語 223

後記 229

參考文獻 233

引言

當前，世界正經歷一場更大範圍、更深層次的科技革命和產業變革。雲計算、大數據、物聯網、人工智能、區塊鏈等新一代信息技術的突飛猛進，讓各類數據急劇累積、快速流動。在由實體物質構成的現實世界背後，一個以數據為基本單元的數字世界正在形成。數據逐漸成為繼煤炭、石油之後的又一重要生產要素。國家競爭焦點已從資本、土地、人口、資源轉向數據空間，全球競爭版圖也將分成兩大新陣營：數據強國與數據弱國。

數據已滲透到每一個行業和領域，它改變著我們的生活，推動社會發生著一場深刻變革。從國家治理角度，大數據提升了治理效率、重構了治理模式、破解了治理難題，掀起了一場國家治理革命；從經濟增長角度，大數據是全球經濟低迷環境下的產業亮點，是戰略新興產業的最活躍部分；從國家安全角度，數據空間沒有國界邊疆，大數據能力成為大國之間博弈和較量的利器。大數據呼嘯而來，攪動世界風雲。順應趨勢、積極謀變的國家和企業將乘勢崛起，成為新的領軍者；無動於衷、墨守成規的組織將逐漸被邊緣化，

失去競爭的活力和動力。毫無疑問，大數據正在開啟一個嶄新時代。然而，需要正視的是，在機遇頻現的同時，也有很多矛盾和問題亟待處理和解決。

大數據時代的第一重矛盾是爆炸性增長的數據量和普通人的數據分析能力之間的矛盾，即數據焦慮。在數據稀缺的時代，人們希望數據多多益善，而在數據爆炸的時代，人們則希望數據越精準越好、越有序越好。海量數據的複雜性不僅體現在數據的體量上，更體現在多源異構、多實體和多空間之間的交互動態性上，以至於我們不能用傳統的方法去描述與度量這些數據。數據分析者、研究者與決策者再一次陷入結構性短缺、集團化佔有、分散化使用和有限性分享等新的困局之中，這直接影響了大數據預測的便捷性、準確性和客觀性。

大數據時代的第二重矛盾是數據孤立與數據融合之間的矛盾，即數據割據。到目前為止，人類形成的大數據，更多的是以個體、領域、行業為單位，往往是彼此割裂、互不相通的數據。各主體按照自己的需求搜集數據，形成統籌自己的「條」，造成了巨大的資源浪費，也在無形中為自己樹立了壁壘。即便能把這個「條」做到極致，它們也是孤單的，就像孤島上的流放者一樣，未形成合力，無法應對數據汪洋的變幻莫測。

大數據時代的第三重矛盾是數據垃圾和人類認知之間的矛盾，即數據擁堵。人類越來越認識到大數據的價值，數據被大量的發現、啟封和挖掘。但在人類看到數據價值的同時，太多無用的數據呈現在人類面前。在小數據時代，數據越大，數據越有價值；在大

數據時代，數據越大，價值越小。僅僅依靠現有的認知水平，人類很難從這無盡的數據海洋中挖掘出無盡的數據價值。舍恩伯格所說的大數據四維特徵——海量、價值、速度、結構，都將成為大數據發展的致命弱點。因為任何一個事件、任何一個設備都將產生無限的數據垃圾。無邊界的數據垃圾將給人類造成認知障礙，數據擁堵將會成為人類社會發展的重要問題。

正是在這樣的時代大背景下，「塊數據」應運而生。塊數據以某一具體的空間為出發點，發揮人類的主觀能動性，推動數據共享開放，實現數據融合，保證數據的指向性和針對性，淘汰垃圾數據，挖掘出更高、更大的數據價值。可以說，塊數據就像人類在數字世界的基地，是我們對這個新世界認知的起點。這是一次新的拓荒，這是一次新的探索。當在數字世界中建立起越來越多的「基地」，並最終連成一片形成新的國度時，就意味著新文明的誕生——數字文明時代最終到來。



上篇
塊數據的定義與特徵

第一章 塊數據的出現——時也，勢也

2013 年被人們稱作「大數據元年」。相對於大數據元年的到來，大數據的應用在更早時就已出現。2011 年 6 月，麥肯錫諮詢公司發佈了《大數據：下一個競爭、創新和生產力的前沿領域》的研究報告，在初步介紹大數據產業發展、實際運用情況的基礎上做了對未來的展望。大數據的產生和應用是人類進入信息社會新階段的重要標誌。必須看到，人類文明誕生之初，數據就存在於人類生活的方方面面，只不過當時的人類無法發現和獲得這些數據信息，有限的數據傳播是通過書籍、教育、工藝傳承等手段實現的。互聯網是人類有史以來最偉大的發明之一，它不僅使人類把海量數據連接起來，而且使數據的價值得以挖掘，大數據的海量性和關聯性將可為人類所掌控。但大數據價值的發現不會是一步到位的，既有歷史傳承，又要創新發展。

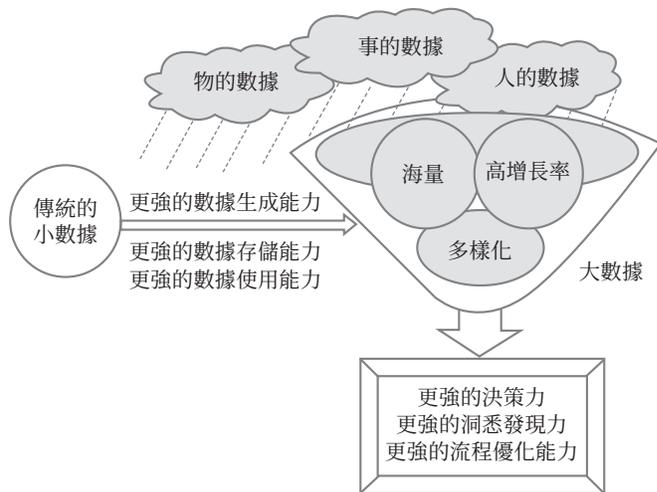


圖 1-1 大數據的概念、形成與價值

(一) 條數據——越長大越孤單

到目前為止，人們所討論和應用的數據幾乎都是被某個行業或領域掌握的「條數據」。無論是傳統企業掌握的，如銀行卡、會員卡所匯聚的企業內部數據，再到各級政府實施的信息化工程所掌握的衛生、教育、交通、財政、安全等部門數據，還是互聯網企業存儲的電子商務、互聯網金融等行業數據，都可以定義為「條數據」，即在某個行業和領域呈鏈條狀串起來的數據。

1. 條數據從哪裡來

企業信息化產生了企業內部的數據，政府部門電子政務「金字工程」產生了各個部門獨立的數據，電子商務平台企業壟斷了自

身電子商務交易的數據，這些數據被限定在某個行業或領域，而且一般不對外公開，像是自己家的寶貝，生怕被別人「偷」去。

工商銀行是中國最大的商業銀行，擁有非常龐大的服務網絡和營業網點。其實，工商銀行也是一家「數據公司」，所擁有的數據資源非常龐大。從 20 世紀 90 年代開始，工商銀行實施的信息化工程經歷了幾個發展階段。第一個階段，銀行業務從手工處理轉變為個人台式計算機的電子化處理，銀行的各種賬務數據被精確地記錄下來，存儲在各個業務系統的數據庫中，但這些數據庫彼此是獨立的。第二個階段，銀行業務從個人電腦的電子化處理向網絡化、集中化系統處理轉變，銀行的客戶、產品、賬戶數據被整合到一起，數據被「大集中」，集中存儲在總行的數據中心，這為銀行內部數據的匯聚奠定了基礎。現在，工商銀行的數據化進入了第三個階段，銀行業務從網絡化處理向互聯網化、移動化和物聯網化處理轉變，客戶在網上銀行、電話銀行、手機銀行、微信銀行及銀行卡等各種新渠道所產生的數據匯聚到了總行數據中心，形成了工商銀行的大數據，但這些數據被工商銀行牢牢掌握著，仍然是企業內部的「條」數據，很難被其他企業和社會所共享。

電信運營商也是擁有大數據的「富豪」。過去，人們的通信方式主要是打電話，固定電話業務曾經成就了中國電信「老大哥」的地位，移動電話業務則讓中國移動成為全球最大的移動運營商之一。但隨著互聯網的發展，人們把更多的時間花在了上網看新聞、登騰訊 QQ（一款由騰訊公司開發的即時聊天軟件）聊天和刷微信朋友圈上，智能手機成了最方便的上網終端。隨著這些數據業務的

發展，電信運營商自然成為主要的數據流量通道。通過光纖寬帶網絡、移動寬帶網絡，三大運營商積澱了大量的互聯網數據，但運營商更多的是關心流量而不是流量背後的數據，彼此的數據很難共享，也很難向其他行業開放，同樣屬於典型的「條」數據。

政府擁有很多數據，權威性也較高，但是當前政府的數據也是「條」、「塊」分割的。國家部委的電子政務項目被稱為一個個「金字工程」，如國稅總局的金稅工程、海關總署的金關工程、公安部的金盾工程、原衛生部的金衛工程等，這些工程項目基本上是由各個部委獨立來承擔。所以，這些項目完工後就形成了各個部門內部的業務系統，也形成了部門內部的「條」數據資源。自上而下，部委的「條」數據也造成了地方政府的「條」數據。如公安部門信息化起步較早，建立了較為完整的全國人口數據庫，但這些人口數據很難被社保、計生、民政等部門共享，以致出現公安的人口數據與計生、民政的人口數據不一致的情況。很多地方政府在建設人口、法人等基礎數據庫和跨部門的數據交換與共享平台時，遭遇到了重重阻力，很多政府業務部門聲稱共享數據需要得到上級相關部委的同意，或者以數據安全和信息保密為由，拒絕提供部門內的數據給其他部門。這使得跨部門的數據交換與共享收效甚微。

互聯網企業與電子商務公司是大數據時代的佼佼者，但它們手中的數據很多也是「條」數據。阿里巴巴是全球貿易領域最大、最活躍的網上交易市場和商人社區，2014年其零售業務總交易額就超過2.3萬億元，承載了中國全社會消費品零售總額的8%以上。通過電子商務交易，阿里巴巴積累了大量產品、客戶和交易行為數

據。憑借這些數據資源，阿里巴巴豪邁地提出了向數據技術轉型的戰略，推出了淘寶指數、數據羅盤等數據產品。但阿里巴巴把這些數據當作自身的財富和核心競爭力，怎肯輕易開放給別人？說到底，這些數據仍然是阿里巴巴的「條」數據。

百度是全球最大的中文搜索引擎公司，每日處理來自中國大陸的搜索量多達50億次以上，這些搜索量來自包括PC（個人計算機）、手機等終端設備，內容涵蓋了政治、經濟、軍事、文化等方方面面，以至於形成了「有問題找度娘」的網絡現象。在這個過程中，百度積累了大量的網頁數據、搜索內容和搜索行為數據，每天百度後台服務器要處理的數據量達幾十PB（拍字節），相當於幾千萬部高清電影的數據量。為了挖掘和利用這些數據，百度專門成立了百度大數據部門，也推出了百度輿情、百度司南、百度精算、百度代言人等數據產品。但是，這些數據只有部分對外開放，還必須經過百度授權、通過專門的接口才能獲得。

再以新浪、搜狐、網易、騰訊四大微博的數據平台為例，四家公司的數據各自為政、相互獨立。相對封閉的數據環境使很多層面的具體分析都有很大的局限，要分析微博用戶行為也只能基於各自獨有的數據。比如針對重疊或交叉用戶，要弄清楚什麼特徵的人群會對什麼平台具有較高的忠誠度，什麼特徵的人會嘗試使用不同平台，同一個人在不同平台上的使用風格是否一致，在不同賬號下的活躍程度是否相同，影響因素又是什麼？如果局限在彼此分割的數據「條條」裡面，很多問題都是無法深入分析的。至少在商業領域，還是被條數據的「陰霾」所籠罩著！

2. 給條數據畫個像

綜上所述，條數據可以定義為在某個行業或領域呈鏈條狀串起來的數據。但這些數據被困在一個個孤立的條上，相互之間卻不能連接起來。



圖 1-2 條數據

這個定義可能比較抽象，通過對條數據特徵的描述，能夠讓「他」更加具象化。概括來講，條數據有四個基本特徵：

(1) 領域單一。由於產生於某一個特定行業領域，條數據信息量比較單一，更多的是就醫療談醫療、就教育談教育，局限在某一小圈子裡，視野相對不開闊，對事物的預測、對規律的把握準確性還有進一步提高的空間。

(2) 數據封閉。條數據更多地被少數企業、行業或者部門獨自佔有，不共享、不開放。例如，淘寶網內部掌握了淘寶網的客戶購物數據、京東商城掌握了京東商城的客戶購物數據、一號店掌握了一號店的客戶購物數據，但這些數據之間彼此是封閉的，想獲得這些客戶購物數據的完整內容，就變得比較困難。

(3) 數據壟斷。在條數據時代，數據被視為私有的資源或者資產，數據本身構成了企業或者某個部門的核心競爭力。所以，大家都在不停地獲取數據、爭搶數據、壟斷數據。在互聯網公司向大數據公司轉型的過程中，一個個數據托拉斯（商業信托）也隨之誕生了。

(4) 源自事務流。由於條數據被視為私有的，「他」是與企業的營銷、生產等商業活動緊密關聯的，企業或者組織更加重視採集生產經營活動中的事務流數據，比如產品、客戶、銷售額等數據，往往忽視了與人關聯的行為數據、人文數據和社會活動的數據。

3. 條數據改變了什麼

行業數據的累積和分析已經促進了自身的變革，從行業數據中我們能夠全面清晰地掌握行業整體狀況、業界最新動態，使我們對事物的精準性預判、降低生產生活成本等工作都提高到一個新的層次上。

企業資源規劃（ERP）系統是企業信息化的核心系統，它能夠實現企業內部業務與管理流程的信息化。通過分析業務處理流程中產生的產品、物料、客戶等企業各類數據，企業可以以銷量定產量，以產量定物料需求量以及成品庫存量，最大限度地縮短採購週期，提高庫存周轉率，提高客戶滿意度。

金盾工程是較早啟動的電子政務金字工程。經過十幾年的建設和應用，公安系統的信息化水平得到顯著提升，極大地方便了民警戶籍辦理、偵查辦案等工作的開展，提高了保障人民生命和財產

安全的能力。舉個例子，通過公安系統建立的人口數據庫系統，輸入嫌疑人的身份證號或者一張照片，就可以通過全國聯網快速搜索到嫌疑人的詳細數據，極大提高了辦案效率。

淘寶網是全國最大的網絡零售平台，提供了豐富的商品和廠家數據，百姓日常生活所需要的各種消費品幾乎都能在淘寶網上找到賣家；淘寶網也提供了每個廠家的產品銷量數據，讓用戶能隨時掌握產品的暢銷程度；淘寶網上的客戶評論數據則說明了廠商的信用和商品的客戶滿意度，讓用戶建立起對商品的信心。在這些數據的支撐下，百姓在淘寶網的購物更加便捷、放心。

4. 條數據成長的煩惱

但是，條數據具有很多天然的缺陷，這使得條數據的作用受到很大限制，產生了很多問題。

(1) 數據孤島。條數據難以在更大範圍進行數據交換、共享，使得數據和信息系統成為一個個孤島。數據孤島成了這個階段的普遍問題，企業內各環節存在著數據孤島，企業間存在數據孤島，政府機關之間也存在數據孤島。在很多地方，有多少個委、辦、局就有多少個信息系統，每個信息系統都由部門自己的信息中心管著，有自己的數據庫、自己的操作系統、自己開發的應用軟件和用戶界面，完全是一個個獨立的小島，造成需要多次重複輸入數據、存在大量的垃圾數據、無法保證數據的一致性等困難。在相當一部分企業內部，業務功能交互與信息共享往往不夠，使企業的物流、資金流和信息流脫節，造成賬賬不符、賬物不符，不僅給正常的財務核算帶來難題，而且很難及時發現問題、有效監控業務過程，極

易滋生計劃失控、庫存過剩、採購與銷售環節的暗箱操作等現象，導致無效勞動、資源浪費和效益流失等嚴重後果。更重要的是，在條數據狀態下，企業缺乏數據二次發掘的能力，也缺乏按業務條線要素自動檢索、分析行為規律的本領。無論是政府機關，還是企業內外，孤立的信息系統無法有效地提供跨部門、跨系統的綜合性信息，一個個的數據孤島更無法使各類數據實現其潛在價值，局部的信息、割裂的數據也難以提升為管理知識，為決策提供支持只能流於空談。

(2) 數據資產壟斷。由於條數據的單維度性和封閉性，數據資產被少數企業所壟斷，難以開放和流動，資產價值難以發揮，這不僅浪費了數據資源，而且會阻礙商業、社會治理和服務民生等領域的創新。

這裡，還涉及數據的法（定）權利問題，比如，數據採集是否有支付對價，數據安全、授權、合理補償的責任與義務何在，以及法律約定等問題。目前存在著明顯的法權不對等、不公平現象。

大家知道，數據本身往往蘊含著巨大的商業價值，哪怕是對小數據、條數據的初步分析，也能夠從中發現商機、改進管理、提升服務，正因為數據的這種價值，行業、企業和政府都在竭盡力量採集數據、佔有數據和利用數據。這其中很大一部分數據被大家認為是公共數據（當然也可以理解為屬於某家網站，只是網站都按照行業規則把數據開放了出來），當然也有很多數據是企業內部服務器上的日誌數據，還有各種交易數據，這部分數據企業一直沒有開放，被認為是屬於企業的。政府和企業也有很多線下採集的數據，

出於各自利益的考慮，這部分數據被很有限地開放，大部分沒有開放。政府部門說，為了安全起見，這些數據只屬於我，阿里巴巴說淘寶上的數據是我的，新浪也說微博數據是我的……於是各個企業和組織挖空心思採集數據、獨佔數據，哪怕重複工作，造成巨大浪費。BAT 甚至四面出擊，希望將天下所有的數據裝入自己的口袋中，但靠自身的力量，這做得到嗎？

對數據的肆意侵佔，產生了一個更為嚴重的問題！數據割裂和數據壟斷出現了。今天的大數據是怎麼來的？主要來自互聯網。為什麼有互聯網？因為數據需要在不同計算機、不同用戶、不同國家之間進行交流和共享。數據本來就在那裡，就應該在一個共享開放的互聯網大池子裡，應該是屬於全球互聯網的，也是屬於大家的，沒有必要每家互聯網企業再去複製互聯網中的一個個數據。但出於商業或利益或其他層面的考慮，互聯網中的數據被圈了起來，貼上了私人財產的標籤。於是，如今的大數據被各類商業機構佔有、控制了，成了這些機構的資源、資產和核心競爭力，成了它們手中的寶貝，不肯輕易共享，也不肯開放。這樣的趨勢延續下去，大數據就被大大小小的數據商壟斷著、割裂著，成了一個個新的「數據孤島」，與開放、連接和共享的理念背道而馳，勢必讓大數據的價值大打折扣，甚至可能因此被毀掉。

目前看，大數據壟斷所造成的大數據資源的貧富差距太大，大數據的交易還是「富人」之間的遊戲，普通草根階層如中小企業只能享受微不足道的數據紅利。大數據領域的很多創新創業企業也因為缺乏數據，創新創業的積極性被極大打擊。長此以往，大數據

鴻溝會越來越深！

(3) 數據預測失真。在數據預測領域有很多「井底之蛙」，過錯可能並不在「青蛙」，實在是「井」太深了，這隻「青蛙」能掌握的就是有限領域內的有限數據，根據這些數據做出的「科學」預測也就失去了科學性，以偏概全，甚至出現重大偏差。例如，僅僅考慮工業部門的產能增長數據，很容易預測經濟發展持續增長的趨勢，但如果忽略了環保部門的環境數據、人力資源和社會保障部門的勞動力數據，就可能造成數據預測失真。

(二) 塊數據——坐擁天時和地利的寵兒

1. 突破條數據瓶頸的一些探索

過去幾年，以 Google 為代表的企業，大力發展基於空間信息系統的大數據，其目的是通過地圖信息的載體把各種條數據聯在一起，從某種意義上來講，它是塊數據的一個初步嘗試，但是由於數據量和開放性等原因，仍然具有很大的局限性。解決問題的根本出路，是真正意義上塊數據的產生。塊數據的產生，打破了傳統信息的不對稱和物理區域、行業領域對信息流動的限制，通過對不同類型、來源信息的集成、挖掘、清洗，極大地改變了信息的生產、傳播、加工和組織方式，進而給各個行業的創新發展帶來新的驅動力，推動各類產業徹底變革和再造。人類將以塊數據為標誌，真正步入大數據時代。以往很多數據價值的外部性難題，將可以通過塊數據以市場化和行政化手段得到解決。

2. 塊數據的概念

所謂塊數據，就是以一個物理空間或行政區域形成的涉及人、事、物的各類數據的總和。（塊數據不只是「條集合」，更是「條集構」。）

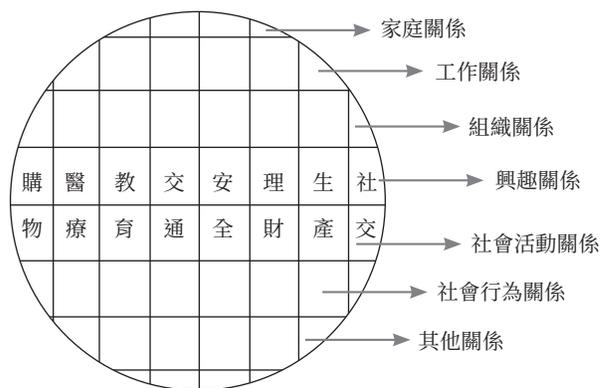


圖 1-3 「人」的塊數據

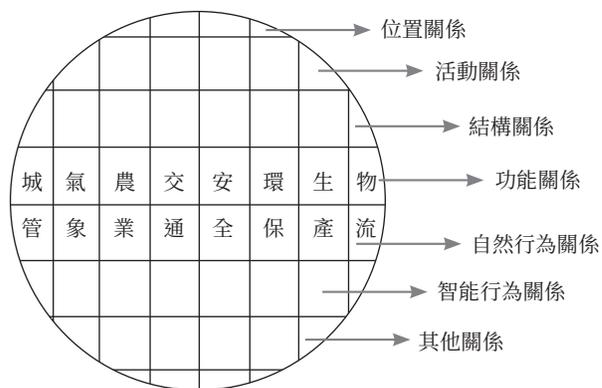


圖 1-4 「物」的塊數據

人的數據、物的數據，以及人與物、物與物的關係數據，構成了各種活動和事件的數據。

3. 塊數據的形成

開放、共享、連接是塊數據形成的基本機制，在一個個「塊」內形成開放、共享、連接的塊數據，再把一個個塊數據開放、共享、連接起來，就會產生更大的塊數據網狀結構。

例如，在社區，通過智慧社區建設，建立起社區居民、家庭和居委會的數據採集和社交網絡平台，它能夠採集居民、家庭、樓宇的基本數據，也能夠匯集物業、民政、醫療、養老、公安、文化、教育等各個領域的數據，還能夠聚合社區金融、超市、農產品供應等商務數據，就會形成一個共享、開放的社區塊數據。

在城市中，通過鋪設高速光纖寬帶網絡，搭建高速移動寬帶基站，建設無縫銜接的 Wi-Fi 接入節點，將使得整個城市的網絡「高速公路」更加暢通，數據將被更加快捷地傳輸和連接，這會有力推動各類社會企業和個人動態數據的「塊」上集聚。通過建設城市基礎數據庫、數據交換共享平台和跨部門協同應用，能夠將人口、法人、地理信息、宏觀經濟等數據開放、共享並連接起來；通過公共數據平台和跨部門協同應用系統，能逐步將各個部門的數據匯集起來，進一步整合城市網絡平台上採集的企業和個人的行為數據，將逐步形成城市塊。

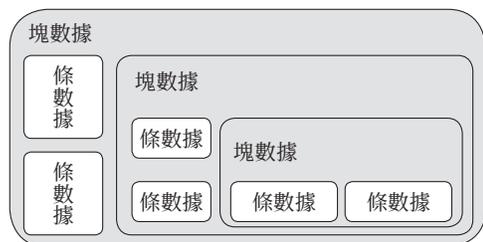


圖 1-5 條數據與塊數據

4. 條數據的融合

塊數據就像一個計算機的主板，它建立起了一個開放、共享、連接的數據基地，而各個行業和部門的條數據就像一個個可插可拔的板卡，它們只有融合和集成到主板上，才能發揮數據資產真正的功效。加快各個行業雲的應用示範，通過跨部門、跨行業的政務雲、民生雲、產業雲、環保雲和文化雲等共享工程示範，建設起行業應用雲平台，能推動計算、應用和數據資源從「條」到「塊」上的融合與集聚。

正如學科交叉通過跨學科知識匯聚獲得學科創新的新優勢一樣，條塊融合也給予了數據匯聚優勢，在匯聚而成的數據海洋背後，那些被發現的規律和知識就是「黃金」。

(三) 「塊」思辨——塊數據哲學體系初步構建

1. 對塊數據幾個範疇的理解

(1) 塊集合。如前所述：塊數據是以一個物理空間或行政區

域形成的涉及人、事、物的各類數據的總和。首先，塊數據是數據集合，包括：點數據（單點數據、多點數據）、條數據（單條數據、多條數據）、面數據（單一層面、多個層面），只是在大數據時代及其既有技術環境下，面數據與塊數據同步發展成形，塊數據是點線面數據的總和及組構。

其次，塊數據在集合過程中，既有數據空間的填充，也有空間數據的重構；既有集合過程中的組構，也有組構過程中的集合；既有新數據的匯集，也有基於原有數據組合後的衍生數據。

再次，塊數據在塊組團過程中是空間網狀，而不是簡單堆砌，具有明顯的網線、節點、脈絡，及其自身內在的邏輯運行規律。類似一個物理學的「電子雲」，或「能級層」。

(2) 塊規模。對有關塊規模的思考，按照本書研究的領域，更多是從實踐角度考慮其現實可行性。塊數據的重點是關注一個人類活動相對集中的區域，及其相應的行為數據。主要是考慮人類相對較大聚落密度時的數據特徵，也可以理解為，相對較為關注空間、密度及複雜行為。當然，這並不意味著塊數據規模就一定是一個超級大都市，但是，如果我們僅說地球村、大經濟體，在當前的技術環境下，也可能會對塊數據的操作性能有較大影響。我們也思考了「塊」的時間序列、數據的域、多個條數據的權重、數據組合的膨化、數據衍生的變異等涉及塊數據規模的問題。對更大的「塊」，建議讀者考慮「塊數據組團」、「超級塊數據鏈接」、「塊篩選」、「塊流動」。對於更細的「塊」，現有數據思維及技術已無障礙。

(3) 塊邊界。這是一個具有理論和現實雙重意義的難題。首

先，可以稱之為一個相對獨立的塊數據的前提是：這個塊數據要具有「塊」定義域內清晰的宗旨和功能；其次，「塊」的邊界可以是相對富有彈性且動態的，因而是智慧的；再次，「塊」間聯繫，是基於「塊」中心外法線方向，持續與其他「塊」連接並催化有價值的數據流，這些數據流有些是一次數據流——原生數據交換，有些是二次數據流——經過加工，或者經過不同「塊」優選整合後的數據集合。

2. 塊數據的辯證思維

一是，「**條條關係入塊**」。如何理解條條關係，我們的觀點是：條條數據是帶著「相互關係」、「相互關聯」入塊的，而不僅僅是條數據獨立入塊。換句話說，入塊條數據及其人類行為的掌控者，需要具有善意溝通的理念，讓渡一定的自我主權，遵守塊數據的法律及規則，踐行塊數據道德操守。

二是，「**條塊關係組構**」。條數據在塊數據中既是基本要素，又具有能動性。條在塊中可以科學地、智慧地尋找自己的位置和角色，同時，也在塊中科學地、智慧地響應其他條的需求和約束，遵守塊的統一規則。

三是，「**塊塊關係互聯**」。塊數據在塊間的交換，既可以是條狀態連接，更應當是加工、選擇、組合、衍生後的塊數據組團交換，在塊內外形成高效的、優化的數據掃描與選擇性交換，從而真正達到動態地、全面地、聯繫地完成數據識別與價值交換。「塊」間的緊密程度，既取決於塊數據的流量、質量、黏性，更取決於塊數據的價值、時效、導向。

3. 塊數據對人類進步的影響路徑

塊數據對人類進步的影響大致將遵循如下路徑：人類行為記錄——數據存儲與處理——塊數據影響人類行為——新的行為記錄。在這一過程中，人類將在「數據塊」的框架下，形成人類行為與數據的交互影響及人類自身的進步歷程。畢竟，人類可以是數據的，但本質上，數據是人類的。可以預期，不遠的未來，我們將擁有如下基於塊數據的「塊」：

「塊社會」將會以塊數據為並行的虛擬社會，形成人類更加智慧、理性、快樂、和諧的法治社會主義。

「塊生活」使百姓在「塊」域中借助塊數據，在標準化和技術化的基礎上，選擇更加個性化的生活方式。

「塊產業」將喚醒並催生出一批傑出、睿智的企業家，他們既可以成為對塊數據選擇性應用的惠及者，也可以成為對「塊」域新型數據產業及傳統產業的貢獻者。

「塊富裕」是塊數據從必然王國走向自由王國的重要目標之一，人類潛在的互聯互通精神與時下熱燥的互聯致富思維，最大的區別就在於，真正的「互聯」不是以殘酷競爭催生首富、極富為目的，而是以「眾入」、「眾籌」、「眾享」、「眾富」為目標，當然，在人類社會進步史上，這也是對「均富」的揚棄。

「塊文化」則是基於塊數據應用的新型文化模式，人們對歷史是選擇性傳承而不是割裂，將來，幾歲的孩子都知道「改寫歷史」與「篡改歷史」意思的區別；人們對思想是價值型交換而不是灌輸；人們對藝術是參與性享受而不是膜拜；人們對學習是社會認可的選

擇性偏好而不是「惡補」。塊數據第一次使人們不再擔憂「機器人」掘墓，第一次不再猶豫「個性化」缺失。再也不會有人相信「絕對權威」，但是人人願意「服從社會」、「共享文化」。

「塊哲學」的到來是一個客觀的歷史進程，是人類思想與智慧在大數據條件下的自然昇華。塊數據將極大地改變人類的精神境界和思考模式，我們相信基於塊數據的思考與實踐也將呼喚新的哲學框架與思維工具。該感謝誰？我們不知道也不需要知道，我們需要的是儘快行動起來，迎接塊數據的挑戰並享用塊數據的智慧。

4. 塊數據的三大自我智慧

塊數據時代的到來，既是一個難得的機遇，也是一個複雜的挑戰。我們覺得，塊數據時代，至少需要擁有三大自我智慧。

一是**邊界識別技術**。在設計階段就應當關注塊數據的邊界識別技術，這將有助於提高「塊」內效率，也將有助於優化「塊」間價值交換，還將有助於實時識別並懲罰惡意攻擊。

二是**條間避讓智慧**。在龐雜的塊數據網格內，當某些時候有些條數據有礙於「塊」的宗旨和主要功能時，塊數據應當具有系統的自動避讓功能。塊效率是為了解決條功能和條效率的問題，塊數據也要解決塊的自效率問題。

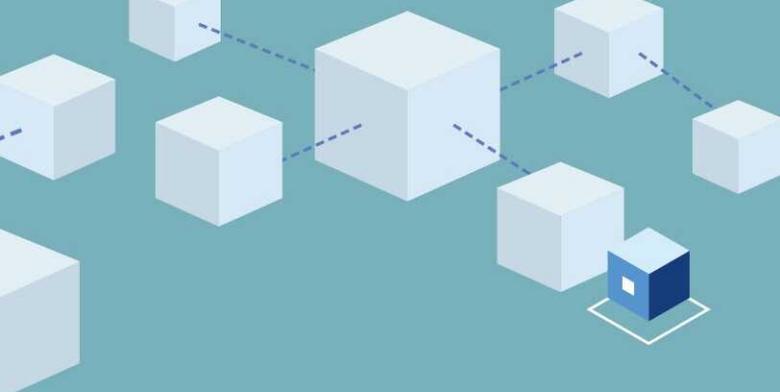
三是**塊的重構與自我修復能力**。塊數據的基因就包含對條數據的組合、糾偏、選擇。因此，「塊」的重構與自我修復能力，也是塊數據自我糾偏的智慧和執行力所在。塊數據本身就應當是一個高度自適應網系。

第二章 塊數據素描——「金礦」是怎麼煉成的

點數據是離散的孤立數據，條數據是單維度下的數據集，塊數據則是點數據和條數據的「塊」上集聚。一般來講，塊數據比條數據的4V即Volume（大量）、Variety（多樣）、Value（價值）、Velocity（高速）的特徵更為明顯。塊數據如同一個計算機的主板，建立起了一個開放、共享的數據基地。各個行業和部門的條數據如同一個個可插拔的板卡，它們只有跨界、融合並集成到「主板」上，才能發揮數據資產真正的價值。在這種融合的過程中，塊數據打破了點數據、條數據中存在的數據分散、數據孤島和數據壟斷等問題的限制，為大數據發展找到了新的突破口。

（一）高度關聯性：「你」「我」「他」與「們」

數據的關聯度越高，信息量越大，價值也就越大。塊數據的



塊數據強調的是數據、算法、場景融合應用的價值體系，是大數據時代的價值觀和方法論。數據、算法、場景是治理科技的三大核心要素。

——日本《日中商報》

共享是新一輪科技革命和產業變革的關鍵力量，借助共享，人類文明必將走向更高階段。可以說，基於數據力與數據關係提出的「共享價值理論」，是繼剩餘價值理論之後頗具革命性的理論。

——南非《非洲時報》

互聯網帶來了超越空間的資訊傳遞、共享與價值交換、增值，卻面臨著從無界、無價、無序走向有界、有價、有序的挑戰。《塊數據 3.0》提出的主權區塊鏈，為從信息互聯網、價值互聯網向秩序互聯網的演進帶來了暢想空間。

——英國《華商報》

面對超數據時代帶來的數據擁堵難題，《塊數據 4.0》運用激活數據學有效挖掘、分析大數據背後人的運行規律，把超數據從「厚」做到「薄」，從「大」做到「小」，為大數據時代清除認知障礙，平衡利益矛盾，讓不確定性對抗確定性成為可能。

——西班牙《僑聲報》

《塊數據 5.0》圍繞構建以人為原點的數據社會學理論與方法，創新性地提出數據進化論、數據資本論、數據博弈論的理論體系，研究和探索人與技術、人與經濟、人與社會的內在關係，以此分析人的行為、把握人的規律、預測人的未來。

——美國《中美郵報》

ISBN 978-988-8568-72-7



9 789888 568727 >



專業出版 國際銷售

紅出版文化平台

加入我們：www.red-publish.com

ModE.

上架建議：資訊管理

定價：港幣 148 元正 / 新台幣 596 圓正