

大数据智能区块链 DAB 项目 白皮书

2017 年 8 月

1	大数据时代背景	4
1.1	大数据.....	4
1.2	区块链.....	4
1.3	人工智能.....	5
2	中国实体经济的供给侧结构改革	6
3	大数据智能链==推动中小企业成长的区块链	8
4	大数据智能链构建企业成长新生态共同体	9
4.1	传统企业的成长困境.....	9
4.2	信息化技术应用的困境.....	9
4.3	深度理解供给侧结构改革，重塑企业成长新路径.....	10
4.4	千亿市场的蓝海.....	11
5	大数据智能链，构建企业成长新生态	12
5.1	大数据智能链五链模型.....	12
5.2	区块链构建产业互联互通.....	12
5.3	助力倾听市场声音.....	13
5.4	改善产业链效率.....	13
5.5	推动产业供给侧结构改革.....	14
6	大数据智能链体系	15
6.1	大数据智能链架构.....	15
6.2	大数据智能链技术特点.....	15
6.2.1	区块链技术特点.....	15

6.2.2	大数据技术特点.....	19
6.2.3	人工智能技术特点.....	22
6.3	大数据智能链业务架构.....	23
6.3.1	平台助力产业融资.....	23
6.3.2	区块链电商平台推动企业转型.....	23
6.3.3	B2B 平台改变产业链结构.....	24
6.3.4	供应链管理改善产业链效率.....	24
6.3.5	大数据构建企业数字神经系统.....	25
6.3.6	企业大脑推动供给侧结构改革.....	25
7	执行团队.....	26
7.1	核心管理团队.....	26
7.2	专家团队.....	29

1 大数据时代背景

1.1 大数据

2003 年，Google 公司为了解决其搜索引擎中大规模 Web 网页数据的处理，研究发明了一套称为 MapReduce 的大规模数据并行处理技术；

2006 年，大数据技术的标志性产品成为一套完整独立的软件，起名为 Hadoop；

2010 年 9 月— Hive(Facebook) 脱离 Hadoop，成为 Apache 顶级项目。

2010 年 9 月— Pig 脱离 Hadoop，成为 Apache 顶级项目。

2011 年 1 月— ZooKeeper 脱离 Hadoop，成为 Apache 顶级项目。

2013 年 10 月 YARN 已经发展出 Hadoop2.2.0 稳定版。

十年磨一剑，大数据技术终于发展成熟稳定。

2015 年 3 月，阿里巴巴董事局主席马云，我们正在从 IT 走向 DT 时代；

2015 年 5 月开始，贵阳举办“大数据博览会”，至今连续举办了三届，“大数据”已成为贵阳的一张名片。

2015 年 8 月 31 日，国务院以国发〔2015〕50 号印发《促进大数据发展行动纲要》。

1.2 区块链

2014 年，在比特币纷争中，做为比特币底层技术的区块链得到了业界、学界和传统机构的关注：“比特币的最大价值并非数字货币本身，而是其背后的分布式记账方式——区块链。”区块链技术的出现，解决了互联网 TCP / IP 协议所不能解决的信息传递的真实性以及在互联网上进行价值存储和传输的问题。

《经济学人》将区块链描述为商业社会的“信任的机器”。而“信任”二字就是区块链技术的最大价值所在,也是区块链最具有颠覆意义的功能之一——改变信用产生机制。通过区块链，交易双方可在无需借助第三方信用中介的条件下开展经济活动，从而降低资产能够全球范围内转移的成本。

这种全新的底层协议的构建模式，将推动互联网从信息互联网向价值互联网到秩序互联网迭代升级，将会通过互联网的应用逐渐传导至社会经济生活的各个环节，带来众多领域的创新，甚至会颠覆和重塑很多现有商业模式、行业运行和治理体系。

党中央国务院高度重视信息化工程，先后印发《国务院关于积极推进“互联网+”行动

的指导意见》、《促进大数据发展行动纲要》。在国务院印发的《“十三五”国家信息化规划》中将区块链列入国家信息化规划，并将其定为战略性前沿技术。

2016年8月，工信部在北京组织召开“区块链技术和产业发展论坛筹备会暨白皮书编写启动会”，宣布开启制定行业标准，这意味着官方对区块链技术的正式接纳。

新加坡金融管理局（MAS）在对外公布的信息中称新加坡将在私有以太坊区块链上发行本国货币的数字代币。瑞士联邦委员会（Swiss Federal Council）正在着手为金融科技制定法律框架。在金融机构方面，英国巴克莱银行正在与金融科技创新公司和金融行为监管局（FCA）合作开展专注于区块链的项目，瑞士土地登记机构 Lantmaterie 正式开始利用区块链技术进行土地和房地产登记。自2017年初以来，多个国家，包括巴西，都已经开始利用区块链技术在网络上推动土地和房地产所有权。

1.3 人工智能

2016年3月，AlphaGo 的横空出世，让整个产业的目光又关注到了人工智能。在随后的时间中，各大 IT 巨头纷纷推荐人工智能战略，积极招揽人工智能的顶尖人才。

谷歌收编了 DeepMind，推出了机器学习开源框架 TensorFlow；

微软转型人工智能，推出了机器学习开源框架 CNTK；

百度从微软挖来了陆奇，推出 Apollo “(阿波罗)” 平台；

阿里推出了 PAI2.0 机器学习平台；

腾讯推出 Angel 第三代机器学习平台；

2017年7月8日，国务院以国发〔2017〕35号印发《新一代人工智能发展规划》。

2 中国实体经济的供给侧结构性改革



十年间，中国经济从 2.75 万亿美元，成长为世界第二大经济体，规模达到 11 万亿美元。随着中国经济的增长，一大批企业已随之成长起来。

2015 年 11 月 10 日，习近平主持召开中央财经领导小组第十一次会议提出，在适度扩大总需求的同时，着力加强供给侧结构性改革。

2016 年 3 月 5 日，政府工作报告中，再次出现“供给侧结构性改革”。

2016 年 5 月 25 日，国务院总理李克强在贵阳出席中国大数据产业峰会暨中国电子商务创新发展峰会开幕式并发表致辞。

李克强在致辞中说，当今世界，信息化浪潮席卷全球，大数据、云计算、物联网等蓬勃发展，使互联网时代迈上一个新台阶。今天的中国要把握住世界科技革命的历史机遇，按照创新、协调、绿色、开放、共享的发展理念，加快创新型国家建设，着力推进结构性改革尤其是供给侧结构性改革，推动信息化与实体经济深度融合发展，不断提升全要素生产率，努力保持经济中高速增长、迈向中高端水平。

李克强表示，要推动新经济发展和传统产业转型升级，必须坚持改革开放，使中国经济发展方式从过度依赖自然资源转向更多依靠人力资源开发。自然资源是有限的，但人力资源是无限的，蕴藏着无穷的创造力，是实现可持续发展的最大资源。我们坚持创新驱动发展战略，制定实施网络强国战略、国家大数据战略、“互联网+”行动计划、《中国制造 2025》等，出台了一系列重大举措，在促进创新和更好发挥人力资源优势方面的效应正在显现。以大数据为代表的创新意识和传统产业长期孕育的工匠精神相结合，使新旧动能融合发展，并带动改造和提升传统产业，有力推动虚拟世界和现实世界融合发展，打造中国经济发展的“双引

擎”。

2016年6月27日，国务院总理李克强出席第十届夏季达沃斯论坛开幕式时表示，中国经济正处于新旧动能接续转换、经济转型升级的关键时期。我们将坚持发展第一要务，坚持稳中求进工作总基调，落实创新、协调、绿色、开放、共享的发展理念，在适度扩大总需求的同时，坚定不移地推进供给侧结构性改革，抓好去产能、去库存、去杠杆、降成本、补短板，推动发展从过度依赖自然资源向更多依靠人力资源和创新驱动转变，使中国经济持续保持中高速增长，迈向中高端水平。

企业与信息化的融合，已经进行了多年，取得丰硕的成果。

在经济结构转型时期，在信息技术飞跃发展的时代，实体经济需要什么样的信息技术，我们认为有三个回答：

实体经济需要跟上时代的信息技术

实体经济需要与成长匹配的信息技术

实体经济需要低成本的信息技术。

总结起来，实体经济需要的是能够推动企业成长的信息技术。

3 大数据智能链==推动中小企业成长的区块链

当我们把目光放在经济的海洋中，只能发现了一个个的巨无霸企业；在时间的长河中，我们能看到无数企业的成长和兴衰。透过企业的层层关联，我们看到产业的兴旺发展和产业链的分工与整合。

著名供应链管理专家马丁·克里斯多弗曾说：“市场上只有供应链没有企业”，“真正的竞争不是企业与企业之间的竞争，而是供应链和供应链之间的竞争”。

市值全球第一的苹果公司，在 iPhone 手机出来之前，是电脑提供商。现在提起苹果产业链，大家通常都会指的是苹果手机产业链。

产业链上核心企业的成长，会带动整个产业链的发展；产业链上分工协作的企业成长，也推动产业链提高竞争力。

供应链治理机制是供应链治理的核心，信任治理机制是供应链中的基本治理机制。产业链之间协同的关键在于产业链各环节的相互信任。

基于区块链技术，我们创新的提出，建立服务于整个产业链的大数据智能链，推动产业和企业成长。

4 大数据智能链构建企业成长新生态共同体

4.1 传统企业的成长困境

随着企业和行业的发展，企业的业务规模在扩大，企业的业务种类在增加，企业的各类合作伙伴在增加。业务规模的扩大，意味着企业要处理的数据规模也相应的增大；企业业务种类和合作伙伴的增加，意味着企业内部的协作、企业之间的协作也相应的变得更加复杂。

这些行业发展要求大数据、人工智能、区块链等信息化技术，帮助企业更快、更加智能、更加有效的处理数据和增强协作；也要求这些信息化技术能够整合形成新的信息技术基础设施，直接提供产品给客户使用。

大数据、人工智能、区块链等新信息技术，在应用上还处于比较初期的阶段，成本巨大，价格昂贵。在前期，只有少量的大型企业和政府才有能力支撑和使用，因此，我们看到的大数据应用、人工智能应用、区块链应用，都普遍存在于政府、金融、电信、互联网巨头这些大型企业。

国内大部分的中小型企业，其年销售额在 1 亿至 10 多亿规模，这样规模的企业发展，是需要信息化技术的支撑，但又不能在这部分投入成本过大，每年能有销售额 5% 左右的信息技术投入，已经是一笔不少的投入了。

因此，这些中小型企业，既希望能够应用到新的信息化技术，同时又希望不要花费大多数的成本，还可以有定制开发适合自身发展的信息化系统。

年销售额过亿到十多亿的公司，围绕自身的业务发展，已经有不少的业务合作伙伴，形成了一条成规模的产业链。但公司限于规模和发展，其信息化程度通常只能满足自身业务的发展，难以对产业链进行扶制，也制约了企业的进一步发展。

服务于全产业链的信息化系统，能够帮助中型企业在信息上对产业链上的公司进行支持，并且基于信息化的支持和产业链业务数据的收集，可以进一步发展产业链金融服务，帮助整个产业链做大做强。

4.2 信息化技术应用的困境

区块链技术在应用过程中，存在几个痛点。

其一，区块链技术的高昂成本。目前一条区块链的搭建成本在百万元以上，并且，基于区块链的应用，还需要传统的数据库、大数据等技术一起，才能完成整个应用的技术基础设施搭建，此外，还需要定制开发应用软件。基于区块链的应用软件开发，至少是几百万，甚

至是上千万元的支出。这样的高昂的成本，使得基于区块链的技术应用，目前只能是大型企业机构才能使用。

其二，区块链技术的应用开发要求过高。目前，区块链技术属于单独发展阶段，区块链技术与传统数据库技术、大数据技术、人工智能等技术并没有进行整合，尚未形成统一的技术架构来支持技术人员进行开发。应用软件开发人员在开发应用时，需要掌握众多的技术能力，形成复杂的技术架构来支持业务应用。导致应用开发的周期长、成本高、技术架构复杂、稳定性差，难以按期投入使用，严重影响项目的效益。

区块链的技术发展，要求区块链技术与传统数据库技术、大数据技术、人工智能技术进行紧密配合，形成一套整体的技术架构和技术基础设施，降低系统架构的复杂度，降低软件开发成本，缩短应用软件开发周期，提高系统性能。

人工智能的应用，同样存在成本巨大的问题。人工智能热，首先是炒热的人工智能的人才，刚毕业的研究生，年薪就在 50 万元起；人工智能的机器学习平台产品，价格甚至有千万元之巨；人工智能的开发服务，起步就是 20 万/人月。

人工智能的产品，大多倾向于智能硬件、机器人等，真正能够利用企业数据，做出人工智能的应用少之又少。其中既有成本原因，也有数据的原因。一般的初创人工智能企业，还在研究技术，来不及也接触不到企业的业务数据，更谈不上为企业提供人工智能的产品和业务解决方案。

大部分的中小企业，其基本的业务信息化建设尚未完整；大数据平台的建设和整合，才刚刚提上规划。企业的应用开始与信息化技术的发展脱离，企业的成长速度开始落后于信息技术的发展速度。

4.3 深度理解供给侧结构改革，重塑企业成长新路径

供给侧结构性改革，就是要用改革的办法矫正供需结构错配和要素配置扭曲，解决有效供给不适应市场需求变化，使供需在更高水平实现新的平衡；就是要扩大有效供给，提高供给结构适应性和灵活性，改善供给品质，创造新供给、培育新需求；就是要发挥市场配置资源的决定性作用，盘活过剩产能沉淀的劳动力、资本、土地等生产要素，让生产要素从低效率领域转移到高效率领域，从已经过剩的产业转移到有市场需求的产业，进而实现资源优化再配置，提高全要素生产率。

站在产业链角度分析，提供给市场的产品和服务，是供给侧的成果；供给侧结构性改革，

就是要改革产品的产业链体系，将以往产业链简单的供应链体系，整合形成支持产业发展和企业成长的新型产业供需链、产业产品链、产业技术链、产业数据链、产业金融链。

要让整个产业链，倾听终端市场的声音，让终端市场的需求，真实有效的传递到产业链的每个环节；

要让整个产业链，针对市场需求，作出统一有效的反应和反馈，提供出符合市场需求和趋势的产品和服务；

要让整个产业链的运作，有高效和高质量的。

这一切，都依托服务整个产业链的大数据智能链的支撑和服务。

4.4 千亿市场的蓝海

2015 年末，全国规模以上中小工业企业 36.5 万家。即主营业务收入在 2000 万以上的企业有 36.5 万家。我们的目标客户是年收在 1 亿至 20 亿之间的企业，在目标范围内的企业家数，至少是一万家以上。

以产业链为目标的信息化系统是一片蓝海，以每家企业构建系统为业务收入的 5% 计算，每年服务费为 2% 计算。整个市场规模在 1000 亿以上。

这片蓝海市场，足以支撑培育起数家百亿规模的企业。

5 大数据智能链，构建企业成长新生态

5.1 大数据智能链五链模型



供需链：改变产业链的供需结构

产品链：改善产业链的产品构成和产品结构

技术链：改变产业链应用技术的范围和群体

数据链：增加产业链的数据和扩大产业链的数据共享对象

金融链：改善产业链的融资结构，扩大产业链的金融服务对象。

5.2 区块链构建产业互联互通

2017年5月24日，由中信出版集团出版、大数据战略重点实验室最新理论研究成果《块数据 3.0：秩序互联网与主权区块链》一书，在“2017 中国国际大数据产业博览会”新闻发布会上首发。

《块数据 3.0》探索了互联网的发展演变，提出区块链的发展和應用將推动互联网实现由信息互联网向价值互联网和秩序互联网的华丽转型。互联网的发展是从信息互联网、价值互联网到秩序互联网的梯度跃升。信息互联网让人看到了区块链对于便利人与人沟通、减少信息不对称的价值。价值互联网让人看到了区块链对于物质和服务增值、数据资产增值、社会价值体系重构的潜力。而秩序互联网让人们看到了借由区块链等技术手段创新社会组织方式、治理体系、运行规则的前景。

《块数据 3.0》围绕互联网治理这一问题，指出区块链将构建有序的互联网世界，推动秩序互联网的到来。区块链加快了互联网从低级向高级形态的演进，改变了互联网的游戏规

则，为互联网治理提出了解决方案。将区块链拓展到互联网治理中，有助于形成一个跨企业、跨组织、跨个体的从事经济社会活动的信任体系，构建一种“主观为人、客观为己”的社会价值形态，推动构建基于价值共享、规则共识和行为共治的信息文明新秩序，即秩序互联网。描绘了由区块链等技术手段创新社会组织方式、治理体系、运行规则的前景。

《块数据 3.0》指出基于信任的契约精神是秩序互联网的基础。信任是网络社会有序运行的基石和润滑剂。虚拟社会是一个高度开放和自由的社会，它既给不同交往主体提供了良好的交流平台，同时也由于其活动的虚拟性、信息传递速度的迅捷性、后果的模糊性和过程的时空分离性，放大和加深了信任危机，加大了信用成本，造成互联网的无序。而区块链作为信任的制造机器，推动人们进入秩序互联网阶段，构建人人共享秩序，消除信任鸿沟。如果说信息互联网解决了无界问题，价值互联网解决了无价问题，那秩序互联网则是要解决互联网信息的无序问题。

由区块链技术构建的产业互联互通，降低了产业链各环节的信任成本，从而降低产业链的交易成本，改善了产业链的效率，提高了产业链间的信息质量。

5.3 助力倾听市场声音

服务整个产业链的大数据智能链，连通产业的供需两端以及整个产业链环节，把市场变化的信息第一时间传递给供给和需求两端。

从某种程度上，大数据智能链就像一个市场的传感器，随时感知搜集市场需求结构的变化，同时传递给供给端。在产业链整合的协同下，整个产业生产结构将发生变化，从传统制造业模式的“我生产什么产品，市场就卖什么产品”，转变为“市场需要什么产品，企业就生产什么产品”，配合产业链整合，就能高效的实现制造的大规模个性化生产，真正实现产品制造的 C2M 模式。

产业链的整合，除了核心龙头企业外，还支持产业链的规模整合，同类企业上到大数据智能链上，使大数据智能链能够收集到的市场需求信息的范围进一步扩大，对市场变化的信息把握进一步加大。

5.4 改善产业链效率

大数据智能链，支持核心企业对产业链进行整合。包括整合产业链中的小、散资源，也包括渠道资源、技术资源，以及金融、政策等要素。

以零部件为例，龙头企业并不直接与产业链中的小、散资源合作，而是向一定规模的中小型下订单，因为订单规模很大，这就倒逼小、散资源创新经营方式，变“小散”为规模化生产，加入到产业链的供需结构中。既保障了龙头企业的产业链安全、效率，也改善了产业链零部件环节的供给结构和生产效率。

产业链整合，有了龙头企业的参与，以及区块链金融的解决方案，与银行、投资机构、投资资金等金融资源方的合作也变得容易，为“产业+金融”的合作扫除障碍。而资金得到保障后，后续产业规模化、品牌化等难题的解决就有了坚实的基础。

大数据智能链通过提供金融服务、供应链金融解决方案以及整套完整的供应链管理系统，连接产业链内部供需各环节，以及外部各类资源，支持产业链的顺利整合，大幅提高了整个产业链的效率。

5.5 推动产业供给侧结构改革

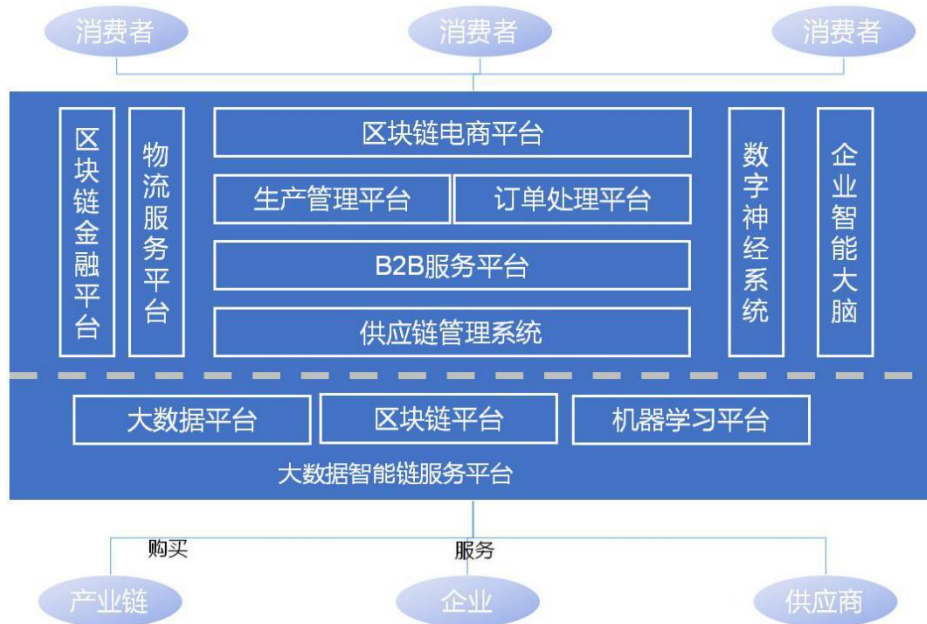
产业链上各环节倾听到市场需求变化，想要生产出市场需要的产品，还需要产业链各方的配合与支持：

产业链整合，吸纳金融、科技、产业、营销等多个领域参与，使得靠产业链各环节单打独斗的产业，进入深度分工、延伸拓展、协作推进的新阶段。

大数据智能链，支持产业链整合延伸至金融、科技、营销等领域，协助产业链设计、制造出符合市场需求的新产品，并通过营销、物流等服务，送达终端客户；改善产品品质，创建新的品牌，从而深度改变了整个产业链的产品供给结构、生产结构、产品结构，推动产业的供给侧结构改革。

6 大数据智能链体系

6.1 大数据智能链架构



大数据智能链，创新的将大数据、区块链、人工智能（机器学习）等新技术，整合形成一个高效统一的数据存储和计算基础设施。

大数据智能链，根据产业链的五链模型，给出九大解决方案，推动产业的发展和企业成长。

6.2 大数据智能链技术特点

6.2.1 区块链技术特点

➤ 分层架构

大数据智能链，依托井通科技的区块链底层平台技术，将区块链信任的传递分为 5 个层次：网络层、区块层、数据层、价值层以及合约层，不同层级服务不同的业务场景，从而实现区块链上不同层次的业务场景。

第一层是网络层，实现了点对点去中心化的数据传输，建立了数据传输的信任。

第二层是区块层，实现了区块链的基本功能，是不可篡改的且有时间性的区块构造所在，并以区块形式记录所有交易信息。这一层次是所有区块链的必备功能。将其功能独立可以便于模块化，也可以满足一些基础性应用的需求。与传统的关系型数据库和 NoSQL 相比，

它更适合于构建基于大批量用户数据一致性的应用，包括征信、溯源、防伪等。

第三层是数据层，基于区块链层并允许所有节点使用自己的数据库技术将信息写入区块，实现了大数据和区块链两种技术的融合。

第四层是价值层，实现了数字资产生命周期管理，以及资产的发布、交易、互联、交换、冻结和授权等功能。

第五层是合约层，在区块链系统上构建了智能合约。这一层提供智能合约功能，并基于价值层实现复杂的商业逻辑计算功能。

这种分层架构实现了真正的满足商业用途的稳定安全的技术架构，可以有效防止 DAO 事件的发生。

➤ 智能合约创新

1) 合约的异步调用

合约的创建和合约功能的调用，都由交易（TX）发起，如果执行的过程中需要修改帐号的余额，则会发起交易并发到下层的 TX，所有这些交易都将在 TX 执行验证并记录在底层的区块链中。这样的处理有以下优点：

a. TX 层的交易不受合约的影响

b. TX 层的跟合约相关的交易以 **metadata** 的方式保存合约的代码 和状态，合约的具体状态对应于对合约相关功能的调用及参数，TX 层的状态 **hash** 保证信息的一致性。

c. 合约层的执行由多个合约节点 **contract validator** 执行，采用 **PBFT** 的方式分别执行并进行共识

d. 每个合约节点采用 **VM** 执行代码

e. 合约节点保存合约执行的 **storage**

2) 合约的分片处理。

当前以太坊智能合约的执行速度受到共识算法的影响，不能满足商业需求。在大数据智能链中，执行智能合约的方式是通过分片技术实现。智能合约服务器都可以在运行时进行配置，以处理不同部分的智能合同。整个系统吞吐量可以比传统方式快 **10 倍-100 倍**。

➤ 基于 IPFS 协议的文件数据存储

IPFS 用基于内容的寻址替代传统的基于域名的寻址，用户不需要关心服务器的位置，不用考虑文件存储的名字和路径。我们将一个文件放到 IPFS 节点中，将会得到基于其内容计算出的唯一加密哈希值。哈希值直接反映文件的内容，哪怕只修改 **1 比特**，哈希值也会完全不同。当 IPFS 被请求一个文件哈希时，它会使用一个分布式哈希表找到文件所在的节点，

取回文件并验证文件数据。

IPFS 是通用目的的基础架构，基本没有存储上的限制。大文件会被切分成小的分块，下载的时候可以从多个服务器同时获取。IPFS 的网络是不固定的、细粒度的、分布式的网络，可以很好的适应内容分发网络的要求。这样的设计可以很好的共享各类数据，包括图像、视频流、分布式数据库、整个操作系统、模块链、8 英寸软盘的备份，还有静态网站。

IPFS 和区块链能够完美结合，用户可以使用 IPFS 来处理大量数据，然后把对应的加密哈希存储到区块链中并打上时间戳。这样就无需将数据本身放在链上，不但可以节省区块链的网络带宽，还可以对其进行有效保护。关于文件的安全性，一方面可以加密后存入 IPFS，另一方面也可利用 IPFS 实现文件分布式共享。

IPFS 弥补了现有区块链系统在文件存储方面的短板，将 IPFS 的永久文件存储和区块链的不可篡改、时间戳证明特性结合，非常适合在产业互联网方面加以应用。

➤ 基于 BFT 的 RAFT 共识协议

拜占庭有 10 个将军要攻击一个繁华强大的城邦，由于某种原因他们要分开行军和驻扎。这个城邦虽然没有拜占庭强大，但也足以打败少于半数的拜占庭联军。拜占庭的将军们必须有超过半数一起攻打，不然就会失败，因此他们需要商定是否在约定时间攻击。他们需要面临的问题是，信使可能没有到达，也可能某几个将军是叛徒，会发不同的决定给不同的将军，误导他们使得最终无法统一行动。

这其实是个分布式系统部分节点下线或发生错误的问题，涉及分布式算法，也是区块链共识机制可能依赖到的底层技术。

专业术语叫 BFT(Byzantine Fault Tolerance)算法/协议。

Raft 协议的工作方式

在一个由 Raft 协议组织的集群中有三类角色：

1.Leader（领袖）

2.Follower（群众）

3.Candidate（候选人）

就像一个民主社会，领袖由民众投票选出。刚开始没有领袖，所有集群中的参与者都是群众，那么首先开启一轮大选，在大选期间所有群众都能参与竞选，这时所有群众的角色就变成了候选人，民主投票选出领袖后就开始了这届领袖的任期，然后选举结束，所有除领袖的候选人又变回群众角色服从领袖领导。这里提到一个概念「任期」，用术语 Term 表达。

关于 Raft 协议的核心概念和术语就这么多而且和现实民主制度非常匹配，所以很容易理解。

Leader 选举过程

在极简的思维下，一个最小的 Raft 民主集群需要三个参与者（如：A、B、C），这样才可能投出多数票。初始状态 ABC 都是 Follower，然后发起选举这时有三种可能情形发生。下图中前二种都能选出 Leader，第三种则表明本轮投票无效（Split Votes），每方都投给了自己，结果没有任何一方获得多数票。之后每个参与方随机休息一阵（Election Timeout）重新发起投票直到一方获得多数票。这里的关键就是随机 timeout，最先从 timeout 中恢复发起投票的一方向还在 timeout 中的另外两方请求投票，这时它们就只能投给对方了，很快达成一致。

选出 Leader 后，Leader 通过定期向所有 Follower 发送心跳信息维持其统治。若 Follower 一段时间未收到 Leader 的心跳则认为 Leader 可能已经挂了再次发起选主过程。

Leader 节点对一致性的影响

Raft 协议强依赖 Leader 节点的可用性来确保集群数据的一致性。数据的流向只能从 Leader 节点向 Follower 节点转移。当 Client 向集群 Leader 节点提交数据后，Leader 节点接收到的数据处于未提交状态（Uncommitted），接着 Leader 节点会并发向所有 Follower 节点复制数据并等待接收响应，确保至少集群中超过半数节点已接收到数据后再向 Client 确认数据已接收。一旦向 Client 发出数据接收 Ack 响应后，表明此时数据状态进入已提交（Committed），Leader 节点再向 Follower 节点发通知告知该数据状态已提交。

在这个过程中，主节点可能在任意阶段挂掉，看下 Raft 协议如何针对不同阶段保障数据一致性的。

1. 数据到达 Leader 节点前

这个阶段 Leader 挂掉不影响一致性，不多说。

2. 数据到达 Leader 节点，但未复制到 Follower 节点

这个阶段 Leader 挂掉，数据属于未提交状态，Client 不会收到 Ack 会认为超时失败可安全发起重试。Follower 节点上没有该数据，重新选主后 Client 重试重新提交可成功。原来的 Leader 节点恢复后作为 Follower 加入集群重新从当前任期的新 Leader 处同步数据，强制保持和 Leader 数据一致。

3. 数据到达 Leader 节点，成功复制到 Follower 所有节点，但还未向 Leader 响应接收

这个阶段 Leader 挂掉，虽然数据在 Follower 节点处于未提交状态（Uncommitted）但

保持一致，重新选出 Leader 后可完成数据提交，此时 Client 由于不知到底提交成功没有，可重试提交。针对这种情况 Raft 要求 RPC 请求实现幂等性，也就是要实现内部去重机制。

4. 数据到达 Leader 节点，成功复制到 Follower 部分节点，但还未向 Leader 响应接收

这个阶段 Leader 挂掉，数据在 Follower 节点处于未提交状态（Uncommitted）且不一致，Raft 协议要求投票只能投给拥有最新数据的节点。所以拥有最新数据的节点会被选为 Leader 再强制同步数据到 Follower，数据不会丢失并最终一致。

5. 数据到达 Leader 节点，成功复制到 Follower 所有或多数节点，数据在 Leader 处于已提交状态，但在 Follower 处于未提交状态

这个阶段 Leader 挂掉，重新选出新 Leader 后的处理流程和阶段 3 一样。

6. 数据到达 Leader 节点，成功复制到 Follower 所有或多数节点，数据在所有节点都处于已提交状态，但还未响应 Client

这个阶段 Leader 挂掉，Cluster 内部数据其实已经是一致的，Client 重复重试基于幂等策略对一致性无影响。

7. 网络分区导致的脑裂情况，出现双 Leader

网络分区将原先的 Leader 节点和 Follower 节点分隔开，Follower 收不到 Leader 的心跳将发起选举产生新的 Leader。这时就产生了双 Leader，原先的 Leader 独自在一个区，向它提交数据不可能复制到多数节点所以永远提交不成功。向新的 Leader 提交数据可以提交成功，网络恢复后旧的 Leader 发现集群中有更新任期（Term）的新 Leader 则自动降级为 Follower 并从新 Leader 处同步数据达成集群数据一致。

综上穷举分析了最小集群（3 节点）面临的所有情况，可以看出 Raft 协议都能很好的应对一致性问题，并且很容易理解。

6.2.2 大数据技术特点

在区块链的各类应用中，区块链的不可篡改特性，被发挥得淋漓尽致；同时，那些需要在使用过程中修改、调整的数据，就需要采用其他数据访问技术来处理。

基于此，大数据智能链的存储计算架构中，直接整合了大数据的标志技术 Hadoop 系统，包括各类技术项目。

➤ HDFS

Hadoop 系统其他组件都是在 HDFS 的基础上组合或者使用的。HDFS 具有高容错性、适

合批处理、适合大数据处理、可构建在廉价机器上等优点，缺点是低延迟数据访问、小文件存取、并发写入、文件随机修改。

➤ HBase

它是一个高可靠、高性能、面向列、可伸缩的分布式存储系统，利用 Hbase 技术可在廉价 PC Server 上搭建大规模结构化存储集群。

HBase 的表有以下特点：

- 1) 大：一个表可以有上亿行，上百万列。
- 2) 面向列：面向列表（簇）的存储和权限控制，列（簇）独立检索。
- 3) 稀疏：对于为空（NULL）的列，并不占用存储空间，因此，表可以设计的非常稀疏。

HBase 提供的访问方式有命令行 shell 方式，java API(最高效和常用的)

HBase 的使用场景：

需对数据进行随机读操作或者随机写操作；

大数据上高并发操作，比如每秒对 PB 级数据进行上千次操作；

读写访问均是非常简单的操作，比如历史记录，历史订单查询，流量通话清单的查询。

➤ Hive

Hive 最初用于解决海量结构化的日志数据统计问题的 ETL(Extraction-Transformation-Loading) 工具，Hive 是构建在 Hadoop 上的数据仓库平台，设计目标是可以传统 SQL 操作 Hadoop 上的数据，让熟悉 SQL 编程的人员也能拥抱 Hadoop（注意。是数据仓库。不是数据库啊。）

使用 HQL 作为查询接口

使用 HDFS 作为底层存储

使用 MapReduce 作为执行层

Hive 就是基于 Hadoop 的一个数据仓库工具，是为简化 MapReduce 编程而生的，非常适合数据仓库的统计分析，通过解析 SQL 转化成 MapReduce，组成一个 DAG(有向无环图)来执行。

➤ Flume

Flume 是一个高可用的，高可靠的，分布式的海量日志采集、聚合和传输的系统，Flume 支持在日志系统中定制各类数据发送方，用于收集数据；同时，Flume 提供对数据进行简单处理，并写到各种数据接受方(可定制)的能力。

Flume 就是一个数据管道，支持很多源(source)，sink(目标)，比如拉取 nginx 日志可以拿

这个工具简单一配就可用。

➤ YARN

YARN 是 Hadoop 2.0 中的资源管理系统，它的基本设计思想是将 MRv1 中的 JobTracker 拆分成了两个独立的服务：一个全局的资源调度器 ResourceManager 和每个应用程序特有的应用程序管理器 ApplicationMaster，该调度器是一个 "纯调度器"，不再参与任何与具体应用程序逻辑相关的工作，而仅根据各个应用程序的资源需求进行分配，资源分配的单位用一个资源抽象概念 "Container" 来表示，Container 封装了内存和 CPU。此外，调度器是一个可插拔的组件，用户可根据自己的需求设计新的调度器，YARN 自身提供了 Fair Scheduler 和 Capacity Scheduler。

应用程序管理器负责管理整个系统中所有应用程序，包括应用程序的提交、与调度器协商资源以启动 ApplicationMaster、监控 ApplicationMaster 运行状态并在失败时重新启动等。

➤ ZooKeeper

随着计算节点的增多，集群成员需要彼此同步并了解去哪里访问服务和如何配置，ZooKeeper 正是为此而生的。ZooKeeper 顾名思义就是动物园管理员，它是用来管大象 (Hadoop)、蜜蜂(Hive) 和 小猪(Pig) 的管理员。ZooKeeper 是一个分布式的，开放源码的分布式应用程序协调服务，以 Fast Paxos 算法为基础实现同步服务，配置维护和命名服务等分布式应用。

➤ Kafka

Kafka 是一种高吞吐量的分布式发布订阅消息系统，它可以处理消费者规模的网站中的所有动作流数据。这种动作（网页浏览，搜索和其他用户的行动）是在现代网络上的许多社会功能的一个关键因素。这些数据通常是由于吞吐量的要求而通过处理日志和日志聚合来解决。对于像 Hadoop 的一样的日志数据和离线分析系统，但又要求实时处理的限制，这是一个可行的解决方案。Kafka 的目的是通过 Hadoop 的并行加载机制来统一线上和离线的消息处理，也是为了通过集群来提供实时的消费。

➤ Spark

Spark 是专为大规模数据处理而设计的快速通用的计算引擎。Spark 是 UC Berkeley AMP lab (加州大学伯克利分校的 AMP 实验室)所开源的类 Hadoop MapReduce 的通用并行框架，Spark，拥有 Hadoop MapReduce 所具有的优点；但不同于 MapReduce 的是 Job 中间输出结果可以保存在内存中，从而不再需要读写 HDFS，因此 Spark 能更好地适用于数据挖掘与机器学习等需要迭代的 MapReduce 的算法。

➤ Spark Streaming

构建在 Spark 上处理 Stream 数据的框架，基本的原理是将 Stream 数据分成小的时间片断(几秒)，以类似 batch 批量处理的方式来处理这小部分数据。Spark Streaming 构建在 Spark 上，一方面是因为 Spark 的低延迟执行引擎（100ms+），虽然比不上专门的流式数据处理软件，也可以用于实时计算，另一方面相比基于 Record 的其它处理框架（如 Storm），一部分窄依赖的 RDD 数据集可以从源数据重新计算达到容错处理目的。此外小批量处理的方式使得它可以同时兼容批量和实时数据处理的逻辑和算法。方便了一些需要历史数据和实时数据联合分析的特定应用场合。

➤ Storm

Storm 是个实时的、分布式以及具备高容错的计算系统。Storm 也可以处理大批量的数据，然而 Storm 在保证高可靠性的前提下还可以让处理进行的更加实时；也就是说，所有的信息都会被处理。Storm 同样还具备容错和分布计算这些特性，这就让 Storm 可以扩展到不同的机器上进行大批量的数据处理。它同样还有以下的这些特性：

- 易于扩展。对于扩展，你只需要添加机器和改变对应的 topology（拓扑）设置。Storm 使用 Hadoop Zookeeper 进行集群协调，这样可以充分的保证大型集群的良好运行。
- 每条信息的处理都可以得到保证。
- Storm 集群管理简易。
- Storm 的容错机能：一旦 topology 递交，Storm 会一直运行它直到 topology 被废除或者被关闭。而在执行中出现错误时，也会由 Storm 重新分配任务。
- 尽管通常使用 Java，Storm 中的 topology 可以用任何语言设计。

6.2.3 人工智能技术特点

为产业链服务的大数据智能链，也需要提供人工智能的应用，因此，需要提供机器学习平台和机器学习开发框架。为此，我们集成了 Google 的 TensorFlow。

TensorFlow 是谷歌 2015 年开源的一个人工智能平台。就如命名一样，TensorFlow 为张量从图的一端流动到另一端计算过程。TensorFlow 是将复杂的数据结构传输至人工智能神经网络中进行分析和处理过程的系统。TensorFlow 可被用于语音识别或图像识别等多项机器深度学习领域，它可在小到一部智能手机、大到数千台数据中心服务器的各种设备上运行。

TensorFlow 是一个深度学习框架，支持 Linux 平台，Windows 平台，Mac 平台，甚至手机移动设备等各种平台。

TensorFlow 提供了非常丰富的深度学习相关的 API，可以说目前所有深度学习框架里，提供的 API 最全的，包括基本的向量矩阵计算、各种优化算法、各种卷积神经网络和循环神经网络基本单元的实现、以及可视化的辅助工具、等等。

TensorFlow 并不仅仅是一个深度学习库，只要可以把你的计算过程表示成一个数据流图的过程，我们就可以使用 TensorFlow 来进行计算。TensorFlow 允许我们用计算图的方式还建立计算网络，同时又可以对网络进行操作（具体计算图是什么意思，后面会有详细的介绍）。用户可以基于 TensorFlow 的基础上用 Python 编写自己的上层结构和库，如果 TensorFlow 没有提供我们需要的 API 的，我们也可以自己编写底层的 C++ 代码，通过自定义操作将新编写的功能添加到 TensorFlow 中。

TensorFlow 可以在 CPU 和 GPU 上运行，可以在台式机，服务器，移动设备上运行。

TensorFlow 采用非常易用的 python 来构建和执行我们的计算图，同时也支持 C++ 的语言。我们可以直接写 python 和 C++ 的程序来执行 TensorFlow，也可以采用交互式的 ipython 来方便的尝试我们的想法。后续会支持更多流行的语言，比如 Lua，JavaScript 或者 R 语言。

TensorFlow 提供了所有开源的深度学习框架里，最全的算法库，并且在不断的添加新的算法库。这些算法库基本上已经满足了大部分的需求，对于普通的应用，基本上不用自己再去自定义实现基本的算法库了。

TensorFlow 的官方网站，提供了非常详细的文档介绍，内容包括各种 API 的使用介绍和各种基础应用的使用例子，也包括一部分深度学习的基础理论。

6.3 大数据智能链业务架构

6.3.1 平台助力产业融资

加密货币也是货币，具备货币的某些特性。货币是从商品中分离出来固定充当一般等价物的特殊商品，它在经济体内循环。

从某种意义上讲，一个产业链就是一个货币循环体。

区块链技术的加密数字代币，在一个产业链内流通，是可能的。

因此，我们创新的将依托区块链技术的融资体系，引入到产业链服务平台中，帮助整产业链的企业，更便利、低成本、可信赖的完成各类融资。

6.3.2 区块链电商平台推动企业转型

电商平台具有诸多的优点，包括：

将传统的商务流程电子化、数字化，一方面以电子流代替了实物流，可以大量减少人力、物力，降低了成本；另一方面突破了时间和空间的限制，使得交易活动可以在任何时间、任何地点进行，从而大大提高了效率。

具有的开放性和全球性的特点，为企业创造了更多的贸易机会。

重新定义了传统的流通模式，减少了中间环节，使得生产者和消费者的直接交易成为可能，从而在一定程度上改变了整个社会经济运行的方式。

一方面破除了时空的壁垒，另一方面又提供了丰富的信息资源，为各种社会经济要素的重新组合提供了更多的可能，这将影响到社会的经济布局 and 结构。

依托区块链技术，服务于产业链而建立的电商平台，还具有新的特点。

产品的信息的真实性；

产品生产全过程的全程可追溯；

物流信息的全程可追溯；

从而让消费者放心购真品，让企业远离假货困扰。

6.3.3 B2B 平台改变产业链结构

B2B 平台是企业间的电子商务，即企业与企业之间通过互联网进行产品、服务及信息的交换。

依托区块链技术，只有加入产业链的企业，才可以进入 B2B 平台。

这样，以往产业链之间复杂、无处记录、难以计量的产业链各环节之间的交易，将在区块链上呈现出透明、清晰的记录。

它将大大降低产业链内部的交易成本；

它将产业链内部各环节的交易信息，清晰的传递给产业链各环节，大大降低了各环节对市场信息的收集成本，改善了产业链各环节收集信息的质量。从而改善了产业链各企业的决策效率和决策质量，极大的降低整个产业链的成本。

在一个产业链集中的平台交易，市场的信息告诉产业链各企业，各自的竞争优势，而各企业的优势强化和劣势改变，直接改变了整个产业链的竞争格局，从而改变了产业链结构，使得整个产业链进一步发展为灵活面向市场的高效产业链。

6.3.4 供应链管理改善产业链效率

供应链管理是一个已经进行了多年的课题。

供应链管理就是协调企业内外资源来共同满足消费者需求，当我们把供应链上各环节的

企业看作为一个虚拟企业同盟，而把任何一个企业看作为这个虚拟企业同盟中的一个部门时，同盟的内部管理就是供应链管理。只不过同盟的组成是动态的，根据市场需要随时在发生变化。

有效的供应链管理可以帮助实现四项目标：缩短现金周转时间；降低企业面临的风险；实现盈利增长；提供可预测收入。

供应链管理可以有效的改善产业链的效率。

6.3.5 大数据构建企业数字神经系统

在大数据时代，企业和整个产业需要关心的数据，有三个方面的来源。

第一，企业和整个产业链建立的各类系统，记录的企业经营管理各类正式的、规范化的数据或者是如邮件系统中，非正式化的数据

第二、企业内部发生的各类信息，但没有正式系统加以采集、记录、整理、分析和利用。

第三、互联网上各类与企业、产业相关的信息。

企业的数字神经系统，就是要把这三类信息都采集、整合到一个平台上，为企业了解整个市场变化信息，为后续的决策支持作好数据准备。

6.3.6 企业大脑推动供给侧结构改革

在人工智能时代，我们发现标杆企业都有一个神奇的系统，名字叫做“企业大脑平台”。

这个企业大脑平台就是企业建设出来，为决策服务的信息系统，人工智能系统；

企业大脑平台，需要利用企业数字神经系统收集整理的各方面数据，分析判断市场需求变化趋势，并结合企业自身资源能力，迅速的给出建议和决策。这些决策，最终都会导致企业产品的设计、制造、原材料采购、物流、服务等各环节的变化。

产业链上各个企业的变化，最终就是产业链的深度分工整合，引发产业的供给侧结构改革。

7 执行团队

7.1 核心管理团队



武源文

无锡井通网络科技有限公司 CEO

区块链金融协会副会长

区块链与大数据技术应用专家

中关村大数据产业联盟副秘书长

曾任长江大数据交易所总裁、贵阳大数据交易所董事

《区块链世界》、《区块链与大数据》等书主创



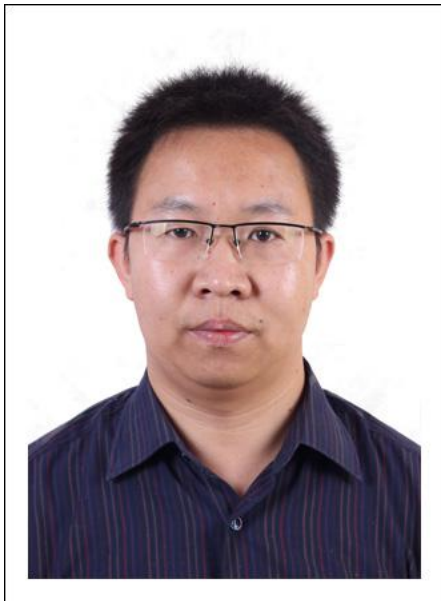
温江凌

北京井立通科技有限公司 总经理

资深解决方案专家、系统分析师、业务架构师

软件行业超过 22 年的工作经验

产业互联网的积极践行者



李孟杰

北京井立通科技有限公司 CTO

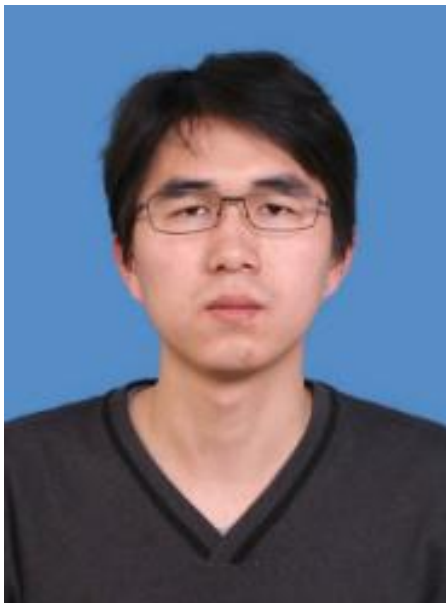
大数据专家、资深项目经理

十多年技术开发经历，在互联网、大数据技术等技术领域具备丰富经验，并主导参与了多个大数据、互联网项目。



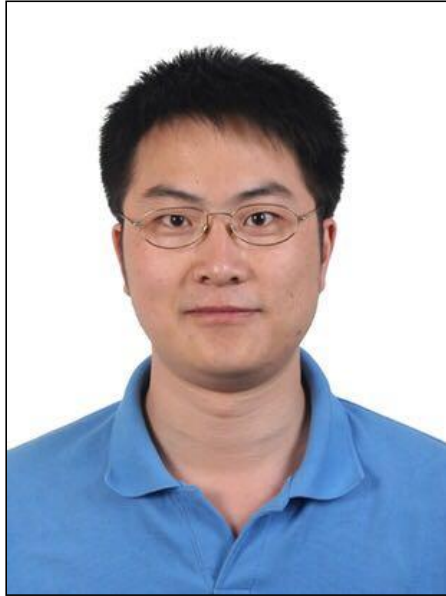
李正鹏

硅谷博士级领域标杆人物,南达科他州立大学博士, 硅谷区块链底层技术专家, 拥有 20 余年算法研究和软件开发的经验, 负责底层架构, 算法研究和软件接口协议开发, 曾在美国参与 NASA, NOAA 的多项国家级科研模型开发, 独立承担过多个大型软件工程, 具有丰富的系统架构和软件开发实施经验。



杨建新

清华大学计算机系硕士, 区块链底层架构师, 密码学专家, 井通底层平台国内负责人, 编制 API、SDK 等相关标准, 参与某世界 500 强企业福利汇、企业级钱包、供应链金融等项目建设。



蔡维佳

原支付宝大数据专家。清华大学硕士，大数据可视化专家，目前就职于井通科技，2006年开始在IBM中国研究院从事信息可视化与分析的设计与开发工作，2012年在淘宝从事从事数据分析与数据可视化工作，大规模交易网张分析与可视化的工作，2014年在支付宝从事新一代BI系统建设，以及可视化报表的开发与设计工作，同时承担着阿里及支付宝内部数据大屏的设计与开发工作，具备多年区块链研究与实践经验。

技术方面精通C、Java语言，Javascript(nodejs)语言，Flash(Actionscript)语言，HiveQL(SQL)语言，熟悉C++、R、Python、Lisp、Prolog、Erlang、Golang，熟练掌握各种网站及互联网公司网络架构，熟悉数据处理各种流程及Nosql系统。

7.2 专家团队



刘文献院长

区块链金融协会会长、贵阳众筹金融交易所董事长，北京特许经营权交易所董事长，北京师范大学珠海分校特许经营学院创始院长，北京航空航天大学北海学院创建者，贵州财经大学大数据金融学院副院长，北京两岸金融研究院执行院长，世界众筹大会发起人。刘文献院长是我国著名特许经营专家和学术带头人，是我国世界众筹之都众筹金融生态体系的架构师之一和实践家。他被誉为“中国特许经营第一人”和“中国众筹金融之父”，主编或合著有新世纪特许经营丛书及众筹金融丛书，包括《中国特许经营大全》、《特许经营概论》《众筹之都》《众筹金融生态》《解放众筹》《众筹的解放》《众链》《区块链与大数据》等近五十本特许经营、区块链、众筹等教科书、畅销书。刘文献院长一直是个前沿学科创新者、大市场大格局实践家和新金融大生态建筑师，他被英国众筹中心巴里教授认为：是少数洞悉未来新经济，并能够帮助设计、计划和构建某些新金融生态世界的建筑大师之一。

周 沙



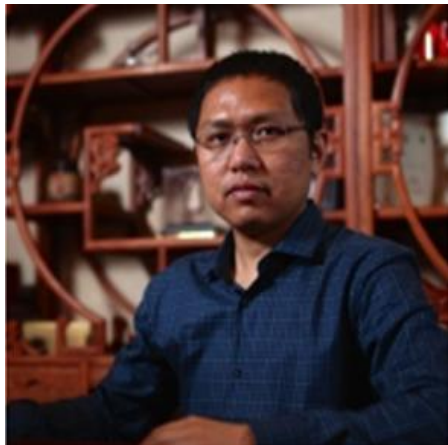
硅谷 20 年工作经验；最早的区块链开发研究专家；井通科技创始人；

《区块链世界》和《区块链+大数据》主要作者



陈小虎

美国硅谷区块链技术大咖，18 年技术工作经验，MOAC 公有链 CTO,FINTECH、人工智能、3D 等领域的顶级科学家，美国加州大学河滨分校硕士，曾在职理光相机，Nextengine，Waytools 等公司，所创建 MOAC 公有链是可提供鲁棒且可扩展的平台，在区块链领域内首先提出异步智能合约及分流技术，最近 5 年致力于区块链技术的开发。



赵国栋

中关村大数据产业联盟秘书长、北京大数据研究院副院长，汇冠股份独立董事，北京邮电大学经济管理学院特聘导师，国务院学位办大数据专业硕士特聘导师、中国计算机学会大数据专家委员会委员、曾获资本市场新财富金牌分析师团队第一名。

出版有《大数据时代的历史机遇》《产业互联网》等著作。系统性提出“联盟三论”等原创

思想：论企业战略，提出以数据资产为核心的企业战略发展理论；论区域经济，提出以产业互联网为统摄的产业升级、产业聚集理论；论组织变革，提出以中间市场为特征的组织变革理论，指导区域经济和企业发展。被誉为中国大数据、产业互联网的引导者和启蒙者。担任多家上市公司的高级顾问和独立董事。