

# The Application of Summingbird Cloud Computing Platform in Energy Internet

Zhijian Liu, Mingyu Wang\*

Faculty of Power Engineering, Kunming University of Science and Technology, Kunming Yunnan  
Email: 248400248@qq.com, \*1506759936@qq.com

Received: Dec. 10<sup>th</sup>, 2015; accepted: Dec. 27<sup>th</sup>, 2015; published: Dec. 30<sup>th</sup>, 2015

Copyright © 2015 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

This text analyzes the practical application of the cloud computing on the energy internet. Firstly, the relative theory and practice works are introduced in domestic and overseas. Secondly, this paper summarizes possible data types and its source, and then presents using the Summingbird real-time data processing platform as the cloud computing platform of Energy Internet, meanwhile the feasibility and superiority of Summingbird are discussed. Finally the topological structure named “Energy Internet and two Cloud Computing platforms” (EI & TC), which is an information system framework based on the cloud computing, is presented. Hoping the cloud computing platform will play an important role in the developing of Energy Internet.

## Keywords

Energy Internet, Cloud Computing, Summingbird, EI & TC

---

# Summingbird云计算平台在能源互联网中的应用

刘志坚, 王明禹\*

昆明理工大学电力工程学院, 云南 昆明  
Email: 248400248@qq.com, \*1506759936@qq.com

---

\*通讯作者。

收稿日期: 2015年12月10日; 录用日期: 2015年12月27日; 发布日期: 2015年12月30日

## 摘要

本文针对云计算技术在能源互联网的实际应用进行分析, 首先介绍了国内外关于能源互联网的理论 and 实践工作, 其次总结能源互联网中的数据类型及来源, 并提出以Summingbird实时数据处理平台为能源互联网的云计算平台, 然后分析了利用Summingbird平台可行性和优势, 最后提出了“一网两云”的基于云计算技术的信息系统构架方案, 望云计算平台在能源互联网发展中起到重要作用。

## 关键词

能源互联网, 云计算, Summingbird, 一网两云

## 1. 引言

当今, 由于化石能源储量的逐年减少, 使得可再生资源的需求迅速扩大。日本、美国和中国相继提出到 2050 年实现可再生能源在能源供应量分别占能源总量的 100%, 80%和 60%~70%的目标。然而因可再生资源自身的特性, 提高分布式电源的利用率及合理储存等方面存在诸多问题, 及时的信息和能源共享成为实现可再生资源有效利用的主要手段, 因此“能源互联网”的电网形式应运而生。

多种能源管理和共享一定伴随着海量数据信息的处理与计算, 由于云计算对海量数据的计算和存储具有灵活性和稳定性的优势, 所以它将成为能源互联网的主要数据处理手段, 可实现信息系统与电力系统的物理设备有效协调运行。作者杰瑞米·里夫金在《第三次工业革命》一书中描述未来理想的能源互联网[1]。在能源互联网中, 人们在生活工作场合生产和使用清洁能源, 并且将多余的能源分享给其他用户, 就像如今在互联网中分享信息一样。

文献[2] [3]提及的美国国家科学基金项目。“未来可再生电力能源传输与管理系统”(The Future Renewable Electric Energy Delivery and Management, FREEDM)项目类比网络技术提出了能源路由器的概念, 使得人们对信息系统和物理设备深度融合的能源互联网有了更加具体的了解。

在欧洲, 德国在 2008 年提出了 E-Energy 计划[4]。该计划充分的利用信息通信技术解决未来电力系统的分布式能源利用问题和数字联网的供电可靠性问题, 到 2020 年完成信息网络与电力系统基本融合, 互联网可以对能源网络的物理系统进行高效的协调控制。日本提出了“电力路由器”的概念[5]。在互联网基础上, 日本对其国家电网逐渐改造为多个自主区域, 通过电力路由器统筹和调度一定范围区域的电力, 以达到各区域电能的互联的目的。

国内学者经过对能源互联网的深入分析与研究, 提出了相关的网架结构和关键技术[6]-[11]。并将云计算技术作为解决能源互联网数据信息处理和储存的主要手段。

文献[12]-[16]介绍了 Twitter 的实时信息处理平台 Summingbird 以及相关的云计算技术。它由 Hadoop 和 Storm 两个云计算平台整合而成, Hadoop 提供批量延时数据处理, Storm 提供流式实时数据处理。尽管该平台起初用于社交网站信息处理, 但是正如开发者所言, 该案例也适用于其他互联网数据处理业务。

本文讨论了将云计算技术应用到能源互联网的一种方案。针对能源互联网中存在实时数据和非实时数据的情况, 提出了应用兼具批处理和流处理能力的 Summingbird 数据处理平台作为能源基础云计算平台; 根据能源互联网的层次结构, 提出了在广域电网和局域电网中分别建立“广域云”和“局域云”的云计算拓扑结构。

## 2. 能源互联网中的数据

如图 1 为能源互联网基础架构示意图。能源互联网络中的电能路由器和电能交换器是其中的关键器件：电能路由器相当于互联网中的路由器，是负责广域电网和局域电网之间的信息与能量的互联枢纽；电能交换器相当于互联网中的局域网交换机，负责局域网中的数据 and 能源快速互联分享。能源互联网包括以输电线路为主干线路和大型发电厂组成的“广域电网”，以及以配电网、分布式能源及微网等单元组成的“局域电网”。

当然，电能稳定的输送和共享，要依靠电能路由器和电能交换器等设备的协同合作，实时数据处理和预测性评估就显得尤为重要，尤其是在能源互联网通信环境中，数据量可能达到 PB 级甚至更大。这些数据本文分为两类：时间跨度较大的非实时数据和实时数据。下面分析能源互联网存在的几个主要的数据来源：

### 1) 新型清洁能源数据

能源互联网建设的主要目的之一是提高新型清洁能源利用率，以实现电力系统资源多元化。在此过程中，对新型清洁能源的电能监控和分配显得尤为重要。信息系统将采集相关数据予以分析并给出调控策略。以风力发电为例，风力发电系统的实时数据包括：实时风机发电量、风机运行状态数据、实时气象状态数据等数据，以及非实时数据包括：如中长期的发电量、各季度的风力资源数据、中长期风机运行状态数据等。

### 2) 电能用户数据

方便快捷的能源共享是能源互联网主要特征之一，电能用户不仅是电能消费者，也可能以电能供给者的身份参与电能交易。同时能源互联网为用

户提供多种个性化用电方案，以减少“局域网”中接入“广域网”的电能容量为目的，降低用户的用电成本。信息系统要实时处理用户实时电能需求信息、实时电价数据、电能交易信息、即插即用、电能双向流动计量、分布式储能装置监控等信息数据。同时也要对用户长期电能需求数据、用户家用电器使用情况和电能交易数据等进行记录，以供信息系统为用户提供电能使用方案。

### 3) 网络运行管理数据

数据化、智能化的能源互联网要求数据平台可高效的对物理系统的信息进行实时监测与处理，如电

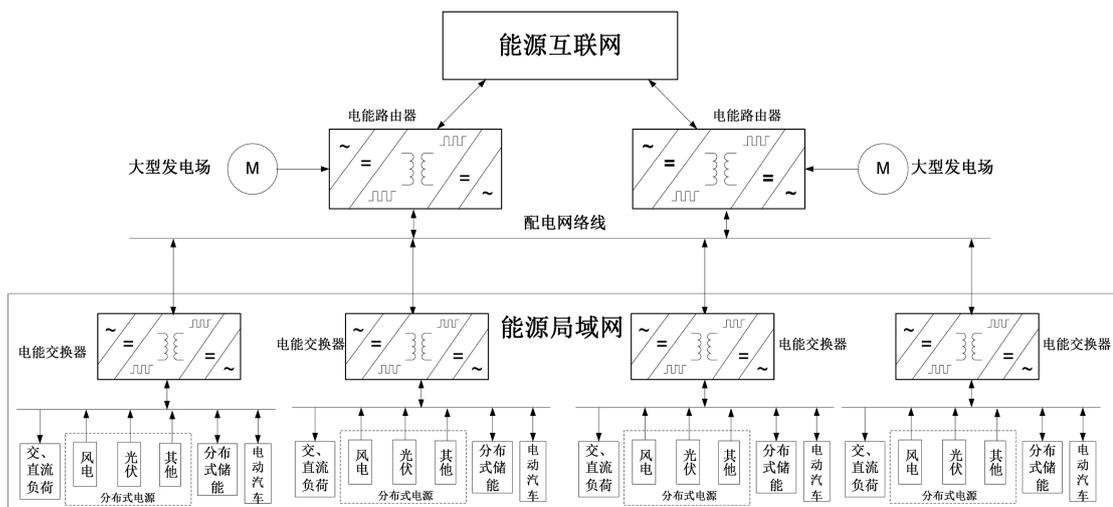


Figure 1. Schematic diagram of energy internet basic architecture

图 1. 能源互联网基本构架示意图

力系统各节点的电流、电压、有功功率、无功功率和设备状态等信息；还要求数据平台可以进行紧急情况下的智能决策、超实时仿真和扰动识别等，以保证能源网络的稳定性、可靠性、鲁棒性、安全性。宏观上也要对中长期负荷电能需求、设备长期的运行状态和分布式电源数量进行计算、预测等。

从以上分析可知，实时数据和非实时数据并存，要求能源互联网的数据平台在高效处理海量数据的同时，也要兼具可靠的数据存储及批量处理数据的能力。

### 3. 能源互联网中的 Summingbird 平台

Summingbird 平台是 Twitter 实时消息处理平台，该平台每天要可靠、高效地处理 2 亿多用户的共 5 亿条推文，因此它具有很强的实时数据和批量数据处理能力。下面介绍该平台的结构原理，并阐述将其应用于能源互联网的优势所在。

在 Summingbird 平台出现之前，Hadoop 是大家较为熟悉的云计算平台，其优点是可以进行批处理，数据吞吐量较大，但是实时性较差，处理大规模历史性数据需要几个小时甚至几天时间。而 Storm 擅长进行流处理，快速响应并处理用户数据，但是数据的吞吐量不大。为了解决以上问题，将流式处理平台 Storm 与批处理平台 Hadoop 通过 API 连接并进行整合，取长补短，并基于 Lambda 建立了 Summingbird 云计算平台。其中 Lambda 是一个实时大数据处理构架，该构架分为三层：批处理层、服务层和加速层，它整合了离线计算和实时计算，并提出了融合不变性、读写分离和复杂性隔离等融合构架原则。

图 2 为基于 Lambda 框架的 Summingbird 计算平台。该框架中除了 Hadoop 和 Storm 两个数据处理模块，同时应用了其他的关键技术包括：持久数据存储 Kafka、分布式文件存储系统 HDFS、分布式数据库 HBase、本地 SQL 查询引擎 Impala。

1) Kafka 是一种分布式的存储器，存储方式是冗余存储，因此可以通过延时处理或再处理方式避免下游出现的错误，但是 Kafka 的数据存储不是永久的，而是几个小时(足够满足系统处理意外故障的时间)，当该存储器的内数据失去时效性就会被删除。

2) HDFS 是一种分布式数据存储系统，也是批处理层 Hadoop 的主要存储组件，以 key-value 的形式存储数据，具有高容错性特点。该系统对高吞吐量做了优化，较为适合 GB 级及以上大文件的读写，在 Summingbird 平台中主要负责批处理视图(Batch view)的存储。

3) HBase 是一个实时的、面向列的数据库，适合非结构化数据存储，可以根据系统的需求动态扩展数据存储空间，从而为实时视图(Online view)的存储提供了条件。

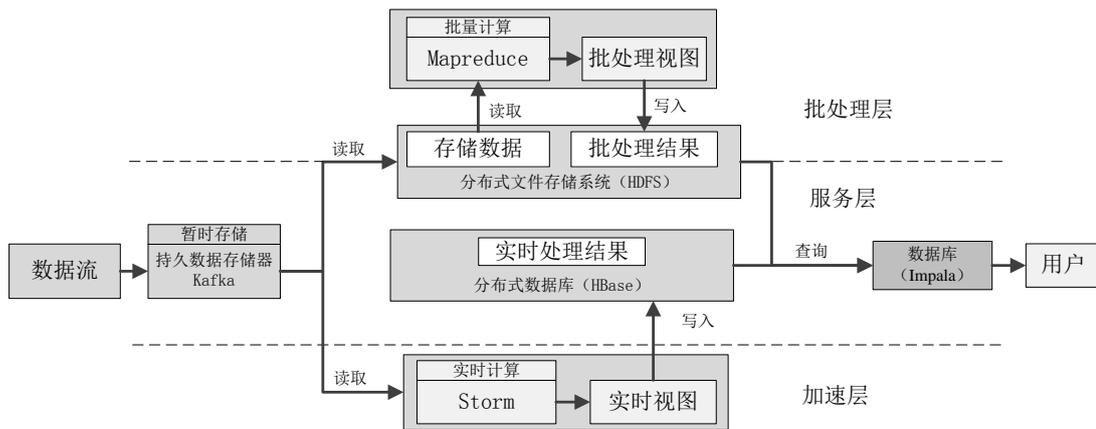


Figure 2. The Summingbird platform based on the Lambda framework

图 2. 基于 Lambda 架构的 Summingbird 云计算平台

4) Impala 是一个本地 SQL 查询引擎, 提供了一个面向批量和实时查询的统一查询数据库, 为用户提供快速查询服务, 可以对 HDFS 和 HBase 的数据提供直接查询, Impala 是 Summingbird 云平台的结果查询工具。

Summingbird 平台的具体工作原理是: 批处理层和加速层分别通过持续数据存储器 Kafka 读取数据。批处理层的 HDFS 存储器对数据进行永久存储, Mapreduce 由 HDFS 读取数据进行批处理, 并生成 Batch view, 然后服务层对 Batch view 的数据建立索引, 以 key-value 的形式存储到 HDFS 系统中, 以供用户查询。当批处理的计算速度不能满足用户的要求时, Summingbird 平台就调用流处理方式, 利用 Storm 对数据进行流式数据处理, 快速的生成 Online view, 同样建立对应的 key-value, 写入 HBase 数据库中, 以供用户实时查找, 此时的 Storm 相当于对云平台处理数据进行加速。最后的结果由 Impala 统一查询。将 Summingbird 数据平台应用于能源互联网, 具有如 2.1~2.3 所述优势。

### 3.1. 数据安全性

Summingbird 数据平台中的数据安全性是指数据存储安全和数据隐私安全。设备数字化智能化的能源互联网系统中, 能源消费、能源交易、设备控制等行为都依赖可靠地数据处理与存储。Summingbird 云平台中, 从数据传输到系统, 再到聚合计算以及结果存储的各个环节, 持续存储器 Kafka、分布式存储系统 HDFS 和分布式数据库 HBase 均使用冗余存储机制, 保证了数据存储的安全性。数据隐私安全方面, 能源共享主要依靠区域之间的能源需求信息实时交互, 在此过程中, 人们会担心机密或隐私数据信息的泄露。2015 年 Apache Storm 官方宣布采取身份验证协议 Kerberos、多租户调度、Hadoop 的安全整合和用户隔离机制等措施, 使 Summingbird 更加安全, 保障了能源互联网的信息数据的完整性和安全性。

### 3.2. 实时查询和数据挖掘

如前文所述, 能源互联网中产生实时数据和非实时数据, Summingbird 可以完全满足能源互联网对数据处理的要求。首先, Summingbird 的流式处理和批处理均为分布式计算, 将任务“化整为零”分配到集群的各个计算节点上, 快速完成计算任务, 实现能源互联网用户的实时查询需求。然后, 批处理层周期性的对数据进行挖掘, 为包括清洁能源在内的发电预测、用户个性化定制、负荷电能需求曲线和电能交易等提供数据支持。其次, Summingbird 的三大应用信息流处理、连续计算和分布式远程程序调用, 以及其它子项目将会为能源互联网提供更加高效数据处理服务。

### 3.3. 基于虚拟机的动态扩展能力

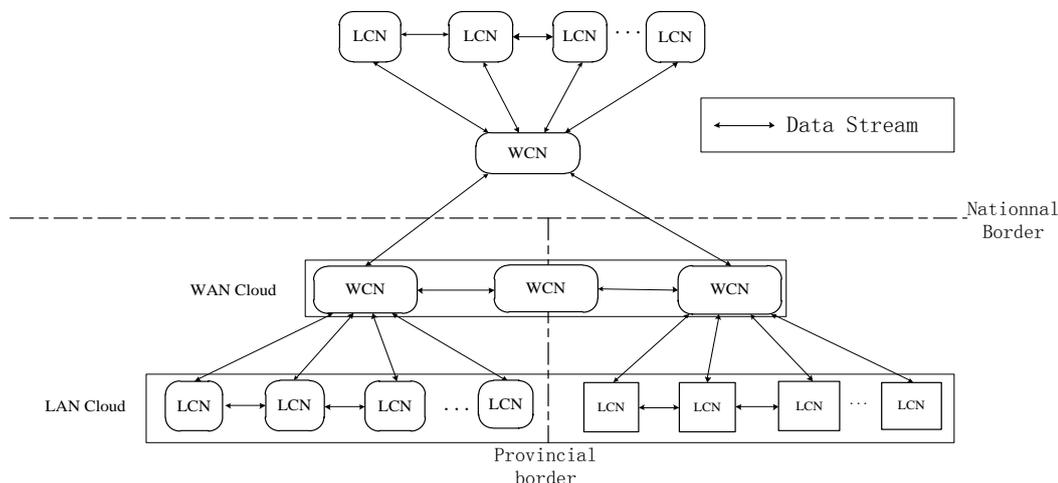
将来能源互联网中的分布式电源、储能设备和电动汽车等间歇性源荷设备数量动态变化, 数据处理平台也要能够动态扩展, 以满足数据处理需求。利用 Summingbird 云计算平台的 IaaS 层的虚拟化技术和 PaaS 层的弹性虚拟资源调度模型, 可以很方便地解决动态扩展问题。IaaS 层对内存、CPU、硬盘和网络等进行管理, 将分散的计算资源集中虚拟化后添加到资源池, 以供 PaaS 层调度; PaaS 层通过弹性均衡算法, 并根据各个虚拟机的计算负载情况, 实时的创建或销毁虚拟机, 从而实现云计算平台集群动态扩展。

由此可见, Summingbird 云计算平台能够满足能源互联网对数据计算和信息存储的基础要求, 可以为能源互联网的构建提供技术支持。

## 4. Summingbird 云计算平台构建

### 4.1. 云计算平台拓扑结构及功能

如图 3 所示的一个基于 Summingbird 云计算技术的“一网两云”云平台为能源互联网提供数据处理



**Figure 3.** The topology structure of cloud computing in energy internet  
**图 3.** 能源互联网中的云计算拓扑结构示意图

服务。根据能源互联网区域和层次的不同, 可将其划分为广域电网和局域电网, 相应地, 云平台的构建也相应的包括“广域云”和“局域云”两种。

在能源互联网中, 以“广域电网”(指区域及以上的电网)的数据中心为基础建立“广域云”(WAN Cloud), 以电能路由器为广域云节点(WAN Cloud Node, WCN)。根据分布式电源、储能设备和电动汽车等间歇性源荷的接入规模情况, 将配电网划分为多个局域云(LAN Cloud), 以电能交换器作为局域云节点(LAN Cloud Node, LCN)。根据该节点所管理区域的数据处理能力需求, 在 WCN 和 LCN 上装设相应数量的服务器, 并服务器上搭建 Summingbird 云计算集群, 以处理来自该区域的智能传感器、电能交换机和电能路由器等设备采集的数据。

“局域云”主要功能为局域电网中众多由分布式电源、储能设备组成的城市配电网或社区微网提供云计算服务, 例如: 电能双向管理、用户电能交易、提供定制用电方案、即插即用响应和社区负荷预测等方面的数据处理服务。协调该区域的源荷设备, 以使局域网成为一个能源自治单元。

“广域云”主要功能为能源互联网中的广域电网的源、网、荷、储的实时协调和电能调度提供云计算服务, 例如: 区域及以上电网之间的电能交换和共享、新型能源电能并网和企业电能交易、国际跨境电动交通工具电能计费。

如图 4 所示, 各个云服务器都要具备云计算的基础设施层、运行环境层、服务应用层和服务访问层。其中基础设施层、运行环境层为云计算平台的运行提供基本的软硬件环境; 服务访问层是为云计算平台用户提供与能源互联网的交互手段, 例如: 无线终端、Web 浏览器、专用客户端等; 业务应用层则要根据广域电网和局域电网的功能异同, 分别为广域云和局域云部署定制应用软件和通用应用软件。

## 4.2. 广域云对局域云的管理

局域云和广域云之间应建立高效的数据传输通道, 以使广域电网对局域电网之间能够实时通信。广域云对局域云的两个方面进行管理: 一方面是计算资源的监测, 另一方面是局域电网运行数据的分析。

### 1) 计算资源方面

局域云的计算能力具有局限性, 在电能使用高峰期时, 一些局域云可能因该区域的计算负荷过载而导致数据处理缓慢, 广域云要通过对其所管辖的局域云的计算资源进行实时的监测, 动态均衡各局域云的计算任务, 从而在提高能源互联网计算资源利用率的同时加快局域云的数据处理速度。



Figure 4. Basic architecture of cloud compute platform  
图 4. 云计算平台的基本层次示意图

2) 局域电网的运行数据分析

虽然局域电网是一个电能自治单元，但是当各局域电网的电能共享后，仍存在电能短缺或剩余时，就需要广域电网的对局域电网的电能进行实时的调度，此时广域云对局域云的相关数据统计分析，甚至结合能源互联网的其他广域网的电能需求数据，进行电能的优化调度，实现电能能源互联网中的实时分享。

两个云平台会增加能源互联网的建设投资，但该方案具有能源分享灵活性、信息处理实时性、控制系统鲁棒性等优点。

4.3. 应用实例

1) 场景一

能源互联网中，局域电网的电能用户可能拥有分布式新型能源和分布式储能等设备，并使用两类家用电器，一类是非灵活供电型家电，指在不能选择时间段供电，而要根据用户要求立即供电的家用电器，例如：电冰箱、电脑、电视机等；另一类是灵活供电型家电，指在电能盈余时使用电能，而电能缺额可以不安排供电的家用电器，例如：洗碗机、洗衣机、园林喷灌系统等。局域云利用 Summingbird 平台根据实时的天气预测情况，预测一定时间(小时为单位)内清洁能源发电量，同时对用户电能使用习惯进行分析，结合习惯分析和预测发电量，制定该时间段内的电能使用方案。在发电高峰期时，充分发挥用户家用电器对清洁能源的消纳能力，优先安排灵活供电型家电进行工作，其次是存储电能备用，无法消纳的电能通过局域电网或广域电网与其他电能缺额用户分享电能。

2) 场景二

实时电价是指在很短时间(以小时为单位)对电能用户的用电量进行定价。能源互联网中，电能作为一种商品，电价高低是电能交易的关键指标之一，电能需求影响实时电价。峰谷电价已经不适合能源多样化的能源互联网，能源互联网的实时电价的制定要考虑广域网中的集中式新型能源发电量、化石能源

发电量和局域电网电能需求和经济运行成本等方面的因素, 建立能源互联网经济运行模型。因此各“局域云”利用 Summingbird 对管辖区域的负荷历史数据进行分析, 对下一个小时的负荷进行预测, 负荷预测数据提交到广域云。广域云中的 Summingbird 平台对广域电网中的可利用电源进行实时统计, 然后根据统计结果和经济成本及环境成本等因素, 对电力系统潮流进行超实时仿真, 快速生成潮流优化结果, 并将之作为下一个小时电能调度方案的依据, 最后根据调度方案对电能用户提供基准参考电价, 构建一个开放透明的电能交易平台。

## 5. 结论

根据能源互联网的数据特点, 引用 Summingbird 数据处理平台作为能源互联网的主要云计算平台。同时考虑到能量互联网的特点, 提出了“一网两云”的基于云计算平台的能源互联网的网络架构模型。

当然, 作为初步构建的 Summingbird 云计算平台, 其具体应用方面还不太成熟, 特别是数据安全的技术方面有待于提高, 但 Apache Storm 正在建立相关的项目对该平台进行改进与完善, 相信未来的云计算平台会为能源互联网提供优质的数据处理服务, 实现信息系统和物理设备的深度融合, 发挥其应有作用。

## 参考文献 (References)

- [1] Rifkin, J. (2013) *The Third Industrial Revolution: How Lateral Power Is Transforming Energy, the Economy, and the World*. Palgrave Macmillan, New York, 31-46.
- [2] Crow, M.L. (2010) Intelligent Energy Management of the FREEDM System. *IEEE Power and Energy Society General Meeting*, 1-4. <http://dx.doi.org/10.1109/pes.2010.5589992>
- [3] Huang, A.Q. (2009) Renewable Energy System Research and Education at the NSF FREEDM Systems Center. *IEEE Power and Energy Society General Meeting*, Calgary, 7, 1-6.
- [4] Vermesan, O., Blystad, L.-C., Zafalon, R., Zafalon, A., et al. (2011) Internet of Energy-Connecting Energy Anywhere Anytime. In: Meyer, G. and Valldorf, J., Eds., *Advanced Microsystems for Automotive Applications*, Springer-Verlag.
- [5] Boyd, J. (2013) An Internet-Inspired Electricity Grid. *IEEE Spectrum*, 50, 12-14. <http://dx.doi.org/10.1109/MSPEC.2013.6395291>
- [6] 孙宏斌, 郭庆来, 潘昭光. 能源互联网: 理念、架构与前沿展望. *电力系统自动化*, 2015, 39(19): 1-8.
- [7] 曹军威, 孟坤, 王继业, 杨明博, 陈震, 李文焯, 林闯. 能源互联网与能源路由器[J]. *中国科学: 信息科学*, 2014, 44(6): 714-727.
- [8] 田世明, 栾文鹏, 张东霞, 梁才浩, 孙耀杰. 能源互联网技术形态与关键技术[J]. *中国电机工程学报*, 2015, 35(14): 3482-3494.
- [9] Xu, Y., Zhang, J.H. and Wang, W.Y. (2011) Energy Router: Architectures and Functionalities toward Energy Internet. *IEEE International Conference on Smart Grid Communications*, Brussels, 31-36.
- [10] 盛万兴, 段青, 梁英, 孟晓丽, 史常凯. 面向能源互联网的灵活配电系统关键设备与组网形态研究[J]. *中国电机工程学报*, 2015, 35(15): 3760-3769.
- [11] 杨方, 白翠粉, 张义斌. 能源互联网的价值与实现架构研究[J]. *中国电机工程学报*, 2015, 35(14): 3495-3502.
- [12] Boykin, O., Ritchie, S., O'Connell, I. and Lin, J. (2014) Summingbird: A Framework for Integrating Batch and Online Mapreduce Computations. *Proceedings of the VLDB Endowment*, California, 13, 1441-1451.
- [13] Alex Holmes, 著. 梁李印, 宁青, 杨卓萃, 译. *Hadoop 实战*[M]. 北京: 电子工业出版社, 2015.
- [14] 安德森, 著. 卢誉声, 译. *Storm 实时数据处理*[M]. 北京: 机械工业出版社, 2014.
- [15] 许辉阳, 李劼, 罗霄翔. 面向业务的云计算 IaaS 研究[J]. *移动通信*, 2011(9): 20-24.
- [16] 徐新坤, 王志坚, 叶枫, 邱振瑜. 一个基于弹性云的负载均衡方法[J]. *微电子学和计算机*, 2012, 29(11): 29-32.