



基于数据挖掘的电力云资源规划调度

庞松涛

(浪潮集团 北京 100142)

摘要:随着虚拟化技术的应用,数据中心的资源利用率已经得到一定程度的提高,但是云资源通常还是根据用户提出的需求预先分配,资源利用率仍然有待提高。为了进一步改进资源利用率,云中心的实际资源需求可分为“周期资源”及“峰期资源”。“周期资源”可以通过历史规律,利用递阶成分负载模型进行分析,预测出数据中心大部分时间的资源需求;“峰期资源”主要是满足短时间内的各种高峰资源需求,这种需求应用了排队论及随机均衡算法模型并根据实际应用的需要动态启用、分配和收回资源。通过使用资源收集与分配守护进程对云中心资源需求进行规划调度实验,效果明显,从而为云中心资源利用率的提升及节能减排提供了一种有效途径。

关键词:数据挖掘;智慧电力云;云资源调度

doi: 10.11959/j.issn.1000-0801.2015084

Resource Scheduling for Cloud Data Center Based on Data Mining in Smart Grid

Pang Songtao

(Inspur Group Co., Ltd., Beijing 100142, China)

Abstract: Since the wide use of virtual technology, the resource use rate in cloud data center has been improved effectively than ever before. However, there is still a large space for improvement due to the resources usually are pre-started and pre-allocated by the user demand rather than the actual needs. In order to allocate available resource more accurately, two algorithms were proposed to meet the needs of the daily use in most of time. The available virtual resources would be arranged according the forecast using the algorithms of hierarchical composition of loading and the peak resources needs would be dynamic allocated using the algorithms of stochastic equilibrium and queuing theory. The results of experiment via the system based upon above theories show that the solution provides a kind of very effective advanced means for the optimal use of resources and energy saves.

Key words: data mining, smart grid cloud, cloud resource scheduling

1 引言

随着智能电网的发展,区域内各地区网络连接愈加紧密,电网计算不但需要高性能,同时也越来越需要跨区域、

跨电压等级的协作,并共同构建电网的基础数据,完成各类分析计算任务。传统采用各网省专业技术人员集中工作的方式,拼合调整形成的运行数据,缺乏协同机制,工作效率不高且数据是由各级部门分散维护,定期合并汇总,其

收稿日期:2014-09-15;修回日期:2015-03-11

论文引用格式:庞松涛. 基于数据挖掘的电力云资源规划调度. 电信科学, 2015084

Pang S T. Resource Scheduling for Cloud Data Center Based on Data Mining in Smart Grid. Telecommunications Science, 2015084



及时性和完整性较差。而应用电力云计算中心,可以有效实现全网数据的统一管理、资源的按需分配以及全局数据的智能分析应用等。

合理调配云计算资源可以有效地减少云计算平台的运行成本。近年来,随着云计算基础架构的扩展,能源消耗问题引起越来越多的关注,据统计,数据中心的能耗成本占到了运维总成本的 50% 左右,包括电力、机房占地、带宽等各方面。在产生高能耗的同时,传统数据中心中 80% 的服务器 80% 的时间其 CPU 占有率低于 20%, 资源存在极大的浪费。国家电网已经建立了北京亦庄、上海、陕西三大数据中心,覆盖了我国北方、南方以及我国西部内陆地区,随着国家智能电网与特高压输电工程的进一步推进,这些数据中心将为我国电力安全与电力信息化的长足发展奠定基础。所以,实现数据中心的资源优化管理,建立智慧电力云具有实际意义。基于上述原因,本文重点研究如何优化云资源利用,动态启动关闭虚拟设备,在不影响云服务的前提下合理提供云资源并尽可能节省电力能源消耗。

2 云资源规划调度优化

云计算将计算任务分布在资源池上,各种应用系统能够根据需要获取计算力、存储空间和各种软件服务。原来的一台物理服务器可以虚拟成多台虚拟服务器,虚拟化后的资源变得更可量化。虚拟化技术将底层的计算资源切分或合并成多个运行环境,以软件的方式模拟硬件,通过软件的方式逻辑切分服务器资源,形成统一的虚拟资源池,创建虚拟机运行的独立环境。这种逻辑结构提供了灵活、可变、易配置、可扩展的平台服务,并且可以实现灵活有效的分布存储和计算,从而整体上为实现强大的计算和海量数据存储能力提供基础保障。为简化起见,本文以 CPU 资源作为分析探究对象。

与传统方式相比,云计算已经在更小粒度上管理分配资源,改善云资源利用及能耗,但是虚拟资源管理如果完全按照云用户申请启用云,还会存在大量的活动资源被闲置,资源及能耗依然还存在浪费,因为用户对资源把握不准,申请资源往往都超过实际需要。要做到更有效地利用资源,需要对云资源需求有更准确的把握,并提前做好资源规划。云计算的容量需求随着时间的推移而逐渐增加,总需求并不是简单地单调增加,而是存在一定程度的波动。为了尽可能避免闲置资源的浪费,同时保证云服务质量,云资源被划分为两部分,第一部分称为“周期资源”,主

要是为了满足云中心大部分资源需求,保证云中心平稳运营,该资源需求可根据历史数据分析找出,用于满足云中心相对平稳的云资源需求;第二部分资源称为“峰期资源”,该资源主要满足各阶段高峰资源需求。通过资源划分,可以更加精细管理云资源,利用周期资源应对基本需求,再根据峰值需求,动态启用或关闭峰期资源。当然,虚拟资源的启用和关闭需要时间,而且资源启动及关闭对应用系统影响程度不同,所以在实际应用中,可根据用户应用系统的实际情况,调整资源管理的时间粒度。基于数据挖掘的云资源规划调度技术框架如图 1 所示。

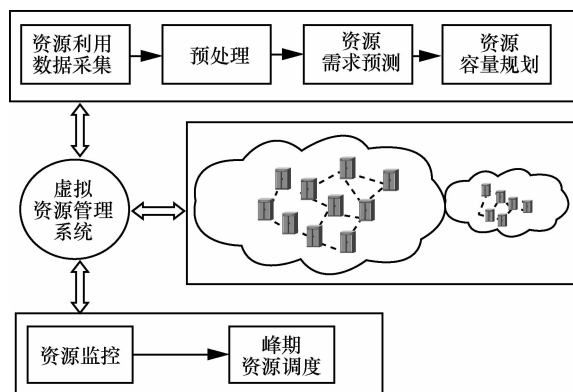


图 1 云资源规划调度技术框架

资源规划调度原理步骤说明如下:

- 通过虚拟资源管理系统采集云资源利用的历史数据;
- 通过预处理生成预测模型可用的数据;
- 利用数据挖掘技术分析资源利用规律;
- 利用分析预测的结果规划资源需求,并区分资源周期需求和峰期需求;
- 利用虚拟化管理系统分阶段配置并启用虚拟资源,保证在非资源高峰需求阶段提供正常云服务;
- 系统通过虚拟管理系统实时/及时采集云资源利用信息;
- 当周期云资源利用率超过指定的阈值,提前启动规划好的后备峰期资源,为了更好利用资源,峰期资源可根据负载的增加而逐步启用;
- 利用相关算法为新增负载分配虚拟资源。

本文在以下部分进一步给出了实现方法。

3 云周期资源需求规划方案

首先,需要尽量对云周期资源需求进行准确估计。指数平滑算法是目前常用的一种时间序列分析预测法。该方法可通过计算指数平滑值,配合一定的时间序列预测模型

对某个变量的未来变化进行预测。通过分析,把电网信息系统云资源需求分为以下组合成分。

- 长期趋势:系统负载时间序列随时间的变化具有逐渐增加的长期变化趋势。
- 季节变动:时间序列的季节变动指其在固定时间内呈现出的有规律的变动,电网系统负载的季节变动具有递阶特性,一般其变动规律可以分为周和日两个层次。
- 不规则变动:即电网系统负载序列中由于随机因素影响所引起的变动。

基于以上分析,参照 Holt-Winters 平滑算法,本文提出一个递阶成分负载模型,它可表示为: $y_{i,j,k}=w_i d_{7+j} h_{56+8+j+k} + \delta_{56+8+j+k}$ ($i=1,2,3,\dots, j=0,1,2,\dots,6, k=0,1,\dots,7$)。

其中, w_i 代表第*i*周的负载量, d_{7+j} 代表第*j*周的负载量在第*j*日的分布系数,即当日访问量相对于整周访问量的百分比, $h_{56+8+j+k}$ 代表第*i*周第*j*日的负载量在第*k*时段的分布系数,即该段时内访问量相对于整日访问量的百分比, $\delta_{56+8+j+k}$ 代表第*i*周第*j*日第*k*时段的随机扰动,其均值为零。在本文模型中,每天被分为8个时段。

根据以上资源需求的特点, w_i 变化将具有长期趋势,而 d_{7+j} 与 $h_{56+8+j+k}$ 的变化则主要呈现季节规律。为了形象说明上述特点,给出一组虚拟负载数据及其规律分解,分别如图2和图3所示。

综上所述,可以采用二次指数平滑法对周访问量进行预测。具体算法及推导如下。

令 $s_i^{(1)}$ 和 $s_i^{(2)}$ 分别代表 w_i 的一次和二次指数平滑值,满足:

$$s_i^{(1)} = a\bar{w}_i + (1-a)s_{i-1}^{(1)}, \quad s_i^{(2)} = as_i^{(1)} + (1-a)s_{i-1}^{(2)} \quad (1)$$

其中, $\bar{w}_i$ 为 w_i 的无偏估计值,取为: $\bar{w}_i = \sum_{j=0}^6 \sum_{k=0}^7 y_{i,j,k}$ 。

由以上可推出:

$$\sum_{j=0}^6 \sum_{k=0}^7 y_{i,j,k} = w_i + \sum_{j=0}^6 \sum_{k=0}^7 \delta_{56+8+j+k} \quad (2)$$

可以假定: $\delta_{56+8+j+k}$ 的均值为零,故有 $E\bar{w}_i = w_i$ 。

另一方面,若以如式(3)所示的方程描述周访问量的趋势:

$$\hat{w}_{i+m} = a_i + b_i m \quad (3)$$

其中, a_i, b_i 是参数变量, m 是从*i*周开始的预测周数,则:

$$a_i = 2s_i^{(1)} - s_i^{(2)}, \quad b_i = \frac{a}{1-a} (s_i^{(1)} - s_i^{(2)}) \quad (4)$$

进一步假定 $\exists t_1, t_2 > 0$,使得:

$$d_{i+7\tau} = d_i, \quad \forall \tau < t_1, \quad h_{i+8\tau} = h_i, \quad \forall \tau < t_2 \quad (5)$$

因此,可利用一次指数平滑法对它们进行预测:

$$\hat{d}_{i+7} = a\bar{d}_i + (1-a)\hat{d}_i, \quad \hat{h}_{i+8} = a\bar{h}_i + (1-a)\hat{h}_i \quad (6)$$

其中, $\bar{d}_i$ 为 d_i 的估计值,可以取为: $\bar{d}_{7+j} = \frac{\sum_{k=0}^7 y_{i,j,k}}{\bar{w}_i}$

($i=1, 2, 3, \dots$)的最小二乘均值即满足离差 $(\bar{d}_i - \bar{d}_{7+j})$ 的平方和 $\sum (\bar{d}_i - \bar{d}_{7+j})^2$ 最小的 $\bar{d}_i$ 值。 $\bar{h}_i$ 为 h_i 的估计值,可以取为: $\bar{h}_{56+8+j+k} =$

$\frac{y_{i,j,k}}{\bar{w}_i \bar{d}_j}$ ($i=1, 2, 3, \dots$)的最小二乘均值即满足离差 $(\bar{h}_i - \bar{h}_{56+8+j+k})$ 的平方和 $\sum (\bar{h}_i - \bar{h}_{56+8+j+k})^2$ 最小的 $\bar{h}_i$ 值。

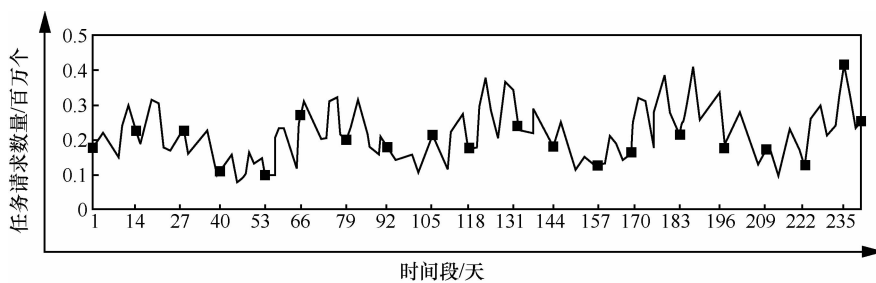


图2 虚拟负载数据

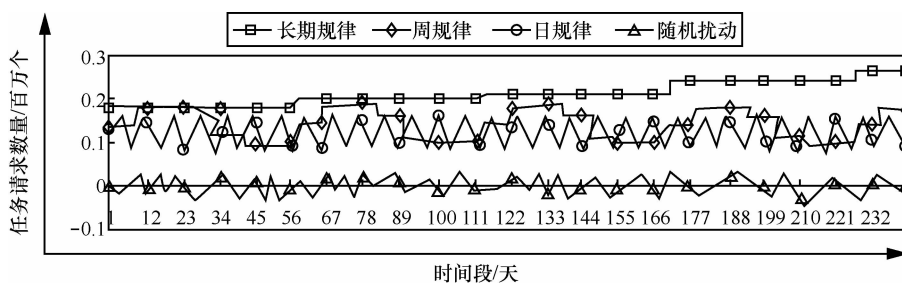


图3 负载规律分解



以上估计值可保证:

$$E\bar{w}_i\bar{d}_{7ij}=E\sum_{k=0}^7Y_{ij,k}=w_id_{7ij} \quad (7)$$

$$E\bar{w}_i\bar{d}_{7ij}\bar{h}_{56i+8j+k}=EY_{ij,k}=w_id_{7ij}h_{56i+8j+k} \quad (8)$$

4 云峰期资源需求调度方案

峰期资源主要用于“调峰”,即主要用于满足各阶段云资源的需求高峰。云虚拟机实例的启动关闭遵循原则是:当周期资源利用率达到给定阈值时,启动新的虚拟机实例,而周期资源利用率少于给定阈值时,终止虚拟机实例。当然启动和关闭操作不能太过于频繁,要留足够的时间,启停频率可以根据实际应用设定阈值使得频率控制在小时级别。另外,由于增加的资源需求可能来自不同应用、不同任务,所以也需要通过合理的算法进行资源分配。本文采用随机均衡策略。在此引入一个简单模型。

首先,假定每个节点上只运行一个服务组件;每个用户请求对应一个组件实例,所有实例的优先级相同;用户请求到达时间与服务时间都满足泊松分布;平台因任务调度而消耗的资源可以忽略,则 $\forall r>0$,当 Δt 足够小时,任意节点正在服务的用户数从 r 转移到 $r-1$ 、 r 、 $r+1$ 的概率与单队列单服务台简单服务系统(M/M/1)中的情况完全相同,如图4所示。

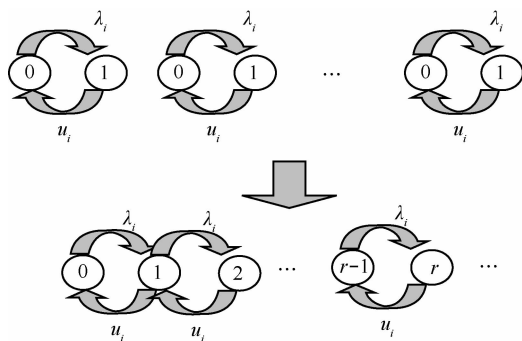


图4 节点状态转移示意

根据排队论的研究结果推出最佳随机均衡策略是按到达率平方根的比例把整个系统的冗余服务率分配到各个节点。

峰值变化对资源需求做出调整,其实时性要求相对较高,为此在虚拟资源管理监控系统上运行两个守护进程 negotiator 和 collector,其中 collector 负责收集负载及任务队列信息,并对上述信息列表进行周期性更新,negotiator 负责新增虚拟资源的分配,同时完成以下各步工作。

(1)negotiator 采集当前资源利用信息。

(2)发现资源利用超过阈值时,发送资源需求到任务队列。

(3)negotiator 确定更新各资源请求任务的优先权。优先权随等待排队中的任务数增加而增加。

(4)negotiator 从具有最高优先权的资源请求开始查询,并收集等待任务所需的资源信息。

(5)negotiator 选择满足要求的虚拟资源,分配给需要的应用。

negotiator 将重复以上步骤,直到为所有的等待任务找到合适的虚拟资源为止。

5 实验与小结

为了得出实验结果,在云中心配置了一个实验环境,该实验环境以3台8路服务器TS850为主,外接一个集中存储服务器,8路服务器和集中存储之间用SAN交换机连接。8路服务器上安装虚拟化软件,生成的虚拟机外挂存储服务器上划分的虚拟盘,最后安装了浪潮云海OS,该软件可以实现跨平台的云资源管理及实时监控,通过浪潮云海OS,首先对某一云中心采集了周、日的云资源运行数据,通过对周访问量、日访问量依次进行不同粒度层次的平滑计算,获得未来某个时段的负载预测值。把前3周的负载数据作为观测值,用它们对第4、5周各个时段的负载进行预测,并将预测结果与观测值(实际上是根据模型产生的计算值)进行比较。图5为每周负载量及其预测值(前3周为观测值,后2周为预测值)之间的比较,图6和图7分别为周负载分布、日负载分布与它们的预测值之间的比较,图8为根据模型产生的附载计算值与按照上述方法得出的负载预测值之间的比较。

相比 Holt-Winters 平滑算法,本文针对电力系统负载的特点,将每一个层次的变化规律分而治之,不仅能够充分反映负载变化规律的递阶循环特性,预测过程也比较简单,每一个层次只需确定一个 α 参数。

另外,在峰期资源调度上,在云海OS的技术上增加服务节点控制与状态信息收集器。该服务由一个主控台和一个后台探测器线程组成,服务对象/组件列表放在 negotiator 中,并通过浪潮云海OS对上述列表进行操纵,从而实现对各节点上的服务器(包括云端的虚拟服务器)进行激活(activate)、去活(deactivate)等控制。在实验环境中,虽然可以根据服务任务需求的情况自动启动或关闭资

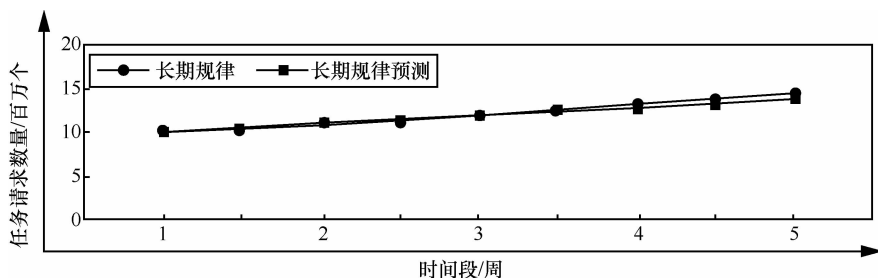


图5 周负载量及其预测值

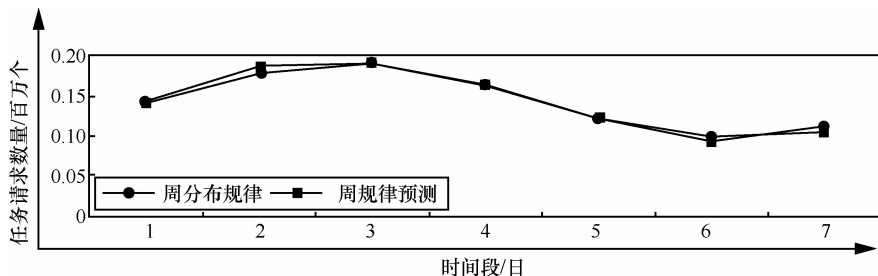


图6 周负载在各日的分布及其预测

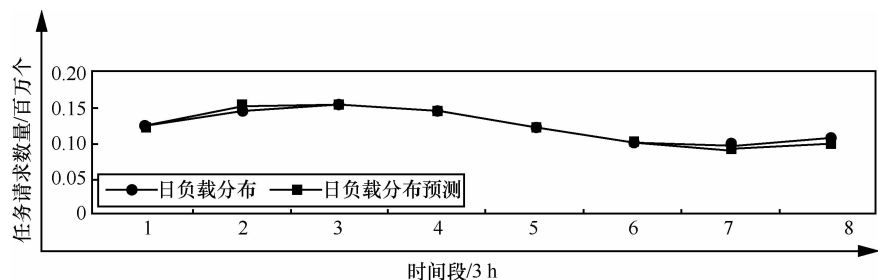


图7 日负载在各时段的分布及其预测

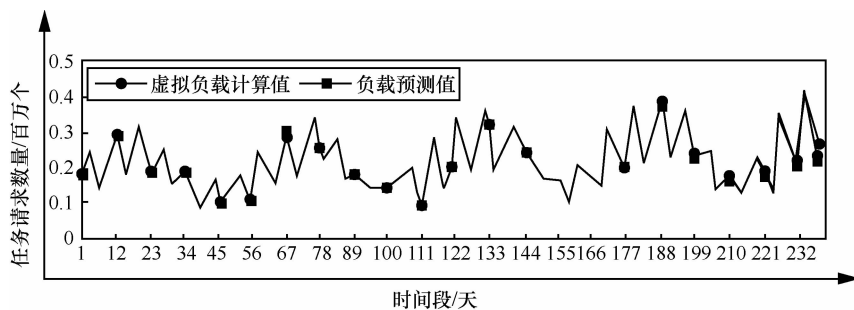


图8 虚拟负载计算值及其预测值

源,但是该资源调度还需要和实际应用相结合,可以针对具体应用,进一步实验得出合理的资源启动关闭频率。

总之,传统通过用户对资源的需求来开启云计算资源的方式在资源利用上还是比较浪费,通过实验结果得出,对云资源利用的历史数据进行挖掘,找出云资源的利用规律,不仅可以实现按需分配资源,进一步可以更好地做到按需开启关闭峰期资源,从而更好地实现云资源的充分利用,建立生态云数据中心。由于电力系统对云的资源需求有比较明显的规律,所以通过数据挖掘,资源预测及规划

相对比较准确。为此,给大规模电力云数据中心带来的经济效益会更加明显。

参考文献

- 1 张大华,丁辉,吴向阳等. 面向智能电网的电力云计算架构. 电力通信管理暨智能电网通信技术论坛论文集, 北京, 中国, 2011
Zhang D H, Ding H, Wu X Y, et al. Cloud computing architecture for smart grid. Proceedings of Electric Power Communication Management and Smart Grid Communications Technology Forum, Beijing, China, 2011



- 2 王利赛, 杨明玉, 孙月琴等. 电力云研究综述. 电力信息化, 2011(5)
Wang L S, Yang M Y, Sun Y Q, *et al.* Summary research on electric power cloud computing. Electric Power Information Technology, 2011(5)
- 3 王鹏. 云计算的关键技术与应用实例. 北京: 人民邮电出版社, 2010
Wang P. Examples of key techniques and application on cloud computing. Beijing: Posts and Telecom Press, 2010
- 4 徐国祥. 统计预测和决策. 上海: 上海财经大学出版社, 2011
Xu G X. Statistical forecast and decision making. Shanghai: Shanghai University of Finance and Economics Press, 2011
- 5 陆传赓. 排队论. 北京: 北京邮电大学出版社, 2009
Lu C L. Queuing Theory. Beijing: Beijing University of Posts and Telecommunications Press, 2009
- 6 Kalekar S P. Time series forecasting using holt-winters exponential smoothing. Kanwal Rekhi School of Information Technology Tech Rep, 2004
- 7 Dostal Z. Optimal Quadratic Programming Algorithms: With Application to Variations to Variational Inequalities. New York: Springer, 2009
- 8 Jiang Y, Perng S C, Li T, *et al.* ASAP: a self-adaptive prediction system for instant cloud resource demand provisioning. Proceedings of ICDM, Vancouver, Canada, 2011
- 9 杨巨龙. 大数据技术全解. 北京: 电子工业出版社, 2014.
Yang J L. Big Data Technology Entire Solution. Beijing: Publishing House of Electronics Industry, 2014
- 10 苏跃进, 韩英杰. 基于分布式电站群的智能数据分析中心建设与实践. 电信科学, 2013(11): 122~126
Su Y J, Han Y J. Construction and practice of intelligent data analysis center based on distributed power plant. Telecommunications Science, 2013(11): 122~126
- 11 耿亮, 吴燕, 孟宪楠. 电力数据挖掘在电网内部及各领域间的应用. 电信科学, 2013(11): 127~130
Geng L, Wu Y, Meng X N. Application of power data mining within and between grid fields. Telecommunications Science, 2013(11): 127~130
- 12 赵云山, 刘焕焕. 大数据技术在电力行业的应用研究. 电信科学, 2014(1): 57~62
Zhao Y S, Liu H H. Research on application of big data technique in electricity power industry. Telecommunications Science, 2014(1): 57~62
- 13 黄章树, 刘晴晴. 基于云计算服务模式的数据挖掘应用平台的构建. 电信科学, 2012(1): 53~57
Huang Z S, Liu Q Q. Construction of data mining application platform based on the service models of cloud computing. Telecommunications Science, 2012(1): 53~57

[作者简介]



庞松涛,男,历任卡西欧浪潮电子有限公司通信事业部副部长,山东浪潮商用系统有限公司市场部部长,浪潮电子信息产业股份有限公司西南大区总经理、西北大区总经理,浪潮电子信息产业股份有限公司市场推进部总经理、渠道发展部总经理、市场管理部总经理等职,现任浪潮集团副总裁。从事云计算、服务器及行业管理工作。

2015084-6

2015年1月通信业主要指标完成情况(二)

指标名称	单位	本月末到达	比上年末净增	本月净增
固定电话用户合计	万户	24 814.6	-128.5	-128.5
城市电话用户	万户	17 956.6	328.6	328.6
农村电话用户	万户	6 858	-457.1	-457.1
移动电话用户合计	万户	12 9048.5	439.1	439.1
其中:3G用户	万户	48 504.5	-21	-21
4G用户	万户	11 787.2	2 058.9	2 058.9
互联网宽带接入用户	万户	20 192.8	144.5	144.5
其中:xDSL用户	万户	8 680.2	-258.5	-258.5
移动互联网用户	万户	88 014.1	492	492
固定电话普及率	部/百人	18.3		
移动电话普及率	部/百人	94.5		

(来源:工业和信息化部运行监测协调局)