

“高端容错计算机系统关键技术与应用”项目由浪潮集团完成,项目成果为天梭K1系统,这是近年来计算机领域获得的第一个国家科技大奖。

让关键数据跑在自主平台上

四项核心突破



“一个跳步,64路互联”

浪潮集团设计实现了“双翼可扩展多处理器紧密耦合共享存储体系结构”,解决了多处理器的互联问题,可支持64颗处理器。该技术是国际上唯一一个只需要一级跳步的64路互联系统,获得全球PCT专利。

在体系结构研究领域,数据经过一个协同芯片就称为一个跳步。跳步越多,系统的延迟也就越多,最终造成“内存墙”问题,这是大型服务器系统面临的普遍性难题,也是制约大型服务器线性扩展的瓶颈,一级跳步的全互联保证了系统具有良好的线性扩展性。处理器间的互联通道能够随着处理器数量的增长而不断增长,而且整体互联结构相对简单。

浪潮还创建了“三级目录二级缓存一致性协议”,保证处理器数量增加时,系统通讯规模和复杂度仅有较小增加,上述两项技术结合起来,使得系统性能可以近线性提升,达到了国际领先水平。

处理器协同芯片组

处理器协同芯片组,是计算机中唯一复杂度可与处理器相比的芯片部件。

Intel目前最高端的通用处理器至强E7集成的晶体管数量为14亿个,引脚数1567个。

浪潮研制的芯片组集成了4.5亿个晶体管,有2577根管脚,在中国没有集成电路产业,没有设计生产如此规模芯片的先例的情况下,该芯片达到了国际领先水平:支持64路处理器互连;报文处理能力40亿条/秒;聚合带宽达到1056Gbps;延迟小于200纳秒。

软硬一体的容错技术

浪潮创建了多层次冗余与故障管理容错系统模型。

天梭K1在芯片、模块、固件、操作系统、业务应用层面全面采用冗余设计以杜绝单点故障,采用故障诊断、隔离和恢复的全流程自动处理技术,系统可自我修复,整体可用度超过99.9994%。

而且天梭K1的容错技术经受了实际应用的检验,2010年8月,第一台天梭K1样机在建设银行新疆分行启动上线测试,承担中间业务系统,成功替代国外产品。

中国首款通过Unix 03认证

浪潮研制成功了中国第一款通过Unix 03认证的操作系统K-UX,通过UNIX 03认证,意味着K-UX进入了关键应用软件俱乐部。

Unix 03全称为Open Group Unix 03认证,该认证是关键应用软件的入门牌。目前,Unix依然是关键应用软件的事实标准,不仅关键应用软件大部分是基于Unix标准开发,而且关键应用软件在Unix平台运行时间也最长、最成熟,因而,要开拓关键应用市场,打入Unix关键技术生态圈是必须的前提条件,但是该认证复杂大,门槛高,目前只有IBM、HP、苹果和Oracle四家厂商的产品通过了该认证。

在研制过程中,浪潮开发3200余项Unix特性,成功通过The Open Group Unix 03国际标准化认证60000余项测试,并对该标准提出七项修改,全部被纳入新的测试标准。

本版文章由本报记者刘燕撰写

高端容错计算机是一类高性能、高可靠的高端服务器,是信息化建设不可或缺的重要基础装备。天梭K1使中国成为继美国、日本之后第3个有能力研制32路高端容错计算机的国家,让中国的关键数据可以跑在自主平台上,是我国信息安全自主可控战略布局的重要技术成果。

关系着一个时代的核心装备

移动互联网、云计算、大数据等新一代信息化技术已经为人耳熟能详,这些技术应用的发展正在把人类社会从工业社会推向下一个社会形态——信息社会,信息社会与工业社会最大的区别就是——信息将成为继劳动力、资本和原材料之后的第四个生产要素,信息的流通效率将决定社会经济的运行效率。

目前,信息化已经对社会各领域高度渗透,这种趋势不仅体现在微信、门户、搜索等人们日常生活中接触最多的互联网服务,更体现在行业信息化领域。

行业信息化是支撑一个社会运行的核心基础设施,目前,银行、电信、证券、保险、电力等在国民经济和社会运行中具有基础性作用的关键行业,已经高度依赖行业的信息化系统,从银行存取款、打电话、乘坐飞机火车,到电力调度、公共安全保障等一切个人和经济社会活动都离不开行业信息化系统的支持。

高端容错计算机是驱动各个行业数据和信息流动的“心脏”,在社会信息流动中具有基础性地位。各个行业信息化系统中关键模块和核心数据处理模块都运行在高端容错计算机上,比如银行的储蓄业务系统、证券的交易系统和报价系统,电信领域的计费系统,能源领域里的电力调度系统,民航领域的机场进出港系统等,高端容错计算机一旦出现问题,往往会严重伤害用户的实际业务,甚至会给社会造成严重影响。

计算机产业皇冠上的宝石

高端容错计算机是现代计算机产业皇冠上的宝石,



产业技术门槛高于PC、手机和一般服务器,耗资巨大,风险大,周期长。例如IBM针对中低端高端容错计算机的eX5架构,研制费用就高达8亿美元,研制周期长达5年,超出了一般企业、甚至政府的承受范围,这也是中国主机产业空白的重要原因。

高端容错计算机是市场利润的主体,也是整个软硬件产业的核心,中国信息企业要实现转型升级,没有高端容错计算机技术的支持几乎是不可能的。

IT产业巨头都有完善的高端容错计算机技术,例如IBM的power、HP的superdome等,并且以此为基础形成了封闭而成熟的技术系统,持续获得高额利润。根据IBM公布的财报,其高端容错计算机在2010年—2011年营收分别增长16.4%和8.4%,IBM硬件业务在产业利润日趋微薄的大形势下,仍然取得了高达

38.5%利润率。

清除核心层面的信息安全隐患

欧美日等发达国家一直将高端容错计算机相关技术作为国防战略技术,对中国实行严格的技术封锁。美国的《出口管理条例》(EAR)和瓦森纳组织,高端容错计算机的核心技术都在禁运名单中。

一方面是技术封锁,另一方面是产品大量向中国倾销。长期以来,金融、电信、能源的核心系统都运行在国外的平台上,这意味着进口产品的技术完全不可控,对于用户来说,就是一个黑洞,存在的后门、逻辑陷阱或其他恶意程序,难以发现和清除。如果有人蓄意利用后门谋求达成非正当目的,尤其是在有政府介入的情况下,

自主产业推动国家信息战略布局

高端容错计算机获国家科技进步一等奖

1月9日,2014年国家科学技术奖励大会在人民大会堂举行,国家电网智能电网创新工程、高端容错计算机、H1N1大流行防控技术等一批对国家长期科技发展和国民经济全局具有重要意义的创新成果分别被授予国家科技进步一等奖。

其中,“高端容错计算机项目”是十一五期间国家863计划重大专项,由浪潮集团承担。2010年8月,浪潮研发成功我国首台具有自主知识产权的高端容错计算机,并将其正式命名为浪潮天梭K1系统。该产品让中国成为继美国、日本之后第3个有能力研制32路高端容错计算机的国家,中国的关键数据从此可以运行在自主平台上,为我国信息化建设自主可控战略布下关键一局。

当前,信息化已经对社会经济各个领域高度渗透,成为支撑各个行业正常运行的基础设施,高端容错计算机是各个行业核心系统,尤其是交易数据处理不可替代的承载平台,在金融、电信、电力等行业应用广泛,是驱动各个行业数据和信息流动的“心脏”。此前,我国高端容错计算机全部依赖进口,进口产品技术完全不可控,使我国关键行业面临数据被窃取、业务被摧毁的风险,

严重威胁国家战略安全。随着天梭K1对进口产品的不断替代,核心平台层面的隐患将被逐步解决。国产高端容错计算机产业的突破和发展为国家信息安全自主可控奠定坚实的基础。

天梭K1系统是中国第一台高端容错计算机,可扩展32颗处理器,可用性超过99.9994%,每分钟完成几百万次交易事务处理,其上运行的K-UX是全球第五个、国内唯一通过UNIX 03认证的UNIX操作系统。

2014年,天梭K1完成了金融、电力、公安、交通等12个行业市场的应用突破,建设银行、农业部、胜利油田、北京市财政局、广州白云机场、洛阳银行都在核心业务中用天梭K1替代了进口产品。目前,天梭K1已经获得广泛应用,形成了产业规模。

与此同时,浪潮主导成立了国产主机系统产业联盟,成员已从最初的16家发展到58家,覆盖了芯片、整机、软件等各个产业环节,建立了一条完整的国产高端容错计算机产业链,推动了整个计算机产业的发展。

在过去的几年中,浪潮从无到有建立自主技术体系,并从零开始建立产业体系,经历了超乎想象的困难。欧美日等发达国家一直将高端容错计算机相关技

术作为国防战略技术,对中国实行严格的技术封锁。美国的《出口管理条例》(EAR)和瓦森纳组织,严禁高端容错计算机相关技术流向中国,中国在高端容错计算机领域的相关积累一直处于空白状态。

而且,高端容错计算机研制工程规模大,技术起点高,这也是我国计算机产业一直没有涉足该领域的主要原因。例如天梭K1有32颗处理器,256个内存插槽,为这些模块建立基于数据一致性的互联和通讯机制,是计算机领域公认的关键难题,全球具有该项技术开发能力的公司不到5家。

高端容错计算机项目总设计师、浪潮集团首席科学家王恩东说,天梭K1系统是400多位浪潮工程师耗费4年多时间研发成功的。在研制过程中,浪潮成功突破“双翼可扩展多处理器紧密耦合共享存储体系结构”“三级缓存二级目录一致性协议”和“软硬一体化的容错技术体系”等3项核心技术,成功研制1个核心部件——处理器协同芯片组以及中国第一个Unix操作系统的K-UX,建立了较为完整的自主技术体系,结束了中国信息化受制于人的尴尬局面,保障了国家信息安全。

中国是人口最多的国家,也是发展最快的经济体,

将对国家安全造成重大威胁。因此高端容错计算机的自主对于我国的信息安全至关重要。

为此,国家863计划设立了十一五期间863重大专项“高端容错计算机系统关键技术与应用”,并列入“战略竞争”项目。

浪潮经过4年多时间的科技攻关,研制成功天梭K1高端容错计算机系统,为中国关键数据和系统运行在自主平台上提供了技术方案,也为实现国家信息安全战略的全面自主可控奠定了基础。

建立产业体系比技术突破更难

虽然高端容错计算机的研发十分困难,但后期的应用推广也不易。

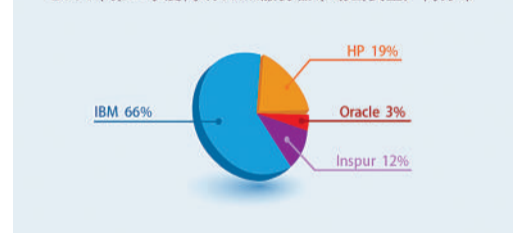
高端容错计算机需要与相应的软硬件配套才能实现应用,天梭K1必须同现有的主流软硬件产品相互兼容,但主流的配套软硬件厂商以及垄断中国市场的高端容错计算机厂商均为美国企业,他们结成了稳固的利益联盟,对国产企业采取排斥的态度。不仅如此,高端容错计算机承担的都是客户的关键业务,系统风险很容易转化为业务损失。

为打消客户疑虑,浪潮让最典型的用户从研发阶段就参与进来,这样的方式被当时的研发工程师形容为“长工把地主请到了田间地头来监工”。

金融是高端容错计算机最大的市场,也是应用需求最苛刻的行业。浪潮选择了建设银行,建设银行将工作委托给技术实力最强的新疆分行,新疆分行因此全程参与了天梭K1的研制工作。2011年8月,该行中间业务系统迁移完成,取代了IBM进口小型机。

2013年9月,浪潮在工信部和银监会的指导下牵头成立了国产主机系统产业联盟,覆盖了芯片、主机、数据库、中间件等各个产业环节,初步构建起中国的高端容错计算机产业链。有此依托,浪潮天梭K1的应用局面很快打开,在金融、保险、民航等事关国计民生的关键领域大显身手,已经在12个行业全部实现了应用,占据了12%的市场份额。与K1相配套的相关产品,市场份额也随之不断扩大。

2014年第三季度高端UNIX服务器市场出货量厂商分布



据全球权威调查公司IDC发布的数据,浪潮在高端Unix服务器市场的份额已经达到12%。

从无到有建立技术体系



王恩东
高端容错计算机项目总设计师、浪潮集团首席科学家

在国外技术严密封锁、产业链高度垄断的情况下,浪潮依靠自主创新在高端容错计算机体系结构、操作系统、处理器协同芯片和硬件系统方面实现重大技术创新与突破,从无到有建立高端容错计算机技术体系并开创高端容错计算机产业体系。

跨过“8”这个门槛

破解核心技术是浪潮遇到的第一个难题,也是一个表面看来很简单的问题,那就是如何让天梭K1系统处理器数量超过8颗。

高端容错计算机不仅可以通过处理器升级来获得性能提升,还可以通过增加处理器数量来提高性能,但是当处理器数量超过8颗,就会碰到“内存墙”的问题。

据浪潮集团首席科学家王恩东介绍,处理器要协同工作,相互之间必然要通讯,而且处理器之间不仅要进行数据传递,更要保持数据一致,任何数据异步都可能造成系统停机,这使得处理器通讯远比一般的通讯技术更为复杂。随着处理器数量的增加,处理器间的通讯强度会呈指数形式增加,甚至是失控。另一方面,通讯对处理器的消耗十分惊人,处理器完成一次本地通讯需要的时间是完成一次计算任务的200—300倍,如果是相邻节点的通讯则需要400—600倍的时间,如果采用一般的处理器协同技术,当处理器数量超过8颗以后,大量的处理器资源都被通讯占用,继续增加处理器数量时,性能提高就会十分有限,其直接的技术表现就是处理器

读取内存数据延时过长,这个现象在专业领域被称为“内存墙”。

天梭K1有32颗处理器,256个内存插槽,已经远远超过了初始的技术瓶颈,破解天梭K1面临的“内存墙”问题,只能创建一个全新的处理器互联结构和通讯机制,这是业界几十年来一直难以破解的难题,全球具有该项技术开发能力的公司不到5家。

提供99.999%的高可用性

高端容错计算机是专门承担关键应用系统的系统平台,一旦出现停机事故,将会严重伤害客户业务,甚至在社会经济层面造成影响,所以,能够提供99.999%的高可用性,可以7×24小时不间断运行是高端容错计算机最基本的应用需求。

王恩东说:“从可用性的角度看,研制高端容错计算机就像用积木来搭一个摩天大厦,必须开发一套完整的容错技术来消除每个可能出现的问题,保证系统运行的连续性。”

高端容错计算机大约有几十万电子器件,目前品质最高的工业级电子器件平均无故障时间在10000小时左右。另外,高端容错计算机的BIOS等底层代码有几千万行,最权威的Coverity Analysis集团数据,目前企业软件的缺陷密度为0.72,也就是平均每1000行代码中会有0.72个缺陷,而任何一处不良代码都是潜在的隐患。因此浪潮需要开发复杂系统内核容错技术,快速准确地发现、隔离和恢复数

百种软硬件故障,保障系统承载的关键应用稳定可靠运行。

操作系统10倍的复杂性

操作系统是硬件资源的管理者,是实现高端容错计算机系统功能、保证系统性能和可用性的关键子系统。高端容错计算机需要适配的操作系统,无论是性能、可靠性还是安全性,普通操作系统不能满足高端容错计算机的技术需求。

复杂度是高端容错计算机操作系统面临的第一个技术挑战。高端容错计算机处理器核、内存容量、I/O设备等资源的规模是通用服务器的十倍以上,由此使得操作系统资源管理复杂性极大提升。例如,100颗核心时的通讯复杂性约为10颗核心的100倍。在这样复杂的硬件条件下,操作系统需要具有精准、高效的线程调度、内存分配回收以及中断处理机制,避免或减少跨节点数据通讯,避免或降低多处理器对同资源请求的冲突情况,因为这些情况一旦发生,一次操作会占有几百次,甚至是上千次的处理器计算周期,大幅拉低系统性能。

王恩东表示,中国已经掌握了国际领先的32路高端容错计算机核心技术,我们正在开发性能更高、可靠性更强的64路系统高端容错计算机系统,同时也将进一步拓展应用规模,天梭K1的目标市场份额是30%以上,把自主化进程向更深层次推进。