

DOI:10.13364/j.issn.1672-6510.20140163

基于 Hadoop 的校园物联网数据处理系统研究

熊聪聪, 吉苏杰, 王兰婷

(天津科技大学计算机科学与信息工程学院, 天津 300222)

摘要: 针对校园各物联网应用系统处理海量数据的性能差、数据的存储和运维成本高以及设备扩容升级困难等问题,设计了一种基于 Hadoop 的数据处理系统,为构建校园云数据中心、实现校园的智慧化服务提供有益的参考方案.文件处理模块针对海量结构化小文件的处理需求提出改进方案,对比实验表明该方案在降低集群主节点的内存消耗和提高小文件访问效率方面优于现有方案.

关键词: 校园物联网; Hadoop; 数据处理; 结构化小文件

中图分类号: TP399 **文献标志码:** A **文章编号:** 1672-6510(2015)05-0072-06

Campus Network Data Processing System Based on Hadoop

XIONG Congcong, JI Sujie, WANG Lanting

(College of Computer Science and Information Engineering, Tianjin University of Science & Technology, Tianjin 300222, China)

Abstract: According to the problems of poor performance of massive data processing, high cost for storage, operation and maintenance, difficult expansion and upgrading of equipment existing in the single campus application system, a data processing system was designed based on Hadoop, which can provide a useful scheme for building campus cloud data center and realizing campus wisdom service. A feasible improvement program of the system file processing module was proposed in order to meet the processing needs of massive structured small files. Experiments show that the improved program is more efficient than the existing solutions to reducing the memory consumption of the primary node and can improve the accessing efficiency of small files.

Key words: campus network; Hadoop; data processing; structure small file

物联网在高校中的应用现状是根据不同的应用场景构建局部物联网实现服务,而真正意义上的校园物联网^[1]是在校园内部,利用成熟的射频识别(radio frequency identification, RFID)、红外感应、激光扫描等传感技术构造一个覆盖整个校园的物体互连网络,实现不同应用需求下的一体化智能管理.基于并行计算和分布式存储的云计算^[1]最适合这一应用需求.

目前,云计算在校园物联网中的应用研究都处于通用应用需求和构建方案的探讨阶段,并未对核心功能模块的实现机制进行深入研究.例如:文献[2]提出基于物联网和云计算构建智慧校园的整体服务方案,

论述了实现这一方案需要关注的技术问题和功能需求;文献[3]设计了校园物联网云平台的总体架构,但均未对数据处理的中心工作机制进行研究讨论.

本文在现有研究基础上设计并实现了一种基于 Hadoop 的面向校园物联网服务的数据处理系统,旨在提升系统的运算处理性能,进而提高校园物联网智慧化服务质量.

1 现有校园物联网应用系统的问题分析

校园物联网中的典型应用有校园一卡通、图书管理、实验设备管理等.当前的实际状况是不同的应用

收稿日期: 2014-12-17; 修回日期: 2015-03-04

基金项目: 天津市科技型中小企业创新资金资助项目(12ZXCXGX33500); 国家自然科学基金资助项目(61272509)

作者简介: 熊聪聪(1961—),女,四川人,教授, xiongcc@tust.edu.cn.

系统各自形成一个局部网络, 搭建专门的 Web 服务器和数据库服务器, 同时开发配套的后台管理系统, 安排技术人员进行管理、维护和升级. 这种模式存在以下问题:

(1) 数据存储和管理的运维成本高, 硬件升级复杂. 在购买存储设备和软件定制的模式下, 往往需一次性投入大量资金, 一旦完成则无法后续动态调整, 随着设备的更新换代, 落后的硬件平台很难重复利用, 而配套的管理软件亦需要不断升级来与之相适应, 有时甚至需要重构, 这可能导致不可控的局面.

(2) 随着数据量的增长, 数据的处理和分析任务随之增加, 单个服务器的 CPU 运算速度和内存容量都是十分有限的, 在面对大规模数据集的计算任务时, 系统很快会出现性能瓶颈, 容易导致任务堆积, 运行速度缓慢.

(3) 存储空间扩容不便. 在扩容时, 传统服务器往往要停止应用服务, 先将数据拷贝至别处, 待换上大容量设备后, 再将数据迁移回来, 这种方法既繁琐, 又给用户带来不便.

基于 Hadoop 的数据处理系统设计和部署则不存在以上问题, 其降低了数据存储的运维成本, 将垂直式扩容转变为灵活的水平式, 可以有效解决单一服务器面对海量计算任务的性能瓶颈问题.

2 系统架构设计

Hadoop 是一个开源的分布式计算框架, 通常在大量廉价的硬件设备上部署集群, 其显著优势为低成本、高可用性、扩容灵活以及良好的移植性. Hadoop 由许多子项目构成, 其中两大核心构件为 Map Reduce 编程模型和分布式文件系统(HDFS)^[4].

系统内部架构分为访问层、应用服务层、基础管理层和存储层, 其架构如图 1 所示. 访问层是平台的入口, 通过接口的方式封装了下层的各类服务, 接收各种通信协议下的数据请求和响应. 应用服务层是系统的核心实现部分, 既是业务处理的关键层, 也是与基础管理层通信的桥梁. 主要业务包括: 数据统计与分析、ECA 规则响应、文件处理等. 另外, 系统内置预留的应用服务 API, 以便业务拓展. 基础管理层是平台的骨架层, 除了 MapReduce 和 HDFS 以外, Hadoop 框架还具有多个子项目, 它们协同工作, 提供了一套全面的数据管理和存储机制. 存储层是数据平台的资源池, 由 HDFS 统一分配和管理, 所有数据

分散持久化存储在集群的各个数据节点上.

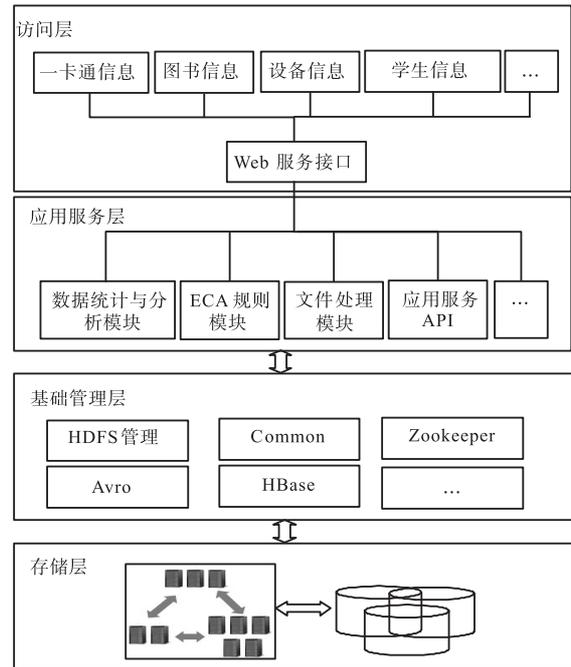


图 1 系统架构

Fig. 1 System architecture

3 核心模块设计

3.1 数据统计与分析模块

传统单节点服务器在进行信息统计、日志分析或数据挖掘等工作时, 将任务推送至堆栈中, 串行执行, 当任务量较多时, 排队等候时间较长. 而基于 MapReduce 并行计算模型编写的 Map 和 Reduce 处理函数则不存在此问题. 从并行计算的角度, 集群的 Master 主控节点充当 JobTracker 角色, 其中 JobTracker 和 NameNode 通常在同一节点上, 负责各种计算任务的调度与分配. 集群的 Master 主控节点接到任务请求后, 与 HDFS 的 NameNode 节点通信, 获取所需文件的所有块位置信息, 同时, 将 Map 任务发送给各个任务子节点并行执行. TaskTracker 根据任务要求对分配到本地的数据进行组合、排序和合并等处理, 然后将状态和完成信息报告给 JobTracker, 由 JobTracker 产生最终的结果文件. 这种并行工作模式的效率明显提高. MapReduce 模块的工作机制见图 2.

3.2 ECA 规则模块

事件控制行为(event control action, ECA)规则模块是实现数据平台事件监测功能的核心, 物联网中的

事件监测过程是:感知设备端向数据平台发送实时状态数据,平台通过 ECA 规则模块对数据进行事件描述、条件判断、逻辑处理,然后回发响应结果,达到监控感知设备状态、及时通知或预警的目的.例如:在一卡通应用中,用户在页面中设置各种 ECA 规则后保存,如余额不足提示、挂失报警、账单定期通知等,系统会为用户建立单独的 ECA 规则库,当该用户的一卡通被感知设备扫描时,感知数据将发送给系统,系统调用该用户的 ECA 规则库进行查询和合法性判断,最后返回结果并执行响应动作.

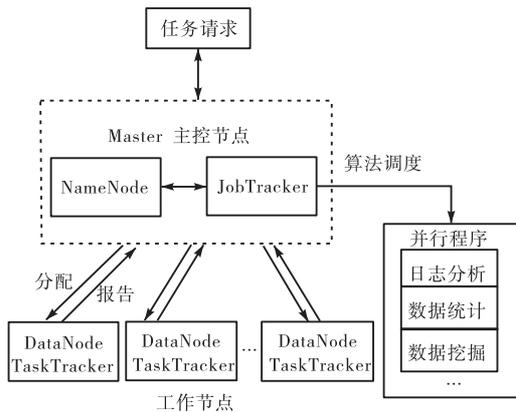


图2 MapReduce模块的工作机制

Fig. 2 Working mechanism of MapReduce module

ECA 规则实现了业务逻辑与动作执行的动态分离,满足多变环境下的松耦合需求.通过 ECA 规则模块,数据平台能够对特定条件下的事件进行主动式响应服务^[5],在整个业务生命过程中无需用户操作或外部应用的干预.其工作原理如图3.

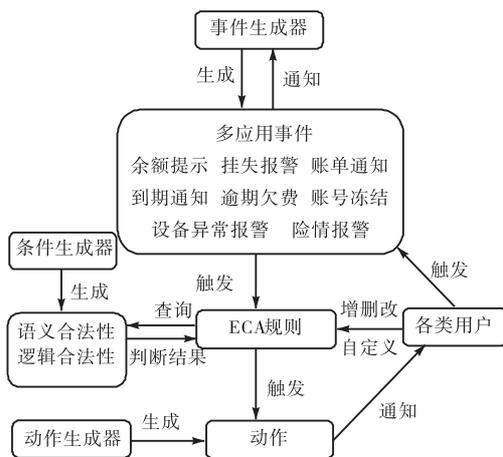


图3 ECA 模块工作原理

Fig. 3 The working principle of ECA module

3.3 文件处理模块

当前主流的分布式文件系统都是针对优先考虑

流式访问的大文件而设计的,忽略了小文件的存储和访问^[6].因此,系统的文件处理模块是影响系统性能的关键之一.

3.3.1 现有小文件处理方案分析

Har 方案的主要思想是通过构建一个层次化的文件系统将大量小文件合并打包成*.har 格式的特殊文件,并在 NameNode 上建立二级索引.这种方案明显改善了 NameNode 内存消耗问题,但存在以下不足之处:(1)小文件归档后,源文件仍在磁盘上,而且不能自动删除,造成了存储资源的重叠以及增加了管理员的额外工作量;(2)合并耗时长,在对小文件进行随机访问时,需要与 NameNode 上的二级索引文件通信,需要占用较多的时间资源.

Sequence File 方案的原理是通过键值对 (key-value) 这种数据结构将小文件序列化后,统一合并起来进行存储.这种方式除了解决了内存消耗问题外,同时还可与 MapReduce 很好的契合.最大的不足是,它没有建立小文件到 Sequence File 的索引,不支持内部小文件的随机访问,所以每次读取都必须遍历整个 Sequence File,导致查找文件的时间比 Har 方案的二级索引还要慢.

针对以上不足,出现了 Map File^[7]方案,它是带有部分索引的 Sequence File 升级版,但这种方式仍然需要在索引间隔区域内部进行一次遍历,随机读取的性能还达不到最理想.

此外,余思等^[8]从负载均衡和小文件合并规模的角度考虑,设计了一种基于层次分析的预测算法来进行负载均衡优化,但并未解决小文件的查找效率问题.王涛等^[9]从用户访问行为出发,通过改进概率浅语义分析模型,提出了一种改进方案,但该方案仅适合于文件之间关联性较强的云存储系统.

3.3.2 改进方案设计

针对校园物联网的各个子应用系统中存在海量结构化小文件的实际情况,基于小文件合并的思想设计了更适用的解决方案,从而进一步降低 NameNode 内存消耗和提高文件的随机访问效率.改进方案的主要原理是在客户端和系统集群的 HDFS 之间增加一个处理层,该层首先对结构化的小文件进行数据交换,统一格式后,记录下每个小文件的元信息和位置信息,保存在临时文件中,然后建立小文件到合并文件的索引文件,与 Har 方案不同的是,索引文件并不放置在 NameNode 中,而是与该合并文件进行绑定,存储到 DataNode 中,在客户端要求访问合并文件内

部的小文件时,系统在获取数据块位置后,将绑定的索引文件发送至客户端本地进行索引,这样既不占用NameNode内存,又将文件的索引任务从系统集群转移到了客户端上,访问效率明显提高.该方案的工作原理见图4.

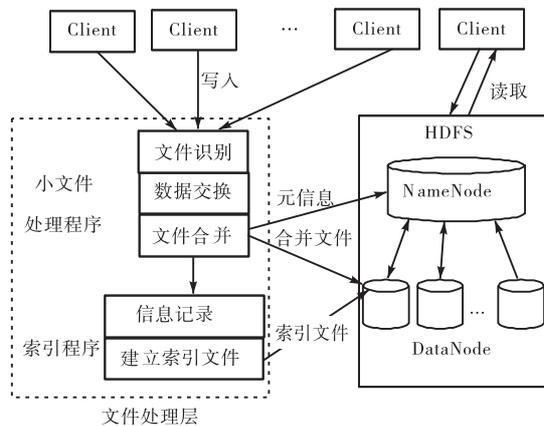


图4 改进方案的工作原理

Fig. 4 Working principle of the improved scheme

XML是目前通用的具有国际统一标准的结构化文件格式,可根据实际需求灵活建立索引机制^[10].对于校园一卡通、图书管理、教学设备管理等应用场景,结构化的感知数据文件都支持与XML文件的数据交换.因此,系统改进方案的具体实现过程可描述如下:

(1)文件识别.对于客户端传来的文件,首先判断文件类型,若为小文本文件,则判断该文件是否支持XML数据交换,若支持则转(2),若不支持则警告用户文件特殊,询问是否继续,若继续则跳转至(5),否则结束本次操作;若为非结构化小文件或大文件,则跳转至(5).

(2)文件处理.先将文件转换为XML格式,然后查询NameNode内存中是否有等待状态的XML文件,若无则新建XML文件,通知NameNode建立元数据信息,并分配DataNode块位置,建立状态标识符,值为“0”,表示等待状态,然后跳转至(3);若已有等待状态的XML文件,则直接跳转(3).

(3)建立索引信息.询问等待状态的XML文件所在的DataNode是否存在状态标识符为“0”的索引文件,若不存在则新建1个XML索引文件,同样建立数值为“0”的等待状态标识符,将文件发送给处理程序,而后进行信息统计;若已存在则直接调用,然后统计文件信息,包括文件编号(用以唯一标识文件)和文件大小.

(4)文件合并.计算并判断待写入的小文件与待合并的XML文件总计是否已达到64MB,若已达到则将XML文件与其所属的索引文件的状态标识符均改为“1”,表示已满,然后跳转到(2);若小于或等于64MB,则标识符不变,顺序地将小文件的内容写入待合并的XML文件中,以“<FileNumber>...</FileNumber>”标签为一个小文件的存储区域,记录下该小文件的起始位置偏移量和文件长度,并发送给其索引文件.

(5)上传至HDFS.若为一般大文件,通知NameNode建立元数据并分配数据块,将文件写入到DataNode;若为合并文档,则将合并文档与索引文件一起发送给DataNode进行管理维护.

4 系统实验与分析

实验在系统集群上进行,Hadoop版本为0.20.1,JDK版本为1.6.0_24,操作系统为Ubuntu 12.04.HDFS集群由4台服务器组成,为保证单节点性能平衡,每台服务器的硬件配置均相同,双核Intel Xeon E5 506 2.13 GHz处理器,2 GB内存,200 GB SATA硬盘.其中1台作为客户端服务器和NameNode节点,其余3台为DataNode节点,数据块为默认设置的64 MB,副本数为3.

4.1 统计测试与分析

为了验证在执行适合于MapReduce模型的任务时,Hadoop集群相对于单服务器的性能优势,分别在不同节点数量的集群和单服务器上对不同数据量的系统日志文件进行关键词汇(例如:Error、Warning等)统计测试,实验结果见表1.

表1 关键词汇统计测试结果

Tab. 1 Key words statistical test results

运行条件	运行时间/s				
	1 MB	10 MB	100 MB	1 GB	10 GB
单服务器	22	29	51	425	1 326
集群(1个节点)	33	37	54	431	1 340
集群(2个节点)	37	51	72	411	953
集群(3个节点)	35	47	77	399	630

从表1结果可以看出:当文件较小时,在节点多的集群上反而没有在节点少的集群上运行快,因为其在MapReduce操作上耗费了较多时间;然而随着文件大小的增加,MapReduce的性能就充分体现了出来,集群的节点数量增加后处理速度明显提高.可以推断,MapReduce在处理更大规模的数据时,效率提

高的更显著。

4.2 改进方案的性能分析

实验所用海量结构化小文本文件由程序模拟生成,文件数据结构是一卡通应用系统的数据报表及高校图书馆馆藏资源的检索信息,共模拟生成了 10^6 个消费信息和馆藏信息小文件,单个小文件的大小约为 32 KB,数据量共计为 30.52 GB。

以 NameNode 内存使用情况和对小文本文件随机读取的时间效率为指标,考察了改进方案的性能,并与其他方案进行对比分析.其中,以文件数量为变化因子,对比在 HDFS 原方案、Har 方案和改进方案下 NameNode 的内存使用情况,结果见图 5.对于文本类型的小文件,Har 方案的小文件合并机制比 SequenceFile 或 Map File 方案更有优势,因此选取 Har 方案进行对比.使用 bin/hadoop jar hadoop-0.20.2-test.jar nnbench 测试命令,可以获得 NameNode 的内存使用情况。

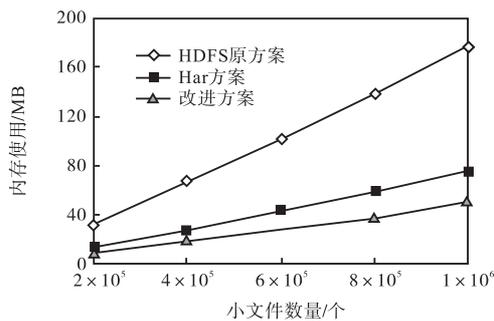


图 5 NameNode 内存使用情况

Fig. 5 Memory usage of NameNode

由图 5 可知:不处理而直接存储小文件的 HDFS 原方案占用内存最多,这是因为 NameNode 中为每个小文件都建立了元数据信息;Har 方案对 NameNode 的内存消耗问题有明显改进,在存储 10^6 个小文件时的内存消耗仅为原方案的 42.6%;而改进方案存储 10^6 个小文件的内存消耗仅为 HDFS 原方案的 28.7%,为 Har 方案的 67.4%,这是因为除了合并大量小文件外,还将索引文件转移到了 DataNode 上,由 DataNode 开销内存去管理,从而降低了 NameNode 内存使用率。

在客户端的检索程序中,随机访问集群中某合并文件上的小文件,分别比较 HDFS 原方案、MapFile 方案以及本文优化方案下的检索时间,实验结果见图 6. MapFile 方案的检索效率比 Har 和 SequenceFile 高,因此选取 MapFile 方案进行对比.在不同文件数量下,分别选取 10 个时间段内的小文件,每个时间

段随机输入 10 个编号值,则每种方案下随机检索共 100 个小文件,为减小偶然误差影响,去除各时段内检索用时最少和最多的各两个值,然后计算平均检索时间。

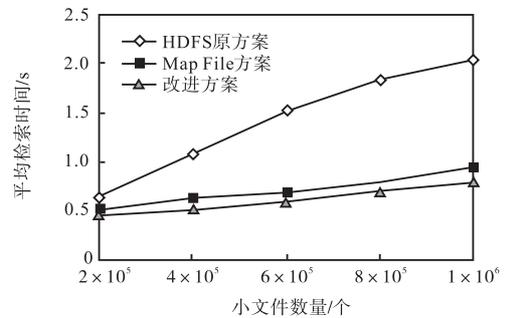


图 6 平均检索时间曲线

Fig. 6 Diagram of average search time

由图 6 可知:当数据量为 10^6 个小文件时,改进方案的检索时间比 HDFS 原方案减少了约 1.3 s,比 Map File 方案减少了约 0.15 s,在不同数据量下,改进方案的检索时间均有明显提升,而且随着文件数量的继续增加,检索时间的优化效果会更加明显。

从两组实验结果可知,改进优化方案对小文件大量消耗 NameNode 内存和内部小文件的随机访问效率这两个性能指标均有明显的改善和提高。

5 结 语

对海量多源异构的物联网感知数据的处理是实现智慧化校园服务的关键部分,核心模块主要面向 3 个方面的处理需求:日志分析、数据统计等高并行需求的计算任务;面向用户的智能化事件监控;海量结构化小文件的存储与访问服务.这些需求既是本文的研究意义所在,也是构建校园物联网云数据中心必须关注和解决的重要问题.本文设计并实现了一种基于 Hadoop 的面向校园物联网服务的数据处理系统,针对处理海量结构化小文件的需求对文件处理进行了改进.不足之处是,本文并未对并行计算任务的负载均衡和 ECA 规则响应的实时性进行探讨,有待进一步研究。

参考文献:

- [1] 李开复. 云计算[J]. 中国教育网络, 2008(6): 34.
- [2] 冯志杰. 云计算在校园物联网中的应用研究[J]. 青岛大学学报:工程技术版, 2014(3): 44-48.
- [3] 吕倩. 基于云计算及物联网构建智慧校园[J]. 计算机

科学,2011(S1):18-21,40.

- [4] Brock Noland. Mounting HDFS[EB/OL]. [2012-01-20]. <http://wiki.apache.org/hadoop/MountableHDFS>.
- [5] 朱达. 基于事件的服务协同及通信服务提供技术研究[D]. 北京:北京邮电大学,2011.
- [6] 周国安,李强,陈新,等. 云环境下海量小文件存储技术研究综述[J]. 信息网络安全,2014(6):11-17.
- [7] Visser W, Havelund K, Brat G, et al. Model checking programs[J]. Automated Software Engineering, 2003, 10(2):203-232.
- [8] 余思,桂小林,黄汝维,等. 一种提高云存储中小文件存储效率的方案[J]. 西安交通大学学报,2011,45(6):59-63.
- [9] 王涛,姚世红,徐正全,等. 云存储中面向访问任务的小文件合并于预取策略[J]. 武汉大学学报,2013,38(12):1504-1508.
- [10] 孔令波,唐世渭,杨冬青,等. XML 数据索引技术[J]. 软件学报,2005,16(12):2063-2079.

责任编辑:常涛

(上接第41页)

的工艺路线如下:选择一定浓缩程度的卤水(密度约为 1.228 g/cm^3) 在适当低温($-12 \text{ }^\circ\text{C}$ 左右)条件下,先行冷冻析出芒硝,然后归集析出芒硝后的卤水,浓缩后卤水中 CO_3^{2-} 含量变大,再利用低温天气进行冷冻析出泡碱.析出泡碱后的卤水 CO_3^{2-} 含量低,降低了卤水日晒蒸发过程中碳酸锂析出的可能性,同时还实现了冷冻母液中有开采价值的 Li^+ 和 K^+ 的富集.

参考文献:

- [1] Baumgartner W, Gross A. The global market for electric vehicles[J]. Business Economics, 2000, 35(4):51-56.
- [2] 郑绵平,刘喜方. 中国的锂资源[J]. 新材料产业, 2007(8):13-16.
- [3] 高峰,郑绵平,卮贞,等. 盐湖卤水锂资源及其开发进展[J]. 地球学报,2011,32(4):483-492.
- [4] 郝勇,张启海,李广汉,等. 西藏结则茶卡和龙木错盐湖卤水协同提锂研究[J]. 无机盐工业,2013,45(6):27-29.
- [5] 张永生,卮贞,卜令忠,等. 富锂碳酸盐型卤水在系列冷冻温度下组成的演变[J]. 海湖盐与化工,2001,30(1):3-6.
- [6] 中国科学院青海盐湖所分析室. 卤水和盐的分析方法[M]. 2版. 北京:科学出版社,1988.
- [7] 宋月月,王学魁,董景岗,等. 拉果错湖水 $5 \text{ }^\circ\text{C}$ 蒸发析盐规律研究[J]. 天津科技大学学报,2014,29(1):42-45.
- [8] 吴敬礼,王学魁,董景岗,等. 拉果错盐湖卤水 $15 \text{ }^\circ\text{C}$ 等温蒸发析盐规律研究[J]. 天津科技大学学报,2014,29(3):49-52.

责任编辑:周建军

(上接第47页)

- [25] 赵辉,齐义泉,王东晓,等. 南海叶绿素浓度季节变化及空间分布特征研究[J]. 海洋学报,2005,27(4):45-52.
- [26] 柴扉,薛惠洁,侍茂崇. 南海升降流区域分布及形成机制分析[J]. 中国海洋学文集:南海海流数值计算及中尺度特征研究,2001(13):117-128.
- [27] Vaillancourt R D, Marra J, Seki M P, et al. Impact of a cyclonic eddy on phytoplankton community structure and photosynthetic competency in the subtropical North Pacific Ocean[J]. Deep Sea Research Part I: Oceanographic Research Papers, 2003, 50(7):829-847.
- [28] Partensky F, Blanchot J, Vaulot D. Differential distribution and ecology of Prochlorococcus and Synechococcus in oceanic waters: A review[J]. Bulletin de l'institut d'océanographie, 1999, 19:457-476.
- [29] Glover H E, Prézélin B B, Campbell L, et al. A nitrate-dependent Synechococcus bloom in surface Sargasso Sea water[J]. Nature, 1988, 331(6152):161-163.

责任编辑:郎婧

【科研成果简介】

氨基酸及核苷代谢调控技术及产业化应用

在全球各主要发达国家加速推进生物技术发展的背景下,我国于2010年明确将生物产业列入国家战略性新兴产业。近十年,我国生物产业快速发展,年均增长率为25.1%,比同期GDP增速高出1倍以上。2012年,我国生物产业产值达2.4万亿元,预计到2020年可接近10万亿元。据统计,发酵工业创造的产值约占生物产业总产值的50%。作为生命体蛋白质的基本组成单位,氨基酸在人和动物的营养健康方面发挥着重要的作用。目前已广泛应用于医药、食品、饲料、保健、化妆品、农药、肥料和制革等领域。氨基酸及其衍生物的种类已由20世纪60年代的50种左右发展到现在的1000余种,全球每年的需求量已超过800多万吨。

氨基酸工业是发酵工业的支柱产业之一,当前全球氨基酸工业已发展成为品种繁多、门类齐全的庞大产业群。我国是氨基酸生产大国,2013年我国氨基酸总产量达到400万吨,同比增长10%,继续稳居世界第一。产品以大宗氨基酸为主,其中谷氨酸及其盐年产已达240万吨,占世界总产量的70%以上,赖氨酸(122万吨)和苏氨酸(27万吨)产量也居世界前列。近些年,高附加值氨基酸产品的需求快速增长,带动了一些小品种氨基酸和高附加值氨基酸衍生物产品的快速发展。虽然我国大宗氨基酸的生产技术水平已达到国际先进水平,但其他高附加值氨基酸产品与国外先进水平仍有不小的差距。从我国的氨基酸生产现状分析,提升现有氨基酸产品的生产技术水平,加快氨基酸深层次加工和新产品开发是实现可持续发展的必然途径。核苷类物质亦在生命代谢过程中扮演着极其重要的角色,具有广阔的市场前景。

天津科技大学陈宁教授所带领的科研团队长期从事氨基酸及其衍生物和核苷类物质代谢工程及发酵工程方面的研究工作。主要包括:(1)氨基酸及核苷生产菌种的定向改造、发酵过程优化控制及产业化应用。针对我国氨基酸工业存在的菌种产酸水平和转化率低、性状不稳定等不足,在代谢控制发酵原理指导下,综合利用诱变育种、基因工程、组学、合成生物学、高通量筛选等技术,对现有谷氨酸、赖氨酸、亮氨酸、缬氨酸、异亮氨酸、色氨酸、肌苷、鸟苷、腺苷等氨基酸及核苷生产菌种的遗传特性、代谢网络及组学特性进行系统地分析,阐明其代谢途径及调控机制并理性地设计和改造其代谢途径,对其代谢和遗传性能进行优化。通过对生产菌的生理和遗传特征以及反应器特性的研究,建立基于基因水平的分子尺度、代谢调节的细胞尺度以及工艺控制的反应器尺度的多尺度发酵过程优化及放大的关键理论、方法和技术体系。将获得的菌种和发酵工艺应用于实际生产中实现产业化。本研究开发的多种氨基酸生产菌发酵水平处于国内甚至国际领先水平且已实现产业化。如自主开发的谷氨酸温度敏感突变株及其发酵工艺已在山东菱花集团、河北梅花集团、宁夏伊品集团等大型氨基酸企业推广应用;发酵法生产分支链氨基酸在无锡晶海氨基酸公司和山东鲁洲氨基酸公司推广,分支链氨基酸代谢调控技术及产业化项目获得2013年天津市科技进步奖开发类二等奖;发酵法生产鸟苷项目在湖南赛康德公司推广;发酵法生产色氨酸、苏氨酸、肌苷等项目已在多家公司成功推广。上述项目的推广打破了发达国家对我国氨基酸、核苷的技术和市场的垄断,有力地推动了氨基酸和核苷技术转化为生产力的进程,为我国氨基酸和核苷工业做出了突出贡献。(2)氨基酸及核苷发酵产物高效分离纯化技术。研发膜分离、工业色谱、连续浓缩结晶等分离提取新技术及其在不同发酵产物提取工艺过程中的应用;针对生产菌株的代谢途径及发酵过程中发酵产物及副产物特性研发合理的分离提取工艺,降低产品分离过程中的资源和能量消耗,实现发酵产品的高效分离。采用微滤、纳滤、反渗透等膜技术以及离交法等方法及不同分离提纯工艺对氨基酸进行分离,分别获得了高纯度的分支链氨基酸、谷氨酸、苏氨酸、色氨酸及鸟苷等。该研究在山东菱花集团、河北梅花集团、宁夏伊品集团、无锡晶海氨基酸有限公司和山东鲁洲氨基酸有限公司等多家企业得到推广。(3)抗逆菌株的选育及新型发酵产品生产菌株的开发。针对我国氨基酸和核苷生产菌株抗逆性差、工业耗能大、新型氨基酸及氨基酸衍生物种类相对较少等问题,采用系统生物学和合成生物学的理论和技术选育或构建耐高温、耐酸碱的抗逆的苏氨酸、色氨酸、谷氨酸等生产菌株,达到提高菌种生产能力和抗逆性、降低生产成本、节约能耗的目的;构建并定向改造氨基酸衍生物、利巴韦林、 α -酮戊二酸的方法等新型发酵产品生产菌,延伸微生物制造产业链,大力发展新产品,为微生物制造技术奠定了理论基础。

近年来,我国氨基酸工业取得了很大的进步。但也要看到,与世界氨基酸产业强国比较,我国氨基酸领域的技术创新能力相对薄弱,新型氨基酸产品及衍生物开发也存在明显不足。产品多元化、生产高效化、质量高端化和应用扩大化是氨基酸工业的长期发展目标。当前形势下,新技术应用和新产品开发是我国氨基酸工业实现可持续发展的两个最重要的因素。因此,提高技术研发和自主创新的能力是氨基酸工业发展不变的主题。虽然我国氨基酸工业发展存在一定的问题,但相信自主创新体系的完善将极大增强我国氨基酸工业的核心竞争力,促进我国氨基酸工业向着良性健康的方向发展。

供稿:生物工程学院 张成林