西藏羊八井ARGO宇宙线实验

陈刚，程耀东
中国科学院高能物理研究计算中心 北京 100049

摘 要： 本文首先介绍西藏羊八井ARGO宇宙线实验的基本状况，然后详细探讨了ARGO实验的网格计算模型，包括数据传输与监控、作业管理、应用移植等。最后，对网格的基本功能与网络连接状况进行测试，并提出展望。

关键字： 羊八井 ARGO 宇宙线实验 网格 数据传输

Grid Computing of Yangbajing-ARGO Cosmic Ray Experiment

Gang Chen, Yaodong Cheng
Computing Center, Institute of High Energy Physics, Chinese Academy of Sciences, Beijing 100049 China

Abstract: Yangbajing-ARGO experiment, located at Yangbajing, is to study cosmic rays, sub-TeV gamma ray sources and GeV Gamma Ray Burst. This paper describes the grid computing system implemented for the experiment. The grid consists of data transfer and monitor system, job management system. The grid computing system was deployed among the sites at Yangbajing, IHEP in Beijing and INFN in Italy. The tests and service challenges were carried out and shown in the paper.

Keywords: Yanbajing, ARGO, Cosmic Ray experiment, Grid computing, data transfer

1.引 言

YBJ-ARGO实验是中国与意大利合作项目，是世界上一个具有特色的宇宙线观测实验。该实验采用“地毯式”探测器，全天候对宇宙线及天体粒子进行观测，以采集海量的数据。物理学家对这些海量数据进行分析，筛选出有物理意义的结果。海量数据的传输、共享和分析使得YBJ-ARGO实验面临巨大挑战。有效地利用分散资源，提升资源应用效能，成为解决问题的唯一可行途径。网格技术[1]为现有的资源共享提供一个大型的分布式协作式的构架，并逐渐成为解决如高能物理实验、破解基因密码等数据量极大的科学工程计算问题的最直接、最有效的途径。为此，YBJ-ARGO实验启动了网格应用项目，以整合各个参与单位的资源，满足数据传输和处理的需求。

2.ARGO宇宙线实验介绍

中科院高能物理研究所粒子天体物理重点实验室是中国科学院的一个开放实验室，是我国研究宇宙线与高能天体物理的重要实验基地。该实验室与日本、意大利合作在西藏羊八井（YBJ）建成了宇宙线观测站。该观测站主要由ASγ阵列、地毯式全覆盖阵列(YBJ-ARGO)和宇宙线强度检测装置等组成。其中，YBJ-ARGO由中国科学院高能物理研究所（IHEP）等国内单位与意大利的国家核物理研究院（INFN）等单位合作建成。该阵列由RPC探测器组成的cluster铺设而成，面积约10000平方米。YBJ是羊八井的中文拼音YangBaJing的缩写，ARGO本为希腊神话中的一个巨怪，身上长着100只眼睛，50只睁着，50只闭着，交替睁闭，总不休息。以YBJ-ARGO命名这项合作，意味着羊八井“地毯”全天候地进行宇宙线及天体粒子物理的观测工作。羊八井宇宙观测站，位于西藏拉萨市西北90公里的青藏和中尼公路交叉点附近，念青唐古拉山脚下的一个长约70公里、宽约7-15公里的小盆地内，海拔4300米，气候条件得天独厚，在严寒的冬季也可正常进行工作。这个北半球海拔最高的宇宙线实验站已成为国际上知名的宇宙实验站。羊八井实验基地的科学目标包括寻找γ点源、弥散γ观测、高能γ暴、反质子丰度测定、太阳宇宙线、超高能宇宙线成分能谱以及寻找暗物质候选粒子等国际天体物理前沿课题[2]。

YBJ-ARGO实验采用边建设边实验的方式。从2004年开始就进行数据采集。到2007年底，实验全部建成投入运行。羊八井宇宙线观测实验基地每年将采集约200TB以上的原始数据。

根据宇宙线物理研究的特殊性，从探测器采集的原始数据需要进行预处理，产生物理学家能理解的数据。预处理过程需要对原始数据进行过滤，剔除物理学家不感兴趣的数据（或称为本底数据），然后对过滤后的数据进行重建，成为具有物理意义的所谓‘事例’。使物理学家在‘事例’的基础进行相关的物理研究。预处理过程对数据进行了过滤和压缩，使体积大大减小，约为原始数据体积的20%。这种新产生的数据叫做事例重建数据。因此每年的重建数据约为40TB。数据处理需要约相当于400个目前最快的CPU的处理能力，用于实验数据的预处理和物理模拟；数据存储每年需要约240TB的存储空间。

图1：羊八井YBJ-ARGO宇宙线实验大厅

如何将海量的原始数据从西藏羊八井传回到位于北京的中科院高能物理研究所，并由合作单位分享和分析处理，是YBJ-ARGO宇宙线实现面临的首要问题。传统的方法是人

工运送磁带，但是效率很低。在网络中心的帮助下，羊八井与高能所之间建立了带宽为155Mb/s的专线网络，因此实现了从羊八井到高能所的高速数据传输。高能所和INFN独立进行数据的重建，模拟和分析。随着实验不断深入，实验数据的不断积累，这将对双方的计算处理能力、数据传输能力和数据存储能力提出了巨大的挑战。为了能最大限度地优化计算资源和存储资源的使用，构建一个完整统一的数据处理平台，YBJ-ARGO实验决定建立数据网格[3]，采用先进的网格技术和高速的国际网络来满足未来的需求。

3. ARGO实验的网格系统

3.1 网格技术简介

网格的本质特征就是实现各种资源的共享。网格技术不仅可以用于建立高性能的本地计算环境，还可以将参加实验的各合作单位的计算资源整合在一起，形成一个超大规模的计算平台，以满足超强计算处理能力和海量数据存储空间的需求。

网格系统结构可以分成三个层次：资源层，中间件层和应用层。网格资源层是构成网格系统的硬件基础，包括各种计算资源和网络设备，它实现了网格资源在物理上的连通。网格中间件层是指一系列工具和协议软件，其功能是屏蔽网格资源层中计算资源的分布，异构的特性，为网络应用层提供统一、透明的使用接口。网格应用层是用户需求的具体体现，是指在网格中间件的支持下，网络用户可以使用其提供的工具或环境开发各种应用系统。

当前，存在多种网格中间件，比如Globus、gLite、Unicore、Naregi、GOS、CROWN等。考虑到与国际高能物理网格兼容等因素，ARGO实验采用gLite网格中间件[5]。

3.2 ARGO网格模型

目前，参与ARGO实验的站点目前主要有三个：YBJ、INFN、IHEP，基本上成对称形式，其中YBJ是数据产生站点，INFN和IHEP分别为数据处理和分析站点，这两个站点对称地对YBJ传送来的原始数据进行存储，同时两个站点利用统一的网格平台提供的计算能力分别进行数据重建和模拟，而且两个站点重建数据和模拟数据将不断保持同步，也就是说，一个站点产生的重建数据和模拟数据将及时地传送到另一个站点，尽快保持两个站点的数据存储处于相同状态。在这种网格模型下，两个站点可以做到有效地利用计算资源，而且为原始数据、重建数据以及模拟数据提供了可靠的安全保护措施，在INFN和IHEP两个站点的数据互为稳定的备份关系。

在上述的网格应用模型下，网格系统的建设主要包括：INFN和IHEP两个站点的计算资源的共享、站点之间的可靠有效的数据传输以及站点的数据存储和获取。

在计算资源共享方面，网格中间件已经提供了相对成熟的技术实现。首先，YBJ-ARGO实验建立了虚拟组织（VO），即ARGO VO，只有加入该VO的用户才可以使用网格系统中的资源，每个用户的身份通过由IHEP或INFN认证中心颁发的用户证书来确认；其次，使用网格的信息监测和索引服务（BDII）发布各站点的资源状况，及时提供各站点的CE（Computing Element），WN（Work Node）以及SE（Storage Element）等网格组件的信息；另外，使用网格调度系统RB（Resource Broker） 进行资源的有效与透明分配，用户只要通过系统提供的UI（User Interface）就能方便地实现对资源的使用。

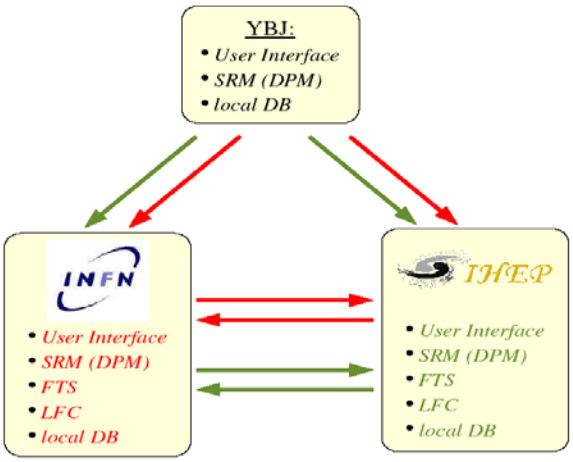

图2：ARGO实验网格应用中的数据存储和传输模型

为了保证可靠的数据传输，YBJ-ARGO实验开发了“Data Mover”[6]。Data Mover基于四个gLite组件：存储单元SE、文件传输服务FTS、网格文件目录LFC以及用户接口UI。Data Mover在各站点之间通过文件传输服务建立可靠的文件传输通道，并在YBJ，INFN以及IHEP三个站点建立文件状态数据库用于监

控和管理数据的传输。YBJ站点主要负责将在线数据文件（DAQ）上传到网格，并记录到文件状态数据库中；接着将文件进行排队，选择合适的FTS通道进行文件传输；在传送过程中监控文件传输的状态，进行各种容错处理，及时更新数据库的信息。INFN与IHEP的站点的状态数据库都存有YBJ状态数据库的本地备份，因此需要及时地根据YBJ的数据库进行更新，这样可以保证两个站点的用户可以及时获得原始数据文件的状态。IHEP与INFN站点负责将YBJ传来的原始数据注册到网格文件目录LFC上，方便用户的读取。在重建与模拟数据的传输中，IHEP与INFN将分别同时进行数据的重建与模拟，两地的数据存储信息如果不能做到及时地同步，就可能造成工作的重复以及资源的浪费，因此需要IHEP和INFN两地的文件状态数据库需要进行时时地同步，从而两个站点通过相互的备份使得两地的文件存储能尽快趋于一致。

为了监控数据传输状态和各个站点的运行情况，YBJ-ARGO实验基于JAVA开发了一个图形界面，如图3所示。

3.3 应用软件移植

网格中间件以及相关的工具提供了很好的资源共享平台，实现计算的网格化需要将应用软件移植到这个平台。YBJ-ARGO应用软件的移植主要关注蒙特卡罗模拟程序和原始数据重建软件。模特卡罗模拟程序用于估算探测器的性能和数据采集的数据率。该过程分为两步：首先调用大气中的宇宙线产生器CORSIKA，然后调用ARGOG程序与YBJ-ARGO探测器进行模拟。原始数据重建程序使用复杂的算法计算出宇宙线的主要参数，比如宇宙线的入射角度、能量等。

模拟与数据重建应用非常相似，都需要提交大量的作业，计算作业都需要输入运行号、输入或输出文件名、能量范围、刻度文件等，其中大部分可以动态获得（可以通过查询数据库直接获取，或通过其它参数间接计算）。向网格系统提交这些作业与向本地集群提交的主要不同点是：

- 向网格提交作业需要作业描述语言JDL文件；

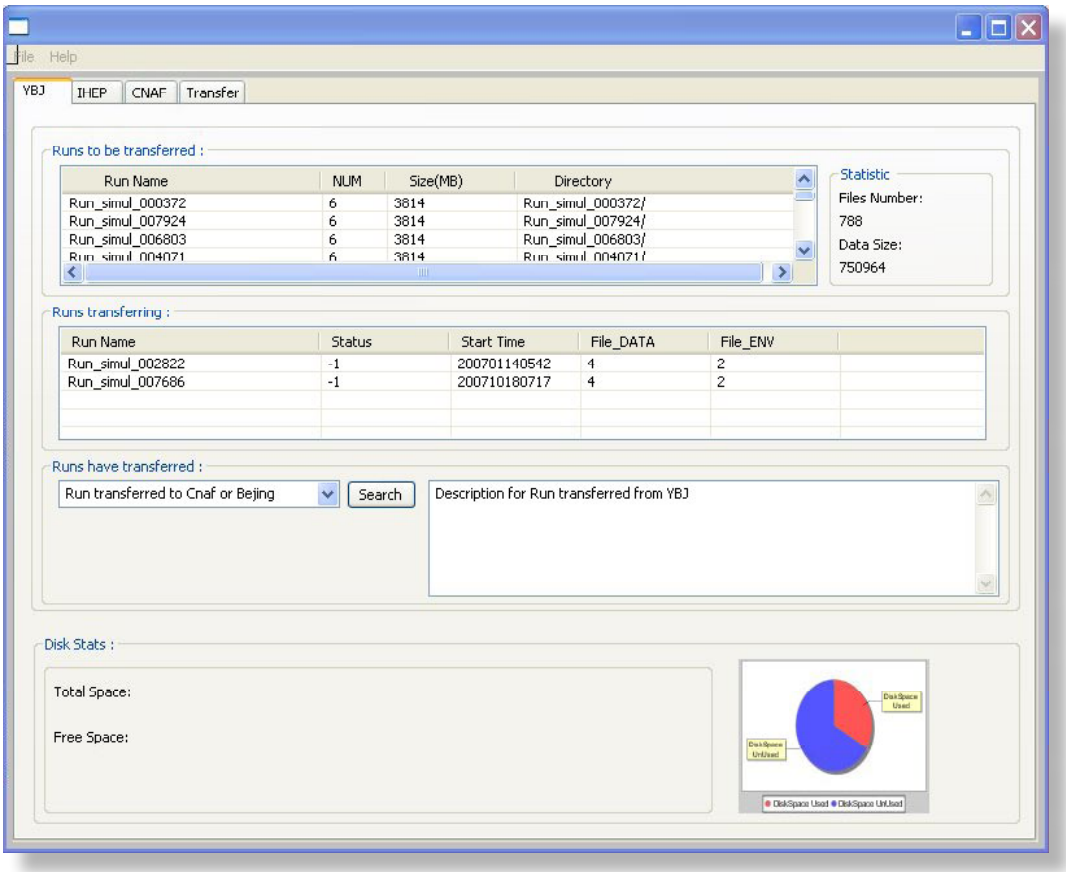

图3：数据传输的GUI界面

■ 作业的脚本文件需要更加通用，比如脚本最好不要依赖绝对路径名，而使用相应的环境变量；

■ 由于因为资源代理RB动态的把作业调度到任何一个合适的计算单位CE上执行，使得跟踪作业比较困难。

实际上，网格环境下的作业提交也可以看成是本地集群系统的一个特例。基于这个前提，YBJ-ARGO实验开发了通用的作业提交程序，可以同时向集群和网格系统提交作业。该程序调用相应的Perl脚本、配置文件和模板文件。应用软件的移植实际上就是通过修改模板文件，通用的作业提交程序就可以自动产生相应的JDL文件并向网格系统提交作业。为了跟踪作业的执行状况，网格作业提交后产生的作业ID被保存在数据库中。

4. 网格功能测试

对于YBJ-ARGO实验的网格模型，在IHEP站点做了实验和测试。网格应用的第一步就是搭建所需的网格环境，安装包括CE、UI、WN、RB以及SE、LFC、BDII等基本组件。实验的内容主要包括：建立ARGO虚拟组织（VO）以及颁发用户证书、YBJ-ARGO软件在网格系统上的安装和发布、YBJ-ARGO数据在网格环境中存储和传输以及YBJ-ARGO实验重建和模拟作业的提交和结果文件的获取。

高能所计算中心已经建立了CA认证中心（IHEP CA），该中心可以颁发用户，主机或服务证书，并已经成为APGridPMA和EuGridPMA的成员之一。ARGO VO的建立则通过VOMS服务进行，该服务可以将属于YBJ-ARGO实验的用户加入到ARGO VO中，并且进行角色的分配，本实验主要分配两种角色：实验软件管理者（ESM，Experiment Software Manager）和普通用户。ESM主要负责在ARGO VO管辖的站点进行软件安装，并且将所安装的软件版本及时地公布在网格信息系统上，这样用户可以通过查询网格信息系统了解软件安装情况，选择合适的站点进行作业的提交。普通用户则是用于作业的提交以及资源的使用。

YBJ-ARGO软件在网格系统上的安装和发布使用gLite中间件的两个工具lcg-ManagerSoftware and lcg-ManageVOTag。其中，lcg-ManagerSoftware负责将YBJ-ARGO软件安装到CE的指定安装目录上，安装的软件可以为各WN所共享，而lcg-ManageVOTag则负责将安装的YBJ-ARGO软件的状态以及版本发布到信息系统，这样用户可以很方便地查询到所安装的软件版本。

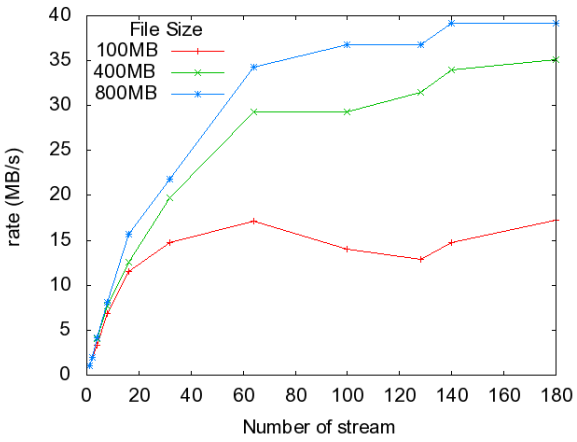

图4：从UI上载到SE的传输速度随文件大小的变化（横轴为文件大小MB，纵轴为传输速度MB/s）

用户对文件的存储和获取主要通过文件目录服务LFC，传输协议基于GridFTP。所用的传输工具为lcg-cr。在传输过程中，性能将受到并行流的多少、文件大小等参数的影响。图4记录的是数据传输速率与数据文件大小以及并行流的关系。可以看出，使用40个以上的流来传输大于100MB的文件可以轻松达到20MB/s，传输800MB文件的性能要优于400MB与100MB的文件，这主要是由于文件越大，其中网络连接等开销在整个传输中所占的比例越小。当并行流的数目达到100个时，性能基本上达到饱和。实际上，并行流的数目并非越多越好，这与网络延迟与网络带宽有非常密切的关系。在实际应用中，合适的并行流的值一般需要根据大量的测试来决定。图中曲线的波动是由于广域网的带宽拥塞引起的，因此对于海量数据的传输任务，高带宽的专用网络有利于改善网格系统的性能。

YBJ-ARGO实验还利用edg-job-submit对ARGO重建作业和模拟作业进行了提交，并通过lcg-cr进行重建结果数据文件的下载，最终对重建结果与PC下的结果进行比较分析，结果完全一致，表明了网格上的重建作业的完成结果是可靠的。

5. 网络状况

为了支持羊八井ARGO实验的数据传输，中

国科技网(CSTNet)、中国教育和科研计算机网(CERNET)、中欧跨欧亚信息网络(TEIN2)和中欧学术网高速互联网(ORIENT)等项目共同合作。当前的网络拓扑如图5所示。

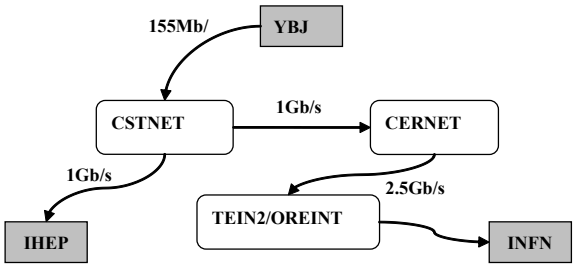

图5 羊八井ARGO实验数据传输网络拓扑图

羊八井宇宙线观测实验基地采集的原始数据通过网络同时向IHEP和INFN传输，即保存两份拷贝。YBJ与IHEP之间直接通过155Mb/s的CSTNET专线，传输性能如图6所示。YBJ与INFN之间通过CSTNET、CERNET、TEIN2/ORIENT等网络，有较大的延迟，传输性能如图7所示。如果从IHEP到INFN进行数据传输，则速度要快很多。

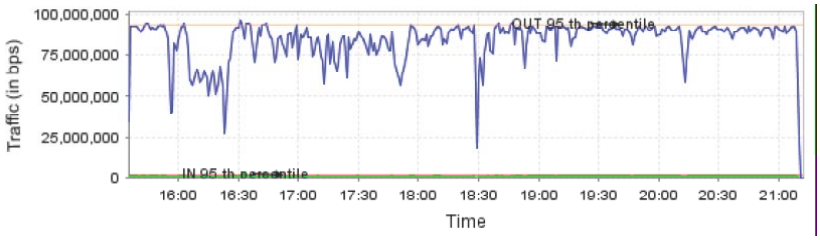

图6：从YBJ到IHEP的网络传输性能(2008年4月23日的统计结果)

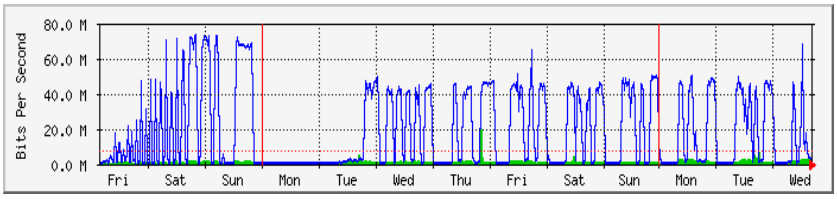

图7：从YBJ到INFN的网络传输性能 (2008年3月27日到4月3日的统计结果)

6. 总结与展望

本文对YBJ-ARGO实验的基本状况和网格应用模型进行了描述，并且对站点的网格基础设施建设做了介绍，包括VO的建立、数据的存储和传输、作业的提交、结果的获取等。网格应用模型的建立以及相关工具的开发，为YBJ-ARGO数据处理实现网格化奠定了基础。但是YBJ-ARGO实验的网格系统仍处于起步阶段，站点以及资源数量有待增加，国际网络链路需要得到改善。可以期待，随着YBJ-ARGO网格系统规模的不断扩展，数据的传输与处理效率将会得到极大提高，从而帮助物理学家更快的产生更高质量的科学成果。

网格技术通过十多年的发展已经趋于成熟，当前网格的发展已经从技术的研发向应用推广和部署转移。如何针对应用的特点选择适用的网格中间件，是目前应关注的重点。从本文介绍的工作可以看出，将一个特定的科学计算应用部署到网格平台上，首先对应用的特性与需求进行细致的分析并选择合适的中间件，然后基于中间件开发应用的接口，实现应用的部署。网格应用的成功与否，另一个关键因素是必须根据应用的实际情况建立严格的运行机制，确保系统的正常高效运行。

高速广域网是网格平台建设的基础。只有借助高速网络才能将跨地域的计算资源整合起来，形成真正意义上的网格系统。由科学院网络中心帮助架设的羊八井到北京的专用光纤网络将羊八井宇宙线观测基地与网格数据中心链接起来，彻底改变了科学研究的数据获取与分析模式，为羊八井宇宙线观测实验提供了重要的基础。

7. 致 谢

本工作是中国科学院科学数据库及信息化支持的项目。特此感谢！

参考文献

[1] Ian Foster, Car Kesslman.The Grid 2[M]. 北京: 电子工业出版社, 2004:224-262.

[2] 何会海等. 羊八井ARGO实验的RPC性能. 高能物理与核物理. 2004, 28（04）:422.

[3] A. Chervenak, I. Foster, C. Kesselman, C. Salisbury and S. Tuecke.The data grid: towards an architecture for the distributed management and analysis of large scientific datasets, Journal of Network and Computer Applications 23 (2001), pp. 187–200.

[4] EUChinaGrid. web: 2008.4, http://www.euchinagrid.org.

[5] gLite. web: 2008.4.http://glite.web.cern.ch/glite.

[6] A.Budano, P. Celio, R. Gargana, F. Galeazzi, F. Ruggieri, C. Stanescu, Y.Q. Guo, L. Wang, X.M. Zhang. A Solution for Data Transfer and Processing Using a Grid Approach. BIO-ALGORITHMS & MED-SYSTEMS, no.5 2007 Grids in Science, Krakow, Poland: oct.2007.

作者信息

陈 刚

中国科学院高能物理研究所。1994年毕业于中科院高能物理研究所和瑞士苏黎世联邦技术大学（ETHZ），获博士学位。2003年起任高能物理研究所计算中心主任，主持建设用于高能物理的高性能计算平台。2004年与欧洲粒子物理中心以及意大利建立合作关系，开始在中国建立高能物理网格站点，为LHC等高能物理实验提供网格平台支撑。2005年当选国际高能物理计算协调委员会（IHEPCCC）委员。2006年起任中国科学院科学数据库专家委员会副主任。

程耀东

中国科学院高能物理研究所。2006年毕业于高能物理研究所并获博士学位。2006年至2008年为高能所博士后从事海量数据存储技术研究和网格技术研究。