

联系信息

北京市, 海淀区
蓟门里东 2 楼 104
邮政编码: 100088

电话: 13911635527
邮箱: gnawux@gmail.com
主页: <http://wangxu.me/blog/>

个人介绍

接近 4 年的 Hadoop HDFS 开发经验, 中国移动研究院大云团队核心开发成员, 负责 NameNode 元数据管理和 DataNode 磁盘管理相关等方面的开发项目, 与 Hadoop 开源社区中的国内外开发者保持密切联系。超过 10 年的 Linux 使用与维护经验, 熟练使用 C/C++, Java 等多种开发语言, Python, Perl 等脚本语言及各种 Unix 工具进行开发与日常维护工作。

工作意向

位于北京的 Apache Hadoop 平台相关的开发与维护工作, 主要是 HDFS 相关的工作; 或分布存储系统相关的研发工作。

当前工作

Hadoop 相关工作

中国移动研究院研究员/项目经理 (2007 年 7 月至今), 负责 Hadoop HDFS 相关开发维护工作以及开源社区交流工作, 包括 Hadoop 的评估测试、问题修复、功能改进等 (Apache JIRA ID gnawux):

- 2010-2011 年, 为 Trunk 提交磁盘刷新功能补丁 HDFS-1362;
- 2008 年, 提交补丁 HADOOP-4742, 修复 Hadoop 0.18 丢失数据块副本的严重 Bug;
- 2008 年 4 月至今, 维护中国移动研究院的 Hadoop 分支, 包含多 NameNode 同步补丁;
<http://github.com/gnawux/hadoop-cmri>
- 2008-2009 年, 开发并维护中国移动研究院的 Hadoop 测试工具 (Python);
<http://code.google.com/p/hadoop-test/>
- 2010 年 9 月, 参加 Hadoop in China, 介绍中国移动研究院的 Hadoop 开发情况;
<http://wangxu.me/blog/p/445>
- 2009 年 11 月参加 Apache Roadshow Asia 2009, 介绍中国移动研究院在 HDFS 方面的工作;
- 开发工作部分记录: <http://gnawux.info/hadoop/>

其他工作

- 【Linux 集群管理】 2010 年至今, 中国移动通信研究院大云平台系统管理员, 负责系统配置维护、性能优化, 涉及到虚拟化平台和 Hadoop 平台。
- 【云存储】 2010 年至今, 项目经理与设计人员, 负责中国移动研究院的云存储项目。
- 【书】 负责中国移动研究院云计算技术相关图书编写工作。
- 【译】 业余时间翻译 Cassandra: The Definitive Guide 一书, 预计将于 2011 年 4 月完成, 由图灵图书出版。

过往工作

Linux 及其他开源相关工作

- 【书】《Debian 标准教程》, 2004-2008 年持续更新、完善, 最终于 2009 年 9 月由人民邮电出版社出版。
- 【活动】LinuxFB/北邮 BBS Linux 每月聚会活动联合发起人和组织者。
- 【文章】 常年在业余时间写/译关于 Linux Kernel, 文件系统, 虚拟化, NoSQL 等技术文章, 发于个人网站 <http://wanxu.me/blog/>, 其中关于 Zookeeper 的译文发于《程序员》2010 年第 2 期。
- 【L^AT_EX】 2007 年制作并发布了北邮研究生毕业论文 L^AT_EX 模板第一版的代码、样例和说明。
- 【编辑】 2004 年曾担任 LinuxFocus (随 Linux 文档项目 (LDP) 发布的一份自由多语言杂志) 志愿编辑, 负责管理中文部分的页面, 同时还负责翻译和撰写技术文章。

培训讲师 2005 年至 2007 年
在北京、广东、武汉、天津等地为多家运营商、开发商进行 Linux 和 3G (WCDMA/TD-SCDMA) 的核心网、业务及无线接入网相关员工培训和认证培训。

软件开发人员 2002 年至 2004 年
参与星河亮点通信软件有限公司 PHS 路测软件开发 (Visual C++, 核心开发人员) 和移动 Agent 相关国家项目 (Java, 项目组长)。

教育背景

北京邮电大学

电信工程学院, 工学博士, 2002 年 9 月至 2007 年 7 月
• 导师: 张平 教授

北京工业大学

电子工程系, 工学学士, 1998 年 9 月至 2002 年 7 月

参考

部分技术文章

Hadoop 相关技术文章 (<http://wangxu.me/blog/p/tag/hadoop>)

- DataNode 的磁盘管理以及 HDFS-1362: <http://wangxu.me/blog/p/516>
- HDFS 退服节点的方法: http://labs.chinamobile.com/mblog/225_18378
- dfs.datanode.max.xcievers: hadoop 的一个无文档的参数:
http://labs.chinamobile.com/mblog/225_18094

系统管理与版本管理相关文章, 多数是管理过程中的问题分析处理与案例和脚本记录:

- 用 SVM 同步 SVN 仓库及解决锁问题: <http://wangxu.me/blog/p/523>
- 同时操作多台主机的命令行: <http://wangxu.me/blog/p/166>
- 一行的日志分析脚本: <http://wangxu.me/blog/p/145>

部分翻译文章

NoSQL、Hadoop 与 Java 相关技术文章翻译:

- Cassandra 实例 <http://wangxu.me/blog/p/383>
- Observers: 让 ZooKeeper 更具可伸缩性: <http://wangxu.me/blog/p/320>
- Avro: 大数据的数据格式: <http://wangxu.me/blog/p/279>
- Java SE 6 Hotspot TM 虚拟机垃圾回收调优: <http://wangxu.me/blog/p/209>

Linux 内核与图形相关以及其他软件相关技术文章翻译:

- LWN: 空指针的乐趣(1): <http://wangxu.me/blog/p/39>
- KernelTrap: Linux: The Journaling Block Device: <http://wangxu.me/blog/p/151>
- LWN: GEM vs. TTM: <http://wangxu.me/blog/p/124>
- LWN: What is RCU, Fundamentally?: <http://wangxu.me/blog/p/122>

注: 所有译文的英文原文都来自于互联网, 并在开头明确注明原文发布时间、链接、作者和翻译时间, 包括 LWN、Cloudera 博客等在内的大部分文章在翻译时取得了发布者的授权。