**Computer Science and Application** 计算机科学与应用, 2019, 9(8), 1483-1488
Published Online August 2019 in Hans. http://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2019.98166

Hans汉斯

# Research on NetCDF Files Publishing Technology Based on Web Services Interface

**Fengjuan Cui[1,2], Jing Lin[1], Peng Yang[1]**

[1]North China Sea Data and Information Service of State Oceanic Administration, Qingdao Shandong
[2]Provincial Key Laboratory of Marine Ecological Environment and Disaster Prevention and Mitigation, Qingdao Shandong
Email: 407177515@qq.com

Received: Jul. 20th, 2019; accepted: Aug. 2nd, 2019; published: Aug. 9th, 2019

## Abstract

**NetCDF is a general network data format widely used in the field of ocean and atmosphere, with good self-description and platform independence. NetCDF now has been the standard of data exchange in the field of the oceanic atmosphere. Because a large number of NetCDF files have to be distributed through the network, to achieve fast file selection and preview of NetCDF files by avoiding large file download with limited bandwidth limitation, the Internet cloud service technology provides a solution to this problem. This paper makes use of SOA framework and NetCDF Java to provide Web Service interface to extract the necessary information of NetCDF files, and partly solves the contradiction between network bandwidth limitation and large file download. The application of the result has some universal applicability.**

## Keywords

**NetCDF, Web Service, Java, SaaS**

# 基于网络服务接口的NetCDF文件发布技术研究

**崔凤娟[1,2]，林　婧[1]，杨　鹏[1]**

[1]国家海洋局北海信息中心，山东 青岛
[2]山东省海洋生态环境与防灾减灾重点实验室，山东 青岛
Email: 407177515@qq.com

收稿日期：2019年7月20日；录用日期：2019年8月2日；发布日期：2019年8月9日

## 摘　要

**NetCDF是广泛应用于海洋大气等领域的通用网络数据格式，具有良好的自描述性和平台无关性，已经成**

文章引用: 崔凤娟, 林婧, 杨鹏. 基于网络服务接口的 NetCDF 文件发布技术研究[J]. 计算机科学与应用, 2019, 9(8): 1483-1488. DOI: 10.12677/csa.2019.98166

**为事实上的海洋大气领域数据交换标准。在大量NetCDF通过网络进行发布的前提下，要实现有限的带宽限制下避免大型文件下载操作来实现对NetCDF文件的快速选取和数据预览，互联网的云服务技术为这种需求提供了解决方法。本文利用SOA框架结合NetCDF Java提供Web Service服务接口实现对NetCDF文件必要信息的快速抽取，部分的解决了网络带宽限制和大文件下载之间的矛盾，成果应用具有一定的普遍适用性。**

## 关键词

**网络通用数据格式，Web服务，Java，PaaS平台**



## 1. 引言

NetCDF (Network Common Data Format，网络通用数据格式)是一种网络通用的数据文件格式，也是支持这种数据格式的一套软件库(Software Library)的通称。它是由 NSF (National Science Foundation，美国国家科学基金会)资助项目 Unidata 计划执行过程的开发成果，设计之初的目的是为该计划下的各种应用提供通用的数据访问方法，NetCDF 所涉及的资料的类型包括单点的观测值、时间序列、规则排列的网格数据以及卫星或雷达的图像数据存储与交换。

NetCDF 数据文件具有自我描述、平台无关、支持随机访问、支持追加、共享性好、版本延续性好的优点[1]，而且能够存储海量的面向阵列(array-oriented)数据，现在已经成为许多数据采集软件生成文件的标准格式，被广泛用于陆地、海洋和大气科学。

目前，NCEP(National Centers for Environmental Prediction，美国国家环境预报中心)发布的再分析资料，NOAA (National Oceanic and Atmospheric Administration，美国国家海洋和大气管理局)发布的 COADS (Comprehensive Ocean-Atmosphere Data Set，海洋与大气综合数据集)均采用 NetCDF 作为标准[2]，NetCDF 已成为一种跨学科的标准数据格式[3]。

## 2. 问题描述

尽管 NetCDF 是为了网络文件共享的目标所设计，但实际上更多的应用还是利用网络这个途径使用 ANSI 或 UNICODE 编码的文件进行数据交换，这样的数据文件一般以“.nc”为扩展名，可以通过 HTTP 或 FTP 等方式下载并在本地进行应用。

由于 NetCDF 文件具有广泛的适应性和良好的扩展性，近年来网络上以 NetCDF 为格式共享的文件持续呈爆发性增长。在 IPCC (Intergovernmental Panel on Climate Change，海洋与大气综合数据集)第五次评估报告(AR5)的 CMIP (Coupled model Intercomparison Project，耦合模式比较计划)中，各模式数据中心贡献了超过 1.5 PB 的 NetCDF 数据[4]。以 Unidata 网站上的 NetCDF 样例数据为例，NetCDF 数据文件的大小从 2.8~281.4 M 不等。而实际应用中，一些模式输出的 NetCDF 文件可以达到 2G 甚至更大，或者是输出一系列的几百 M 字节的数据文件。当这些数据文件通过网络进行发布的时候，用户需要将 NetCDF 文件的下载到本地，然后利用 NetCDF 程序包扩展开发的各类应用软件对文件进行操作，100 M 以内的数据文件操作尚可，过大的文件操作就显得非常不方便。

为了解决海量数据应用和单个及系列大文件使用的矛盾，NetCDF 的应用人员和技术人员已经使用了很多先进的方法提高数据用户的应用体验。早期的 NetCDF 数据发布往往利用关系型数据库对网络的 NetCDF 文件发布过程进行处理，将 NetCDF 文件根据其变量、维度和数据内容拆分成不同的数据表，并提供了数据库接口用于远程调用，用户也可以利用接口抽取一部分数据，从而避免了大数据量下载所带来的等待和延迟等问题。但是这种方法也具有相当的缺点，首先，NetCDF 的原始资料多种多样，维数多且结构复杂，使用数据库进行管理很困难；其次，气象水文资料的数据量大，关系型数据库难以满足网络环境下大量的多维数据存取的要求；最后，许多用户在下载数据后，还要手工把这些数据再转换成 NetCDF 格式才能在自己的模式或应用中使用，使用过程非常不方便。

XML 结构的元数据的出现也为 NetCDF 的应用提供了一定的便捷途径。用户上传成果数据的同时发布系统就在后台对文件进行解析，在生成数据表索引的同时将变量、维度和数据的说明抽取到一个结构相对简单的 XML 文件中作为元数据，更进一步的也可以抽取部分样例数据或生成 NetCDF 数据的缩略图。这种方式可以让用户对要下载的 NetCDF 文件具有一个清晰和直观的认识，清华大学地球系统科学研究中心的数据共享发布系统就采用了类似方法来提高用户体验[5]，但这种方式没有从本质上解决 NetCDF 文件的下载到应用的瓶颈问题。

借鉴互联网上媒体文件(如电影等媒体)发布的技术和文件下载技术，也有人探索使用 SM (Stream Media，流媒体技术)和 P2P (Point to Point，对等互联网技术)来提高 NetCDF 文件使用过程中的下载速度。

Unidata 组织也在 NetCDF 的网络应用上也进行了扩展开发，最典型的就是利用 OPeNDAP (Open-Source Project for a Network Data Access Protocol，开放网络数据访问协议)协议支持获取数据子集的特性，通过 Web 应用系统对 NetCDF 等结构文件的支持，用户可以使用检索字符串在服务器端完成对 NetCDF 文件的抽取并获取需要的数据[1]。这种方式可以让专业的数据使用人员快速熟悉并获取数据，但不完全适合非计算机专业的人员在应用系统开发过程中的应用。

## 3. 解决方法

互联网云服务技术的发展为海量大型 NetCDF 文件的存取技术提供了新的思路，NetCDF 的数据获取服务和对数据更进一步的应用服务可以归为云 SaaS (Software as a Service，软件即服务)这一层次中。

借鉴成熟系统设计的经验，NetCDF 数据发布和应用系统在获取新的数据发布时首先要生成各个 NetCDF 文件的索引和元数据进行发布，在实际应用过程中利用 Java 的 Web Service 服务技术和多线程技术在云端对大型的 NetCDF 文件进行预处理并提供一定的网络服务接口。接口可以根据用户请求内容在云端对数据文件直接进行处理和抽取并返回信息，其处理和抽取过程可以是同步的，也可以是异步的。

一个典型的 NetCDF 的发布、检索和应用流程如图 1 所示：成果发布用户上传的文件经预处理后形成数据库中的表索引，以及原始文件和 3 类辅助性文件，成果应用用户可以由“Web 检索页面”进入并检索数据，通过 HTTP/FTP 服务获取 XML 格式的元数据文件，实际使用和分析可以直接通过 HTTP 和 FTP 方式获取较小的 NetCDF 样例文件。当服务器的 NetCDF 文件过大或对数据结果不确认的情况下，则可以通过 Web Service 接口根据需要获取更详细的数据片段进行分析后再决定是否下载等操作。

为了更好的与互联网服务技术进行结合，需要确定适合 NetCDF 文件发布的 Web Service 及开发环境，以及与该开发环境匹配的 NetCDF 软件库。

### 3.1. NetCDF Java 简介

如前所述，NetCDF 也是一套支持 NetCDF 格式数据读写操作的软件包的名称，这套软件包提供了支持包括 C、Fortran77、Fortran90、C++、Java 和 Python 等多种语言的调用接口。

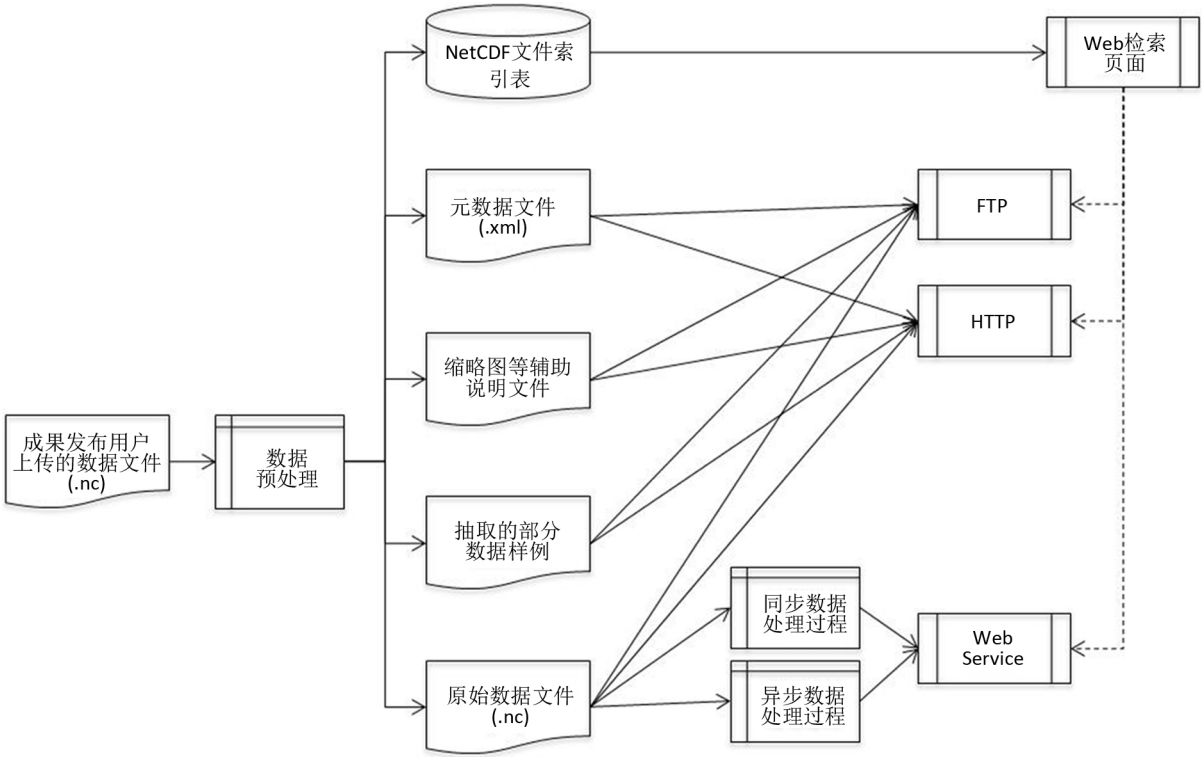


**Figure 1.** NetCDF file processing and publishing process diagram
**图 1.** NetCDF 文件处理和发布流程示意图

NetCDF Java 库具备了对 CDM (Common Data Model 通用数据模型)的较为完整的支持，提供了一系列的 Java API 接口来实现对 NetCDF、OPeNDAP 和 HDF5 (Hierarchical Data Format，层次数据结构)文件格式的处理，接口涵盖了我们对本地 NetCDF 文件操作时所需要的全部功能。NetCDF Java 库仅支持对 NetCDF-3 格式文件的写入，如果要写入 NetCDF-4 格式的文件，还需要编译后的 C 程序包支持，TDS (THREDDS Data Server)等开源数据服务系统就是基于 NetCDF Java 所开发的[6]。

### 3.2. SOA 和 Web Service

SOA (service-oriented architecture，面向服务的体系结构)是一个组件模型，它将应用程序的不同功能单元通过服务之间定义良好的接口联系起来，接口采用中立的方式进行定义并独立于实现服务的硬件平台、操作系统和编程语言，这种具有中立的接口定义的特征称为服务之间的松耦合[7]。

Web Service 技术是 SOA 的具体实现。Web Service 是自描述、自包含的可用网络模块，可以包含一个或多个具体的业务功能。Web Service 基于一些常规的产业标准以及已有的一些技术，如 XML (Extensible Markup Language，可扩展标记语言)、SOAP (Simple Object Access Protocol，简单对象访问协议)、WSDL (Web Services Description Language，网络服务描述语言)和 UDDI (Universal Description, Discovery and Integration，通用描述、发现与集成服务)等，容易部署和维护[8]。Web Service 技术能使得运行在不同机器上的不同应用无须借助附加的、专门的第三方软件或硬件，就可相互交换数据或集成。随着互联网应用服务技术的发展，许多开源的 Web Service 支持平台已经被广泛应用于各类系统开发过程中，如 JBoss SOA 和 Apache CFX 等。

利用 Web Service 技术可以直接提供对单个 NetCDF 的数据获取操作，也可以将多个 Web Service 组合起来形成一个服务链来提供更多的应用支持，例如直接形成图件或加载在在线地理信息平台上。Web

Service 技术延续了 NetCDF 文件的自我描述和平台无关性等优点，使得对 NetCDF 文件的应用用户可以以较小的转换成本完成对数据文件的应用。此外，利用 Web Service 技术发布 NetCDF 数据，还可以使不熟悉 NetCDF 的通用软件开发人员直接利用原有的系统代码获取数据并在原有的平台上进行展示和操作，降低了 NetCDF 文件利用的技术门槛。

## 4. 应用实例

以混合坐标大洋环流模式(HYCOM)为例，水平方向采用麦卡托网格，分辨率为 0.5˚的表层温盐场和流场全球模拟结果存储在服务器上的数据量达到 7 G 左右，模式成果数据的应用人员每次都要花费大量时间在数据下载和比较过程中。利用 Apache Tomcat 服务器作为载体，部署基于 Servlet 的 Web 应用程序和 Apache CXF 的 Web Service 服务，以 Web 页面提供查询入口，通过查询结果可以完成元数据查看和下载，也可以进入到 Web Service 的服务页面。每个(组) NetCDF 文件的 Web Service 服务页面提供了一系列的接口方法支持用户以不同的方式和数据量获取与该文件相关的全部或部分数据，从而避免了大量的数据文件下载和挑选的过程，可以有效的提高 NetCDF 成果数据的应用效率。

在系统的接口定义过程中，需要先定义与 NetCDF 对应的类，并将 NetCDF 的操作分为面向“维(Dimensions)、变量(Variables)、属性(Attributes)和数据(Data)”四类接口(Interface)，在每类接口下按“定义、查询、提取和写入”分别定义针对性操作方法(Function)，典型代码如图 2 所示。

```
import java.util.List;

@WebService
public interface IDimensionsFactory {
    public List<Dimensions> getDimensions();
}
```

```
import java.util.List;

@WebService(endpointInterface = "DimensionssFactoryImpl", serviceName = "DimensionsUtils")
public class DimensionssFactoryImpl implements IDimensionsFactory {

    public DimensionssFactoryImpl() {
    }

    public List<Dimensions> getDimensions() {
        List<Dimensions> list = null;
        return list;
    }

}
```

**Figure 2.** Web service interface code diagram for NetCDF
**图 2.** NetCDF 的网络服务接口代码示意图

## 5. 结论

应用实践证明，利用 Web Service 技术调用后台同步和异步的 NetCDF 文件操作，结合 NetCDF 数据文件和元数据索引技术提供数据的共享功能，可以有效的减少模式成果对比和应用过程中的数据传输量，加强了多模式输出数据的实际对比分析能力，可以一定程度上解决网络带宽对大型 NetCDF 数据发布和应用的限制，从而为海洋大气与水文环境等信息发布提供了新的方法和途径，也为海洋大气水文环境要素等数据密集型信息基础设施(CI, CyberInfrastructure)理念的实践提供了依据[9] [10]。

## 参考文献

[1] 杨猛. 基于 OPeNDAP 协议的海洋数据文件共享平台设计与实现[D]: [硕士学位论文]. 青岛: 中国海洋大学,

2011.

[2] 孙建伟, 孙昭晨, 陈轩, 耿红. netcdf 格式数据的创建及应用[J]. 交通标准化, 2010, 3(15): 31-34.

[3] NetCDF (Network Common Data Form). http://www.unidata.ucar.edu/software/netcdf

[4] Nemoto, T. and Kitsuregawa, M. (2013) Surveying Systems of Global Climate Change Simulation Data and Access Logs to Them. *IEEE International Congress on Big Data*, Santa Clara, 6-9 October 2013, 342-346.

[5] 徐灏, 白玉琪. 海量 NetCDF 空间数据的分布式协同分析环境设计与实现[C]//中国地理学会地图学与地理信息系统专业委员会, 中国地理信息产业协会. 第六届全国地理信息科学博士生学术论坛论文集. 2014: 408-410.

[6] https://www.unidata.ucar.edu/software/thredds/current/netcdf-java

[7] 苏文鹏. 多框架 Web Services 统一调用组件的设计与实现[D]: [硕士学位论文]. 济南: 山东大学, 2014.

[8] 尉飞新. 用 Web Services 实现基于 SOA 的企业应用集成研究[D]: [硕士学位论文]. 上海: 同济大学, 2007.

[9] National Science Foundation, Cyberinfrastructure Council (2007) Cyberinfrastructure Vision for 21st Century Discovery. Alexandria.

[10] Yang, C., Raskin, R., Goodchild, M., *et al*. (2010) Geospatial Cyberinfrastructure: Past, Present and Future. *Computers, Environment and Urban Systems*, **34**, 264-277. https://doi.org/10.1016/j.compenvurbsys.2010.04.001