

董贵莹,曹敏杰,张丰,等. Argo 资料协同管理方法研究[J]. 海洋学研究,2017,35(3):1-8,doi:10.3969/j.issn.1001-909X.2017.03.001.

DONG Gui-ying, CAO Min-jie, ZHANG Feng, et al. Research on collaborative management method of Argo data[J]. Journal of Marine Sciences, 2017,35(3):1-8,doi:10.3969/j.issn.1001-909X.2017.03.001.

# Argo 资料协同管理方法研究

董贵莹<sup>1,2</sup>,曹敏杰<sup>3,4</sup>,张 丰<sup>1,2</sup>,杜震洪<sup>\*1,2</sup>,刘仁义<sup>1,2</sup>,吴森森<sup>1,2</sup>

(1. 浙江大学 浙江省资源与环境信息系统重点实验室,浙江 杭州 310028;2. 浙江大学 地理信息科学研究所,浙江 杭州 310028;3. 卫星海洋环境动力学国家重点实验室,浙江 杭州 310012;  
4. 国家海洋局 第二海洋研究所,浙江 杭州 310012)

**摘要:**Argo 资料已成为当前海洋和大气科学领域基础研究的重要数据来源。但由于其剖面元数据与观测数据混合存放的特点,现有的共享平台无法实现浮标漂移轨迹与剖面图的实时绘制。因此本文提出一种结构化与半结构化并存的 Argo 资料协同管理方法,通过分析 Argo 资料结构组成与特点,将结构化数据与半结构化数据分离提取;然后利用关系型数据库对半结构化类型属性的扩展支持,建立剖面元数据与观测数据间的关联关系;并利用分表存储,降低数据量快速增长对单数据表带来的存储压力。最后通过对近 20 a 的全球 Argo 资料解析建库结果进行分析,证明该方法具有良好的可扩展性和高效的轨迹数据获取效率,能够支持浮标漂移轨迹和剖面图的实时绘制。同时,该方法也可为特征类似的剖面观测数据管理提供技术参考。

**关键词:**Argo 资料;剖面;结构化与半结构化;协同管理

**中图分类号:**P71; TP311.52

**文献标识码:**A

**文章编号:**1001-909X(2017)03-0001-08

**Doi:**10.3969/j.issn.1001-909X.2017.03.001

## 0 引言

国际 Argo (Array for Real-time Geostrophic Oceanography)计划是由美国、法国和日本等国的科学家于 1998 年推出的全球大洋观测计划<sup>[1]</sup>。我国已于 2002 年正式宣布参加国际 Argo 计划的组织实施。截至 2016 年 3 月 31 日,该计划已在全球海洋投放约 12 000 个浮标,提供近 150 万条温、盐剖面。随着浮标和传感器技术的发展,Argo 浮标在采样密度、传感器测量精度等方面都得到了显著的提高,且增加了溶

解氧、叶绿素等多种类型传感器<sup>[2]</sup>。Argo 浮标获取的垂直剖面数据时间序列可以反映大尺度海水温、盐变化,对研究海水热盐储量/输送、大洋海水特性等具有重要意义<sup>[3-5]</sup>。浮标在“随波逐流”过程中的实时位置信息直接体现了海水的流动状态,对研究海洋环流、中尺度涡、湍流等具有重要意义<sup>[2,6-8]</sup>。Argo 资料已成为当前海洋环境和气候变化研究的重要数据来源。

Argo 观测资料主要通过光盘和网络下载进行免费共享,如中国 Argo 实时资料中心、法国海洋开发研究院 (IFREMER)、日本海洋地球科技厅 (JAM-

收稿日期:2016-12-16

修回日期:2017-03-31

基金项目:国家自然科学基金项目资助(41101356,41101371,41171321);国家科技基础型工作专项项目资助(2012FY112300);海洋公益性行业科研专项经费项目资助(201305012);测绘地理信息公益性行业科研专项项目资助(201512024)

作者简介:董贵莹(1992—),女,天津市人,主要从事海洋 GIS 相关研究。E-mail:413521577@qq.com

\* 通讯作者:杜震洪(1981—),男,副教授,主要从事地理信息服务、时空大数据、高效能地学计算、海洋 GIS 研究。

E-mail:duzhenhong@zju.edu.cn

STEC)等均提供上述服务。这些途径获取到的 Argo 浮标资料按日期分类压缩,下载后需根据研究区域逐个筛选,由于 Argo 观测网覆盖全球且观测密度较大,Argo 资料数据量呈指数增长,这种数据提取方式会越来越难以实施。目前已存在的一些基于网络的 Argo 数据共享平台(如 Argo Information Center、Argo Portal 等),可就浮标的属性信息进行查询(如浮标编号、投放位置等),并对查询所得结果提供浮标的历史剖面、剖面时序、浮标漂移轨迹和区域内浮标分布等图。但这些图的灵活性、可读性有限,如 PNG 类型的剖面图无法根据个人需求获取观测点的准确观测值;产品图浪费了存储资源且不利于更新,如浮标分布图、浮标轨迹图等,理论上浮标每更新一次地理位置就需将所涉及的产品图重新绘制。考虑到剖面资料的增长速度(约每月 1 万条),理想的解决方案是按需绘制、实时绘制,但由于 Argo 剖面资料本身的数据特点——元数据与剖面观测数据混合存放,使得浮标漂移轨迹与剖面图的实时绘制无法实现。

由于 Argo 资料内部同时存在结构化与半结构化数据,传统的数据管理方式仅依靠关系型数据库管理元数据属性字段,无法同时有效地管理半结构化剖面观测数据<sup>[9-11]</sup>。近年来,随着互联网的发展,半结构化数据管理的重要性与日俱增,越来越多的关系型数据库也大力优化了对半结构化数据的支持,比如 JSON(JavaScript Object Notation)数据<sup>[12-13]</sup>,使得关系型数据库在保持原有的高查询效率的同时具备了一定的半结构化数据管理能力。

基于以上背景,本文从分析 Argo 剖面资料本身的数据特点出发,分析其结构特性,提出一种结构化与半结构化并存的 Argo 资料协同管理方法,并建立 Argo 数据库。该方法综合考虑了 Argo 剖面资料数据量的增长趋势与观测剖面类型的变化趋势,能够满足长期的存储与检索需求,同时能够更加便捷地实现浮标漂移轨迹和剖面图的实时绘制。

## 1 Argo 资料来源与结构组成

### 1.1 资料来源

中国 Argo 计划启动实施之初,即在杭州建立了中国 Argo 实时资料中心(China Argo Real-time Data Center,以下简称为 C-ARDC),负责对我国布放的各种类型的剖面浮标观测数据进行接收、处理和分发<sup>[14]</sup>。本文所使用的全球 Argo 资料由 C-ARDC 提供,收集了 1997 年至今全球海洋中所有 Argo 剖面浮

标的观测资料。数据格式参考 C-ARDC 发布的“全球 Argo 剖面浮标资料集(V2.0)说明”<sup>[15]</sup>。

### 1.2 Argo 资料结构组成

C-ARDC 提供的 Argo 剖面浮标资料为 dat 类型的文本数据。Argo 资料包括元数据与剖面数据两部分,其中剖面数据又分为表头数据和实际观测数据(图 1)。元数据包含对每个 Argo 浮标的详细说明,包括浮标的技术参数、传输定位卫星信息、布放信息、传感器信息以及观测周期信息等内容。剖面数据以文件的方式存储,表头信息存放该条数据的描述信息,如浮标号、循环号、经纬度和观测日期等;实际观测数据记录了海水温度、压强、盐度和溶解氧浓度等观测值及其对应的校正值和观测值质量控制标记等。2.0 版本的 Argo 剖面浮标资料新增了叶绿素观测数据,同时还增加了多深度轴的剖面数据。

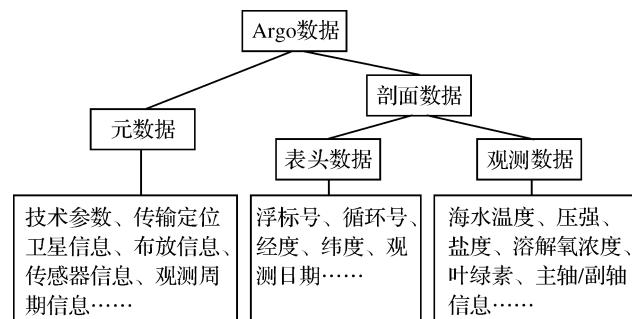


图 1 Argo 数据结构

Fig. 1 Argo data structure

按照数据结构可以将数据大致分为三类:结构化数据、非结构化数据和半结构化数据。C-ARDC 提供的 Argo 资料数据以统一的格式存储于文本文档中。其中元数据字段类型、字段数目在一定时间内固定不变,属于结构化数据,适宜存储于关系型数据库中,可以用二维表结构来逻辑存储,需要通过对数据进行分析并设计合适的表结构和表间关系。剖面数据中的表头数据与元数据相同,属于结构化数据,组织存储方法类似。剖面数据中的观测数据则由于不同浮标类型以及浮标装载的传感器不同,且采样间隔不均匀等多种因素影响,无法做到固定不变,但剖面观测数据的数据组成具有统一化标准,因此属于半结构化数据。

## 2 结构化与半结构化数据协同管理方法

针对 Argo 资料数据结构化与半结构化并存的特性,首先设计数据分类和提取规则,对结构化数据和

半结构化数据进行分离;之后针对所提取的结构化数据部分对关键字段进行信息抽取,针对半结构化数据部分设计合理和可扩展的 JSON 数据结构,并将其作为结构化数据的一个字段存储于数据库表中;最后根据整个组织存储结构设计数据库表结构和表间关系。

## 2.1 元数据质量梳理

C-ARDC 提供的 Argo 元数据字段是按照数据操作人员的习惯进行设计的,存在数据分类混乱、语义一致性不高、信息分散等缺点,不便于直接进行存储。因此需要对原始数据进行元数据质量梳理。待梳理的数据可以分为三类:普通说明性数据、剖面数据和空间数据。

普通说明性数据存在两个问题:(1)多个来源,(2)重复定义。其中第一个问题以项目所属国家为例,国家信息并未存放于 Argo 元数据文件中,而是在每 3 个月一次的数据更新时附带一份国家信息文件,内含每个浮标的所属国家信息。元数据文件以浮标编号命名,每个浮标拥有一个元数据文件(如 5900586\_meta.dat)。国家信息文件中所有浮标的所属国家均存储于一个文件中,每行一个浮标。在数据预处理时应将多个来源的浮标元数据信息进行归并,属于同一个浮标的元数据通过浮标编号进行唯一匹配。第二个问题通常是由于某些原因浮标元数据文件中的字段名会出现大小写混乱和末尾多出空格的情况,这种情况下应该在数据入库前,对数据字段进行统计分析,并根据分析结果重新入库,如“CHINA ARGO PROJECT”和“china argo project”应为同一个项目名称,尽量提高入库数据的语义一致性。

剖面数据来自于浮标所产生的剖面文件中的统计信息。由于浮标元数据与剖面数据分开存放,要查找有关某浮标的剖面统计信息,如最新一个剖面位置、最后一个剖面产生的时间等,将非常麻烦。在数据入库时,可提前设计好一些待用的统计字段,使用数据库的触发器功能,在剖面数据更新时,对其浮标元数据的相关统计字段同步更新。

此处所涉及的空间数据为元数据辅助数据,根据前期对数据获取需求的调研发现,某些海区或某些海峡是科学研究所常被选取的区域。由于海洋中的海区、海峡范围通常为复杂多边形,因此,平台在存储 Argo 元数据的同时,也对海域范围元数据进行预存储。提前获取“兴趣海区”范围,并利用 PostGIS 空间数据库引擎,将矢量的海区范围转储为空间字段。设计“兴趣海区”数据表,存储海区范围、海区名称等。

## 2.2 剖面观测信息提取

浮标工作期间所采集的海水剖面数据以文本的形式存储于原始资料文件中。剖面观测数据与剖面元数据一起存储于以浮标编号和循环号命名的文件中(如 1900727\_258.dat)。其中剖面观测数据存放于文件的末尾,分为两部分:列说明数据和具体观测数值矩阵。列说明数据行数通常为 3 的倍数,每 3 行代表一个采样要素。如 1900727 号浮标的第 258 条剖面具有压力、盐度和温度三个观测要素,因此列说明数据有 9 行。不同浮标的列说明数据的行数和内容可能会不同。观测数值矩阵每行代表一个采样点,每列之间以空格隔开,每一列对应列说明数据的一行。不同浮标或同一浮标的不同剖面,其观测数值矩阵的行数都可能不同。

通过观察和分析,剖面观测数据属于半结构化数据,不适宜也无法存储于关系型数据库中。为避免在使用过程中对同一文件的频繁访问而导致的大量重复性工作,需提前对剖面观测数据进行提取和组织。半结构化数据的组织形式以 JSON 和 XML 两种最为流行,均能够被多种编程语言及可视化平台支持。JSON 是一种轻量级的数据交换格式。基于 JSON 的地理信息数据交换方法,相较于传统的 XML 标记语言具有更为精简的结构<sup>[16-17]</sup>,并且能够同时被多种关系型数据库兼容和利用,因此本文选用 JSON 作为剖面观测数据的存储格式。

自 2015 年 12 月起,Argo 剖面资料格式版本升级为 2.0,增加了叶绿素观测数据及多深度轴的剖面数据,能够同时观测高分辨率剖面和低分辨率剖面,有些浮标还有第三条剖面。为了适应 Argo 浮标剖面数据观测要素、深度轴的灵活变化,设计了一种“总剖面数据-多剖面-多观测要素”的树状 JSON 结构(图 2)用于组织提取出的剖面观测数据,并增加统计信息,如最大值(maxValue)、最小值(minValue)。其中 type 为该剖面的要素组合类型;profile<sub>i</sub> 为第 i 个观测剖面;element<sub>i</sub> 为第 i 个观测要素的观测值数组;data<sub>i,j</sub> 为第 i 个观测要素第 j 层深度的观测数值。

## 2.3 数据库设计

Argo 数据库应至少具有两张表:Argo 元数据表(Argometadata)和 Argo 剖面数据表(ArgoProfile)。其中 Argo 元数据表存储预提取的 Argo 浮标元数据,Argo 剖面数据表存储浮标剖面元数据及其剖面观测信息数据的资源链接。良好的表结构及表间关

系不仅能够增强数据库本身的健壮性和稳定性,也能够更好地支持数据的获取与更新操作。本文针对 Argo 数据本身的属性特征,主要从表拆分-建立表间关系、分表存储和构建索引三个方面进行优化。

(1)由于 Argo 元数据表中属性字段较多,根据数

据库三范式将所属项目相关属性抽出,新建项目表。由于所属国家经常被用于查询,因此将国家作为冗余字段直接存储于元数据表中,减轻 Join 操作带来的时间消耗。最后对分离后的多张表设置合理的主键及表间外键关联(图 3)。

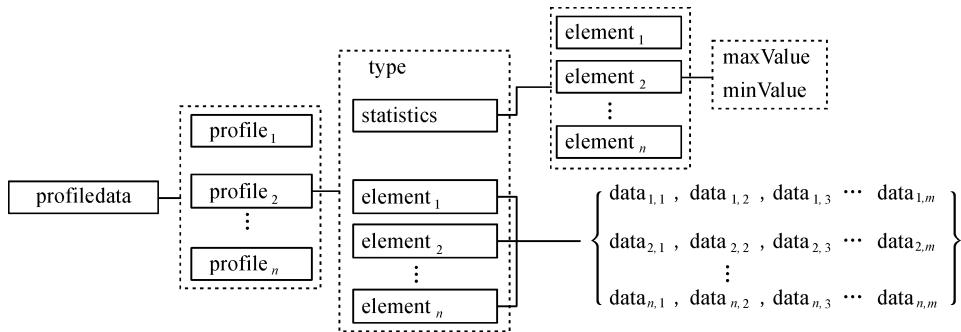


图 2 2.0 版本 Argo 剖面数据 JSON 格式结构

Fig. 2 JSON data structure of Argo profile (V2.0)

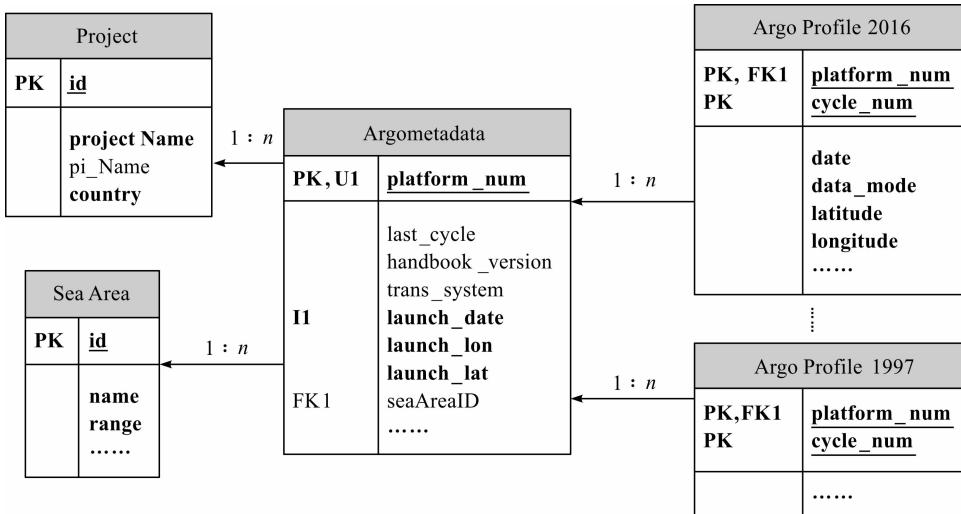


图 3 Argo 数据库模型图

Fig. 3 Diagram of Argo database model

(2)由于 Argo 剖面数目逐年增多,且每年增长量在十万数量级,因此对 Argo 剖面数据采取分表存储的策略。如数据库收录了自 1997 年至今的全球 Argo 剖面资料,需建立 1997—2016 共 20 张表,具有相同的表结构(图 3)。该策略后被证实能够适用于多种对象持久化技术及多线程查询。

(3)数据库的索引好比一本书的目录,适当地建立索引能够加快数据库的查询速度。Argo 数据表字段较多,对哪些字段建立索引,应该选取哪种类型的索引均会影响索引的使用效果。以浮标元数据表为例,本文根据对查询需求的调研和统计,确定浮标编

号、投放国家、所属项目名称、浮标类型、通讯方式、投放时间和海域范围 7 个字段,根据数据库索引最左前缀原则,按照搜索频率对索引字段进行排序后,建立聚簇索引。该索引在根据浮标编号进行查找、多组合查询、时间范围查询等实例中均提高了效率。

### 3 实验结果分析

本文选用了 PostgreSQL 关系型数据库,并利用其 9.4 版本对 JSON 类型字段的扩展支持管理剖面观测数据的 JSON 结构数据,利用 C-ARDC 提供的近 20 a 全球 Argo 资料及其说明文件,使用 Java 编程语言

编写数据提取与批量入库程序,建立了 Argo 数据库,并从可用性与可扩展性两方面对实验结果进行分析。

### 3.1 可用性分析

使用 2.2 节提出的基于 JSON 的剖面数据格式(图 2)对所收集的剖面资料进行剖面观测信息提取,其中部分浮标在上升和下沉过程中进行了 2 次采样,

由于传感器装在浮标的顶部,下沉过程的采样会受浮标尾流的影响而产生误差,因此需剔除下沉采样剖面。提取成功 142 万余条剖面,失败 0 条。统计结果(表 1)表明,本文所提出的剖面模型能够支持目前全球 Argo 剖面浮标资料集<sup>[15]</sup>中多深度轴的、多种观测要素的观测剖面信息提取。

表 1 1997—2016 年 Argo 剖面资料类型及数量统计

Tab. 1 Type and quantity statistics of Argo profile data from 1997 to 2016

深度轴数	观测要素			剖面数	总剖面数
	第一深度轴	第二深度轴	第三深度轴		
1	T			17 983	
	S,T			1 332 255	1 400 683
	S,T,O			50 445	
2	S,T	S,T,O		13 692	
	S,T	O		5	
	S,T	C		1 700	21 565
	S,T	S,T,O,C		6 168	
3	S,T	O	C	2 934	2 934

注:T 为温度,S 为盐度,O 为溶解氧,C 为叶绿素

基于所提取的剖面观测数据,于中国 Argo 实时资料中心机房服务器建立全球 Argo 数据库,利用 FSH(Flex+Spring+Hibernate)框架搭建“Argo 资料共享服务平台”。结果表明本文所提出的协同管理方法,能够支持属性查询、浮标漂移轨迹及剖面图的实时绘制(图 4~图 6)。该平台可通过以下链接访问:<http://101.71.255.4:8090/flexArgo/out/index.html>。

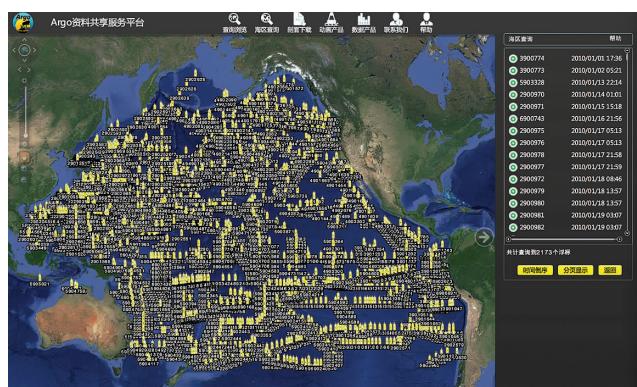


图 4 太平洋海区 Argo 浮标分布

Fig. 4 Distribution of Argo floats in the Pacific Ocean

### 3.2 可扩展性分析

#### 3.2.1 存储可扩展性分析

已入库数据量统计结果(图 7)表明:从 2008 年开



始,Argo 计划每年产生的剖面数已经超过 10 万条。随着 Argo 全球海洋实时观测网趋于稳定,剖面数年增长率降低,年剖面数趋于稳定。如将所有剖面元数据存储于同一张数据库表中,单表数据量将达到百万级,且逐年增长(图 8)。本文所提出的分表存储方法,能够将单表存储压力由百万级降低至十万级,数据库中各年份表间数据量分布趋于均衡,具备较好的可扩展性,满足 Argo 剖面数据的长期存储需求。

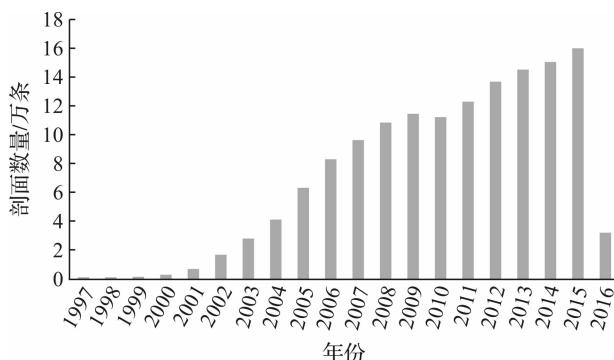


图 7 1997—2016 年间入库剖面数据量时间序列分布  
Fig. 7 Annual amount of profile data imported into Argo database from 1997 to 2016

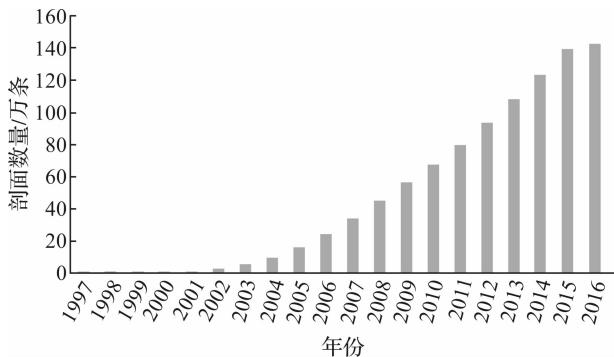


图 8 1997—2016 年间入库剖面总量逐年统计结果  
Fig. 8 Total amount of profile data imported into Argo database each year from 1997 to 2016

### 3.2.2 检索可扩展性分析

本文利用 2004—2015 年间剖面数据,建立分表存储与单表存储数据库,分析其轨迹数据获取效率随时间的变化关系。

选取 1900343 号浮标,针对不同年份区间的数据进行轨迹数据检索,查询语句均为“select \* from tablename where platform\_number = '1900343'”。平均时间消耗如图 9 所示,其中横坐标查询年数  $n$  的含义如下:

分表:取  $n$  个表(profile2004、profile2004+1……profile2004+n-1)分别执行查询语句,语句间使用 union 进行连接。

单表:取  $n$  个表(profile2004、profile2004+1……profile2004+n-1)中数据总和,建立新表 profileall\_n,并对 profileall\_n 表执行查询语句。

测试结果表明:分表与单表的查询时间消耗均与查询年数基本呈线性正相关;相同查询年数,分表较单表查询效率低(图 9)。

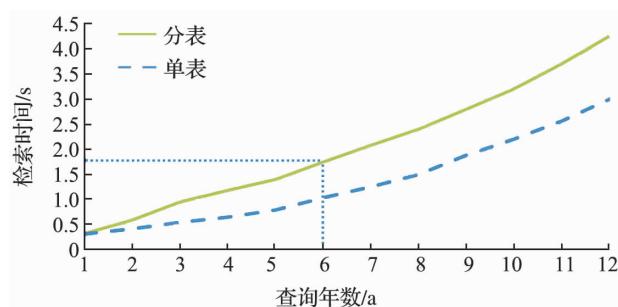


图 9 浮标轨迹检索时间消耗曲线  
Fig. 9 Elapsed times used for drift trajectory query

但 Argo 浮标的寿命受其所用电池影响,约为 4 a<sup>[18]</sup>,图 10 为已入库的浮标寿命统计结果(除活跃浮标与无剖面浮标,样本浮标总数为 7 253 个),可见 96% 的浮标使用寿命在 6 a 以内。因此在实际应用中,可以根据浮标的投放日期及最后一个观测剖面生成日期,确定该浮标剖面的年份区间,该区间通常小于或等于 6。

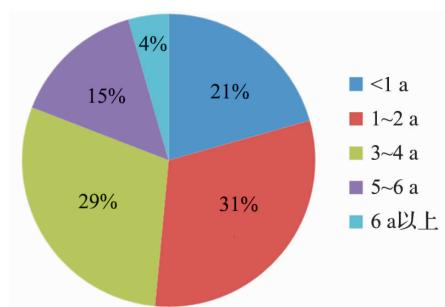


图 10 Argo 浮标寿命统计  
Fig. 10 Statistics of lifetime of Argo floats

利用按年份分表存储的方法,可以仅检索浮标存活年限内的数据表,进而将大部分浮标的检索区间控制在 6 a 以内。

在图 9 所示实验中,查询年数为 6 a 时分表存储检索时间消耗约 1.8 s,由于各年份表内数据量趋于

稳定,即连续 6 a 内数据量趋于稳定,可以推断出在连续 6 a 内进行轨迹数据检索时间,即检索寿命在 6 a 以内的浮标的轨迹数据时间消耗稳定在秒级。而单表存储需对 12 a 的数据进行检索,所需时间为 3 s,且随着时间的增长,该检索时间消耗同步增长。较单表存储,分表存储的方法在浮标轨迹数据检索方面,具有更好的可扩展性。

而对于其余 4% 存储年份较长的浮标,不能有效地减少其检索区间,导致检索效率较单表存储并无明显优势。

#### 4 结语

针对 Argo 资料结构化和半结构化数据共存的特性,本文设计并实现了一种协同管理方法,适用于已有的全球 Argo 剖面资料集,并已在中国 Argo 实时资料中心建立了共享服务平台,提供了一体化数据检索及浮标漂移轨迹和剖面图的实时绘制功能。同时该方法具备较好的可扩展性,能够符合长久的存储与检索需求。

由于结构化和半结构化并存的特征普遍存在于各种剖面观测数据中,如钻孔数据<sup>[19]</sup>、走航数据<sup>[20]</sup>、三维地震剖面数据<sup>[21]</sup>等,该方法也可为上述剖面数据管理提供技术参考。

#### 参考文献(References):

- [1] XU Jian-ping. Argo global ocean observation system [M]. Beijing: China Ocean Press, 2002.  
许建平. 阿尔戈全球海洋观测网大探秘[M]. 北京: 海洋出版社, 2002.
- [2] STEPHEN C R, HOWARD J F. Fifteen years of ocean observations with the global Argo array[J]. Nature Climate Change, 2016, **6**(2): 145-153.
- [3] ZHANG Chun-ling, XU Jian-ping. T/S distribution and variation in the Pacific based on Argo observations Part I: Temperature [J]. Marine Science Bulletin, 2014, **33**(6): 647-658.  
张春玲,许建平. 基于 Argo 观测的太平洋温、盐度分布与变化(I): 温度[J]. 海洋通报, 2014, **33**(6): 647-658.
- [4] ZHANG Chun-ling, XU Jian-ping. T/S distribution and variation in the Pacific based on Argo observations Part II: Salinity[J]. Marine Science Bulletin, 2015, **34**(1): 21-31.  
张春玲,许建平. 基于 Argo 观测的太平洋温、盐度分布与变化(II): 盐度[J]. 海洋通报, 2015, **34**(1): 21-31.
- [5] YANG Sheng-long, JIN Shao-fei, HUA Cheng-jun, et al. Spatial-temporal distribution of bigeye tuna *Thunnus obesus* in the tropical Atlantic Ocean based on Argo data[J]. Chinese Journal of Applied Ecology, 2015, **26**(2): 601-608.
- 杨胜龙,靳少非,化成君,等. 基于 Argo 数据的热带大西洋大眼金枪鱼时空分布[J]. 应用生态学报, 2015, **26**(2): 601-608.
- [6] YUAN D L, ZHANG Z C, CHU C P, et al. Geostrophic circulation in the tropical North Pacific Ocean based on Argo profiles[J]. Journal of Physical Oceanography, 2014, **44**(2): 558-575.
- [7] XU L X, LI P L, XIE S P, et al. Observing mesoscale eddy effects on mode-water subduction and transport in the North Pacific[J]. Nature Communications, 2016(7): 1-9.
- [8] WU L X, JING Z, RISER S, et al. Seasonal and spatial variations of Southern Ocean diapycnal mixing from Argo profiling floats[J]. Nature Geoscience, 2012, **4**(6): 363-366.
- [9] WANG Shuai, XU Cong-fu, CHEN Ya-fang. WebGIS based Argo data-sharing service system[J]. Marine Sciences, 2011, **35**(3): 32-36.  
王帅,徐从富,陈雅芳. 基于 Web GIS 的 Argo 数据共享服务系统[J]. 海洋科学, 2011, **35**(3): 32-36.
- [10] WANG Xian-ling, QIN Bo, LIU Pei-shun. Argo data sharing system based on grid technology[J]. Computer Engineering and Design, 2009, **30**(15): 3 634-3 637.  
王显玲,秦勃,刘培顺. 基于网格技术的 Argo 数据共享系统[J]. 计算机工程与设计, 2009, **30**(15): 3 634-3 637.
- [11] YANG Feng, ZHOU Jian-zheng, DU Yun-yan, et al. Multi-source Argo data sharing and visualization based on Web Service [J]. Science of Surveying and Mapping, 2011, **36**(5): 62-64.  
杨峰,周建郑,杜云艳,等. 基于 Web 服务的多源 Argo 数据共享与可视化应用研究[J]. 测绘科学, 2011, **36**(5): 62-64.
- [12] Oracle XML DB Developer's Guide[DB/OL]. [2016-12-16]. [http://docs.oracle.com/cd/B19306\\_01/appdev.102/b14259/toc.htm](http://docs.oracle.com/cd/B19306_01/appdev.102/b14259/toc.htm).
- [13] The PostgreSQL Global Development Group. PostgreSQL 9.4. 10 Documentation[DB/OL]. [2016-12-16]. <https://www.postgresql.org/docs/9.4/static/index.html>.
- [14] LIU Zeng-hong, WU Xiao-fen, XU Jian-ping, et al. Fifteen years of ocean observations with China Argo[J]. Advances in Earth Science, 2016, **31**(5): 445-460.  
刘增宏,吴晓芬,许建平,等. 中国 Argo 海洋观测十五年[J]. 地球科学进展, 2016, **31**(5): 445-460.
- [15] Description about the global Argo profile data set(V2.0) [EB/OL]. [2016-12-16]. [www.argo.org.cn](http://www.argo.org.cn).  
全球 Argo 剖面浮标资料集(V2.0)说明[EB/OL]. [2016-12-16]. [www.argo.org.cn](http://www.argo.org.cn).
- [16] HAN Min, FENG Hao. Research of GIS data exchange method based on JSON[J]. Science of Surveying and Mapping, 2010, **35**(1): 159-161.  
韩敏,冯浩. 基于 JSON 的地理信息数据交换方法研究[J]. 测绘科学, 2010, **35**(1): 159-161.
- [17] ZHANG Hu-yin, QU Qian-song, HU Rui-yun. Data exchange model based on JSON[J]. Computer Engineering and Design, 2015, **26**(12): 3 380-3 384.

- 张沪寅,届乾松,胡瑞芸.基于 JSON 的数据交换模型[J].计算机工程与设计,2015,26(12):3 380-3 384.
- [18] FEDER T. Argo begins systematic global probing of the upper oceans [J]. Physics Today,2000,53(7):50-51.
- [19] WANG Ji-min, LÜ Qing, WAN Ding-sheng. Geological section modeling system based on borehole data[J]. Journal of Hohai University Natural Sciences,2009,37(4):463-466.  
王继民,吕庆,万定生.基于钻孔数据的地质剖面建模系统[J].河南大学学报:自然科学版,2009,37(4):463-466.
- [20] WU Qing-song, WANG Qi, YING Jian-yun, et al. Real-time vi-
- sualization of the DGPS data based on Matlab and its application in ship-mounted observation[J]. Journal of Marine Sciences, 2016,34(2):60-64.  
吴清松,王琪,应剑云,等.基于 Matlab 的 DGPS 数据实时可视化与走航观测[J].海洋学研究,2016,34(2):60-64.
- [21] WEI Jin-lan, ZHENG Hong-ming, XU Qun-zhou. Data management and application system for vertical seismic profiling[J]. Journal of Computer Applications,2011,31(2):239-241.  
魏金兰,郑鸿明,徐群洲.垂直地震剖面数据管理和应用系统[J].计算机应用,2011,31(2):239-241.

## Research on collaborative management method of Argo data

DONG Gui-ying<sup>1,2</sup>, CAO Min-jie<sup>3,4</sup>, ZHANG Feng<sup>1,2</sup>, DU Zhen-hong<sup>\*1,2</sup>, LIU Ren-yi<sup>1,2</sup>, WU Sen-sen<sup>1,2</sup>  
(1. Zhejiang Provincial Key Lab of GIS, Zhejiang University, Hangzhou 310028, China; 2. Institute of GIS, Zhejiang University, Hangzhou 310028, China; 3. State Key Laboratory of Satellite Ocean Environment Dynamics, Hangzhou 310012, China; 4. Second Institute of Oceanography, SOA, Hangzhou 310012, China)

**Abstract:** Argo data have become an important data source in basic researches of ocean and atmosphere sciences. But with the hybrid storage of profile metadata and its observations, it is difficult for existing shared platforms to realize the real-time visualizing of the drift trajectory and profile chart. A method of collaborative management of structured and semi-structured Argo data were proposed. Based on the analysis of Argo data structure and characteristics, Argo data were separated as structured and semi-structured data. By means of extended support of relational database to a semi-structured property, a relationship between profile metadata and observation data was established. Faced of the rapid growth of data, a multi-table storage method was used to reduce storage pressure of single data table. Finally, an Argo database was set up and global Argo data in period of 1997-2016 are parsed and imported. The results indicate that the method has good scalability of both storage and query, as well as high efficiency of float trajectory query, and it is able to support real-time visualization of float trajectory and profile chart. Our result will offer a reference for management of other similar observation profile data.

**Key words:** Argo data; profile; structured and semi-structured; collaborative management