

# 气象大数据资源汇交可视化分析研究

赵卓宁<sup>1, 3</sup> 李湘<sup>2, 3</sup> 舒红平<sup>1, 3</sup> 何文春<sup>2</sup> 孙超<sup>2</sup>

(1 成都信息工程大学, 成都 610225; 2 国家气象信息中心, 北京 100081;

3 中国气象局-成都信息工程大学气象软件工程联合实验室, 北京 100081)

**摘要:** 针对气象大数据资源汇交多维可视化分析的复杂性问题, 结合气象大数据汇交业务特点, 构建了气象大数据资源汇交的业务对象模型、可视化分析维度、可视化监控指标、可视化设计框架和典型分析场景, 为气象大数据资源汇交分析提供了有关流量、质量和效率的汇交可视化分析方案, 以及面向过程、数据、服务等典型分析场景。

**关键词:** 气象大数据, 数据汇交, 可视化分析

**DOI:** 10.3969/j.issn.2095-1973.2018.01.028

## Research on Visualization Analysis of Meteorological Big Data Collection

Zhao Zhuoning<sup>1, 3</sup>, Li Xiang<sup>2, 3</sup>, Shu Hongping<sup>1, 3</sup>, He Wenchun<sup>2</sup>, Song Chao<sup>2</sup>

(1 Chengdu University of Information Technology, Chengdu 610225 2 National Meteorological Information Centre, Beijing 100081 3 CMA-CUIT Meteorological Software Engineering Joint Research Centre, Beijing 100081)

**Abstract:** In this paper, we constructed the business object model, visual analysis dimension, visualization monitoring indicators, visual design framework and typical analysis scenarios of meteorological big data resources, according to the complexity of multidimensional visual analysis of meteorological big data resources and combined with the characteristics of meteorological big data submissions. It provides the visualization solution of traffic, quality and efficiency for the data exchange in meteorological big data, as well as the typical analysis scenarios such as process, data and service.

**Keywords:** meteorological big data, data collection, visualization analysis

### 0 引言

气象大数据资源汇交是气象主题相关的数据采集、吸纳、存储的过程, 是气象大数据产品和服务输出的前序环节, 目的是确保气象相关数据得到长期保藏、有序共享和广泛应用。数据汇交过程中采集、加工、存储和共享等业务作业环节, 将改变资料结构、资产数量和资源品质, 推进数据服务共享、资源价值再生和领域知识涌现等。随着大数据技术在各行业深入应用, 推动了行业信息资源供给的流动性, 跨行业、跨部门的信息资源汇交业务日益增多, 数据资产供给交易、数据服务价值评估、数据产品协同生产等业务级评估需求日益旺盛。统筹管控数据汇交业务, 依赖于对作业各环节业务处理状态和资源品质的跟踪

和评估。传统海量、实时汇交业务多数基于专用业务监控功能, 面向大数据汇交则需拓展适用于跨行业、跨部门数据汇交的监控功能, 并可视化分析业务处理过程及结果<sup>[1]</sup>, 为持续改进的数据汇交过程提供信息反馈, 这对完善气象大数据平台建设和应用开发具有重要的支撑作用。

### 1 气象大数据信息资源及其汇交

#### 1.1 气象信息资源与气象大数据资源

气象大数据资源是以气象为主题的广义气象信息资源, 具有跨行业、跨领域、跨部门、多学科交叉的主题特征, 包涵气象相关的科学、工程、技术、业务、服务、组织、人员等领域的信息资源。以气象科学专业知识服务系统 (<http://k.data.cma.cn/>) 为例, 截止2017年3月16日, 已累积数据条目140758条, 涉及气象基础数据、气象数据产品、专业文献、院士专家、科研机构、科研报告、标准规范、气象专利、气象百科、新闻资讯等10大类数据。

根据气象行业标准《气象资料分类与编码》(QX/T 102-2009) 气象信息资源有14大类, 即地面、

收稿日期: 2017年9月11日; 修回日期: 2018年2月22日  
第一作者: 赵卓宁(1972—), Email: zhaozn@cuit.edu.cn  
资助信息: 四川省科技厅应用基础项目(2017JY0196); 中国气象局-成都信息工程大学气象软件工程联合实验室开放课题(201703); 信息共享与数据挖掘四川省高校重点实验室开放课题(201609)

高空、辐射、海洋、农业气象和生态气象、大气成分、卫星、雷达、气象灾害、数值预报产品、历史气候代用、科学试验和考察、气象服务产品和其他资料。国内气象资料主要通过全国地面宽带网、大院骨干网等国内通信系统收集。其中，地面气象宽带网收集省级上行资料，大院骨干网收集国家级业务单位资料。这些气象资料汇交所形成的气象资料和气象服务产品，通过CMACast卫星广播、实时数据推送、数据库接口访问（API）、国家级气象业务内网（http://idata.cma）、中国气象数据网（http://data.cma.cn）等五种方式，向国家级业务单位、省级业务单位以及其他企事业单位用户和社会公众用户提供服务，日均数据服务量已达TB级。

## 1.2 气象大数据资源汇交的复杂性

气象大数据资源包括气象行业内的气象信息资源和来自网络的其他行业数据资源。气象大数据资源汇交是气象大数据处理的一部分，由加工、元数据和存储三部分组成（图1）。实线表示气象业务数据流，将来自气象观测的数据经采集、传输、加工、存储和服务形成气象产品的过程。虚线表示其他行业数据流，将行业外部数据源经通信网络接收或前置机采集后，加载到数据仓库或各类存储中，再经面向主题的场景加工，形成气象大数据产品的过程。

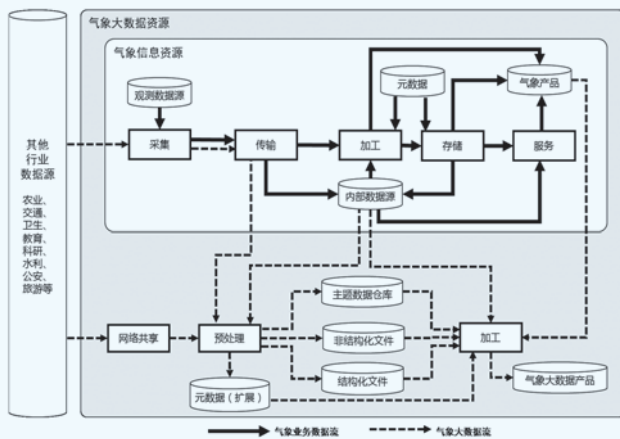


图1 气象大数据处理过程

Fig. 1 Meteorological big data collection process

气象大数据信息资源数据采集范围和服务对象具有跨行业跨部门、跨领域的扩展性，导致气象大数据资源汇交具有过程复杂性：1) 接入的数据来源差异大，预处理工作量复杂，数据接入的方式多样，包括前置机、网页爬取、数据接口输入、通信系统接入、应用系统输入、手工输入等；2) 气象元数据在数据汇交中的分类标注作用有限，需要拓展和引入涉及其他领域的元数据定义；3) 大数据汇交包括从外部进入内部的

汇入，以及从内部向外部的汇出，其中汇入数据形式零散、类型繁杂，汇出数据主题明确但维度和关联复杂。

## 1.3 气象大数据汇交的对象模型

气象大数据汇交就是将“汇交资源”变成“气象大数据资料”的过程。气象大数据汇交的对象模型（图2）是对这一过程中所涉及的客观实体关系进行的静态描述，是可视化分析的模型基础<sup>[2-3]</sup>。从汇交业务视角看，气象大数据汇交的对象模型由汇交资源和气象大数据汇交两个业务领域组成。

### 1.3.1 汇交资源

汇交资源描述了以“汇交资料”对象为中心的对象关系图谱。描述了“数据生产者”从其“业务系统”中提供多个“数据包”的汇交资源结构。一个“数据生产者”拥有多个“业务系统”，每个业务系统提供多个“数据源”，每个“数据源”包含“设备”、“数据员”信息，每个“数据源”可提供多个“数据包”。一个“数据包”由多个“汇交资料”组成，“数据包”是进行数据汇交时发送和接收数据的基本单位。一个“汇交凭证”中包含了多个“数据包”，即一次数据汇交中可以涉及多个“数据包”。

“汇交资料”是一种“气象资料”，也是一种“行业资料”，都可以统称为“资料”，具备“资料”的所有属性。“气象资料”的类型由“气象资料类型”定义描述，“行业资料”由“行业资料类型”定义描述。“气象服务”、“模式数据”都是一种“产品”，而“产品”、“国外资料”都可用“汇交资料”所具备的属性来描述。一个“汇交资料”由多个“要素”组成，这些“要素”继承了“常规气象要素”的特征，也可能继承了“公路交通要素”、“轨道交通”、“闪电定位”等要素的特征，也可以按“要素模板”拓展自定义特征。所有这些“要素”特征都可以统一用“资料要素”这个领域描述。

### 1.3.2 气象大数据汇交

气象大数据汇交描述了“汇交资料”经汇交处理后存储为“气象大数据资料”的数据组织结构与关系。

一个“汇交项目”包含多次“汇交任务”，一次“汇交任务”对应多条“汇交记录”，一条“汇交记录”的详细情况由多条“汇交日志”记录描述。一个“汇交工具”执行多次汇交就对应多条“汇交记录”，一次汇交可能由多个“汇交工具”协同执行，因此，“汇交工具”和“汇交记录”是多对多的关系，这种关系可由“汇交日志”详细记录。

一种“气象大数据资料”是“气象大数据目录”中的一种“元数据”。其中“元数据”就是依据气象

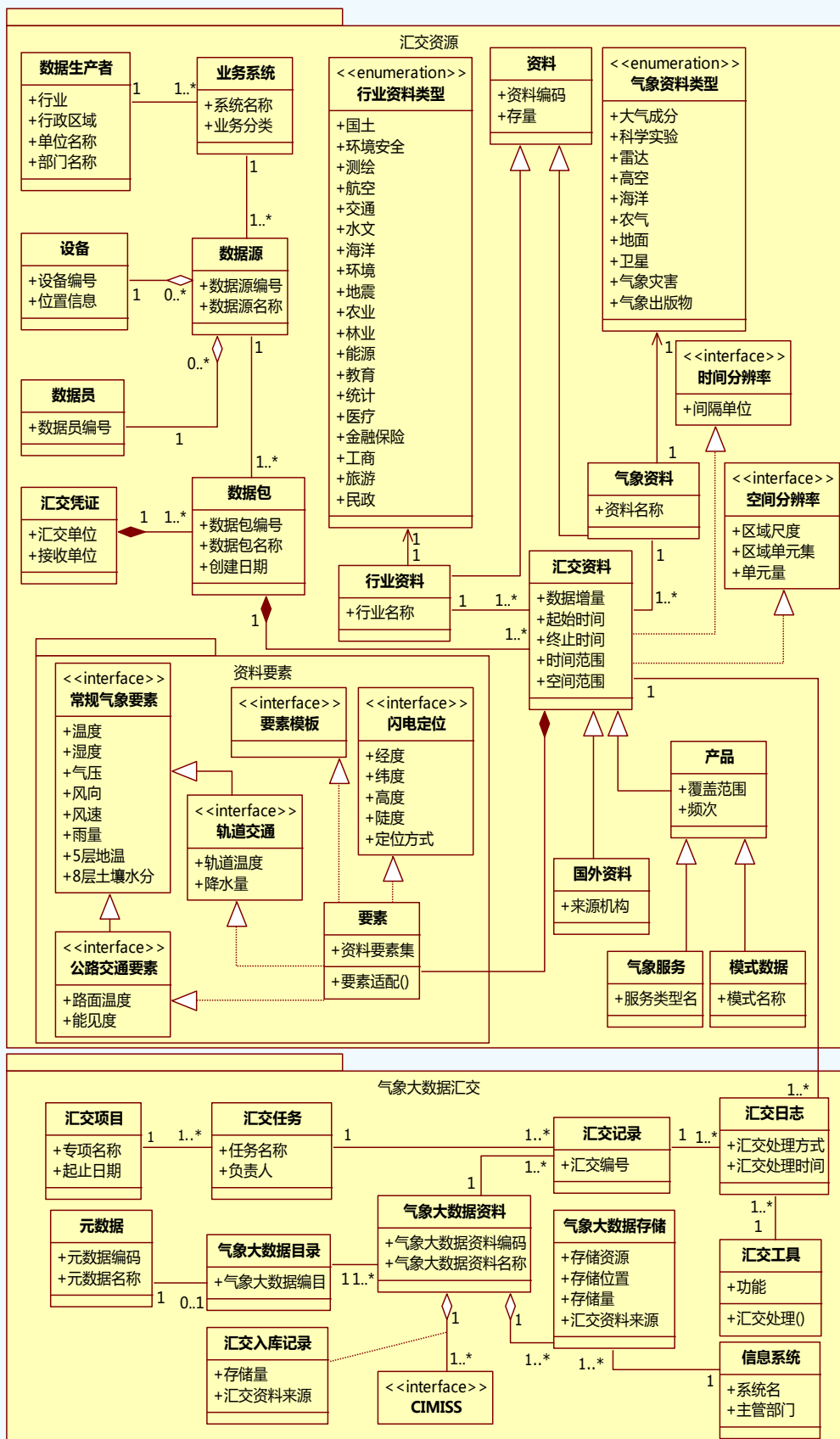


图2 气象大数据汇交的对象模型

Fig. 2 The object model of meteorological big data collection

行业标准《气象资料分类与编码》(QX/T 102-2009)定义的14大类气象信息资源的分类与编码信息。

一个“气象大数据资料”的数据存放可以通过CIMISS接口, 入库到CIMISS系统中, 并记录“汇交入库”。也可以单独在“气象大数据存储”中存储, 并标记存储所属的“信息系统”。

“汇交日志”是“汇交任务”、“汇交工具”、“汇交资料”三个对象经过汇交业务处理后, 建立的可追溯的关系对象, 记录了根据“汇交任务”复用“汇交工具”对“汇交资料”的处理过程。

## 2 汇交可视化分析模型

### 2.1 汇交可视化分析维度

气象大数据资源汇交中各数据包从数据源采集, 被分解为汇交资料后, 经检查、规范化后吸纳, 按气象大数据目录重新编码后存储为气象大数据资源。汇交分析关注数据包在数据源、汇交资料、气象大数据

资料这三种资料状态时的结构、质量, 以及全程处理过程中的质量和效率。因此, 多维分析是汇交可视化分析的特点, 分析维度主要包括基本维度、数据源维度、汇交资料维度、气象大数据维度, 汇交任务维度<sup>[4]</sup>。

基本维度包含时间、空间, 数据源维度包含行业、数据生产者、业务系统、设备、数据员、数据包。汇交资料维度包含气象资料、行业资料、国外资料、产品、服务、要素。气象大数据维度包含元数据、气象大数据目录、存储、信息系统。汇交任务维度包含汇交阶段、汇交任务、汇交项目、汇交工具。

### 2.2 汇交可视化监控指标

汇交可视化监控主要面向数据汇交采集、吸纳和存储这三个业务环节上, 在汇交业务处理发生的事前、事中和事后的全过程, 有关流量、质量和效率的监测指标, 具体如表1所示。

表1 气象大数据汇交可视化分析主要指标  
Table 1 The main indexes of meteorological big data collection visualization analysis

汇交业务	涉及业务对象	流量指标	质量指标	效率指标
采集	事前 数据生产者, 业务系统, 数据源, 设备, 数据员, 数据包	预估量	病毒检出率	逾期时间
	事中 数据包, 汇交资料, 要素	在线处理量	空报率, 漏报率	传输速率
	事后 汇交凭证, 行业资料, 气象资料, 国外资料, 产品	完成量	数据包异常率	单位时间处理量
吸纳	事前 元数据, 气象大数据目录, 汇交工具	预估量	资料对立度	数据包实时解包时间
	事中 汇交任务, 气象大数据资料, 汇交工具	在线处理量	资料差异度	质量实时审核时间
	事后 汇交日志, 汇交工具	完成量	资料同一度	单位时间处理量
存储	事前 汇交项目, CIMISS, 信息系统	预估存储量	可用接口数	请求时间
	事中 汇交入库记录	在线处理量	可用存储量	实时I/O时间
	事后 气象大数据存储	完成量, CIMISS入库量, 链接入库量	有效存储率	单位时间处理量

#### 2.2.1 流量指标

流量指标主要用于监测汇交处理过程的进度。流量指标是有关采集数据包、吸纳汇交资料和存储气象大数据资源这三类汇交业务处理过程中, 对应处理的事前、事中和事后的量化指标有数据包数据的预估量、在线处理量和完成量。

#### 2.2.2 质量指标

质量指标主要用于监测汇交处理过程中有关数据安全、数据业务质量、数据异常、资料规范化、存储有效性等的性能。

采集阶段主要对上报的数据包进行数据安全、业务质量和数据异常的监测。以病毒检出率指标反馈数据安全, 以空报率、漏报率指标反馈数据包业务质量, 以数据包异常率反应上报数据传输情况。

吸纳阶段将数据包解包为汇交资料的过程中, 需要按气象大数据的元数据进行分类提取, 并按气象大数据目录生成汇交任务。采用集对理论<sup>[5]</sup>构建评估数

据包中的数据规范性的指标, 设元数据、气象大数据目录的编码规范为数据标准集合A, 数据包中已存在的各类数据编码规范视为集合B, 则集合A和B组成集对。设某数据包共有N种数据, 其中符合集合A标准的数据类型有S种, 不在集合A标准中的数据类型有P种, 其余的 $F=N-S-P$ 个数据类型既不相对立, 又不为集合A和B共同具有。则 $S/N$ 为资料同一度, 用于表示符合集合A标准的可汇交资料, 资料同一度反应数据包按元数据解包和按气象大数据目录生成汇交任务的比例;  $P/N$ 为资料对立度, 用于表示不能使用集合A标准进行转化的数据包, 资料对立度指标有利于区分新类型数据资料。  $F/N$ 为资料差异度, 指部分符合集合A标准的数据包, 资料差异度指标有利于识别可再细分数据的数据包。

存储阶段将执行吸纳阶段所生成的汇交任务, 这些任务将按项目提交给不同的汇交工具进行处理。对每个汇交工具存储过程进行监测, 主要包括存储前可

用接口数、存储中可用存储量、存储结束后的有效存储率。

### 2.2.3 效率指标

效率指标关注汇交业务的关键环节的执行情况。采集阶段关注事前的延迟上报数据的逾期时间，事中数据包的传输速率以及进度，事后单位时间内采集业务完成的数据量。吸纳阶段关注事前数据包解包为汇交资料的实时解包时间，事中形成资料同一度评价所需要的质量实时审核时间，事后单位时间内汇交资料的处理量。存储阶段关注事前向存储系统的请求时间，以及事中存储过程中的实时读写时间，事后存储处理的单位时间处理量。

除了以上三类汇交可视化监控指标外，也可以按不同维度组合对相关业务对象进行计数、平均等汇总处理，以形成对汇交业务处理对象的总体量化指标，与各分析维度一起组成气象大数据汇交多维分析数据包（表2），可用于支持气象大数据汇交业务的数据仓库构建。

表2 气象大数据汇交多维分析数据包

Table 2 Multidimensional analysis data packets for meteorological big data collection visualization analysis

维度	基本维度	数据源维度	汇交资料维度	气象大数据维度	汇交任务维度
层次	时间	行业	气象资料	元数据	汇交阶段
	空间	数据生产者	行业资料	气象大数据目录	汇交任务
指标		业务系统	国外资料	存储	汇交项目
		设备	产品	信息系统	汇交工具
		数据员	服务		
		数据包	要素		
指标	流量指标（预估量、在线处理量、完成量、预估存储量、CIMISS入库量、链接入库量）；质量指标（病毒检出率、空报率、漏报率、数据包异常率、资料对立度、资料差异度、资料同一度、可用接口数、可用存储量、有效存储率）；效率指标（逾期时间、传输速率、单位时间处理量、数据包实时解包时间、质量实时审核时间、单位时间处理量、请求时间、实时I/O时间、单位时间处理量）；统计指标（计数、平均、最大值、最小值）				

## 3 气象大数据汇交可视化分析

气象大数据汇交分析数据是在采集、吸纳和存储过程中，由大数据汇交任务调度程序审核、记录和汇总而形成的，支持面向过程、面向数据、面向服务、面向行业、面向部门等主题分析，通过多种视图组合，以面向业务对象的方式，借助人眼的感知能力和人脑智能对数据进行交互的可视表达，支持分析任务驱动的时序分析、关联分析或聚类分析。

### 3.1 可视化分析的设计框架

气象大数据汇交可视化分析设计框架（图3）自顶向下分为应用设计层、业务设计层、服务设计层和汇交设计层，支持从汇交业务流程到汇交可视化分析的处理流程和加工设计。

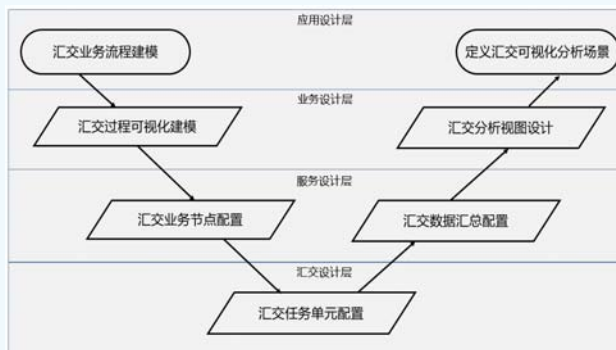


图3 气象大数据汇交可视化分析设计框架

Fig. 3 A design framework for visualization analysis of meteorological big data

应用设计层解决汇交业务过程在系统中的模型映射和业务分析关键场景识别的问题。支持业务主管人员根据现实业务及需求按采集、吸纳和存储三阶段建立业务流程模型，根据不同业务人员业务监控需求面向不同的主题分析定义汇交可视化分析场景。

业务设计层解决汇交过程可视化设计、汇交分析视图设计的环境和工具问题。支持业务技术人员以类似spss clementine12数据流建模的方式，根据业务流程来定义气象大数据包的汇交全生命周期过程，并在各业务过程的关键环节上，以类似Tableau 10的方式提供“对象—结构—质量”、“对象—结构—效率”、“对象—结构—流量”等样式，根据分析场景来定义多维分析视图。

服务设计层解决业务处理逻辑和业务分析逻辑范畴内的业务加工和数据汇总的任务单元组合问题。该层提供类似SOA服务装配的工具，支持业务技术人员完成业务单元封装、业务服务接口发布，支持业务分析人员按汇交指标定义业务数据汇总接口。

汇交设计层解决气象大数据包的采集、传输、加工、存储、审计等数据处理环节上工具调用、日志存储和数据提交等，该层提供了构件化设计工具，以支持对气象大数据资源汇交每一任务单元的具体实现。

### 3.2 典型的可视化分析场景

气象大数据资源汇交分析是典型的多维数据分析，需要强大的汇交数据汇总配置工具和汇交分析视图设计工具支持，这些工具的技术选型依赖于气象大数据资源汇交可视化分析场景的需求。气象大数据主题分析的内容决定了场景可视化的形式，典型的气象大数据资源汇交可视化分析场景有以下四类。

面向过程主题分析是在基本维度的基础上，以汇交任务维度的汇交阶段为主，可从数据源维度上考

（下转263页）

的资源冗余、重复建设现象，并未将气象业务人员彻底从系统运维工作中解放，仍需进行大量基础维护工作。

为了解决以上遗存问题，除已经应用的服务器虚拟化和存储虚拟化技术外，还可引入当前技术成熟度较高，且应用较为广泛的还有容器虚拟化及网络虚拟化技术。通过容器虚拟化，可以使不同应用系统复用操作系统资源，方便业务人员对于应用系统的调试和部署，极大减少部署业务的时间成本和人力成本。网络虚拟化的部署则可以按需对业务进行网络安全策略发放，对不同业务进行安全隔离。未来对于容器及网络虚拟化的应用，将降低业务开发、部署及网络运维方面的工作难度，释放相关业务人员运维压力。

通过服务器和存储虚拟化技术的应用，已在气象

业务应用中取得了巨大效益，解决众多业务对平台资源的需求问题，节省大量机房场地能源成本，实现部分传统平台不具备的功能。未来还将在系统建设中应用其他虚拟化及云计算相关技术，解决应用级使用问题，满足气象业务的发展诉求。

#### 深入阅读

沈文海, 2012. 从云计算看气象部门未来的信息化趋势. 气象科技进展, 2(2): 49-56.

王彬, 2014. 高性能计算技术在气象部门的应用. 计算机工程与设计, 35(4): 1476-1479.

中国信息通信研究院, 2016. 云计算白皮书(2016年).

(作者单位: 国家气象信息中心)

(上接207页)

核汇交资料或气象大数据在流程、质量或效率上的指标。分析主场景以桑基图展示各汇交阶段，基于时序的数据包中各成分在汇交过程中的转化情况。

面向数据的主题分析是在基本维度的基础上，以气象大数据维度的元数据为主，可考核各数据源维度上的汇交资料转化为气象大数据的流程、质量或效率。分析主场景以树图展示各类气象资料通过元数据规范化识别并归入气象大数据目录的吸纳情况。

跨行业、跨部门、跨系统的主题分析是在基本维度基础上，以数据源维度中行业、数据生产者、业务系统三种情况为主，可考核汇交资料维和气象大数据维度上同一度、差异度和对立度等指标，以解析相关行业、部门和系统的数据质量。分析主场景以旭日图展示各数据的细分溯源、数据具体构成以及存储情况。

面向服务的主题分析是在基本维度基础上，对汇交日志数据进行时序显著性、空间显著性、关联关系等深度分析，通过构建数据资源业务价值评估指标，进行相应的服务价值评估。分析主场景以时间序列、地图等为背景，采用点、线、面的不同标识来呈现其显著性和关联性。

## 4 展望

现有数据可视化技术为可视化分析提供了丰富的表现形式，但在气象大数据资源汇交这类过程复杂性的可视化分析领域中，主要挑战来自于业务对象关系、业务观测指标、数据描述维度、应用分析场景的复杂性，但却是决定可视化内容的关键。本文结合气象大数据资源汇交的业务过程，汇交资料与气象大数据间汇交的内在规范性联系，构建了气象大数据资源汇交的数据模型、监控指标、分析设计框架和分析场景，解决了气象大数据资源汇交可视化多维分析的一些基础性问题，但可视化分析效果还需结合可视化技术，在构建集成的、实时的、交互式的气象大数据资源汇交全景工具中改进。

#### 参考文献

- [1] 霍亮, 朝乐门. 可视化方法及其在信息分析中的应用. 情报理论与实践, 2017, 40(4): 111-116.
- [2] 任利敬, 赵正旭, 徐骞. 三维地图式可视化大数据. 测绘与空间地理信息, 2017, 40(2): 63-65.
- [3] 杨彦波, 刘滨, 祁明月. 信息可视化研究综述. 河北科技大学学报, 2014, 35(1): 91-102.
- [4] 洪文学, 王金甲. 可视化和可视化分析学. 燕山大学学报, 2010, 34(2): 95-99, 105.
- [5] 赵克勤, 集对分析及其初步应用[M]. 浙江科学技术出版社, 2000.