



Research  
Intelligent Manufacturing—Article

## Fog-IBDIS——基于雾计算的制造系统大数据集成方法

汪俊亮<sup>a</sup>, 郑鹏<sup>b</sup>, 吕佑龙<sup>a</sup>, 鲍劲松<sup>a</sup>, 张洁<sup>a,\*</sup>

<sup>a</sup> College of Mechanical Engineering, Donghua University, Shanghai 201620, China

<sup>b</sup> School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

### ARTICLE INFO

#### Article history:

Received 22 May 2018

Revised 31 August 2018

Accepted 5 December 2018

Available online 5 July 2019

#### 关键词

雾计算  
工业大数据  
多源数据  
数据集成

### 摘要

在工业领域,如何在多源工业数据的协作分析中保障源数据的私密性与安全性至关重要。本文提出了一种基于雾计算的工业大数据集成与共享方法(Fog-IBDIS),采用云端与边缘端协作的方式,实现工业数据的分布式本地处理,在多源数据的分析中保障源数据的私密性与安全性。首先,在云端设计了任务流程图,将多源数据分析过程分解成多个子任务;其次,设计了子任务管理、编译和运行控制、数据集成传输、基本算法库和管理组件五个模块,实现子任务的本地边缘端处理;最后,本文以大型客机制造过程为例,对Fog-IBDIS的运行过程进行了验证,其通过边缘与云端的协作方式,将多来源数据分析任务分解至本地执行,通过中间结果的传输串联实现工业大数据的分析,可保障原始数据的私密性与安全性。

© 2019 THE AUTHORS. Published by Elsevier LTD on behalf of Chinese Academy of Engineering and Higher Education Press Limited Company This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. 引言

随着互联网与工业融合创新、智能制造时代的到来,工业大数据技术及应用将成为未来提升制造业生产力、竞争力、创新能力的关键要素,是驱动产品设计智能化、生产过程智能化、管理智能化、服务智能化、新业态新模式智能化,支撑制造业转型和构建开放、共享、协作的智能制造产业生态的重要基础,对实施智能制造战略具有十分重要的推动作用[1,2]。从技术应用侧来看,在产品生产阶段,对系统感知得到的大数据进行分析可以提高生产效率[3,4]、系统柔性[5,6]与产品质量[7],具有重要的价值;在产品设计与营销阶段,大数据分析可用于量化客户需求,从而实现营销与设计的精

准决策,提高用户体验。从技术供给端来看,大量数据分析模型[8]、方法[9,10]和平台[11]逐渐成熟。随着技术供应与需求侧的成熟,大数据可能带来的巨大价值正在逐渐被传统产业认可,它通过数据的感知与分析,深入融合传统业务,驱动制造系统的运行优化逐渐向智能化转型[12,13]。丰田汽车基于6家门户网站、5家主流视频和4家汽车垂直媒体的用户行为大数据,通过对用户数据、广告推荐数据、用户观看数据三方面数据的分析,实现广告的精准投放。事实上,大数据的关联与碰撞,带来了产业价值链的全新视角,它将产品制造中的规划、设计、制造、销售、售后运维全流程贯通起来,进一步创造价值,为制造业转型升级带来新动力[14]。其中,工业大数据集成与共享(IBDIS)决定了大数据分

\* Corresponding author.

E-mail address: [mezhangjie@dhu.edu.cn](mailto:mezhangjie@dhu.edu.cn) (J. Zhang).

析的效率，正在成为制造系统大数据分析中的重要基础[15–16]。IBDIS通过定义数据交换模型实现数据提取、转换和加载，来实现多来源数据的分析[17]。然而，由于数据交换中海量数据带来的网络流量负载和源数据的私密性，制造系统中大数据集成与分享仍然充满挑战。

### 1.1. 海量数据交换中的网络流量负荷

制造系统中的大数据集成的首要难题是海量数据长距离传输导致的网络流量负载过重问题。在智能制造系统中，智能装备和物联网设备采集得到的数据量持续增长[18]，制造系统领域累积的数据量每年逾1000 EB，并且预计在未来几年内持续增长[19]。海量数据的处理需要极大的计算能力，大量企业建立云端数据中心，以支持制造系统运行的数据集成和存储[17,20]。但是，当前云计算框架需要将大量源数据传输到远程云数据中心进行进一步分析[21,22]，这会导致网络流量负载过重。工业大数据具有多来源特性，制造系统运行优化中所需的数据（包含订单要求、产品工艺路线、机器状态和计划调度数据等）来自产品数据管理（PDM）系统、制造执行系统（MES）、监督控制和数据采集（SCADA）系统等[23]。因此，IBDIS频繁调度多个源数据库以支持数据分析，极大增加了网络负载。目前，现有的商业云计算模型使用批处理模式，分批次向云端传输大数据，以减轻网络负载，但这依然难以满足工业领域实时优化的需求。

### 1.2. 源数据的私密性

与互联网数据不同，制造业大数据具有私密性的特点。在产业链内部，企业的设计、制造与销售数据一般都存储于企业内部的私有云中，属于商业秘密[24]。数据具备典型的无差异留存特性，交易过后可以低成本无差异地复制原始数据，数据也没有唯一的所有权约束，可以同时交易和传播给多个对象。现有的数据交易模式无论是数据中心的托管交易模式，还是聚合交易模式都难以从根本上解决工业数据的交易源数据私密性问题。

针对上述两个问题，本文提出了一种名为Fog-IBDIS的雾计算架构，它充分利用网络中边缘设备的计算能力来对原始工业大数据进行预处理。在Fog-IBDIS中，数据处理任务被前移到边缘设备，如工业个人计算机和应用服务器。使用Fog-IBDIS，通过在预处理期间减少数据量，并将中间结果进行串联传输实现多源数据集

成，能够在IBDIS中改善时间延迟和减少网络流量负载。此外，企业对中间结果数据的敏感度远远低于源数据，利用Fog-IBDIS仅对中间结果进行传输集成的方式可保护源数据的私密性。

本文其余部分的结构如下。首先，综述了大数据集成和雾计算的国内外研究现状。接着，提出了基于雾计算的工业大数据集成与共享架构Fog-IBDIS，其中包括任务流图（TFG）、任务编译与运行控制、管理组件等。然后，将Fog-IBDIS的功能分成5个模块来实现工业大数据的集成和共享。随后，进行了案例研究，以说明Fog-IBDIS的实施和性能，并讨论了IBDIS与云计算和Fog-IBDIS之间的差异。最后，论述了本文的结论和展望。

## 2. 国内外研究现状

除了互联网大数据具备的大体量、多种类和高速度“3V”特点之外，工业大数据还具有多来源、多维度、多噪声的“3M”特点。如何针对“3V-3M”特性的工业大数据进行建模、集成与处理，一直是学术界与工业界的焦点。Xiang等[25]提出了一种混合制造云架构，实现产品全生命周期中数据的集成与共享。Ma'ayan等[26]开发了一种集中数据管理方法，将系统生物学和系统药理学中的大数据整合到中心数据库中。Mezghani等[27]设计了一种通用的语义大数据平台，设计单一数据池实时管理异构的可穿戴设备数据。这些多样化的医疗状态数据，可用于监控用户的健康状态，并协助医生了解患者的健康状况。在物联网数据的数据管理方面，Jiang等[28]提出了一个数据存储框架，它包括将多个数据库与Hadoop平台相结合以及存储和管理大规模物联网设备收集的各型的数据。借助云计算平台，可以有效地存储和处理快速生成的物联网数据，Chang等[29]整合了几种大数据仓库平台，设计优化数据库结构，实现工业大数据的高性能、高可用性和高可扩展性地集中处理。

鉴于云端服务器集群的计算能力，当前IBDIS的方法通常采用集中式的云计算方式，将所有数据转移到云端进行分析与处理[17]。然而，随着数据量的不断增长，传统的云计算中心化数据集成方法（如Liu等提出的语义集成方法[31]），必然会带来繁重的网络流量负载[30]，对网络带宽提出了新要求。另一方面，集中式

的数据处理方式需要待所有数据收集完毕后才能展开多源数据的集成分析，会降低分析结果的时效性，难以适用于时间敏感型的数据分析任务。传统的数据处理架构遇到了新挑战。因此，需要将云计算框架扩展到网络边缘设备[32]（如智能手机、可穿戴设备和游戏控制器），这种方法称为雾计算，也称为边缘计算[33]。随着传感器和芯片技术的快速发展，雾计算得到了一些先导型的研究。Zhang等[34]研究了一种在协作边缘环境中进行大数据共享和处理的新型计算框架。这种框架能够通过处理靠近数据源的边缘设备数据来改善响应延迟，从而减少向云端的数据传输。Tang等[35]在智能城市中提出了分层分布式雾计算，以支持基础设施组件生成的海量数据的集成，利用雾计算技术，可以快速识别城市中的异常和危险事件，并在不同的时延要求下作出响应。近期很多工作已经证明，与云计算相比，雾计算具有更灵活的架构和更快的响应速度[36–38]。

为解决云计算环境下源数据隐私和网络流量负载问题，本文提出了基于雾计算的制造数据集成与分享解决方案——Fog-IBDIS。使用雾计算技术，我们设计了Fog-IBDIS架构，将算法和模型应用于边缘设备实现数据的本地处理，而不是将数据传输到云端进行分析。在Fog-IBDIS中，边缘服务器可直接实现源数据本地处理以支持低时延迟的数据分析和优化，其仅对处理得到的中间结果进行上传与共享，从而实现源数据隐私保护。通常而言，数据处理后的输出量至少比原始数据的输出量小两到三个数量级。因此，Fog-IBDIS使用雾计算将

大数据处理前移至边缘侧，实现数据分析的去中心化，可解决IBDIS中网络负载与数据私密性问题。

### 3. 基于雾计算的制造数据集成系统架构

制造系统中的大数据具备典型的多来源特性，其包括产品设计数据、制造数据、营销数据、运维数据等，如图1所示。从数据的产生形式来看，制造系统中的大数据可以分为两类：系统数据和物联网数据。系统数据是指由信息系统因业务流转而生成的数据，如电子商务系统（EP）、社交网络平台（SNP）、产品生命周期管理系统（PLM）、企业资源计划系统（ERP）、装备运维系统（MRO）以及供应链管理系统（SCM）。这些信息系统积累了大量的产品研发数据、制造数据、供应链数据、销售数据、客户反馈信息等，是制造系统大数据的重要构成。物联网数据是指由传感器捕获的数据，如射频识别（RFID）捕捉的产品流转信息、温湿度传感器感知的车间环境数据、高速CCD相机捕捉得到的产品表面图像数据等。

在制造系统中，工业数据的集成可分为三类：单源IBDIS（S-IBDIS）、协同IBDIS（C-IBDIS）和多系统IBDIS（M-IBDIS）。在S-IBDIS中，用于分析的所有数据都来自单个数据源，如MES。在这种模式下，因所涉及的数据都是单机存储，可直接实现数据集成。在C-IBDIS模式中，数据是从制造系统中的两个以上数据源获取的。源数据可以通过中间件、虚拟数据库和数据

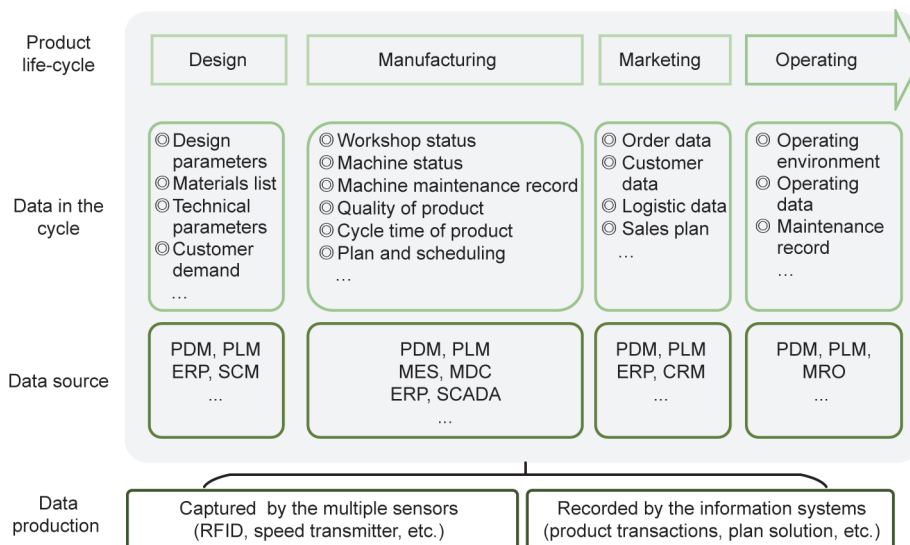


图1. 产品生命周期的工业大数据。CRM：客户关系管理系统；ERP：企业资源计划系统；MDC：制造数据采集系统；MRO：装备运维系统；PLM：产品生命周期管理系统；RFID：射频识别；SCM：供应链管理系统。

仓库技术，使用数据同步/异步复制、数据视图或者接口实现数据集成。在这种模式下，数据仅在单个制造系统或者单个企业内进行交换集成，因此可针对源数据直接集成。M-IBDIS涉及来自不同制造系统中数据源的数据集成，适用于设计多企业产业链的数据分析。在M-IBDIS模式下，为保护数据隐私，企业往往禁止将内部源数据跨企业公开使用。因此，M-IBDIS模型将分析算法迁移至边缘侧，然后聚合中间结果，实现跨企业的数据分析。

为进一步描述数据集成过程，本节提出TFG方法对IBDIS过程进行建模。TFG图是由节点与边构成的二元组，其节点包括数据集节点、操作节点和传输节点。数据集节点是指数据处理中所需的元数据信息；它提供数据集名称、元数据结构 and 提取路径及样例数据。操作节点是用于数据处理的单元，其定义数据分析中的数据输入、输出和分析算法，具体可分为数据清洗节点、数据转换节点、预测节点和聚类节点等。传输节点在两个制造系统之间传输中间结果，以串联多源数据的分析过程。如图2所示，其包含三个IBDIS任务，其中，任务一包含一个数据集节点和操作节点，以处理来自单个数据源“D1”数据的S-IBDIS任务。在任务一中，操作节点“O1”的输出转至任务二，接力完成数据分析任务。任务二是在同一企业内的协同型数据集成任务(C-IBDIS)，其由两个数据源(“D1”和“D2”)与两个操作节点(“O1”和“O2”)实现协作型数据分析。操作节点“O2”的输出由Fog-IBDIS传输至任务三，任务三是跨企业间的M-IBDIS数据分析任务。其在任务二的基础上利用传输节点“T1”将任务二的输出结果传输到任务三，由“O3”“O4”与“O5”三个操作节点完成任务。

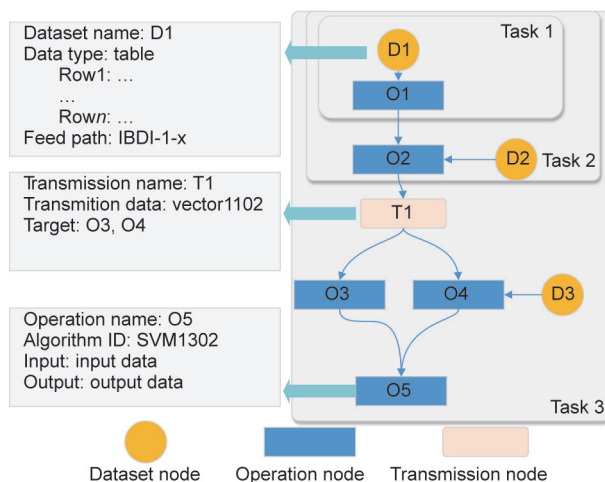


图2. 面向Fog-IBDIS的数据集成任务流程图。IBDI: 工业大数据集成。

TFG图可清晰描述数据处理过程，为企业内单数据源、多数据源与跨企业的数据协作分析提供了数据处理的建模方法。

在Fog-IBDIS架构下，所有数据分析任务都由数据所有者完成，仅通过数据中间计算结果进行数据传输，实现多源数据协作分析。为协调多边缘设备，实现协作数据分析，Fog-IBDIS仍需要雾服务器以控制多边缘设备的协作。雾服务器用于制定与发布TFG图，并指导边缘客户端执行TFG中的任务。如图3所示，雾服务器首先通过操作指令“I1”向边缘客户端发出操作任务，即任务一。此任务包含两个操作节点“O1”和“O2”，它们处理来自两个数据源(“D1”和“D2”)的数据集。接下来，通过数据流“DF4-T1-DF5”将处理结果发送到随后的边缘客户端。边缘客户端根据任务二分析来自“T1”和“D3”的数据，随后通过数据流“DF6”将结果上传到雾服务器。然后，雾服务器将分析结果发送给最终数据用户。在分析过程中，所有源数据处理任务都在源数据边缘端完成，以保护数据隐私，且仅在边缘客户端之间传输数据分析的中间结果，以减轻网络流量负载。

## 4. Fog-IBDIS 的功能模块设计

针对基于边缘计算的大数据融合需求，本节介绍IBDIS的功能模块(图4)，包括：TFG任务管理、编译和运行控制、数据集成运输、基本算法库及管理组件五个部分。

### 4.1. TFG 管理

TFG任务管理是Fog-IBDIS的核心功能，其包含四个部分：数据集节点管理(DNM)、操作节点管理(OND)、传输节点管理(TNM)和边缘管理(EM)。DNM维护TFG中的数据集节点，该节点由数据结构定义、数据格式和提取路径定义组成。通过DNM模块，数据生成器可以编辑并发布数据视图，提供数据集的详细描述信息。在数据视图的帮助下，数据用户可以通过提供的数据视图搜索并预览样例数据。OND为IBDIS中的操作节点提供编辑、打包和校验功能，如数据清理、数据转换与分析等。通过OND模块，数据客户将分析过程的所有操作下放至多个节点中。TNM管理数据分析中的传输过程，包括数据编码、解码、上传和下载。为了提高传输安全性，数据在传输过程中

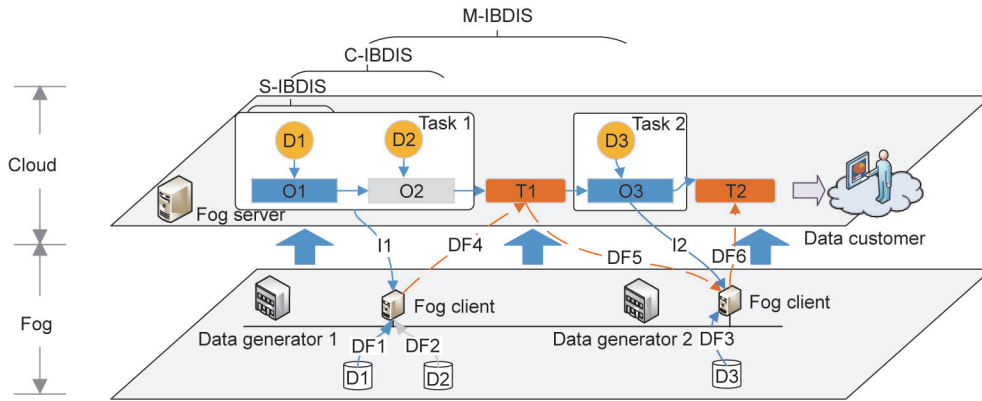


图3. Fog-IBDIS运行原理图。DF: 数据流; I: 指令。

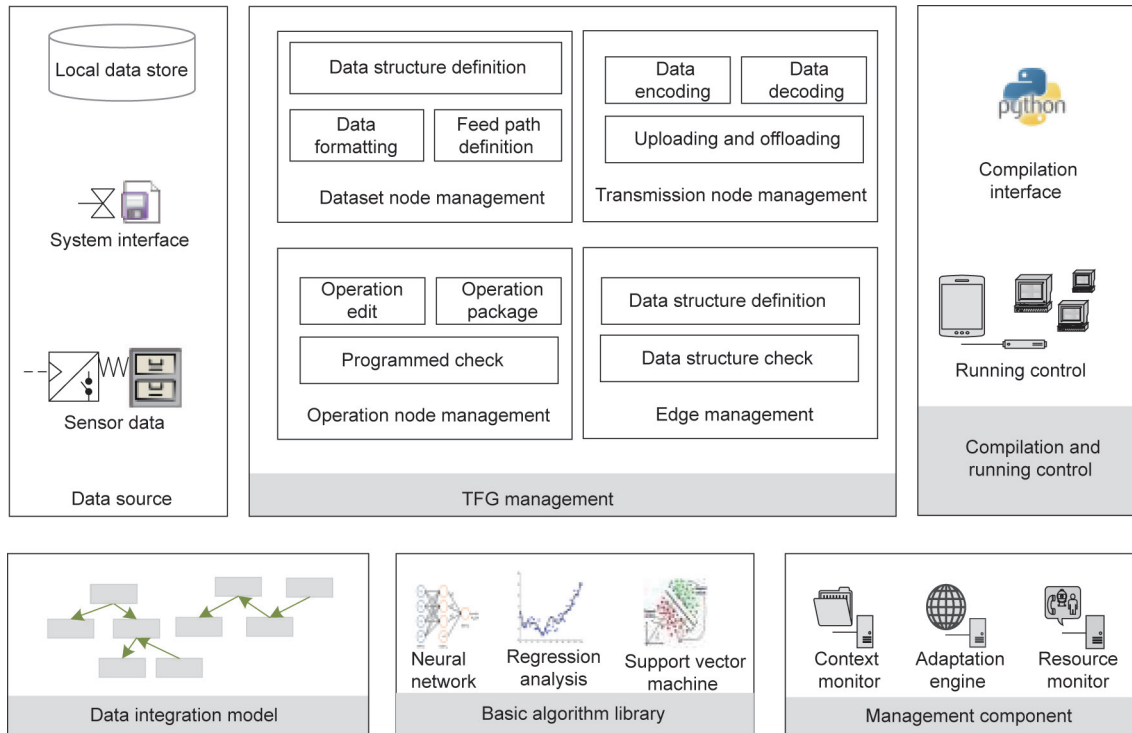


图4. Fog-IBDIS平台功能架构。

被编码为带有私钥的密文。EM模块定义并检查在两个相邻节点期间自动传输的数据结构，以确保数据分析任务的有效性。

#### 4.2. 编译和运行控制

Fog-IBDIS平台提供用于编译和运行的数据分析应用程序的编程接口。所有调试任务都由数据用户使用样例数据完成。在调试之后，IBDIS任务由编译器编译并通过Fog-IBDIS平台发送到边缘客户端。基于TFG，编译的文件分布在属于不同企业的多个Fog-IBDIS服务器中。其运行环境和配置参数由Fog-IBDIS平台确定和调整。

#### 4.3. 数据集成模型

在工业大数据分析过程中，所需的数据存储在不同的业务系统中，为实现源数据与分析所用的元数据一一对应，数据集成模型通过目标数据项、实体、系统和源数据项四层模型来实现元数据的对应描述（图5）。

- 目标数据视图。其包含若干元数据项，分别用于TFG图中不同的数据节点。
- 实体。实体用于标识元数据项之间的关系。实体与所属数据项之间相对应，属于不同实体的数据项通过实体关系链相互关联。
- 系统。系统定义了数据项的存储信息系统，如MES、MDC等。利用系统组件，可以通过信息系统的

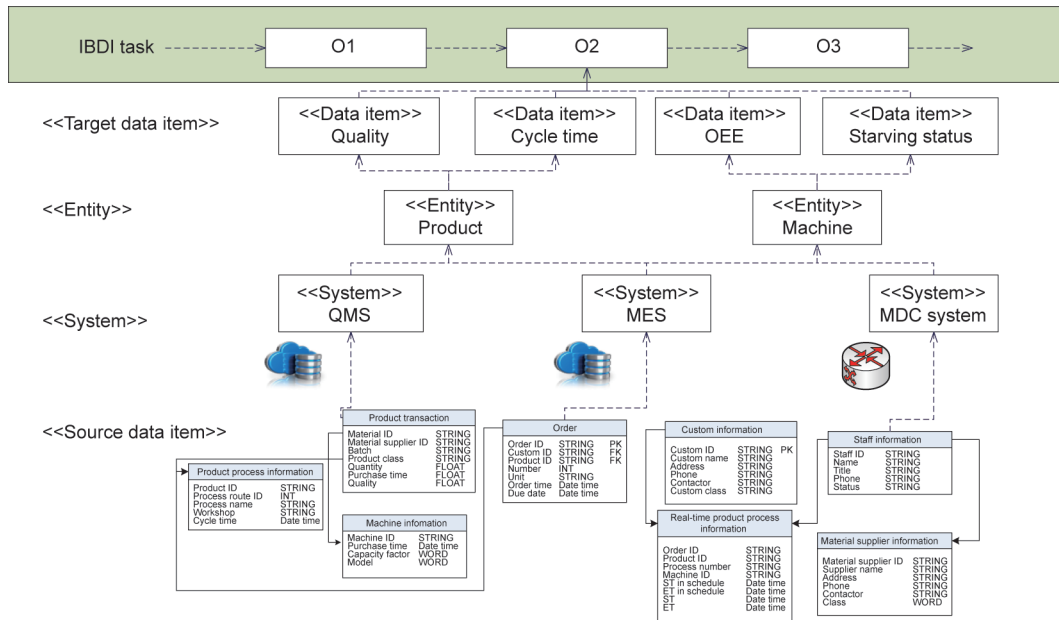


图5. 基于雾计算的IBDIS数据集成4层模型。OEE: 设备总体效率; QMS: 质量管理体系。

不同接口提取数据。

- 源数据项。源数据项定义了各信息系统中的数据结构，包括数据库和接口中的源字段。利用源数据项，数据集成单元可以准确地精确定位提取源数据。

#### 4.4. 基本算法库

基本算法库为Fog-IBDIS中的数据转换的简单数据分析提供底层支持。数据用户可以调用和修改基本算法，以满足数据处理的要求，如异常值分析、缺失值插值、神经网络、回归分析、支持向量机等。

#### 4.5. 管理组件

管理组件定义了边缘客户端所需的软硬件功能要求。通常，完整的管理组件包含三个主要部分：环境监视器、资源监视器和适配引擎。环境监视器和资源监视器用于监视设备的状态信息，包括资源可用性和实时任务执行信息。适配引擎可以实现数据精简和系统兼容。

## 5. 案例验证

为了验证Fog-IBDIS的有效性，以上海某飞机制造厂为对象进行大数据分析案例验证。在该场景中，飞机的装备工艺由研发企业完成设计，由制造厂完成装备，其隶属于同一集团的不同公司。在飞机装配过程中，不同的工艺对定位装置的应力有不同的影响，定位器是固定和支撑平面的关键设备，可以通过分析定位器状态来

推断飞机的状态。在本案例中，研发企业希望通过监控装配过程中机翼的姿态变形来优化工艺路线。该案例需要姿态数据（包含机翼上多个关键姿态点的位置）和过程事务数据（包含每个过程步骤的执行时间）。姿态数据可以从子装配制造系统的MDC系统导出，其为数据生成器。过程事务数据由研发企业的计划和调度组管理的MES导出，MES在本案例中同时起着数据生成器和数据用户的作用。

#### 5.1. 用于过程路径优化的 Fog-IBDIS 数据集成

为集成过程路径优化的数据，设计了具有两个任务的TFG，其包含两个数据集节点、三个操作节点和一个传输节点（图6）。第一个操作节点对机翼的多个关键点位置进行预处理，包括数据清洗和转换。第二操作节点通过对机翼关键点位置进行分析来检测异常的姿态变形。然后通过Fog-IBDIS平台将异常数据片段传送到任务二中的第三个操作节点。在该操作节点中，这些数据片段被映射到过程事务数据中，以便诊断出导致异常姿态变形的原组装过程。

第一个操作节点O1对机翼的姿态状态进行预处理，该姿态状态是测量通过图7（a）所示的三个关节的力的特征而获得的。力的数据由嵌入在定位器中的传感器测得，并由SCADA系统通过开放平台通信技术收集。首先，如图7（b）所示，在数据清洗中检测并恢复数据记录中的空值和异常值。接下来，根据数据集成模型，数据项通过数据立方体进行转换，数据立方体作用

是通过深挖、卷积、分段以及抽取等操作将源数据自定义为目标数据的结构（商业中的数据转换）。例如，我们设计一个包含有关定位器状态的事务数据的数据集。它包含三个维度的记录：机器、场地和时间。立方体的每个单元 ( $M, F, T$ ) 包含  $M$  机器在  $T$  时间时的  $F$  场地的值。在该示例中，大数据分析中需要定位器 LWA, LWF 和 LWO 在时间  $T_4$  处的  $z$  轴力。如图 7 (c) 所示，在这种情况下，设计的操作“立方体切片—卷积—卷积”可以获得目标数据。如图 7 (a) 所示，LWA、LWF 和 LWO 是左翼的三个定位器。

第二个操作节点 O2 检测机翼的异常姿态变形，并

将异常数据片段发送给任务二。在机翼设定角度的偏差（由 E1、E2 和 E3 描述）、使用状态分析模型估计正二面角的偏差（由 E4 和 E5 描述）和扫描角的偏差（由 E6 描述）[图 7 (a)] 这三个关键点位置的数据分析中，使用了位置分析模型来估计。为检测异常情况，通过控制图检测机翼状态的异常偏差，当误差超过控制上下限（如误差要在 5% 以内）时，数据点被认为是异常的。这时检测到的异常数据通过 Fog-IBDIS 传输到任务二。之后偏差数据会被映射到第三操作节点“O3”，该操作节点将通过专家经验对偏差数据和装配过程的组合分析来诊断异常偏差的根本原因。最终，获得了能够改善装配过程并

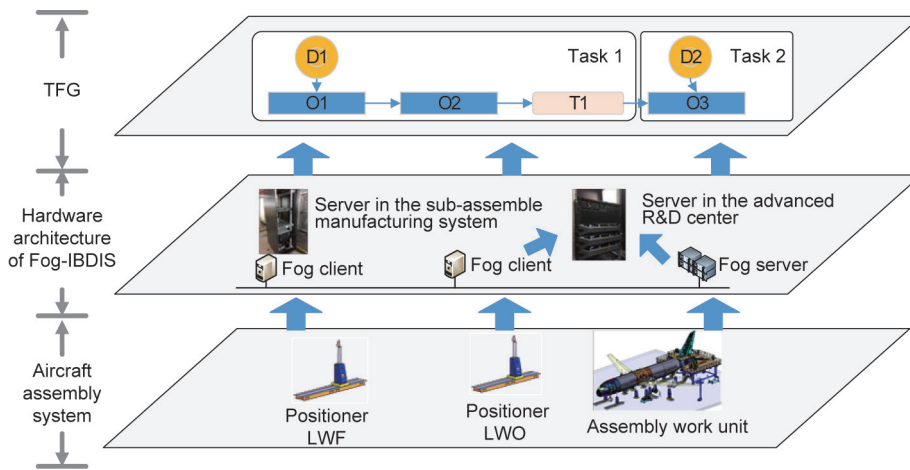


图6. 面向飞机装配工艺路线优化的Fog-IBDIS任务。LWF和LWO为左翼定位器。

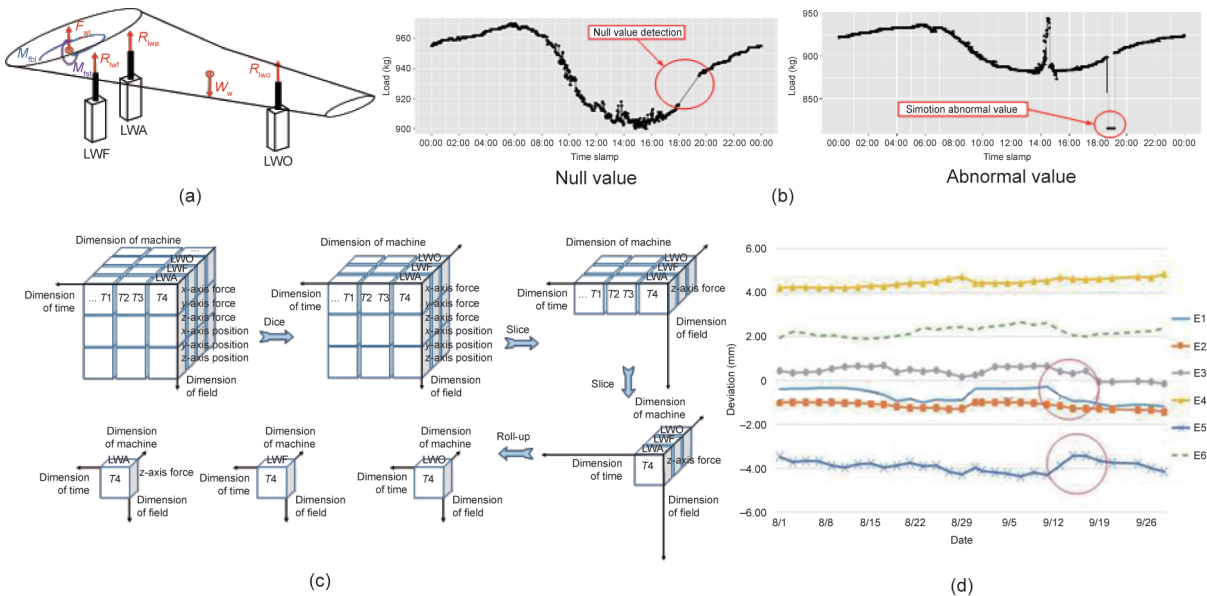


图7. 基于Fog-IBDIS的定位器状态监测与分析。(a) 机翼状态分析模型；(b) 数据清洗；(c) 面向数据处理的数据立方体；(d) 三轴偏差分析结果。Fw1: 机身到左翼的支撑反力；Mfb1: 横向力矩；Mfst: 翼展方向力矩；Rlwf: 定位器LWF到左翼的支撑反力；Rlwa: 定位器LWA到左翼的支撑反力；Rlwo: 定位器LWO到左翼的支撑反力；Ww: 左翼重力；E1、E2、E3表示安装角的偏差值，E4、E5表示上反角偏差值，E6表示后掠角偏差值。

减少姿态变形程度的改进建议。

在此案例中，Fog-IBDIS从飞机制造系统中获取工业大数据，并将数据集成任务分配给雾客户端，以提供分析数据进而优化工艺路线。如图7（d）所示，机翼姿态在9月10日到9月14日期间发生了很大的变化。通过数据分析可以推断出这种变化是由工艺路线的不合理设计引起的：在对飞机左翼的连接孔进行去毛刺清洗之前，移除了夹具连接。在此案例中，2016年8月1日至9月30日收集的源数据量为1.7 TB，而传输的中间结果数据量为160 kB。在本案例研究的数据传输中，传输数据量仅为源数据量的 $9.1 \times 10^{-8}$ 倍，这表明Fog-IBDIS可以大大减少传输数据量，从而减轻网络流量负载。

## 5.2. Fog-IBDIS 的实施

此案例中采用了Fog-IBDIS方法用于集成工业大数据。实施过程（图8）分为如下5个步骤。

第一步：模型设计。在IBDIS中，首先，为了数据转换，通过描述大数据集成模型来定义数据结构。数据模型将同一个分析主体的所有数据项与语义三元组链接在一起，并且定义了目标数据项与源数据项之间的关系。

第二步：TFG设计。根据大数据集成模型，在TFG中定义了所有节点以实现大数据的集成和共享。这些节点分装为两个任务，每个任务都由数据用户用Python编程实现。

第三步：任务调试。数据用户通过数据视图中包含的示例数据实例，根据特定TFG来编译和调试任务。编译应用程序在雾客户端中启动，并进行测试和调试，以便通过Fog-IBDI平台验证TFG和代码的设计。

第四步：任务审核。在程序调试和修改后，任务被提交给数据生成器以获取任务操作的权限。

第五步：任务执行。在任务审查之后，这些任务在Fog-IBDIS平台的协调下在雾客户端中进行执行，以便工业大数据的集成。

## 5.3. 讨论

为进一步论证Fog-IBDIS的有效性，我们仔细比较了Fog-IBDIS和基于云计算的IBDIS。Fog-IBDIS在数据集成框架与基于云计算的IBDIS有如下不同。

分散化。在基于云计算的IBDIS中，物联网设备和信息系统向大数据中心提供数据，所有数据提取、转换和加载任务都由集中式大数据平台完成。尽管大数据中

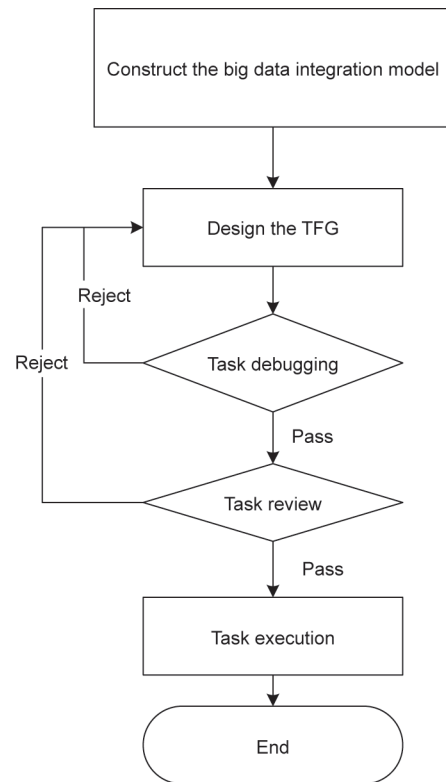


图8. Fog-IBDIS的实现过程。

心具有巨大的计算能力，但分离、索引、存储和清洗大量工业大数据仍然具有挑战性。通过雾计算，Fog-IBDIS将分析案例的大量数据处理任务分散到边缘节点，有效地减轻了制造系统中大数据平台的压力。

延迟。与云计算相比，雾计算对网络边缘的数据进行并行化处理，满足了工业控制实时数据分析的要求。通过使用Fog-IBDIS，工业大数据可在边缘端进行处理，然后传输到云端以进行进一步分析；这能够支持不同级别时间延迟要求的数据分析。

数据隐私。在具有云计算的数据共享服务中，共享的数据可以轻松地被复制、共享和二次交易。由于服务器直接共享源数据，因此无法保证源数据隐私。在工业大数据中，源数据通常是严格保密的，只允许在制造系统中使用。与基于云计算的IBDIS（将所有源数据传输到云数据中心）不同，Fog-IBDIS仅将中间结果传输到下一个雾客户端，然后将分析结果上传给数据用户。此外，Fog-IBDIS引入了任务审查机制，允许数据生成器检查源数据安全性，这可以保护数据集成中的数据隐私。

网络流量负载。在云计算中，所有源数据都传输到大数据中心，这些大数据中心配备了分布式存储平台，如Hadoop分布式文件系统。制造系统中的一些源



数据体量巨大，如通过激光扫描得到的大量图像，此外，需要大量带宽来上传这些图像。使用Fog-IBDIS，只通过网络上传中间处理结果，通过数据清洗、重采样和数据转换，减少了大部分源数据的体量，从而有效降低了网络通信的成本。此外，与云计算的批处理不同，Fog-IBDIS通过将数据上传到不同数据点进行网络流量分散化，降低了网络的峰值负载。

## 6. 结论

针对工业大数据分析中源数据的私密性与海量数据传输带来的网络负载问题，本文提出了基于雾计算的大数据集成与共享架构Fog-IBDIS。与之前的基于云计算的大数据集成研究不同，本文采用雾计算构建Fog-IBDIS平台，将IBDIS任务分解为若干独立运行的子任务，由数据所有者进行数据的处理和分析，以保障源数据私密性。此外，Fog-IBDIS将数据处理任务前移至系统内的边缘节点，从而将集中式数据处理模式变为分布式任务执行。在当前架构下，仅需分析中间结果可在分布式边缘客户端之间传输，可极大减少传输数量，减轻网络流量负载。随着边缘设备运算能力的进一步增强，雾计算将在数据私密性强、实时性要求高的工业领域得到进一步的应用。在未来的研究中，本科研发团队将在航天发动机制造、纺织等领域进一步研究基于边缘计算的工业数据分析方法。

## Acknowledgement

本文得到国家自然科学基金重点项目（51435009）的资助。

## Compliance with ethics guidelines

Junliang Wang, Peng Zheng, Youlong Lv, Jingsong Bao, and Jie Zhang declare that they have no conflict of interest or financial conflicts to disclose.

## References

- [1] Hughes D, Ueyama J, Mendiondo E, Matthys N, Horr  W, Michiels S, et al. A middleware platform to support river monitoring using wireless sensor networks. *J Braz Comput Soc* 2011;17(2):85–102.
- [2] Jiang P, Ding K, Leng J. Towards a cyber–physical–social-connected and service-oriented manufacturing paradigm: social manufacturing. *Manuf Lett* 2016;7:15–21.
- [3] Wang JL, Zhang J. Big data analytics for forecasting cycle time in semiconductor wafer fabrication system. *Int J Prod Res* 2016;54(23):7231–44.
- [4] Wang JL, Zhang J, Wang XX. Bilateral LSTM: a two-dimensional long shortterm memory model with multiply memory units for short-term cycle time forecasting in re-entrant manufacturing systems. *IEEE Trans Industr Inform* 2018;14(2):748–58.
- [5] Zhang WJ, Lin Y. On the principle of design of resilient systems—application to enterprise information systems. *Enterprise Inf Syst* 2010;4(2):99–110.
- [6] Zhang WJ, van Luttervelt CA. Toward a resilient manufacturing system. *CIRP Ann* 2011;60(1):469–72.
- [7] Tsuda T, Inoue S, Kayahara A, Imai S, Tanaka T, Sato N, et al. Advanced semiconductor manufacturing using big data. *IEEE Trans Semicond Manuf* 2015;28(3):229–35.
- [8] Lu C, Li X, Gao L, Liao W, Yi J. An effective multi-objective discrete virus optimization algorithm for flexible job-shop scheduling problem with controllable processing times. *Comput Ind Eng* 2017;104:156–74.
- [9] Lei CU, Man KL, Liang HN, Lim EG, Wan KY. Building an intelligent laboratory environment via a cyber–physical system. *Int J Distrib Sens Netw* 2013;9(12):109014.
- [10] Wang JL, Zhang J, Wang XX. A data driven cycle time prediction with feature selection in a semiconductor wafer fabrication system. *IEEE Trans Semicond Manuf* 2018;31(1):173–82.
- [11] Wang W, Chong W, Liu D, Liang HN, Man KL, Han YS, et al. An examination of the internet of things through the data management perspective. *J Platf Technol* 2014;2(2):16–30.
- [12] Lu C, Gao L, Li XY, Chen P. Energy-efficient multi-pass turning operation using multi-objective backtracking search algorithm. *J Clean Prod* 2016; 137:1516–31.
- [13] Lu C, Gao L, Li XY, Xiao SQ. A hybrid multi-objective grey wolf optimizer for dynamic scheduling in a real-world welding industry. *Eng Appl Artif Intell* 2017;57:61–79.
- [14] Kusiak A, Xu GL. Modeling and optimization of HVAC systems using a dynamic neural network. *Energy* 2012;42(1):241–50.
- [15] Zhong RY, Xu C, Chen C, Huang GQ. Big data analytics for physical Internetbased intelligentmanufacturing shop floors. *Int J Prod Res* 2017;55(9):2610–21.
- [16] Zhang W. An integrated environment for CAD/CAM of mechanical systems [dissertation]. Delft: TU Delft; 1994.
- [17] Majkić Z. Big data integration theory: theory and methods of database mappings, programming languages, and semantics. Heidelberg: Springer; 2014.
- [18] Wang G, Gunasekaran A, Ngai EWT, Papadopoulos T. Big data analytics in logistics and supply chain management: certain investigations for research and applications. *Int J Prod Econ* 2016;176:98–110.
- [19] Mourtzis D, Vlachou E, Milas N. Industrial big data as a result of IoT adoption in manufacturing. *Procedia CIRP* 2016;55:290–5.
- [20] Lim JB, Yu HC, Gil JM. An efficient and energy-aware cloud consolidation algorithm for multimedia big data applications. *Symmetry* 2017;9(9):184.
- [21] Kadadi A, Agrawal R, Nyamful C, Atiq R. Challenges of data integration and interoperability in big data. In: *Proceedings of 2014 IEEE International Conference on Big Data*; 2014 Sep 27–30; Washington, DC, USA. Piscataway: IEEE; 2015. p. 38–40.
- [22] Hashem IAT, Yaqoob I, Anuar NB, Mokhtar S, Gani A, Khan SU. The rise of “big data” on cloud computing: review and open research issues. *Inf Syst* 2015;47:98–115.
- [23] Wang JL, Yang JG, Zhang J, Wang XX, Zhang WJ. Big data driven cycle time parallel prediction for production planning in wafer manufacturing. *Enterprise Inf Syst* 2018;12(6):714–32.
- [24] Pan WK, Yang Q, Aggarwal C, Koch C. Big data. *IEEE Intell Syst* 2017;32(2):7–8.
- [25] Xiang F, Yin Q, Wang Z, Jiang GZ. Systematic method for big manufacturing data integration and sharing. *Int J Adv Manuf Technol* 2018;94(9–12):3345–58.
- [26] Ma’ayan A, Rouillard AD, Clark NR, Wang ZC, Duan QN, Kou Y. Lean big data integration in systems biology and systems pharmacology. *Trends Pharmacol Sci* 2014;35(9):450–60.
- [27] Mezghani E, Exposito E, Drira K, Da Silveira M, Pruski C. A semantic big data platform for integrating heterogeneous wearable data in healthcare. *J Med Syst* 2015;39(12):185.
- [28] Jiang L, Xu LD, Cai H, Jiang Z, Bu F, Xu B. An IoT-oriented data storage framework in cloud computing platform. *IEEE Trans Industr Inform* 2014;10(2):1443–51.
- [29] Chang BR, Tsai HF, Tsai YC, Kuo CF, Chen CC. Integration and optimization of multiple big data processing platforms. *Eng Comput* 2016;33(6):1680–704.
- [30] Suárez-Albela M, Fernández-Caramés TM, Fraga-Lamas P, Castedo L. A practical evaluation of a high-security energy-efficient gateway for IoT fog computing applications. *Sensors* 2017;17(9):E1978.
- [31] Liu X, Zhang WJ, Radhakrishnan R, Tu YL. Manufacturing perspective of enterprise application integration: the state of the art review. *Int J Prod Res* 2008;46(16):4567–96.
- [32] Varghese B, Wang N, Barbhuiya S, Kilpatrick P, Nikolopoulos DS. Challenges and opportunities in edge computing. In: *Proceedings of IEEE International Conference on Smart Cloud*; 2016 Nov 18–20; New York, NY, USA. Piscataway: IEEE; 2016. p. 20–6.
- [33] Shi W, Dustdar S. The promise of edge computing. *Computer* 2016;49(5):78–81.

- [34] Zhang Q, Zhang XH, Zhang QY, Shi WS, Zhong H. Firework: big data sharing and processing in collaborative edge environment. In: Proceedings of the 4th IEEE Workshop on Hot Topics in Web Systems and Technologies; 2016 Oct 24–25; Washington, DC, USA. Piscataway: IEEE; 2016. p. 20–55.
- [35] Tang B, Chen Z, Hefferman G, Wei T, He H, Yang Q. A hierarchical distributed fog computing architecture for big data analysis in smart cities. In: Proceedings of the ASE Big Data & Social Informatics; 2015 Oct 7–9; Kaohsiung, Taiwan, China. New York: ACM; 2015.
- [36] Kumar N, Zeadally S, Rodrigues JJPC. Vehicular delay-tolerant networks for smart grid data management using mobile edge computing. *IEEE Commun Mag* 2016;54(10):60–6.
- [37] Liu JQ, Wan JF, Zeng B, Wang QR, Song HB, Qiu MK. A scalable and quickresponse software defined vehicular network assisted by mobile edge computing. *IEEE Commun Mag* 2017;55(7):94–100.
- [38] Park HD, Min OG, Lee YJ. Scalable architecture for an automated surveillance system using edge computing. *J Supercomput* 2017;73(3):926–39