

大数据需求分析

大数据需求分析是指在大数据应用软件开发启动前,以挖掘数据价值、满足业务需求为目的,对业务功能场景、数据资源、非功能需求等进行全面系统性分析的过程。相较于传统软件开发需求分析,大数据需求分析具有更强的业务导向性和数据驱动性,需要更多地从数据角度考虑,进行跨领域的关联分析。同时,大数据应用的需求变化也更加动态,需求分析需要进行不断迭代和优化。大数据需求分析是一个大数据应用项目或产品开发过程中至关重要的一步,通过需求分析,开发者能够准确地了解用户的需求和期望,从而确定项目的目标和范围,确保项目的顺利进行和成功实施。



5.1 需求的来源与特点

5.1.1 业务功能→数据→需求

在进行大数据需求分析之前,首先需要明确大数据分析的核心目标,即将复杂、海量的数据转换为有价值的信息和知识,以支持业务决策或实现特定任务。这一过程涉及对数据本身的理解和对业务功能背景的深刻洞察,数据是业务功能活动的产物,同时也是业务功能活动的推动力,了解业务功能背后的数据意味着深入理解数据是如何与业务活动相关联的,这需要对业务功能的流程、目标和需求有深入的了解,以便确定需要收集和分析哪些数据。此外,这一过程还涉及从需求到实际的数据分析任务的转化,将抽象的需求具体化为可执行的分析步骤,确保分析工作的每个环节都紧密围绕最终目标展开,从而最大化数据的价值。

如图 5.1 所示,业务功能、数据和需求三者之间密不可分。业务功能产生数据,数据为业务功能和需求提供支持,需求则最终落脚到业务功能和数据中。事实上,业务功能和需求决定了需要收集和分析的数据类型和规模,数据分析结果又反馈给业务功能并进一步影响需求。三者之间相互作用和影响,共同推动着大数据应用软件

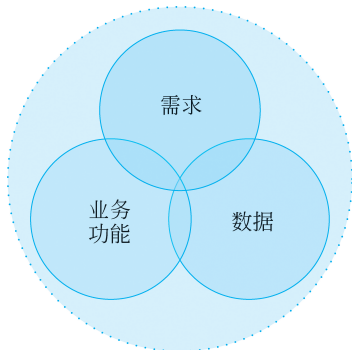


图 5.1 业务功能、数据和需求三者之间密不可分

开发的进程。

与传统软件开发中的需求分析相比,大数据应用软件开发中的需求分析更加注重数据的价值挖掘和业务功能目标的实现。数据是大数据需求分析的基础。大数据应用需要大规模、多样化和高质量的数据支撑,以实现对海量数据的深入分析和挖掘,从而为业务决策和业务功能创新发展提供有力支持。

5.1.2 大数据背景下需求分析的特点

在大数据时代,需求分析不再是简单的需求收集和整理过程,而是一项复杂的、多维度的综合性工作,涉及技术、法律和商业等多方面。相较于传统软件开发中的需求分析,大数据需求分析更加注重数据本身的特性和价值挖掘。这不仅要求开发人员具备跨学科的知识背景,包括数据科学、计算机科学和行业知识,还要求他们具备前瞻性的思维,能够预见数据在未来的潜在价值和应用场景。具体来说,大数据背景下的需求分析具有以下特点。

1. 数据的多样性和复杂性

大数据需求分析所涉及数据具有多样性和复杂性的特点,不仅来源广泛,而且类型各异,涵盖了结构化数据(如数据库中的表格数据)、半结构化数据(如JSON、XML文件)和非结构化数据(如文本、图像、视频)等。在传统软件开发中,需求分析往往聚焦于功能需求和性能需求,而大数据应用软件则需要从一开始就综合考虑数据资源的多样性和复杂性。这意味着大数据需求分析不仅要明确数据从哪里来,还要清楚数据将以何种速度流动、数据存储和处理的体量,以及如何保证数据的质量和准确性。在此过程中,选择合适的数据存储和处理技术成为关键,这往往涉及分布式系统、实时数据流和高效的计算能力。

2. 需求的动态性和不确定性

在大数据应用项目中,需求分析的一个显著特点是其动态性和不确定性。大数据应用项目的开发往往不是一个固定不变的过程,它随时可能受数据质量、业务功能需求变化、技术限制等多种因素的影响。一方面,大数据应用软件的需求往往随着项目的推进和业务环境的变化而不断调整和完善。由于大数据应用软件通常涉及海量数据和复杂的业务场景,需求往往无法一次性确定,需要在软件实施过程中不断调整和优化。另一方面,大数据技术的快速发展也会影响需求分析,新的技术可能会带来新的需求或改变原有需求的实现方式。其中涉及的不确定性来源于多方面,包括数据质量、业务需求的变化及技术限制等因素,这些因素使得需求分析过程充满挑战。

3. 功能的实时性和高效性

随着实时数据分析需求的增长,很多大数据应用软件必须能够处理流式数据,具有实时处理和快速响应的能力。因此,在进行大数据需求分析时,需要明确数据处理的实时性要求,确保系统能够在数据生成的瞬间进行采集、处理和分析。这种需求常见于金融、电子商务、社交媒体等领域,实时的数据流需要及时转化为决策支持。在这种环境下,数据存储和计算架构的高效性显得尤为重要。需求分析需要考虑如何设计系统以满足低延迟、高并发的需求,同时确保数据分析结果的准确性和时效性。例如,电商平台可能需要实时监控用户行为数据,以便实时调整推荐系统,推送个性化商品。因此,需求分析不仅要关注传统的批量数据处理,更要关注如何高效、准确地处理和分析实时数据流。

5.2 大数据需求分析的过程

为适应大数据场景的复杂性和变动性,大数据需求分析需要规范的流程。具体来说,大数据需求分析内容可按照以下顺序进行:业务功能需求分析,描绘业务场景,确定问题、目标、功能概念以及人机交互和视觉呈现需求;数据需求分析,明确关键数据的规模、类型等;模型需求分析,定制选择适合业务场景的模型算法;非功能需求分析,包括性能和安全等方面的约束;测试需求分析,确保软件能满足前述功能、数据、模型以及非功能需求中的规格要求。在此基础上,对需求进行确认、验证和审查。这种明确分步骤的大大数据需求分析策略,可以有效减少需求偏差和遗漏,明确需求之间的优先级和依赖关系,为后续设计开发奠定坚实基础。

5.2.1 大数据需求分析内容

相较于传统软件开发主要关注于业务功能需求,大数据需求分析更加强调从数据出发,根据数据规模和特征细化需求。因此,在其需求内容上扩充了数据、模型等方面的需求,进行了针对大数据场景的内容扩展,这是由大数据应用软件开发面临海量异构数据以及数据驱动业务创新模式的特点所决定的。

从业务功能、数据、模型、非功能性、测试五方面进行需求分析,可以有效弥补传统需求分析在面对大数据场景时的局限。例如,在数据需求分析方面,传统需求分析可能只关注数据量和数据类型,而在大数据场景下,需要更关注数据的实时性、准确性和多样性。通过对数据维度的需求分析,可以明确数据的来源、质量、清洗和整合需求,从而为后续数据处理和分析提供基础。在模型需求分析方面,传统软件开发的需求分析中很少涉及建模方面的需求,而在大数据应用软件中,需要考虑模型的选型、规模、复杂度和准确性等,以适应特定的大数据分析任务。

1. 业务功能需求

大数据应用软件开发中的业务功能需求是指用户基于业务目标提出的对大数据应用软件整体功能和性能的预期描述。业务功能需求从业务出发阐明软件要达成的核心价值,是大数据应用软件开发需求分析的重要指导意见,其他需求必须服务于业务功能需求的实现。充分明确业务功能需求,对指导大数据应用软件开发设计满足业务目标,使软件功能契合复杂业务场景至关重要。

与传统软件开发不同,大数据应用软件开发中的业务功能需求更加强调从海量异构数据出发,以数据驱动业务模式创新。因此大数据应用软件开发需要对业务功能需求进行扩展,不仅要关注实现特定功能,更要描述软件如何从数据中挖掘价值,提供决策支持,并实现相应的可视化展示。这需要根据大数据环境下的数据规模和业务场景的复杂性对业务功能需求进行拓展和细化,以突破传统需求分析在处理充满数据驱动智能需求的复杂业务场景时的局限性。

2. 数据需求

大数据应用软件开发中的数据需求,是对驱动软件实现业务目标所需的多源异构数据的规模、种类、格式、实时性等方面的描述。与传统软件开发不同,大数据应用软件开发数据

需求分析主要基于海量数据,考虑多种结构化、半结构化、非结构化数据的融合处理。准确描述这些海量异构数据的属性,可以指导大数据应用软件开发针对数据复杂性进行功能设计,实现数据采集、存储、处理、分析等模块与数据场景的契合。

与传统软件开发的数据需求侧重功能驱动不同,大数据应用软件开发的数据需求更强调从复杂数据场景出发,根据数据特征推导出功能需求。比如明确用户日志数据的规模、字段、生成频率等属性,可以推导出日志采集、存储、解析、挖掘等功能需求。进行细致和全面的数据需求分析,不仅能提高需求质量,还可以弥补传统方式难以处理海量复杂数据场景的不足,指导大数据应用软件开发满足数据处理实际需求的功能。

3. 模型需求

大数据应用软件开发中的模型需求是指对软件中所需数据分析方法和模型的描述,包括统计分析模型、机器学习模型和数据挖掘模型等。模型需求明确软件需要使用的算法手段来发掘数据价值,是实现数据驱动业务需求的关键所在。准确描述模型需求,对定制选择适合业务场景的数据模型算法以达成业务目标,实现建模功能与海量异构数据分析的有效联合至关重要。

与传统软件开发不同,大数据应用软件开发中的模型需求需要考虑海量异构数据的特点。这需要根据数据的维度和分布等特征选择合适的模型,以突破传统建模方式面对动态、杂乱数据的局限。扩展后的建模需求可以指导定制选择具备处理大数据能力的高效算法,以达到深度数据分析的效果。

4. 非功能需求

在大数据需求分析中,非功能需求是指与大数据应用软件的功能无关的需求,通常涉及软件的性能、可靠性、安全性、可维护性、可扩展性等方面。非功能需求对于大数据应用软件至关重要,因为大数据处理往往需要考虑对海量数据的处理和分析,以及系统的高性能和可靠性。

与传统软件开发相比,大数据应用软件开发中的非功能需求呈现出一些独特的特点,主要体现在以下几方面。

(1) 性能需求的复杂性:大数据应用软件开发由于处理的数据量巨大、数据类型多样,要求其具备高并发处理能力和快速数据存取能力。与传统软件开发相比,大数据应用软件需要更高水平的性能优化,比如通过分布式计算和存储来提高系统的吞吐量和响应速度。

(2) 可扩展性的重要性:大数据应用软件开发常常涉及对海量数据的实时处理与分析,可扩展性成为大数据应用软件开发的一项基础非功能需求。系统需要能够灵活地随业务需求的增长而扩展,通过增加节点或集群来处理更大规模的数据。

(3) 安全性:在大数据应用软件开发过程中,数据量的增加与数据类型的复杂化使得安全性变得更加突出。大数据应用软件开发需要针对庞大的数据集设计更为复杂的安全策略,包括数据加密、匿名化处理 and 合规性管理等,以保护用户的隐私权和数据安全。

5. 测试需求

测试需求是指在大数据应用软件开发过程中,为确保系统或产品的功能、性能和质量达到预期标准而需进行的测试活动的具体要求,它们通常来源于前面所述的用户功能需求、数据需求、模型需求和非功能需求的规格说明。在进行需求分析时,明确并记录测试需求,可

以帮助测试团队在后续的验证阶段进行更有效的测试设计和执行,同时降低项目风险,提高产品质量。

与传统软件开发不同,大数据应用软件开发中的测试需求更加复杂和动态,需要考虑大量的数据、多样的数据类型以及分布式系统的特殊性。在数据方面,需要确保数据在采集、存储和处理过程中的完整性和一致性;在平台架构方面需要关注集群的性能、数据传输速度、负载均衡和容错能力;在安全性方面,需要考虑数据隐私和安全性方面的测试,尤其是在处理敏感数据时,需要遵循相关法律法规;此外,在用户体验方面,需要测试可视化界面的友好性、交互性,以及用户在使用软件时的体验,确保数据分析结果易于理解和使用。

5.2.2 需求确认、验证和审查

在大数据应用软件开发过程中,需求确认、验证和审查是实现项目成功不可或缺的环节。由于数据的复杂性、多样性以及用户需求的动态变化,通过有效的需求确认,可以明确用户的真实需求;通过需求验证,保证需求的准确性和可行性;通过需求审查,可确保需求文档的完整性和一致性。

1. 需求确认

需求确认是在软件开发初期,根据业务目标和用户需求明确软件需实现的功能与特性。对于大数据应用软件而言,需求确认的准确性直接关系到软件的最终效果与业务价值。它的主要目标是澄清用户的真实需求,引导开发团队明晰项目目标,避免由于需求不明确引发的开发偏差和资源浪费。

在需求确认阶段,利益相关者,包括用户、产品经理、系统分析师等,需充分沟通以确保对需求的全面理解。可以采用多种方法,如访谈、问卷、头脑风暴等,收集不同视角下的需求。

2. 需求验证

在需求确认完成后,需要进行需求验证。需求验证是确保需求定义正确、完整和一致的过程。验证阶段旨在消除需求错误,避免在后续的开发过程中造成大量修正,同时确保项目团队和利益相关者对系统的期望一致,并最终满足用户的实际需求。

一般来说,需求验证的方法可以分为静态验证和动态验证。静态验证方法是在不运行软件系统的情况下对需求规格进行检查和评估的过程。这包括审查需求文档、采用形式化方法进行验证,以及模型化和仿真等技术。静态验证主要侧重于发现需求规格中的问题和错误,以确保其准确性和一致性。相比之下,动态验证方法则是在运行软件系统的情况下对需求进行检查和评估的过程。这包括使用原型、模拟和测试等手段验证需求。动态验证方法主要侧重于测试软件系统的功能和性能,以验证系统是否按照需求规格正确地工作。

3. 需求审查

需求审查是针对确认和验证后的需求进行的进一步检查,其目的是确保需求文档的质量和完整性。通过仔细审查需求规格文档,团队可以识别潜在的问题、缺陷和改进点,从而确保最终交付的系统符合用户的期望和业务需求。需求审查通常由包括项目经理、开发人员、测试人员和业务分析师等角色的跨职能团队进行。在大数据应用软件开发中,需求审查

不仅关注需求的功能性,还需要考虑其在大数据环境下的可执行性。

需求审查的方法包括静态分析、评审会议和头脑风暴等。通过组织评审会议,团队成员针对需求文档提出意见和建议,形成更为全面和准确的需求描述。此外,还可以通过头脑风暴来激发团队的创造性思维,发现潜在的需求或改进。需求审查可具体分为以下几个步骤:

(1) 审查准备:在进行审查之前,需要做好充分的准备工作。这包括确保审查文档的完整性和准确性,安排审查会议的时间和地点,邀请利益相关者参与审查过程,并明确审查的目标和议程。

(2) 审查会议:审查通常以会议的形式进行,在会议上团队成员和利益相关者一起仔细审查需求规格文档。参与者应该积极参与讨论,提出问题、建议和改进意见,并确保所有的问题都得到充分讨论和解决。

(3) 问题识别和解决:在审查过程中,可能会发现各种各样的问题和疑虑,包括需求的不一致性、缺失或模糊等。团队需要及时记录并解决这些问题,确保需求规格文档的准确性和完整性。

(4) 审查记录和报告:审查过程中产生的问题、建议和决定应该记录下来,并形成审查报告。这个报告应该清晰地记录审查过程中发现的问题及其解决方案,以便后续跟踪和参考。

在大数据应用软件开发中,审查的重点还应包括对数据治理标准、隐私政策和性能需求等方面的审查。需求审查的充分性能够显著提升软件开发过程的透明度,确保团队对需求的一致理解,为后续的开发阶段奠定良好的基础。

通过以上有效的需求审查过程,团队可以确保需求规格的准确性和一致性,最大程度地满足用户的期望和需求,为后续的系统开发奠定良好的基础。

5.2.3 需求规格说明书的制定

在需求分析的过程中,提交需求规格说明书是重要环节之一。需求规格说明书是分析员在需求分析阶段需要完成的文档,是需求分析的最终结果。它为开发团队提供明确的指导,确保软件开发过程的准确性、一致性和有效性。在大数据应用软件开发中,由于涉及的数据规模庞大、复杂性高,需求规格说明书的撰写更加重要,需要遵循一定的规范和标准,以确保项目顺利进行并达到预期的效果。

在撰写需求规格说明书时,应该遵循一定的结构和格式,采用标准的模板,包括项目背景、项目目标、项目范围、项目进度和交付要求、需求确认和变更管理,以及附录部分,以便用户快速定位所需信息。

1. 项目背景

在需求规格说明书的项目背景部分,应该包括项目的背景信息、目的、范围和目标。例如,描述该大数据应用软件开发项目是为了解决某个行业的数据处理问题,提高数据处理效率和准确性。同时,说明项目的目标是实现特定的功能需求,提供优质的数据服务。

2. 项目目标

项目目标部分应该详细描述项目的具体目标,包括主要目标和次要目标。主要目标可能包括提高数据处理速度,提高数据处理准确性,降低成本等;次要目标可能包括提高用户

体验,提高系统的可维护性等。确保项目目标具体、明确,便于开发团队和相关人员理解和执行。

3. 项目范围

项目范围部分应该描述项目的范围,包括功能范围、技术范围、时间范围等。功能范围描述项目的功能模块和功能点,技术范围描述项目所涉及的技术和工具,时间范围描述项目的时间安排和进度计划。确保项目范围清晰、明确,避免出现范围蔓延和功能不清晰的情况。

4. 项目进度和交付要求

在需求规格说明书中,应明确项目的进度和交付要求,具体包括以下内容。

(1) 项目进度计划: 确保项目有清晰的进度计划,包括关键里程碑、任务分解和时间安排等。项目进度计划应与需求规格说明书相协调,确保项目按时完成。

(2) 交付要求: 明确项目的交付要求,包括交付成果、交付标准和交付时间等。确保项目交付符合客户要求,达到预期效果。

在项目进度和交付要求方面,需求规格说明书应包括项目进度计划表、交付要求表等内容,以便开发团队和相关人员了解项目的进度和交付要求。

5. 需求确认和变更管理

在需求规格说明书中,应设立需求确认和变更管理机制,具体包括以下内容。

(1) 需求确认流程: 明确需求确认的流程和步骤,包括需求提出、需求分析、需求评审和需求确认等环节。确保需求确认的过程清晰、规范。

(2) 变更管理机制: 建立变更管理机制,包括变更申请、变更评估、变更批准和变更实施等环节,确保变更管理的过程有序、可控。

在需求确认和变更管理方面,需求规格说明书应包括需求确认流程图、变更管理流程图等内容,以便开发团队和相关人员了解需求确认和变更管理的具体步骤和流程。

6. 附录

在需求规格说明书的附录部分,应包括相关的补充信息和文档,具体包括以下内容。

(1) 术语表: 列出项目中涉及的专业术语和缩写词的解释,以使用户理解项目中的专业术语。

(2) 参考文献: 列出项目中使用的参考文献和资料,包括相关标准、规范和文档,以便读者查阅和参考。

(3) 其他附录: 包括项目中的其他补充信息和文档,如需求分析报告、用户调研报告等,以便读者了解项目的详细信息。

需求规格说明书的附录部分应包含清晰的术语表、参考文献列表和其他相关附录,以便读者深入了解项目的相关信息。

图 5.2 展示了一个完整的需求规格说明书的目录结构,包括引言、项目介绍、业务功能需求、数据需求、模型需求、非功能需求、测试需求、项目进度和交付要求、需求确认和变更管理等核心部分,此外,还应包含清晰的术语表、缩略词表和其他相关附录,以便相关人员深入了解项目的相关信息。

1 引言	6 非功能需求
1.1 文档目的	6.1 性能需求
1.2 文档范围	6.2 可扩展性需求
2 项目介绍	6.3 可靠性需求
2.1 项目背景	6.4 安全性需求
2.2 项目目标	6.5 可操作性需求
2.3 项目范围	6.6 兼容性需求
3 业务功能需求	7 测试需求
3.1 用户角色和特征	7.1 功能测试需求
3.2 业务功能概念	7.2 数据测试需求
3.3 可视化需求	7.3 模型测试需求
4 数据需求	7.4 非功能测试需求
4.1 数据描述	8 项目进度和交付要求
4.2 数据源与格式	8.1 项目进度要求
4.3 数据采集	8.2 开发阶段要求
4.4 数据预处理	8.3 交付阶段要求
4.5 数据质量要求	9 需求确认和变更管理
4.6 数据处理流程	9.1 需求确认过程
5 模型需求	9.2 变更管理过程
5.1 模型应用场景	10 附录
5.2 算法选择	10.1 术语表
5.3 模型构建	10.2 缩略词表
5.4 模型效果	10.3 附加信息

图 5.2 需求规格说明书样例



5.3 需求分析常用工具

大数据应用软件开发中的需求分析环节在一定程度上与传统软件开发需求分析存在相似性,例如需求调研、需求分析、需求管理等基本活动和流程仍具有相似之处。因此,传统软件开发中常用的需求管理工具、UML 等可以进行适当的移植和应用。然而,由于大数据应用软件开发需求涉及海量数据规模、多源异构数据特性以及数据驱动的业务模式创新,其需求分析必须强调数据视角,并需要采用符合大数据特点的专业化需求工具。

需求引导工具(如思维导图、Excel 表格等)可以帮助从宏观视角理解数据应用场景的全貌,梳理复杂的业务过程,明确软件所需承担的核心功能;数据描述工具(如数据流图)则能够描绘海量数据集的来源、流向和处理等细节,便于准确定位数据需求。这两类大数据需求工具弥补了传统方法在大数据场景中的不足,支持对数据流转过程进行全面、细致的需求分析,确保需求描述的完整性与一致性,从而为大数据应用软件开发实现需求导向的模式提供有力支撑。

5.3.1 需求引导工具——思维导图

在大数据需求分析中,由于数据量庞大且结构复杂,需求分析人员常难以准确把握数据中隐含的需求,从而造成数据分析结果与实际需求之间存在一定的偏差。为了更好地引导大数据需求分析,提高分析的准确性和有效性,需求引导工具应运而生。此类工具可以通过直观的可视化方式呈现复杂的业务流程,进而支持从业务需求向系统功能需求的转化。与传统软件开发不同,大数据应用软件开发的需求分析需应对海量且异构的数据源带来的挑战,仅从用户的初始需求出发难以直接定义对数据的收集、存储、处理和分析等功能。

因此,大数据应用软件开发亟须引入需求引导工具,以实现业务需求与数据需求、模型需求的转换。需求引导工具通过直观的可视化方式,帮助明确在海量数据下需实现的价值目标,并建立业务需求与数据需求、模型需求之间的内在对应关系。例如,当业务部门提出用户行为预测分析的需求时,需求引导工具可以通过与数据采集、建模算法等需求的关联,指导将该业务需求转化为对日志数据收集、模型训练等功能的明确定义。

思维导图是一种图形化呈现思维结构和关系的典型需求引导工具,能够有效应对传统需求分析工具在大数据应用软件开发中因海量异构数据而导致的需求不清晰、难以精准定位等问题。在大数据应用软件开发的需求分析阶段,思维导图通过节点—连线—层次的网络结构,直观地反映复杂业务过程的整体情况,特别适合将业务需求转化为数据需求和模型需求。其可视化表征能够清晰展示大数据环境下多源异构数据与复杂业务的对应关系。例如,对于预测用户行为的业务需求,思维导图可以通过关联日志数据采集和模型训练等功能节点,将该业务需求转化为对日志数据收集和建模算法等具体技术的需求。

在使用思维导图进行需求转换时,应根据业务的重要性合理确定节点的层次,突出核心需求,避免过于关注非关键细节。同时,需要确保需求演变的方向指向技术实现路径,而不仅停留在业务概念层面。

5.3.2 数据描述工具——数据流图

数据描述工具是用于描绘和表达数据特征,是大数据应用软件开发概念需求分析过程中明确数据需求的关键工具。不同于传统软件开发,大数据系统需处理海量、异构、分布广的数据,这使得传统方式难以细致描绘数据特性。仅依赖传统需求分析工具,往往难以准确定位数据的规模和流转过程,导致数据需求分析不全面。而大数据应用软件开发中的数据描述工具采用定制化的图形化方式,能够支持对海量异构数据进行细粒度描绘,从而实现全面的数据需求分析。

数据流图是一种以图形网络的方式描绘数据从源到目标流过程的工具。图 5.3 展示了一个简单的数据流图示例,其通过可视化节点与连接线,直观地反映了大数据应用中的数据采集、数据预处理、模型预测到预测结果的过程。与传统的文档描述相比,数据流图以图形化的方式清晰展现各步骤的数据和处理环节,有助于减少需求分析中的遗漏。数据流图以直观的图形形式表达系统的业务逻辑,与系统的软硬件无关,也并不涉及系统的物理部署环境等,因而具有较高的应用层次。

数据流图分析的主要元素包括外部实体、数据流和数据存储等。外部实体表示系统的数据源或数据终点,反映了系统与外部实体之间的关系,并可表示系统的组成部分。数据处

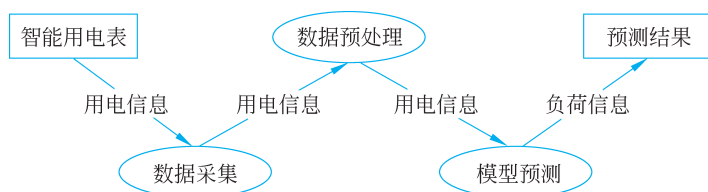


图 5.3 数据流图示例

理表示系统对数据的加工和处理,即如何满足用户需要的操作功能。数据流是指数据在系统中的传播过程,从数据源到终点的流动,明确了系统中数据的处理内容和流向。数据流可以使用矢量图描述,以便能够明确数据流流转的方向。

在大数据需求分析中,构建数据流图时需识别关键数据对象,例如原始数据、中间结果数据和终点数据等,同时明确数据流转过程中的提取、转换、加载等重要处理环节,并通过节点和线条清晰描述流转过程。在大数据应用软件开发需求分析阶段,绘制数据流图能够全面呈现海量异构数据的流转情况,帮助发现需求描述中的疏漏。此外,使用数据流图时应注重展现数据流转的连贯性,突出重点,避免图表过于复杂。综合运用数据流图进行概念需求分析,有助于弥补传统文档方式的不足,支持对大数据应用中的数据进行全面、细致的描述,提升需求分析的质量,从而使软件设计更好地满足数据处理需求。



5.4 案例分析

本节以“用电大数据负荷预测系统”为例,介绍实际项目需求分析案例。需求分析环节的最终目标是根据该软件的需求,形成项目的需求文档。用电大数据负荷预测系统开发过程如图 5.4 所示,需求分析为第一个环节。



图 5.4 用电大数据负荷预测系统开发过程——需求分析环节

本案例中采用了数据驱动的需求快速迭代范式,在第一次需求分析完成后,配合初步的数据探索性分析,识别出缺失、异常的数据和数据之间的相关性,然后根据数据探索性分析的结果,反馈至需求分析,对相关因素进行修正和补充。例如,如果数据探索性分析结果显示某些变量与负荷的相关性较低,则可以适当地调整需求分析中的重点变量。在调整后的基础上,再次进行深入的数据探索性分析,挖掘潜在的特征和规律,进一步识别影响负荷的深层次因素。这一过程可多次迭代,在每次分析中不断优化需求分析的结果,直至形成一个全面、系统、科学的需求分析说明书。由于篇幅限制,本案例中无法展示每次迭代的具体内容,仅介绍最终形成的需求分析内容,主要内容包括需求提出,需求分析,需求确认、验证和审查。

5.4.1 案例需求提出

1. 项目背景

本项目拟通过收集和整合多源异构数据,利用大数据和人工智能分析技术,结合大数据