浙 江 大 学

硕士学位论文开题报告

**（专业学位）**

论文题目： ××××××××××××××××

姓 名： 彭瑶

学 号： xxxxxxxx

专 业： xxxxxxxxx

院 别： xxxxxxxxx

导 师： xxxxxxxx

二零一六 年 九月

# 题目：数据可视化技术在APM中的应用

# 课题来源及类型

随着互联网、移动互联网和O2O的快速发展，企业的业务与IT系统的关系越来越紧密。企业应用已经从业务支撑系统变成了企业业务的核心，因此应用性能问题逐渐凸显。我们需要快速和高质量的开发程序、测试程序、发布程序，并且要做到持续集成、一键发布和无缝回滚。因此，应用性能管理（简称APM）的提出就是为了满足这种需求。APM（Application Performance Management）主要指对企业的关键业务应用进行监测、优化，提高企业应用的可靠性和质量，保证用户得到良好的服务，降低IT总拥有成本。

过去的APM重系统，轻交互，与如今的IT环境已经格格不入，没有面向用户，项目周期过长，运维人员维护成本高，易用性非常低。而现在新一代的APM所强调的是：1、端到端的性能管理可视化；2、面向真实的用户体验，实时的知道甚至预测何时出现问题，在用户还没有察觉之前，就由工程师来解决掉；3、从业务的角度看IT，通过APM分析数据，把业务信息、性能信息放在一起分析，进一步发现数据之间的关联。

但是新一代APM在大型IT系统中的实施，通常都面临着海量甚至巨量的数据处理与展示的挑战。怎样从最终用户体验出发，对海量的应用性能数据提供高效、直观和便利的展示，以及如何通过可视化技术来揭示数据背后隐藏的信息是一个很有意义的课题。基于此，本文主要结合APM海量应用性能数据处理和展示的需求，面向用户体验，设计与实现性能数据的可视化，以提高对应用系统的监控。

# 课题的意义及国内外现状分析

* 1. 课题研究的目的及意义

近几年，APM快速发展，各APM厂家都在主动地调研用户需求，有针对性的开发适合用户要求的功能和服务，研究如何更加贴近用户、贴心服务，从而推动用户需求，同时使APM解决方案也变的更加复杂与多样。

APM数据是天然的大数据，符合4V特征（Volume、Variety、Value、Velocity）。APM对数据有两点要求：1、数据处理要及时，必要时候要做到实时的处理，问题可能随时都会发生；2、数据的分析报告要精确，大量的数据本身是无价值的，按照业务模型进行精确分析、预测才有其价值体现。相对于其他监控数据，应用性能数据与业务的关联度更高，也更有深入分析的价值。如何将系统数据背后的意义讲述给企业的管理者、技术人员及运营人员，让公司基于数据进行有效决策，利用可视化的数据给用户创造更好的服务体验——这就是应用性能管理需要解决的重要问题。

一幅图胜过千言万语。人类从外界获得的信息约有80%以上来自于视觉系统,当大数据以直观的可视化的图形形式展示在分析者面前时,分析者往往能够一眼洞悉数据背后隐藏的信息并转化知识以及智慧。但是，仅靠一幅静态的可视化图像无法支持数据分析的动态过程,用户需要根据需求,与可视化界面中的图形元素进行交互式分析,来实现分析目标。在数据可视化中，最终呈现界面进行交互设计的这一阶段,用户的角色发生了转变,因为能够对界面进行操作,从接受数据变为控制数据,从被动接受信息变为主动探索、分析信息,这也是与信息分析、用户决策联系最紧密的步骤。数据的分析过程往往离不开机器和人的相互协作与优势互补。

* 1. APM的发展历程

从90年代末APM理念出现到今天的产品方案，APM受到技术、市场、产品的冲击与更新到现在，大致分为三个阶段：

第一阶段是以网络为中心，网络速度=应用速度。1996年Tivoli与HP公司开发了从应用程序层面出发的应用响应管理开发包（ARM API 1.0），随后的2.0版本被The Open Group批准为开放标准。1998年提出的面向商业的网络管理（BONM）概念，BONM被定义为以协助网络管理者测量和提高运行网络端到端的性能，它的功能包括监控和故障发现、带宽管理、数据分析和服务水平等级协议（SLA）等，APM这个理念正式作为这种软件技术的一部分提出。由此2003年IETF还专门为APM定义了管理信息库（SNMP-MIB）。

第二阶段是以IT部件／组件为中心，部件／组件健康监控，基础设施可用性监控。这个阶段伴随各种IT基础架构组件的发布，如：网络、系统、中间件、数据库。

第三阶段是以IT应用为中心，高度复杂，交易为核心，面向用户，面向应用生命周期管理。

经过上述这三个阶段近15年的发展，APM领域已经形成了较完善的相关管理标准与解决方案。

2.3数据可视化发展历程

普遍意义上的数据可视化被认为伴随统计学的诞生而出现。其实,用图形图像描绘、记录量化信息的思想,从人们开始观察这个世界进而产生测量、管理的需要的时候就己经出现了。

2.3.1、可视化思想的起源(15世纪——17世纪)

15世纪——17世纪是欧洲中世纪的晚期,这段时间可以被看做是可视化的起始阶段。经济、技术的发展、文艺复兴的到来使人们开始了解人文和科学知识,对地球的新认识则使许多的著名航海家浮出水面,新的国家与地区开始被载入人类史册。天文学、测量学、绘图学等等都快速起步以跟上对未知新世界的探索。三角测量技术、数学函数表相继出现了,人类也开始了对概率论和人口统计学的研宄。这个时期是数据可视化的早期探索阶段。

2.3.2、数据可视化的孕育时期(18世纪)

在此期间,在数学和物理知识成为了科学研宄的基础上,技术己经成为主力,社会管理的精确定量逐渐形成。伴随着早期的统计学的萌芽,社会和科技的进步体现在数据表现的多样化,已经出现了很多现在被广泛使用的图形形式,直方图、柱状图、饼图、圆环图等也己经出现。

2.3.3、数据图形的出现(19世纪前半叶)

在18世纪至丨9世纪前半叶这几十年间,统计图、地图和主题图等等这些如今依旧火热的数据可视化的表达的手法很多都开始被使用了,其中一个重要原因是很多公共领域的数据 始被政府部门重视,因而数据在这一时期极大地丰富起来一一例如关于人口、教育、犯罪、疾病等数据都被系统性的收集和发布,已经从科学技术和经济领域扩展到社会管理领域;另一方面,正在萌芽的计算机、通讯等提供了技术实现的可能性;书籍、报纸等媒体的出现和大量应用使印刷形式替代了手绘。重要的是,数据图形在这一时期在视觉表现上有了极大的进展,表达方式多样化了。现如今所常见的统计图形的样式及其他表现图等等都出现了。柱状图、饼图、地图等等集成为这个年代展示数据信息的一种常用方式。

2.3.4、第一个黄金时期(19世纪中、末期)

前面的所有发展似乎都是为这个黄金时期所做的铺塾,数据可视化迎来了它历史上的第一段辉煌。欧洲逐渐意识到信息据的作用,官方的统计机构也普遍的建立起来了,数理统计成为了一门新的学科,统计学的国际会议对可视化图形制定了分类和标准,各种图形、统计图表等都被广泛的应用和熟知了起来。

2.3.5、低潮期(20世纪前期)

20世纪前期,数理统计成为了数学的一个支派,统计学家们这个时期关注的主要是在准确的数学基础上扩展统计的疆域。数据的量和种类并没有太大的变化,于是黄金时期所出现的数据表示方式就己经够用,所以具有美观性和启发性的图形表达的研究就受到了冷落。

2.3.6、新的黄金时期(20世纪中末期至今)

现代电子计算机的诞生但来了强大的冲击,对数据可视化研究的再次兴起有了推波助澜的作用。计算机对数据分析的影响来自两方面——高分辨率的图形展现和交互式的图形分析一一都是手绘图形无法带来的革命性改变。同时,随着统计应用的发展,数据分析的应用扩展到了各行各业。当二者互相结合之后,就催生了统计计算工具、图形软件工具以及输入输出、显示技术等等。

综上所述，数据可视化的整个发展历程用表格汇总如下：

|  |  |
| --- | --- |
| 时间 | 数据可视化发展历程 |
| 15世纪-17世纪 | 可视化思想诞生初始，数据可视化的早期探索正式拉开序幕。 |
| 18世纪 | 数据可视化初步发展，直方图、柱状图、饼图、圆环图等开始出现。 |
| 19世纪前半叶 | 数据开始得到重视，数据图形出现。 |
| 19世纪后半叶 | 数据可视化第一个黄金时期，图形、图表等被广泛应用。 |
| 20世纪前期 | 前期的可视化表达方式已经够用，图形表达的研究并无新的进展。 |
| 20世纪中后期至今 | 数据可视化依附计算机科学与技术拥有了新的生命力，并在不久的未来大放异彩。 |

2.4国内外研究现状

在数据可视化领域，国内外研究者对数据可视化的方法做了大量研究，主要可以分为以下四个类别：

第 1 类研究从高层的用户目标出发,以用户意图为关注点,例如 North 等人提出的 Perceive, capture, encode, recover, and reuse, Pike 等人提出的 Explore, analyze, browse, assimilate, triage, assess, understand, compare。

第2类从较低层次的用户活动出发,以用户行为为关注点,例如Shneiderman提出的经典的Overview, zoom, filter, details-on-demand, relate, history, extract 这7类任务,Keim提出的动态投影、交互过滤、交互缩放、交互变形、关联与刷新这5类任务,Amar 等人提出的支持分析过程的 find extremum, sort, determine range, characterize distribution, find anomalies, cluster, correlate 等任务,此外还Wilkinson以及 Yi 等人提出的针对可视化元素进行操作的任务集合;。

第 3 类从系统的层次出发,以软件操作为关注点,例如 Chuah 等人提出了针对图形、集合、数据的操作模型,Ward 等人提出了交互基本操作、交互作用域、交互参数模型,Heer 等人提出了基于类的软件设计模式。

第 4 类对多层次任务进行整合,建立了多层任务模型,例如 Brehmer 等人从 Why,How,What 这 3 个方面建立了多层任务模型,并对各种信息可视化与分析任务进行归类;此外,Ren 等人提出了高层、底层、系统层之间的多层映射模型,Schulz 等人从 Why, How,

What, Where, When, Who 这 6 个维度建立了任务形式化描述模型。

# 3. 课题的研究目标、研究内容和拟解决的关键问题

## 3.1 课题研究目标

本文以Athena应用性能管理实际项目为例，从用户体验角度出发，对应用性能监控数据的可视化方法设计与实现进行阐述。

研究目标是：通过对用户需求的分析，结合性能监控数据的特点以及整个APM系统的体系结构和关键技术进行研究，分析和处理现有的大量、复杂和多维的应用性能监控数据，以直观的可视化的图形形式展示给用户，通过可视化图形呈现数据中隐含的信息和规律，并且提供交互能力，允许用户根据自身需求来选择数据的展示内容和展示方式，增强用户体验，通过对数据的可视化，用户能够实时获取系统运行状况，快速定位问题，预先发现性能问题，提高系统的运维开发效率，从而实际设计开发出一套基于用户体验的、 能在故障发生后快速定位问题域、从多个维度进行监控的、始终追求最直观的、最简洁的人机交互界面及实现方案的应用性能监控系统。

## 3.2课题研究内容

本文以Athena应用性能管理项目实践为背景，根据用户的需求，从功能上的角度将应用性能管理系统分为Application Administration, Application Configuration, Metrics analysis & reporting, Alert Workflow四大部分。针对应用性能管理海里监控数据的特点和基于人机交互数据可视化的发展现状，主要站在用户体验的角度，研究基于人机交互的APM的海量监控数据可视化技术。

本文研究的侧重点是主要包括以下两个方面：（1）可视性：选择适当的图形图表来展示数据并可对数据之间的存在的相互关系进行可视化分析。（2）交互性：通过交互的方式来管理和挖掘数据。

本文主要研究内容包括：

首先，明确应用性能管理系统在数据展示方面的需求，整理APM的发展历程和基于人机交互的数据可视化技术的发展历程，研究国内外的发展现状，通过分析与对比APM海里监控数据与其它数据之间的关系，总结出APM海里监控数据的特点，从而判断出应该选择哪些数据来展示以及怎样直观、有效地展示这些数据。

其次，根据系统的整体软件架构，完成各功能模块的设计与实现。将系统监控的数据（elapsed time, exception, garbage collection, heap detail, slow sql）进行可视化展示，提供一些有用的统计信息，同时提供交互能力，提高用户分析数据的能力，使用户通过分析数据之间的关联关系，即时发现问题并定位问题来源、预测未来可能发生的性能。

最后，总结当前几种经典的基于交互的数据可视化方法，针对APM海量监控数据，设计数据可视化的实现方式，通过实验，汇总该数据可视化方式的优缺点，从而提高APM系统中对海量监控数据可视化能力，揭示数据背后隐藏的意义，加速用户定位发现问题来源。

## 3.3 拟解决的关键问题

针对APM系统中海量监控数据的特点，在数据可视化过程中我们需要做到以下两点：（1）数据处理要及时，必要时候要做到实时的处理，因为问题可能随时都会发生；（2）数据的分析报告要精确，大量的数据本身是无价值的，按照业务模型进行精确分析、预测才有其价值体现。为了达到这两点的要求，在排除外界不可控因素情况下，本文需要解决的两个关键问题是：

1、实时的监控数据量多于庞大，在屏幕大小的限制下难以在一个屏幕中展示出所有的数据，可能会导致一部分数据的丢失、重叠、杂波、聚集等问题。

2、挖掘海量监控数据的价值所在，揭示数据背后隐藏的含义，从而快速定位问题域是APM系统的关键所在。如何从用户角度出发，利用交互能力，设计最直观的、最简洁的人机交互界面及实现方案，将不同时间段的数据进行不同层次方面的比较是待解决的另一大难题。

# 4. 课题的研究方法、设计及试验方案，可行性分析

## 4.1课题的研究方法

综合国内外大量研究者的研究成果,基于人机交互技术的数据可视化方法主要可概括为5类:动态过滤技术(dynamic queries)与动态过滤用户界面、整体+详细技术(overview + detail)与Overview + Detail 用户界面、平移+缩放技术(panning + zooming)与可缩放用户界面(ZUI)、焦点+上下文技术(focus + context)与 Focus + Context 用户界面、多视图关联协调技术(multiple coordinated views)与关联多视图用户界面。

4.1.1动态过滤方法

这种方法是在

4.1.2整体+详细方法

4.1.3平移+缩放方法

这种方法是通过平移和缩放的方式来实现，它在缩放的同时能够保持上下文和时间间隔。用户的选择创建一个缩放线图的层次，呈现为一个嵌套的树布局。最大的数据集作为线图的根节点，每一个高层级的节点作为一个新的子节点，堆放在父节点的下面。用颜色和箭头来说明子节点与父节点之间的关系。树结构方便追踪图交互的历史踪迹。

4.1.4焦点+上下文方法

焦点+上下文方法的起源是广义鱼眼视图的提出,它将用户关注的焦点对象(focus)与整体上下文环境(context)同时显示在一个视图内,通过关注度函数(degree of interest function,简称DOI Function)对视图中的对象进行选择性变形,突出焦点对象,而将周围环境上下文中的对象逐渐缩小，这样用户在探索局部信息的同时,就可以保持整体信息空间的可见性。

研究者针对焦点+上下文方法开展了大量研究,如 Spence 等人提出的双焦点变形技术(bifocal display)、Furnas 提出的鱼眼视图及其各种扩展技术、Lamping 等人提出的双曲几何变换技术(hyperbolic geometry)、Yee 等人提出的放射图(radial graph)、Heer 等人提出的关注度树(DOI tree)、Ren 等人提出的动态扇形图 DOI-Wave、Piertiga 等人提出的 Sigma 透镜等。其中,鱼眼视图的研究最为广泛,如文本鱼眼菜单 Fisheye Menus、搜索引擎结果鱼眼列表的 WaveLens、PDA 手持设备鱼眼日历 DateLens、图像鱼眼等。鱼眼视图也应用于密集网络节点的可视化,如密集树图多焦点Ballon 技术、大规模树结构的嵌套圆鱼眼视图等。大数据环境下，焦点+上下文方法因其能在突出关注的焦点的同时保持上下文整体视图的连贯性,将为密集型可视化界面和强调上下文关联的搜索分析行为提供有力的支持。同时,将焦点与上下文之间单纯的距离概念拓展到语义层面,结合挖掘与学习算法计算语义距离来动态获得与焦点语义相关的上下文,并做出智能自适应性可视化反馈,也将是焦点+上下文方法的研究重点。

4.1.5 多视图关联协调方法

## 4.2课题设计方案

本文主要从认知、可视化、人机交互的综合视角出发，根据监控数据的不同提供了不同的展示方式：

1、以观察和跟踪数据为主要目标，强调实时性、时效性的数据将其展示为时序图表

（1）将整个数据集呈现在时序图上

（2）通过用户的选择来查看详细信息

（3）查看选择区域上的aggregate数据（最大值，最小值等）

（4）将snapshot放在同一张图上，重叠起来，比较之间的关系

以Slow sql为例：在时序图表上，用户可以移动鼠标在某一时刻的位置，查看该时刻的基本信息：sql语句，执行时间和执行次数，点击该位置可以查看该时刻的具体信息，如该SQL语句的执行情况以及调用该SQL

2、以分析数据为主要目标，强调数据的统计信息，用可以检索、与用户交互式的图表——条形图或饼图

（1）将数据量相对小的那部分数据聚合成一类，避免各类数据差距悬殊，用户点击这一小块区域来查看这部分的统计信息

（2）用户移动鼠标来查看其基本信息语句的应用代码调用堆栈。可以导出snapshot，将其放在同一张图上进行对比。

## 4.3课题可行性分析

可行性研究分析是软件开发过程中最重要的一环，分析的目的是为了对现有问题进行总结，验证项目实施是否是可行的，确定能否在规定的时间完成该项目。

时间可行性分析：基于Athena项目团队的前期积累，目前已有一个正在运行的版本，整体架构设计已完成，具体需求明确，后台拥有大量的监控数据，有足够的时间用现有前端可视化技术进行应用性能监控数据的可视化展示与分析，时间上是可行的。

技术可行性分析：数据可视化技术发展起源早，本身拥有非常丰富的理论知识体系以及大量的研究成果。随着近年来，大数据技术的发展，使得数据可视化技术愈发成熟，目前拥有许多用于数据可视化的前端库，它们提供许多丰富图表样式，包括折线图、条形图、饼图、甘特图、各种不同的压力表图、漏斗图、热图，甚至世界、各州、各个国家的地图。同样地，大部分的图表包含2D及3D的视图，图表是完全可客制化的，标籤、字体、边界等等全部都是可以依使用者需求做相应改变。除此之外，他们也重视交互性，提示框、向下延伸资料、可点选的图例关键字、缩放及上下捲动，以及一次按键进行图表输出或列印。这些数据可视化的图表工具易于学习，容易修改并做延伸。

基于以上两点,我们认为本文的研究方法与设计方案切实可行,可操作性高,研究结果也具有很高的实用性,课题具有一定的研究价值。

# 5.课题计划进度和预期成果

## 5.1计划进度

|  |  |  |
| --- | --- | --- |
| 开始时间 | 结束时间 | 主要工作内容 |
| 2016年09月 | 2016年10月 | 查阅文献资料,编写课题开题报告 |
| 2016年10月 | 2016年11月 | 进行第一次问卷调查工作,明确系统需求,搭建系统架构、进行设计系统 |
| 2016年11月 | 2016年12月 | 适当调整系统设计,基本实现系统需求 |
| 2016年01月 | 2016年03月 | 撰写论文正文 |

## 5.2预期成果

期望产出：实现现有APM系统中的海量监控数据的展示方式，利用人机交互技术，让它能够在保有集成和展示能力的同时，能够更高效地处理海量、实时的数据流。