

## 图书情报学研究前沿：社会网络软件绘图

经济与管理学院 图书情报与档案管理 马昕远 202261113054

### 一、VOSviewer 绘图:

## 使用 web of science 核心合集

检索式: patent analysis (主题) and technology transfer (摘要)

查找主题是专利分析且摘要包括技术转移的英文文献；检索到从 2012 年到 2016 年的文章 405 篇，做词共现分析。通过筛选，去除没有意义的词，如：paper、person 等，剩下共计 105 个关键词来做词共现分析。这里的关键词是来源于文章的标题和摘要，可以在 VOSviewer 中设定。

(1) 聚类视图:

圆圈和标签组成一个元素，元素的大小取决于节点的度、连线的强度、被引量等，元素的颜色代表其所属的聚类，不同的聚类用不同的颜色表示，通过该视图可以查看每个单独的聚类。比如：词共现分析的聚类视图可以发现研究热点的关联分布，如图 1。

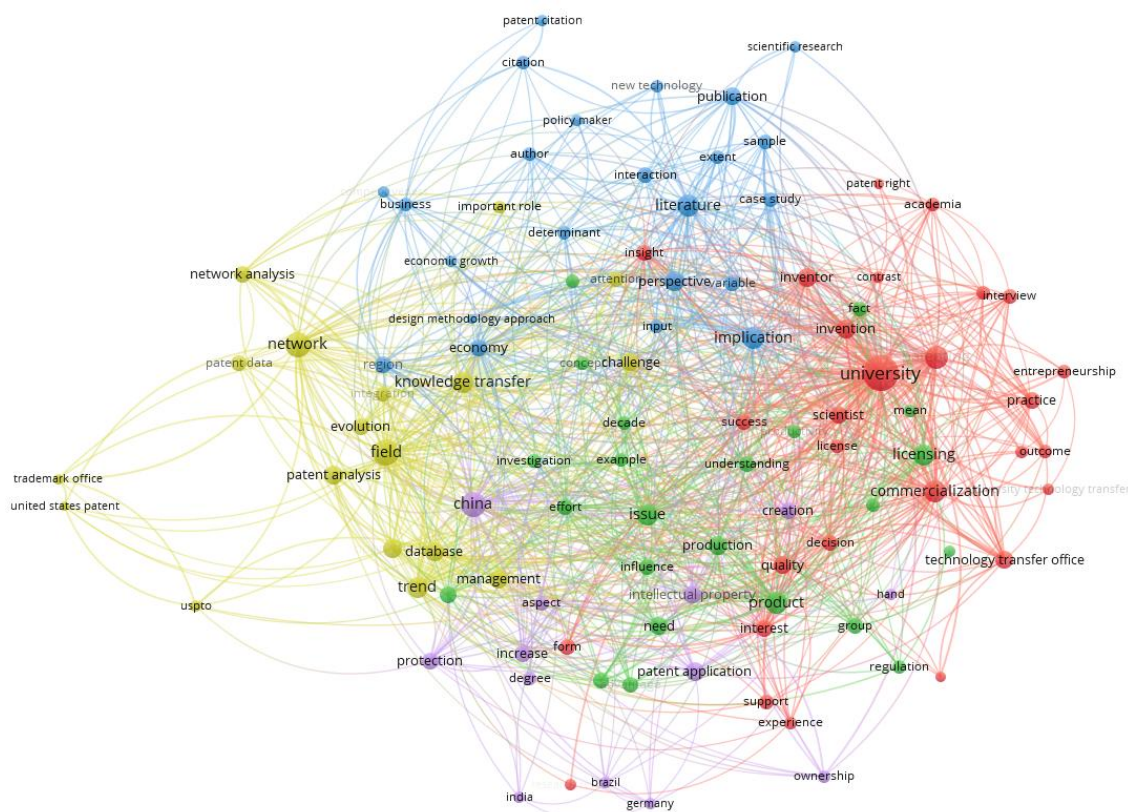
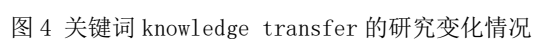
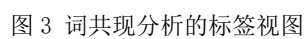


图 1 聚类视图







(3) 密度视图:

图谱上每一点都会根据该点周围元素的密度来填充颜色，密度越大，越接近红色；相反，密度越小，越接近蓝色。密度大小依赖于周围区域元素的数量以及这些元素的重要性。密度视图可用来快速观察重要领域以及某一领域知识及研究密度情况，如图 5 所示。



图 5 密度视图

总结：该软件还可以分析 Reforks、Scopus、PubMed、RIS、EndNote 等格式的文件数据，用来做关键词共现、作者共现分析。

## 二、Gephi 绘图:

查找与自然语言处理相关的文献，提取摘要并用 jieba 做分词处理，分析与自然语言处理相关的研究方向的关键词有哪些。为了图更容易观察，本实验仅选取 200 篇文献举例子，经筛选合并提取出 70 条与自然语言处理相联系的关键词，使用 Gephi 作图。

如图 6 所示，本次实验利用 Gephi 中邻接矩阵连入度和连出度的原理，除自然语言处理之外还发现人工智能、计算机视觉、深度学习等研究方向的度较大，也就是自然语言处理与人工智能、计算机视觉、深度学习等方向研究较为密切。但是数据量越大，说服力越强，图形也会变得越混乱，不易观察。

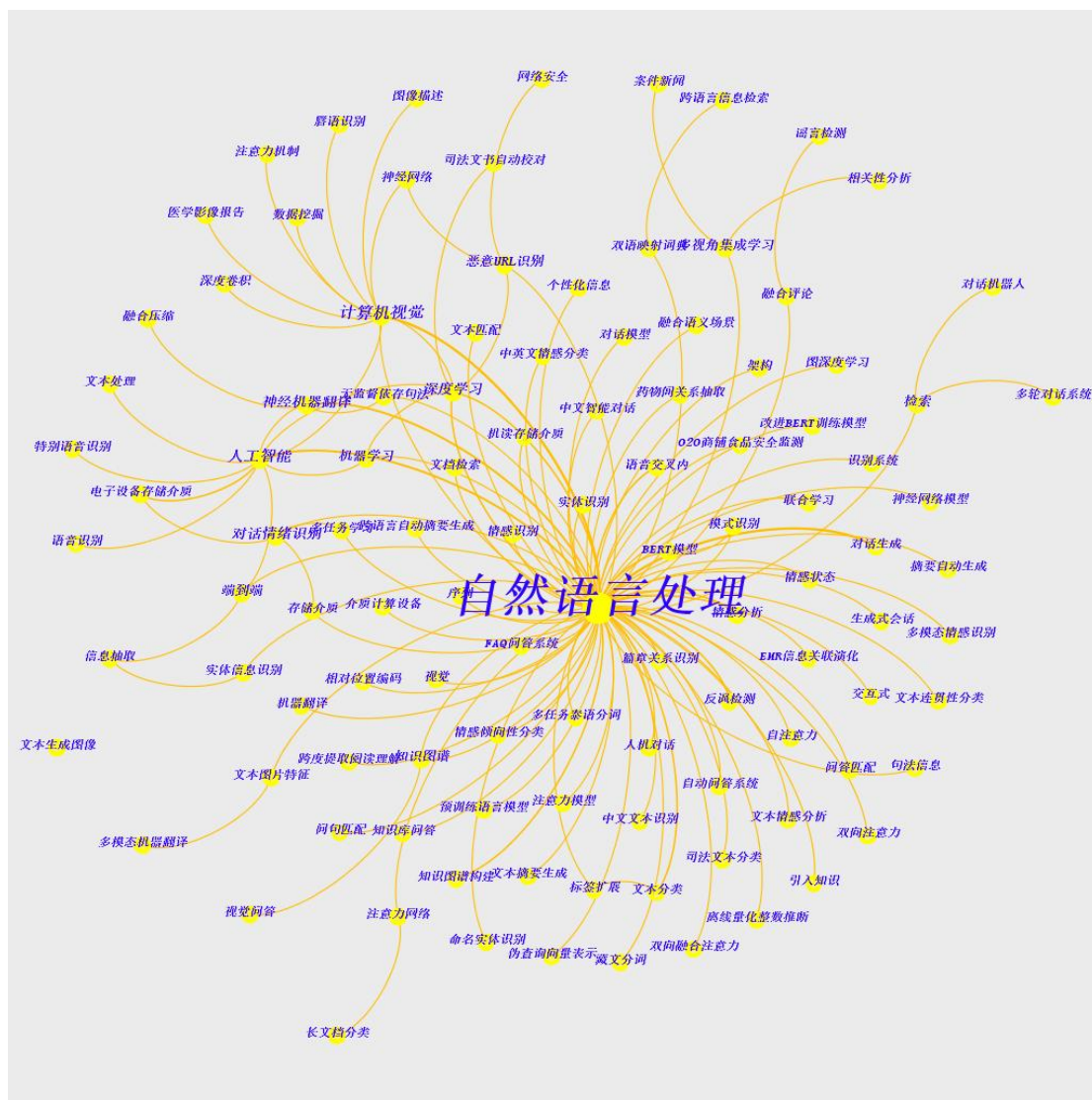


图 6 专利摘要分词分析

### 三、 形式概念分析(FCA)——Concept Explorer 绘图：

概念格，又叫做形式概念分析，它提供了一种支持数据分析的有效工具。概念格的每个节点是一个形式概念，由两部分组成：外延，即概念所覆盖的实例（对象）；内涵，即概念的描述（属性），该概念覆盖实例的共同特征。

每个概念格都是圆形，黑色下半圆代表该概念格上有至少有一个对象，蓝色上半圆代表该概念格至少有一个属性。

构建概念格的本质即分析对象拥有哪些属性：obj0:attribute0;attribute1,例子如图 7 所示，本实验选取人工智能领域文献 300 篇，给每篇文献按照软件要求取简单别名，经过分词、筛选、人工核验，选出关键词，分析与人工智能相关的上下位词。数据量越大图形越复杂，概念格越多，导致图看不清概念节点，因此本实验选择的数据总量较少，图会更容易看懂，但

普适性相对有限。本实验的概念格基本框架如图 7 所示。

分析上下位词如图 8，可以得知关键词“人工智能”是“自然语言处理”、“神经网络”、“图像处理”、“存储介质”、“数据处理”等关键词的上位词

可以展示每个概念格都有哪些对象，如图 10.

obj13:人工智能;对话系统;多轮情绪分析;对话系统  
obj37:人工智能;控制科学与工程;点云配准  
obj46:人工智能;蛋白降解;深度学习  
obj53:人工智能;视觉描述;卷积神经网络;视觉场景生成  
obj66:人工智能;影像分类;自适应医学影像分类  
obj68:人工智能;半结构文本;信息提取  
obj75:人工智能;肺部影像区域分割;医学  
obj120:人工智能;图像分类  
obj133:人工智能;预测有效动作  
obj170:改进;强化学习电熔镁炉工况识别;人工智能  
obj192:人工智能;自然语言处理;自对比学习;嵌套关系抽取  
obj201:人工智能;面部动作组合检测;情绪识别  
obj208:计算机应用;人工智能;预训练语言模型;语义解析  
obj222:人工智能;视频文本相似度确定;电子设备、存储介质  
obj247:人工智能;目标识别应用;机器视觉;果实检测产量估计

图 7 形式概念分析的格式

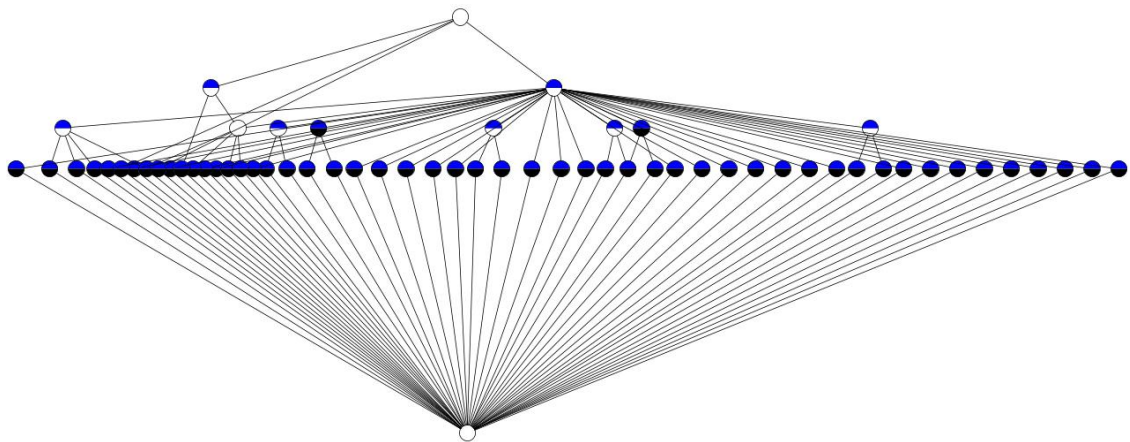


图 8 基本框架



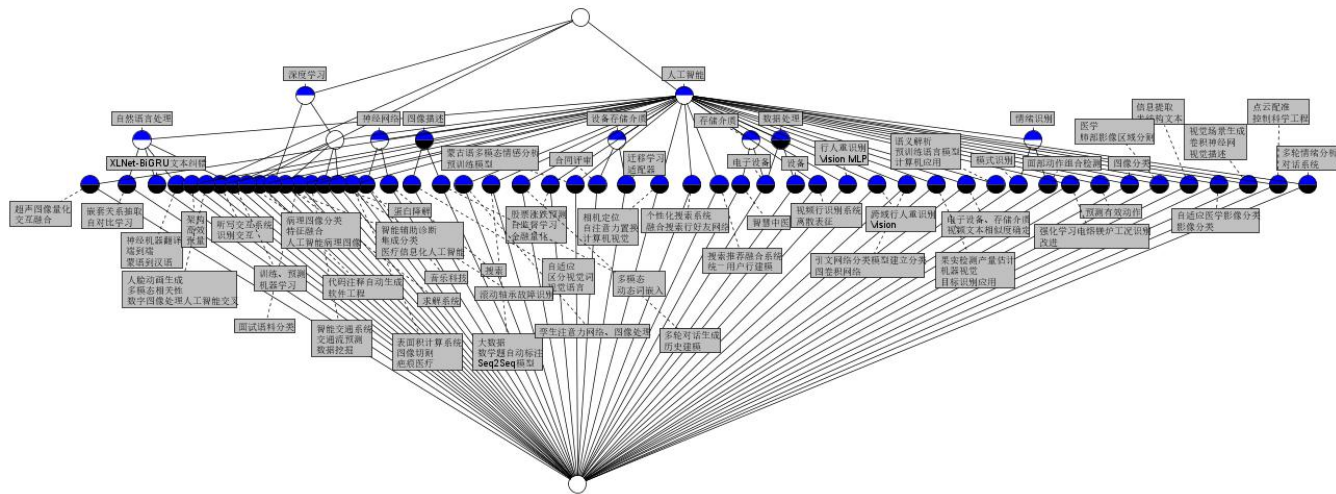


图 9 上下位词分析

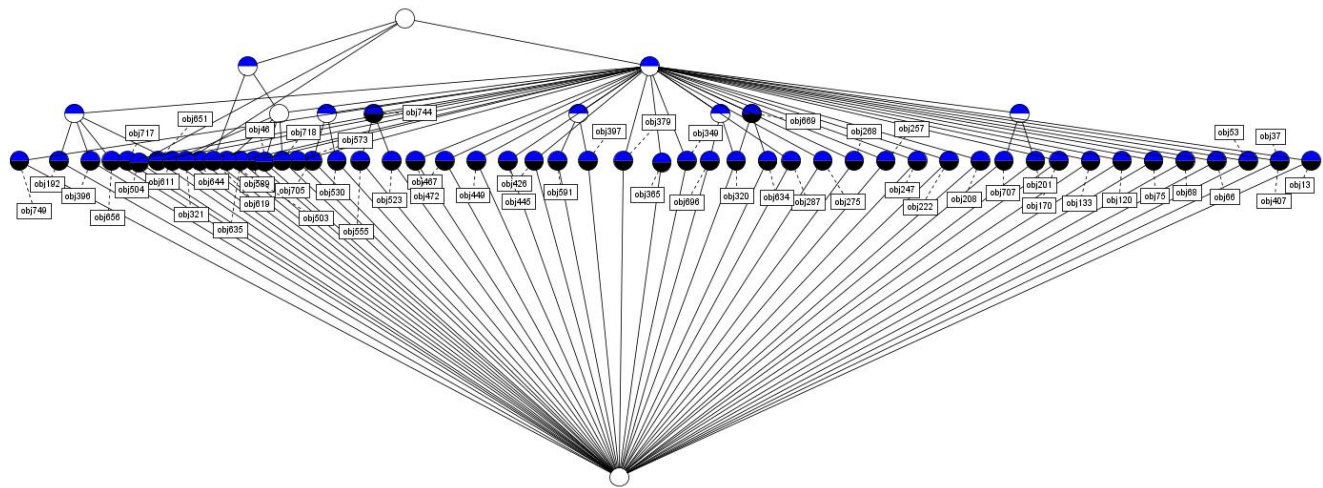


图 10 每个概念格拥有的对象

#### 四、 CiteSpace 绘图：

在中国知网以技术转移为主题、专利为关键词检索文献，时间限制 2016——2022，期刊选择 CSSCI，经筛选后保留 200 篇文章。

作者共现见图 11，机构共现见图 12，名字越大代表发表的文献越多；边的颜色越靠近蓝色代表发文时间距今越远。

CiteSpace, v. 6.1.R3 (64-bit) Basic  
November 22, 2022 at 6:00:49 PM CST  
CSID: C:\Users\LL\Desktop\CN\data  
Timespan: 2016-2022 (Slice Length=1)  
Selection Criteria: g-index (k=25), LRF=3.0, L/N=10, LBY=5, e=1.0  
Network: N=183, E=98 (Density=0.0059)  
Largest CC: 5 (2%)  
Nodes Labeled: 1.0%  
Pruning: Pathfinder

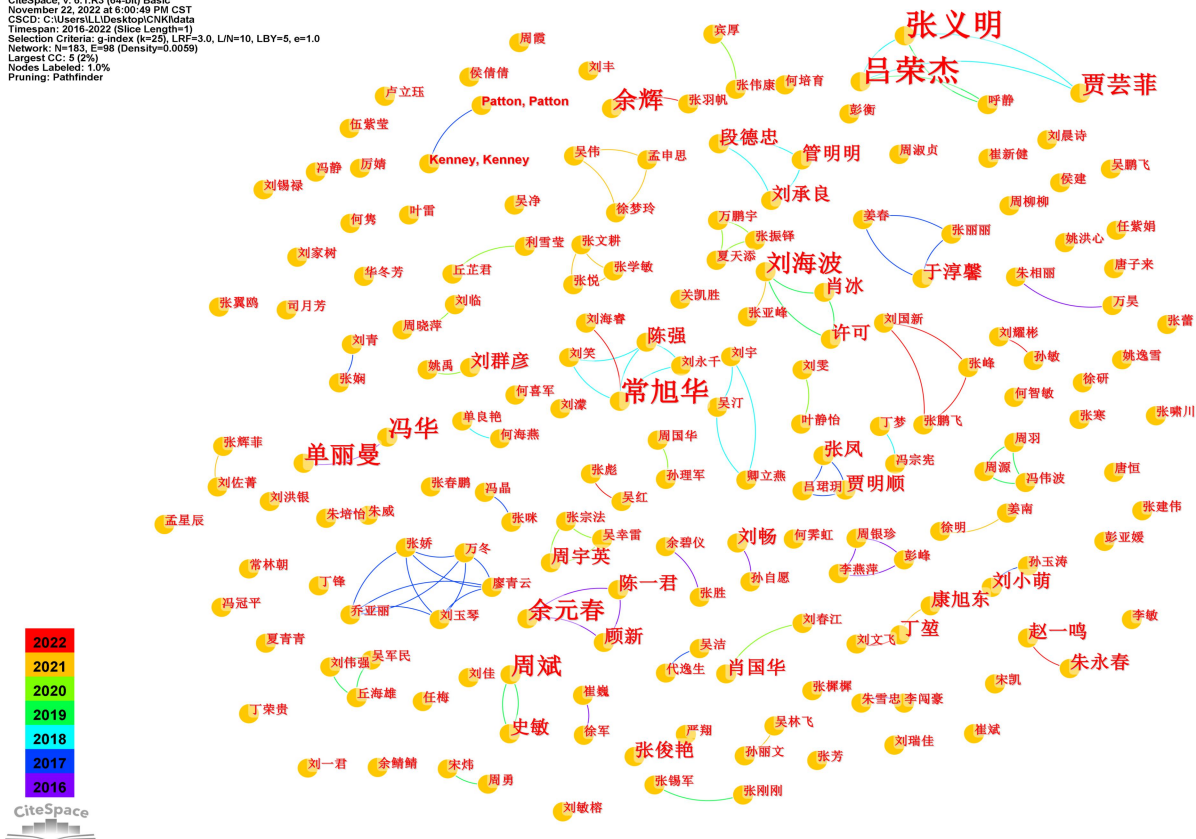


图 11 作者共现

CiteSpace, v. 6.1.R3 (64-bit) Basic  
November 22, 2022 at 6:31:13 PM CST  
CSID: C:\Users\LL\Desktop\CN\data  
Timespan: 2016-2022 (Slice Length=1)  
Selection Criteria: g-index (k=25), LRF=3.0, L/N=10, LBY=5, e=1.0  
Network: N=171, E=102 (Density=0.007)  
Largest CC: 14 (8%)  
Nodes Labeled: 1.0%  
Pruning: Pathfinder



图 12 机构共现



## 五、 其他软件绘图：

以下是在之前系统运维实施的工作中做需求分析、系统架构分析绘制的结构图，阐述内容清晰，可读性好。绘图软件是公司内部开发的，非公司 IP 无法使用。图能用 Visio 等应用复现。

以下每张图仅供本次作业观看，请勿外传！

图 13，监控大数据系统总体架构：这是我工作两年中投入心血最多的系统，该系统实现变电站设备监控数据的智能分析，系统总体框架如图所示。

图 14，标签体系架构：数据标签能够通过对数据信息分析得到的高度精炼的特征标识，方便机器或人对数据的识别能够提供统一业务化的标签给下游使用方，而不是直接提供原始数据表。

图 15，数据存储的内容：系统所需外部数据源的各类数据首先将会进入清理区，然后经过数据辨识、清洗、加工、转换后，根据数据类型及应用需求存储在关系型数据库、时序数据库、分布式文件系统，各类数据库存储的内容如图所示。

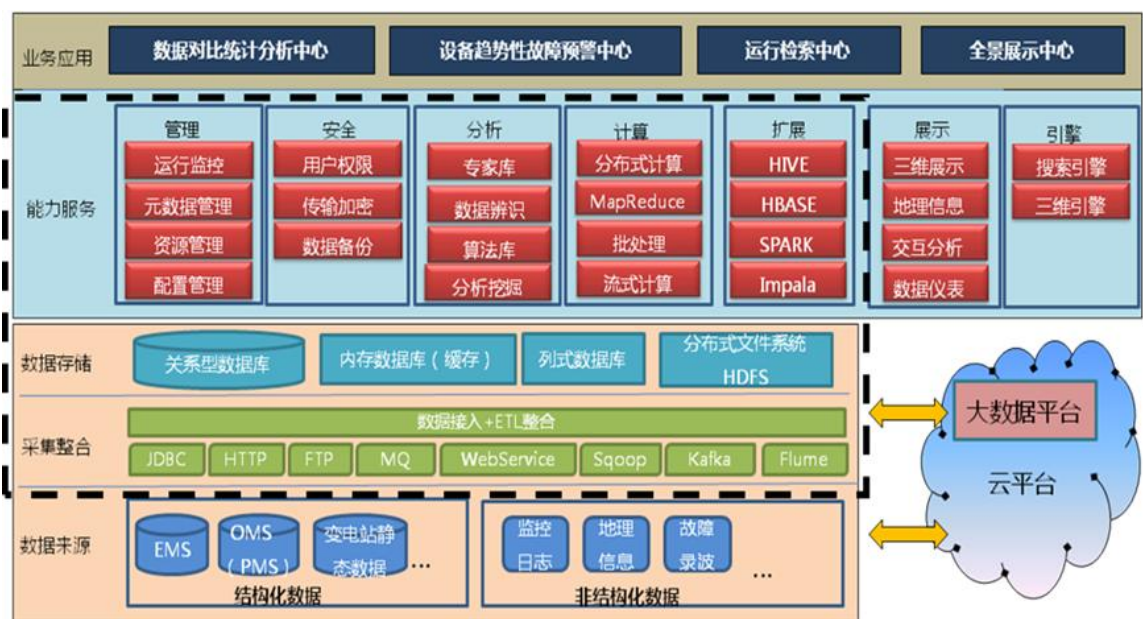


图 13 监控大数据系统总体架构

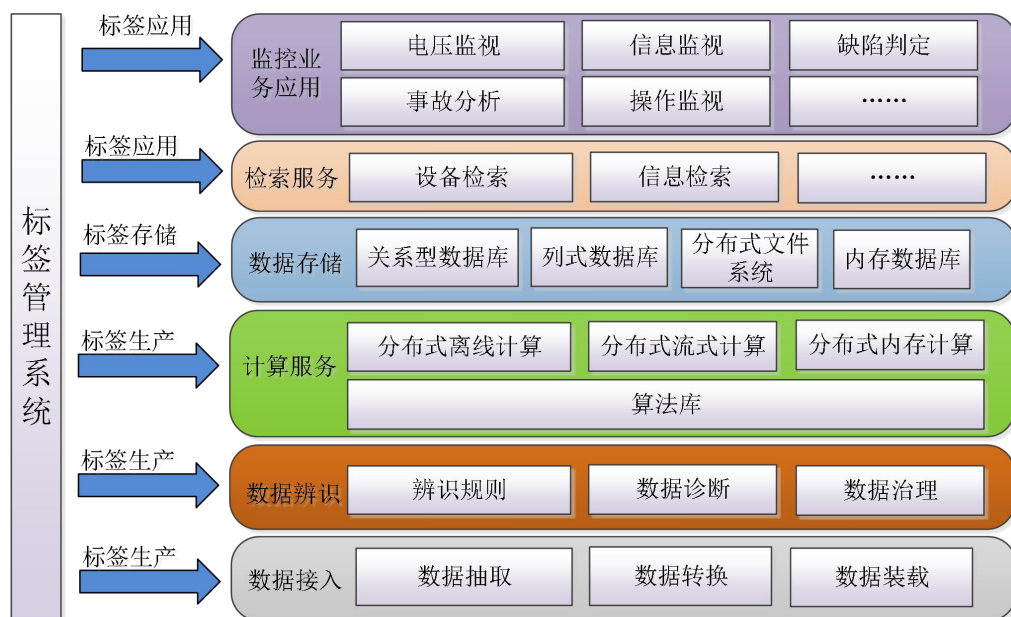


图 14 标签体系架构



图 15 各类数据的存储

一、二、三、四部分的图已上传 Github:

<https://github.com/NJTechMXY/Library-and-Information-Frontier>