



易联智能分布式图数据库 产品白皮书

中网数据（北京）股份有限公司
www.isinonet.com

目录

1	引言	1
1.1	数据存储和分析挖掘现状	1
1.2	图数据库的行业分析	2
1.2.1	图数据库发展现状	2
1.2.2	图数据库特点	2
2	产品概述	3
2.1	产品简介	3
2.2	适用范围	3
2.3	客户收益	3
2.4	产品架构	4
3	产品优势	6
4	产品功能	7
4.1	图数据服务	7
4.1.1	数据加载	7
4.1.2	数据检索	7
4.2	图深度关联分析	7
4.2.1	关联分析	8
4.2.2	路径分析	8
4.2.3	时间线分析	8
4.2.4	图片实体识别	9
4.2.5	文字实体识别	9
4.3	图数据可视化	10
4.3.1	图谱视图	10
4.3.2	图结构布局	11
4.4	安全策略	12
4.5	管理工具	12
5	应用案例	1
5.1	某军开源情报分析系统	1
5.2	某单位网络空间攻击线索发现紧急任务	2

1 引言

1.1 数据存储和分析挖掘现状

当前处于高速发展的互联网大数据时代,各种结构化和非结构化等多元异构数据大量出现,比如一些社交网络和电商数据每天都可能形成 TB 级别的海量数据,数据信息的关联越来越复杂,如何高效存储海量数据,如何获取数据之间的关系,如何才能真正利用数据,将数据进行融合分析,从而产生业务价值?

通常情况下会采用关系型数据库来管理和分析数据,但是面对这样巨大的数据量,传统关系型数据库往往难以很好地处理。我们的社交信息都是由关系组成,当海量数据处理涉及到关系数据时,传统的数据库存在建模难、性能低、查询难等问题,显的力不从心,它不能很好处理关系。

目前使用传统关系型数据库处理关系时主要存在以下问题:

- **数据存储方式单一,不灵活**

传统关系型数据库(例如 MySQL)以表为建模单元。通常每种实体都有一个表(个人、地点、事物、类别等)。在实际应用中,关系型数据库会有几十个甚至几百上千个表。用表记录数据之间的关联关系,数据存储方式单一且不灵活。

- **实时查询性能低,数据响应慢**

在传统的数据库中,每个表采取物理方式进行存储,建立两个表之间的关联较慢。随着不同表之间关系的数量和深度的增加,出现海量高并发读写需求时,实时关系查询的效率将急剧衰减,数据响应慢,甚至崩溃。

- **数据深度关联复杂,数据查询分析难**

关系型数据库按表格存储,查询时需要扫描行、过滤行,最后再进行数据聚合,它不能进行探索性或预测性的分析。比如:某些用户如何关联?如何分析社交网络好友的动态?传统数据库只能按一系列的表存储起来,但是把这些图关联起来却很难。

- **业务数据处理缓慢,不能满足敏捷处理需求**

某些情报分析、案件调查等业务需求的不断激增,需要数据支持的应用程序要随时满足业务需求,快速响应需求。传统数据库处理复杂关系慢,不能满足业务数据敏捷处理需求。

1.2 图数据库的行业分析

1.2.1 图数据库发展现状

图数据库善于处理大量的、复杂的、互联的、多变的网状数据，其效率远远高于传统的关系型数据库的百倍、千倍甚至万倍。

据 DB - Engines 的最新报告显示，图数据库是业界近 5 年来**排名增长最快**的类型。政府、军队、集团企业等大型组织均已开始使用图数据库作为传统数据库的补充，**实现对大规模关联数据的实时分析**。

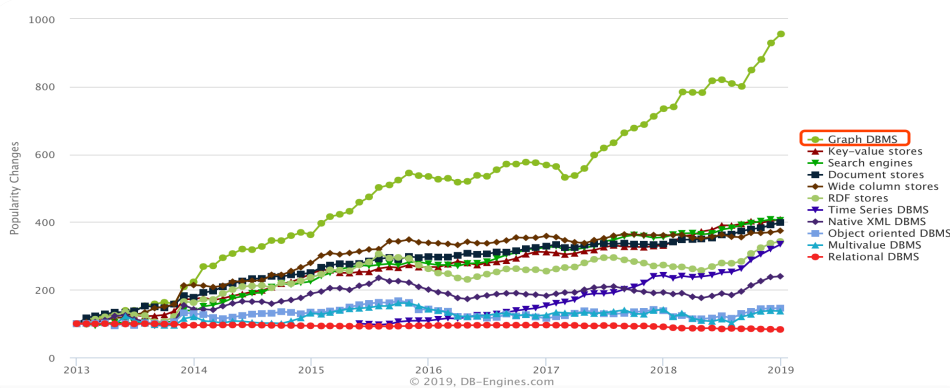


图 1 DB - Engines 数据库发展趋势图

1.2.2 图数据库特点

图数据库主要处理关系数据，能快速分析复杂关系的关联，比如社会网络中的人与人的关系。它使用图结构来实现语义查询，使用点、边和属性来表示和存储数据。它最重要的概念是边（关系），在存储中，它直接将数据项关联起来。利用图结构相关算法可进行深度分析挖掘，比如最短路径查找，N 度关系查找等，快速解决复杂的关系问题。

随着图数据库的普及，越来越多的图数据产品应运而生。目前市场上的图数据库产品拥有独特的功能和特点，但是对于超大规模图的处理仍然存在诸多限制，中网数据公司研发的 iFusionGraph 图数据库有独特的优势，可进行超大规模图存储和分析管理，满足复杂关系处理需求。

传统图数据库特点	iFusionGraph 图数据库特点
大规模图存储管理	超大规模图存储管理
简单无语义的异构数据存储	完整语义 的属性图模型
单一图存储模型	基于 语义 的异构数据存储
静态图谱	动态图谱 + 静态图谱
安全级别较低	单元级安全

2 产品概述

2.1 产品简介

易联 (iFusionGraph) 智能分布式图数据库是一种快速、可靠、易用、安全的下一代超大规模分布式图数据库，内置了可视化的人机结合探索式分析引擎。iFusionGraph 基于动态本体模型实现分布式多源数据智能管理和分析，为用户提供安全可靠的大数据融合存储、大数据关联分析和大数据可视化能力，并通过人机结合探索式分析引擎高效地实现数据关联探索、隐藏线索发掘，未知目标及异常行为发现等关键功能。iFusionGraph 基于分布式图存储和图分析引擎，及可线性扩展的图数据库集群，可支持百亿级边亚秒内数据实时查询响应，满足海量数据的存储和处理需求。

iFusionGraph 可以高效地映射、存储、遍历和分析海量的异构复杂关联数据，借助于简单易用的查询语言快速进行图的查询和分析，并通过可视化图谱多维度深度挖掘海量数据关联分析。产品目前广泛应用于开源情报、网络安全、金融反欺诈、物联网、实时个性化推荐、360 度画像、人工智能和机器学习等领域。

2.2 适用范围

iFusionGraph 作为图数据库产品，在目前大规模多元异构数据爆发的情况下，具有广泛的应用场景。只要是涉及数据融合和数据深度挖掘，都可以使用 iFusionGraph 进行数据关联分析。具体的使用场景如下：

- 政府开源情报决策分析
- 政务公安案件调查
- 网络威胁情报安全
- 企业金融保险反欺诈分析
- 政府互联网舆情监测

2.3 客户收益

- **构建全维知识图谱，助力政府对开源情报的决策分析**

将国内外社交、新闻媒体中的人物、机构、事件等相关情报数据信息，通过情报本体模型，动态生成情报知识库；通过深度关联分析，形成全维关系图谱，助力情报分析人员决策分析。

- **融合管理异构数据，协助案件线索调查**

通过案件本体模型将有关轨迹、档案等基础数据 ,以及其他外部收集的数据 ,如单位信息、银行记录、话单等多元化异构数据有机融合起来,围绕案件、重点人员和嫌疑人,对相关的人员、事件、组织、时间、空间等进行多维度分析,从而全方位支持复杂案件的信息,可以从多个侧面进行分析和对比,助力线索发现和嫌疑人甄别。

- **数据深度关联分析,实现网络威胁攻击事件追踪溯源**

分析实时流量信息,利用机器学习技术,识别网络攻击特征,构建网络安全知识图谱,基于图谱技术进行数据深度关联分析,实现攻击事件的追踪溯源、网络安全态势预警等,及时应对各种新兴攻击和威胁。

- **防止金融保险反欺诈,避免企业经济损失**

利用关系图谱分析技术,通过工商、征信等多维数据源综合评估,识别高危企业、个人、担保圈等对象,防范贷款违约风险;整合借款人的基本信息、借款人的消费记录、行为记录、关系信息、线上日志信息等,构建包含多数据源的统一反欺诈知识图谱,从而实现业务申请和交易阶段的反欺诈,避免企业出现经济损失。

- **数据实时查询响应,快速发现异常舆情信息**

将国内外社交媒体数据,通过社交舆情本体模型进行关联融合,建立舆情知识图谱,提供社交媒体互动效果分析、关联关系挖掘,传播路径分析、话题事件分析和传播效果评估等能力。同时,数据驱动的免开发模式,亚秒级查询与搜索响应,有效提升舆情数据业务开发和分析效率,快速发现异常舆情信息并及时预警。

2.4 产品架构

iFusionGraph 基于分布式图存储引擎、分布式图分析引擎两大引擎,以及图数据服务、图分析可视化界面,iManager 管理三大功能组件构成,提供灵活的定义本体模型,结构化和非结构化融合图存储、图检索和图分析,细粒度的数据安全服务等相关功能。

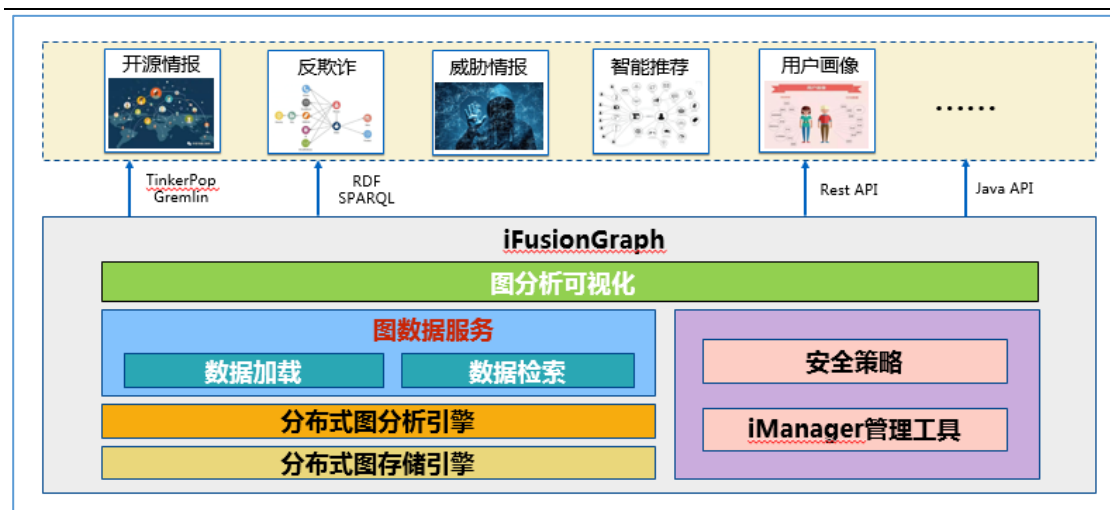


图 1 iFusionGraph 系统架构图

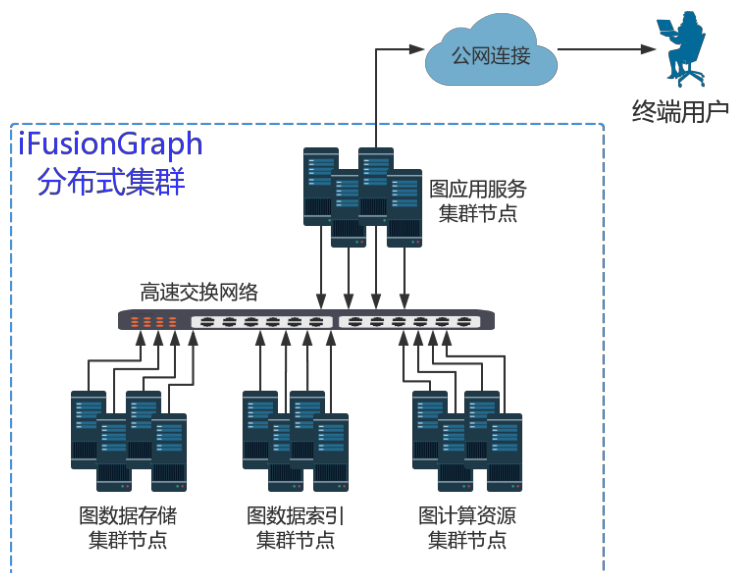


图 2 iFusionGraph 分布式集群示意图

分布式图分析引擎内置典型图算法，更好的支持任务的拆分、分发、调度和监控。

遍历和寻路算法	中心性算法	社区检测算法
并行广度优先搜索	PageRank	标签传播
并行深度优先搜索	度中心性	强连通组件
单源最短路径	亲密度中心性	弱连通组件
多源最短路径	中介中心性	Louvain 模块度
最小生成树	特征向量中心性	节点聚类系数

3 产品优势

相较于其他同类产品，iFusionGraph 具备高可靠、高性能、高安全、简捷易用等四大优势。

- **高性能**

部署简单，性能卓越，支持百亿级边亚秒内数据实时查询响应，可以在图上实时处理遍历多步（10 步以上）的路径探索，步数越多，分析洞察能力越深刻。它拥有更快的图数据加载和图更新速度，以协助构建和更新图形。高性能 iFusionGraph 可以提供更快速地实时深度关联分析。

- **高可靠**

iFusionGraph 分布式图数据库集群可线性扩展，可将数万亿个顶点和边的图存储在多机集群上，支持多用户并行操作。支持 PB 级海量数据存储和处理需求，有更好的计算性能和容错性。

系统稳定性大于 99.99%。支持多个数据副本容灾，当某一个数据副本发生故障时，其他数据副本可正常运行并对关键业务实现接管，达到热备份效果，保证系统的高可用性，使用户对故障零感知。

- **高安全**

拥有安全标签技术，从引擎侧提供了原生的最细粒度单元级访问控制，应用侧无需编程，且综合性能影响低于 10%。

数据的修订溯源技术可保留数据修改痕迹，支持精准的数据版本控制。

支持图查询和编辑的统一共享权限控制，使数据访问更安全。

- **简捷易用**

基于本体论扩展的 PropertyGraph 图模型，原生地支持语义网建模。

支持简单易用的图查询语言 Gremlin 和 SPARQL，提供多种定制化 API 接口，方便地与第三方应用集成，轻松实现基于图的各种查询分析运算。

具备可视化操作界面，丰富的图表直观展现多维度深度关联分析，产品易用性高。

4 产品功能

4.1 图数据服务

4.1.1 数据加载

iFusionGraph 支持 csv、json、rdf triples 等多种结构化数据和图片、文本、音视频等非结构化数据的加载。数据加载支持离线和实时两种方式，加载过程中支持数据校验和数据转换。

4.1.2 数据检索

iFusionGraph 能够根据图数据库中的顶点或边属性类型不同进行不同类型的检索，包括精确检索、范围检索、全文检索、模糊检索、地址检索、图遍历、数据过滤、限制结果数目、结果统计、结果聚合、结果排序和多重关系检索，并支持自定义数据转换和自定义数据过滤。

iFusionGraph 不但支持点和边的相互查找功能，还支持由点按条件筛选查找其关联的边、由点按条件筛选查找与其关联的点以及能查找两点之间的路径等高级功能。



图 2 数据检索

4.2 图深度关联分析

iFusionGraph 可进行图的渐进式探索分析，包括图的关联分析、路径分析、时间线分析、图片实体识别和文字实体识别等功能，为用户提供更深层的数据关联分析。

4.2.1 关联分析

iFusionGraph 图关联分析支持在图谱视图上探索分析任意数据对象的自网络信息，支持目标实体类型过滤、关系类型过滤、关联方向过滤。



图 2 关联分析

4.2.2 路径分析

iFusionGraph 路径分析支持对任意两个数据对象之间进行深层次的路径查找，并可视化到图谱中。

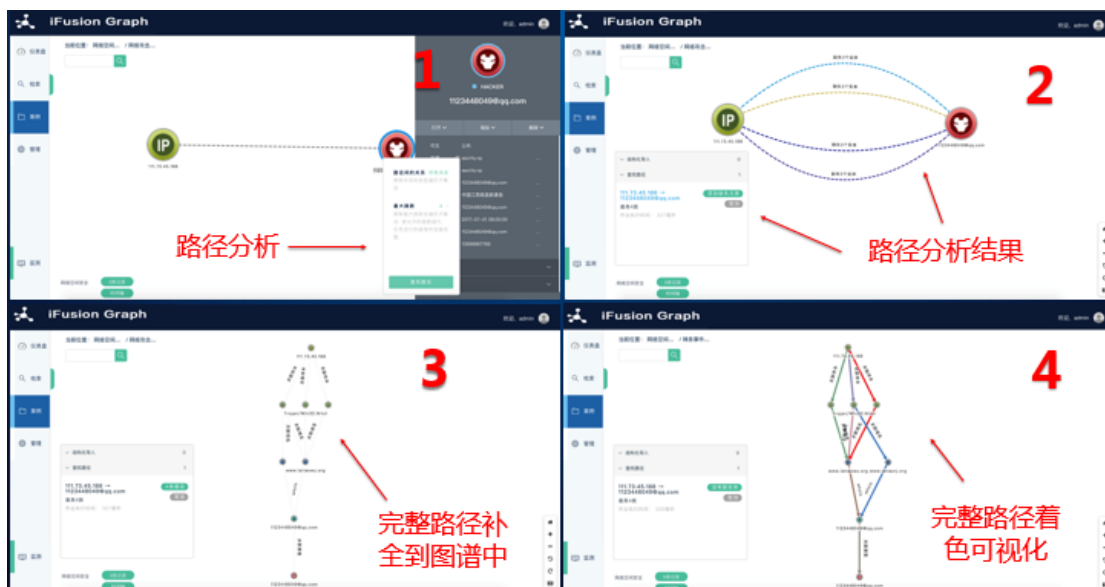


图 3 路径分析

4.2.3 时间线分析

iFusionGraph 时间线分析可自动提取视图中的所有时间数据，以统计直方

的方式展示出来，用户可以直观地发现边界信息，并允许以时间维度进行下钻分析。

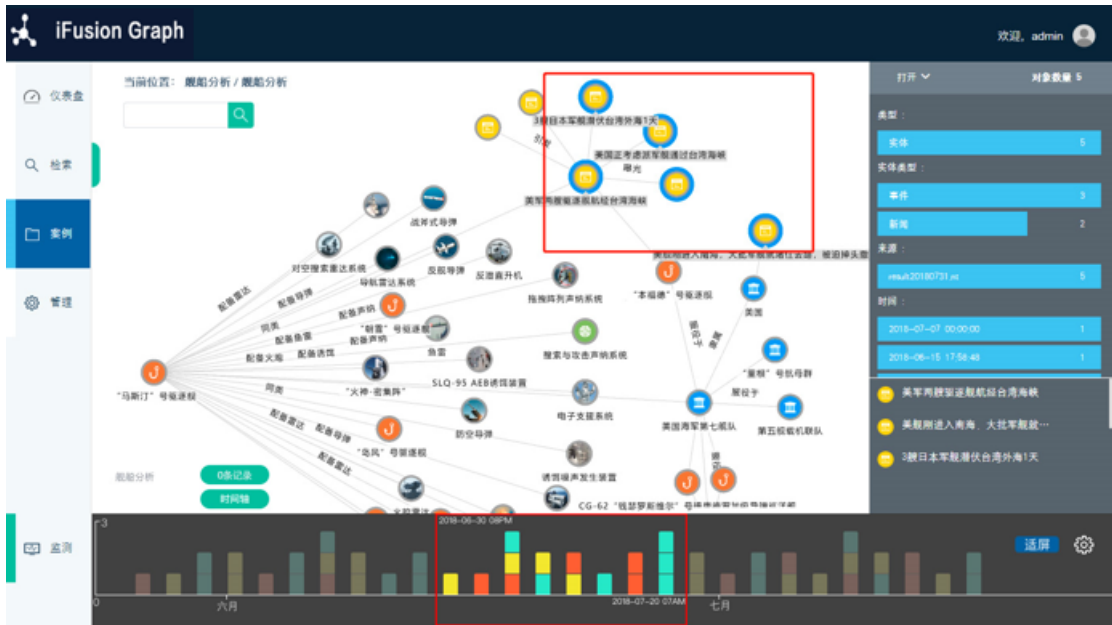


图 4 时间线分析

4.2.4 图片实体识别

iFusionGraph 支持图片的实体自动识别，将识别到的信息与图谱中的数据对象进行物理关联，并在视图上展示出来。



图 5 图片实体识别

4.2.5 文字实体识别

iFusionGraph 支持文本的实体识别，允许用户将识别到的信息与图谱中的数据对象进行物理关联，并在视图上展示出来。

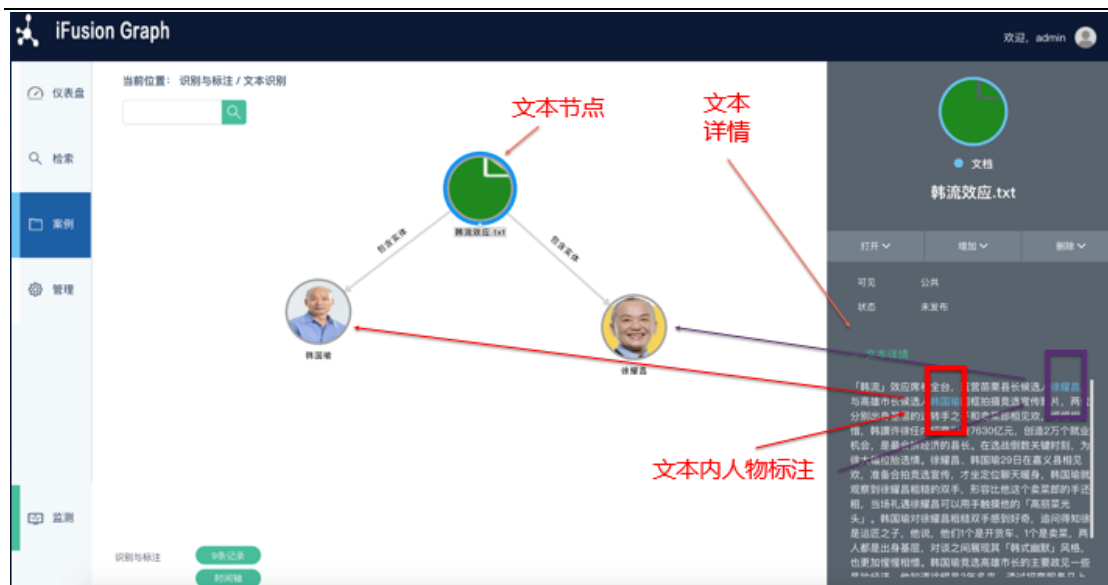


图 6 文字实体识别

4.3 图数据可视化

4.3.1 图谱视图

图谱视图是理解真实世界事物之间复杂联系的最理想的可视化方式。系统提供图谱视图、地图视图、仪表盘视图和统计视图，丰富的可视化图谱更直观展示数据关联分析状态，帮助用户发现异常特征。

- 地图视图：iFusionGraph 提取数据中的 GIS 地理位置信息，并以地图视图进行空间维度的展示和分析。

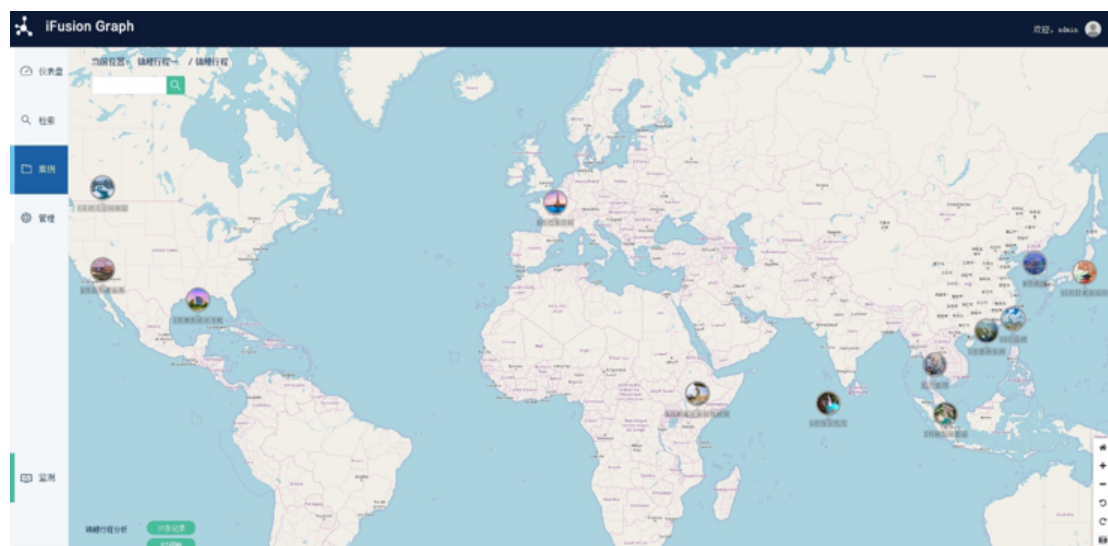


图 7 地图视图

- 仪表盘视图：iFusionGraph 提供了可定制的仪表盘视图，通过组合多种

视图将感兴趣的内容整合到一个页面中。



图 8 仪表盘视图

- 统计视图：选中图谱上的多个对象后，iFusionGraph 会自动生成属性值的统计视图，展示数值的分布情况，易于发现异常特征。



图 9 统计视图

4.3.2 图结构布局

iFusionGraph 提供多种图结构布局方式，通过同心圆布局、分级布局、力导向布局和网络布局方式，可从不同的拓扑角度查看数据间的关联关系，助力于复杂关系的探索挖掘。



图 10 图结构布局

4.4 安全策略

iFusionGraph 修订溯源技术可保留数据修改痕迹，支持精准的数据版本控制。同时支持图查询和编辑的统一共享权限控制，使数据访问更安全。

4.5 管理工具

iManager 管理工具用于图数据库集群系统状态、数据状态监控和图数据库服务启停管理。

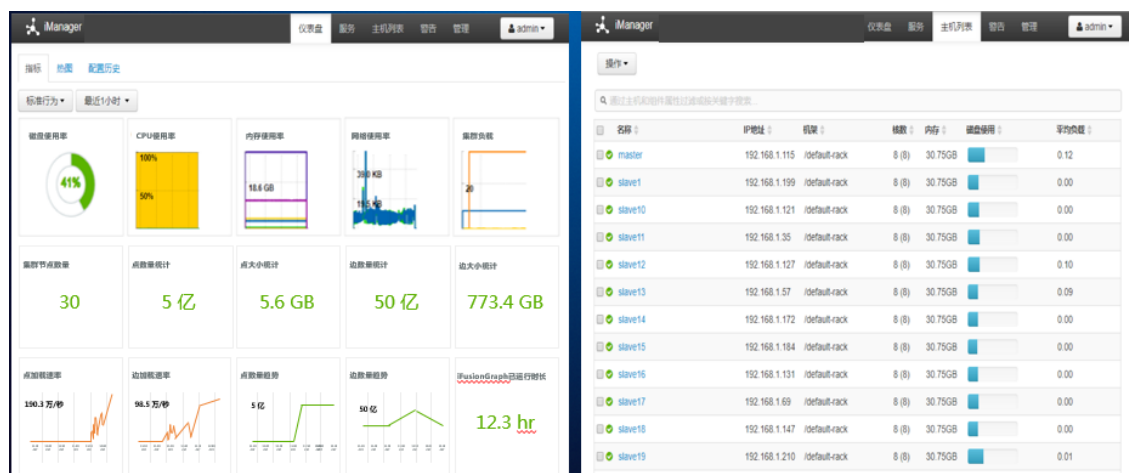


图 11 管理工具

5 应用案例

5.1 某军开源情报分析系统

中国沿海某军需要通过社交媒体及第三方线下情报对海外某方向主要人员做 360°画像，以便协助不同级别的情报分析师快速准确地在海量碎片化的信息中开展深度挖掘和关联预测，协同完成目标对象的全维度分析。

面对庞大的多源异构海量数据，人员关联复杂，分析维度多，数据权限复杂的情况，iFusionGraph 上线后，为 1000 多个社交媒体网站、线下情报、目标人物等数据快速构建了统一的图谱模型。

系统存储了 **1.2PB** 的目标方向的主流新闻网站、论坛、博客等数据，总计 200 亿以上各类数据对象和 800 亿以上的各类关系对象。在**千亿级**的超大图上提供了**秒级**的精确检索、模糊检索、全文检索和地理检索。在**千亿级**的超大图上提供了**分钟级**的路径查找、社群分组等经典图算法，**30** 分钟内刷新态势信息。拥有的**安全标签技术**，允许情报分析专家和普通业务人员看到的不同的情报内容。

针对安全情报领域的多样化数据进行挖掘，将大量孤立、异构的数据进行高效的连接和处理，并转化为高价值情报。平台快速智能的关联关系挖掘，绘制出反映真实世界的社会关系网络图，以灵活交互功能的可视化分析，多维度深层次的探索式分析，使情报分析人员在短时间内找到关键线索。

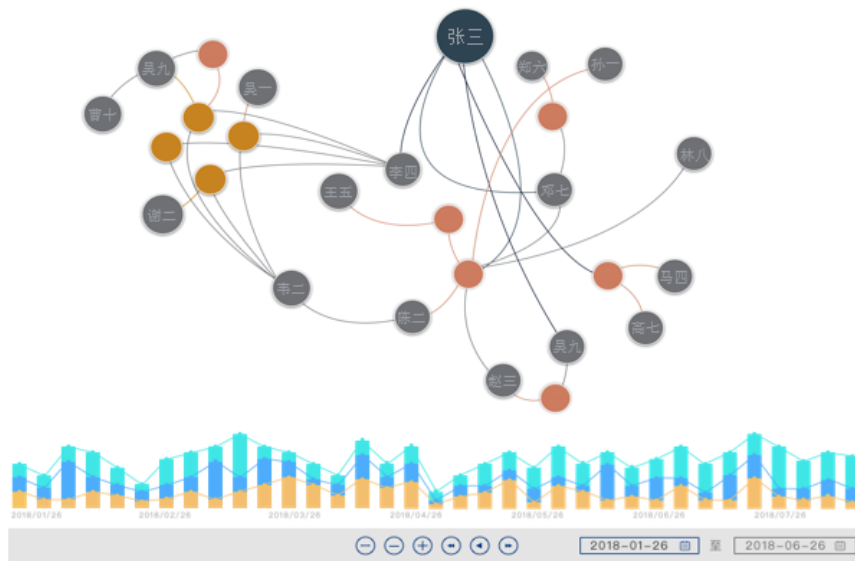


图 1 情报分析系统

5.2 某单位网络空间攻击线索发现紧急任务

某单位收集了 800 万个网络攻击行为。要求 3 天内从攻击行为数据和自有骇客信息库中分析出 XCodeGhost 事件的传播路径和幕后黑手。项目时间短，任务急，数据关系复杂，事件溯源路径长。

基于 iFusionGraph，快速构建了符合数据情况的图谱模型。总计 28 个实体类型，18 个关系类型，97 个属性类型。40 分钟内完成了 820 万实体和 4 千万关系的加载；3 小时内完成了整个事件的追踪，找到了可疑骇客；1 天内提前完成了任务，节省了 66%的时间。

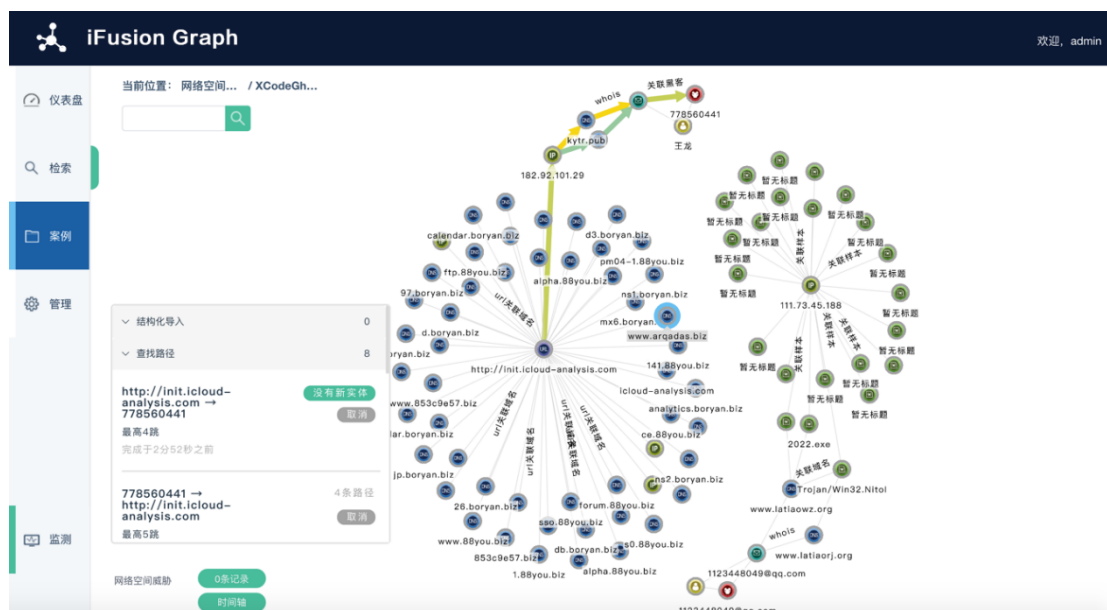


图 2 网络攻击分析系统

地址：北京市海淀区丰秀中路 3 号院 7 号楼中网数据大楼 100094

电话：+86 (10) 58251700 传真：+86 (10) 58251701

E-Mail：contact@isinonet.com

资料版本号：V4.1