

# EYWA: Building a Distributed Graph Engine on Huawei Cloud

Yinglong Xia  
Huawei Cloud America  
09/01/2017



# About Huawei



**170,000**  
employees

**170**  
countries

**\$75B**  
revenue

**14**  
regional HQ

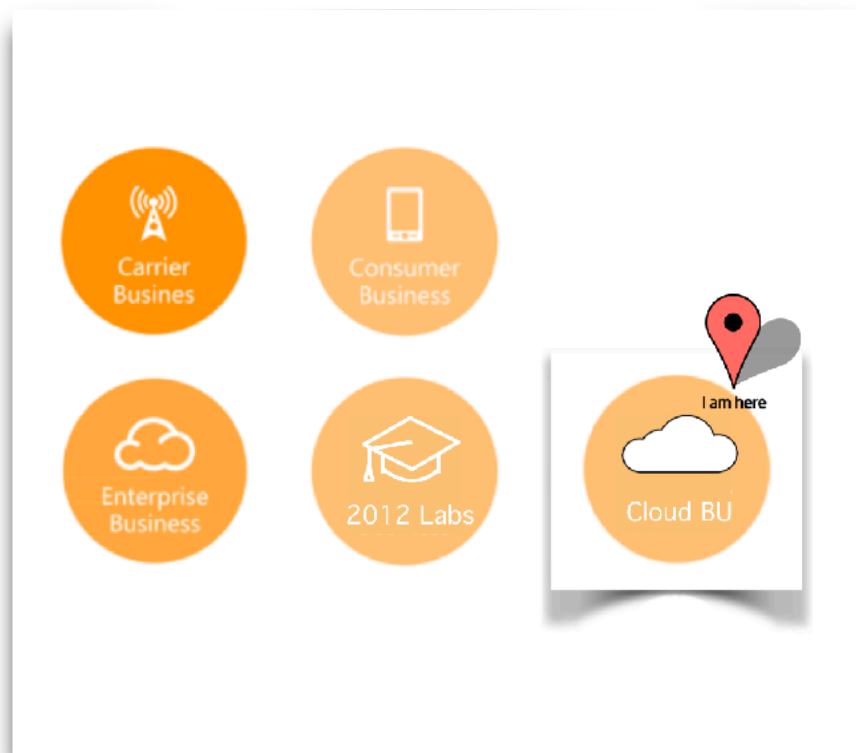
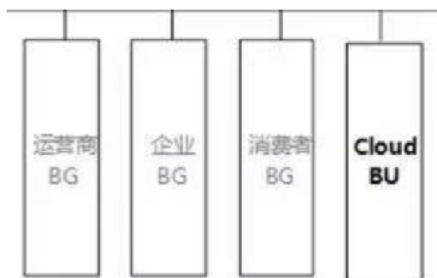
**16**  
R&D centers

**32%**  
YoY growth



# About Huawei - 2

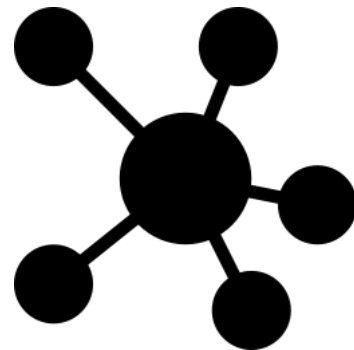
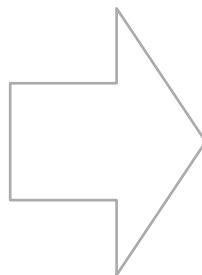
8/28/2017



# Eywa



The Na'vi believe that Eywa acts to keep the ecosystem of Pandora in perfect equilibrium. Eywa emphasizes connection, the connections among the creatures on Pandora and the connections between Na'vi and their ancestors.



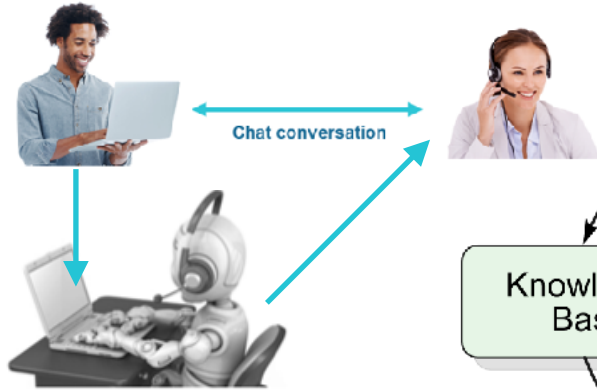
Edge-set sYnergistic Workload Architecture (EYWA) as the in-house graph engine techniques in Huawei

**Connection is the essence of graph**

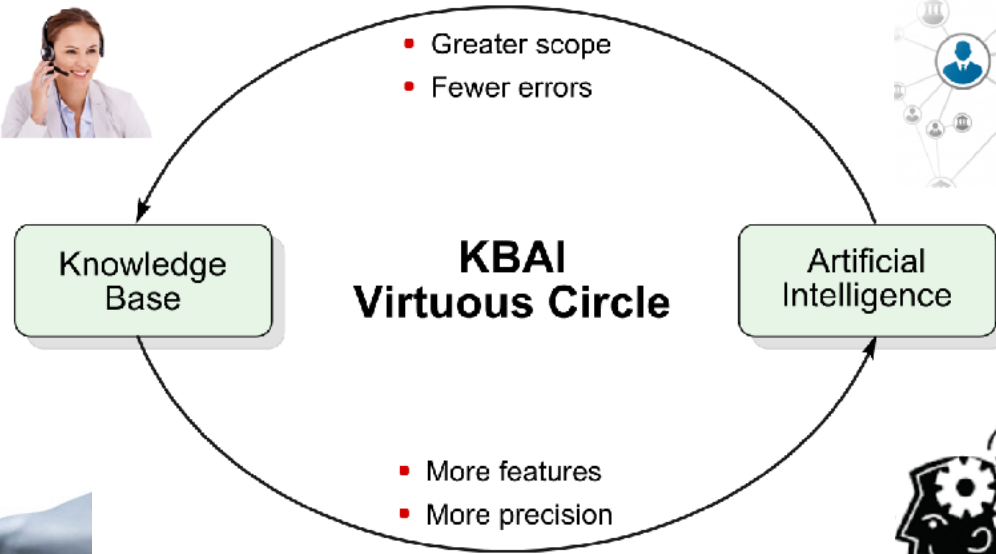
# Graphs in AI: Intelligent Service

Use Case:

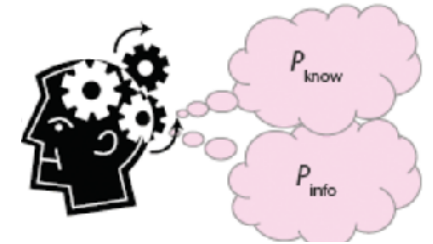
AI assists Technical Services



Human-Machine Collaboration



Private Knowledge Graph



Inference & Reasoning

# Eywa Graph Engine Service

The screenshot shows the Huawei Cloud product page at [www.hwclouds.com/en-us/product/](http://www.hwclouds.com/en-us/product/). The navigation bar includes 'Products', 'Solutions', 'Cooperation and Ecosystem', and 'Support'. The main content area is a grid of service categories. The 'Data Analysis' category is circled in red and contains the following services:

- Data Ingestion Service
- MapReduce Service
- Data Warehouse Service
- Data Pipeline Service
- Machine Learning Service

Other visible categories and services include:

- Compute:** Elastic Cloud Server, Cloud Server Backup Service, Cloud Container Engine, Bare Metal Server, Auto Scaling, Image Management Service, Dedicated Cloud.
- Storage:** Elastic Volume Service, Volume Backup Service, Object Storage Service, Data Express Service, Dedicated Enterprise Storage Service.
- Network:** VPC, Elastic Load Balance, Direct Connect, Virtual Private Network, Domain Name Service.
- Security:** Anti-DDoS, Key Management Service, SSL Certificate Service, Security Index Service, Security Situation Awareness, Application Recognition Service, Security Partners Product Center.
- Management & Deployment:** Cloud Eye, Identity and Access Management, Cloud Trace Service, Integrated Cloud Migration Service, Cloud Catalog Service, Cloud Report Service.
- Application:** ServiceStage, Distributed Message Service, Simple Message Notification, EC Platform Cloud, Distributed Database Middleware, Image Management Service.
- Database:** Relational Database Service, Distributed Cache Service, Document Database Service.
- DevCloud:** ProjectMan, CodeHub, CloudPipeline, CodeCheck, CodeCI, Deployman, TestMan, ReleaseMan.
- Enterprise Applications:** Workspace.

# From Product to Techniques

High usability through clicks



Import graph data

Display result in JSON and visualize it

Usability - Easy to use

Usability - Flexible to use

Perform user queries through query language



legend

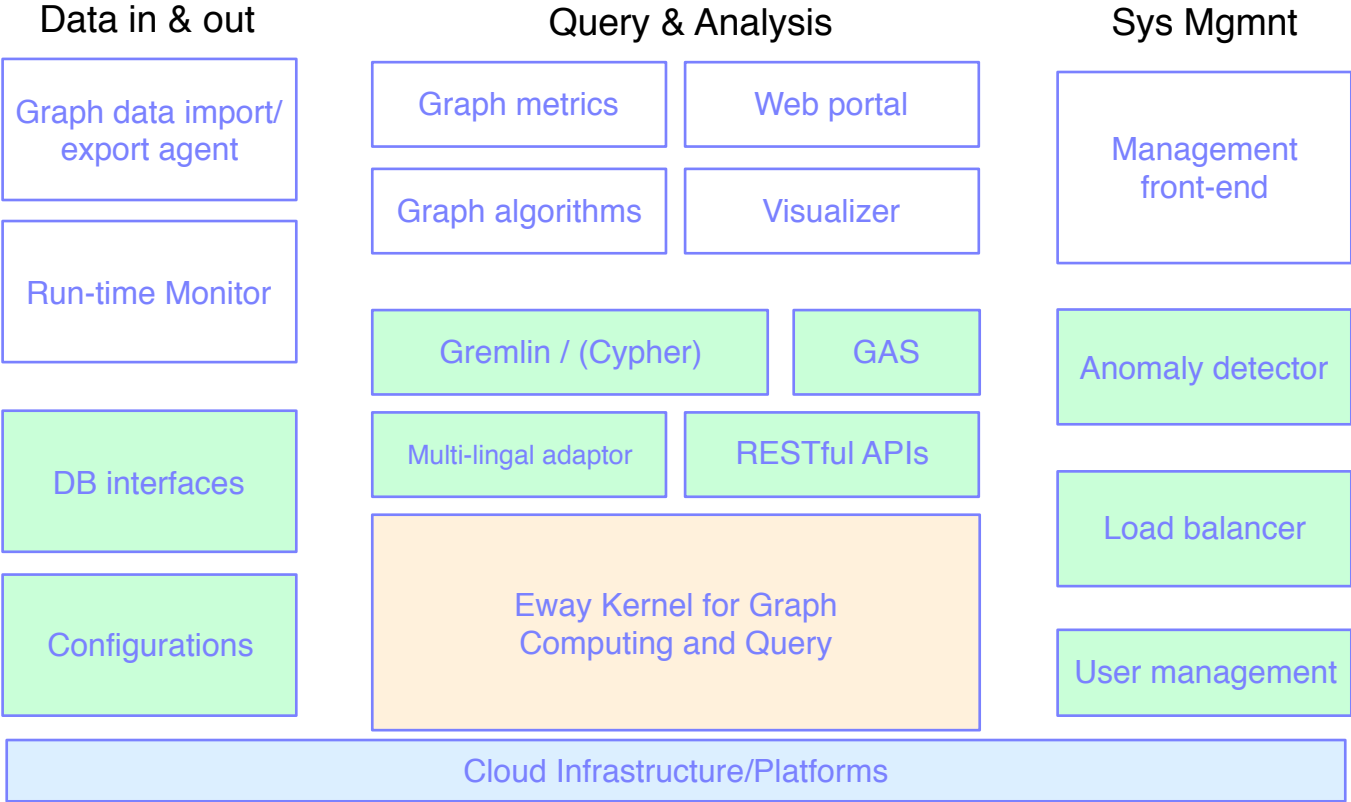
End User

Developers

Kernel

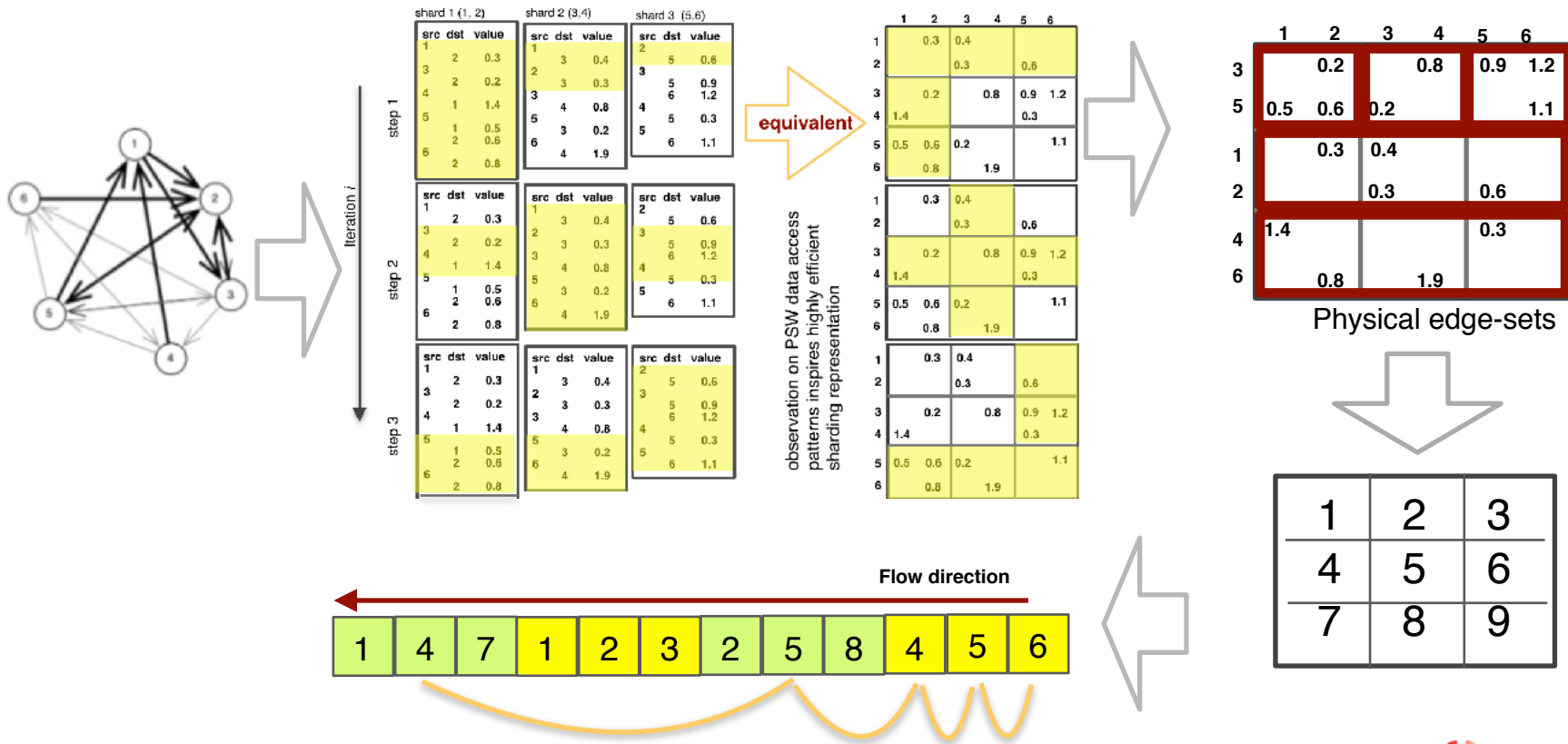
Platform

# Eywa on Cloud - the Architecture





# Trade-off Between Analytics and Query



# Labeled Property Graph Model

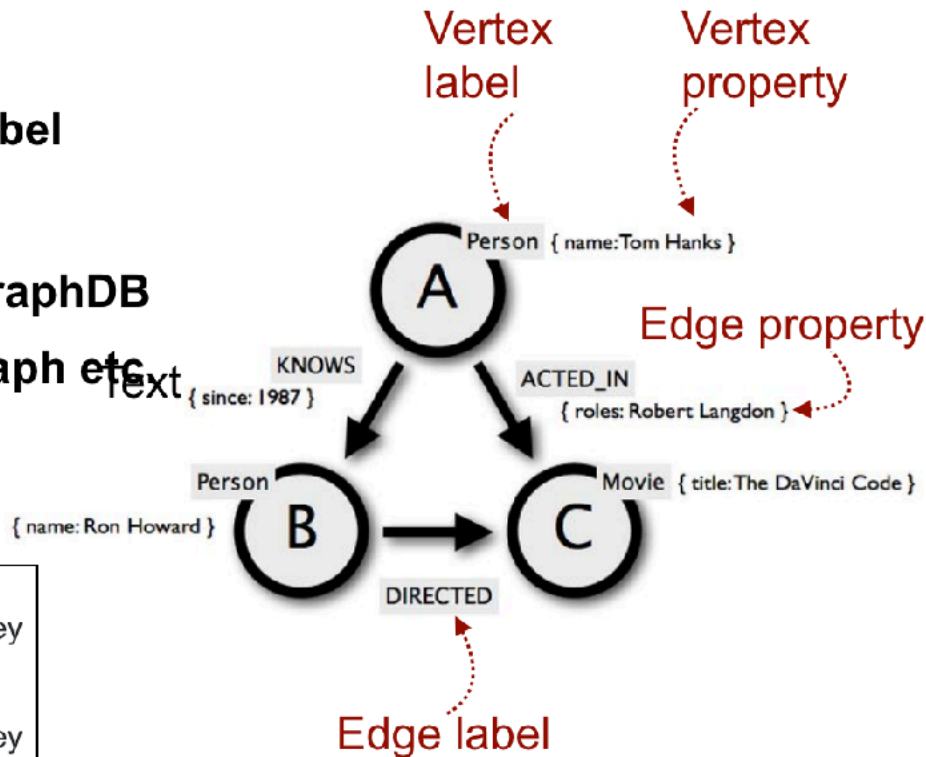
- **Definition**

- **graph + property + label**

- **Industry adoption**

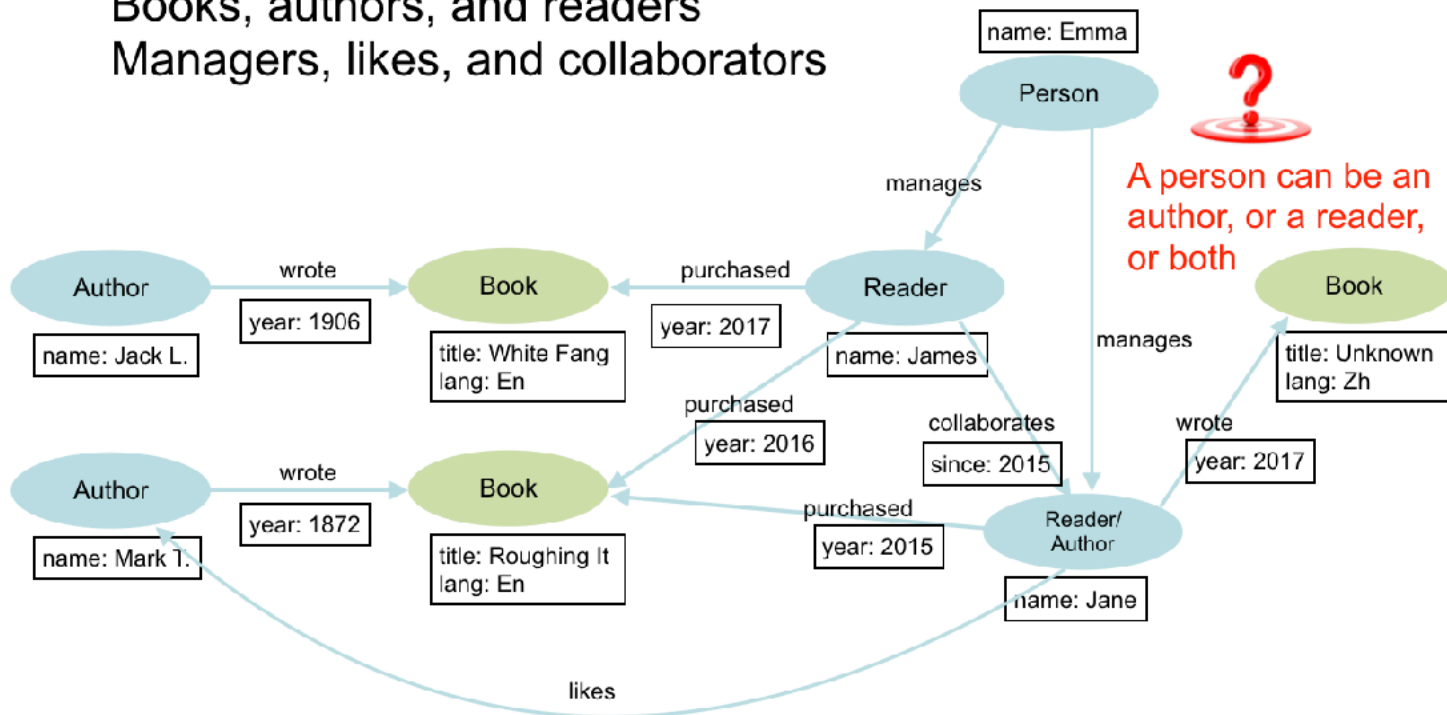
- **de facto model for GraphDB**
- **Noe4j, Titan, DSE Graph etc**

- each vertex has a collection of properties defined by a map from key to value.
- each edge has a collection of properties defined by a map from key to value.



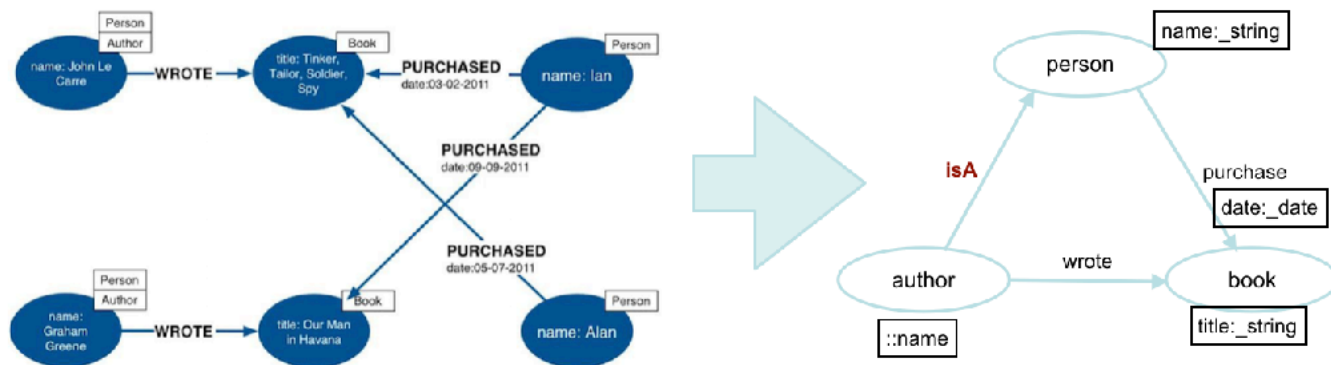
# Challenges - Label Ambiguity

Books, authors, and readers  
Managers, likes, and collaborators

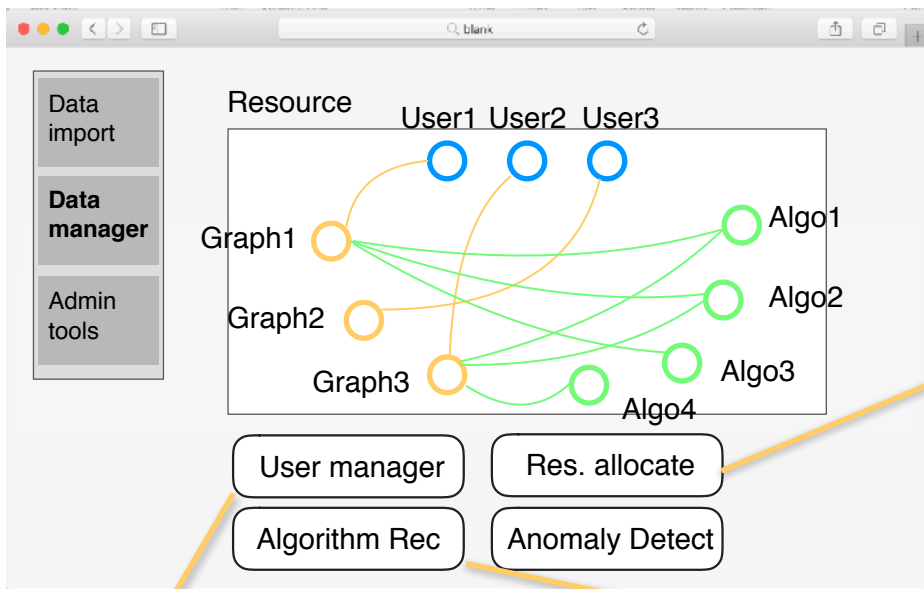


# Graph-based Property Management

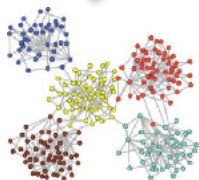
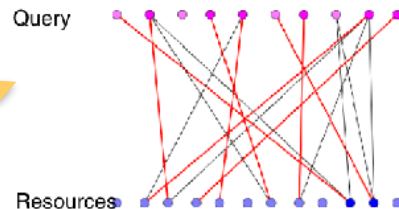
- **Schema graph of a property graph**
  - Vertex — a label representing a vertex type
  - Edge — a label representing an edge type
- **Schema graph ontology**
  - Relevant labels are connected using directed edge



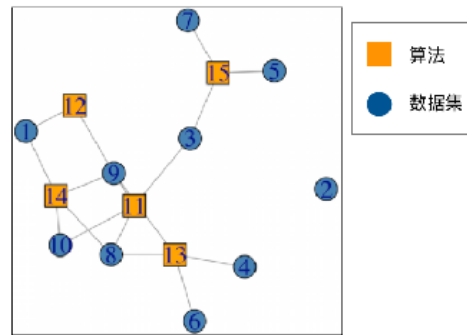
# Graph-based DevOps for Eywa



Use a graph to link algorithms, datasets, and users, which DevOps by grouping users, predicting algorithm usage, and offer helps



Utilize ML and graph algorithms to analyze the above graphs, Eywa enjoys some basic AI capability in its DevOps

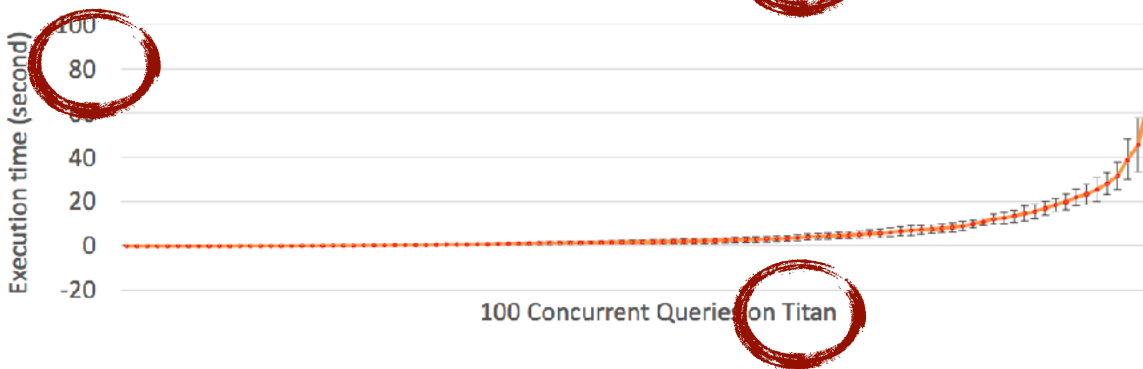
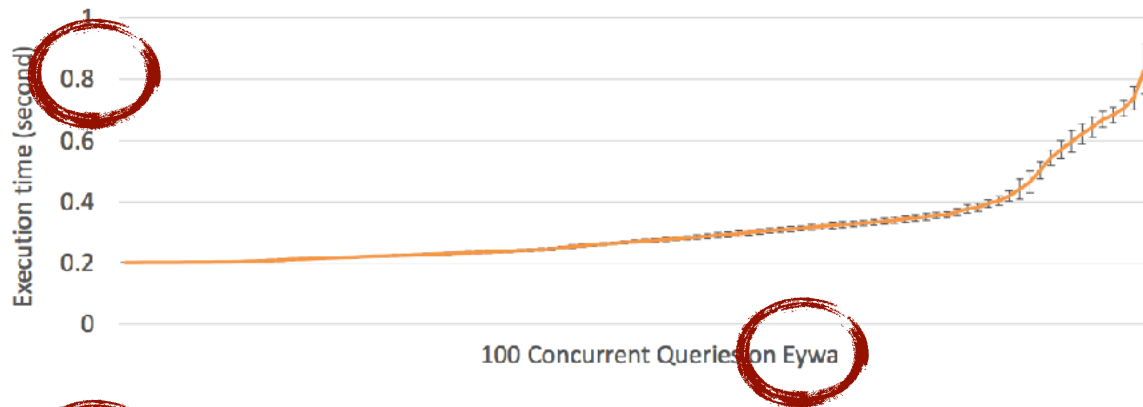


# Experiments

- **Infrastructure**
  - 9 VMs from data center
  - each with 22-core CPU , 125 GB memory
- **Dataset**
  - Orkut: IVI=3.07M, IEI=117M
  - Friendster: IVI=65M, IEI=1.8B
  - Kronecker: IVI=984M, IEI=106.5B
- **Query evaluation**
  - Given a set of vertices, perform 3-hop local traversal

# Experiments - 1

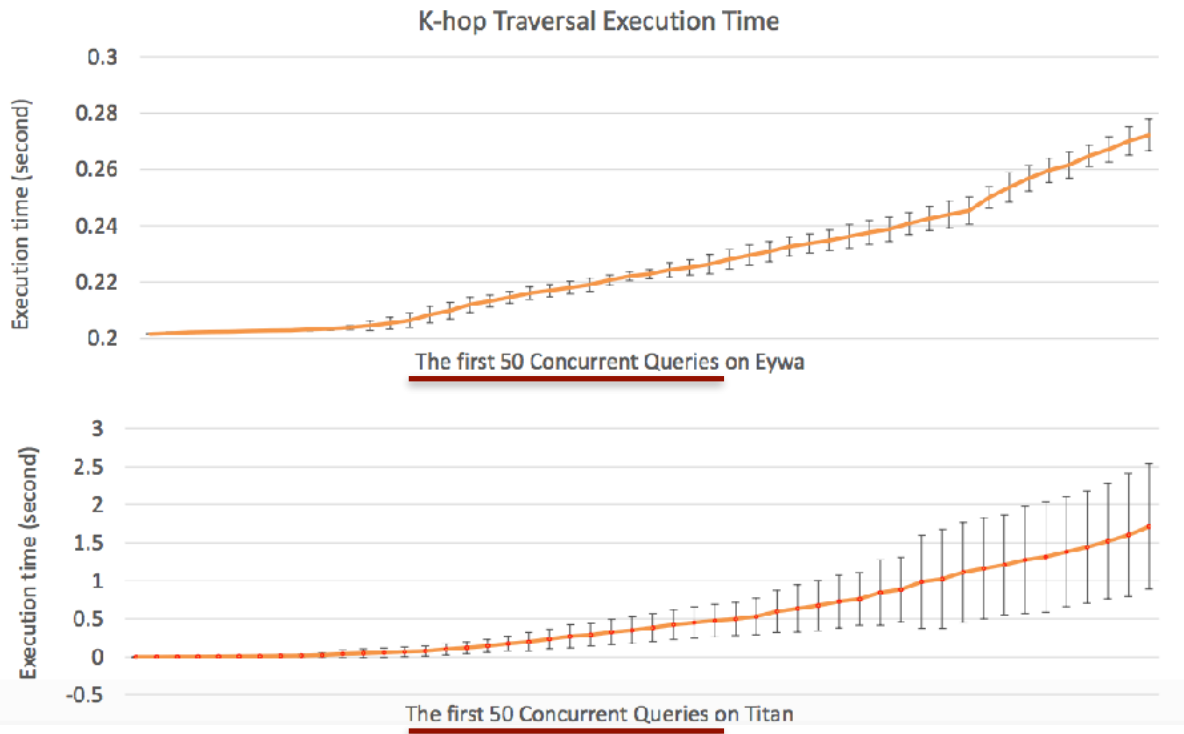
100 Concurrent Queries, each executing 100 consecutive 3-hop traversals with random roots



- Dataset: **Orkut**
- Platform: Eywa vs Titan
- Infrastructure: a single machine
- Test bench: local traversal
- Concurrency: 100 queries

# Experiments - 2

100 Concurrent Queries, each executing 100 consecutive 3-hop traversals with random roots



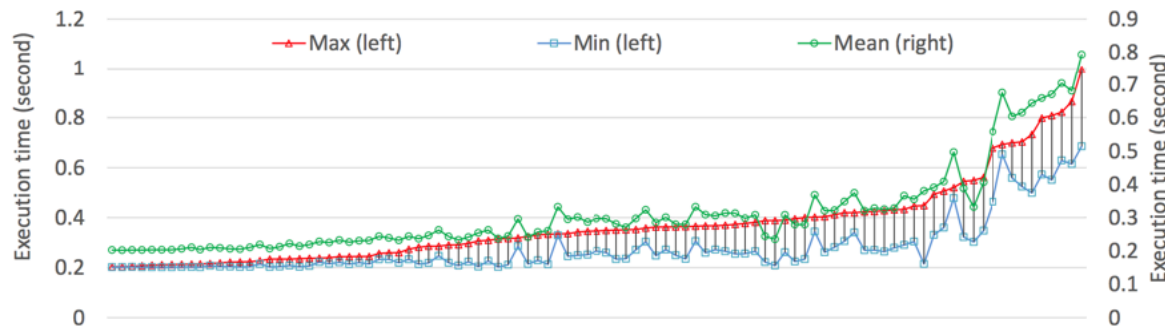
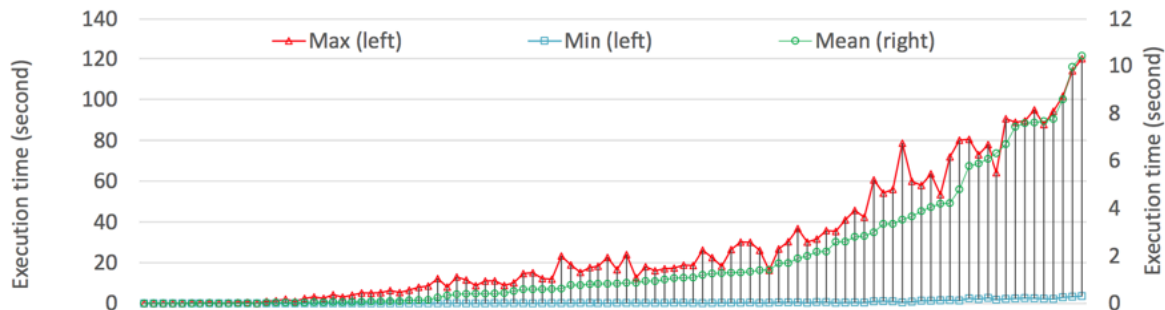
- Dataset: Orkut
- Platform: Eywa vs Titan
- Infrastructure: a single machine
- Test bench: local traversal
- Concurrency: 100 queries



# Experiments - 3

Eywa VS. Titan ( $|V|$ : 3072441,  $|E|$ :117185083)

100 Concurrent Queries, each executing 100 consecutive 3-hop traversals with random roots



- Eywa outperformed the baseline method
- Eywa shows consistent running time

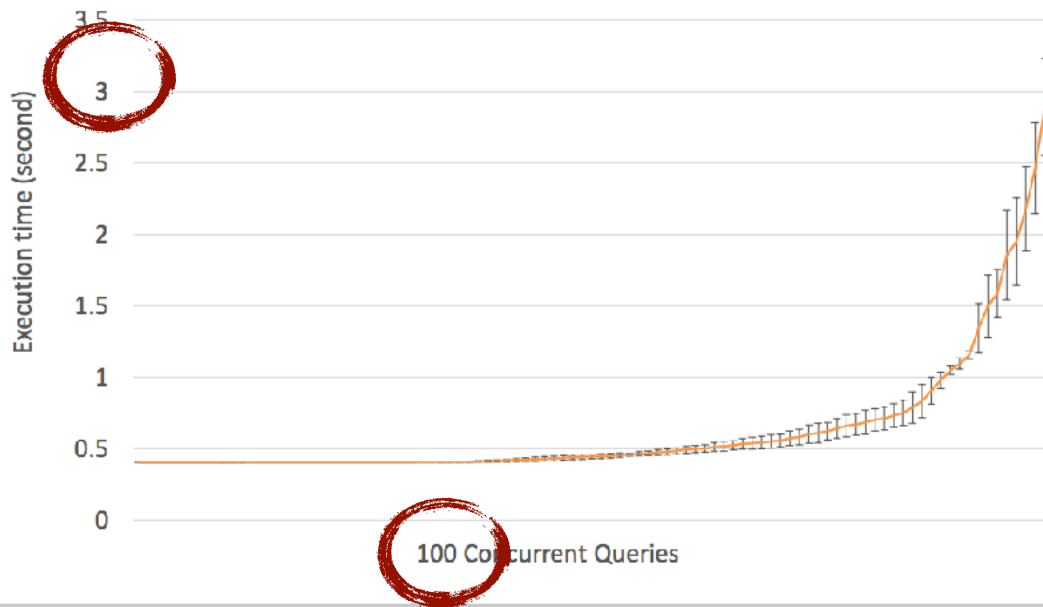
# Experiment - 4

Eywa ( $|V|$ : 984,125,490,  $|E|$ :106,579,558,164)

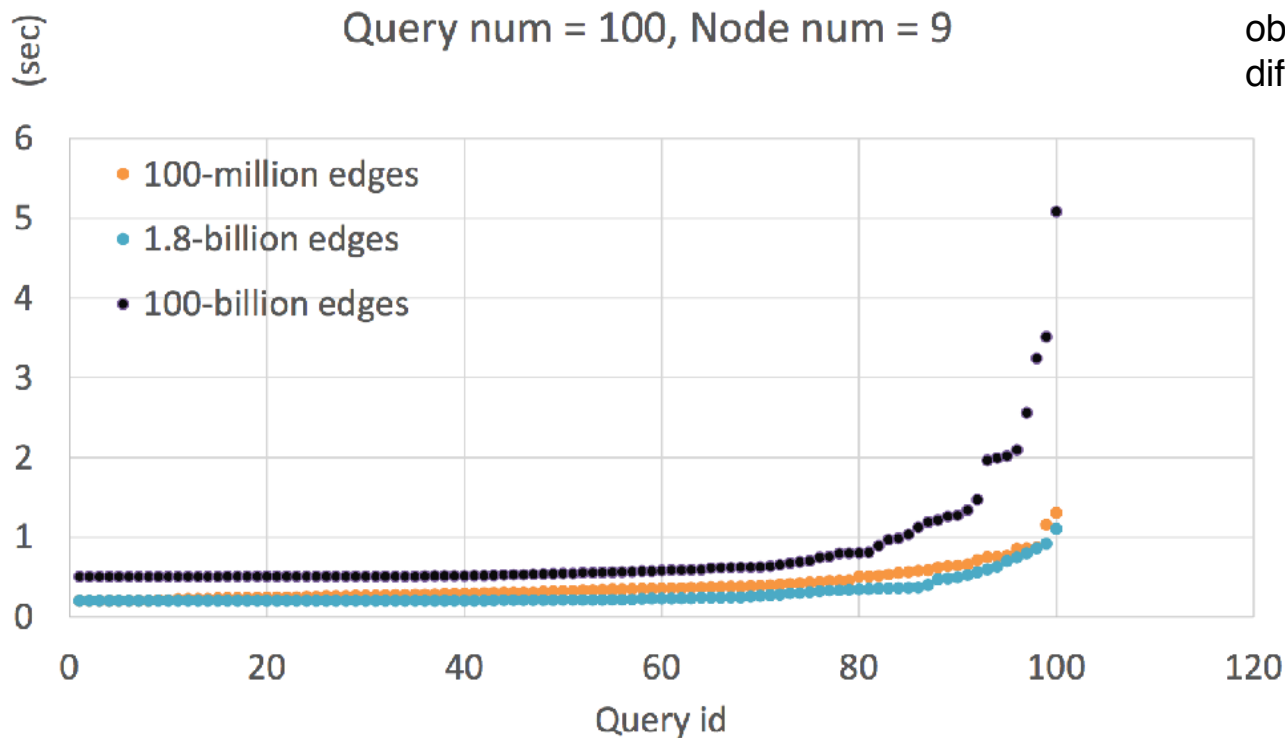
100 Concurrent Queries, each executing consecutive 3-hop traversals with random roots

- Dataset: **Kronecker**
- Platform: Eywa
- Infrastructure: 9 VMs
- Test bench: local traversal
- Concurrency: 100 queries

K-hop Traversal Execution Time

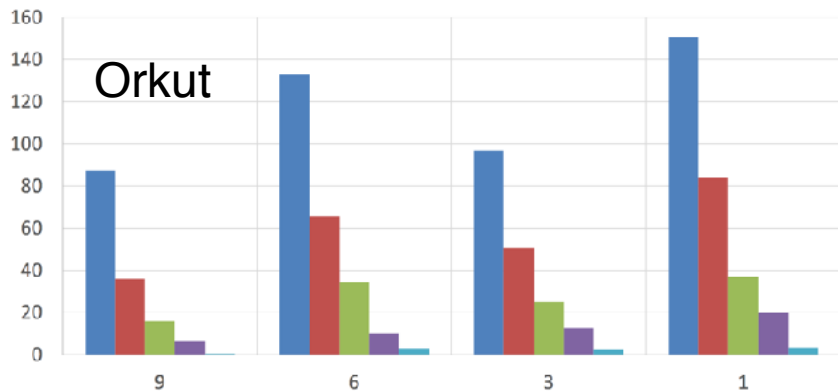
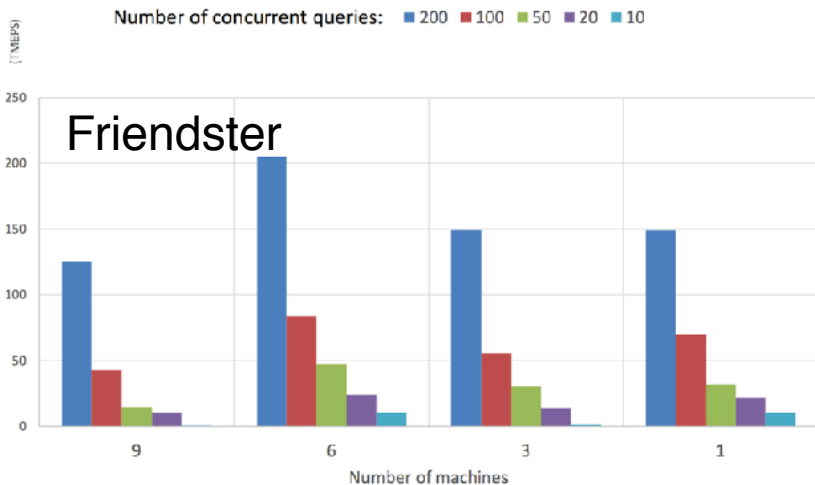


# Experiment - 5



- **Consistent** performance observed across graphs of different scales

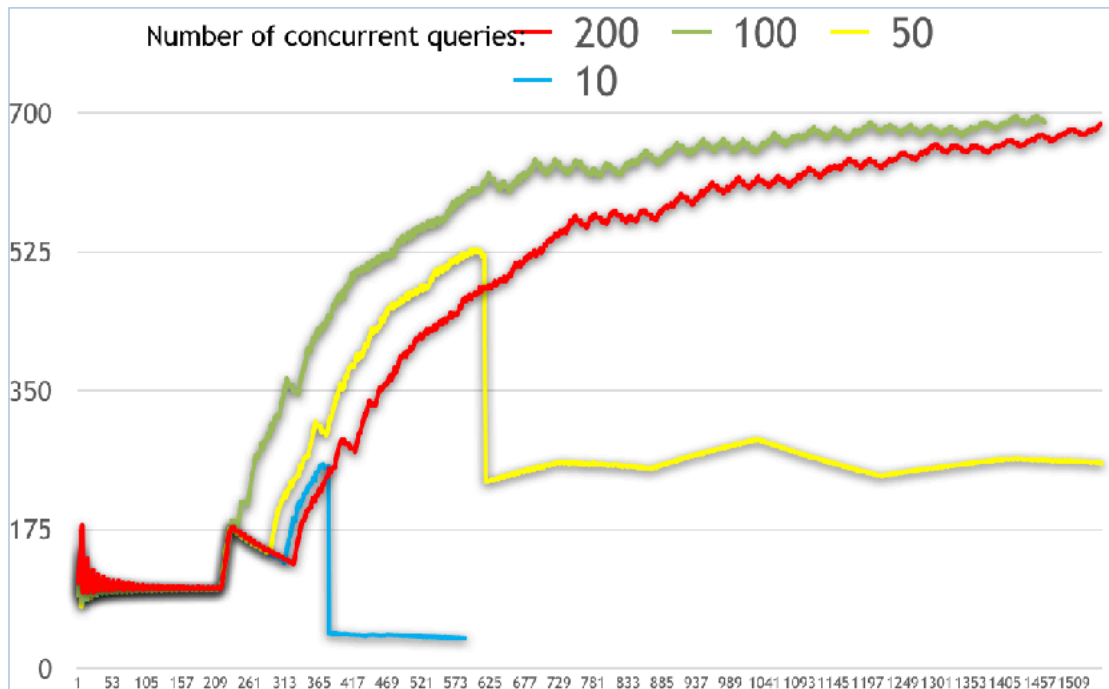
# Experiment - 6



- Consistent performance observed across graphs of different scales
- Better overall throughput for **higher concurrency**
- For **bigger** graph, the **throughput** increases to some extent as the number of VMs increases

# Experiment - 7

## CPU Usage: friendster



- CPU usages seems not high
- Further concurrent queries can be supported
- Each query may incur multiple threads (workers) for distributed communication reasons, which consumes little CPU resources, but occupy a core (OS can swap them out)



**Copyright©2012 Huawei Technologies Co., Ltd. All Rights Reserved.**

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. Huawei may change the information at any time without notice.