

# New LDBC SNB Benchmark Record by Galaxybase

More than **6** times faster and **70%** higher throughput

Chen Zhang

CreateLink Technology Co., Ltd.



# CreateLink Technology Introduction

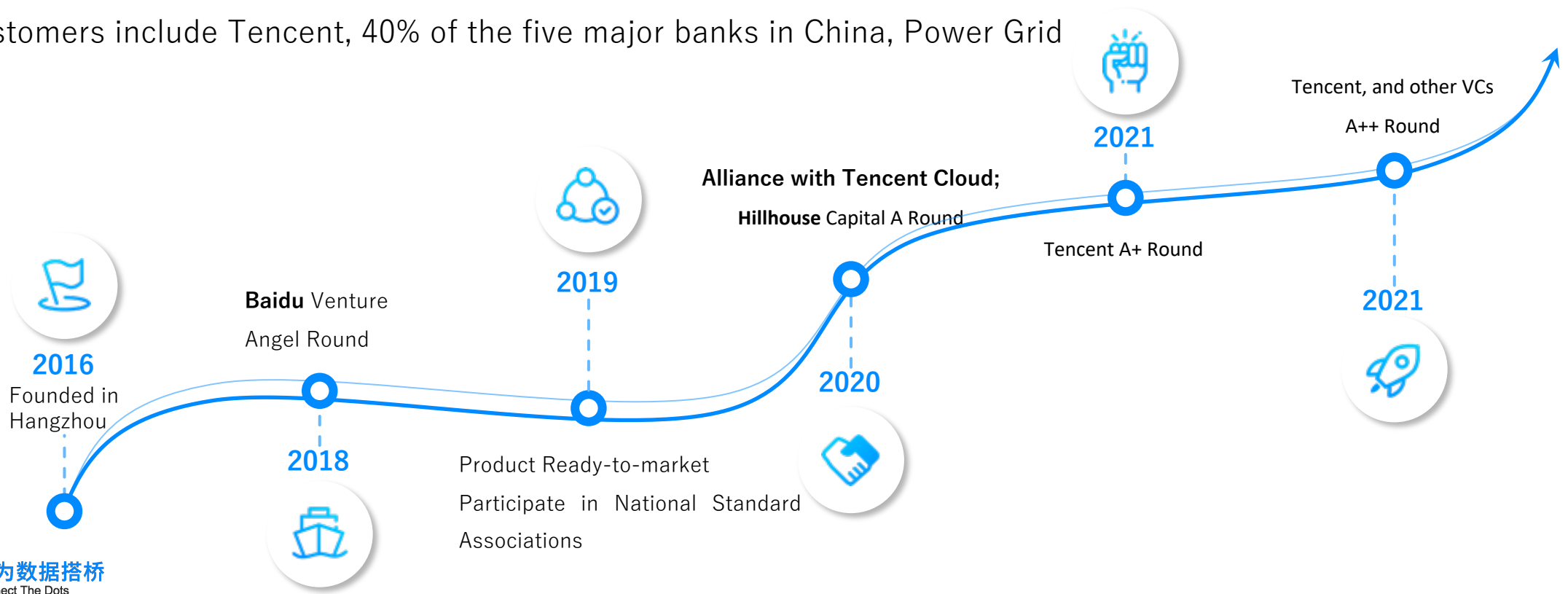


Core product **Galaxybase** is a native high-performance graph platform provide one step solution to large scale graph data store and computation

**Galaxybase** is the current **LDBC-SNB** world record holder

**Galaxybase** is the current **largest-scale graph** record holder, 5-trillion graph

Customers include Tencent, 40% of the five major banks in China, Power Grid



# Enterprise Performance



## Patents

There are a number of patents related to the existing graph technology.

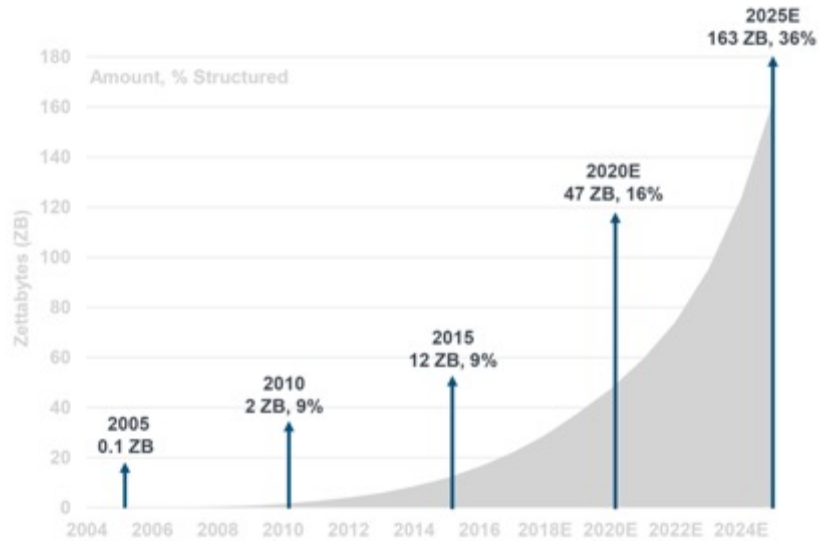


## Industry Recognition

- August 2019 Won the Amazon Cloud 2019 AI Fusion Award.
- September 2019 Appraised as an influential enterprise in the field of Internet + knowledge graph deep mining and an excellent solution enterprise in the field of graph data storage.
- October 2020 Awarded "the most influential enterprise in the field of graph database in 2020" and "the best solution enterprise in the field of graph database in 2020" by CCIDnet.
- December 2020 Awarded the 2020 Data Wind and Cloud Award for Annual Innovative Enterprises.
- January 2021 Won the Excellent Solution Award for IT Services of Commercial Banks.
- July 2021 Awarded "China's New Economy Unicorn & Quasi-Unicorn 2021".
- August 2021 Became a board member International Linked Data Benchmarking Committee (LDBC).
- January 2022 Won the best fundamental product of Tencent Yunqi in 2021.
- January 2022 Awarded "Leading Enterprise of Domestic Commercial Graph Database in 2021" and "Best Domestic Commercial Graph Database Product in 2021" by CCIDnet.
- May 2022 Participated in the development of the benchmark technical specifications for the FIE-1 Energy Industry Graph Database.

# The Value of Graph Technology

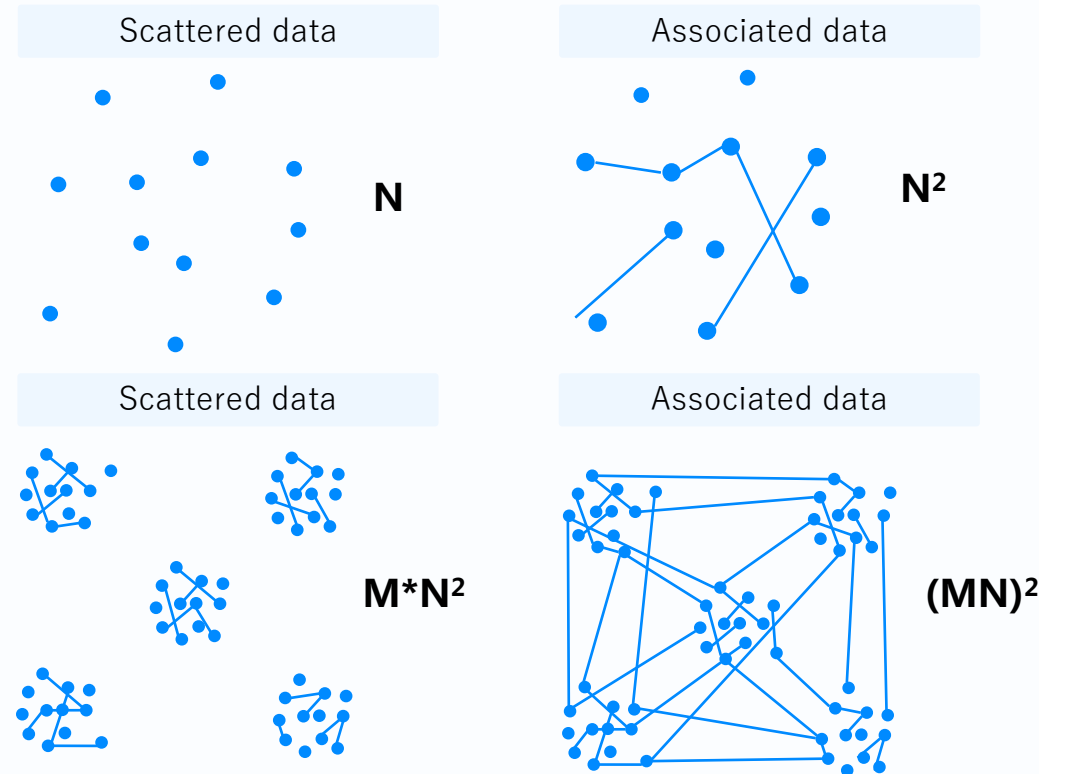
The value of big data lies in the discovery and application of relationships and data patterns, while the graph database is the intuitive storage of data relationships.



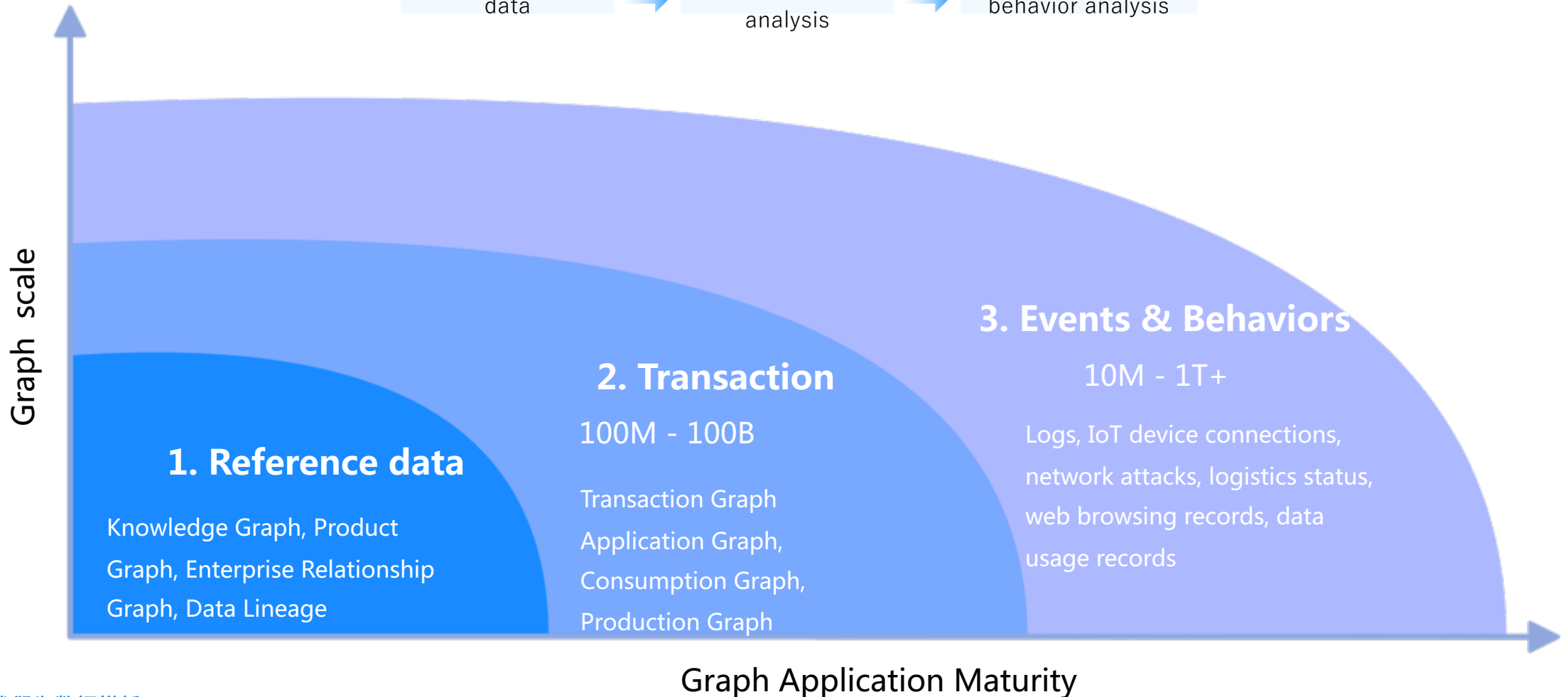
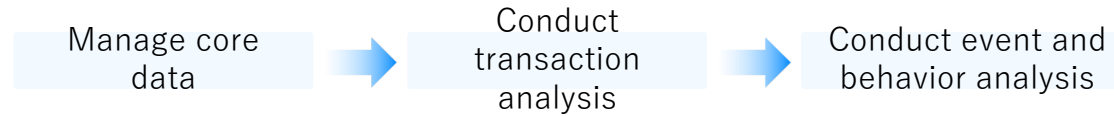
The total amount of global data is expected to increase by more than 80 times in 2025 compared with 2010.

The value of data has been highly valued, from the previous value mining of a small number of isolated data to the value mining of massive global data.

## Exponential Growth of Graph Data Value



# The Development History of Graph Application



Galaxybase has the best Interactive query performance  
and the more complex the query, the greater the performance benefit.

## LDBC SNB

The most complete benchmark of graph databases in social network scenarios so far.

### SF100 Dataset Test Results

Test environment: 24-core 2.5GHz CPU, 372G memory, gigabit network broadband, 2 \* 900GB mechanical hard disk

Item	Quantity	Galaxybase v3.3.0 Fast	TuGraph v0.3.3 Fast	Average response time compare to TuGraph	Faster than TuGraph in response time by
Interactive Complex Query (IC)	14	100%	0%	6.1 times	54.1 times
Interactive Simple Query (IS)	7	100%	0%	4.7 times	11.2 times
Interactive Inserts (II)	8	100%	0%	9.7 times	72.6 times
Total	29	100%	0%	6.7 times	72.6 times

Test data set: SF100 data set, 280 million vertexes, 1.78 billion edges

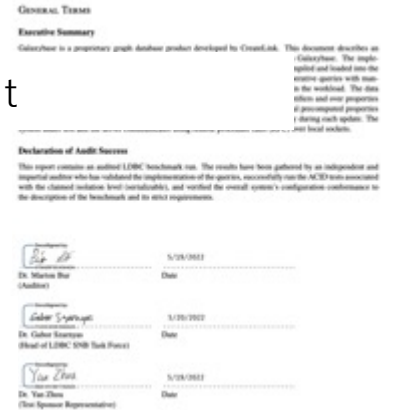
## Breaking LDBC SNB



**70%** increase in throughput

Average **6X** improvement in query performance

Up to **72X** improvement in query performance



Complete test report:  
<https://ldbcouncil.org/benchmarks/snb/>.

# The Galaxybase Advantage-Superior Horizontal Scalability



36Kr

## 最前线 | 创邻科技Galaxybase图数据库完成5万亿规模交易数据智能挖掘





王与桐 · 2022-01-25 [关注](#)

打破世界纪录

2021年11月底工信部印发《“十四五”软件和信息技术服务业发展规划》，明确提出“突破大规模并行图数据处理关键技术”，推动高性能数据库在金融、电信、能源、制造等重点行业关键业务系统应用。图数据库作为以图论为设计原理的数据库管理系统，将现实世界的实体和实体关系抽象表达为顶点和边，擅长海量图数据的高效存储、查询、计算、分析，能有效解决传统数据库技术无法解决的大数据关联难题，在金融风险、精准零售、物流优化、能源调度、生物制药、智能交通、疫情防控等多种新兴领域有巨大的应用价值。其解决了传统技术关联查询效率低、成本高的问题，具有丰富、高效和敏捷的数据处理能力。而伴随全球数字化竞赛日趋白热化，这将直接影响企业数字化、智能化进程。

## Breaking the Scale of Graph Data Processing

### World Record

-  5 trillion large-scale graph distributed storage, real-time online query
-  Covers super nodes with a maximum outgoing edges of more than **10 million**.
-  6-hop deep link query averaging 6.7 s
-  Uses only 50 machine cluster

# Galaxybase: Graph platform for storage, computation, and analytics



## Customer pain points

Slow query speed and poor business experience



**Fast:** Native distributed parallel graph storage with millions of neighbors returned in 1 second.

Poor scalability and difficulty in data increment



**High scalability:** Fully distributed architecture, dynamic online capacity expansion, and efficient support for trillion-level super large graphs.

Large data expansion and high hardware cost



**Efficient data compression:** Optimize resource utilization, 3-5X save on hardware and maintenance costs.

High ETL and time cost for graph computation



**Real-time computation:** Built-in rich distributed graph algorithm and real-time graph analysis without ETL.

Slow start and high learning cost



**Visual analysis without programming:** the business is easy to understand and the technical value is realized quickly.

Graph Storage

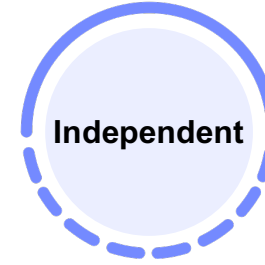
Graph Computation

Graph Analytics

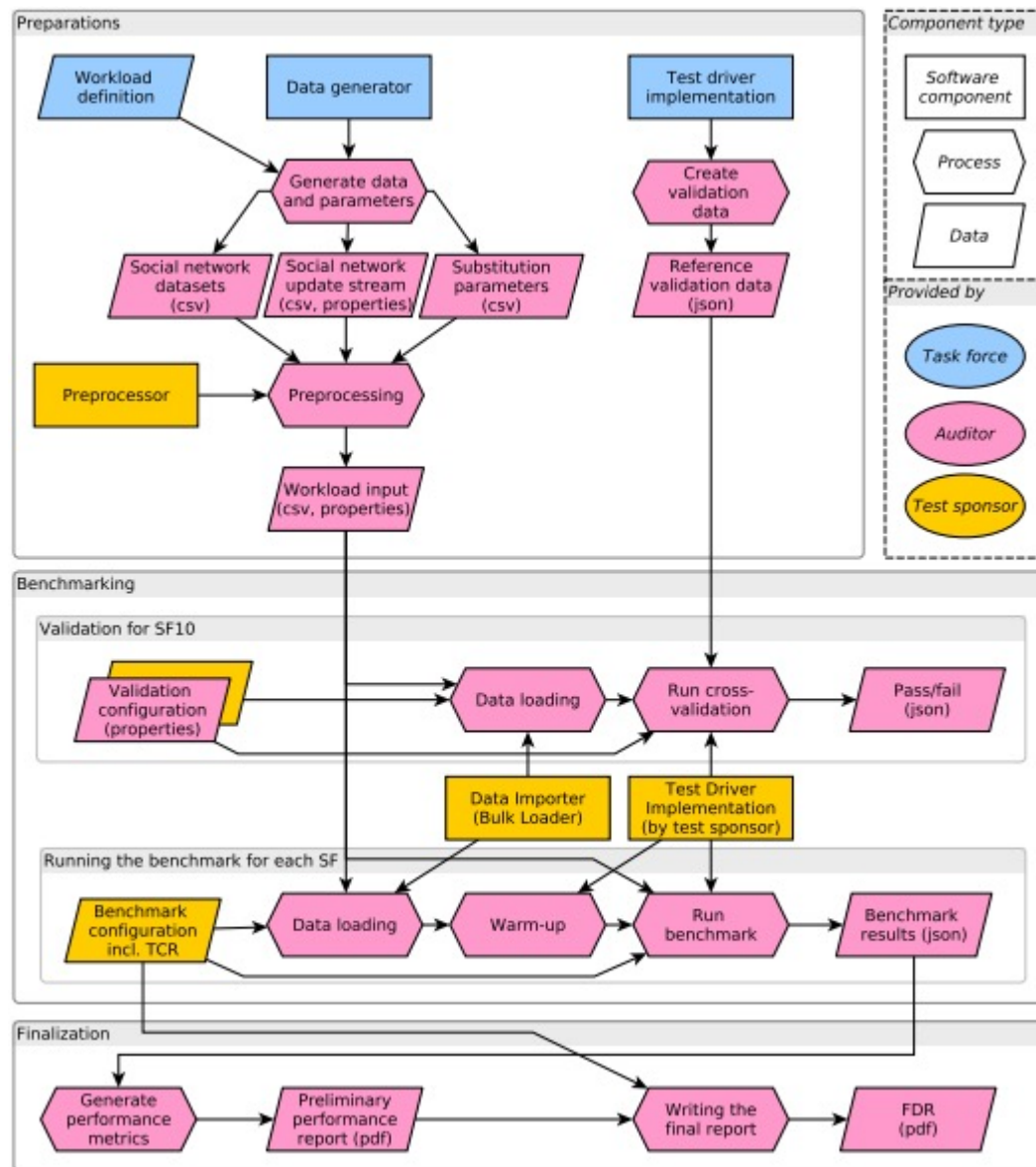


# Why Choose LDBC?

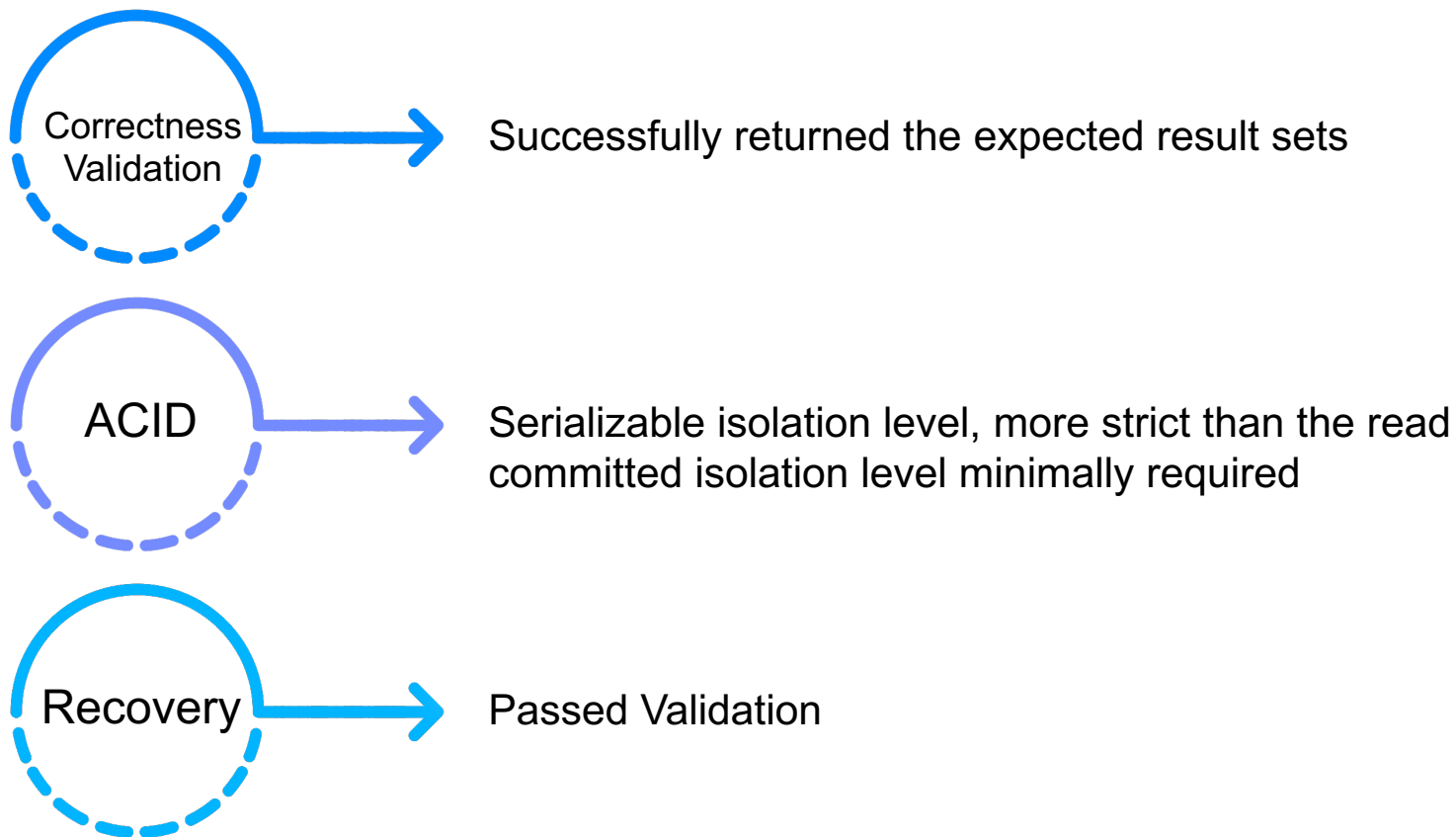
The Linked Data Benchmark Council (LDBC) is a non-profit organization aiming to define standard graph benchmarks to foster a community around graph processing technologies. LDBC consists of members from both industry and academia, including organizations and individuals. Successfully served many leading customers in finance, energy, Internet and other industries, and supports super large graph online with trillions of nodes and edges.



# LDBC Benchmark Procedure



# Galaxybase Test Results

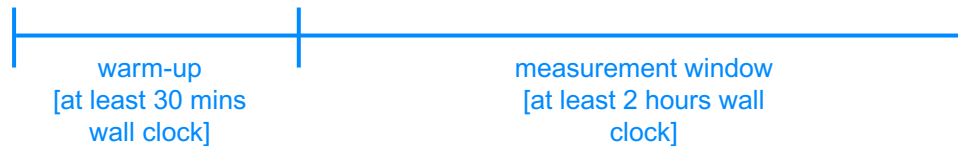


# LDBC Performance Testing Method

95% on-time requirement :

$\text{actual\_start\_time} - \text{scheduled\_start\_time} < 1 \text{ second}$

The measurement window :



The key configuration :

```
thread_count=48  
time_compression_ratio=0.0038  
warmup=19000000  
operation_count=75000000
```

# LDBC Performance Testing Results



## Performance summary:

- Duration
- Operations
- Throughput
- Query on-time compliance

## Detailed performance:

- Count
- Mean
- 50th Percentile
- 90th Percentile
- 95th Percentile
- 99th Percentile

# LDBC SNB Record



Before Galaxybase, the holder of the LDBC-SNB record was TuGraph in 2020.

System details	Previous
Cloud Provider	Amazon Web Services
Instance Type	r5d.12xlarge
Operating System	18.04.1-Ubuntu
CPU Count/Cores/Threads	1/24/48
Memory Size	374GB
Storage	2 x 900GB NVMe SSD

Throughput	Previous
sf30	5436.37
sf100	5010.77
sf300	4855.52

# Galaxybase LDBC SNB Official Audit



In this LDBC-SNB audit, Galaxybase selects the same system configuration with TuGraph.

System details	Previous	Galaxybase
Cloud Provider	Amazon Web Services	Amazon Web Services
Instance Type	r5d.12xlarge	r5d.12xlarge
Operating System	18.04.1-Ubuntu	18.04.1-Ubuntu
CPU Count/Cores/Threads	1/24/48	1/24/48
Memory Size	374GB	372GB
Storage	2 x 900GB NVMe SSD	2 x 900GB NVMe SSD

# Galaxybase LDBC SNB Official Audit



Galaxybase throughput increases by 70%.

Throughput	Previous	Galaxybase	Rate of increase
sf30	5436.37	9285.86	70.8%
sf100	5010.77	8501.21	69.7%
sf300	4855.52	8370.52	72.4%



# Galaxybase LDBC SNB Official Audit SF100 Result



The response time of Galaxybase is faster than the previous record in all aspects.

Data base	Mean response time			50% response time			90% response time			95% response time			99% response time		
	Previ ous	Galax ybase	ratio	Previ ous	Galax ybase	ratio	Previ ous	Galax ybase	ratio	Previ ous	Galax ybase	ratio	Previ ous	Galax ybase	ratio
IC1	137.11	23.10	5.93	70.22	7.04	9.98	253.84	9.80	25.90	792.77	229.63	3.45	959.94	274.70	3.49
IC2	5.73	1.64	3.49	4.58	1.19	3.86	8.44	1.75	4.82	11.88	2.10	5.67	28.65	6.41	4.47
IC3	93.50	54.66	1.71	90.28	50.94	1.77	117.80	66.13	1.78	123.47	73.92	1.67	178.90	140.02	1.28
IC4	8.69	2.91	2.99	7.02	2.26	3.11	15.11	3.98	3.79	17.55	4.70	3.73	34.27	9.81	3.49
IC5	1280.00	230.11	5.56	1306.43	225.29	5.80	1717.95	321.10	5.35	1823.30	361.58	5.04	2000.06	435.10	4.60
IC6	28.89	7.54	3.83	22.22	7.99	2.78	61.35	15.30	4.01	66.50	17.31	3.84	80.12	31.88	2.51
IC7	1.98	0.52	3.81	0.73	0.10	7.12	4.41	0.26	17.28	7.66	0.36	21.16	21.50	4.18	5.14
IC8	1.75	0.54	3.25	0.55	0.13	4.14	3.99	0.21	18.73	7.24	0.29	25.30	20.75	4.22	4.92
IC9	600.30	151.06	3.97	586.75	144.91	4.05	788.29	204.51	3.85	845.09	249.15	3.39	956.48	312.83	3.06
IC10	131.85	50.10	2.63	129.34	48.15	2.69	168.14	59.88	2.81	181.22	66.51	2.72	227.09	126.26	1.80
IC11	41.33	1.12	36.90	40.21	0.71	57.04	49.93	0.92	54.15	53.25	1.09	48.68	92.36	5.34	17.29
IC12	94.91	30.22	3.14	92.60	28.27	3.28	127.37	40.13	3.17	139.82	46.35	3.02	180.70	90.18	2.00
IC13	3.86	1.09	3.53	2.91	0.62	4.72	6.71	1.26	5.33	10.12	1.45	6.97	24.79	5.60	4.43
IC14	34.98	8.06	4.34	6.75	1.33	5.07	108.44	23.92	4.53	127.88	28.99	4.41	186.62	54.06	3.45

# Galaxybase LDBC SNB Official Audit SF100 Result



The response time of Galaxybase is faster than the previous record in all aspects.

	Mean response time			50% response time			90% response time			95% response time			99% response time		
Data base	Previ ous	Galax ybase	ratio	Previ ous	Galax ybase	ratio	Previ ous	Galax ybase	ratio	Previ ous	Galax ybase	ratio	Previ ous	Galax ybase	ratio
IS1	0.58	0.09	6.17	0.28	0.06	4.74	0.91	0.08	11.20	2.22	0.10	23.37	5.53	0.40	13.79
IS2	0.76	0.24	3.25	0.47	0.18	2.61	1.23	0.32	3.87	2.32	0.44	5.29	5.71	0.84	6.82
IS3	1.40	0.68	2.07	0.85	0.36	2.36	3.43	1.26	2.73	5.05	2.99	1.69	7.54	4.74	1.59
IS4	0.59	0.11	5.38	0.29	0.07	4.01	0.94	0.10	9.65	2.27	0.11	20.07	5.60	0.44	12.63
IS5	0.55	0.10	5.76	0.26	0.07	3.94	0.83	0.09	9.14	2.06	0.10	19.99	5.43	0.30	17.99
IS6	0.56	0.10	5.64	0.27	0.07	3.79	0.83	0.10	8.77	2.06	0.11	19.05	5.45	0.30	18.49
IS7	0.65	0.13	4.95	0.37	0.10	3.85	0.96	0.16	5.88	2.17	0.19	11.20	5.58	0.37	15.26
II1	4.24	1.18	3.59	2.29	0.31	7.47	10.54	1.44	7.34	15.72	2.91	5.40	30.09	12.18	2.47
II2	1.89	0.23	8.14	0.90	0.08	11.34	4.25	0.11	39.73	6.35	0.14	45.06	14.48	2.06	7.02
II3	1.93	0.23	8.33	0.92	0.08	11.59	4.35	0.11	40.64	6.41	0.14	44.85	14.60	2.20	6.63
II4	3.34	0.32	10.42	0.96	0.14	6.72	11.62	0.20	56.96	15.40	0.31	50.50	20.82	2.29	9.11
II5	2.13	0.25	8.37	1.16	0.10	11.69	4.45	0.15	28.87	6.68	0.21	31.66	15.15	2.12	7.15
II6	6.50	0.31	21.23	3.18	0.17	19.15	16.69	0.23	72.56	18.92	0.28	67.57	24.25	1.65	14.74
II7	4.15	0.32	13.00	1.24	0.16	7.92	13.84	0.21	66.56	16.82	0.26	65.94	21.83	2.09	10.46
II8	5.68	1.34	4.25	3.24	0.53	6.10	12.54	2.90	4.33	20.13	5.38	3.74	37.07	10.76	3.45

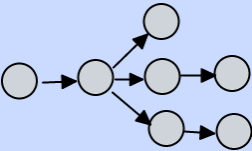

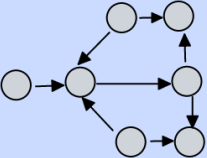
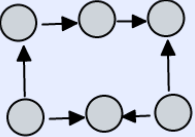
# Galaxybase LDBC SNB Official Audit



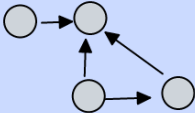
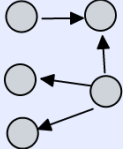
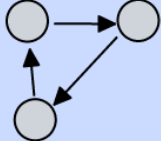
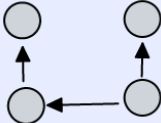
The average response time more than 6 times faster and the best response time is more than 72 times faster.

Item	Average response time compare to Previous Record Holder	Faster than Previous Record Holder in response time by
sf30	6.0 times	49.24 times
sf100	6.7 times	72.6 times
sf300	6.4 times	58.7 times
Total	6.4 times	72.6 times

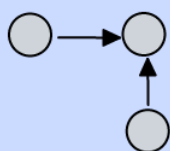
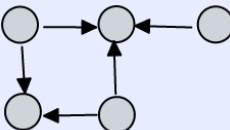
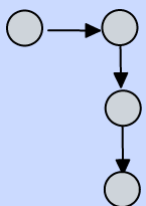
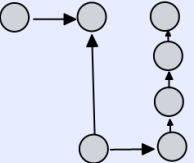
# Test Items

Item	Title	Pattern	Choke Points
IC1	Transitive friends with certain name		Rich join order optimization Intra-query result reuse
IC2	Recent messages by your friends		Interesting orders Late projection Join type selection Dimensional clustering
IC3	Friends and friends of friends that have been to given countries		Rich join order optimization Detecting correlation Flattening sub-queries
IC4	New topics		Join type selection


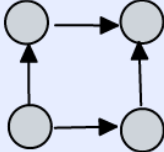

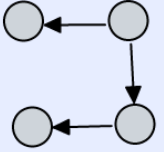
# Test Items

Item	Title	Pattern	Choke Points
IC5	New groups		Sparse foreign key joins Scattered index access patterns
IC6	Tag co-occurrence		Flattening sub-queries
IC7	Recent likers		Late projection Join type selection Scattered index access patterns Flattening sub-queries
IC8	Recent replies		Sparse foreign key joins Scattered index access patterns Intra-query result reuse




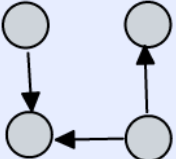
# Test Items

Item	Title	Pattern	Choke Points
IC9	Recent messages by friends or friends of friends		<ul style="list-style-type: none"> <li>Interesting orders</li> <li>High cardinality group-by performance</li> <li>Late projection</li> <li>Join type selection</li> <li>Dimensional clustering</li> <li>Scattered index access patterns</li> </ul>
IC10	Friend recommendation		<ul style="list-style-type: none"> <li>Join type selection</li> <li>Scattered index access patterns</li> <li>Common subexpression elimination</li> <li>Complex boolean expression joins and selections</li> <li>Flattening sub-queries</li> <li>Overlap between outer and sub-query</li> <li>Inter-query result reuse</li> <li>Incremental path computation</li> </ul>
IC11	Job referral		<ul style="list-style-type: none"> <li>Top-k pushdown</li> <li>Join type selection</li> <li>Sparse foreign key joins</li> <li>Scattered index access patterns</li> <li>Complex boolean expression joins and selections</li> </ul>
IC12	Expert search		<ul style="list-style-type: none"> <li>Scattered index access patterns</li> <li>Cardinality estimation of transitive paths</li> <li>Execution of a transitive step</li> </ul>

# Test Items

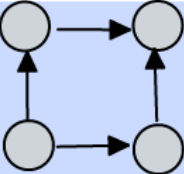
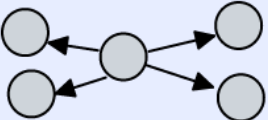


Item	Title	Pattern	Choke Points
IC13	Single shortest path		Scattered index access patterns Cardinality estimation of transitive paths Execution of a transitive step Unweighted shortest paths
IC14	Trusted connection paths		Scattered index access patterns Intra-query result reuse Cardinality estimation of transitive paths Execution of a transitive step Unweighted shortest paths Composition of graph queries
IS1	Profile of a person		Query node Query edge
IS2	Recent messages of a person		Query node Query edge

# Test Items

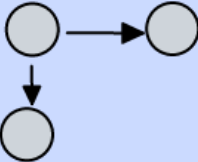

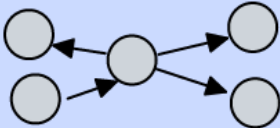
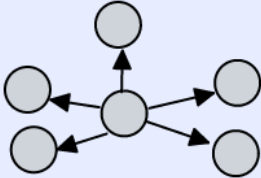

Item	Title	Pattern	Choke Points
IS3	Friends of a person		Query node Query edge
IS4	Content of a message		Query node
IS5	Creator of a message		Query node Query edge
IS6	Forum of a message		Query node Query edge



# Test Items

Item	Title	Pattern	Choke Points
IS7	Replies of a message		Query node Query edge
II1	Add person		Insert node Insert edge
II2	Add like to post		Insert edge
II3	Add like to comment		Insert edge

# Test Items

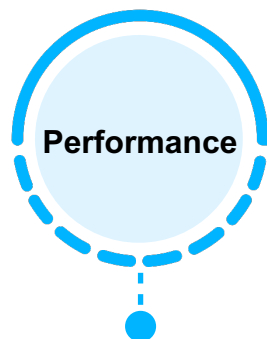
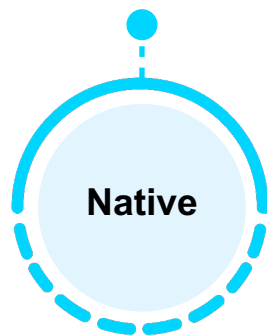
Item	Title	Pattern	Choke Points
II4	Add forum		Insert node Insert edge
II5	Add forum membership		Insert edge
II6	Add post		Insert node Insert edge
II7	Add comment		Insert node Insert edge
II8	Add friendship		Insert edge

# Technical Advantages

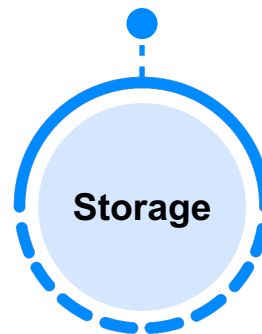
Storage advantage: Galaxybase uses an innovative proprietary native graph datastore.

The core datastore engine does not rely on any third-party open-source components.

To efficiently organize memory data through its proprietary memory allocation and management mechanism.



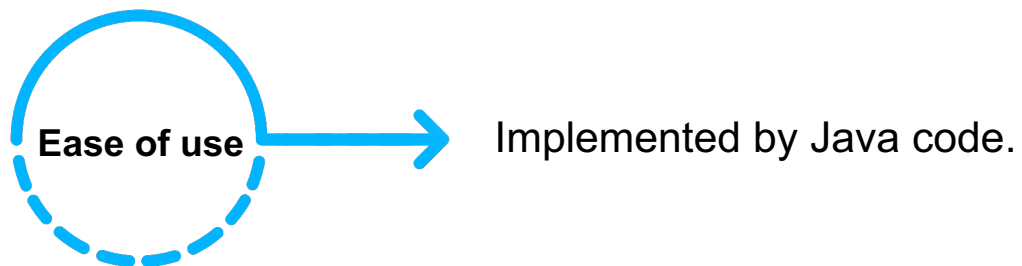
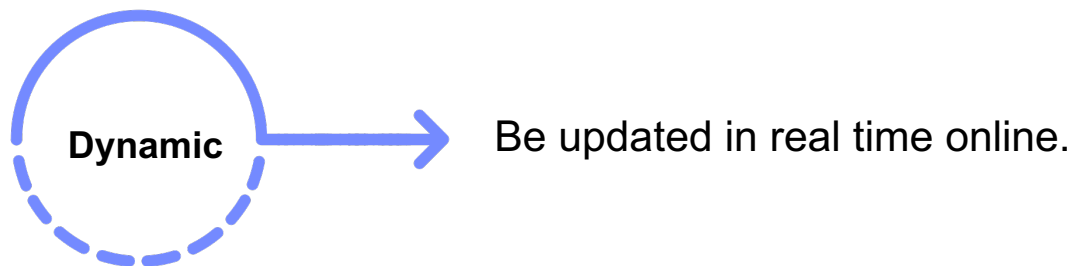
Customized optimization for index-free adjacency of graph data.



Adaptively allocates the number of threads for parallel iterations based on the number of neighbours during neighbour iteration to achieve best utilization of system resources.

# Technical Advantages

Development advantage: PAR (Parameterized Algorithm Routine) API.



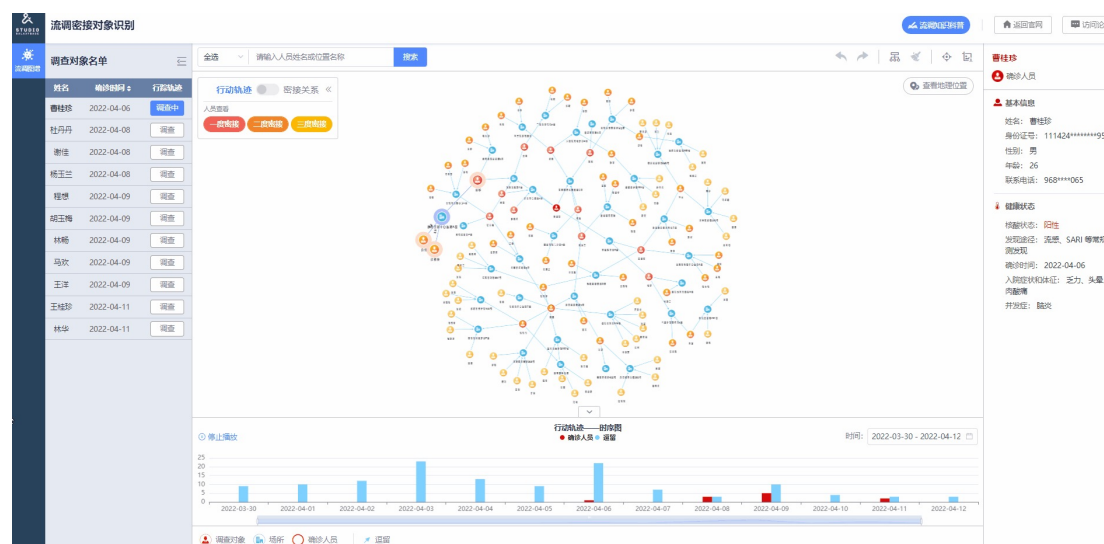
# Galaxybase Advantage-Easy-to-Use Graph Analysis Window

Multiple layout views, custom graph styles and dynamic zooming in/out to meet the requirements of high-precision multi-dimensional analysis;  
Highly complex relationship analysis can also be done by simply dragging and clicking;  
Analysis results can be exported in multiple formats.

## Path Finding



## Time Series Analysis of Community Evolution



# Success Story-IoT Analysis

## Pain Points

Fall short of responding in real time with the change of time and space for intelligent scheduling decision-making, transportation, tourism and so on;  
Huge volumes of mobile data, long processing time, low query efficiency, and high hardware cost

## Solutions

Build an IoT network based on device, WIFI hotspot, and ID information, for analysis of co-appearance and other cases, in support of its banking, government and public security customers;  
300TB original data;  
5.4 billion vertexes, 3.75 trillion edges and 18.77 trillion properties;  
Data loading speed: 15 billion edges/hour, 72.5 billion properties/hour.

## Business Value

- ✓ Efficiently store, query and daily increase data;
- ✓ Reduce cost by 500% (420 TB) and save millions of hardware costs and operation and maintenance expenses.



# Success Story-Fraud Detection



## Pain Points

Insufficient fraud black sample;  
Static expert rules falling short of detecting dynamic and changeable fraud means;  
Long time taken by manual verification

## Solutions

Build a user network of related entities (contacts, phone, address, company, IP, email, etc.) with various types of relationships (guarantor, transaction, family member, work address, etc. ) for varies fraud detection cases.

## Business Value

Save manual audit cost and improve audit efficiency;  
Improve the accuracy of fraud identification and reduce misjudgments;  
Discover risks in advance, take precautions in advance and reduce losses.



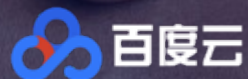
# Adopters and Partners



- Adopters



## Partners





THANKS FOR WATCHING



info@chuanglintech.com

