

The #1 Platform for Connected Data

Neo4j 图数据库平台

俞方桦 博士

Neo4j Beijing Meetup, 18th Mar. 2018

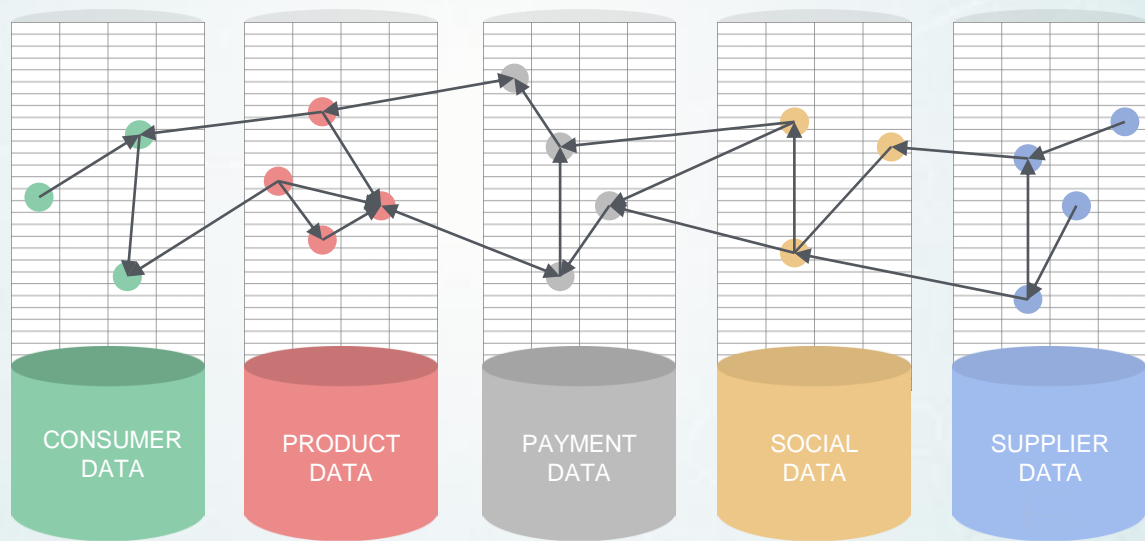


俞方桦 | Joshua Yu

Field Engineering, Neo4j APAC

Joshua.yu@neo4j.com





“(在大数据时代)新的**竞争优势**将完全来自于**理解和**
使用各种关联来产生可执行的洞见。”

“The next wave of *competitive advantage* will be all about *using connections* to produce actionable insights.”

Agenda

- The Graph Database Platform
- The Graph DBMS
- Cypher Query Language
- Analytical UI – Neo4j Browser
- Neo4j Desktop
- Visualization
- Enterprise Integration Patterns
- Data Migration

Agenda (cont.)

- Clustering
- Deployment
- Development
- Operational Support
- User Defined Procedures
- Security
- Business Use Cases
- Technology Use Cases
- The Community

Neo4j – 图数据库平台

Neo4j 图数据库平台



Applications



Business Users

Drivers & APIs
(Python, Java, JavaScript, R ...)

Discovery & Visualization

Neo4j
Desktop



Development
& Admin

Neo4j
Bloom

Neo4j
Browser



Analytics
Tooling



Developers



Admins



Data Analysts



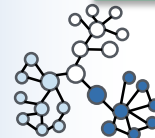
Data Scientists



Graph
Transactions



- Native graph database
- Cypher query language
- Clustering



Graph
Analytics

Data Integration

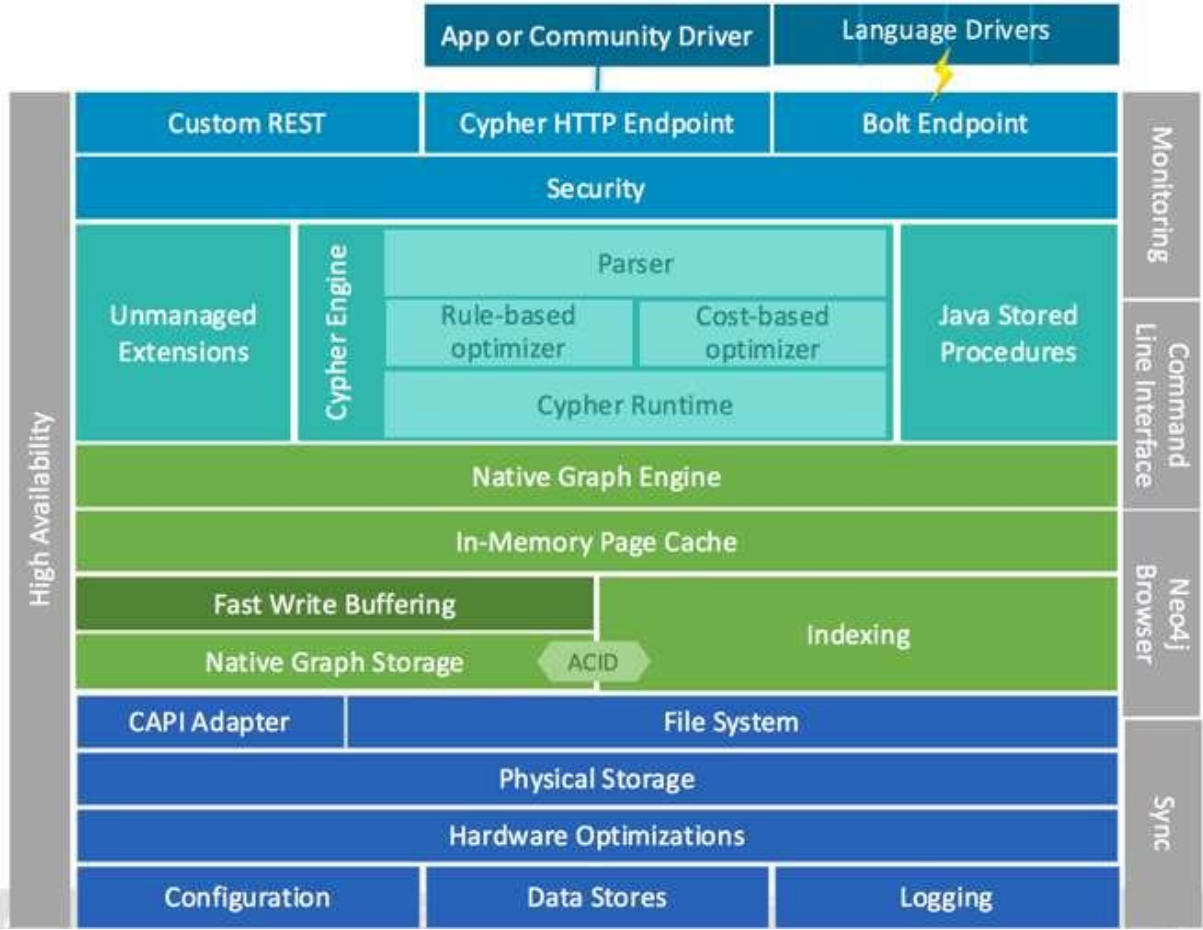


Streaming /
Realtime



Batch / ETL

Neo4j – The Complete Application Architecture



Neo4j: 企业级的图数据库

本地/原生的图存储

Designed, built, and tested for graphs

本地/原生的图处理

For real-time, relationship-based apps

Evaluate millions of relationships in a blink

直观一致的数据建模

Faster projects compared to RDBMS

ACID 交易和数据安全

Fully ACID transactions, causal consistency
and enterprise security

强大的、表达性强的查询语言

Improved productivity, with 10x to 100x less
code than SQL

因果集群应用架构

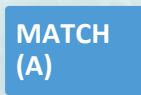
Architecture provides ideal balance of
performance, availability, scale for graphs

内置的数据导入

Seamless import from other databases

支持常用平台和语言的驱动

Fits easily into your IT environment, with
drivers and APIs for popular languages



关键性的架构组件

1

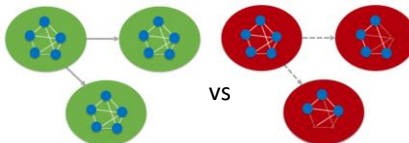
Index-Free Adjacency



In memory and on flash/disk

2

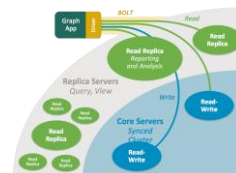
ACID Foundation



Required for safe writes

3

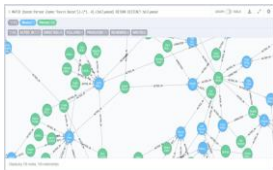
Full-Stack Clustering



Causal consistency

4

Language, Drivers, Tooling



*Developer Experience,
Graph Efficiency, Type Safety*

5

Graph Engine



*Cost-Based Optimizer, Graph
Statistics, Cypher Runtime*

6

Hardware Optimizations



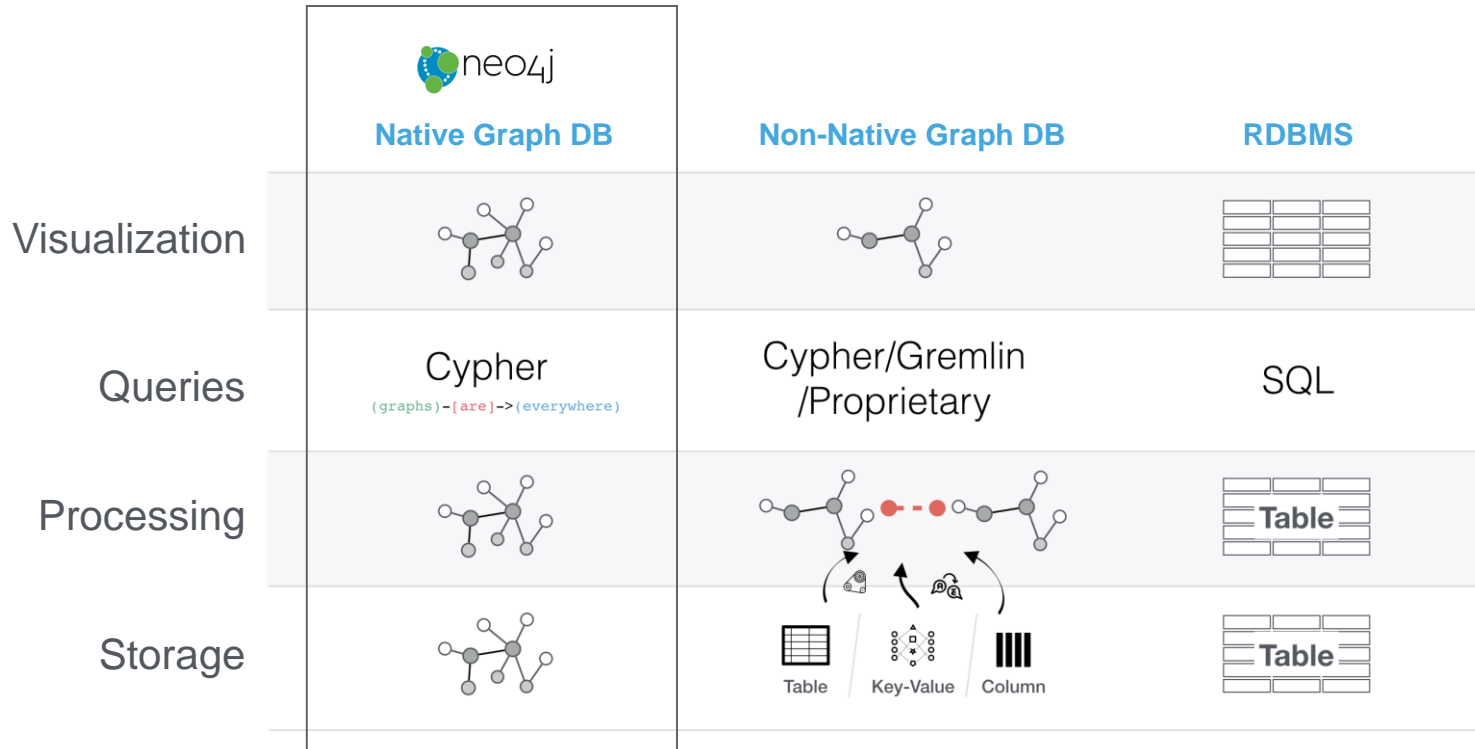
For next-gen infrastructure

Neo4j公司最初是在下面哪一个城市创建的：

- A. London
- B. San Mateo / 硅谷地区
- C. Frankfurt
- D. Malmo
- E. Sydney
- F. Beijing

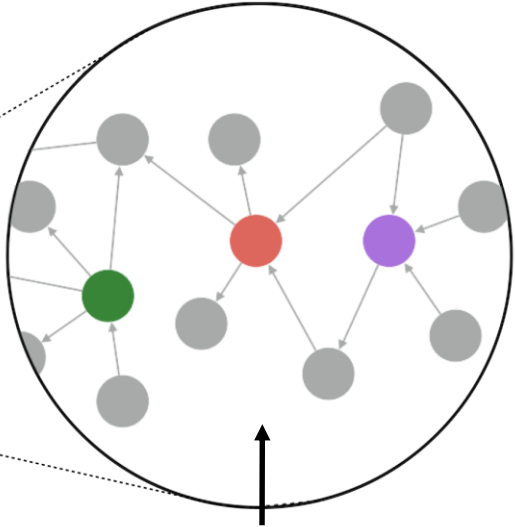
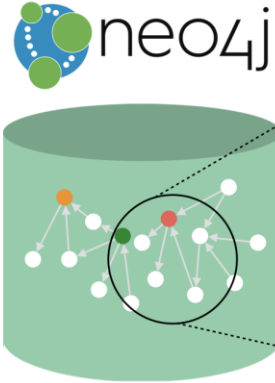
图数据库管理系统

How Neo4j Differentiates from other Databases



Optimized for graph workloads

Native Graph Architecture Advantage

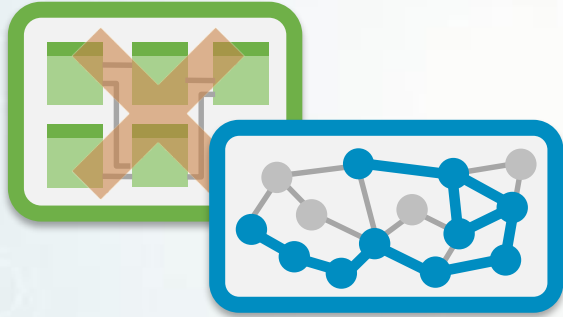


Index free adjacency

Unlike other database models Neo4j *connects* data as it is stored

Index-free adjacency ensures lightning-fast retrieval of data and relationships

Neo4j: Native Graph from the Start



Native graph storage

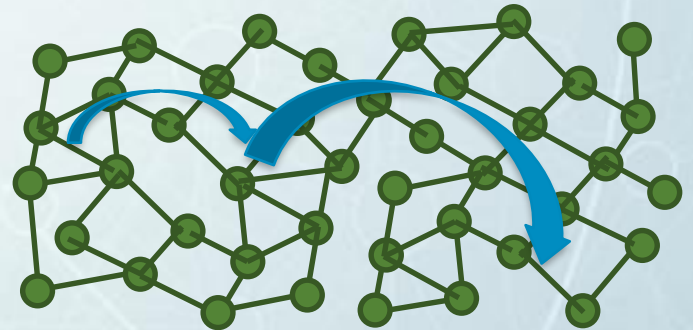
Optimized for real-time reads and ACID writes

- Relationships stored as physical objects, eliminating need for joins and join tables
- Nodes connected at write time, enabling scale-independent response times

Native graph querying

Memory structures and algorithms optimized for graphs

- Index-free adjacency enables 1M+ hops per second via in-memory pointer chasing
- Off-heap page cache improves operational robustness and scaling compared with JVM-based caches
- “Minutes to milliseconds” performance improvement



Traversal Speeds on Amazon Retail Dataset



Social Recommendation Example

Equivalent Cypher Query

```
MATCH (you)-[:BOUGHT]->(something)<-[:BOUGHT]-(other)-[:BOUGHT]->(reco)
WHERE id(you)={id}
RETURN reco
```

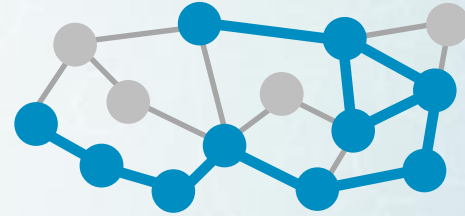


Threads	Hops per second
1	3-4 million
10	17-29 million
20	34-50 million
30	36-60 million

Relationship Queries Strain Traditional Databases



A single query can touch a *lot of data*



Queries can take non-sequential, *arbitrary paths* through data



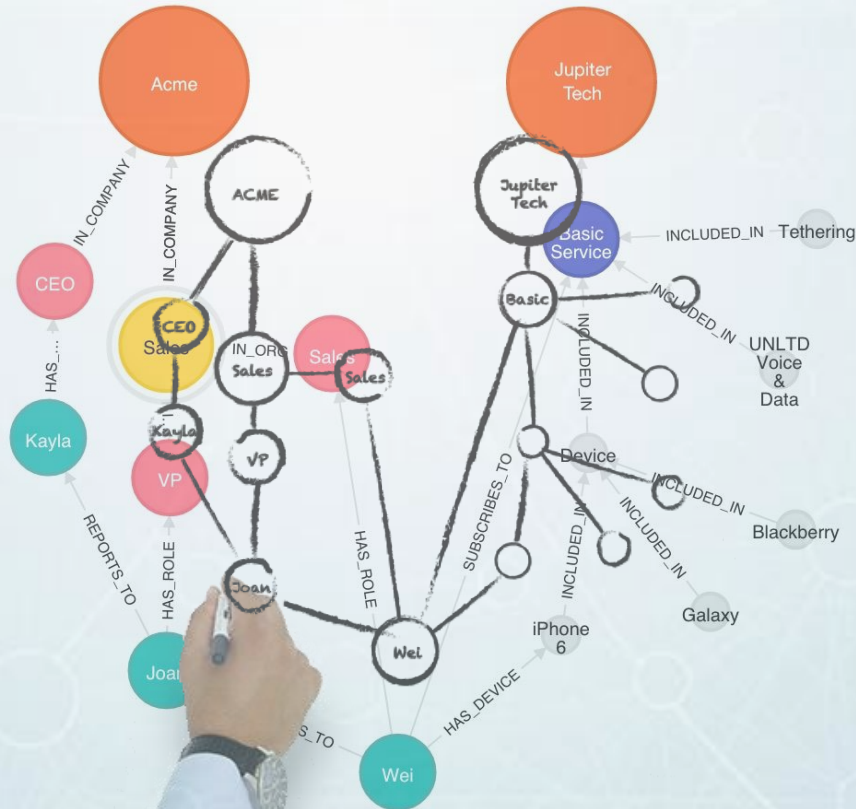
Real-time queries need speed and consistent response times



Queries must *run reliably* with *consistent results*

Neo4j Property Graph

The Whiteboard Model is the Physical Model



A unified view for ultimate agility

- Easily understood
- Easily evolved
- Easy collaboration between business and IT

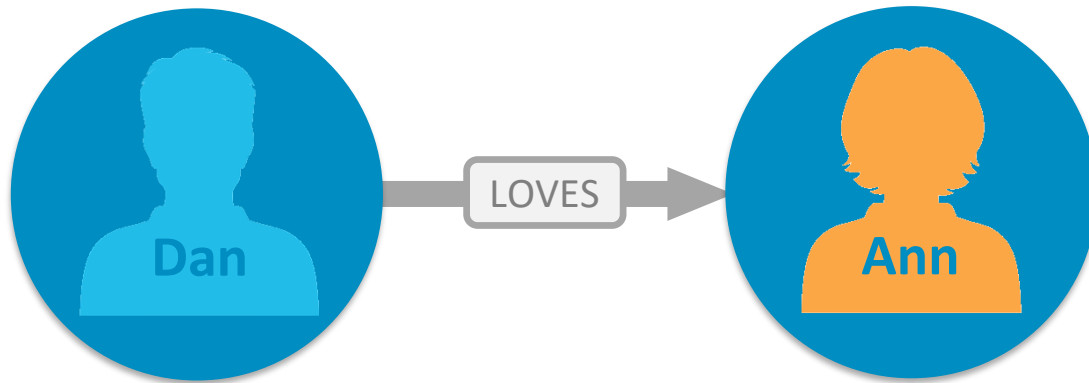
Cypher 查询语言

A pattern matching query language made for graphs.

- Declarative
- Expressive
- Pattern Matching

De facto industry standard query language for graph database.

Cypher: Express Graph Patterns



NODE

Relationship

NODE

```
(:Person { name:"Dan" } ) -[:LOVES]-> (:Person { name:"Ann" } )
```

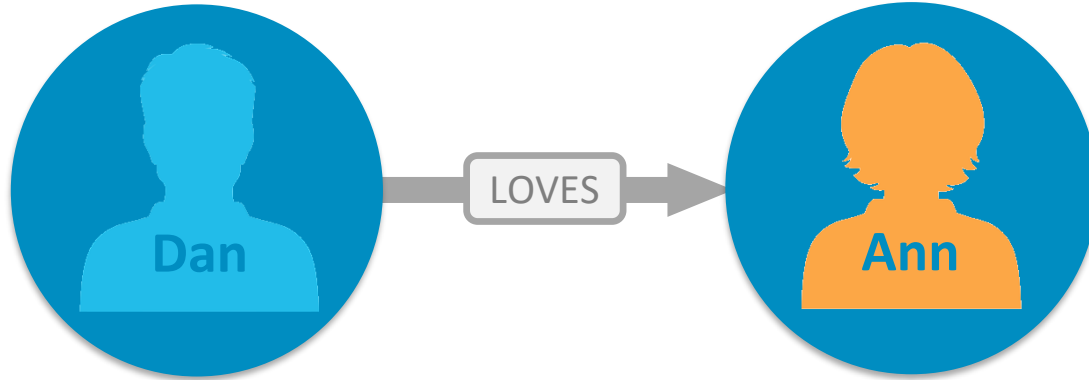
LABEL

PROPERTY

LABEL

PROPERTY

Cypher: CREATE Graph Patterns



NODE

Relationship

NODE

```
CREATE (:Person { name:"Dan" } ) -[:LOVES]-> (:Person { name:"Ann" } )
```

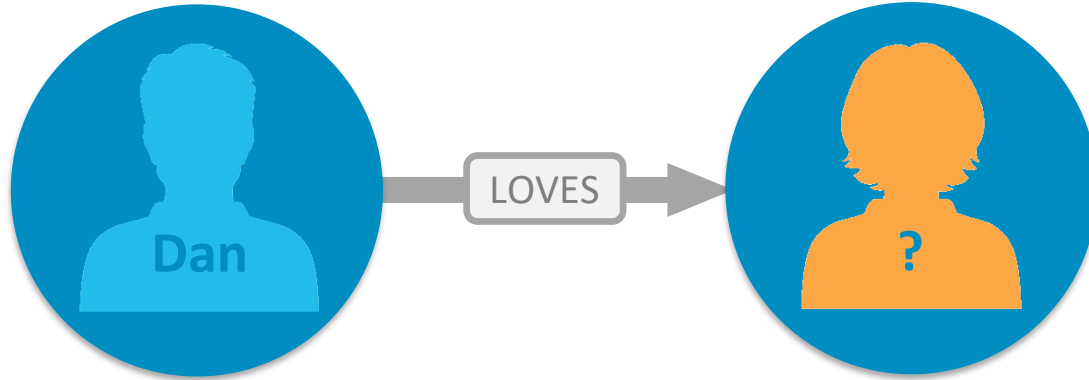
LABEL

PROPERTY

LABEL

PROPERTY

Cypher: MATCH Graph Patterns



NODE

Relationship

NODE

```
MATCH (p:Person { name:"Dan" } ) -[:LOVES]-> ( whom ) RETURN whom
```

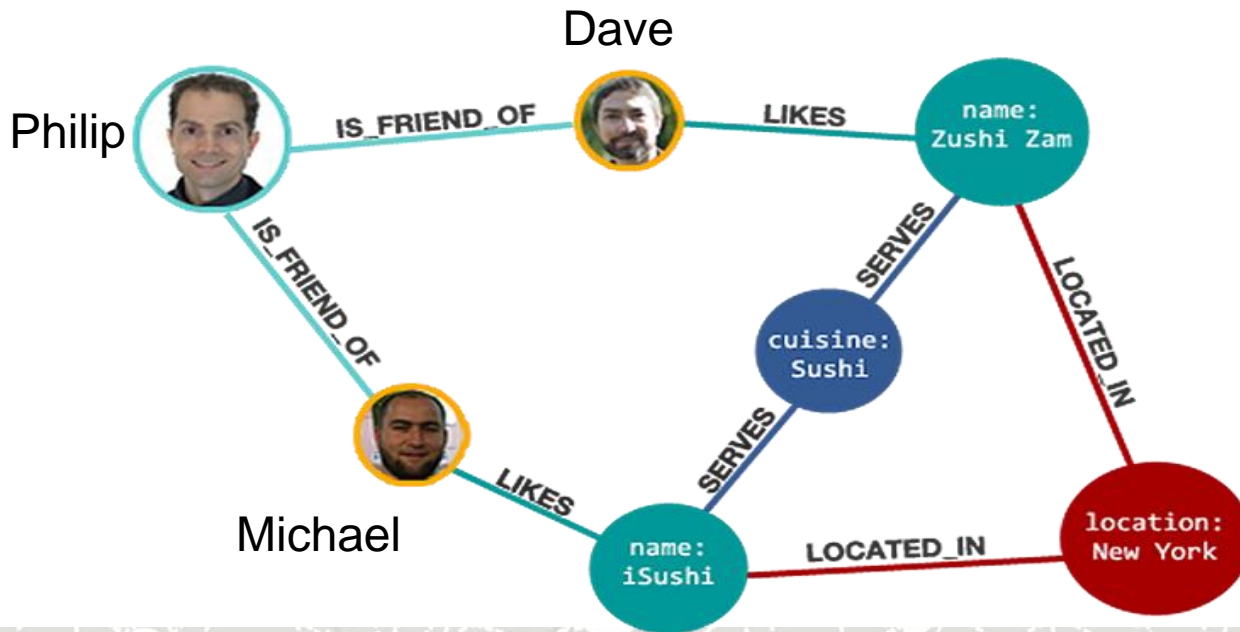
LABEL

PROPERTY

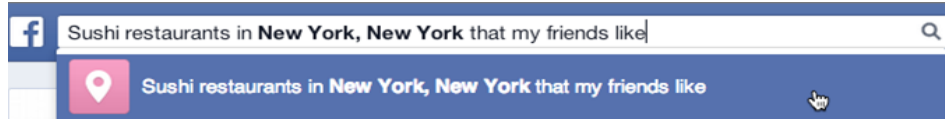
VARIABLE

```
WHERE p.Loves_Person_Id = whom.Id
```

A social recommendation



A social recommendation



```
MATCH (person:Person)-[:IS_FRIEND_OF]->(friend),  
        (friend)-[:LIKES]->(restaurant),  
        (restaurant)-[:LOCATED_IN]->(loc:Location),  
        (restaurant)-[:SERVES]->(type:Cuisine)  
WHERE person.name = 'Philip'  
AND loc.location='New York'  
AND type.cuisine='Sushi'  
RETURN restaurant.name
```



分析工具 – Neo4j Browser

Neo4j Browser



A GUI for power users and developers to access the graph database, and perform administrative tasks using Cypher.

- Running in web browser
- Installed as part of Neo4j database
- Service launched together with database
- To launch, access: `http://<hostname>:7474`

Neo4j Browser – Browse Database Schema



The screenshot shows the Neo4j Browser interface in a web browser window. The address bar shows `localhost:7474/browser/`. The interface is divided into a left sidebar and a main content area.

Database Information Sidebar:

- Node Labels:** ArtWork, Artist, Associate, Curriculum, Entity, Event, EventVenue, Group, Iconographic, Location, Material, MediaCategory, Organisation, Period, Person, Right, RightOwner, Subject, Technique, Venue.
- Relationship Types:** ABOUT, APPLIES_TO, BELONGS_TO, COLLECTED_AT, CONTAINS, CREATED, DEPICTS, EDITION_CREATED_AT, HAS_PARENT, INFLUENCED_BY, INTENDED_FOR, IS_INSTANCE_OF, MADE_BY, ORIGINAL_CREATED_AT, OWNS, PARTICIPATES_IN, PROVENANCE, RELATES_TO, UNKNOWN_ASSOCIATION.

Main Content Area:

- Query input: `$:play start`
- Learn about Neo4j:** A graph epiphany awaits you. Includes a small graph icon and a [Start Learning](#) button.
- Jump into code:** Use Cypher, the graph query language. Includes a code icon and a [Write Code](#) button.
- Monitor the system:** Key system health and status metrics. Includes a heart icon and a [Monitor](#) button.

Copyright © Neo4j, Inc 2002-2017

Neo4j Browser – Analyse Data



neo4j@bolt://localhost:7...
localhost:7474/browser/

```
$ MATCH (a:Artist{name:'Jackson POLLOCK'}) RETURN a;
```

Graph (39) Artist(1) Associate(1) Entity(3) Person(2) RightOwner(1) ArtWork(26) Right(1) Material(2) Technique(1) Organisation(1)

*(58) CREATED(25) INFLUENCED_BY(1) DEPICTS(2) OWNS(1) HAS_PARENT(12) APPLIES_TO(14) USES(2) MADE_BY(1)

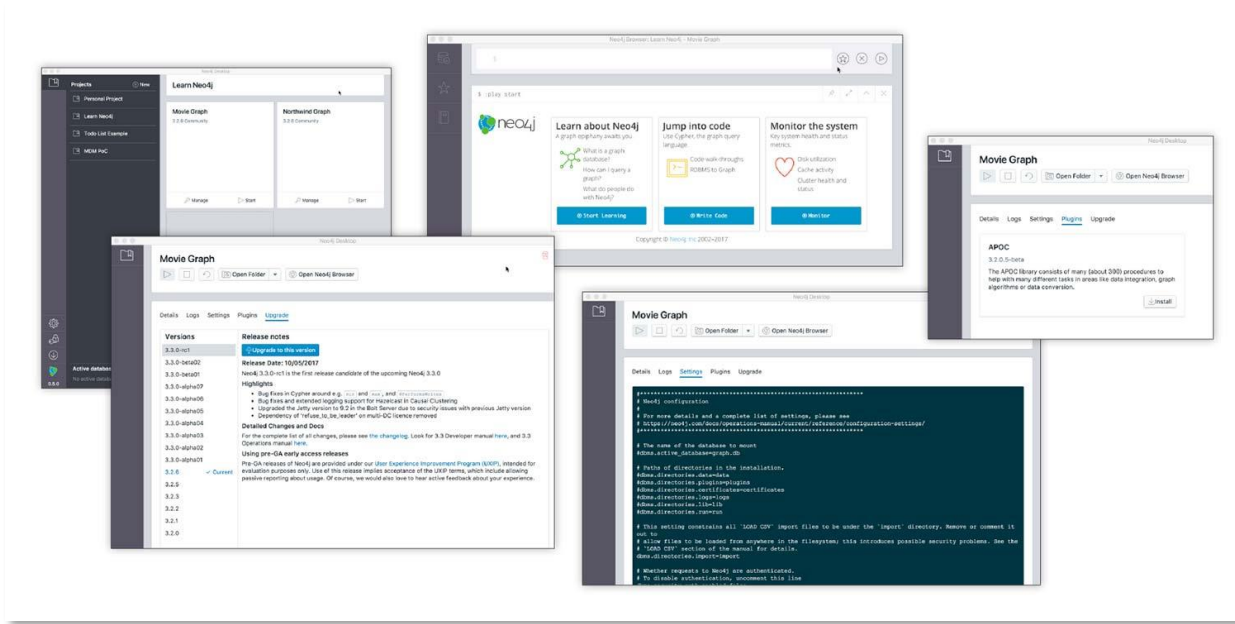
The graph visualization shows a central node for Jackson Pollock (purple circle) with numerous outgoing relationships. Key relationships include:

- Artists:** Created by Sainford MCCOY and BETTY FARGO.
- Materials:** Uses ink and paper.
- Artworks:** Created 25 artworks, including IDs 120467, 120466, 120462, 120464, 120483, 120458, 120449, 120452, 120454, 120450, 144701, 144700, 144702, 144691, 144703, 127789, 127720, 144690, 21430, and 73707.
- Other:** Has parent relationships with several other nodes.

APPLIES_TO <id>: 980469

Neo4j Desktop

Neo4j Desktop (Mac & Windows)



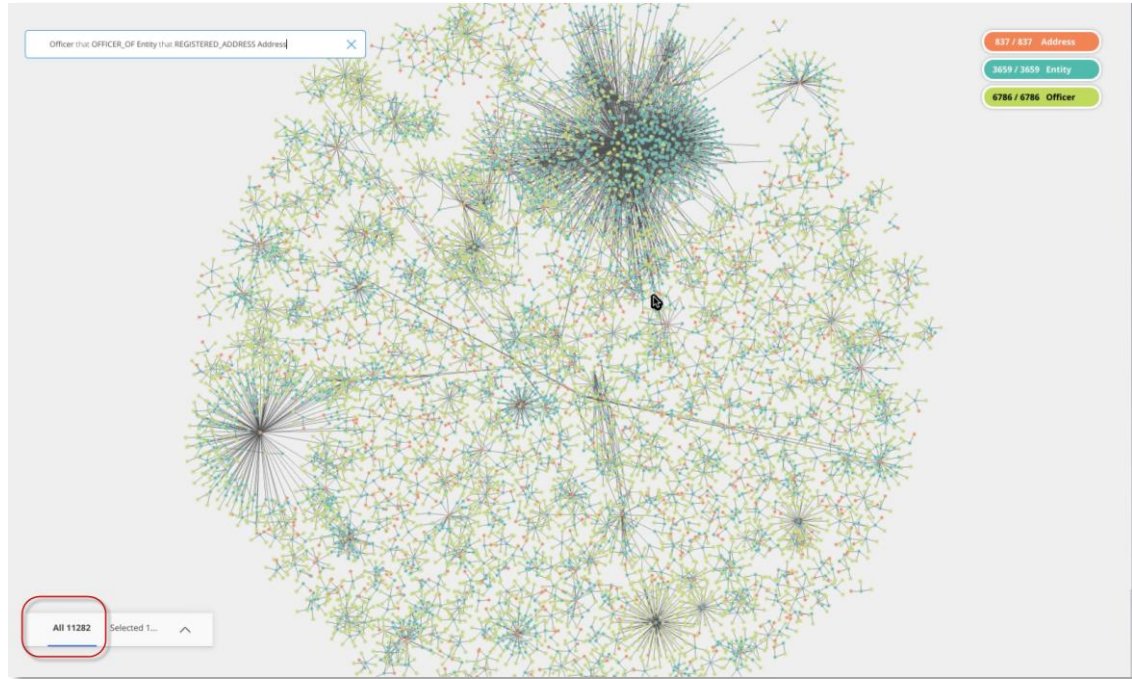
- The new mission control centre for developers
- Connecting to all your Neo4j databases within a single application
- Installing extensions like graph algorithms or ETL components
- Free to download, bundled with an Enterprise version of Neo4j for Developers

Visualization

Neo4j Bloom* – To be released in Q2 18



- A powerful tool for business users to explore data and connections
- Support of natural language input
- Support of editing nodes and relationships on the graph
- And many more...

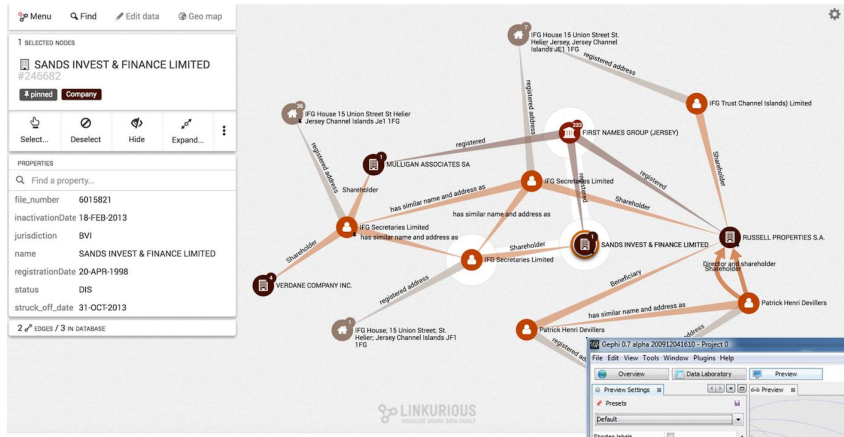


* Note: at this stage all features mentioned are for your references ONLY, and subject to change due to schedule and priority.

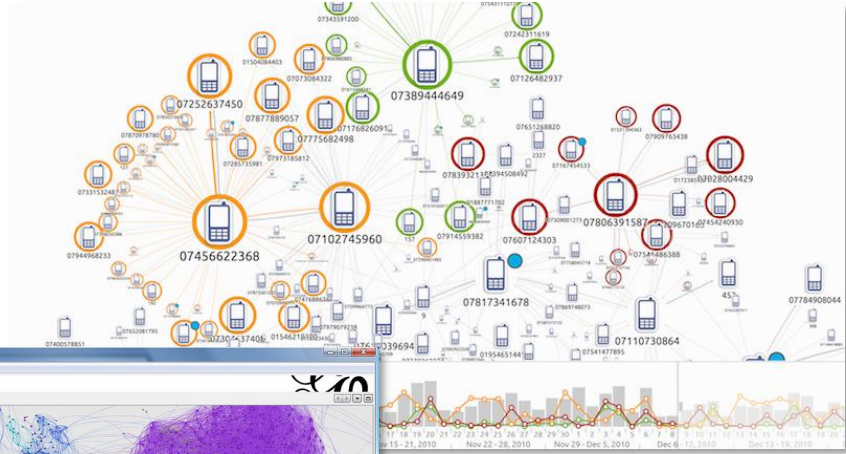
3rd Party Tools



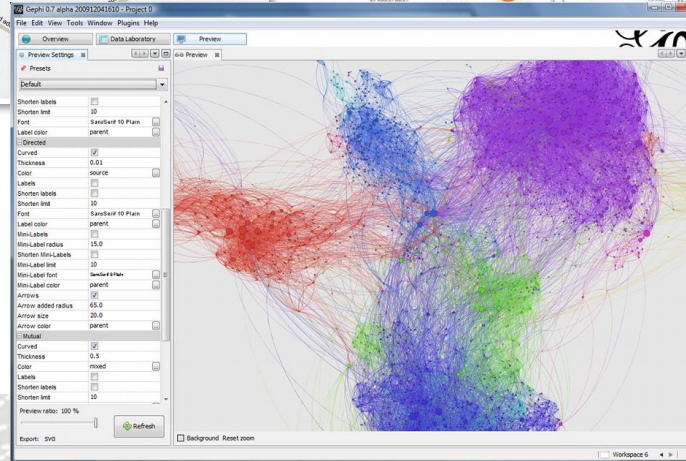
- Linkurious



- Keyline



- Gephi



Open Source Libraries



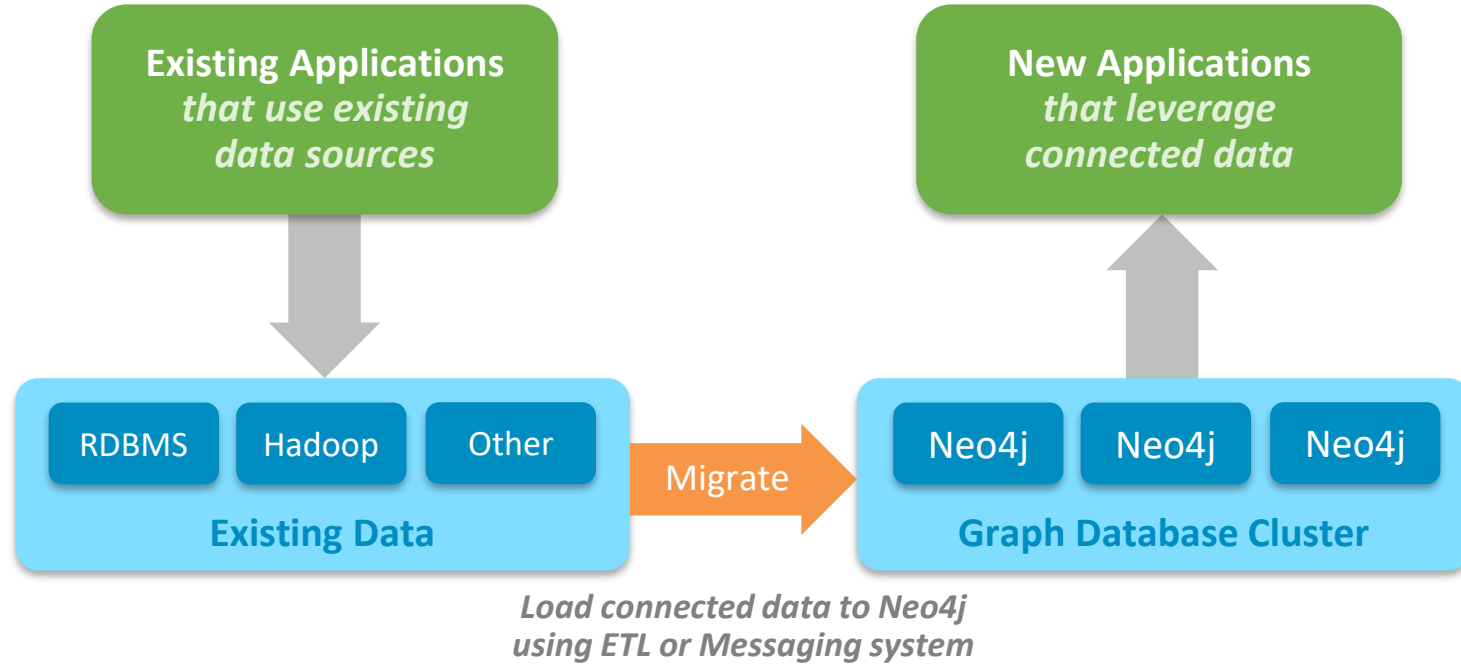
- Sigma
- NeoVis
- D3.js
- Cytoscape

Do a search on Github and you will find several more.

与企业的现有系统集成

Neo4j in your Environment

Usage Pattern 1 – 共存



Neo4j in your Environment

Usage Pattern 2 – 分治



Operational Applications



Operational Applications

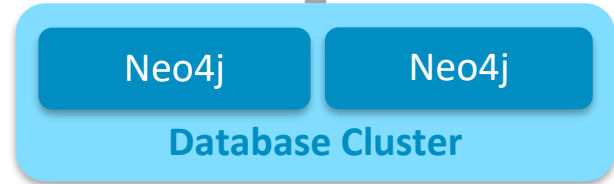


Analytic Applications



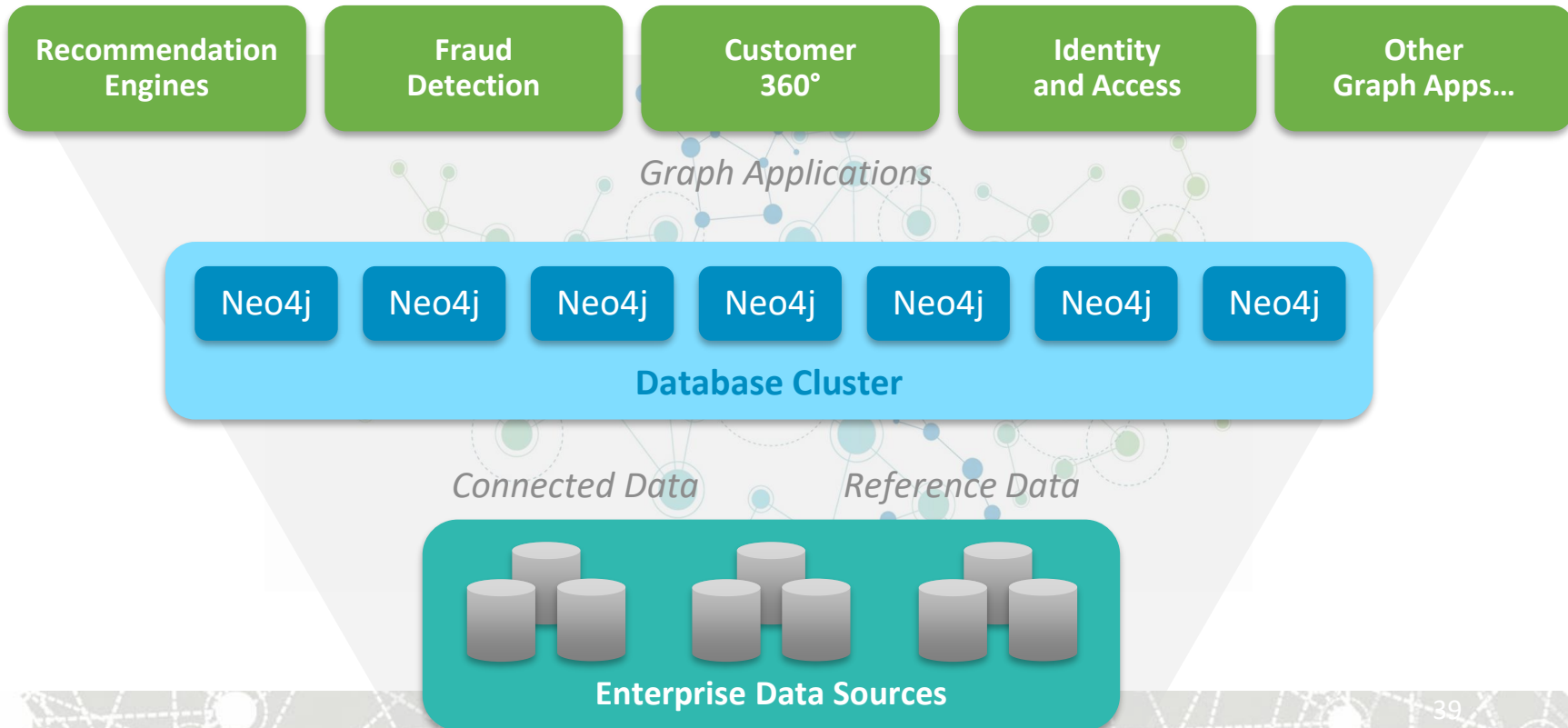
*Custom (D3, Vizjs)
Linkurious
Tom Sawyer*

Graph Visualization Applications



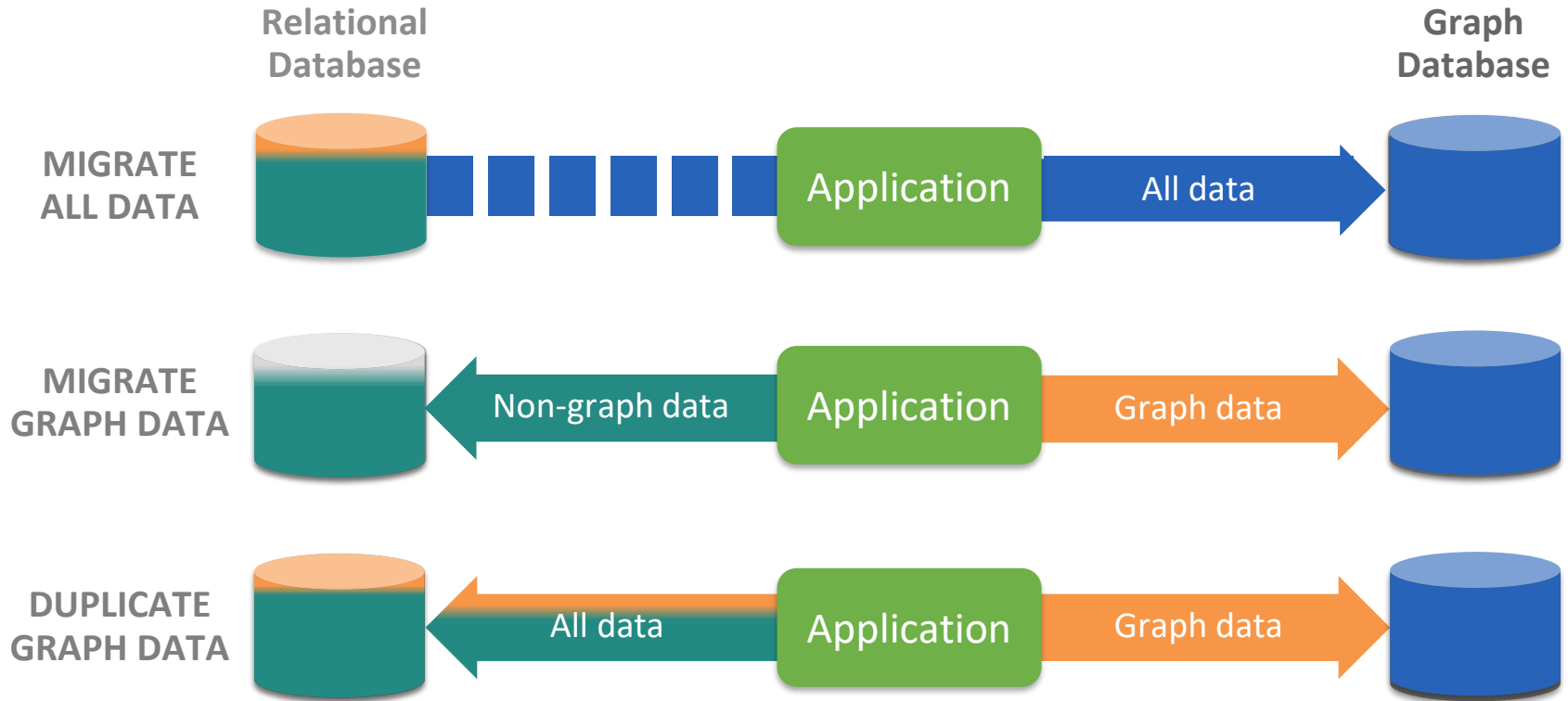
Neo4j in your Environment

Usage Pattern 3 – 联合



数据迁移/更新

Ways to migrate data to Neo4j



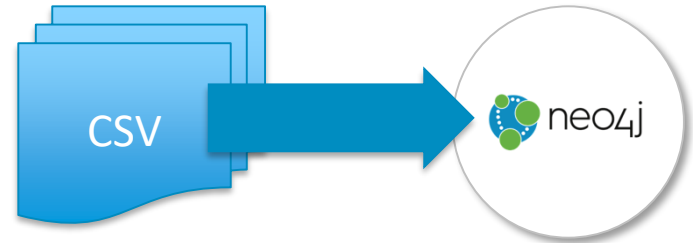
Getting Data into Neo4j: CSV



Cypher-Based “LOAD CSV” Capability

10M

- Transactional (ACID) writes
- Initial and incremental loads of up to 10 million nodes and relationships
- From HTTP and Files
- Power of Cypher
 - Create and Update Graph Structures
 - Data conversion, filtering, aggregation
 - Destructuring of Input Data
- Transaction Size Control
- Also via Neo4j-Shell / Cypher-Shell

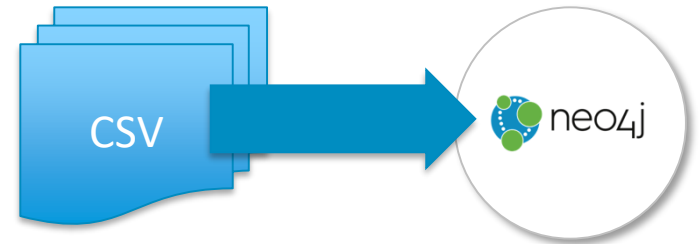


Getting Data into Neo4j: CSV



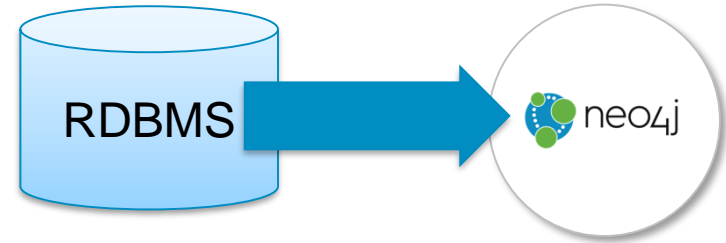
Command line bulk loader - neo4j-admin import

- For initial database population
- Scale across CPUs and disk performance
- Efficient RAM usage
- Split- and compressed file support
- For loads up to 10B+ records
- Up to 1M records per second



Neo4j ETL

- A GUI to create ETL mapping and jobs
- Loaded from Neo4j Desktop
- Reading data using JDBC from sources
- Landing data as CSV files
- Loading data into Neo4j by scheduled jobs



Mapping successful

Inspect your data

Choose a connection and then run the inspection.
Current connection: **Oracle Northwind Database**

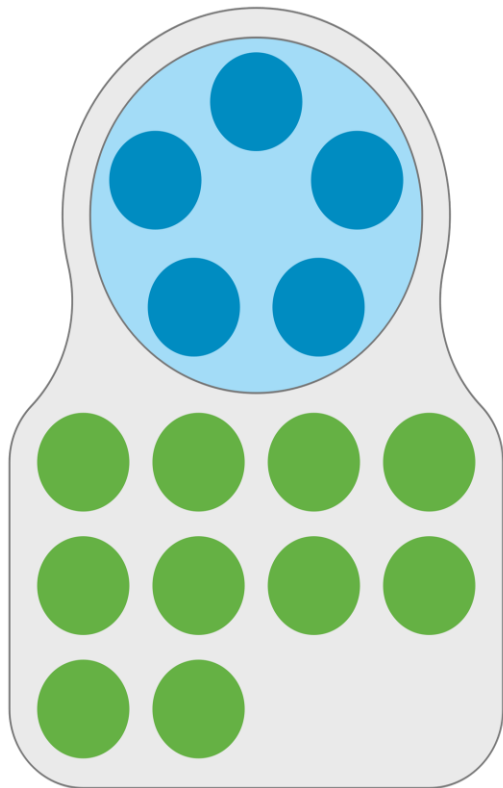
Type	Entity Name	Column Name	SQL Type	Neo4j Type	Primary Key
Node	Order	RegionID	NUMERIC	Double	<input checked="" type="checkbox"/>
Node	Shipper	RegionDescription	VARCHAR	String	<input type="checkbox"/>
Node	CustomerDemographic			Long	<input type="checkbox"/>
Node	CustomerCustomerDemo			Byte	<input type="checkbox"/>
Node	Region				<input type="checkbox"/>
Node	Product				<input type="checkbox"/>
Node	Territory				<input type="checkbox"/>
Node	Customer				<input type="checkbox"/>
Node	Employee				<input type="checkbox"/>
Node	Supplier				<input type="checkbox"/>
Node	Category				<input type="checkbox"/>
Relationship	SUPPLIER				<input type="checkbox"/>
Relationship	CUSTOMER				<input type="checkbox"/>
Relationship	EMPLOYEE				<input type="checkbox"/>

Your connections: Neo4j Exclude oracle | Oracle Northwind Database oracle

Start Mapping

The screenshot shows the Neo4j ETL interface. On the left, a table lists database entities and their mappings to Neo4j types. A dropdown menu is open for the 'Order' entity, showing options for Double, String, Long, and Byte. On the right, a graph visualization shows the relationships between entities like Order, Shipper, Customer, Product, Supplier, Category, Employee, and Territory. At the bottom, there are buttons for 'Start Mapping' and a list of connections.

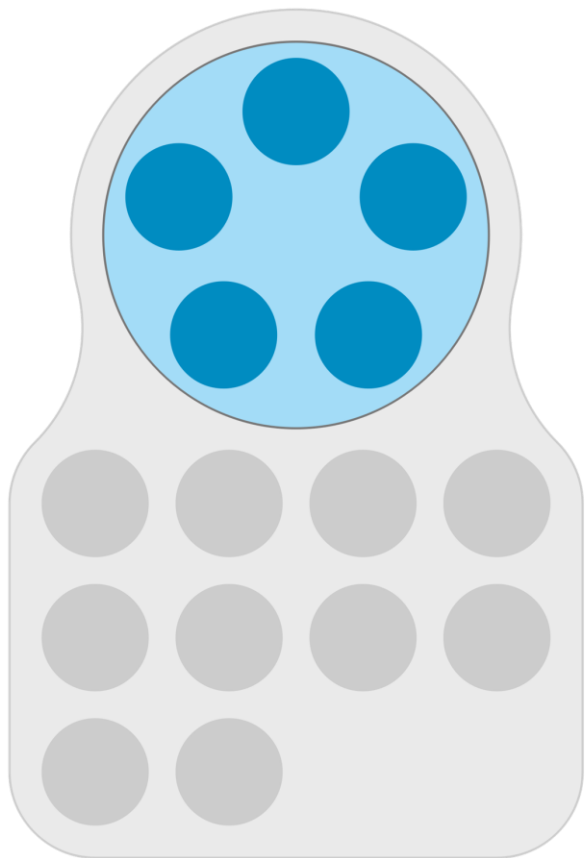
集群(Clustering)



Core

Replica

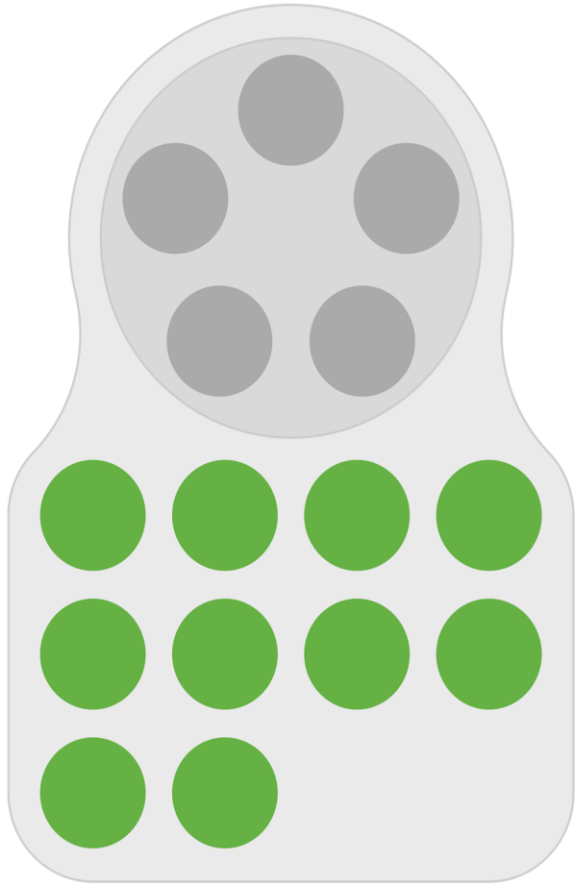




Core



- Small group of Neo4j databases
- Implements Consensus Commit
- Responsible for data safety



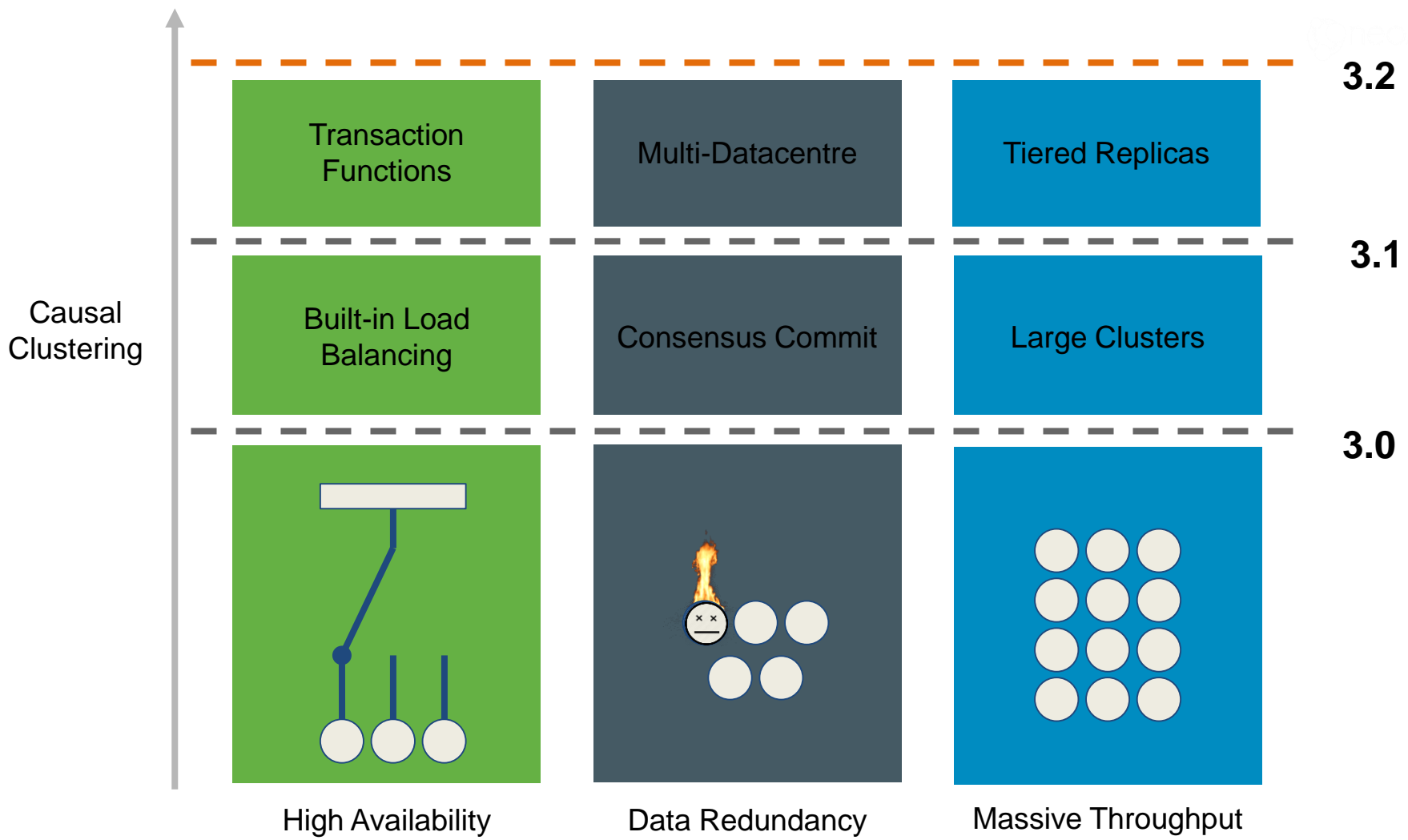
Replica

- For massive query throughput
- Read-only replicas
- Not involved in Consensus Commit
=> less overhead
- Disposable, suitable for auto-scaling

Causal Clustering - Features



- Two Zones – Core + Edge
- Group of **Core Servers** – Consistent and Partition tolerant (CP)
 - Transactional Writes
 - Quorum Writes, Cluster Membership, Leader via Raft Consensus
- Scale out with **Read Replicas**
- Smart **Bolt Drivers** with
 - Routing, Read & Write Sessions
 - Causal Consistency with Bookmarks



部署(Deployment)

Neo4j Supported Platforms



On-Premise Platforms



Linux



IBM POWER



macOS
For Development

Cloud Platforms and Containers



docker

... and others

开发(Development)

Neo4j的编程语言驱动

- Hosted on GitHub
- Apache licensed
- Versioned and released independently of Neo4j
- One driver can support multiple server versions
- Quarterly releases (on average)
- Published to Maven Central, PyPI, npm, etc

Neo4j Drivers

Java

```
<dependency>  
  <groupId>org.neo4j.driver</groupId>  
  <artifactId>neo4j-java-driver</artifactId>  
  <version>X.Y.Z</version>  
</dependency>
```

Python

```
pip install neo4j-driver
```

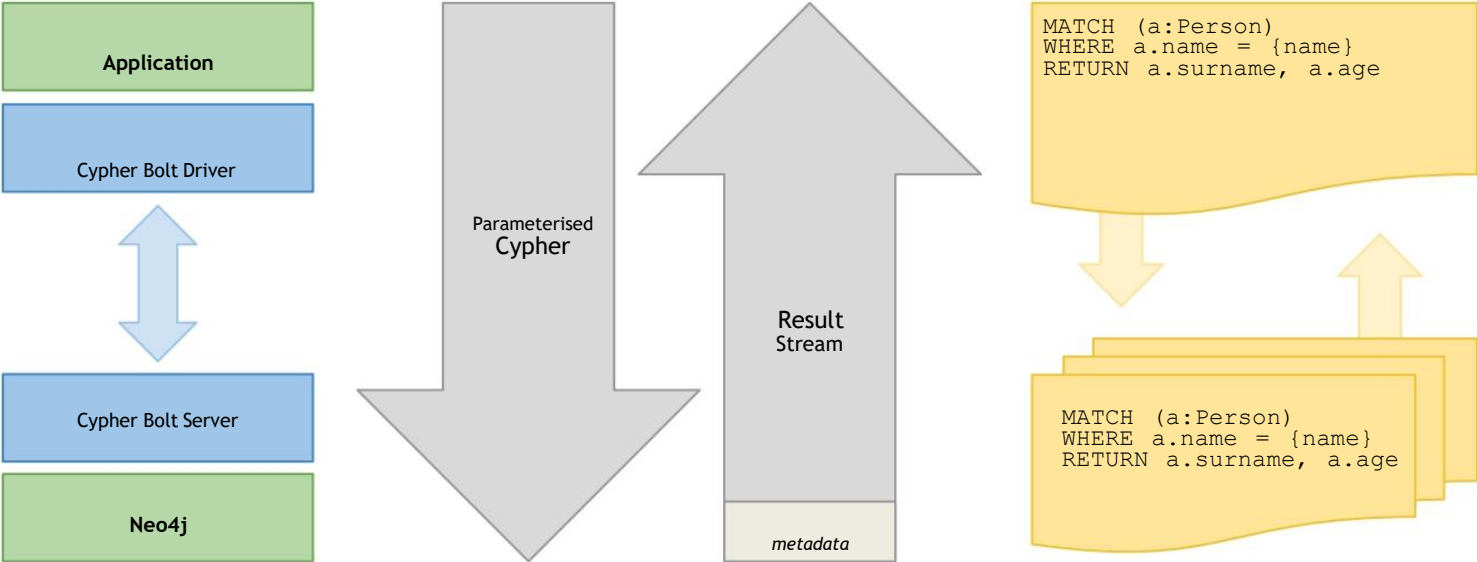
.NET

```
PM> Install-Package Neo4j.Driver
```

Javascript

```
npm install neo4j-driver
```

Data Flow



使用Java访问Neo4j数据库的方法

- Official Driver
- User Defined Functions & Procedures
- Core API
- Neo4j-OGM
- Spring Data Neo4j

← Unmanaged Extensions

Resources

- <https://neo4j.com/developer/java/>
- <https://neo4j.com/docs/java-reference/current>
- <https://neo4j.com/docs/developer-manual/current/extending-neo4j/procedures/>
- <https://neo4j.com/docs/operations-manual/current/security/securing-extensions/>
- <https://projects.spring.io/spring-data-neo4j/>

运营支持

Currently running queries



List the running queries :

CALL dbms.listQueries()

\$ CALL dbms.listQueries()													
Flows	queryId	username	query	parameters	startTime	elapsedTime	connectionDetails	metaData					
	query-88	neo4j	MATCH (m), (n), (o), (p), (q) RETURN COUNT(*) AS count	(empty)	2016-10-05T16:10:31.773+01:00	00:00:11.170	server-session http 127.0.0.1 /db/data/transaction/67/commit neo4j	(empty)					
Code	query-90	neo4j	CALL dbms.listQueries()	(empty)	2016-10-05T16:10:42.943+01:00	00:00:00.001	server-session http 127.0.0.1 /db/data/transaction/69/commit neo4j	(empty)					
Returned 2 records in 10 ms.													

Kill running queries



We can also kill the query manually using the kill query procedure:

\$ CALL dbms.listQueries()								
	queryId	username	query	parameters	startTime	elapsedTime	connectionDetails	metaData
Flows	query-88	neo4j	MATCH (m), (n), (o), (p), (q) RETURN COUNT(*) AS count	(empty)	2016-10-05T16:10:31.773+01:00	00:00:11.170	server-session http 127.0.0.1 /db/data/transaction/67/commit neo4j	(empty)
Text	query-90	neo4j	CALL dbms.listQueries()	(empty)	2016-10-05T16:10:42.943+01:00	00:00:00.001	server-session http 127.0.0.1 /db/data/transaction/69/commit neo4j	(empty)
Code	Returned 2 records in 10 ms.							

```
CALL dbms.killQuery("query-88")
```

Enabling JMX metrics



// default setting for enabling all supported metrics

```
metrics.enabled=true
```

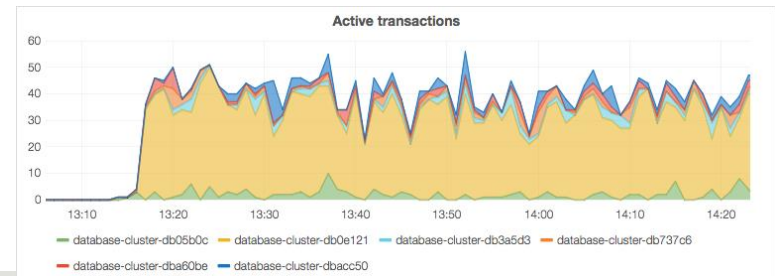
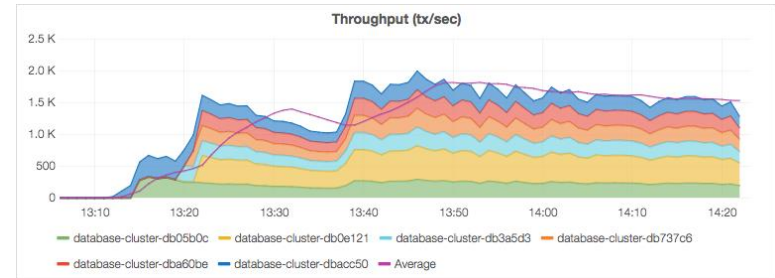
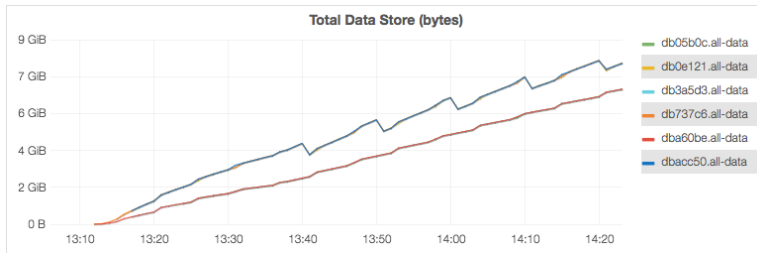
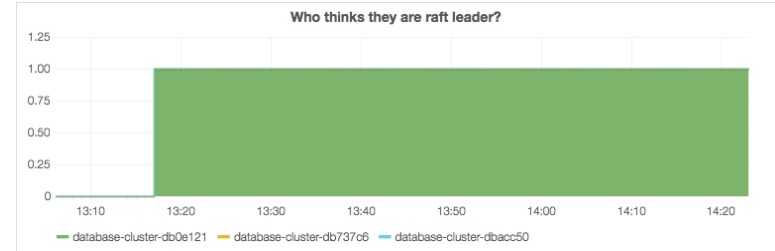
```
metrics.graphite.enabled=true
```

```
metrics.graphite.server=<ip>:2003
```

```
metrics.graphite.interval=<how often to send data, defaults to 3s>
```

```
metrics.prefix=<Neo4j instance name, e.g. wwwneo1>
```

Grafana as the Monitoring dashboard



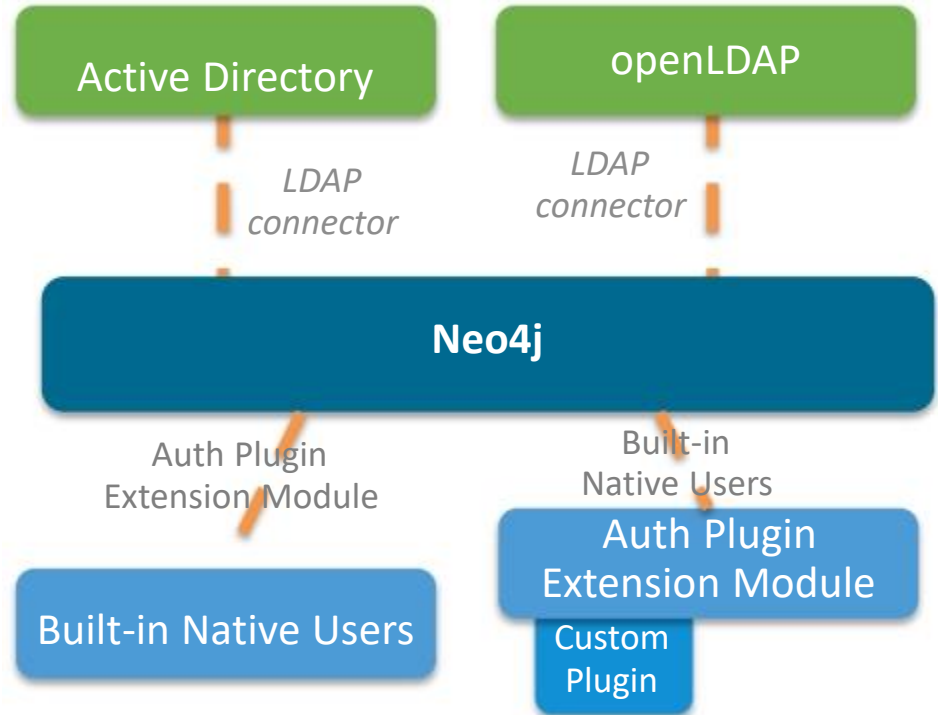
用户自定义过程

安全性

Flexible Authentication Options

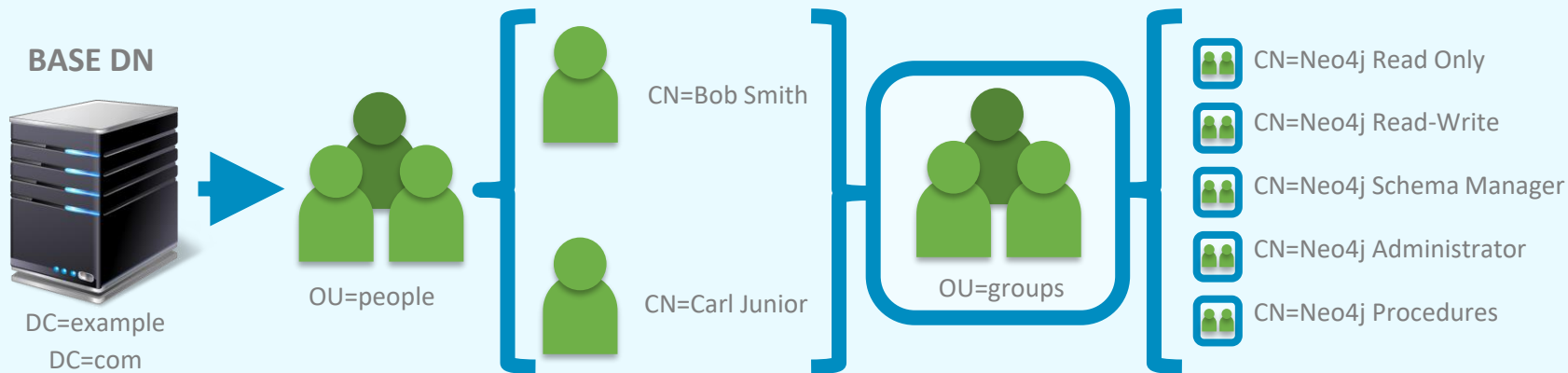
Choose authentication method

- Built-in native users repository
Testing/POC, single-instance deployments
- LDAP connector to Active Directory or openLDAP
Production deployments
- Custom auth provider plugins
Special deployment scenarios



Flexible Authentication Options

LDAP Group to Role Mapping



Configure the actual mapping between groups in the LDAP and roles in Neo4j

```
dbms.security.ldap.authorization.group_to_role_mapping= \  
  "CN=Neo4j Read Only,OU=groups,DC=example,DC=com"      = reader; \  
  "CN=Neo4j Read-Write,OU=groups,DC=example,DC=com"     = publisher; \  
  "CN=Neo4j Schema Manager,OU=groups,DC=example,DC=com" = architect; \  
  "CN=Neo4j Administrator,OU=groups,DC=example,DC=com" = admin; \  
  "CN=Neo4j Procedures,OU=groups,DC=example,DC=com"    = allowed_role
```

./conf/neo4j.conf

业务应用场景

Common Graph Technology Use Cases



Real-Time
Recommendations

Walmart 



Fraud
Detection



Network &
IT Operations



Master Data
Management



Knowledge
Graph

AirBnb



Identity & Access
Management



Neo4j图数据库应用领域广泛，然而以下哪些领域并不适合使用图数据库？

- A. 社交网络分析
- B. 在线推荐引擎
- C. 欺诈团伙检测
- D. 财务报表
- E. 身份管理和访问控制
- F. 媒体数据管理和存储
- G. 知识图谱
- H. 数据地图/元数据管理
- I. 网路和设备管理

技术应用场景

Neo4j Technology Use Cases*

- Integration with Hadoop (and other No-SQL databases)
- Containerization / Docker images of Neo4j
- Cloud deployment: AWS and Azure
- Natural Language Processing
- Integration with Elasticsearch
- Large scale and complex calculation: Spark + Neo4j
- Artificial Intelligence: R + Neo4j
- Internet of Things
- Machine Learning
-

** Demo available on request*

Neo4j的社区

Largest pool of graph technologists



Downloads

10M+

3M+ from Neo4j Distribution
7M+ from Docker



Events

400+

Approximate Number of
Neo4j Events per Year



Trained Developers

50k+

Trained/certified Neo4j
professionals



Meetups

50k+

Number of Meetup
Members Globally

- Neo4j中文社区: neo4j.com.cn
- 中国图数据库技术群
- 中国用户和技术聚会

Finding Help

- Website: <http://neo4j.com>
- Twitter: <https://twitter.com/neo4j>
- Slack: search for neo4j
- Technical questions: <http://stackoverflow.com>
- APOC on Github: <https://github.com/neo4j-contrib/neo4j-apoc-procedures>
- QQ Group: Neo4j Online Community / 547190638 (for Chinese users)

Contact us via email: apac@neo4j.com (Asia Pacific regions)

是时候活动一下筋骨、放松一下大脑了。。。。

THANK YOU!