Graph Computing with JanusGraph

Jason Plurad
Open Source Developer & Advocate

IBM CODE

JanusGraph

- Established in January 2017
- Scalable graph database distributed on multi-machine clusters with pluggable storage and indexing
- Vendor-neutral, open community with open governance
  - Founders: Expero, Google, Grakn, Hortonworks, IBM

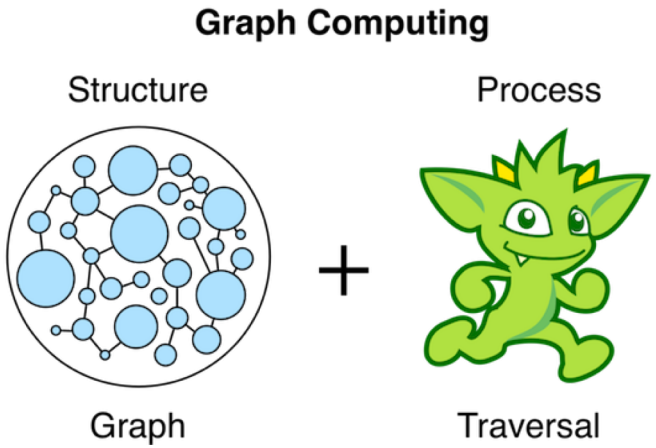
JanusGraph™	
Maintainer	The Linux Foundation
License	Apache
Releases	0.3.0 planned 2Q 2018
https://janusgraph.org	

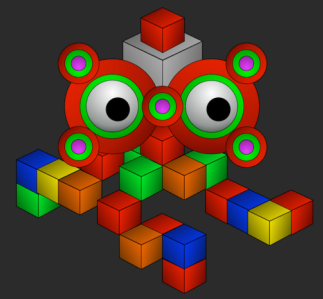
JanusGraph Community

- **Contributors**
  - 49 total
- **Committers**
  - 14 initial, 6 added
- **Technical Steering Committee**
  - 6 initial, 2 added
- **Issues**
  - 287 open, 352 closed
- **Open source projects**
  - Apache Atlas
  - Open Network Automation Platform (ONAP)
- **Member companies**
  - Amazon
  - Huawei
  - Linkurious
  - Netflix
  - Newforma
  - Orchestral Developments
  - Uber
- **In production**
  - Celum
  - Finc
  - G-Data
  - IBM Cloud
  - Times Internet

Apache TinkerPop

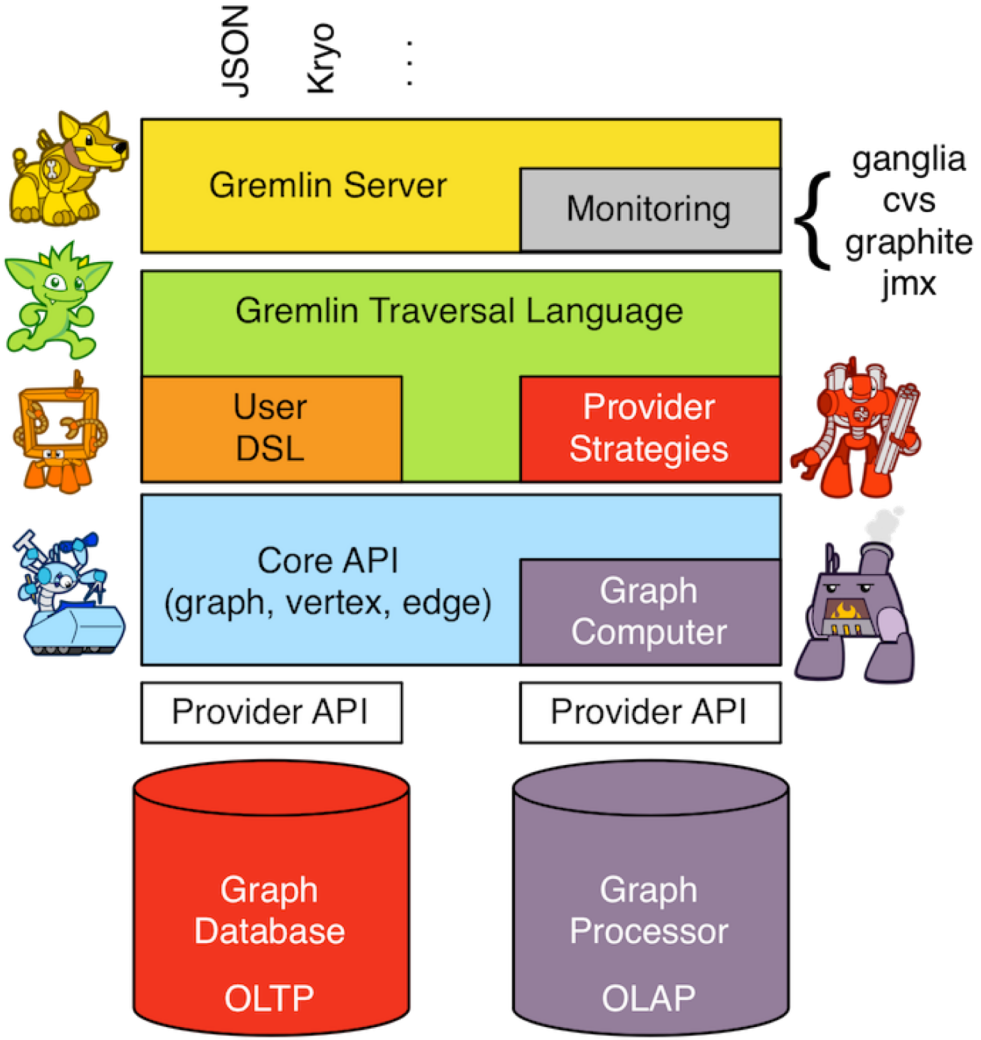
- Established in 2009
  - Apache incubator in 2015
  - Top-level project in 2016
- Open source, vendor-agnostic, graph computing framework
- Gremlin graph traversal language





Apache TinkerPop™	
Maintainer	The Apache Software Foundation
License	Apache
Releases	3.3.3 May 2018
https://tinkerpop.apache.org	

Graph System Integration

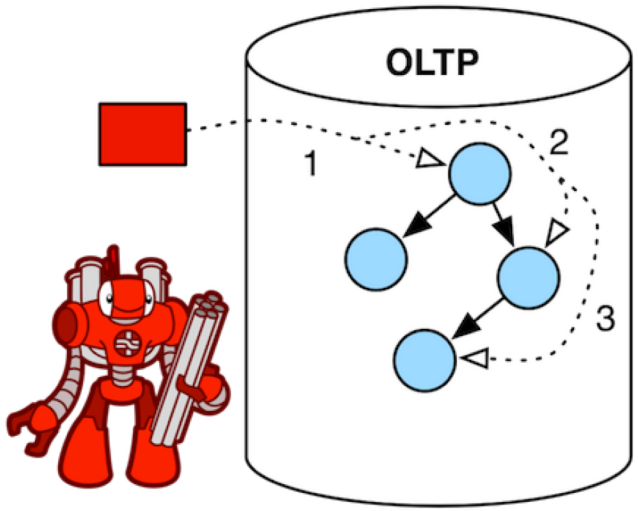


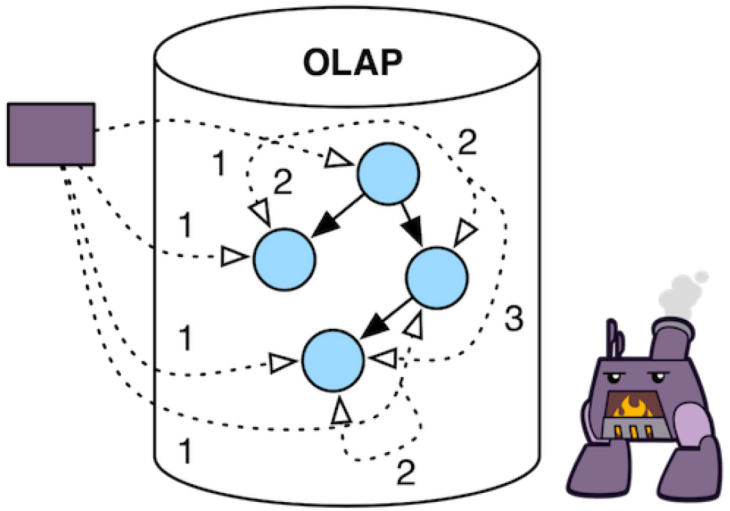
https://tinkerpop.apache.org/docs/current/tutorials/getting-started/

Graph Provider Landscape

- **Vendor**
  - Neo4j
  - OrientDB
  - Stardog
  - DataStax DSE Graph
  - Azure Cosmos DB
  - Amazon Neptune
- **Open Source**
  - Apache S2Graph
  - Blazegraph
  - HGraphDB
  - JanusGraph
  - TinkerGraph
  - UniPop

- **Open Source (OLAP)**
  - Apache Giraph
  - Apache Spark





Graph Data Use Cases

Social network analysis

Configuration management database

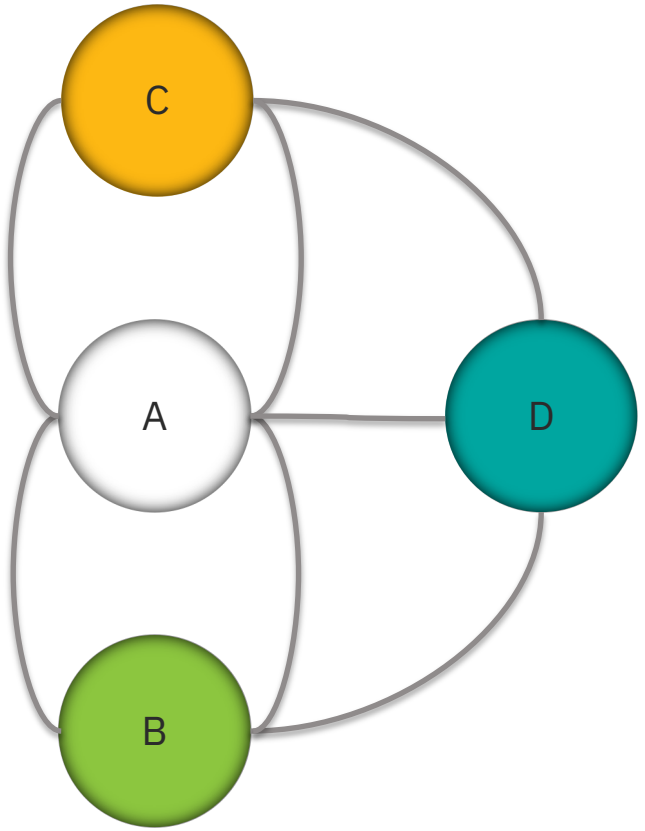
Master data management

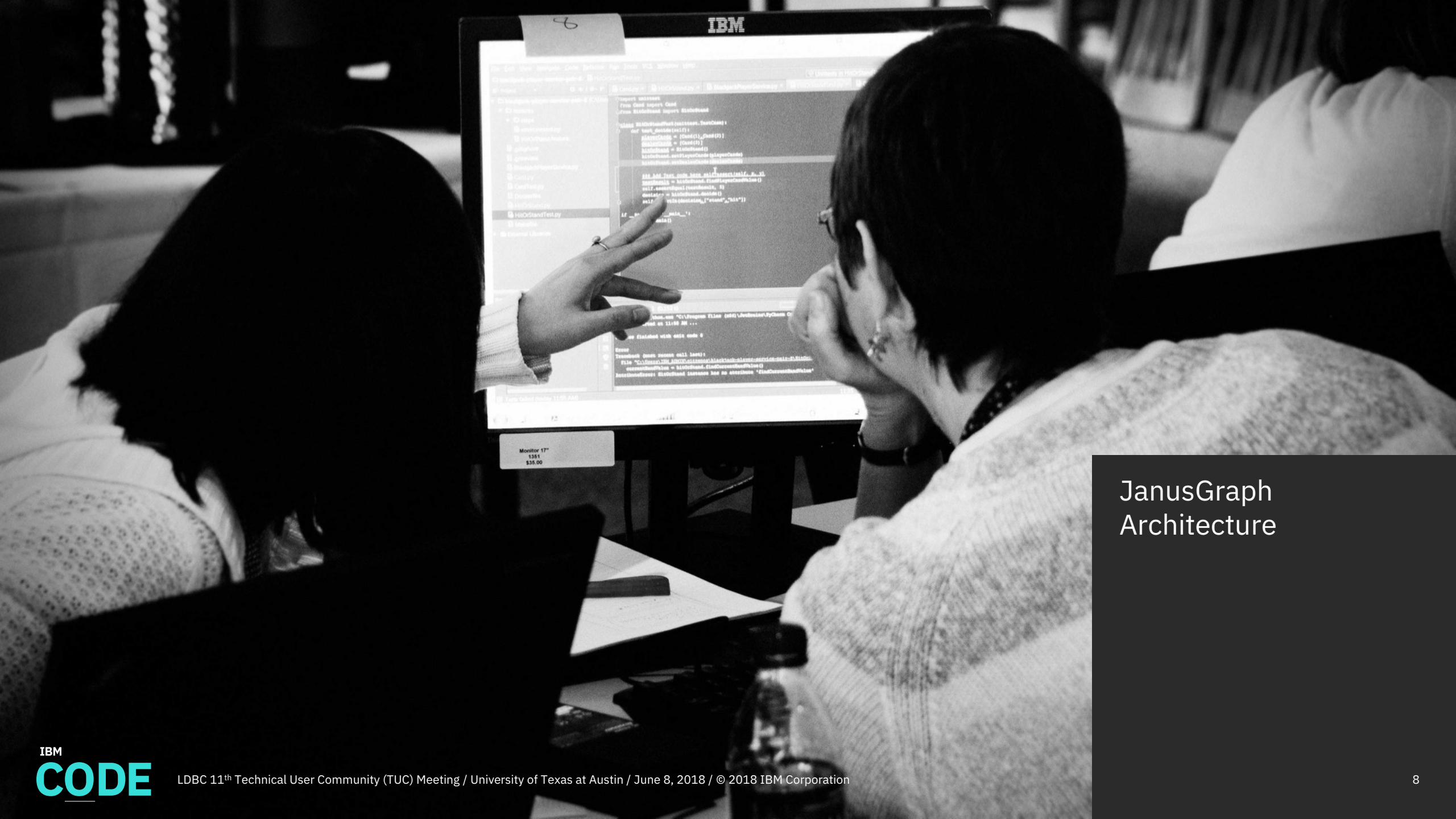
Recommendation engines

Knowledge graphs

Internet of things

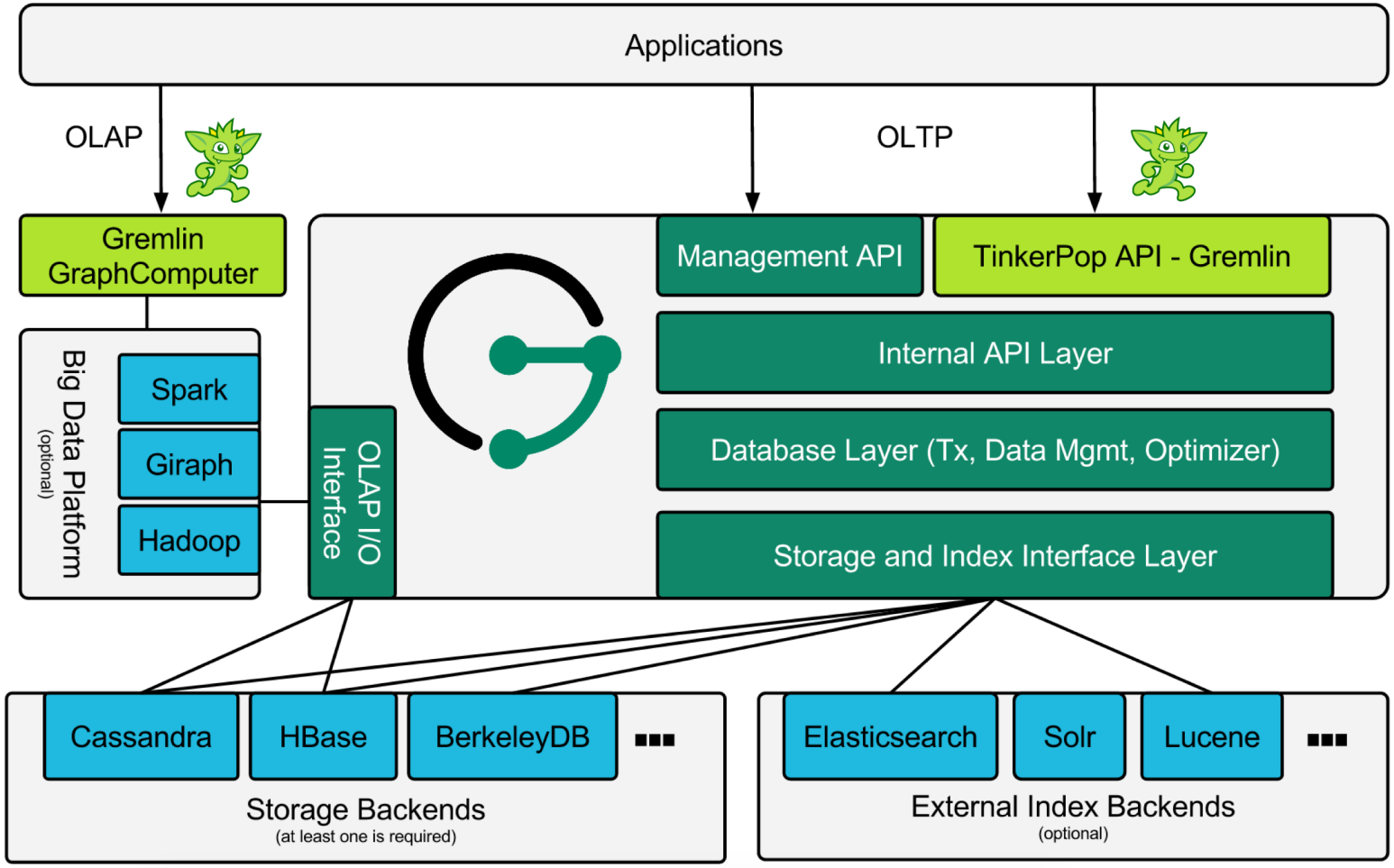
Cyber security attack analysis





JanusGraph Architecture

JanusGraph Architecture



http://docs.janusgraph.org/latest/arch-overview.html

Pluggable Storage Backends

Apache Cassandra

Apache HBase

Oracle BerkleyDB

ScyllaDB

Google Cloud Bigtable

Amazon DynamoDB

Tupl

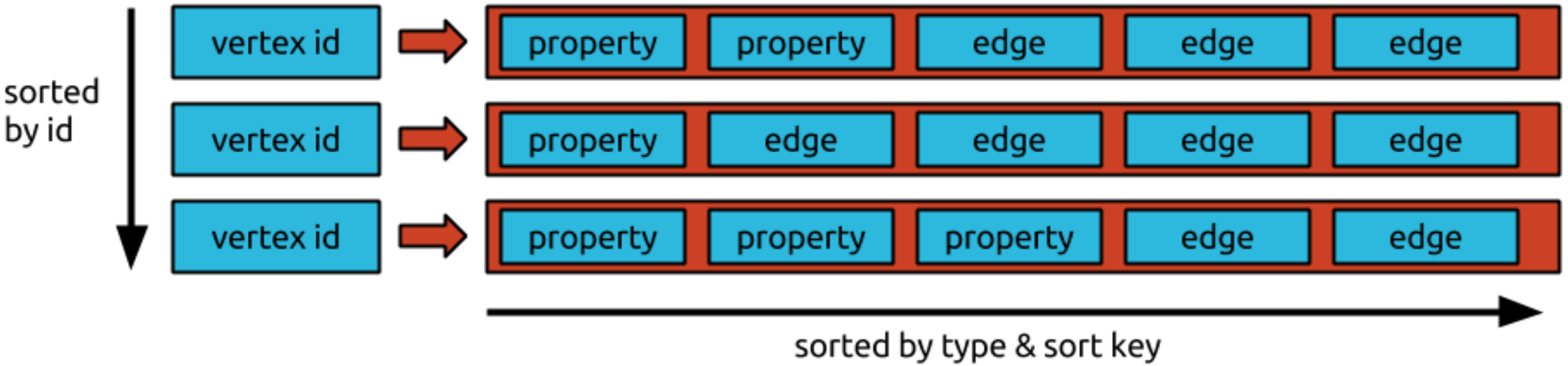
Instagram Rocksandra

Couchbase ForestDB

Apple FoundationDB

YugaByte DB

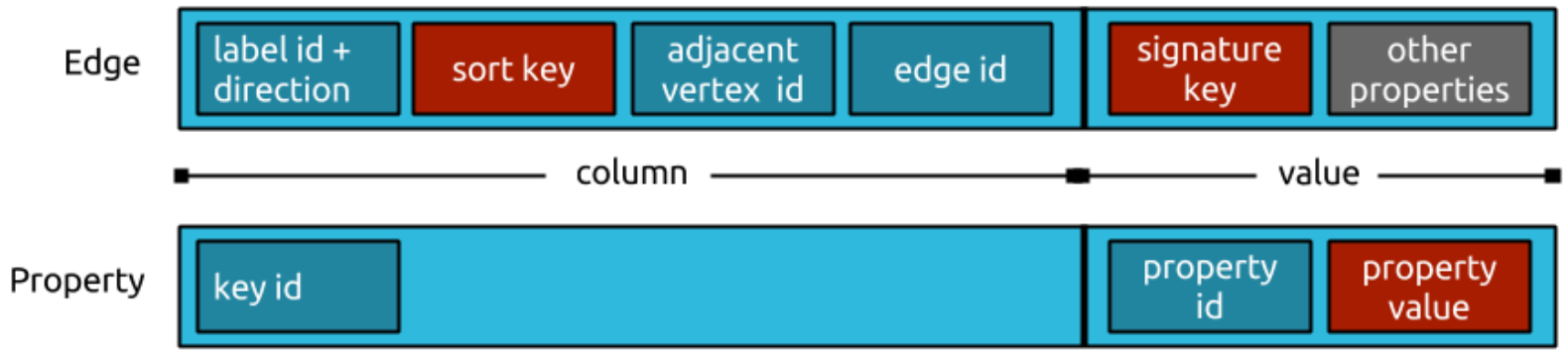
Storage Format



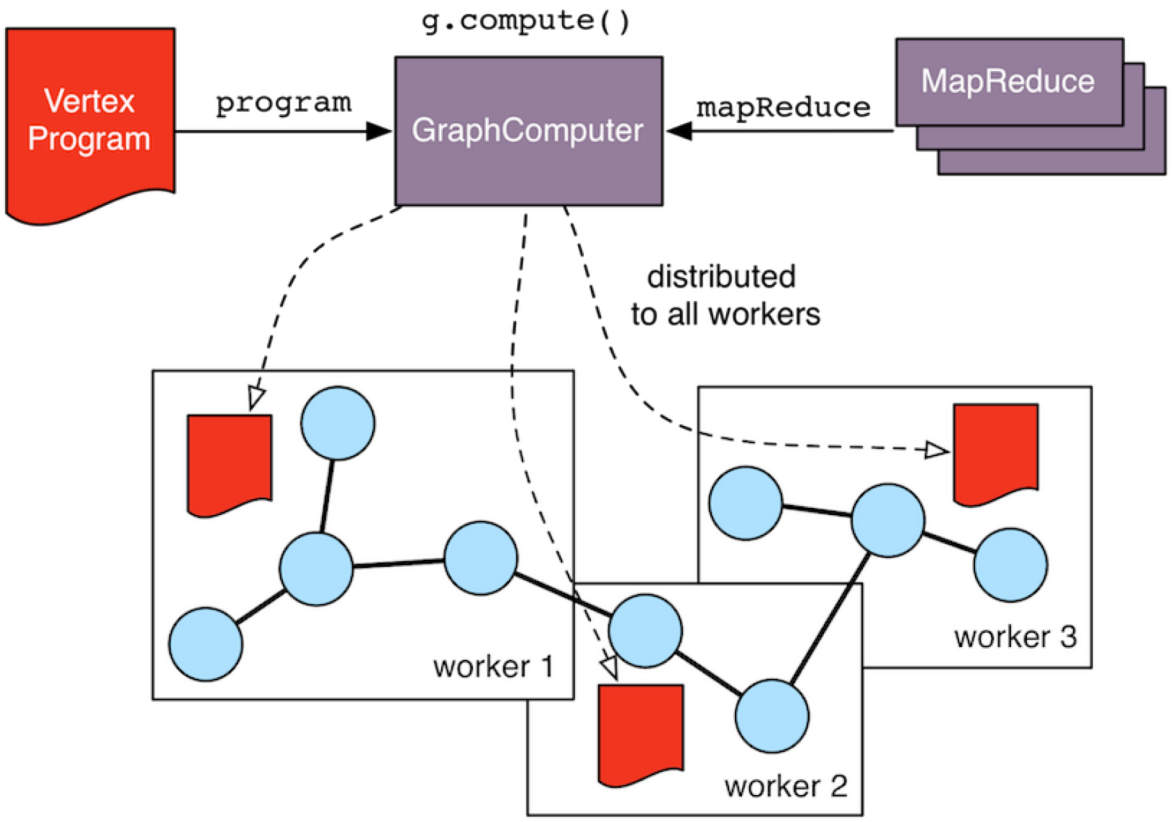
https://docs.janusgraph.org/latest/data-model.html

Indexes

- **Composite index**
  - Exact match search
- **Vertex-centric index**
  - Local index per vertex
- **Mixed index**
  - Full text search
  - Numerical range search
  - Geospatial search



OLAP with Graph Computer and Vertex Programs



- **Input formats:** Cassandra, HBase
- **Execution engines:** Spark, Giraph
- **Vertex programs**
  - Traversal
  - Bulk loader
  - Bulk dumper
  - Page rank
  - Peer pressure

Open Collaboration Towards JanusGraph 1.0

IBM CODE

Towards JanusGraph 1.0

- Benchmarking storage backends
- Bulk loading, serialization formats
- Query profiling and traversal optimization
- Gremlin language variant support
- Platform support (Windows)
- Server clustering support
- Administration console
- Operations tooling and monitoring

Thank you

janusgraph.org

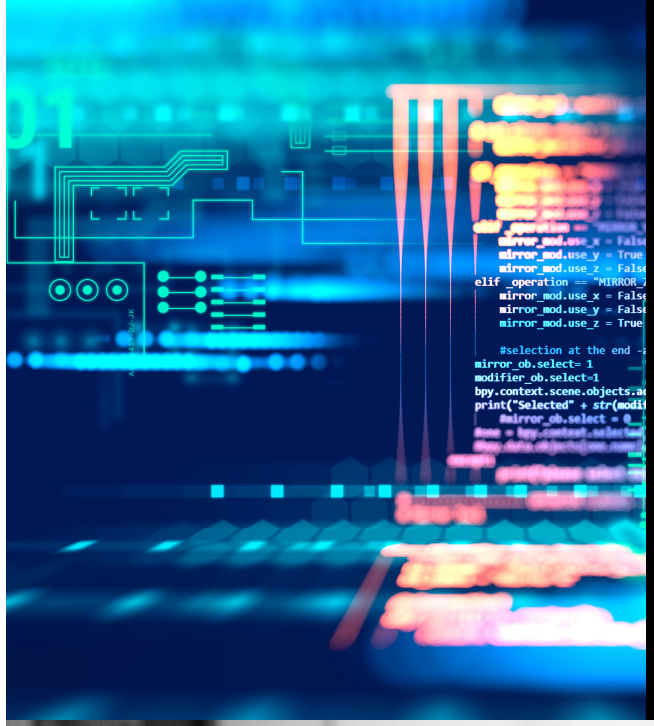
tinkerpop.apache.org

compose.com/databases/janusgraph

github.com/IBM/janusgraph-utils

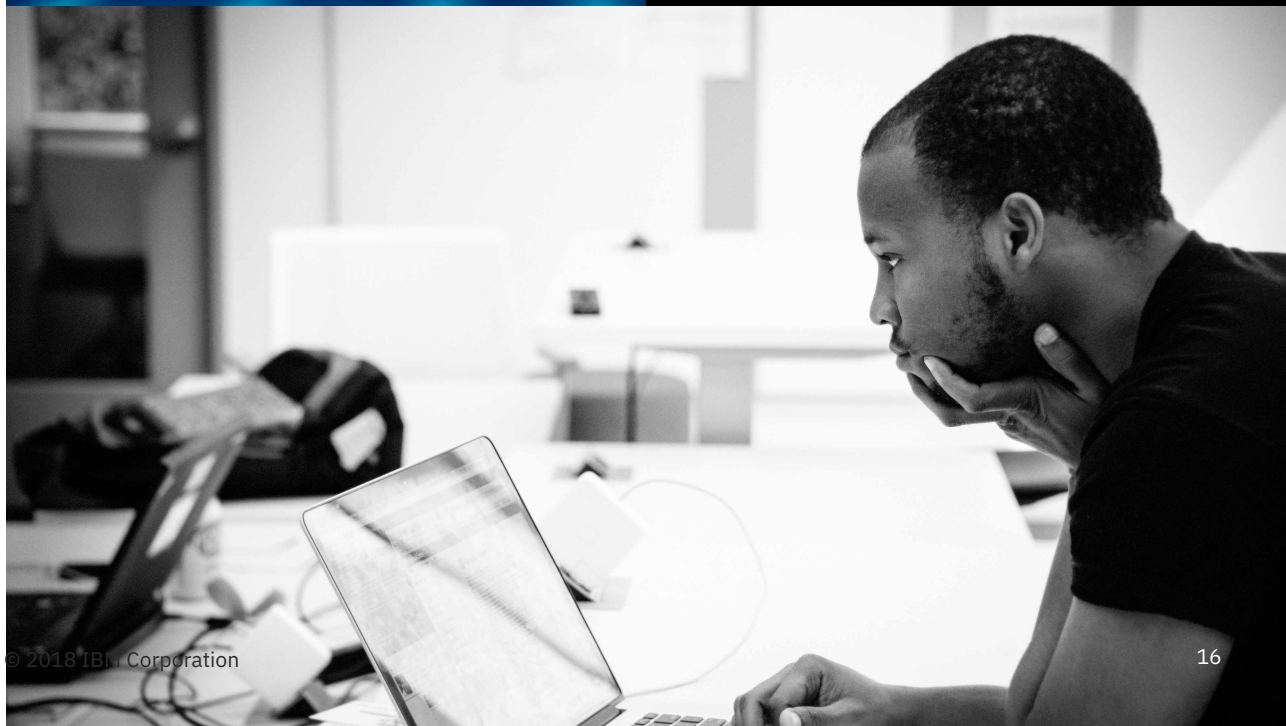
developer.ibm.com/code/patterns

twitter.com/pluradj



Build on IBM Cloud
for free with
no time restrictions

https://ibm.biz/BdZ6ci



IBM CODE

 © 2018 IBM Corporation