



Sharpen Your Insights

Fast, Effortless Analytics and AI on Massive Connected Datasets





COMPANY

- **85 employees:** 70 engineering (15 PhDs) + 15 business
- **Leaders** in graph, AI/ML, distributed/parallel systems
- **Offices** in: Austin, Bay Area, NYC, and Denver



Experienced Team



INVESTORS



Anthony Lin



Lip-Bu Tan



Amarjit Gill



Scott Darling



Tatiana Evtushenkova

LEADERSHIP



Keshav Pingali
CEO, Co-founder
Prof UT Austin,
Academia Europaea,
Fellow ACM, IEEE, AAAS



Chris Rossbach
CTO, Co-founder
Prof UT Austin
MSR, VMware, Canesta

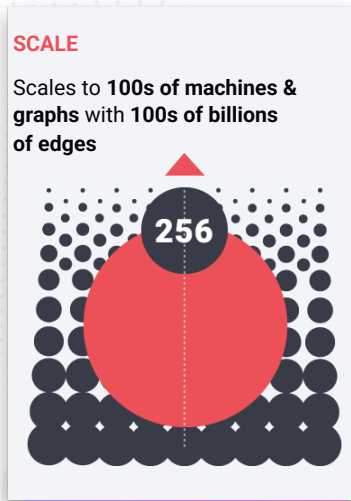
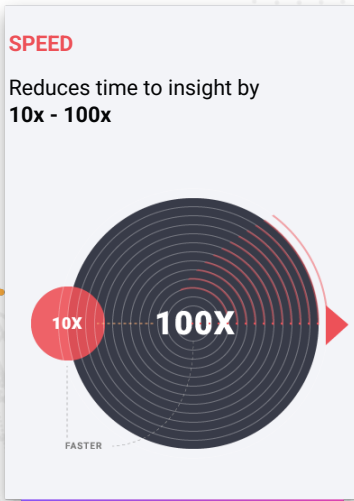


Farshid Sabet
CBO
Intel, Movidius,
Aptina, SanDisk



Katana Graph in a Nutshell

Graph Intelligence Platform for Query, Analytics and AI





Performance Matters!

Problem

- **Root Cause Analysis** : Identify cancer biomarkers using pathway analysis of patient-molecular event-association graph

Result

- Multiomics Knowledge Graph bridges genomic, epigenomic and clinic data silos
- Pathway analysis enables explainability

“

Performance is critical for such dense graph applications, and scalability is pivotal. Our lab is excited to apply this pipeline to other cancer types.

Dr. Kasthuri Kannan, UT MD Anderson Cancer Center

neo4j

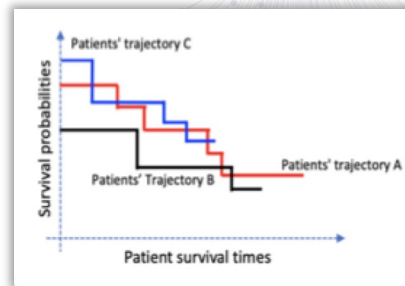
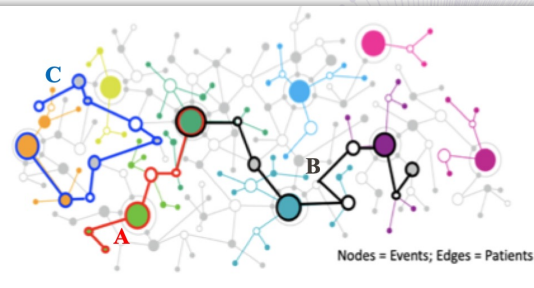
2.5 days



KATANA GRAPH



2 hours





Why Graphs?

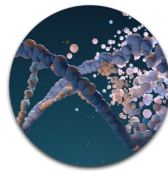
Real World Data is not Tabular



SUPPLY CHAIN



FINANCIAL



LIFE SCIENCE



IoT

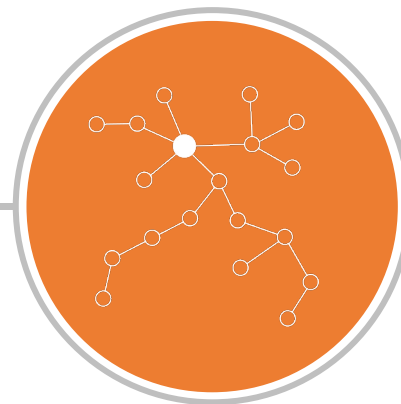


SOCIAL NETWORKS



TELCO

CONNECTED



Real world data is heterogeneous,
multiscale, hierarchical, and schema-free

Graph Applications

Fraud Detection

Anti-Money Laundering

Credit Decisioning

Drug Discovery

Precision Medicine

Genomics

Knowledge Graph

Intrusion Detection

Identity Graph

Recommendation

Customer 360

Supply Chain Optimization

Predictive Monitoring

Electronic Circuit Design Tools

Graphs: Fast-growing Market

“ Graphs are the native format, the most natural data structure of the world’s data.



“ Graph Tech will be used in 80% (2025) of Data and Analytics innovations up from 10% in 2021.

Gartner

“ By the end of this decade, 50% of SQL workloads will move to graphs.



\$155.6B

Data & Analytics
Market size 2025

Gartner

\$55B

Business Intelligence & Analytics
Market size 2027

VMware

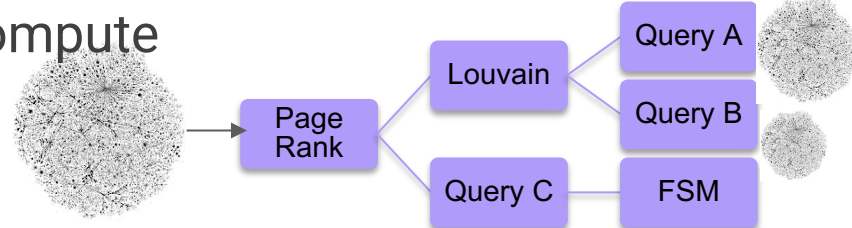
800%

Growth in Unstructured Data
Through 2024

Gartner



Katana: Graph Data Engine for Graph Compute



Database: neo4j, TigerGraph, Neptune

Manage and mutate a corpus of data

Tightly couples data and computation

Dataflow abstraction on top of database: **HARD**

Requires: ETL, orchestration, schema-mapping, provisioning

Dataflow Engine: Katana, Spark

Compose DAGs of arbitrary compute

Loosely coupled, separate storage and compute

Database abstraction on top of dataflow: (relatively) **EASY**

Requires: Shared log, storage layer



Katana: a dataflow engine that supports important graph database functionality

Efficient queries (OpenCypher)

Design for easy composition and orchestration

Design to anticipate integration with other computing frameworks (Dask, RDKit,...)



Competitive Landscape



Performance

10 - 100x
10+ hop queries on large dataset
High-performant OLAP workloads

Neo4J

Degrades >6 hops

AcmeWorks

Degrades >6 hops

AWS Neptune

Degrades >3 hops: OLAP not supported



Scale Out

Verified at +256 machines
(Public clusters: AWS, GCP, Azure)

Not Distributed

Distributed

Not Distributed



Programming Flexibility

C/C++, Python (Analytics, Mining)
OpenCypher (Query)

Java UDFs,
OpenCypher

Proprietary Only

SPARQL,
Gremlin



Graph Partitioning

Supports custom policies (including 1D, 2D and hybrid vertex cuts)
Optimized partition-aware Communication layer

NA

NA



Memory Footprint (cost saving)

Requires small clusters with efficient memory footprint
Allows selective loading of node and edge properties

Limited by single machine memory

Inflexible

Limited by single machine memory

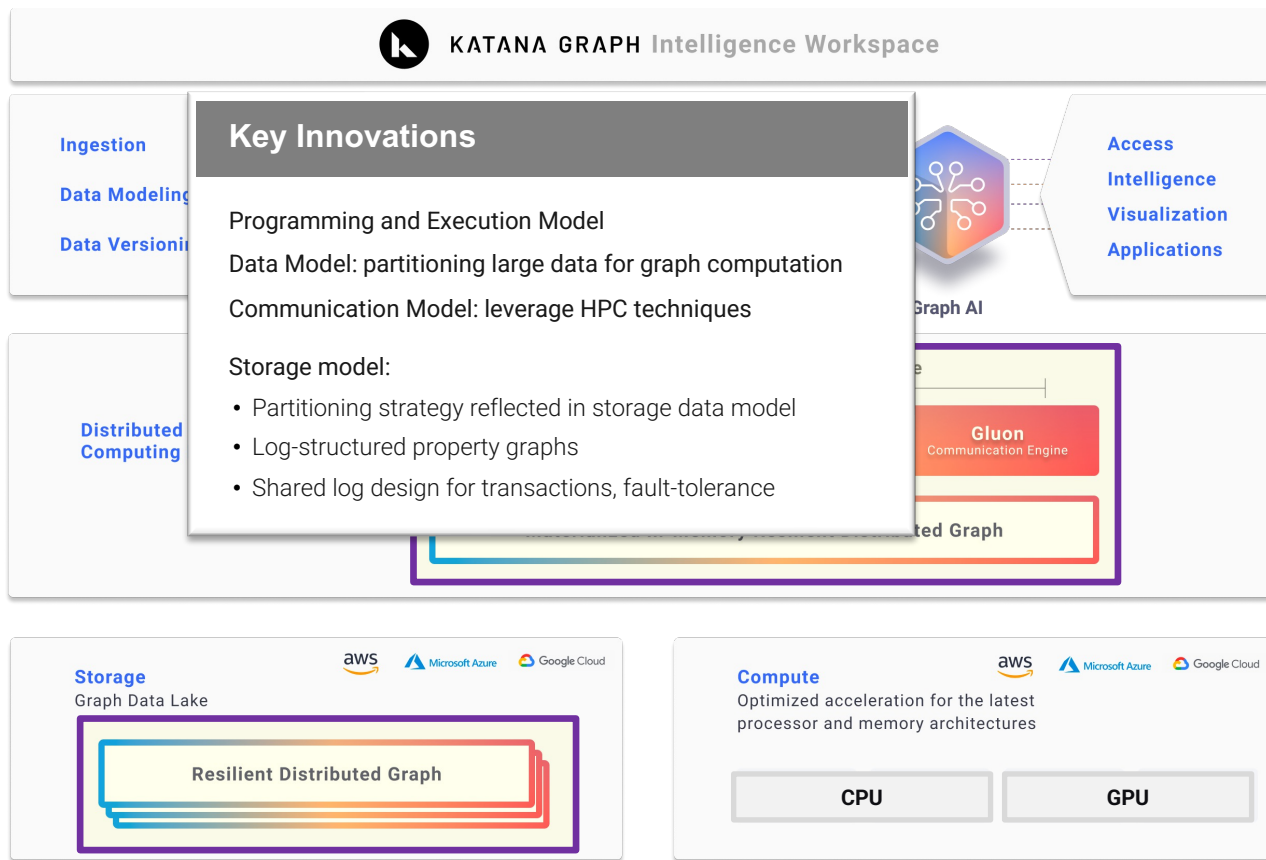


Dataset Size

Designed to handle +trillion edge graphs
Verified on massive graphs (WDC12: 3.5B N and 128B E)

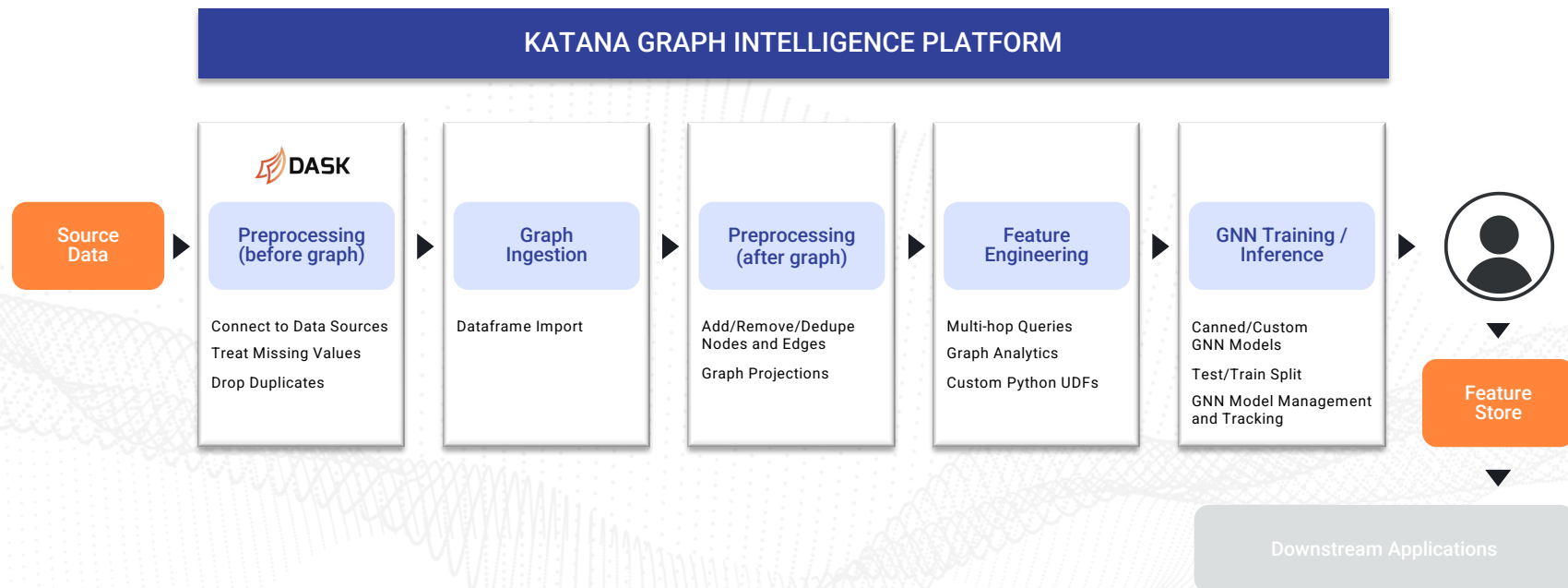


Graph Intelligence Platform – Under the hood



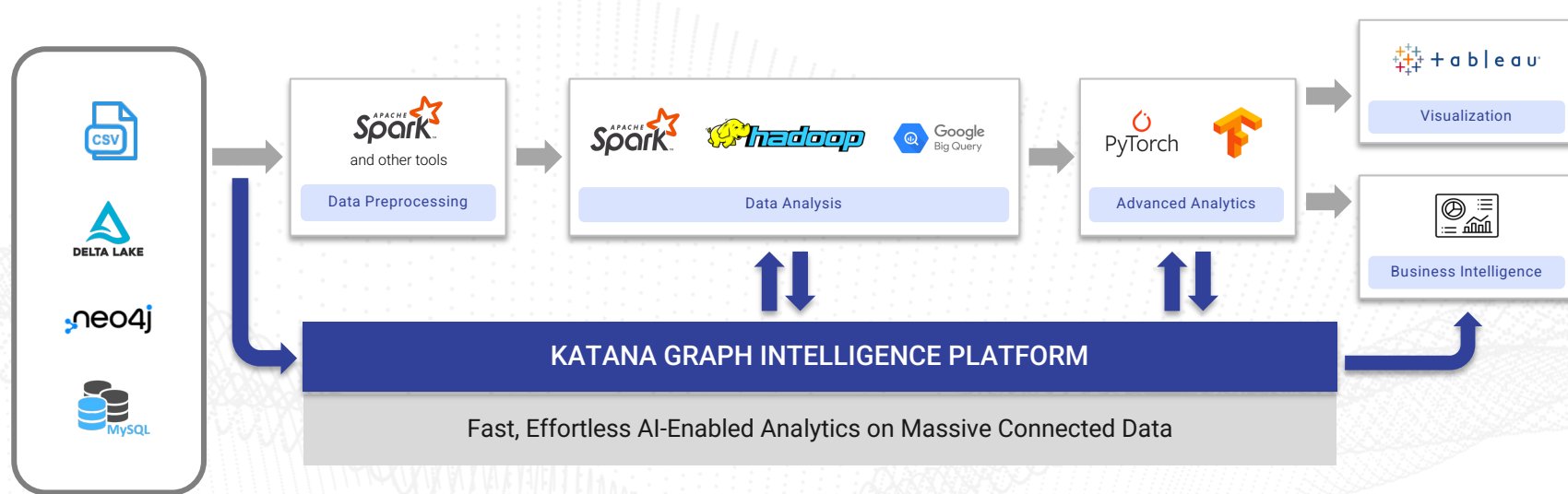


Katana Deployment (I) – Stand-alone Graph ML Pipelines





Katana Deployment (II) – Augmenting Existing ML Pipelines





Use Case : Anti Money Laundering



Katana Graph enables faster time-to-value and deeper insights by combining scalable data persistence and compute

Katana's platform **reduces false positives** and **the cost of compliance** across multiple areas of AML.



Consolidated Representation of Customer Data

Enables at-will large heterogeneous schemas that can serve verity of digital transformation use cases without any application or data remodeling.



Transaction Monitoring

Transaction monitoring is typically rules-based and cannot adapt to changes in fraud patterns. Katana Graph supports seamless integration with deep learning applications to improve model accuracy and reduce false positives.



Identity Resolution

Higher accuracy of identity resolution in short amount of time and with lower cost. On-demand scale-up/down of infrastructure when performing high compute functions.



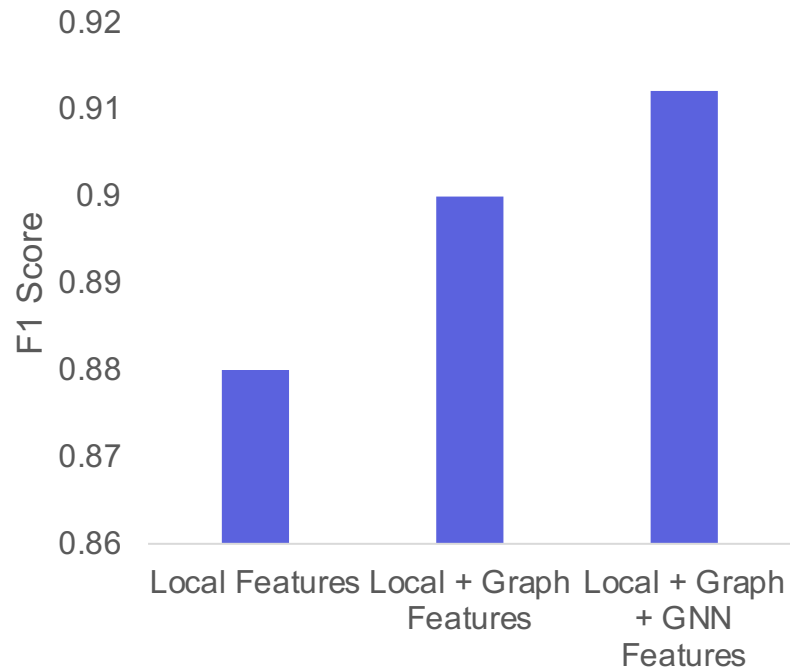
Transaction Monitoring

Reduce customer onboarding time and streamline Customer Due Diligence (CDD) processes by leveraging graph representation of all customer data collected by the bank. Utilize graph embeddings to risk-rate customers for annual reviews.



Graph-based features improve model accuracy on the Elliptic Bitcoin transaction dataset

Model accuracy identifying illicit transactions



Local Features: Transaction-level features for each node.

Graph Features: Features derived from graph query (multiple-hop queries) and graph algorithms (Louvain, Connected Components)

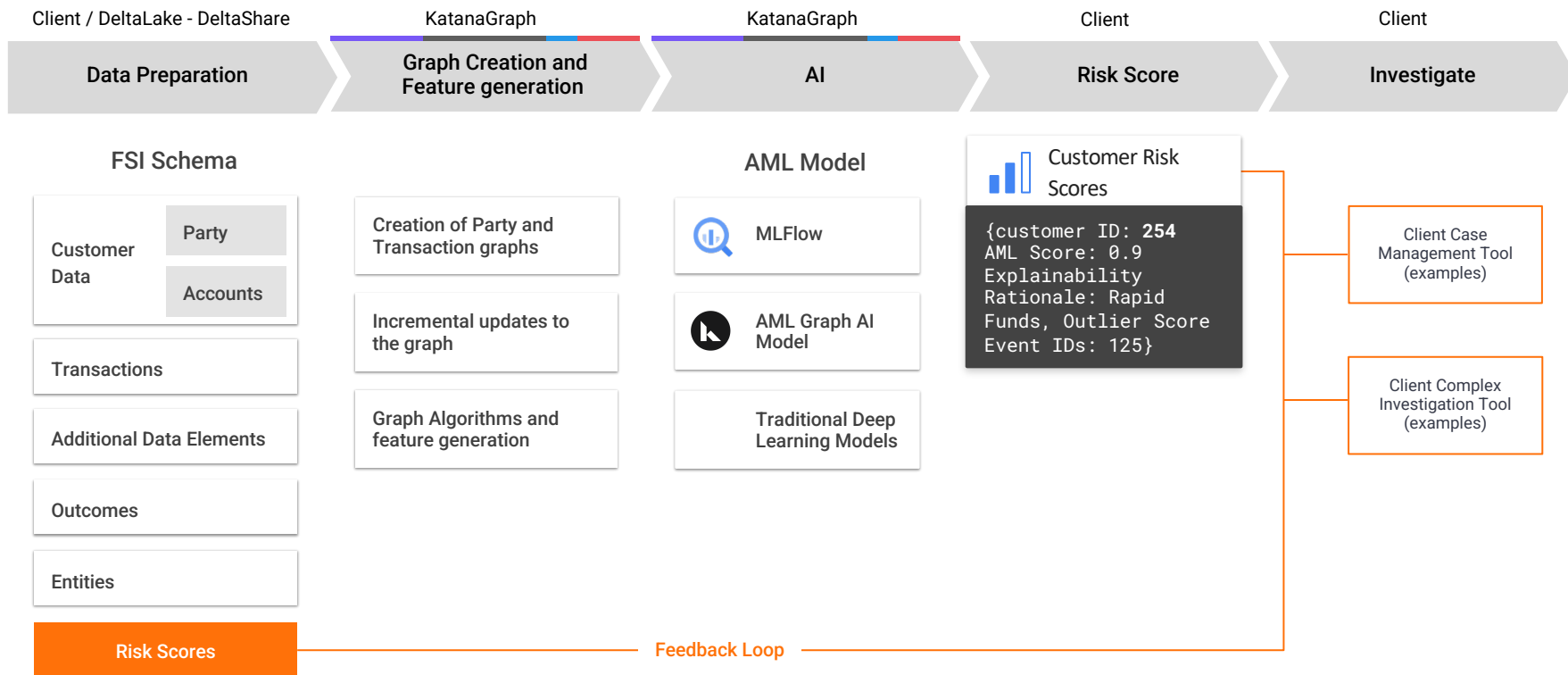
GNN Features: Trained graph embeddings using graph neural network architectures (i.e. GCN)



Katana - Databricks Integration (AML Reference Architecture)

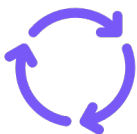
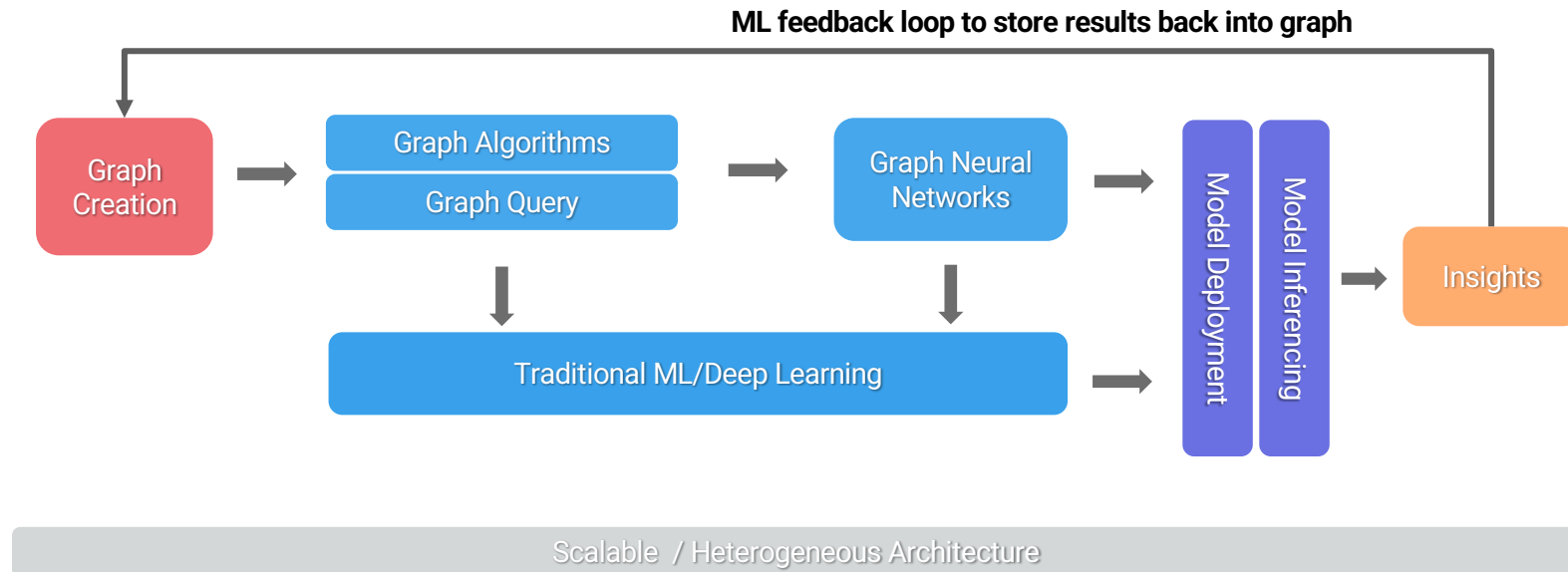


Overview of KatanaGraph – Databricks Integration





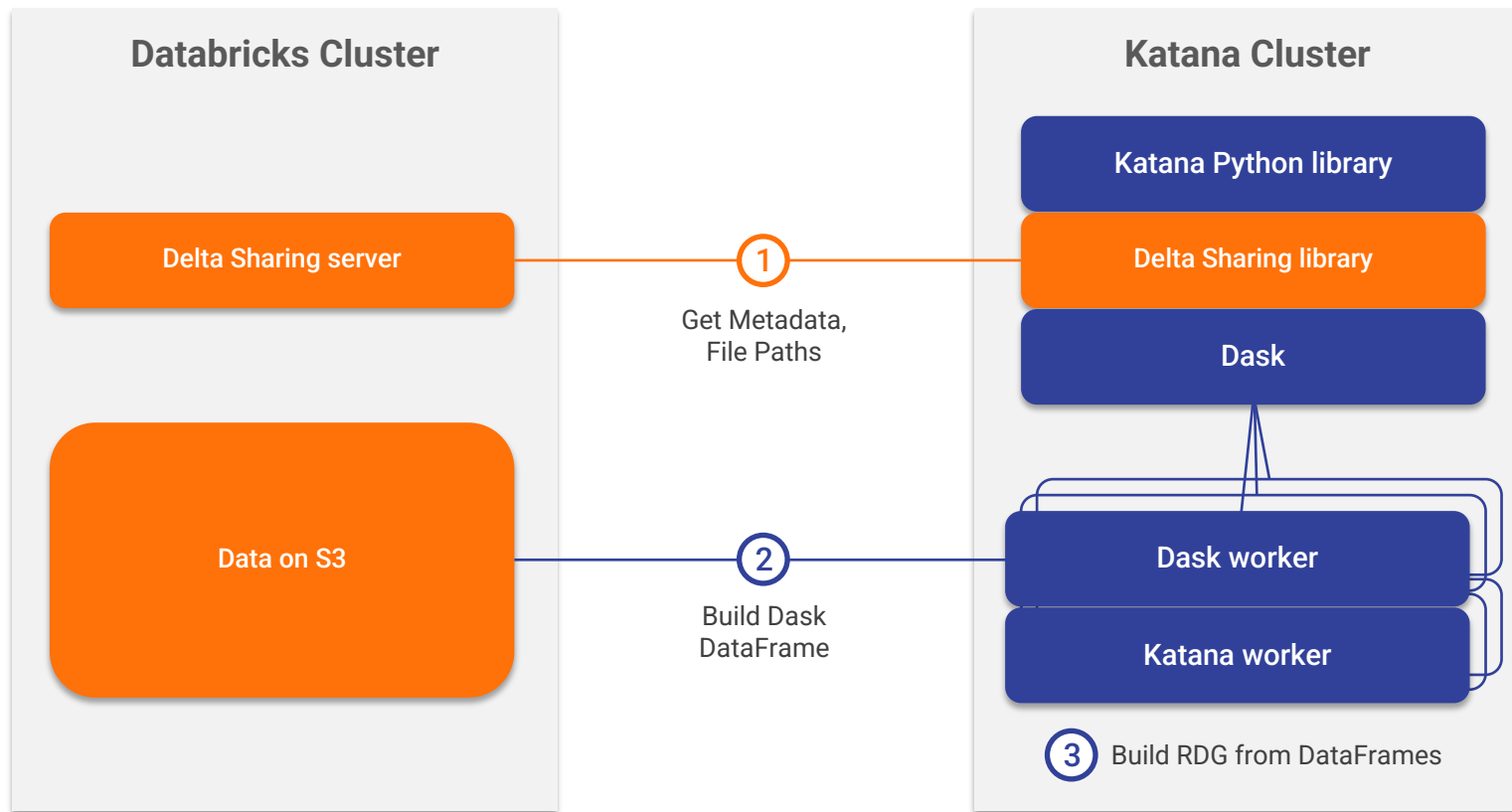
Overview KatanaGraph accelerated deep learning pipeline



Katana Graph's seamless platform enables rapid iteration of pipeline at-scale to produce best graph-based solutions



Data Sharing





Problem

- Prevent consumer fraud in real-time by detecting suspicious transaction patterns.
- Existing ML pipeline suffered from high false positives (30% PR-AUC).

Result

- Boosted accuracy and robustness over industry standard by leveraging graph feature augmentation (> 11% increase in PR-AUC and 2% increase in ROC-AUC)
- Improved ability to detect first-time fraudulent profiles/devices
- Ability to handle large evolving graphs



KATANA GRAPH

+11% PR-AUC



KATANA GRAPH

40B Transactions



While Amazon Neptune failed to load our TBs of graph data, Katana Graph loaded it in minutes...



Problem

- Better estimate the probability that a borrower will default on their loan
- Model improvement can directly boost financial performance

Result

- 6% improvement in accuracy over industry standard MLP model by leveraging graph AI features
- Leads to millions of dollars in increased revenue



KATANA GRAPH

+6% Accuracy

“

Katana provides easy to use end-to-end AI pipeline infrastructure and handles billion-node heterogeneous graph data.

VP at Fortune 500 FinTech



Katana Graph @ Intel ON event (September 2022)



Greg Lavender (He/Him) • Following
Chief Technology Officer at Intel
6h • 🌐

Combining the benefits of [Intel Corporation](#)'s next-generation hardware and software for drug discovery and development - impressive results shared by @Brian Martin using Intel 4th Gen Xeon Scalable processor cluster and [Katana Graph](#)'s Intelligence platform. [#IntelON](#) Watch at: <https://intel.ly/3Clzomk> [#AI](#), [#dataanalytics](#) [#drugdiscovery](#)



KATANA GRAPH [abbvie](#)

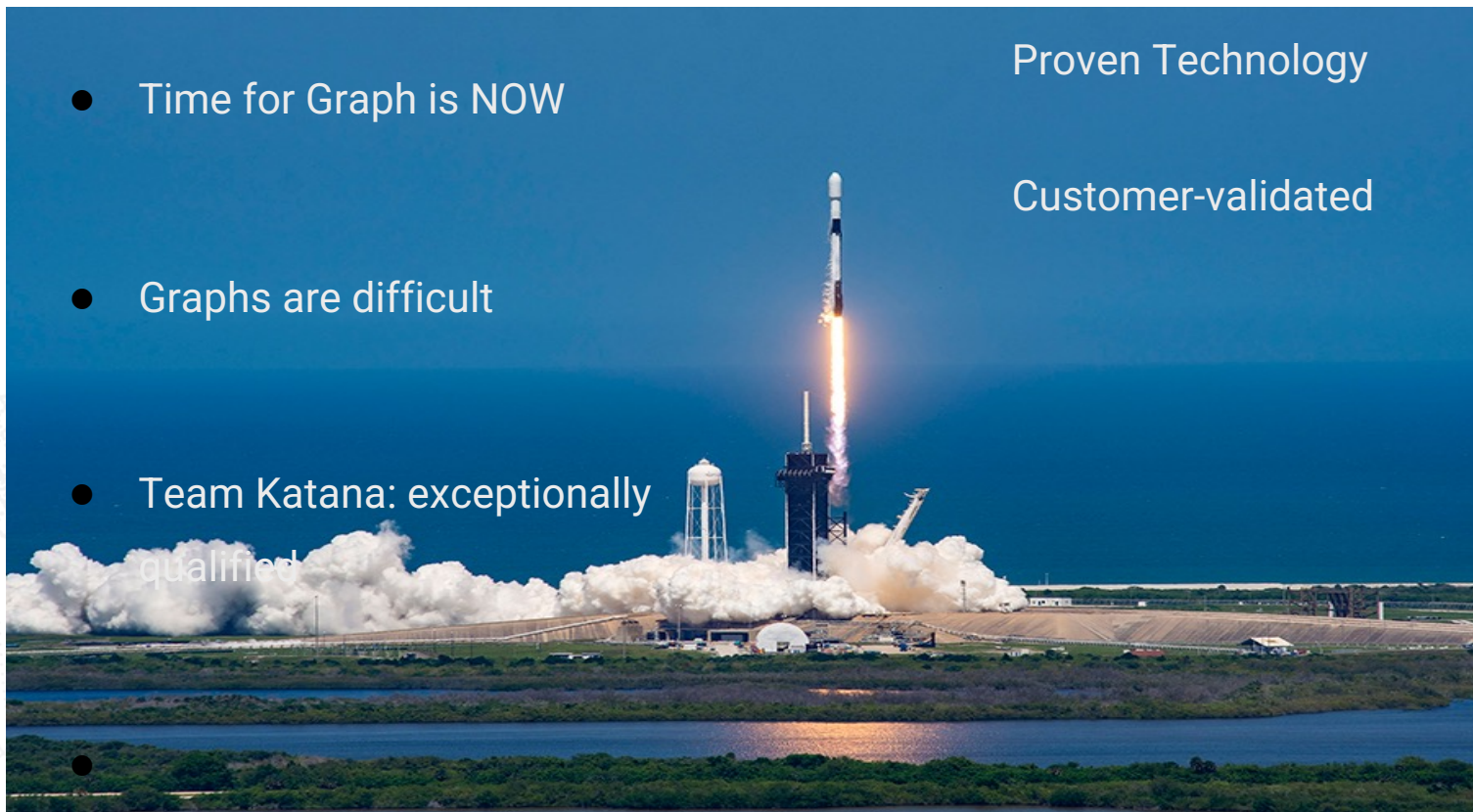




- Time for Graph is NOW
- Graphs are difficult
- Team Katana: exceptionally qualified

Proven Technology

Customer-validated





Demo : KatanaGraph – Databricks DeltaSharing Integration

Use Case : Anti Money Laundering