STARDOG

Transform your Data into an Asset with a Knowledge Graph

Our Presenter

Mike Grove

Founder & VP of Engineering, Stardog

STARDOG

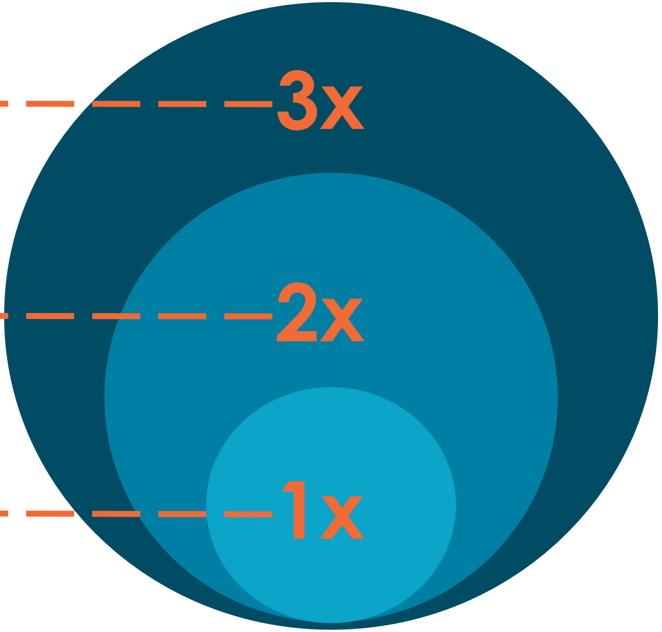
The market acknowledges companies that have successfully leveraged their data

Market vs tangible asset value (“Tobin’s q” ratio)

Companies who directly monetize their data whether through ads, derived data products, etc

Companies recognized as “information-savvy” by leveraging their data successfully

All others; companies with direct 1:1 market-to-book value



STARDOG

Source: Gartner: *The Birth of Infonomics, the New Economics of Information*

Data goals may appear disparate...

Operational efficiencies: supply chain optimization, predictive maintenance

Knowledge management: internal smart search, scientific research applications

Product improvement: recommendation engine, customer 360

Risk reduction & compliance: fraud detection, GDPR, SR-14

But they boil down to the same core goals

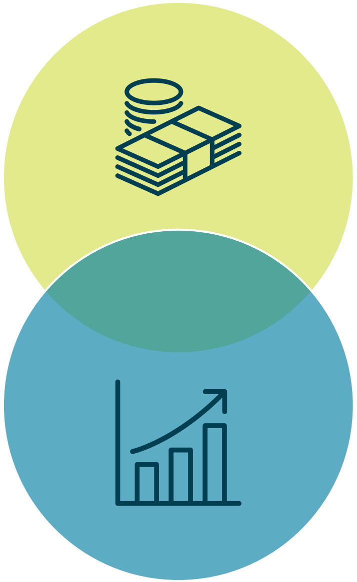
Operational efficiencies: supply chain optimization, predictive maintenance

Knowledge management: internal smart search, scientific research applications

Product improvement: recommendation engine, customer 360

Risk reduction & compliance: fraud detection, GDPR, SR-14

Direct monetization



Indirect monetization

Is enterprise data really an asset?

THE MAZE

It's only an asset if you can access it.

But accessing enterprise data is HARD.

Common barriers

What's preventing you from leveraging your data as an asset?

- Data unavailable; access limited (by social or structural issues)
- Data is collected centrally but you don't know what data you need
- Analysis is hampered by data quality or data preparation time required
- Machine learning is in use but lack of model explainability results in resistance to adoption

Common barriers

Access

Data unavailable; access limited (by social or structural issues)

Data is collected centrally but you don't know what data you need

Analysis is hampered by data quality or data preparation time required

Machine learning is in use but lack of model explainability results in resistance to adoption

- Lines of business manage their own datasets and do not yield access, have incompatible data formats to other LOBs, or both
- Data is in a format that is unreadable by current systems (frequently unstructured data)

Common barriers

Purpose

Data unavailable; access limited (by social or structural issues)

Data is collected centrally but you don't know what data you need

Analysis is hampered by data quality or data preparation time required

Machine learning is in use but lack of model explainability results in resistance to adoption

- Data lake was implemented to be "the single source of truth" within the organization but inventory of data is unclear or inconsistent
- Lack of business problem to solve but mandate to improve data strategy (competitors leveraging data better than you are)

Common barriers

Analysis

Data unavailable; access limited (by social or structural issues)

Data is collected centrally but you don't know what data you need

Analysis is hampered by data quality or data preparation time required

Machine learning is in use but lack of model explainability results in resistance to adoption

- MDM has been implemented but perceived data potential value is in the remaining ~80% of data, no visible path to data reconciliation
- Preparation time is a consistent burden that needs to be reduced to allow data scientists to spend more time on analysis

Common barriers

Adoption

Data unavailable; access limited (by social or structural issues)

Data is collected centrally but you don't know what data you need

Analysis is hampered by data quality or data preparation time required

Machine learning is in use but lack of model explainability results in resistance to adoption

- ML is hampered by limited training data, poor data quality, incomplete or ill-prepared data sets, reducing quality of ML models
- Pushback within organizations to accept directives from ML outputs without the context behind the analysis; need to improve explainability

Common barriers

Access	Purpose	Analysis	Adoption
Data unavailable; access limited (by social or structural issues)	Data is collected centrally but you don't know what data you need	Analysis is hampered by data quality or data preparation time required	Machine learning is in use but lack of model explainability results in resistance to adoption

STARDOG

Company spotlight:

NASA

Knowledge Management

"Stardog enables us to see all the data and all the relationships. It's a 10-to-1 savings."

Andrew Schain,
Program Data Integration Manager,
Exploration Systems Division, NASA

Problem

NASA's Space Launch System is the foundation for human exploration beyond Earth's orbit. Successful mission planning and execution requires navigating largely disconnected and siloed datasets across complex systems. To determine the impact a slight change in acceptable vibration had on humans, NASA engineers needed 6-8 weeks to pull the relevant data manually from each dataset, analyze and qualify it, and then create a report.

Solution

NASA implemented Stardog's knowledge graph to manage, query, and analyze their data. To date, the team has unified 24 data sources with 40 specific record types. This unified view is crucial for knowledge management, quality control, and decision support.

Benefits

NASA has saved countless hours assembling the answers they need from interconnected data. What took weeks to compile now takes seconds. Engineers can now focus on exploring their data, asking new questions, and examining the impact of potential decisions.

Common barriers: Access

Access

Data unavailable; access limited (by social or structural issues)

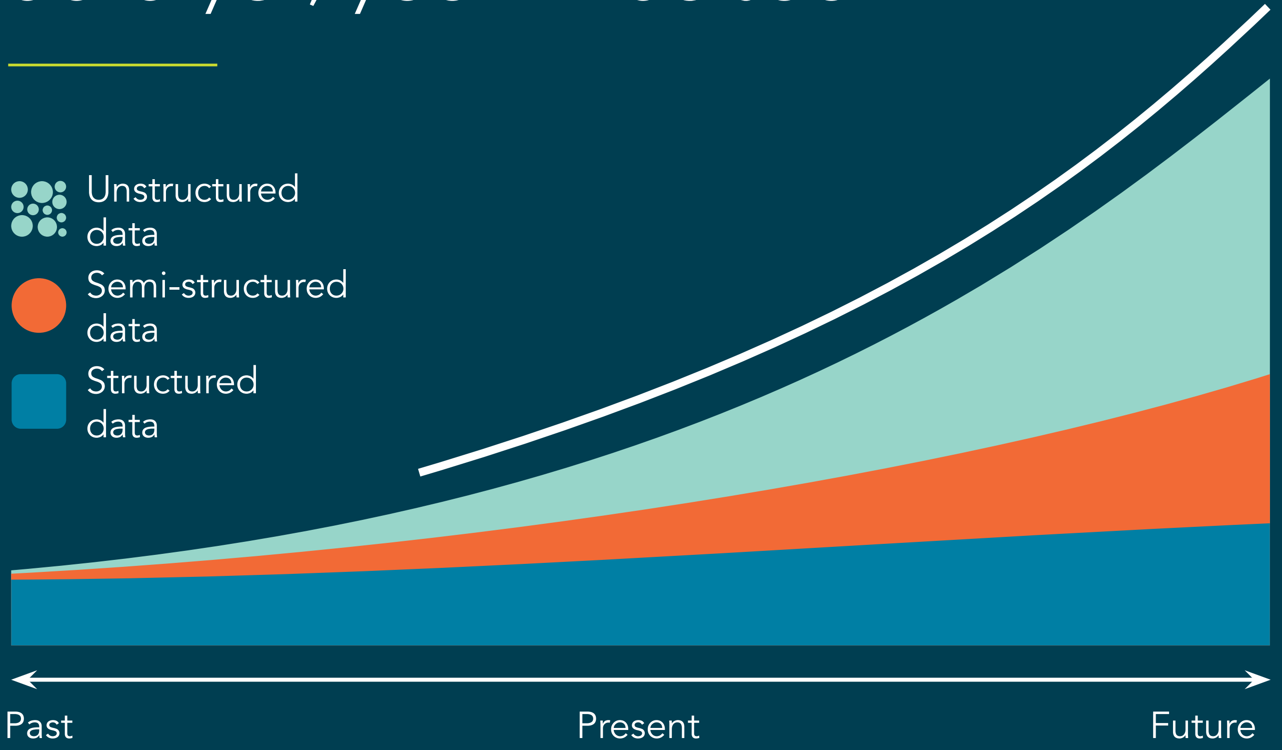
Data is collected centrally but you don’t know what data you need

Analysis is hampered by data quality or data preparation time required

Machine learning is in use but lack of model explainability results in resistance to adoption

- Lines of business manage their own datasets and do not yield access, have incompatible data formats to other LOBs, or both
- Data is in a format that is unreadable by current systems (frequently unstructured data)

If you aren't wrestling with unstructured data yet, you will be soon



30.3% CAGR

in non-relational analytic data stores

PREDICTED BY IDC

800% GROWTH

in data volume, mostly in unstructured

PREDICTED BY GARTNER

Sources: IDC Report, Netowl

Data of all varieties

Structured

Data with defined length and format, relationships defined by indices

Examples: XLS

Sources: Relational databases, your existing data warehouses

Semi-structured

Data not organized in fixed fields or records, but has some hierarchies

Examples: JSON, XML

Sources: NoSQL databases

Unstructured

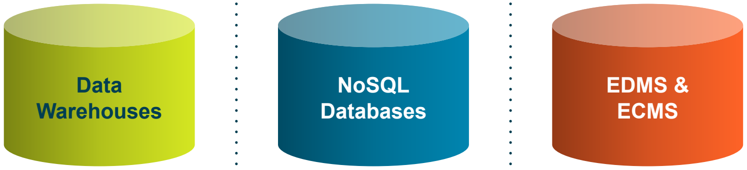
Data without any organization, often text-heavy; also multimedia formats

Examples: PDF, social media content

Sources: Often saved in data lake or locked in applications eg Sharepoint, JIRA

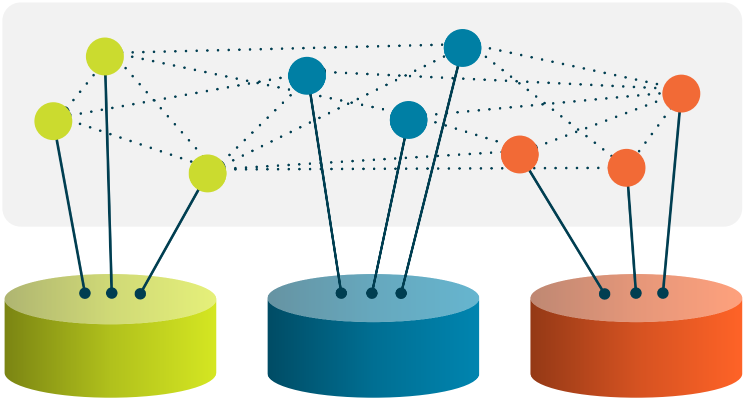
Data content matters more than data location

Database Centric



Most IT organizations have been built to manage data by department and type.

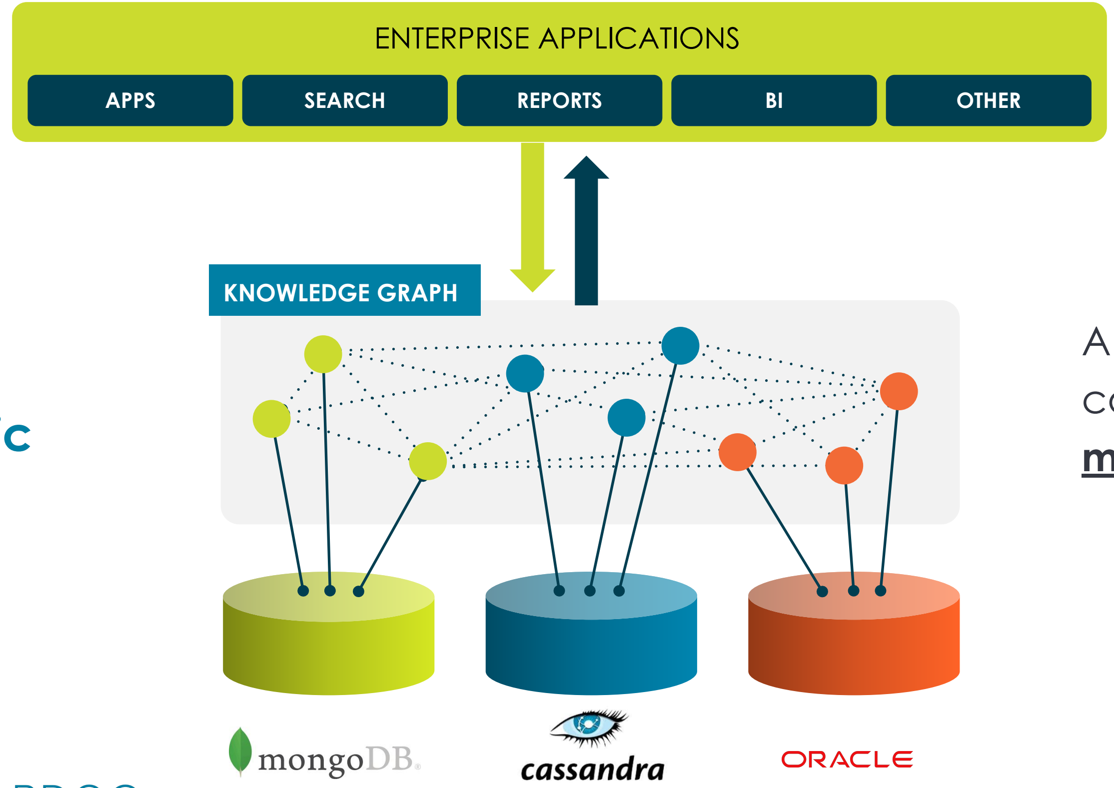
Data Centric



A Knowledge Graph connects all data **without moving or copying it.**

STARDOG

Data content matters more than data location



Data Centric

A Knowledge Graph connects all data **without moving or copying it.**

STARDOG

Company spotlight:

BOSCH

Recommendation Engine

Stardog connects and exposes our global product catalogue – true enterprise scale.”

Dr. Christian Hütter
Enterprise Architect, Power Tools

Problem

Each year, Bosch launches more than 100 new power tools to its global product line, with over 50,000 products in total. Data about these tools is spread across many relational and unstructured sources.

Solution

- Stardog powers a live Product Recommendation engine on BoschTools.com
- Organized Bosch’s content library to create a clean, consistent, customer-facing searchable database -- complete with decades of tools, their optimal usage, associated products, and user manuals

Benefits

As the online marketplace becomes more competitive for retail goods, Bosch has improved its customer experience with searchable content and product recommendations.

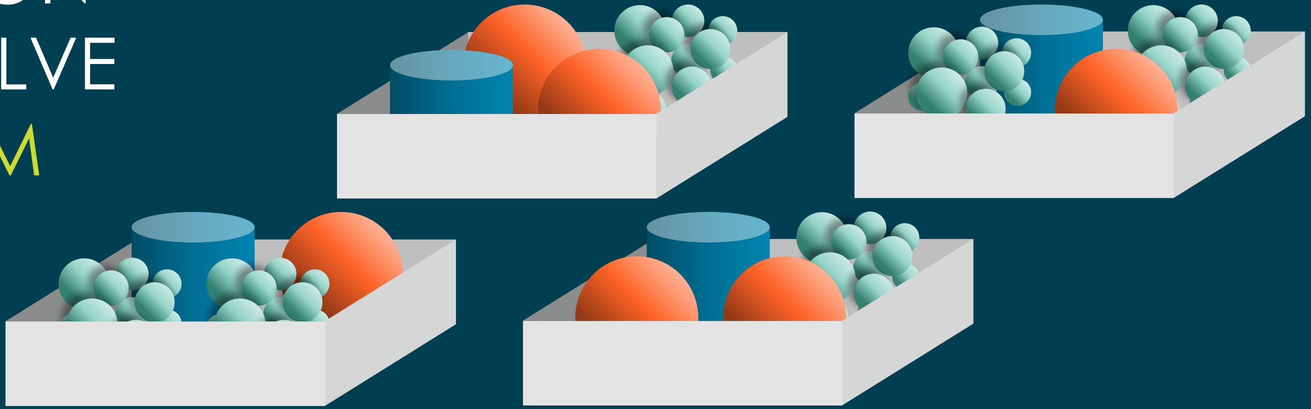
Common barriers: Purpose

Purpose

Data unavailable; access limited (by social or structural issues)	Data is collected centrally but you don't know what data you need	Analysis is hampered by data quality or data preparation time required	Machine learning is in use but lack of model explainability results in resistance to adoption

- Data lake was implemented to be "the single source of truth" within the organization but inventory of data is unclear or inconsistent
- Lack of business problem to solve but mandate to improve data strategy (competitors leveraging data better than you are)

CO-LOCATION DOESN'T SOLVE THE PROBLEM



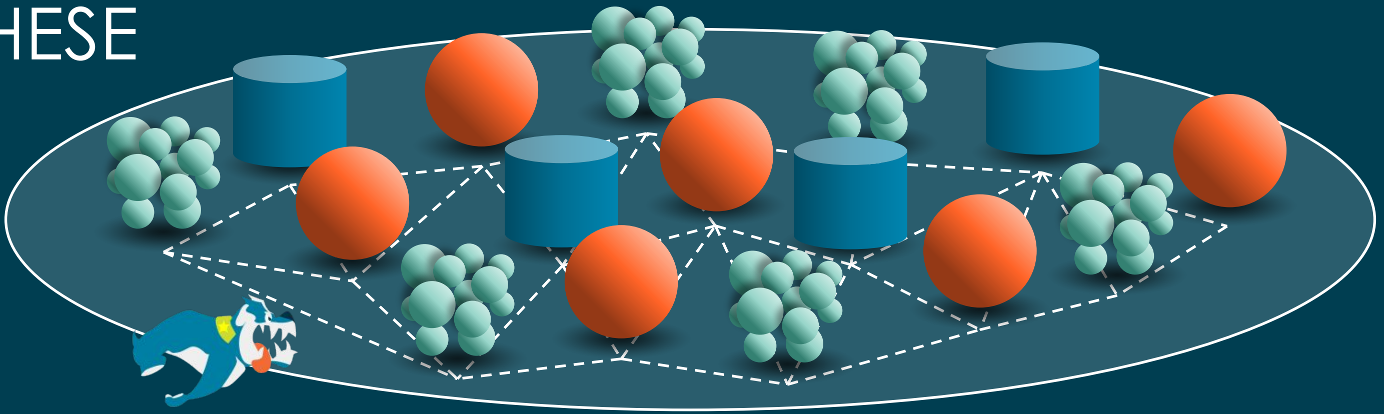
There is always data in another location.

Unstructured data is usually left out.

Data context is critical.

STARDOG

STARDOG SOLVES EACH OF THESE PROBLEMS



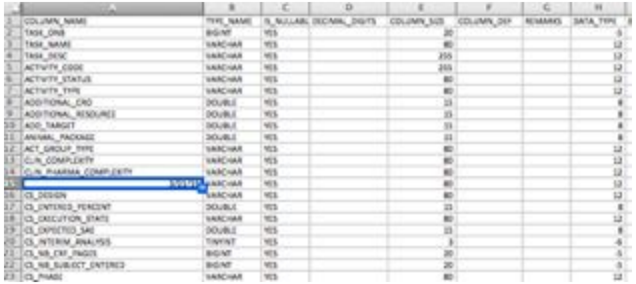
Works in a hybrid environment

Uses graph to work across all data types

Provides valuable context to data

Relational vs Graph

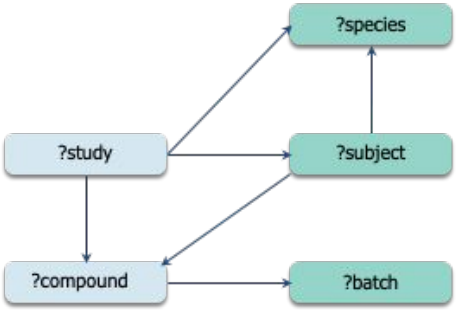
Relational Database



	A	B	C	D	E	F	G	H
1	COLUMN_NAME	TYPE_NAME	IS_NULLABL	DECIMAL_DIGITS	COLUMN_SIZE	COLUMN_DEF	REMARKS	DATA_TYPE
2	TASK_ONB	BIGINT	YES		20			-5
3	TASK_NAME	VARCHAR	YES		80			12
4	TASK_DESC	VARCHAR	YES		255			12
5	ACTIVITY_CODE	VARCHAR	YES		255			12
6	ACTIVITY_STATUS	VARCHAR	YES		80			12
7	ACTIVITY_TYPE	VARCHAR	YES		80			12
8	ADDITIONAL_CRO	DOUBLE	YES		22			8
9	ADDITIONAL_RESOURCE	DOUBLE	YES		22			8
10	ADD_TARGET	DOUBLE	YES		22			8
11	ANIMAL_PACKAGE	DOUBLE	YES		22			8
12	ACT_GROUP_TYPE	VARCHAR	YES		80			12
13	CLIN_COMPLEXITY	VARCHAR	YES		80			12
14	CLIN_PHARMA_COMPLEXITY	VARCHAR	YES		80			12
15	[illegible]	VARCHAR	YES		80			12
16	CL_DESIGN	VARCHAR	YES		80			12
17	CL_ENTERED_PERCENT	DOUBLE	YES		22			8
18	CL_EXECUTION_STATE	VARCHAR	YES		80			12
19	CL_EXPECTED_SAE	DOUBLE	YES		22			8
20	CL_INTERIM_ANALYSIS	TINYINT	YES		3			-6
21	CL_NB_CRF_PAGES	BIGINT	YES		20			-5
22	CL_NB_SUBJECT_ENTERED	BIGINT	YES		20			-5
23	CL_PHASE	VARCHAR	YES		80			12

- Stores data in rows and columns
- Well-suited for structured data queries with pre-defined relationships
- Rigid structure prevents changes later
- Works well for management of transactional data and simple look-ups

Knowledge Graph



- Represents data as entities with relationships with an expressive graph model
- Well-suited to query complex and highly connected data
- Flexible design lets you accommodate new use cases and data sources

Formally transitioning from a relational model to that of linked data was a huge strategic benefit to the bank. We are now able to design and link domain models across organizations and silos.

Executive Director, Top 5 US Bank

Company Spotlight:

Morgan Stanley

IT Portfolio Management

"

An hours-long triage of outages is a 50ms query now."

Richard Viana, Executive Director
Enterprise Infrastructure

Problem

With nearly 60,000 employees, Morgan Stanley's Enterprise Infrastructure team struggled maintaining a single source of truth across the production, operations, and engineering of all global technology assets.

Solution

Stardog powers Morgan Stanley's Impact Data program, an IT Portfolio Management platform used to manage global assets and their associations, roles, and entitlements for Morgan Stanley's global employee base.

Benefits

- Created a single, authoritative view of the entire enterprise and its associated architecture, which enables analytics and reporting
- Platform drives risk management, as well as system & infrastructure cost reporting
- Downstream outage ramifications are now easily identified and quantified

Common barriers: Analysis

Analysis

Data unavailable; access limited (by social or structural issues)

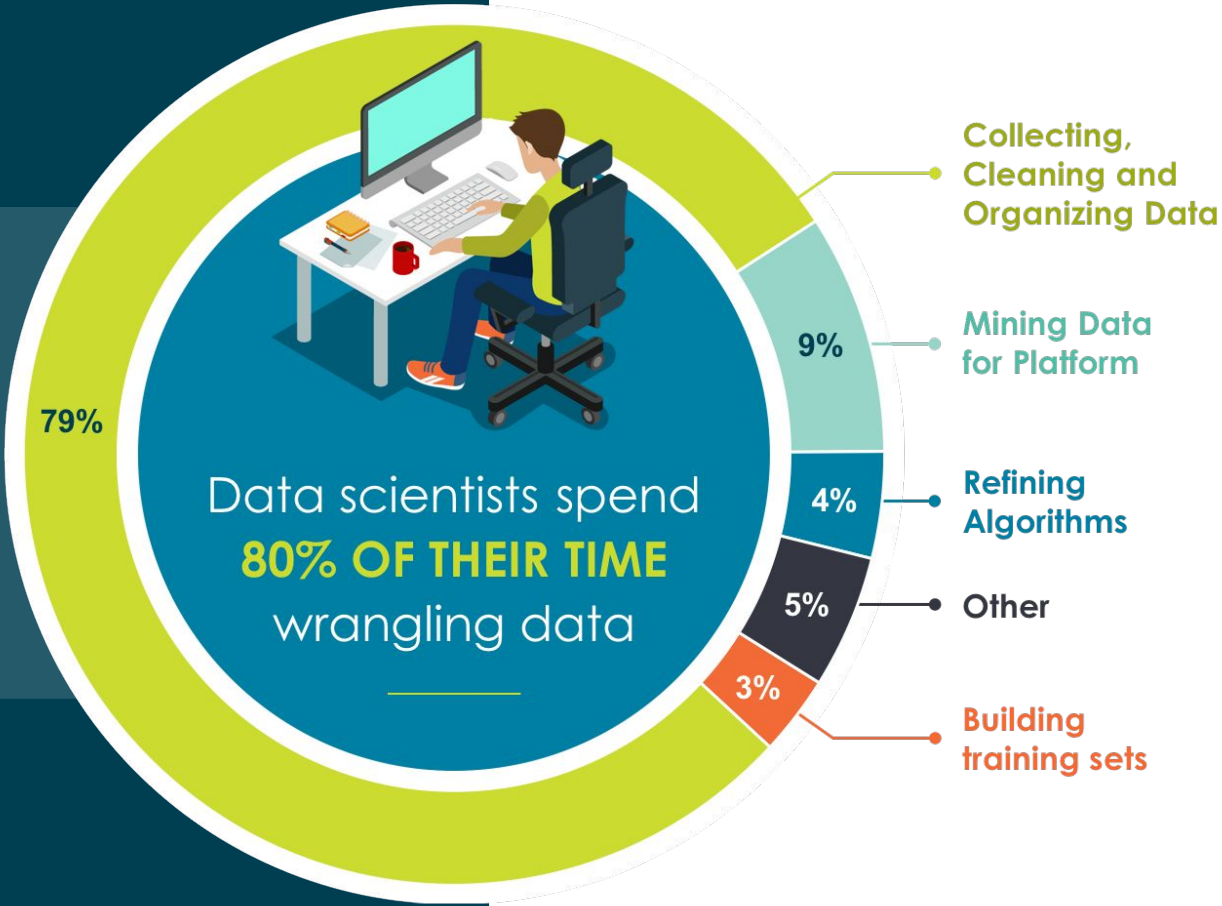
Data is collected centrally but you don't know what data you need

Analysis is hampered by data quality or data preparation time required

Machine learning is in use but lack of model explainability results in resistance to adoption

- MDM has been implemented but perceived data potential value is in the remaining ~80% of data, no visible path to data reconciliation
- Preparation time is a consistent burden that needs to be reduced to allow data scientists to spend more time on analysis

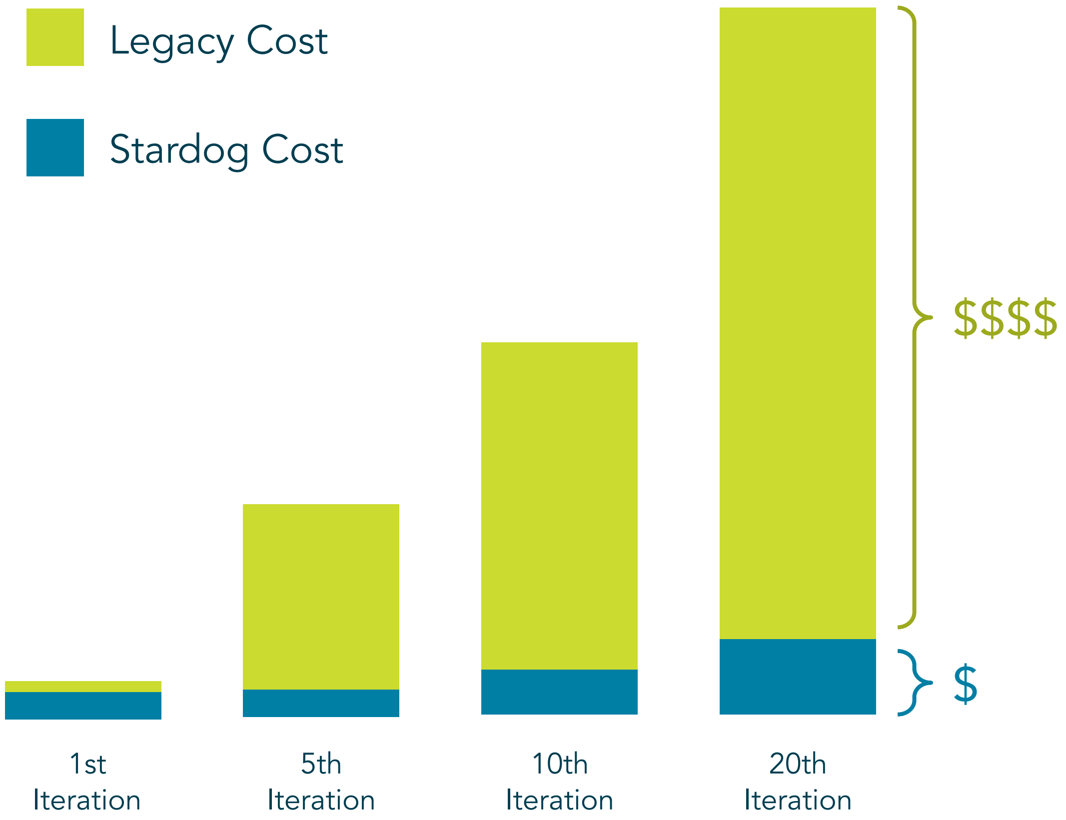
Every large organization struggles with data analysis



STARDOG

Source: Figure Eight

STARDOG DECIMATES THE COST OF ANALYSIS



Data wrangling is a huge source of inefficiency. Stardog changes the game.

- Logical model works with changing data.
- Simple mapping makes it easy to add new sources.

Set the logic once and it's done.

Enterprise "Dark Data" is information collected during the course of business that remains in archives, is not generally accessible, or is not structured sufficiently for analysis. It can include emails, contracts, documents, multimedia, system logs and other overlooked information assets.

Parsing, tagging, linking, or otherwise structuring or extracting usable information from these sources is the greatest immediate opportunity for most businesses among all types of information.

Gartner: Applied Infonomics

Common barriers: Adoption

Adoption

Data unavailable; access limited (by social or structural issues)

Data is collected centrally but you don't know what data you need

Analysis is hampered by data quality or data preparation time required

Machine learning is in use but lack of model explainability results in resistance to adoption

- ML is hampered by limited training data, poor data quality, incomplete or ill-prepared data sets, reducing quality of ML models
- Pushback within organizations to accept directives from ML outputs without the context behind the analysis; need to improve explainability

Leverage a knowledge graph for gaining buy-in

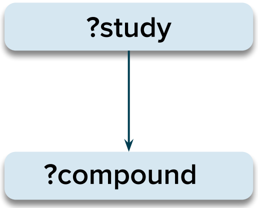
- **Problem**: Machine learning models have high accuracy but low explainability, answers are in a "black box"
- **Solution**: Knowledge graphs allow you to look at the snapshot from the model and ask "What led to this point?" -- bringing context to how the decision was made

"Model interpretation (or explainability) is the ability to explain the decisions of a predictive or prescriptive model to enable accuracy, fairness, reliability, accountability, stability and transparency in algorithmic decision-making."

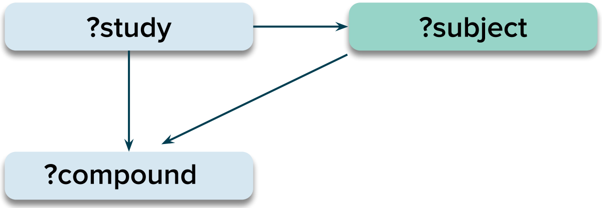
Gartner, Making Machine Learning Explainable

Knowledge graphs enable discoverability

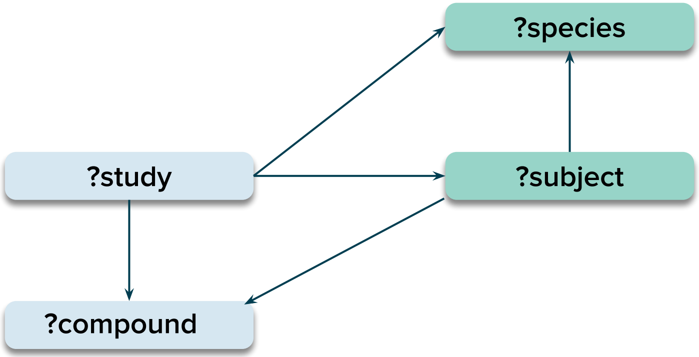
Knowledge graphs enable you to walk through data and understand the connections. You can easily jump from question to question and test new hypotheses.



Which studies included this compound?



Which subjects were also in common with this compound?



What was the species mix of the studies for this compound?

Developing a data strategy

Identify an internal use case

Layer on incrementally

- Start with a key use case with a strong advocate
- Strengthen value of knowledge graph with additional sources; value is accretive over time

Evaluate prime data sources

Look for opportunities in your data

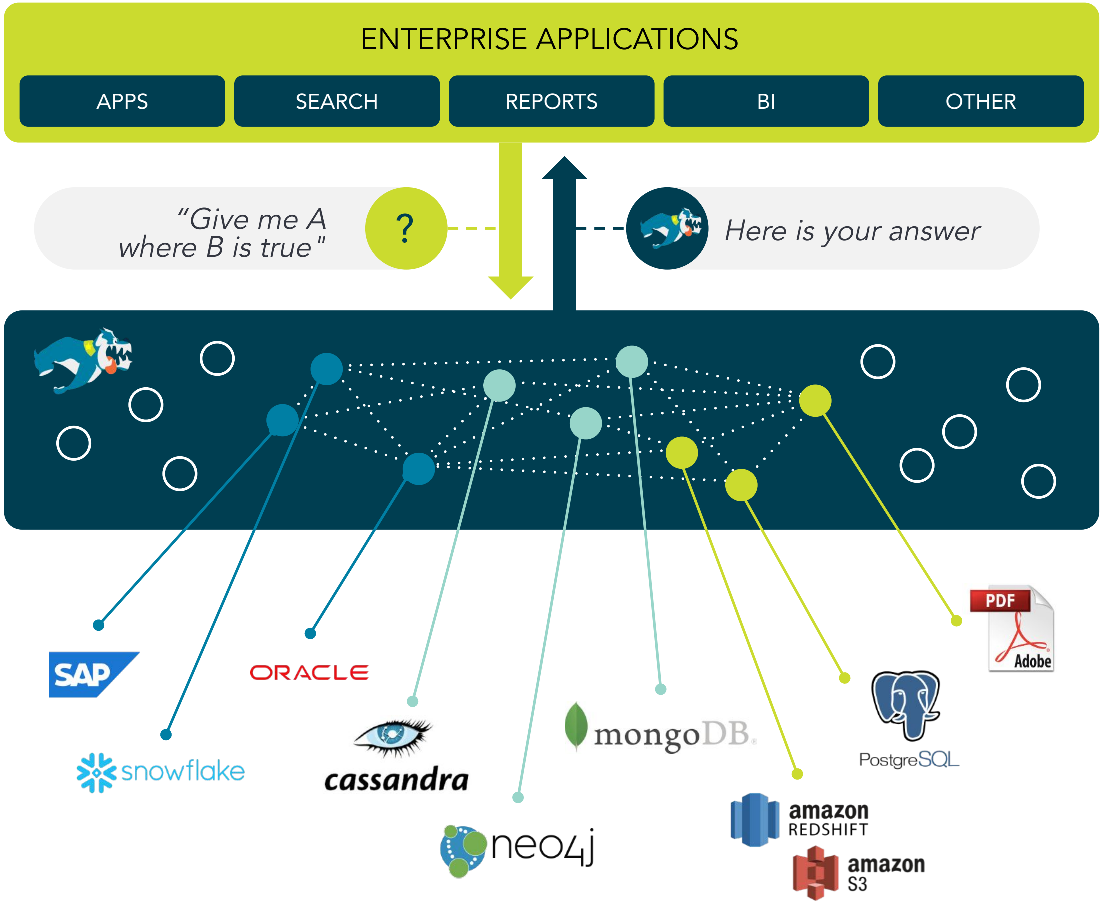
- Proprietary data
- Accurate/precise data (that you can warrant)
- Siloed data - exposing and connecting to relevant data across the org can be very powerful

How Stardog is unique

HOW STARDOG WORKS

Stardog queries data across the enterprise by:

- Using a high-level data model based on semantics
- Mapping data locations and meanings
- Virtualizing and storing as needed
- Running graph queries against the data model to return answers



Our differentiated features

- Support for all data types
- Virtual graph & storage capabilities
- Logical Data Model
- Enterprise Scale and Performance
- AI Native
- Standards-based

STARDOG

MAJOR COMPANIES RECOGNIZE THE VALUE

BOSCH Invented for life
BNY MELLON
NASA
SPRINGER NATURE

Schneider Electric
Morgan Stanley
NATIONAL CANCER INSTITUTE
ERICSSON
IDB

SCHAEFFLER
Debeka
NIH National Institutes of Health
Raytheon
salesforce

NATIONAL BANK OF CANADA
FINRA
Australian Bureau of Statistics
NOKIA

Boehringer Ingelheim
Moody's
DSM BRIGHT SCIENCE. BRIGHTER LIVING.
Basel basel.ch

Our use cases span across your business

OPTIMIZE	GOVERN	PROTECT	MANAGE	KNOW
Supply Chain	Reporting	Risk & Compliance (e.g. SR-14, GDPR)	Asset Management	Customer 360
Expertise Location	Outage Analysis	Fraud Detection	Scientific Research	Recommendation Engines
Predictive Maintenance	Knowledge Base	Anti-Money Laundering	Resource Authorization	Smart Buildings & IoT

STARDOG

Ways to learn more

Download the whitepaper

Read the Knowledge Graphs 101 whitepaper for additional details:

fetch.stardog.com/web-enterprise-knowledge-graph/

Try out a knowledge graph

Download Stardog's knowledge graph to connect and query your own data:

stardog.com/download

Q&A

Thank you!

Knowledge Graph can drive significant business value

Deploying a knowledge graph addresses the following market realities:

Organizations continue to spend 60%-80% of data analysis workflow on finding, accessing, preparing and sharing data for further analysis

- Gartner: *Market Guide for Data Preparation*

Companies today are sitting on "dark data" - defined as data not generally accessible or structured sufficiently for analysis. Parsing, tagging, linking or otherwise structuring and extracting information from this data represents "the greatest immediate opportunity for most businesses among all types of information"

- Gartner: *Applied Infonomics: Seven Steps to Monetize Available Information Assets*

While data is an unquantifiable asset from an accounting standpoint, companies that are leveraging their data to improve their operations and product are recognized by the market with a 2x market-to-book value

- Gartner: *The Birth of Infonomics, the New Economics of Information*

Silos are a reality

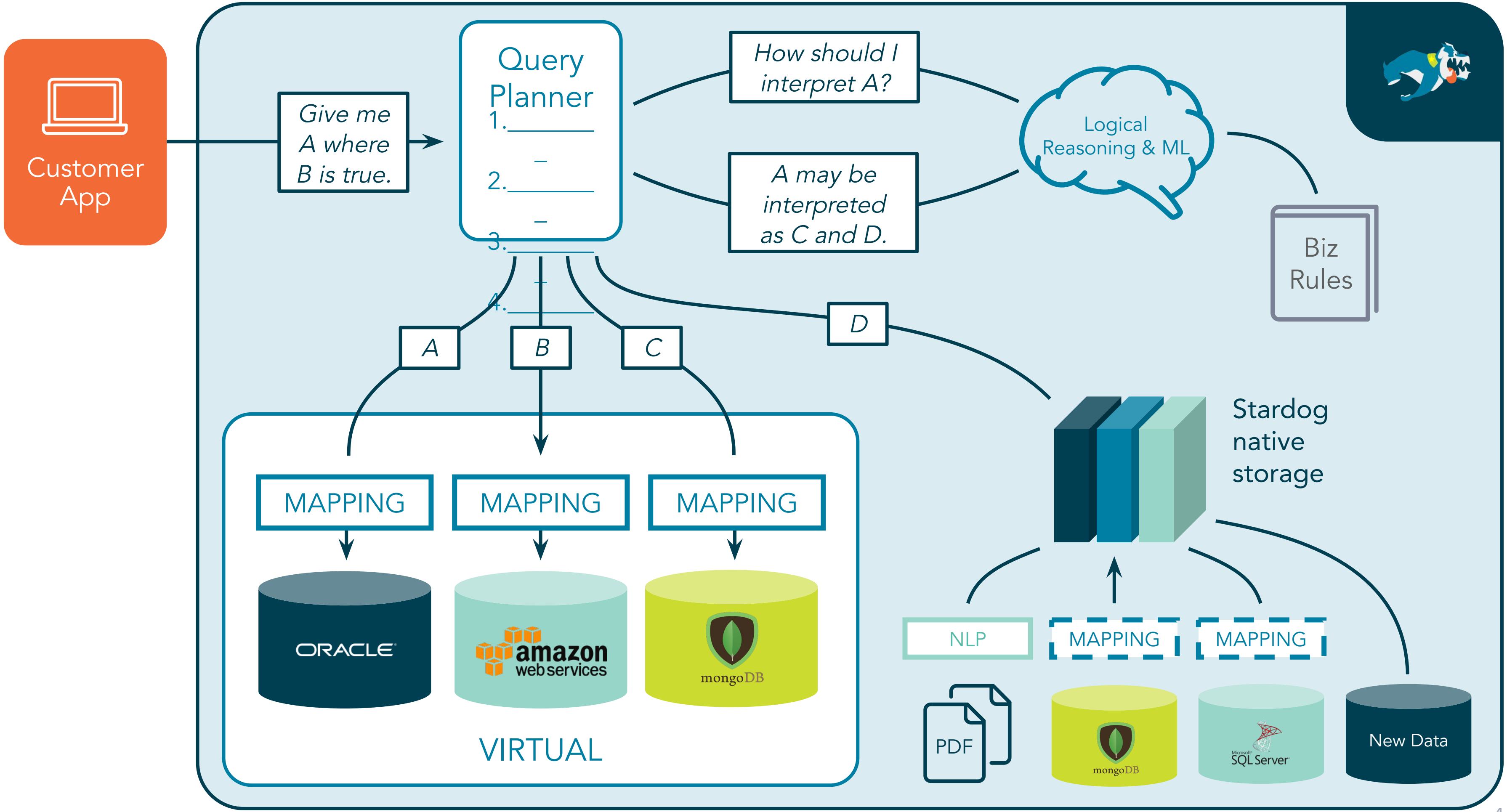
* Big Data Insights and Opportunities, CompTIA

Source: Dun & Bradstreet, Are Data Silos Killing Your Business, May 7 2018

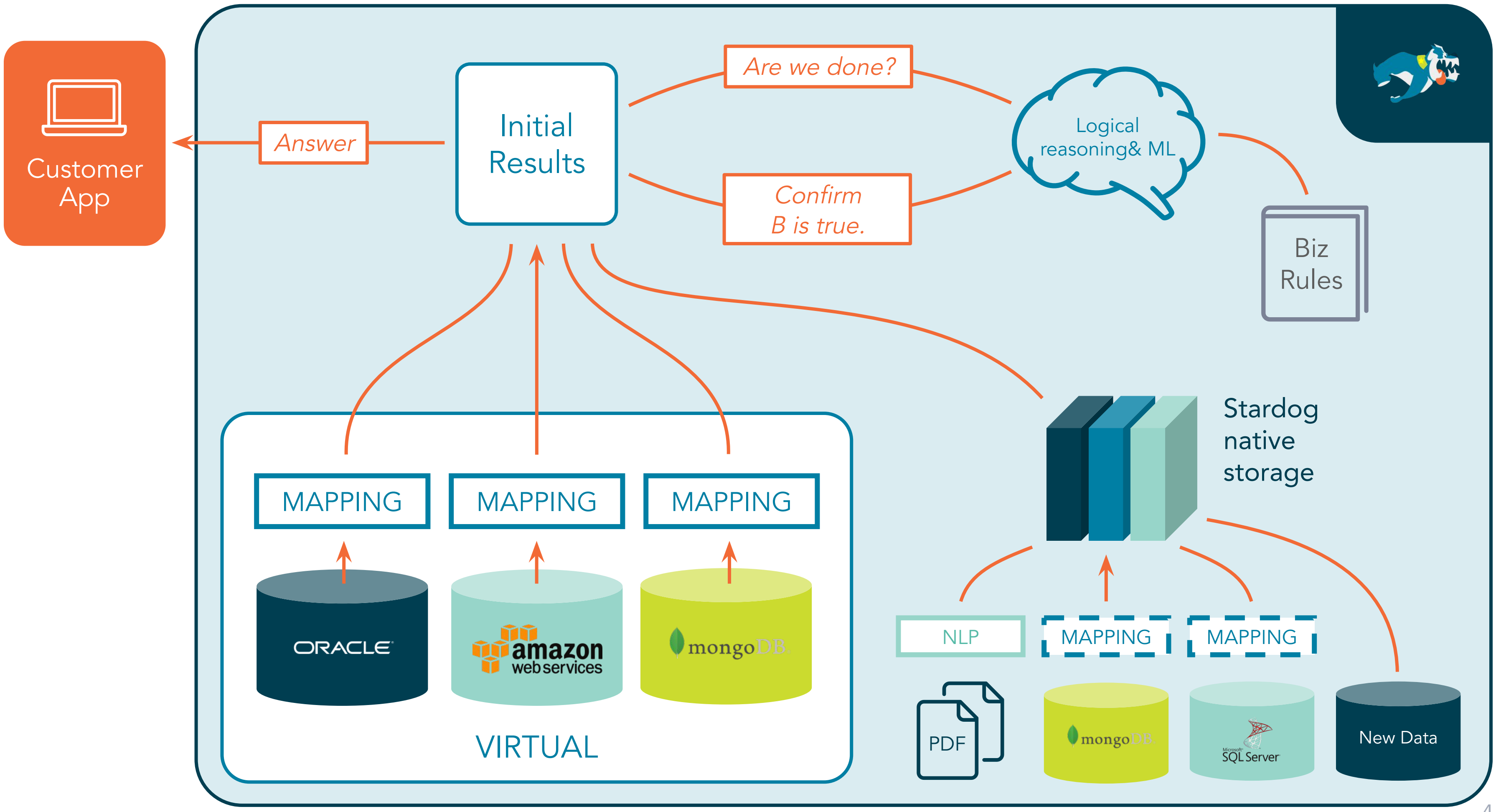
69%

of Customer 360 projects **FAIL**

STARDOG

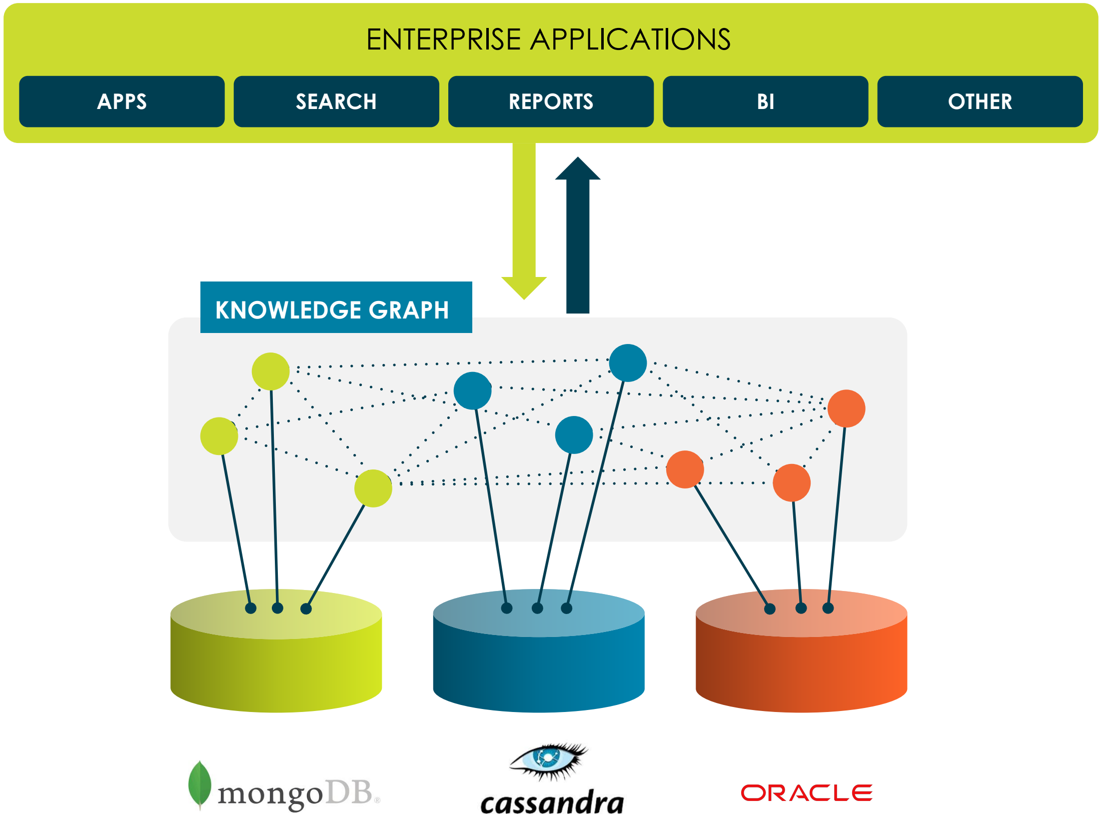


Customer App
Give me A where B is true.
Query Planner
1.
2.
3.
4.
How should I interpret A?
A may be interpreted as C and D.
Logical Reasoning & ML
Biz Rules
A
B
C
D
MAPPING
MAPPING
MAPPING
ORACLE
amazon web services
mongoDB
VIRTUAL
Stardog native storage
NLP
MAPPING
MAPPING
PDF
mongoDB
SQL Server
New Data


Customer
App
Answer
Initial
Results
Are we done?
Confirm
B is true.
Logical
reasoning& ML
Biz
Rules
Stardog
native
storage
MAPPING
MAPPING
MAPPING
ORACLE
amazon
web services
mongoDB
VIRTUAL
NLP
MAPPING
MAPPING
PDF
mongoDB
SQL Server
New Data

What is a knowledge graph?

A knowledge graph leverages graph technology and a declarative model to connect, query, and retrieve data.



Examples of data

STARDOG