Neuron ✸ Sphere®

**How polyglot storage cost me a job,
almost killed data modeling,
(and started my quest for one data model to rule them all)**

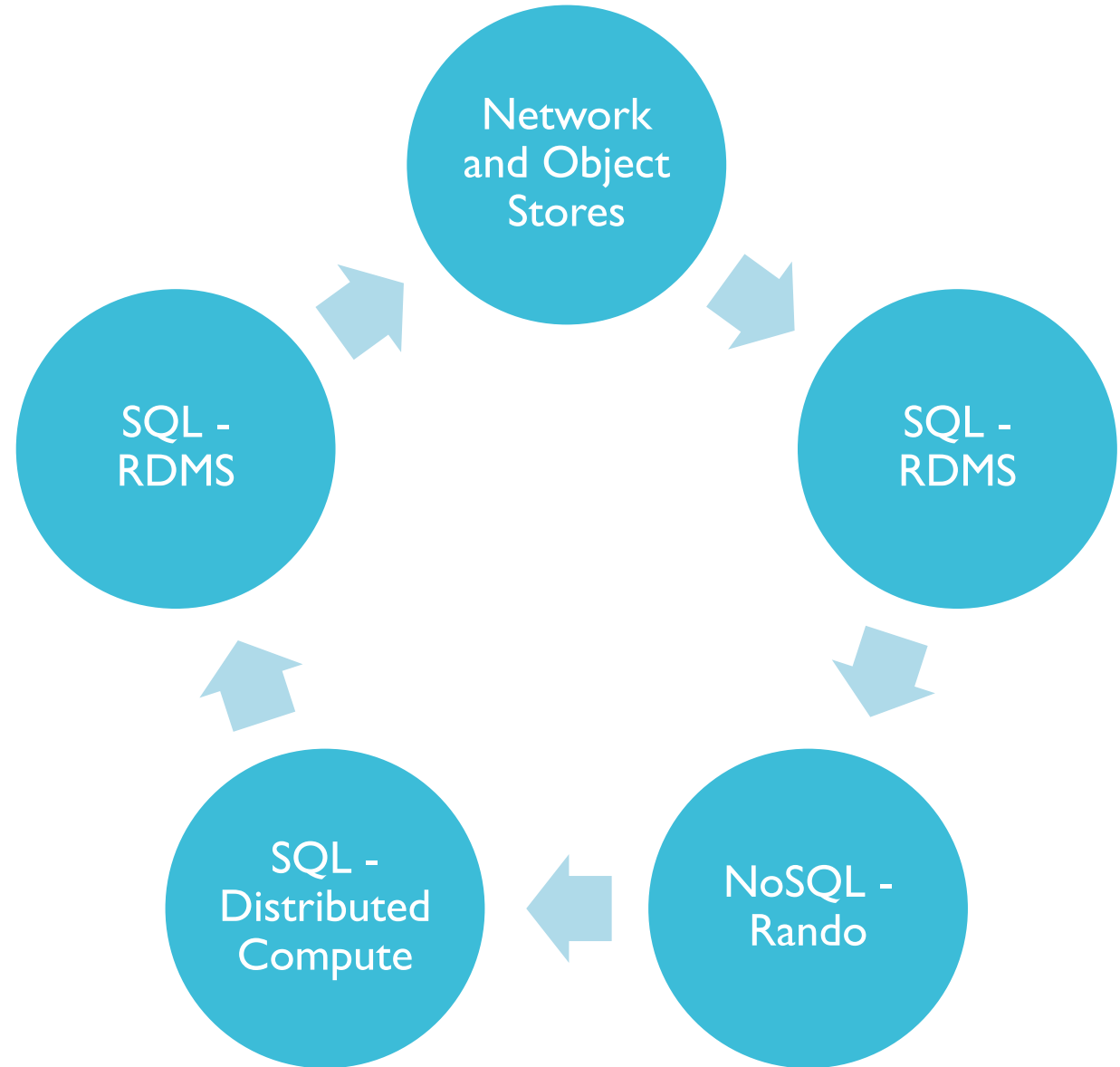Brian Greene @ Data Day Texas 2024

Why am I here?

- Walked with data in primary dimensions
- A casual relationship with infrastructure and machines
- Ignoring the schism between software and data
- Polyglot storage lost me a job*

Neuron Sphere®

History Repeats Itself

You can't escape SQL or MDA

Network and Object Stores
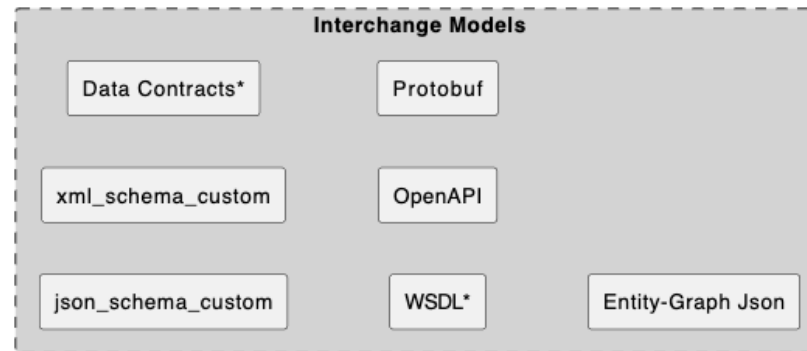
SQL - RDMS

SQL - RDMS

NoSQL - Rando

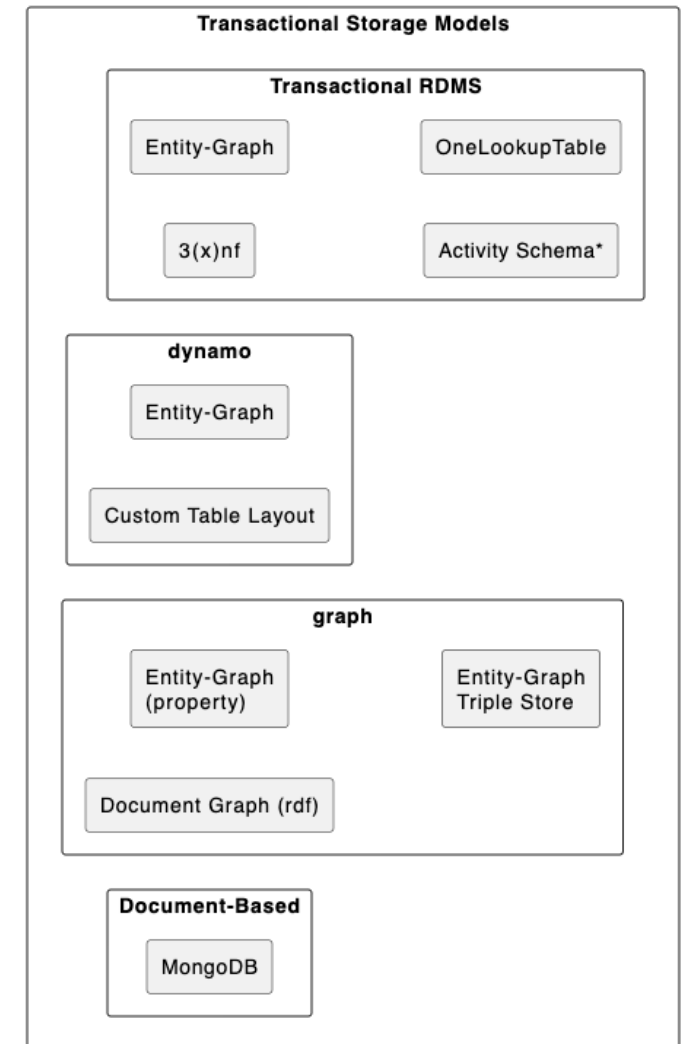SQL - Distributed Compute

Neuron Sphere®

Polyglot and Transfer Models

Polyglot storage helped in one place and brought an exponential new degree of model mismatches or outright unmodeled systems.

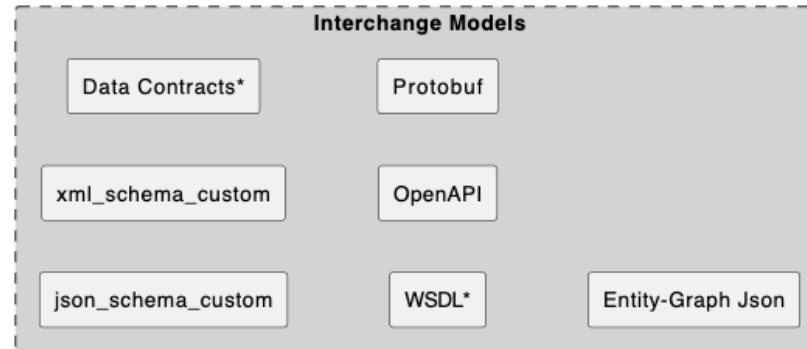An explosion in data transmission requirements and techniques further complicated the issue

Interchange Models

Data Contracts*	Protobuf
xml_schema_custom	OpenAPI
json_schema_custom	WSDL* Entity-Graph Json

N x M model mappings?

Transactional Storage Models

Transactional RDMS

Entity-Graph	OneLookupTable
3(x)nf	Activity Schema*

dynamo

Entity-Graph

Custom Table Layout

graph

Entity-Graph (property)	Entity-Graph Triple Store

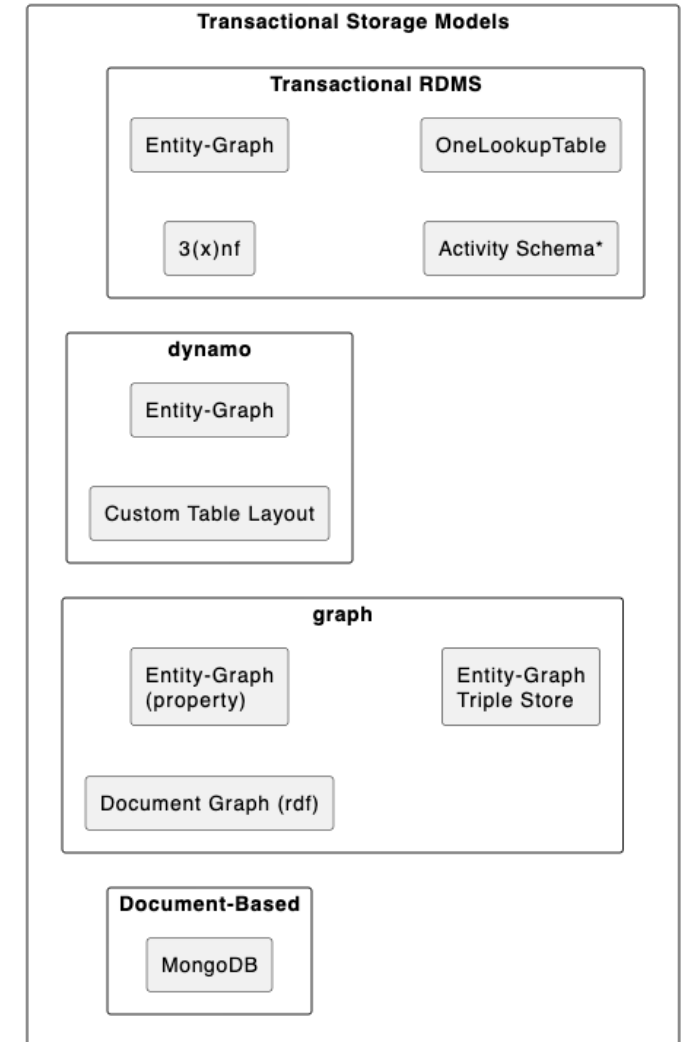Document Graph (rdf)
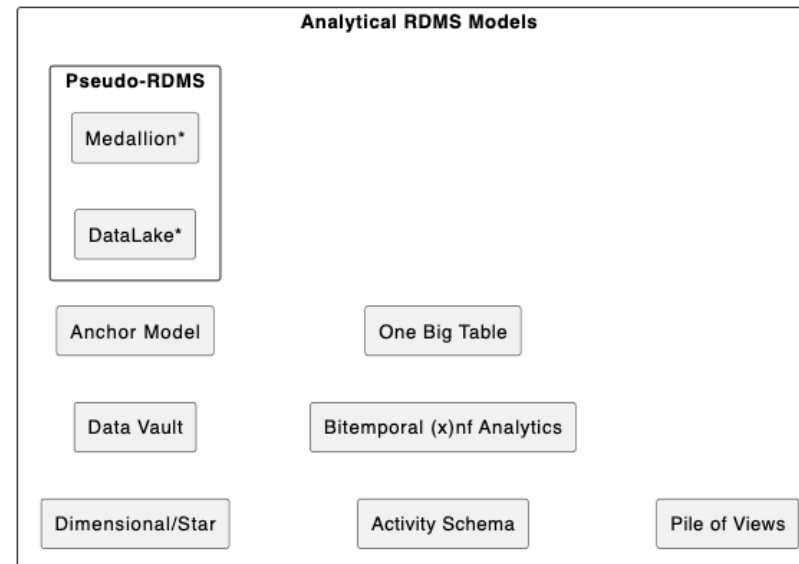
Document-Based

MongoDB

Analytical Models

The challenge is always between the state or message-based data models and those that need to capture history across systems and offer analytical insight

Schema change is inevitable, and different across the modeling ecosystem

Interchange Models

- Data Contracts*
- Protobuf
- xml_schema_custom
- OpenAPI
- json_schema_custom
- WSDL*
- Entity-Graph Json

$N \times M \times A$ *model mappings* ☹

Analytical RDMS Models

Pseudo-RDMS

- Medallion*
- DataLake*

- Anchor Model
- One Big Table
- Data Vault
- Bitemporal (x)nf Analytics
- Dimensional/Star
- Activity Schema
- Pile of Views

Transactional Storage Models

Transactional RDMS

- Entity-Graph
- OneLookupTable
- 3(x)nf
- Activity Schema*

dynamo

- Entity-Graph
- Custom Table Layout

graph

- Entity-Graph (property)
- Entity-Graph Triple Store
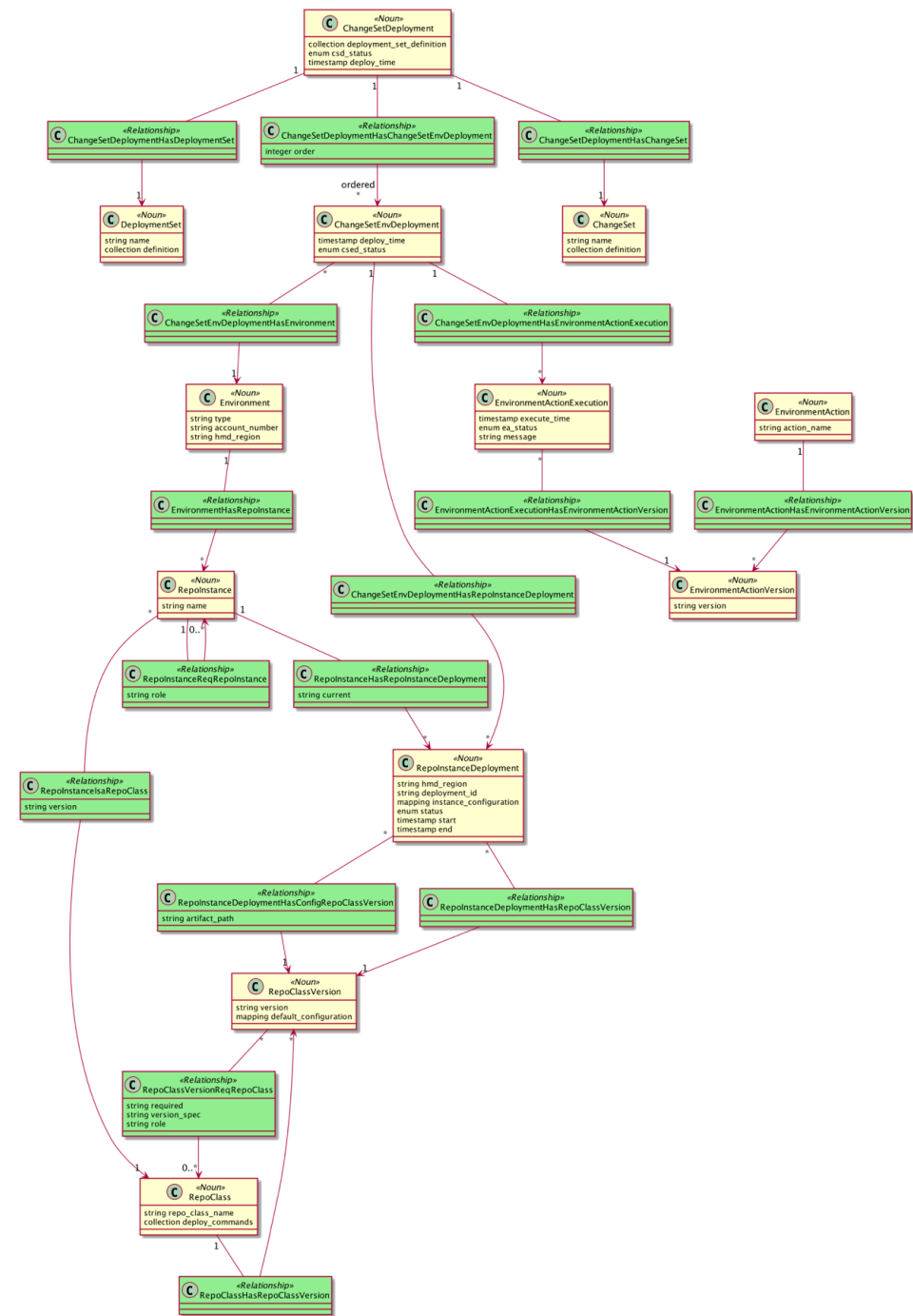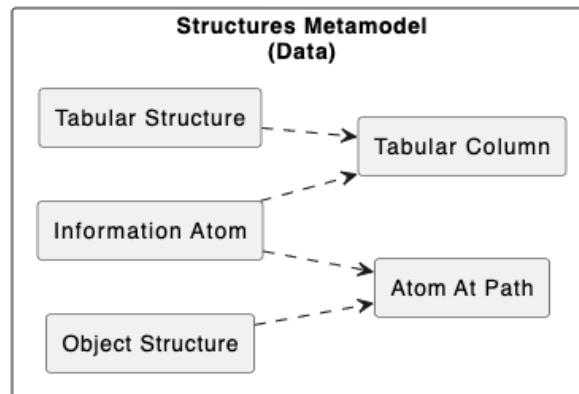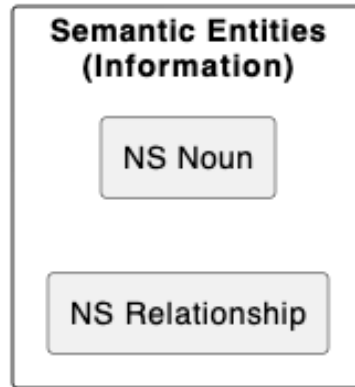- Document Graph (rdf)

Document-Based

- MongoDB

One Model to Rule Them All?

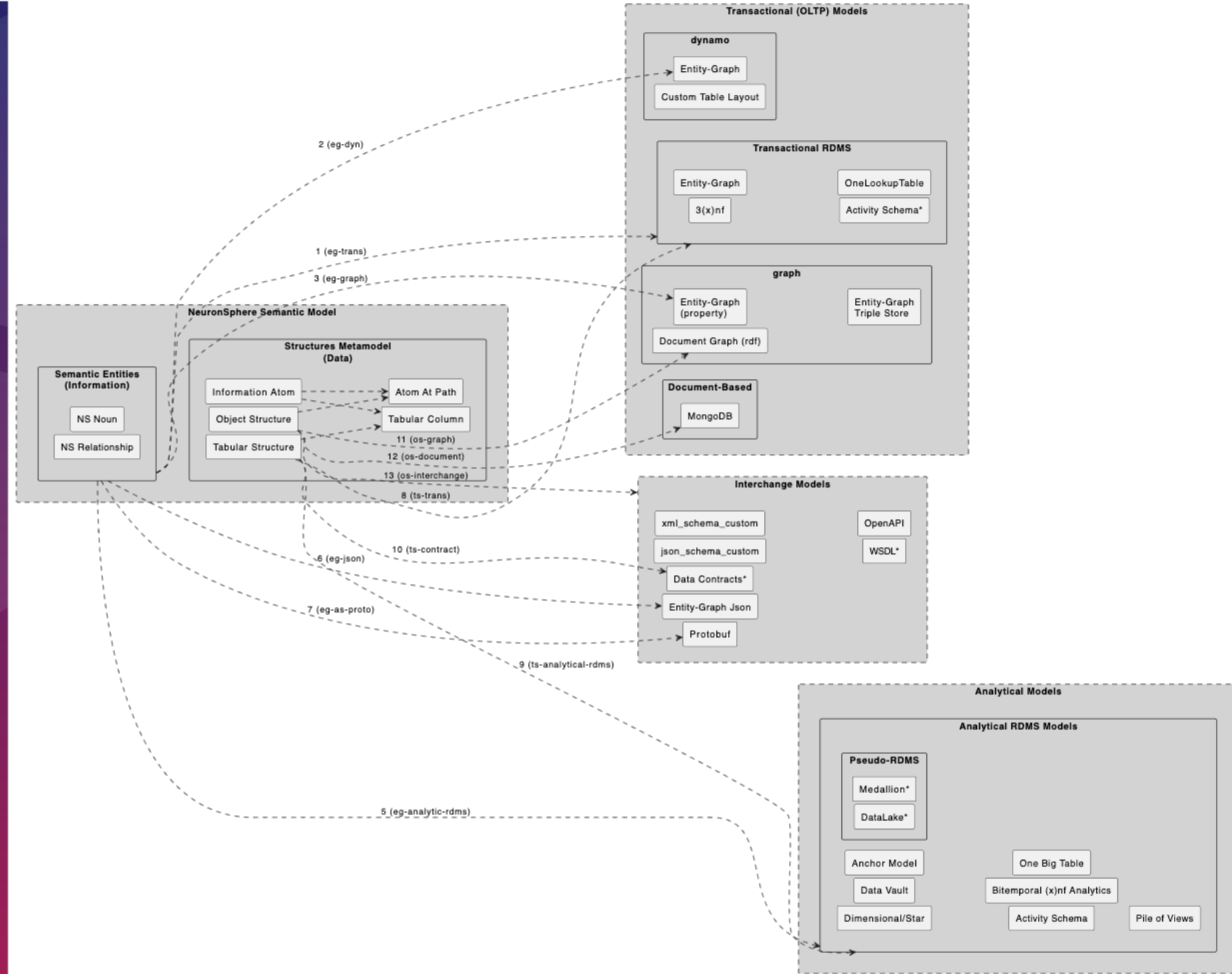Use a labeled property graph to avoid impedance mismatches

Augment the graph as needed for projection to other models
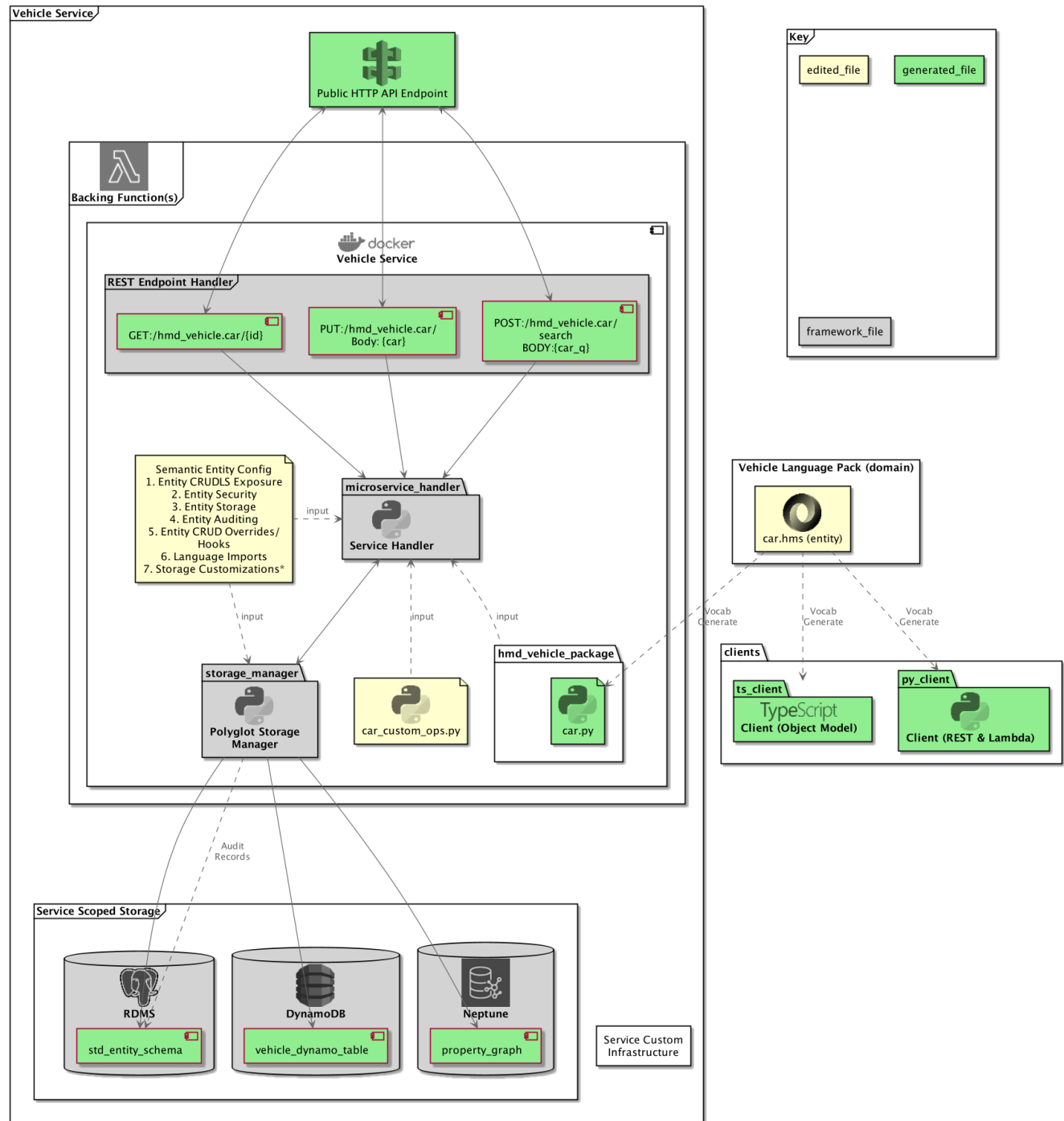
One Model to Link Them All

Augment the graphs as needed for projection to other models

Embrace the polyglot

Entity Storage Managers

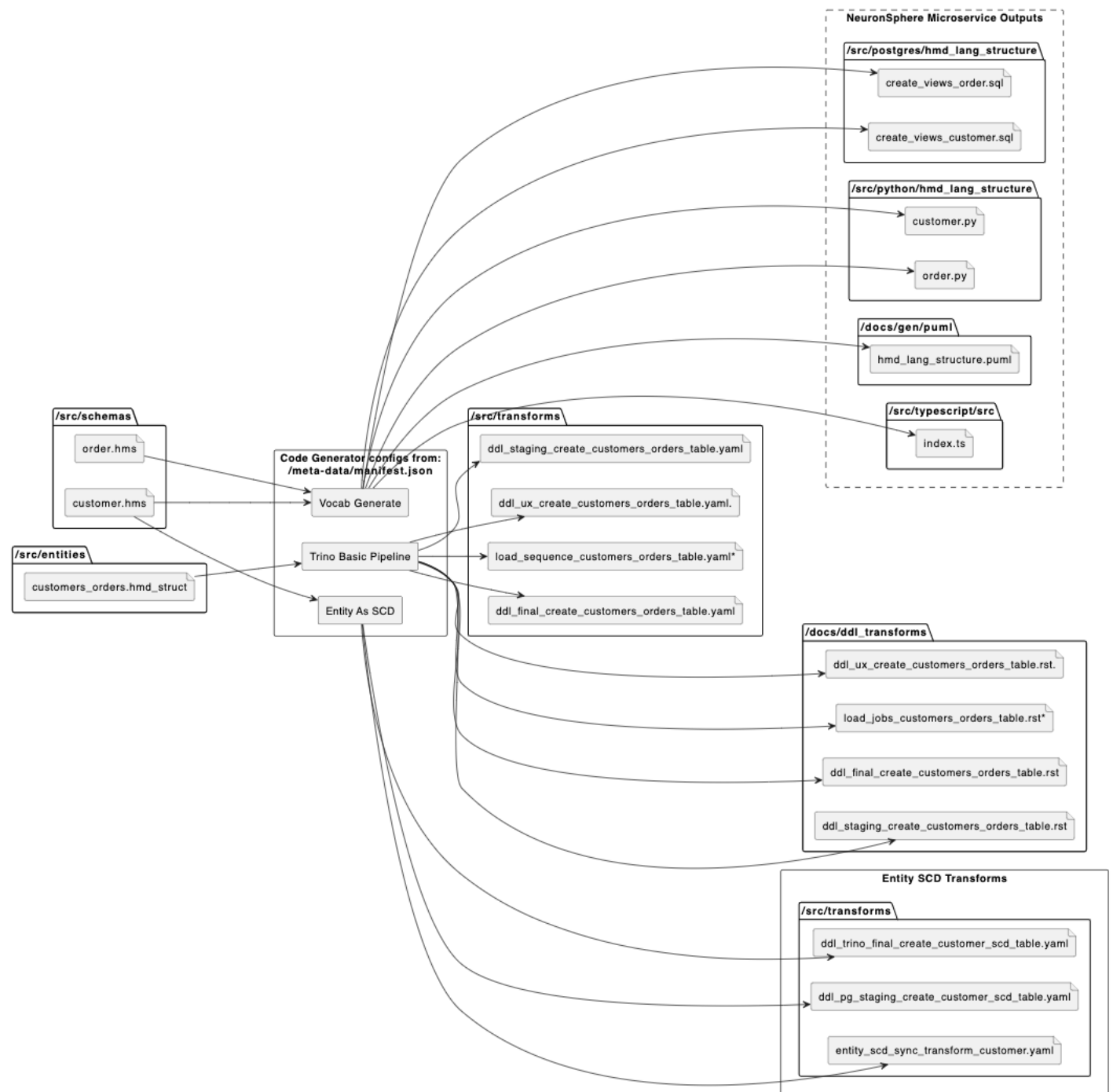An Entity to Polyglot Storage Manager is the MVP
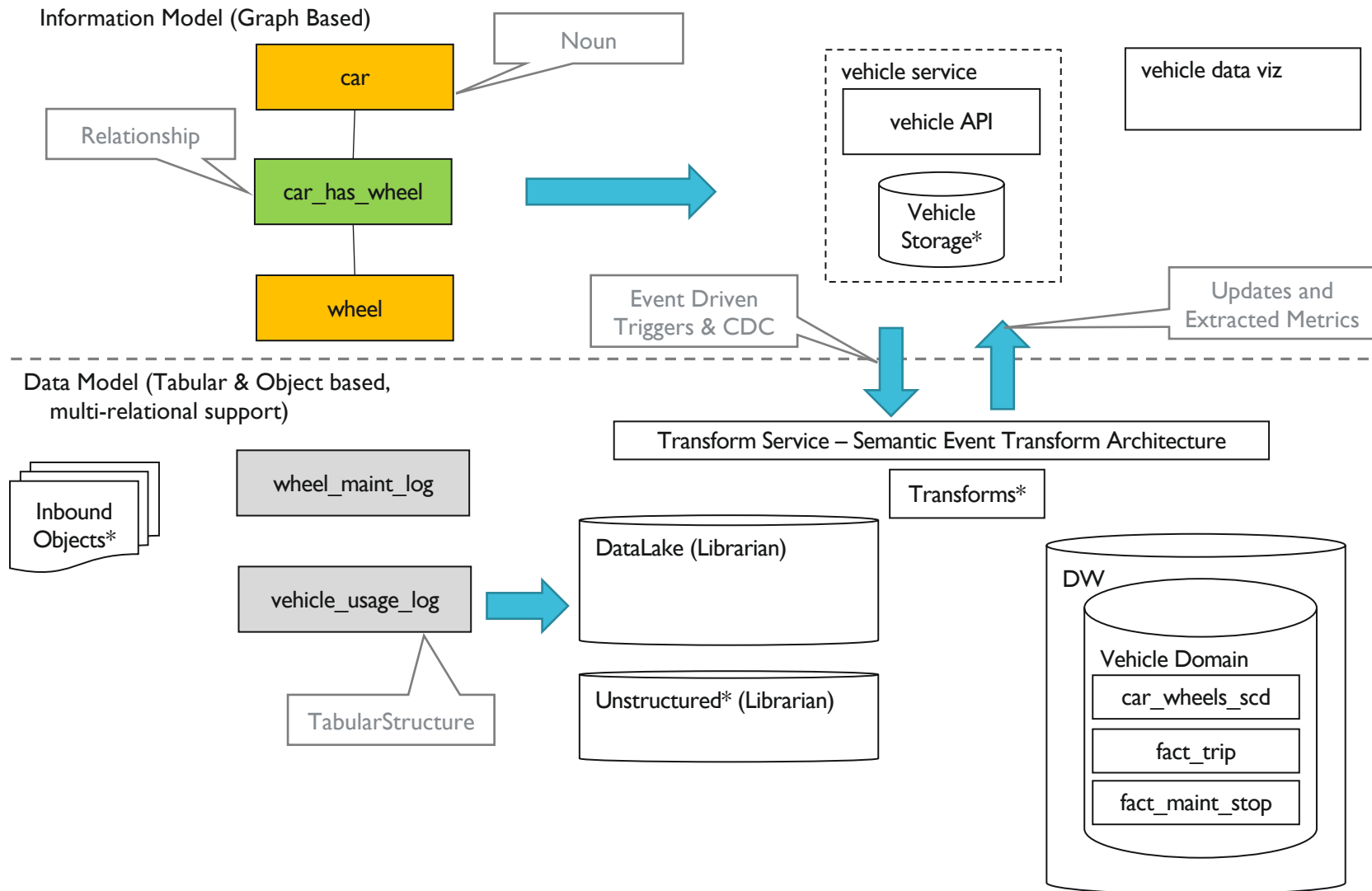

8

Model Driven Everything

Use the same data models and techniques based on metadata-driven development from software to create an end-to-end modeling and development experience from App and API to data warehouse and pipeline.

Generate model-spanning systems

Generate full-stack Data Ecosystems

This multi-layered data modeling enables "Model Driven Functional Data Engineering", with support for polyglot data lineage and graph analytics

Information Model (Graph Based)

- Noun → car
- Relationship → car_has_wheel
- wheel

vehicle service
- vehicle API
- Vehicle Storage*

vehicle data viz

Event Driven Triggers & CDC

Updates and Extracted Metrics

Data Model (Tabular & Object based, multi-relational support)

Inbound Objects*

wheel_maint_log

vehicle_usage_log

TabularStructure

Transform Service – Semantic Event Transform Architecture

Transforms*

DataLake (Librarian)

Unstructured* (Librarian)

DW

Vehicle Domain
- car_wheels_scd
- fact_trip
- fact_maint_stop

Neuron ❂ Sphere®

Elevate your delivery

Model Driven * based on an augmented graph model offers exponential delivery

1. Find a current cross-model problem in your architecture
2. Extract from that model to a common graph & data definition models
3. Drive the creation of new artifacts (data products?) with metadata-driven software and data delivery techniques

Closing Examples
1. RDMS -> graph DB and new services
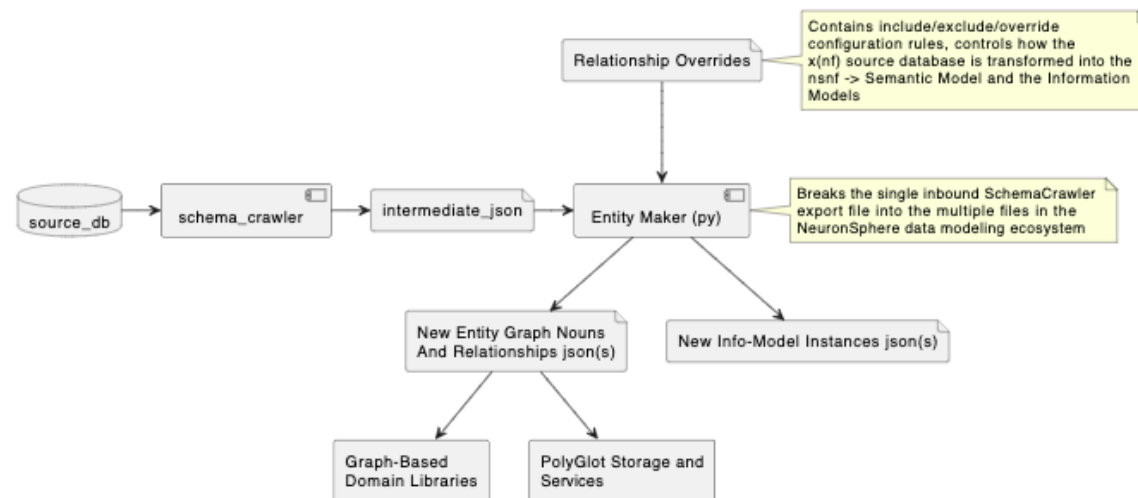
2. Mongo -> Graph, DataLake & RDMS

Platform Engineering for [data]