

intuit.



turbotax



quickbooks



mint

# Knowledge Graph Use Cases @ Intuit

**Jay Yu, Ph.D.**

Distinguished Engineer/Architect, Intuit

Stanford Knowledge Graph Class, 05/26/2020

# Speaker bio

Hands-on **full-stack developer, innovator, strategic thinker, leader** and **evangelist** for new technology and product, *25+ years of experience, 22 patents granted*

Database Engine → Web Services → Tax Engine & Platform → App UI and Products → E2E Data Arch & Strategy → Knowledge Engineering

## Recent focus: How to elevate data to knowledge!

Early-mid 90's



DEPARTMENT OF  
**Computer Sciences**  
UNIVERSITY OF WISCONSIN-MADISON

Parallel Shared-Nothing DB

Late 90's



Teradata  
a division of  NCR

Teradata ORDBMS

Early 2000's



B2Xcom

B2B eCommerce  
Trading Platform

Mid 2000's - now



intuit.

Knowledge Graph  
E2e Data Arch & Strategy  
TT / Mint / Turbo Apps  
Declarative UI Platform  
Data Import Services  
Tax Engine & Platform

# Introduction

# Intuit's mission: Powering the Prosperity of the World

**intuit quickbooks**  
**4.5M** QBO CUSTOMERS

**intuit quickbooks**  
**14M** PAID WORKERS  
 \$180B payroll

**intuit proconnect**  
**>600K** TAX PROS & EXPERTS  
 >30M consumer returns  
 \$50B<sup>1</sup> in refunds

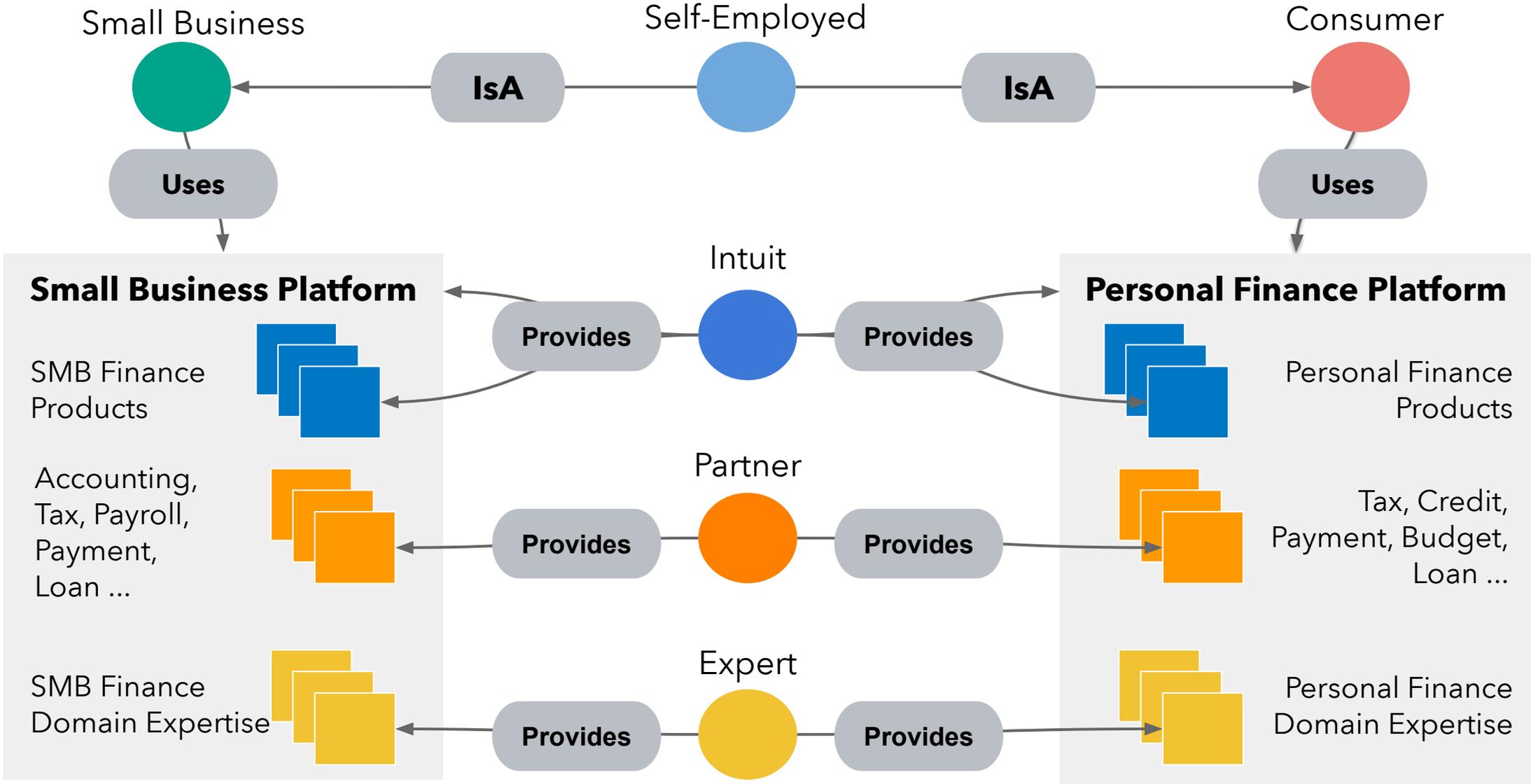


**intuit turbotax**  
**>50M** UNIQUE TAX FILERS  
 41.4M TurboTax units<sup>1</sup>  
 \$82B in refunds<sup>1</sup>

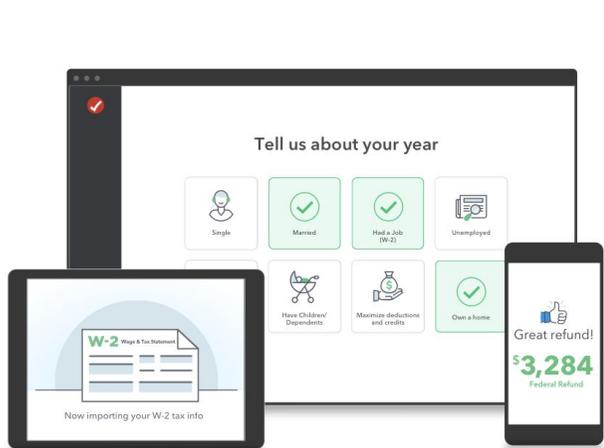
**intuit turbo + intuit mint**  
**37M** REGISTERED USERS

**>20K** FINANCIAL INSTITUTIONS  
 95% consumer coverage

# Intuit's platform described in Knowledge Graph



# AI-Driven Platform Example: A Smart TurboTax Assistant Powered by Intuit's Tax Knowledge Engine and CUI Platform





## TurboTax Mobile App

File smarter from anywhere

Your maximum refund is now at your fingertips with the TurboTax app. Seamlessly switch between your phone, tablet, or computer as you do your taxes.

[Learn more](#)



Knowledge Graph played a huge role in delivering a **personalized experience** to get ***taxes done right*** with ***minimal effort*** and ***high confidence***

# Knowledge Graph definition - our approach

## Webster's Dictionary

### Definition of *knowledge*

- 1 a (1) : the fact or condition of knowing something with familiarity gained through experience or association  
(2) : acquaintance with or understanding of a science, art, or technique
- b (1) : the fact or condition of being aware of something  
(2) : the range of one's information or understanding  
*// answered to the best of my *knowledge**
- c : the circumstance or condition of apprehending truth or fact through reasoning : COGNITION
- d : the fact or condition of having information or of being learned  
*// a person of unusual *knowledge**

### 2 a : **Definition of *graph* (Entry 1 of 4)**

- 1 : a diagram (such as a series of one or more points, lines, line segments, curves, or areas) that represents the variation of a variable in comparison with that of one or more other variables
- b c
- 2 : the collection of all points whose coordinates satisfy a given relation (such as a function)
- 3 : a collection of vertices and edges that join pairs of vertices

### *knowledge*

**What:** information, truth, facts, condition, principles, understanding

**Who:** mankind

**How:** apprehending, reasoning, learning, acquired

**From:** experience, association

### *graph*

**What:** vertices, edges

**How:** edge joins pairs of vertices

# Knowledge Graph definition - our perspective

## *knowledge*

**What:** information, truth, facts, condition, principles, understanding

**Who:** mankind

**How:** apprehending, reasoning, learning, acquired

**From:** experience, association

## *graph*

**What:** vertices, edges

**How:** edge joins pairs of vertices

## Webster's Dictionary

A graph ...

- to represent **data** (facts, information) and **logic** (conditions) into **concepts** as vertices, and **relationships** between concepts as edges,
- created/enriched from **experience** and **learning**,
- used for **understanding** and **reasoning**,
- by **human** and **machine**

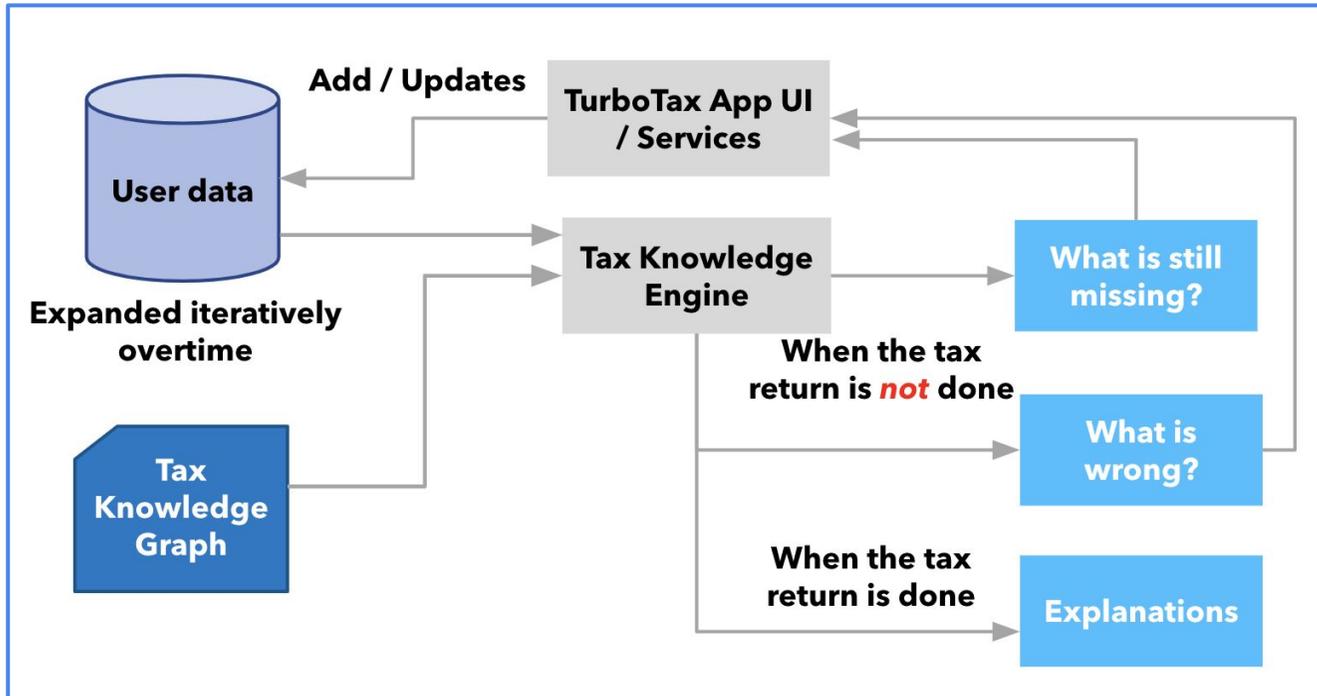
Foundation to bring AI approaches together:

- **Symbolic AI (Explainable / Semantic)**
- **Machine Learning**

**DATA GRAPH + LOGIC GRAPH**

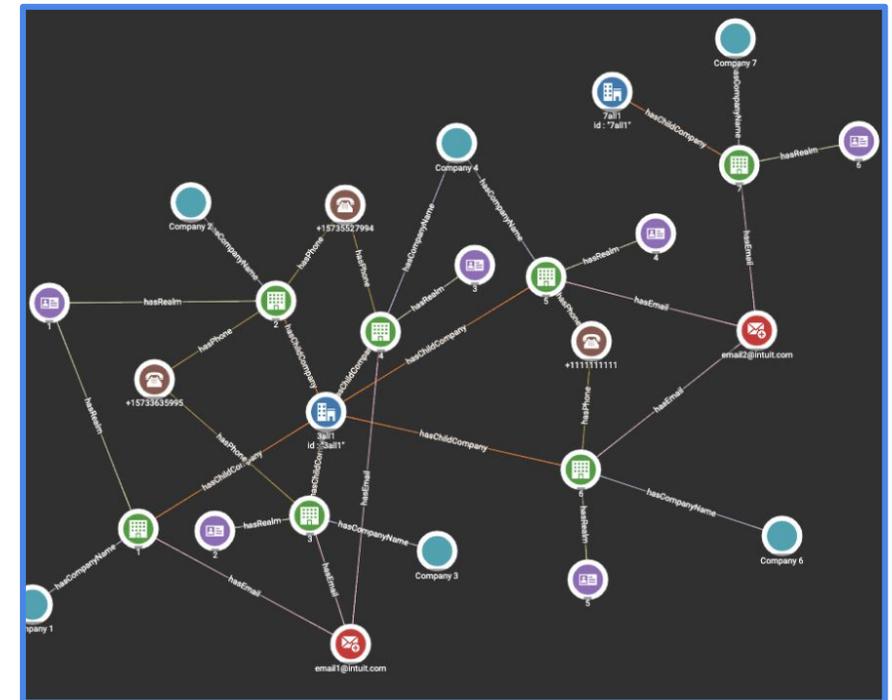
# Data + Logic Knowledge Graph Use Cases at Intuit

## TAX PROGRAMMING (LOGIC GRAPH)



Intuit Knowledge Engine Platform

## ENTITY RESOLUTION (DATA GRAPH)

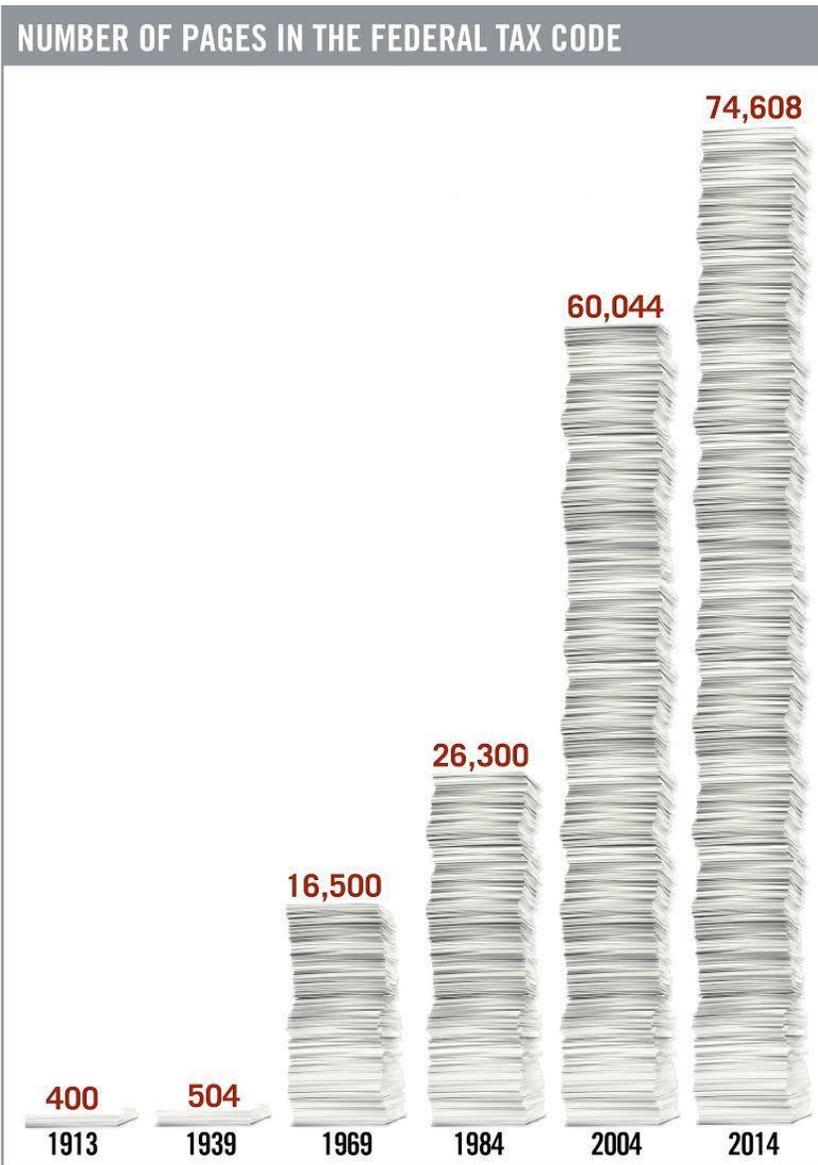


Intuit Data Curation Platform

Other use cases: Fraud Detection, Customer 360 Graph, Company Graph, Product Taxonomy ...

# Use Case #1: Tax Programming (LOGIC GRAPH)

# The big challenge



- 80,000+ pages of instructions and requirements
- Thousands of tax forms can appear in a tax return
- Changes yearly, could change in the middle of a tax filing period
- 50,000+ TurboTax UI screens

How to **scale the tax logic development** for a highly **personalized experience** to get **taxes done done** with **minimal effort** and **high confidence**?

# Conventional approach - procedural programming

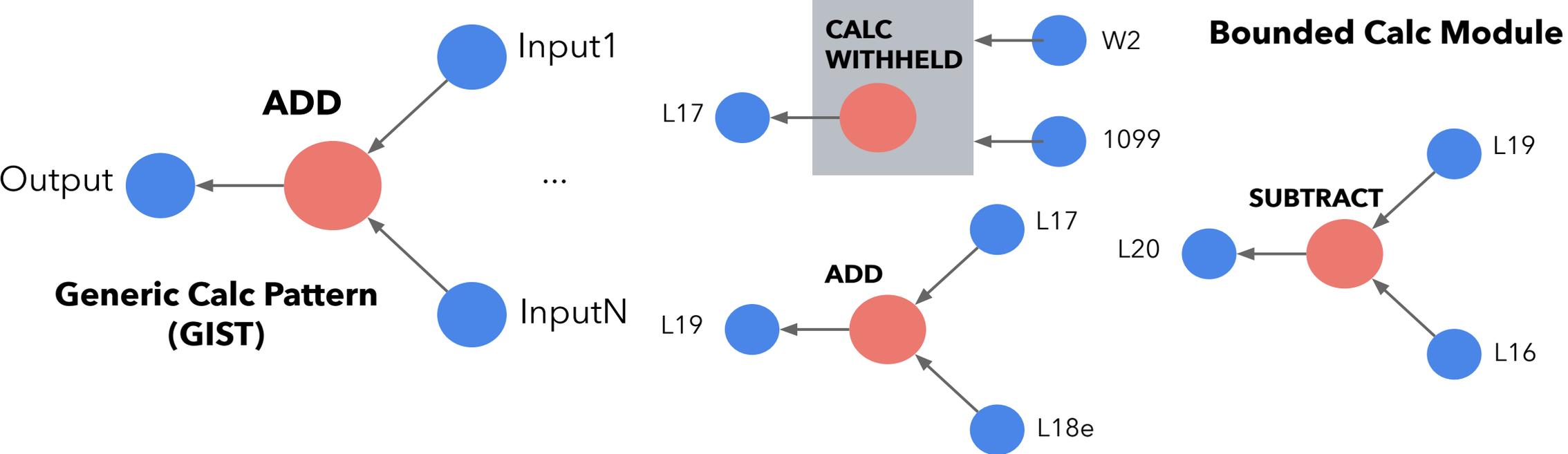
17	Federal income tax withheld from Forms W-2 and 1099 . . . . .
19	Add lines 17 and 18e. These are your <b>total payments</b> . . . . .
20	If line 19 is more than line 16, subtract line 16 from line 19. This is the amount you <b>overpaid</b> .

```
L17 = getTaxWithheld() # from W-2 and 1099
L19 = L17 + L18e # total payments
if L19 > L16:
    L20 = L19 - L16 # refund
else:
    L20 = 0
```

- Procedural programming
- Tops-down sequential execution
- All inputs required to calc
- Implicit explainability hidden in code

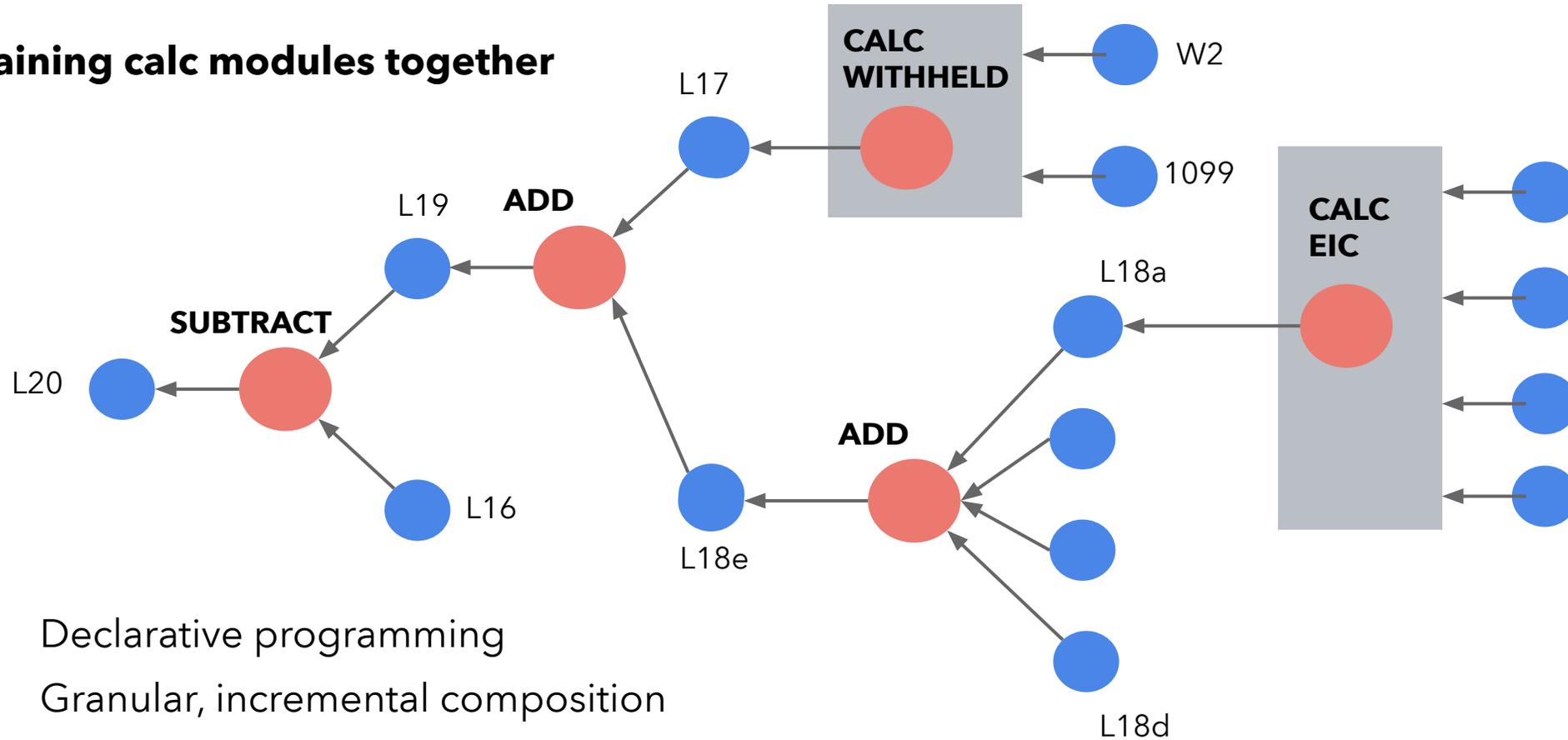
# Knowledge Graph approach

17	Federal income tax withheld from Forms W-2 and 1099
19	Add lines 17 and 18e. These are your <b>total payments</b>
20	If line 19 is more than line 16, subtract line 16 from line 19. This is the amount you <b>overpaid</b>



# Calculation Graph

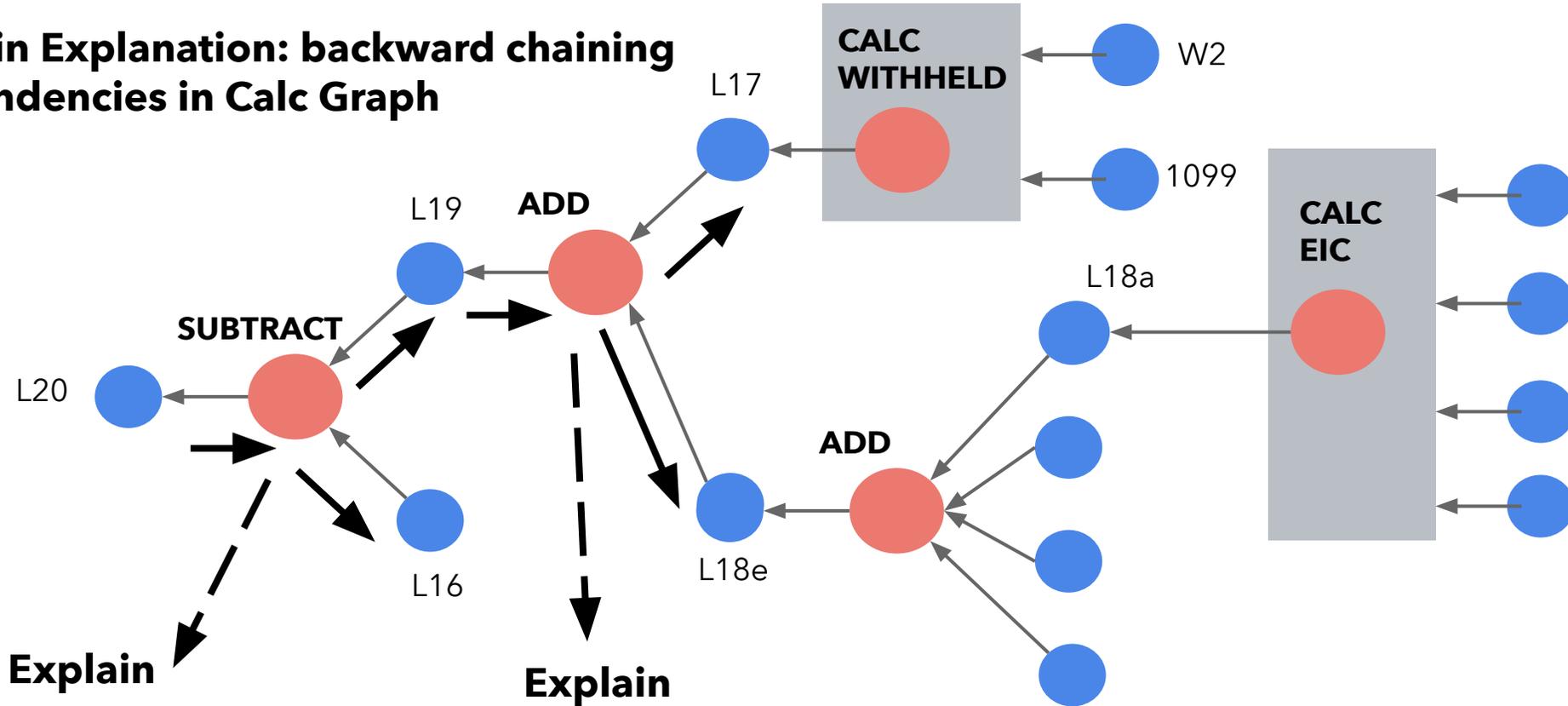
## Chaining calc modules together



- Declarative programming
- Granular, incremental composition
- Visible calc dependency and data flow
- Built-in and intuitive explainability

# Explainability

**Built-in Explanation: backward chaining dependencies in Calc Graph**



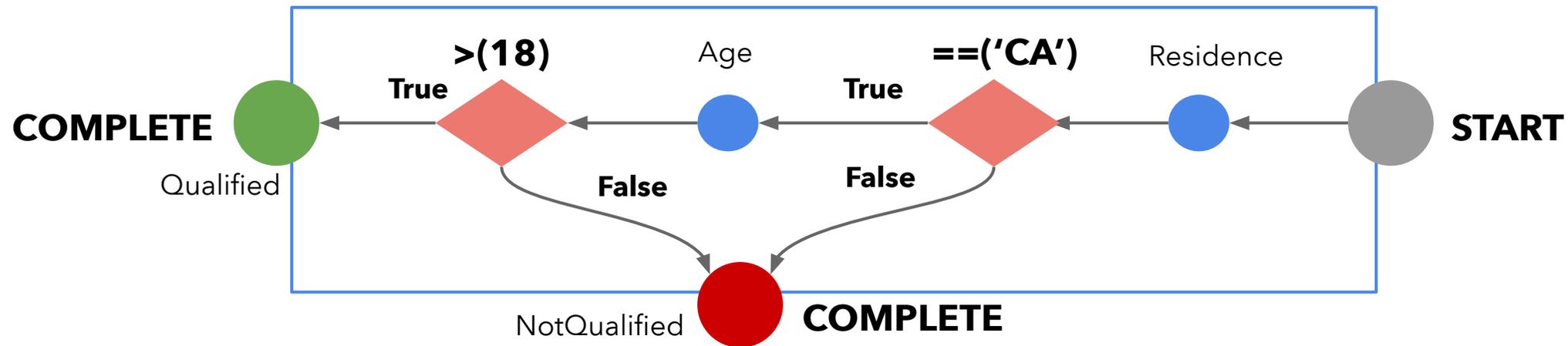
**L20** is the result of **SUBTRACT** **L16** from **L19**

**L19** is the result of **ADD** **L17** and **L18e**

# Completeness Graph

## Sample tax benefit rule:

1. If a person is not a resident of California, he/she is not qualified;
2. Otherwise, he/she must be older than 18 to qualify for the benefit.



What data is missing before we can complete the decision?

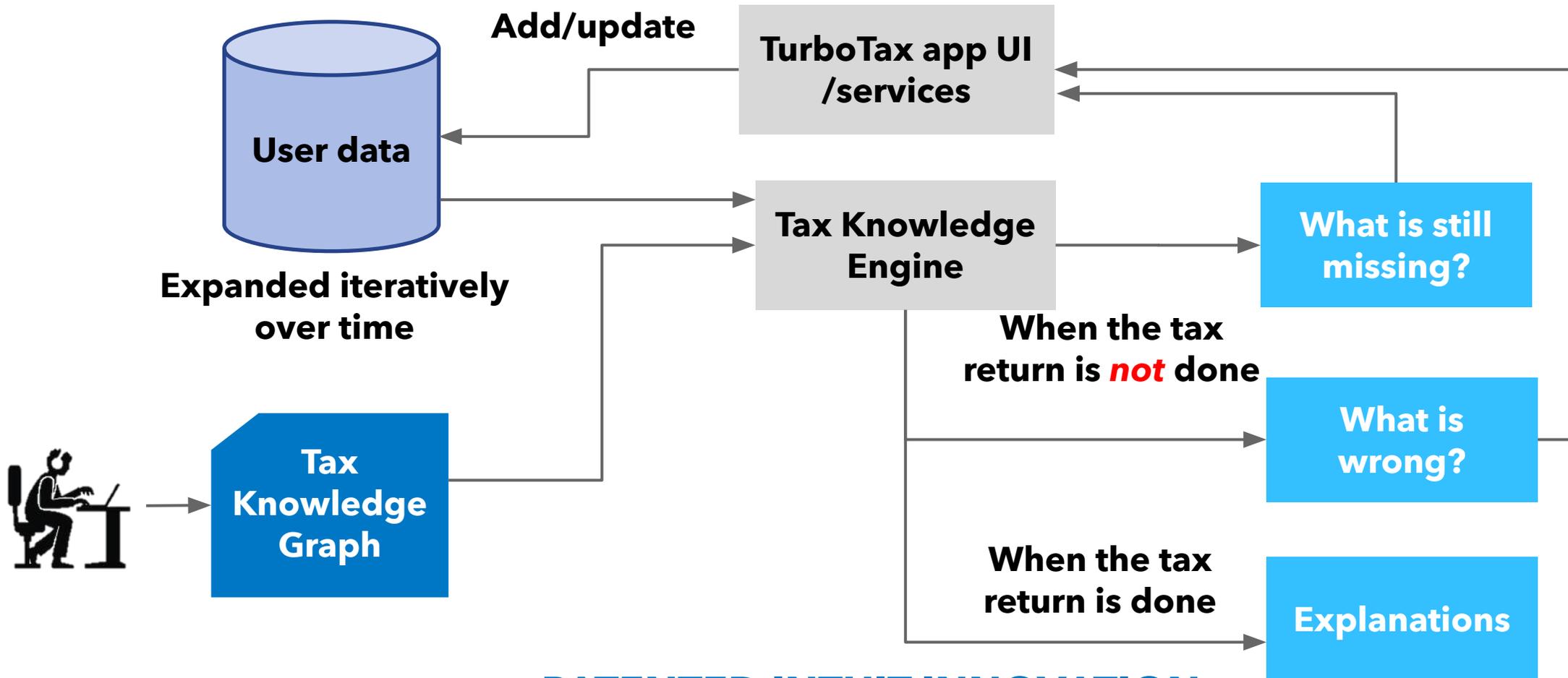
Given age = 17, is residence info necessary to determine qualification?



**Data-driven, personalized experience** that *minimizes questions* for data collection

# Tax Knowledge Engine

Based on Tax Knowledge Graph, at **any moment** and for **any (partial) user data**, be able to tell what's **missing**, what's **wrong**, and **explain** back how it's done.

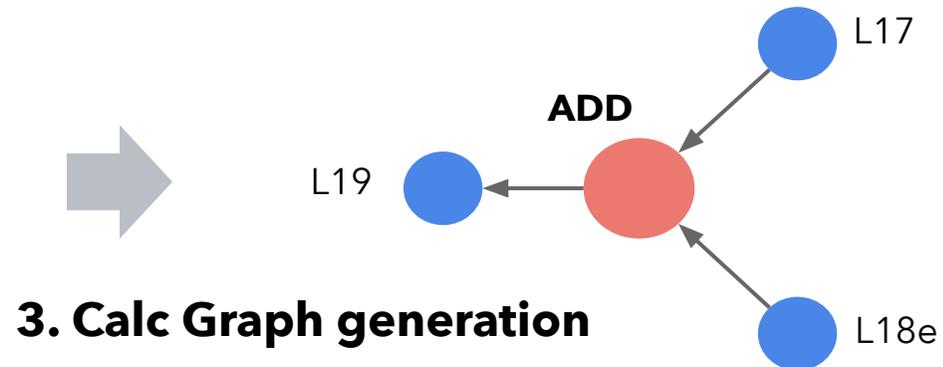
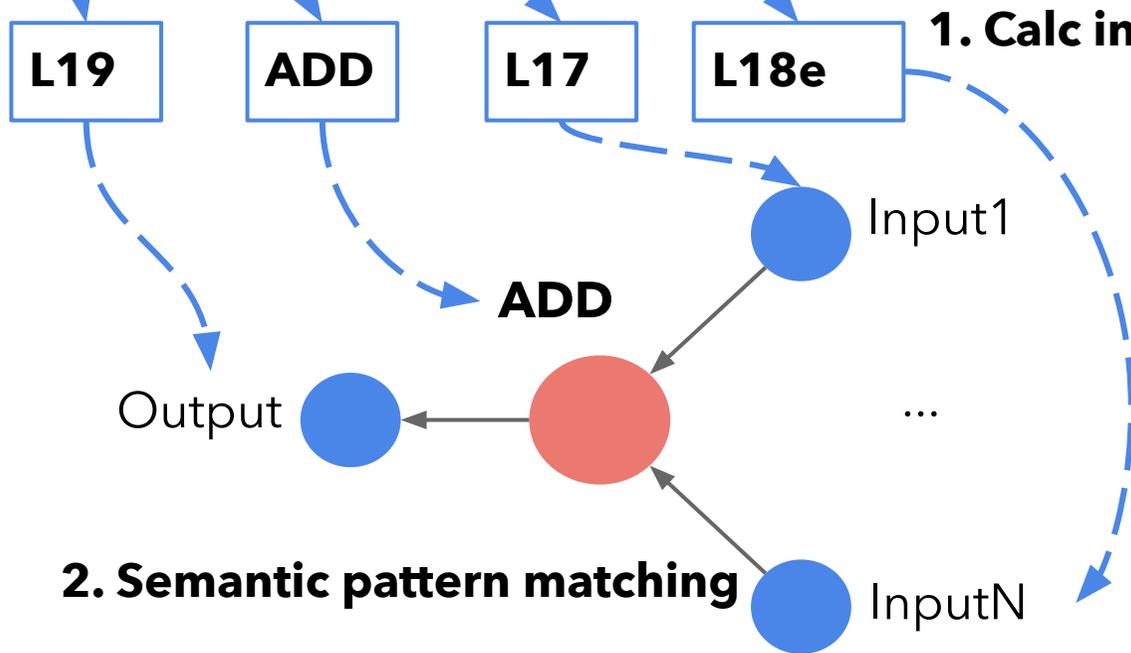


# Automating Calc Graph construction

17 Federal income tax withheld from Forms W-2 and 1099 . . . . .

19 Add lines 17 and 18e. These are your **total payments** . . . . .

20 If line 19 is more than line 16, subtract line 16 from line 19. This is the amount you **overpaid** .



# Comparison between 2 approaches

	<b>PROCEDURAL</b>	<b>KNOWLEDGE GRAPH</b>
<b>Logic</b>	Procedural	Declarative
<b>Execution</b>	Sequential	Dependency-driven
<b>Modularity</b>	By best practice	By design
<b>Explainability</b>	Blackbox	Built-in
<b>Personalization</b>	Hard-coded, limited	Dynamic, automatic
<b>Automation</b>	Not manageable	Manageable
<b>Testability</b>	Harder	Easier
<b>Optimized for</b>	Exceptional complex logic	Most tax logic
<b>Target Developer</b>	Engineers	Domain experts

**Clear Winner!**

# TurboTax Explain Why results

First year production deployment (2015) with initial limited tax topics coverage, used by **16M customers**:

**46%**

**Reduction in care contacts for W-2**

**77%**

**Customers found it helpful**

# Use Case #2: Entity Resolution (DATA GRAPH)

# The big data challenge: quality, silo ...

DATA SILO 1

fName	phone	email
Bob	923-456-7890	xyz@domain1.com
Robert	123-456-7890	xyz@domain1.com

DATA SILO 2

F_Name	Telephone	Email
Robert	(923)456-7890	xyz@domain2.com

DATA SILO 3

Name	mobile	email
Robert	+11234567890	xyz@domain2.com

DATA SILO 4

nm	ph	em
Robet		zyz1@domain2.com

- Different field names
- Multiple values
- Missing value
- Wrong value (by mistake)
- ...

How many customer records across data silos are duplicates?  
How can they be merged ?

# Preparation - data cleansing

fName	phone	email
Bob	923-456-7890	xyz@domain1.com
Robert	123-456-7890	xyz@domain1.com

F_Name	Telephone	Email
Robert	(923)456-7890	xyz@domain2.com

Name	mobile	email
Robert	+11234567890	xyz@domain2.com

fname	phone	email
Bob	9234567890	xyz@domain1.com
Robert	1234567890	xyz@domain1.com

Robert	9234567890	xyz@domain1.com
--------	------------	-----------------

Robert	1234567890	xyz@domain2.com
--------	------------	-----------------

## STANDARDIZATION

# Procedural approach (1 of 2)

fname	phone	email
Bob	9234567890	xyz@domain1.com
Robert	1234567890	xyz@domain1.com

Robert	9234567890	xyz@domain1.com
--------	------------	-----------------

Robert	1234567890	xyz@domain2.com
--------	------------	-----------------

1. Pairwise compare



Customer1	Customer2	Score
1Bob	1_Rober	0.5
1Bob	2_Robert	1.0
1Bob	3_Robert	0.0
1Robert	2_Robert	1.0
1Robert	3_Robert	1.0
2Robert	3_Robert	0.5

3. Repeat



2. Merge



CustomerID	Name	Telephone	Email
1Bob_2Robert	Bob	9234567890	xyz@domain1.com
	Robert		
1Robert_2Robert	Robert	1234567890	xyz@domain1.com
		9234567890	xyz@domain2.com
1Robert_3Robert	Robert	1234567890	xyz@domain2.com

# Procedural approach (2 of 2)

CustomerID	Name	Telephone	Email
1Bob_2Robert	Bob	9234567890	xyz@domain1.com
	Robert		
1Robert_2Robert	Robert	1234567890	xyz@domain1.com
		9234567890	xyz@domain2.com
1Robert_3Robert	Robert	1234567890	xyz@domain2.com

### 3. Pairwise compare



Customer1	Customer2	Score
1Bob_2Robert	1Robert_2Robert	1.0
1Bob_2Robert	1Robert_3Robert	0.5
1Robert_2Robert	1Robert_3Robert	1.5

### 4. Merge



CustomerID	Name	Telephone	Email
Merged_CustID	Bob	9234567890	xyz@domain1.com
	Robert	1234567890	xyz@domain2.com

### Result after converging

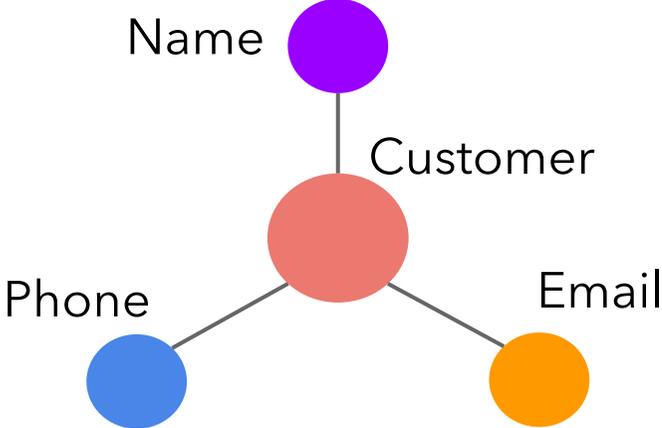
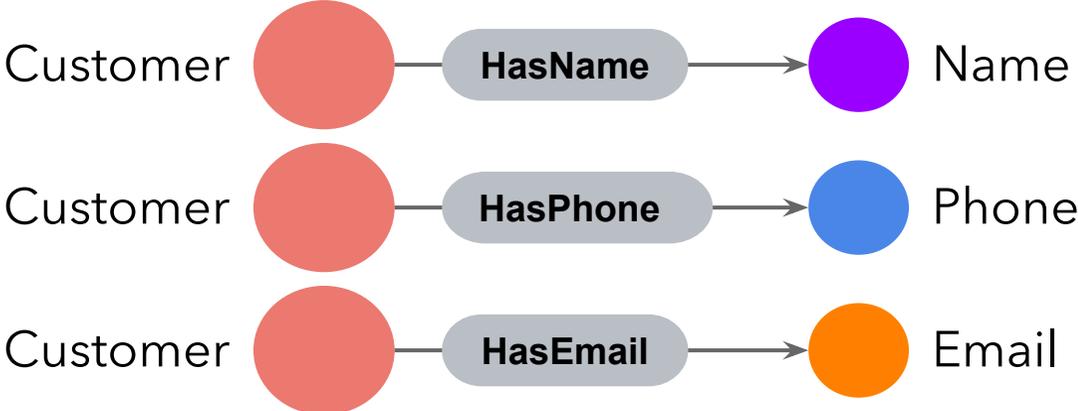
- Purpose-fit procedural logic
- Iterative and sequential execution
- Blackbox processing, hard to explain

# Knowledge Graph approach (1 of 5)

fname	phone	email
Bob	9234567890	xyz@domain1.com
Robert	1234567890	xyz@domain1.com

Robert	9234567890	xyz@domain1.com
--------	------------	-----------------

Robert	1234567890	xyz@domain2.com
--------	------------	-----------------



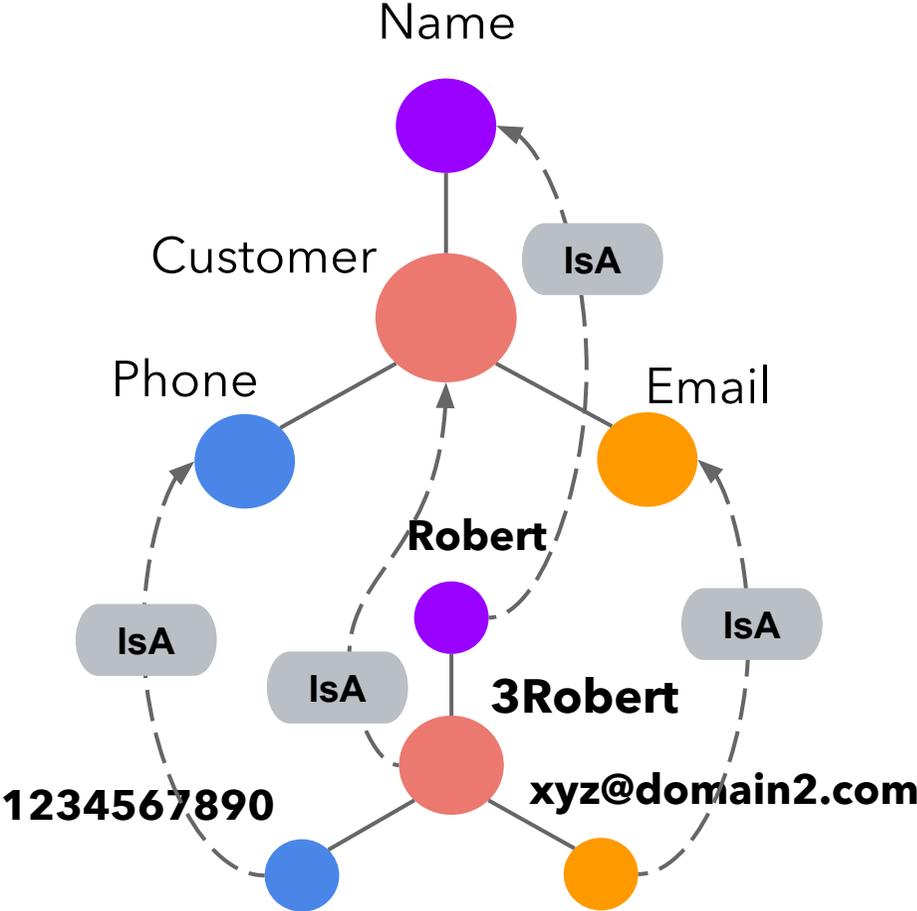
**BUILD ONTOLOGY**

# Knowledge Graph approach (2 of 5)

fname	phone	email
Bob	9234567890	xyz@domain1.com
Robert	1234567890	xyz@domain1.com

Robert	9234567890	xyz@domain1.com
--------	------------	-----------------

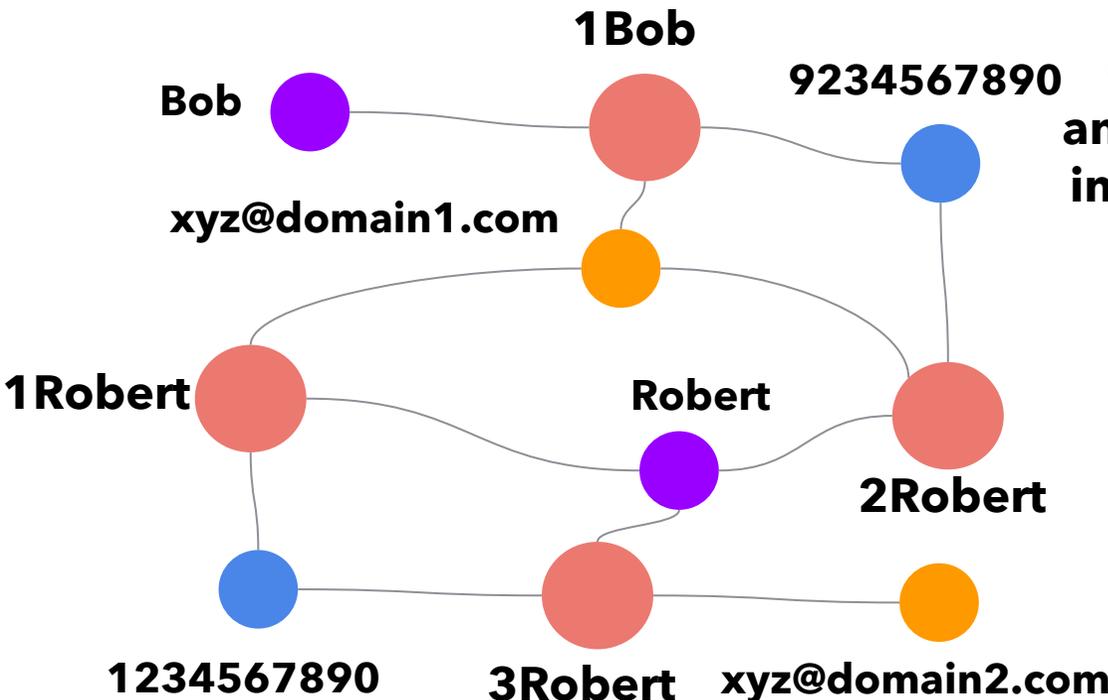
Robert	1234567890	xyz@domain2.com
--------	------------	-----------------



**LOAD KNOWLEDGE GRAPH**

# Knowledge Graph approach (3 of 5)

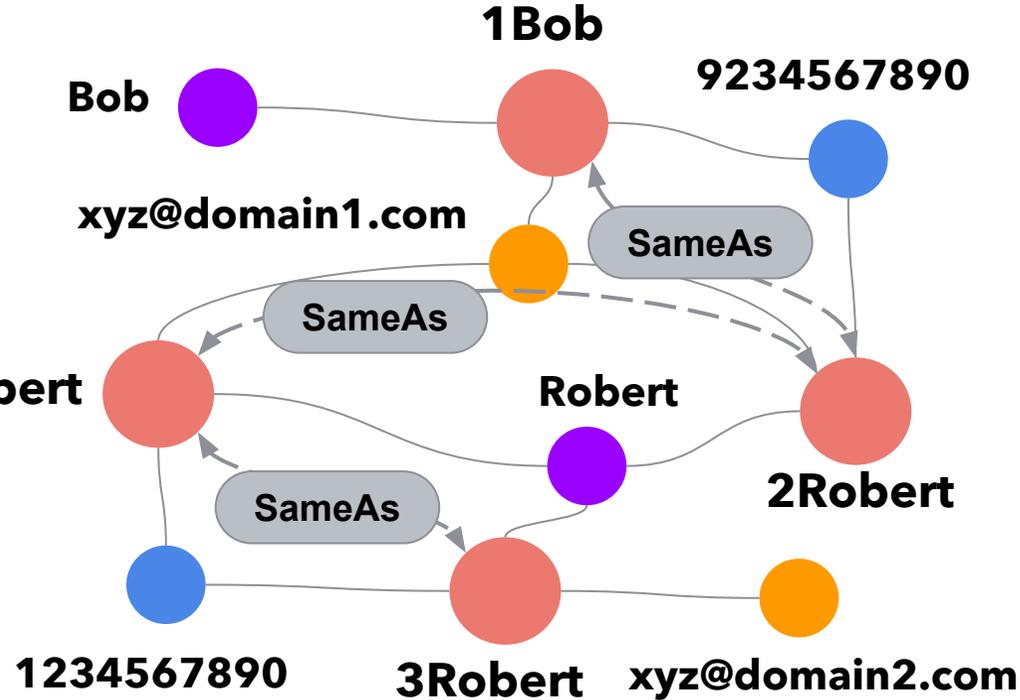
## Initial graph after loading



Graph analysis & inference



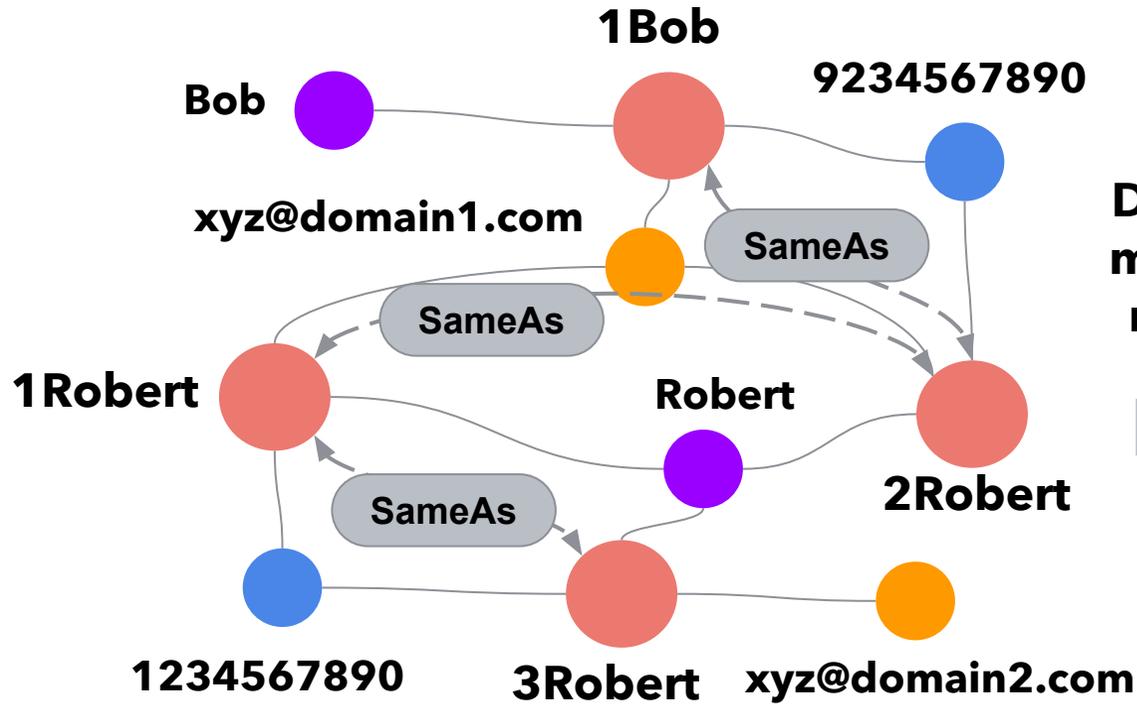
## Enriched graph



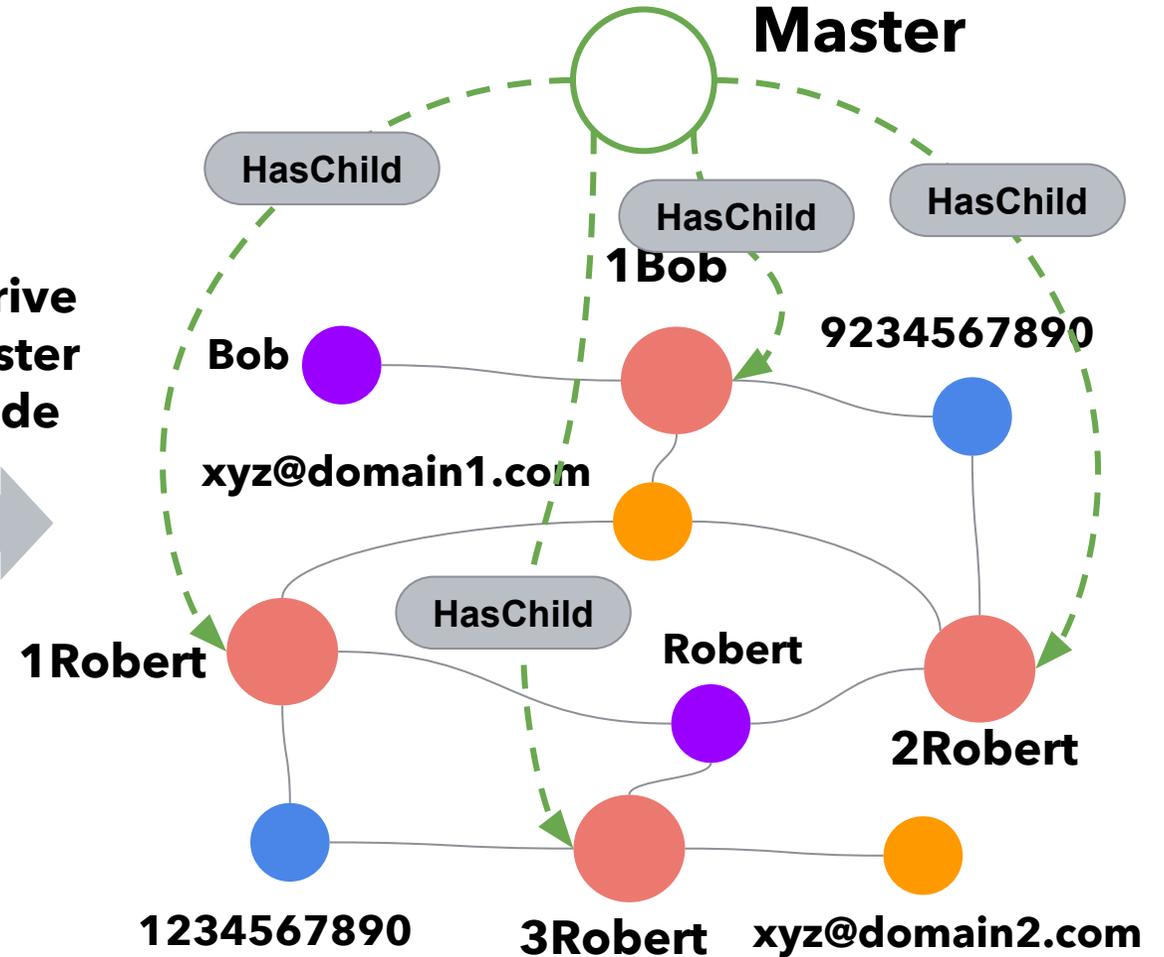
**ENRICH KNOWLEDGE GRAPH**

# Knowledge Graph approach (4 of 5)

## Enriched Graph



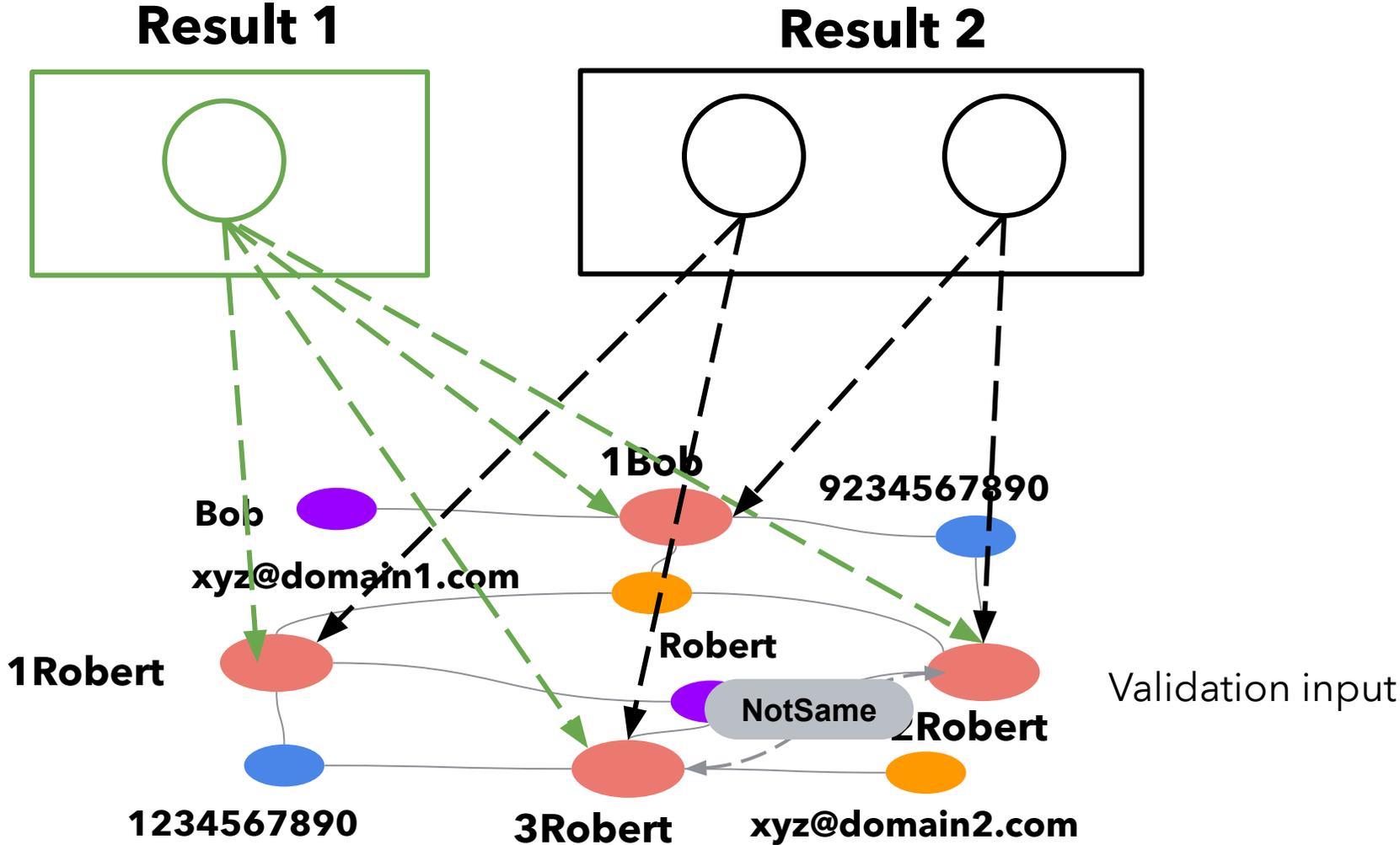
Derive master node



- Semantics and lineage preserved
- Declarative logic (inference, derivation)
- Built-in visual explanation

**DERIVE RESULT: NEW MASTER NODE**

# Knowledge Graph approach (5 of 5)



## FEEDBACK AND VARIATION

# Comparison between two approaches

	<b>PROCEDURAL</b>	<b>KNOWLEDGE GRAPH</b>
<b>Schema</b>	Relational	Graph
<b>Relationship</b>	Relational join	Graph traversal
<b>Logic</b>	Procedural	Declarative
<b>Explainability</b>	Blackbox processing	Built-in visual explanation
<b>Extensibility</b>	Purpose-built	Easy to extend, continuous enrichment
<b>Scalability</b>	Leverage Spark processing	Leverage native parallel graph DB
<b>Tuning</b>	Batch rerun	Rapid iteration
<b>Optimized for</b>	Batch processing (bootstrap)	Near real-time processing (streaming)
<b>Target Developer</b>	Engineers	Domain experts

**Hybrid: Bootstrapping (Procedural) + Real-time updates (KG)**

# Entity resolution result

Initial effort to clean, detect and resolve **180M customer records from 6 sources:**

**19%**  
**Reduction**

**<1%**  
**False positive**

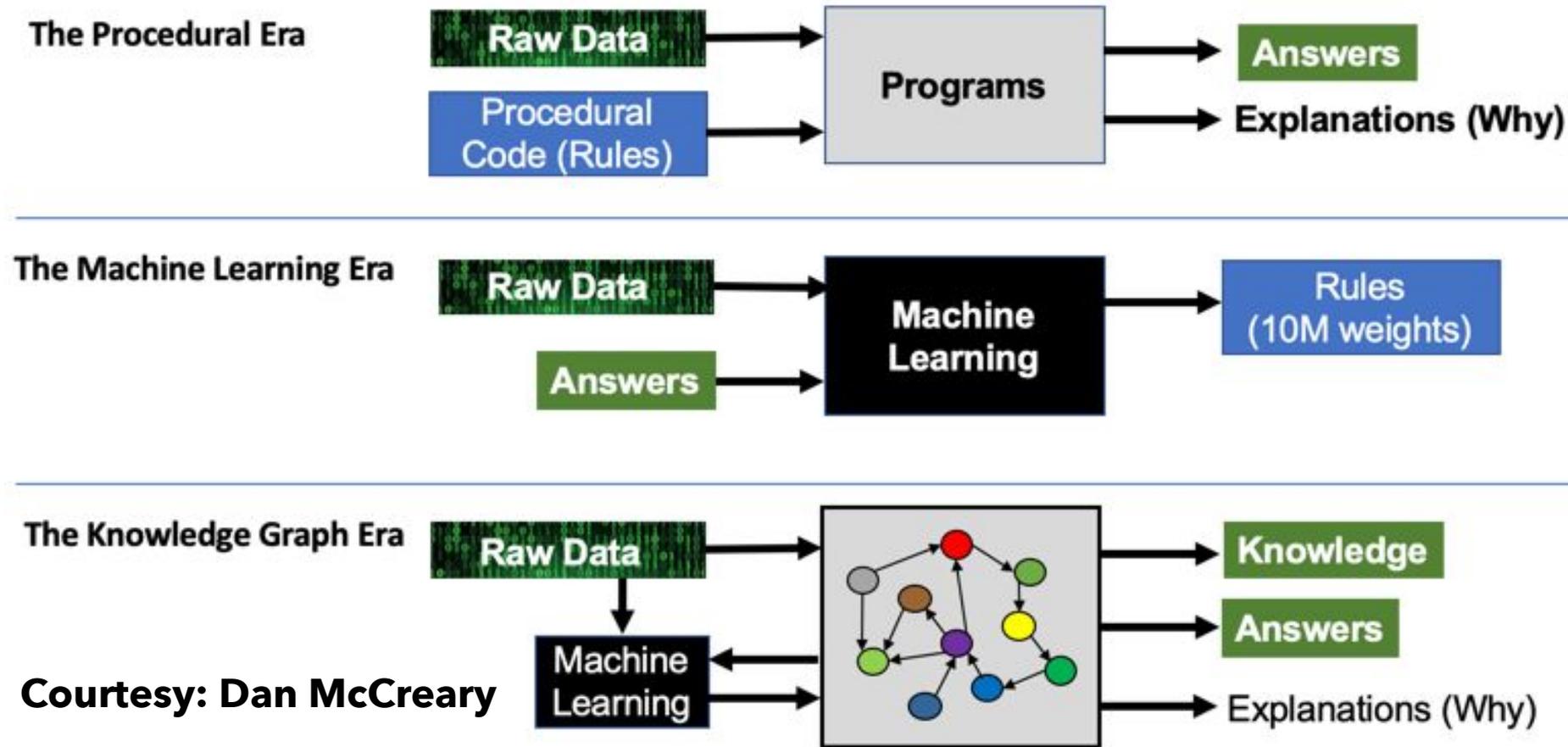
# Summary

# Key takeaways

- Knowledge Graph is a **natural fit** for many use cases.
- Knowledge Graph can be used to **model logic**, beyond data.
- Knowledge Graph can be **automatically created/enriched via AI**.
- Knowledge Graph **makes Intuit products smarter** with tangible customer benefits: **More Money, No Work** and **Complete Confidence**.

**Intuit has been leveraging and innovating on Knowledge Graphs to bring tangible benefits to our customers!**

# Food for thought: KGs are the core of third era of computing



“Knowledge Graphs are at the core of the third era of computing where we use machine learning to continuously enrich our shared knowledge”

<https://medium.com/@dmccreary/knowledge-graphs-the-third-era-of-computing-a8106f343450>

# Thank you!

**Interested in applying Knowledge Graphs to challenging  
personal/business financial problems?**

**Contact [jay\\_yu@intuit.com](mailto:jay_yu@intuit.com)**