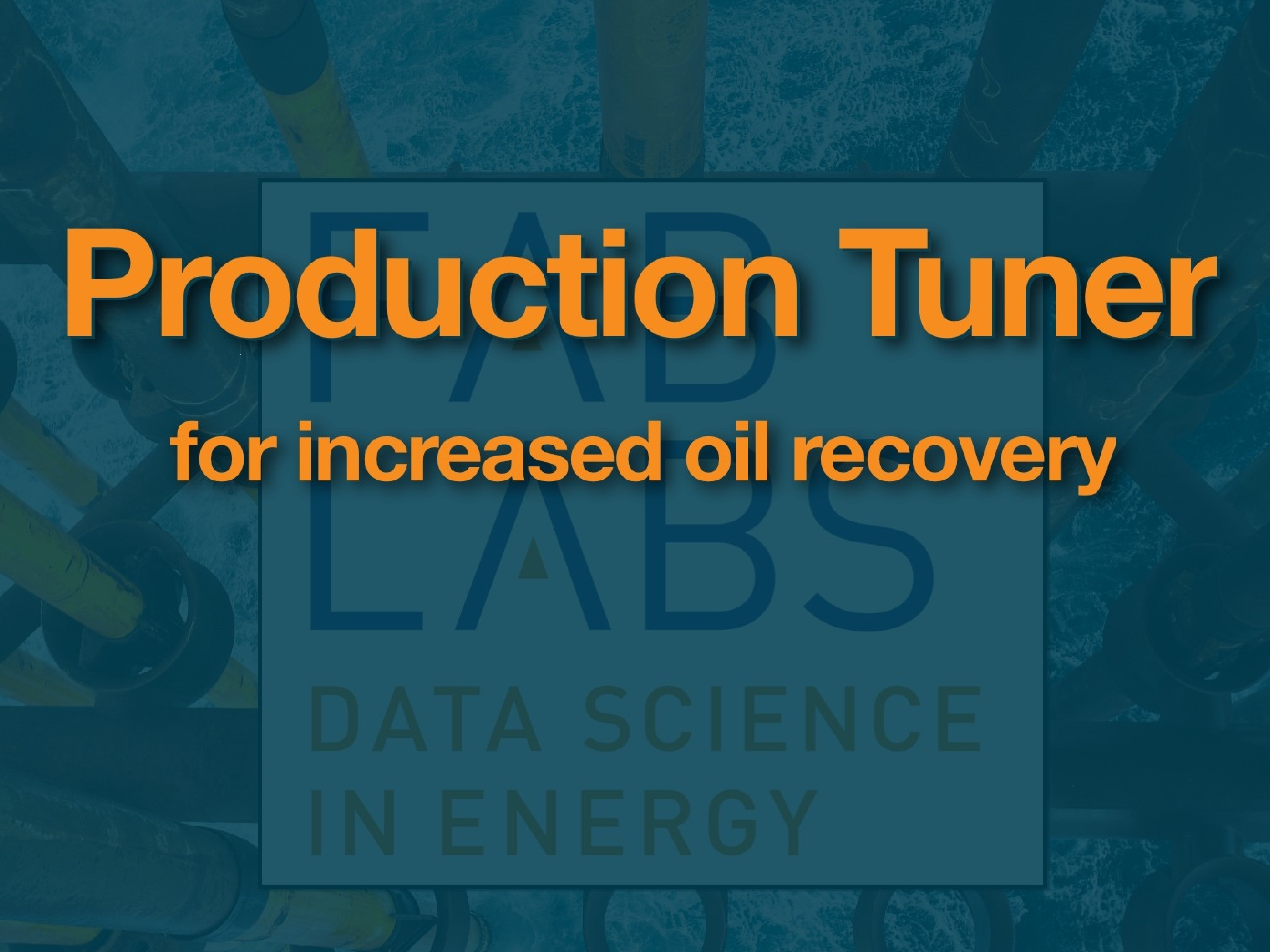
Rich Modelling

Capturing the context of data as it exists in the wild

FAB LABS

DATA SCIENCE
IN ENERGY



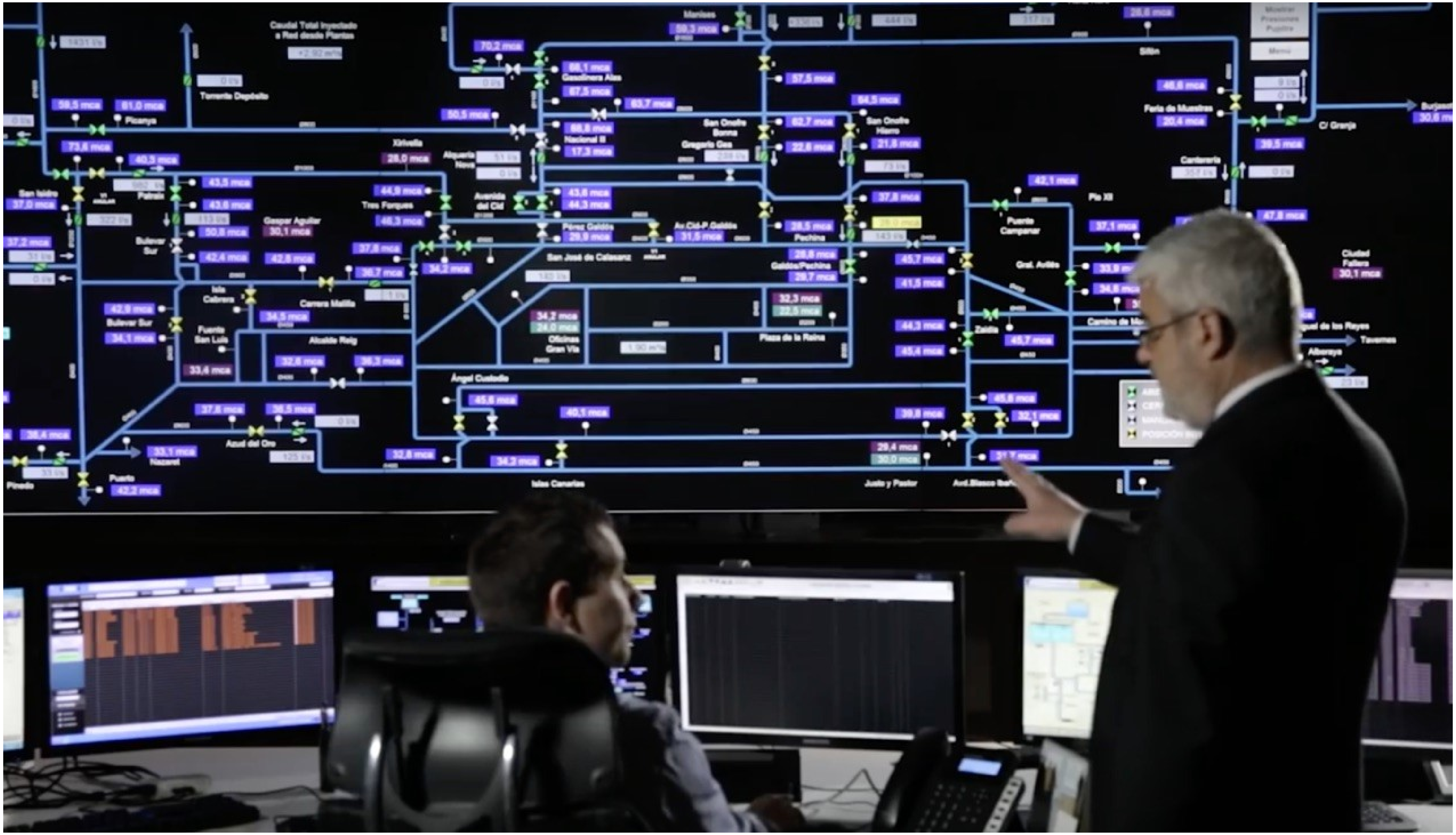
Production Tuner

for increased oil recovery

FAB LABS

DATA SCIENCE
IN ENERGY

State of the art?

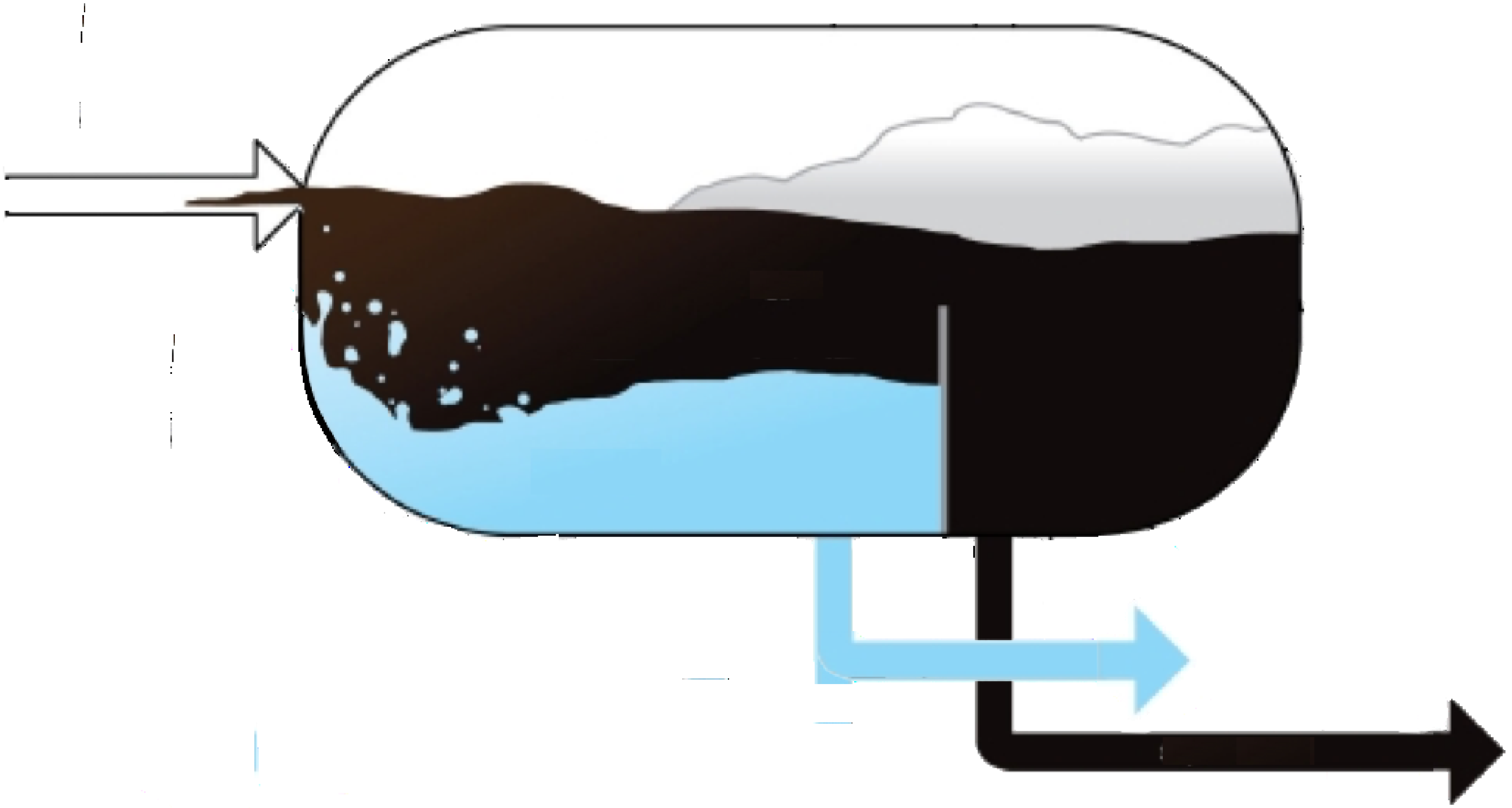


Model => Analysis =>Decision
Short windows of opportunity

FAB LABS
DATA SCIENCE IN ENERGY

Topside Constraints

Multiple wells & separators all with different characteristics = tradeoffs.



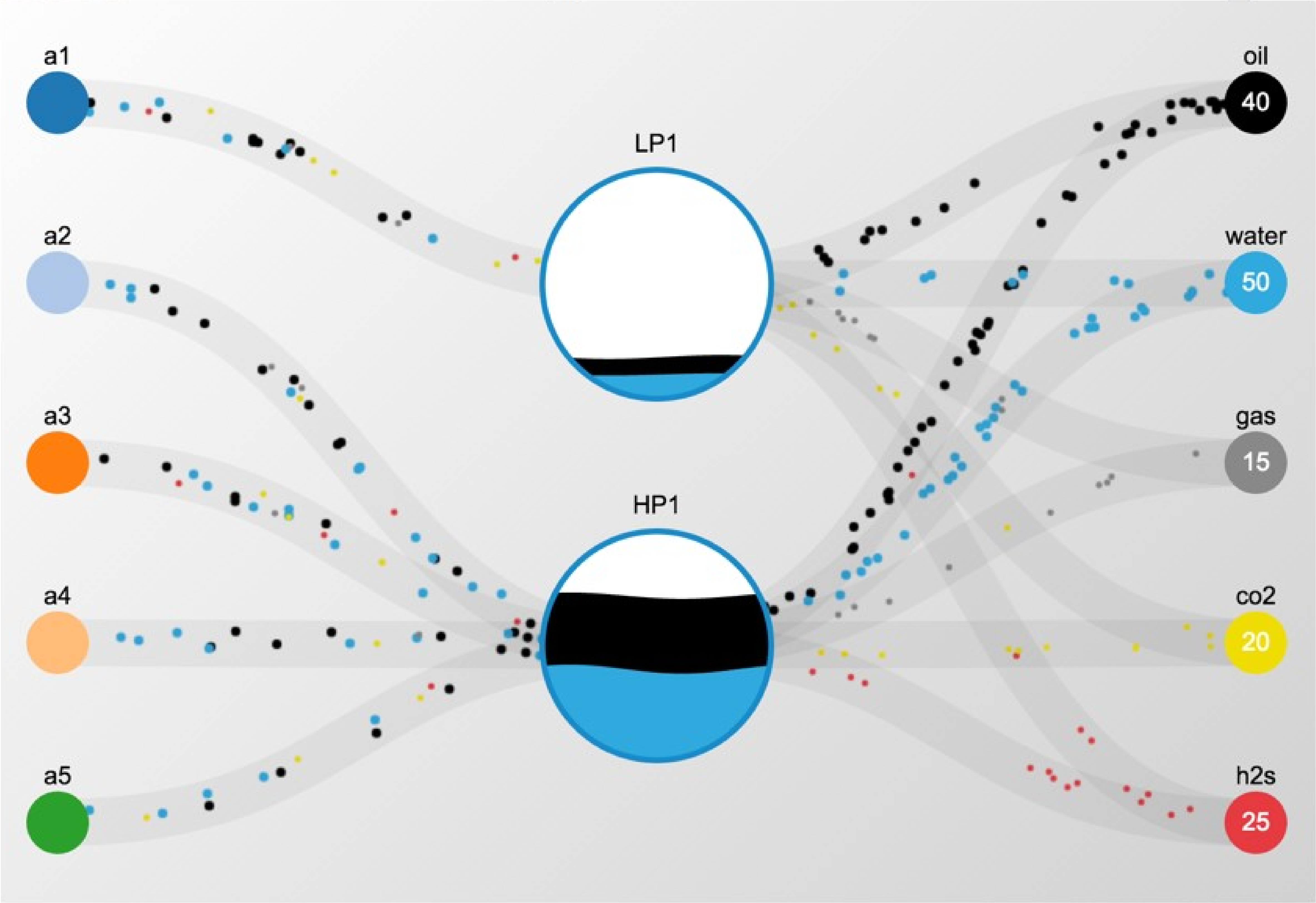
FAB LABS
DATA SCIENCE IN ENERGY

Production Tuner

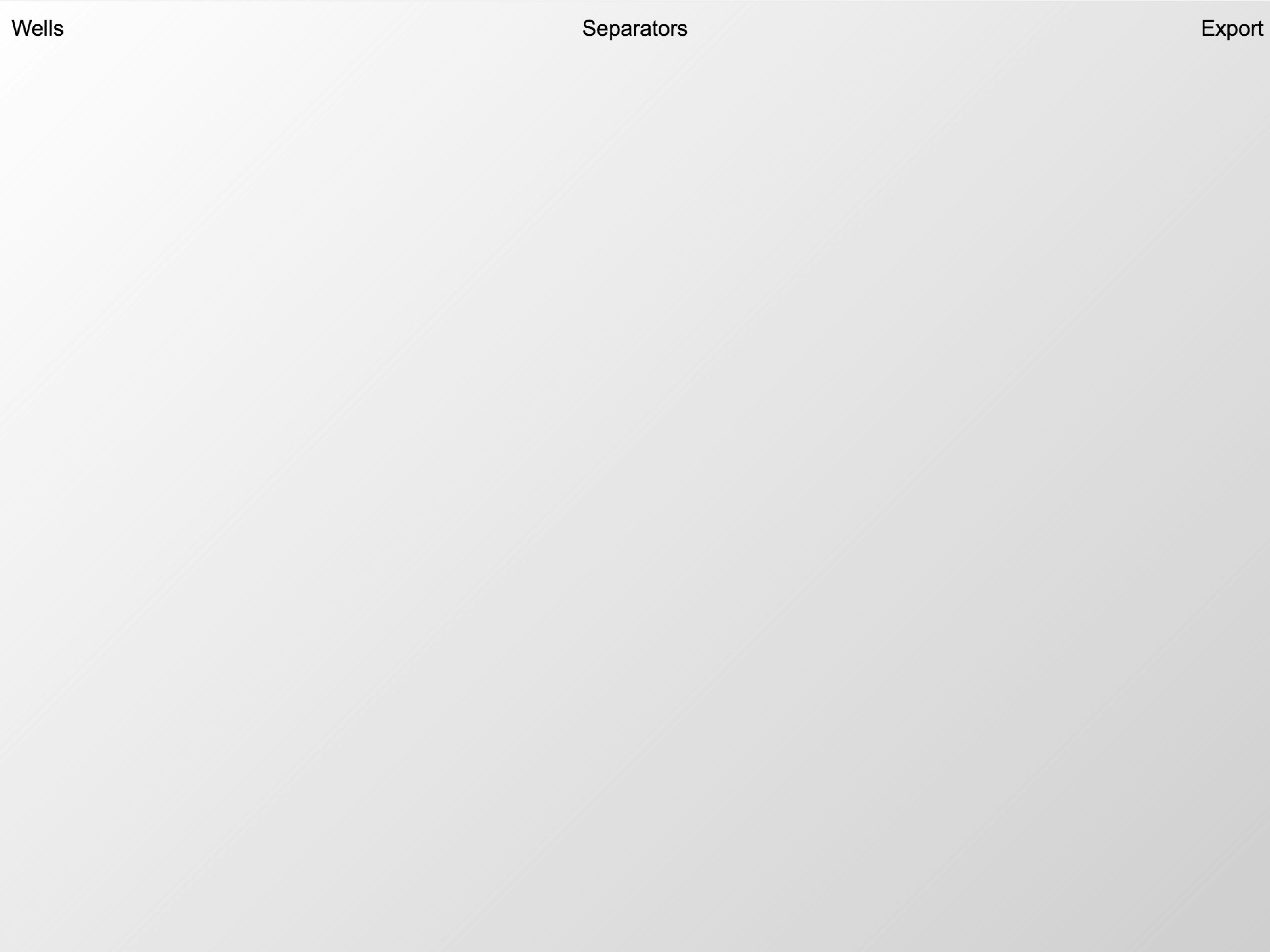
- near-realtime monitoring
- runs millions of scenarios per day
- not more alarms, best opportunities
- cuts execution lead-time to zero*
- sees hidden edge cases

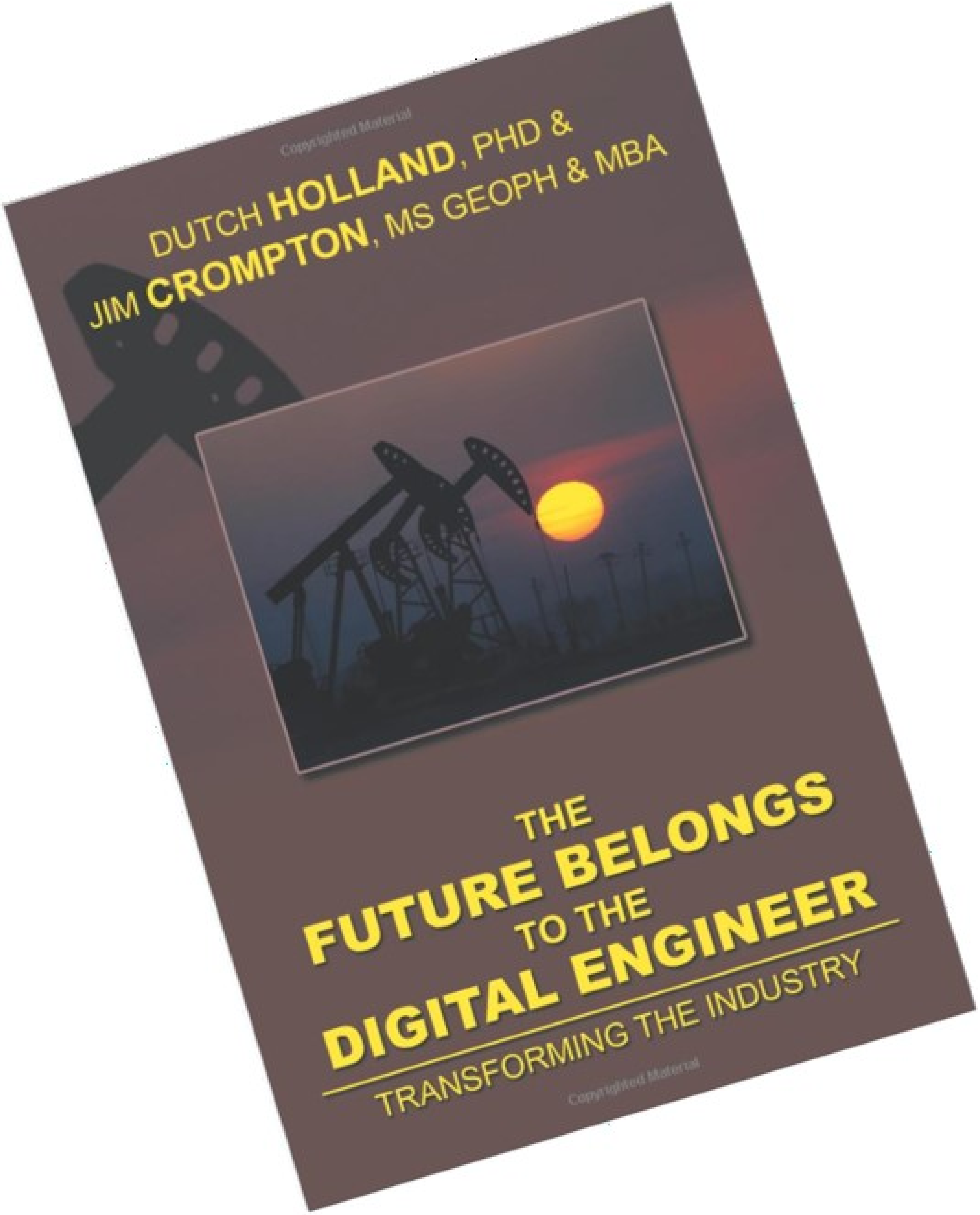
FAB LABS
DATA SCIENCE IN ENERGY

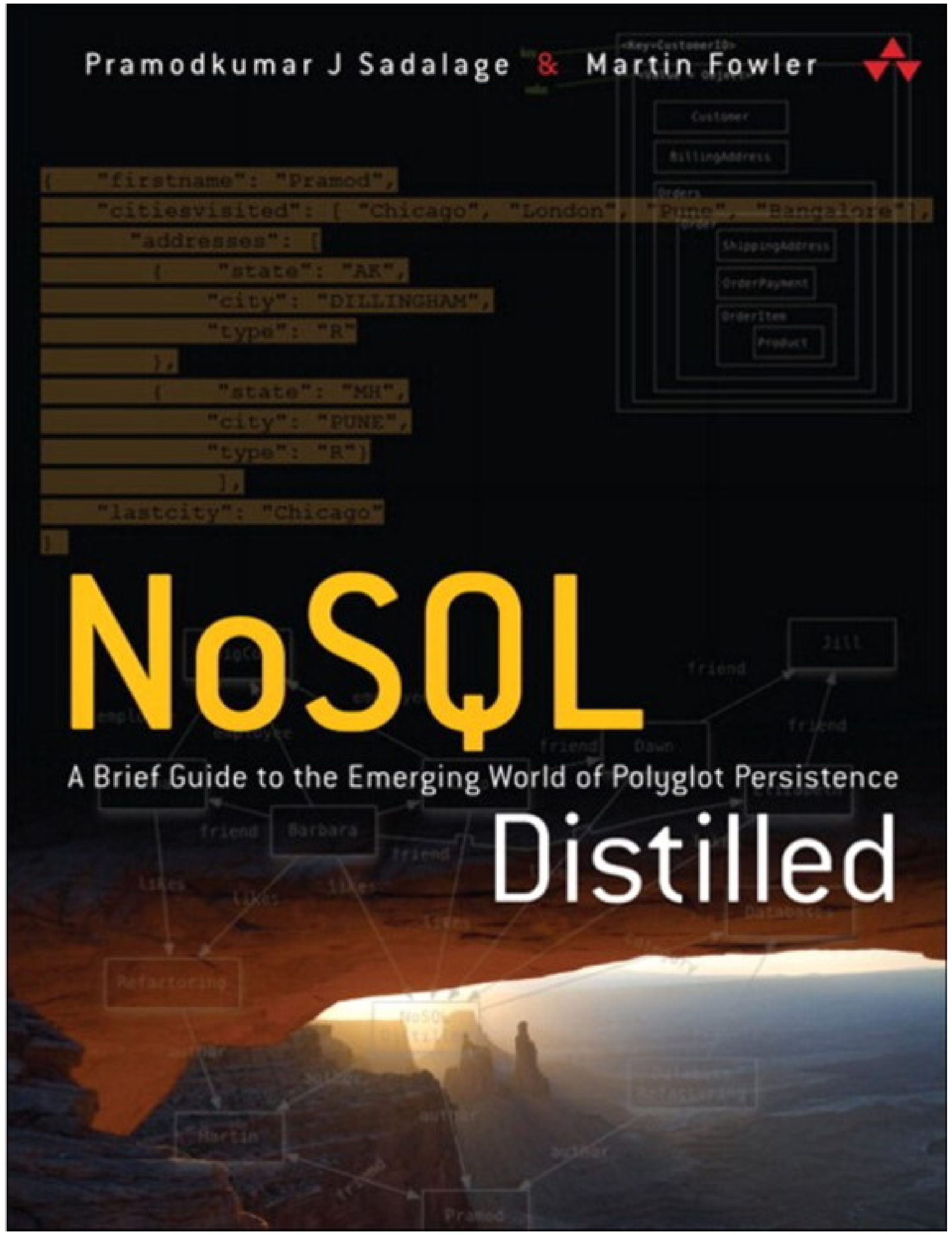
wells separators export

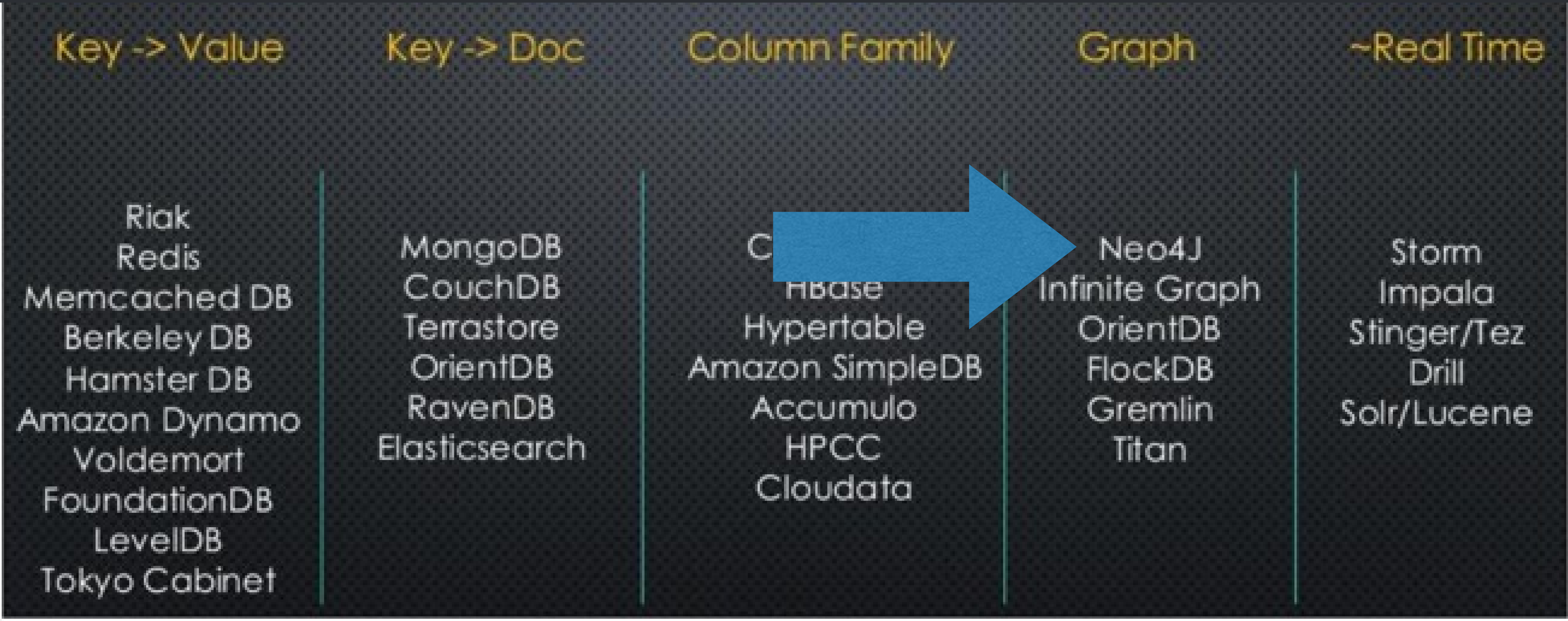


exhaust


Wells
Separators
Export


Copyrighted Material
DUTCH HOLLAND, PHD &
JIM CROMPTON, MS GEOPH & MBA
THE FUTURE BELONGS TO THE DIGITAL ENGINEER
TRANSFORMING THE INDUSTRY
Copyrighted Material

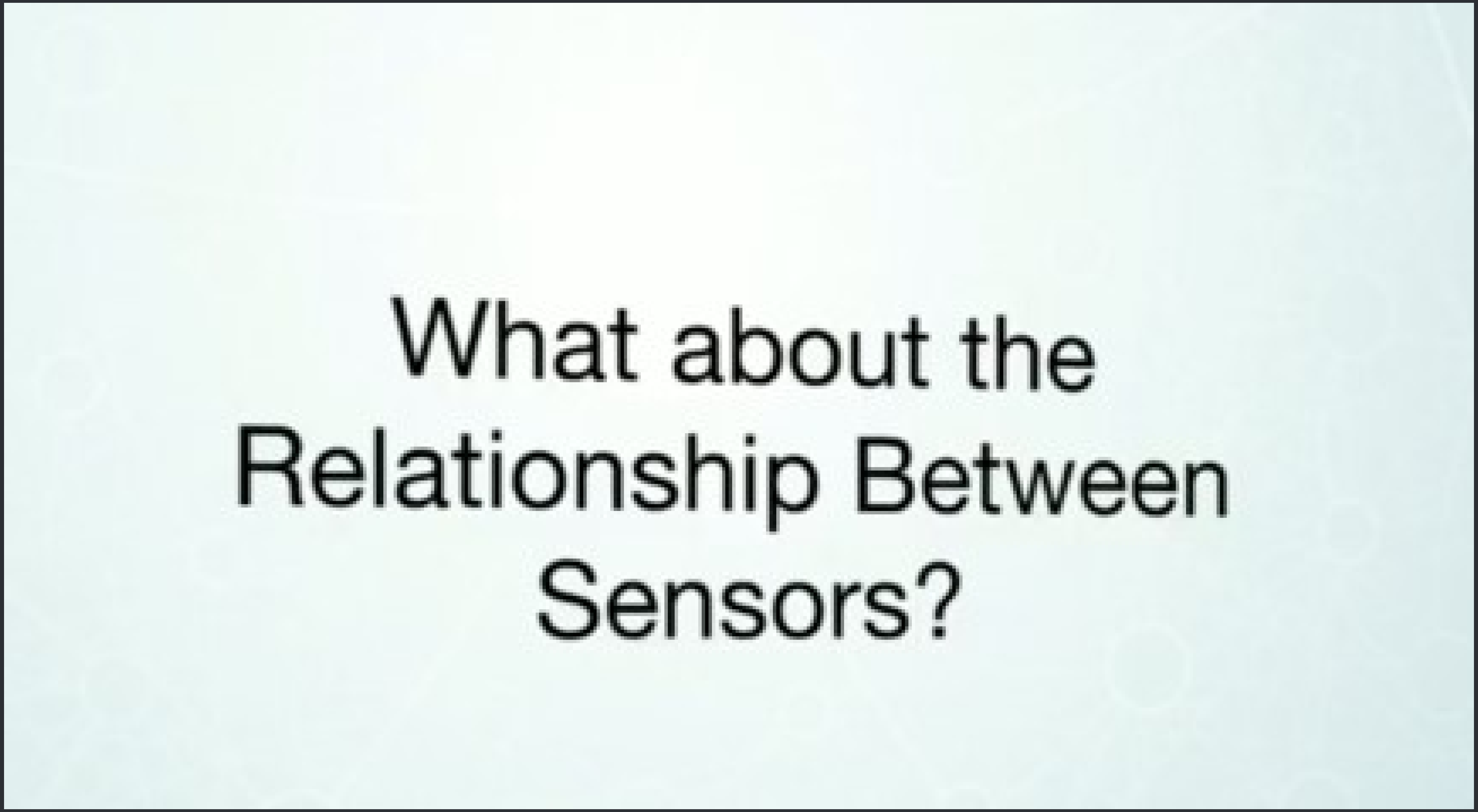

Pramodkumar J Sadalage & Martin Fowler
{ "firstname": "Pramod",
"citiesvisited": ["Chicago", "London", "Pune", "Bangalore"],
"addresses": [
{ "state": "AK",
"city": "DILLINGHAM",
"type": "R"
},
{ "state": "MH",
"city": "PUNE",
"type": "R" }
],
"lastcity": "Chicago"
}
NoSQL
A Brief Guide to the Emerging World of Polyglot Persistence
Distilled



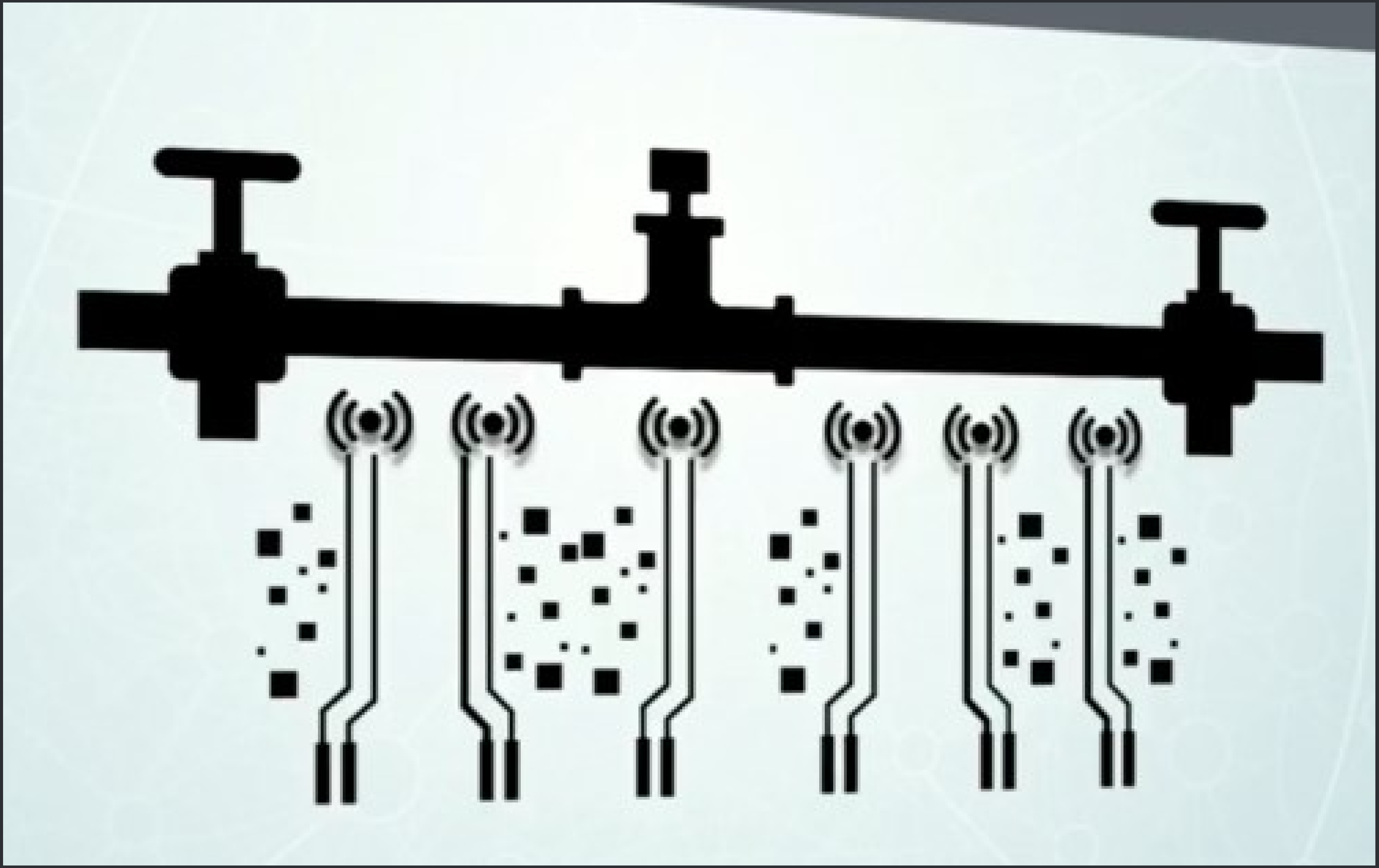
Key -> Value	Key -> Doc	Column Family	Graph	~Real Time
Riak	MongoDB	C	Neo4J	Storm
Redis	CouchDB	HBase	Infinite Graph	Impala
Memcached DB	Terrastore	Hypertable	OrientDB	Stinger/Tez
Berkeley DB	OrientDB	Amazon SimpleDB	FlockDB	Drill
Hamster DB	RavenDB	Accumulo	Gremlin	Solr/Lucene
Amazon Dynamo	Elasticsearch	HPCC	Titan	
Voldemort		Cloudata		
FoundationDB				
LevelDB				
Tokyo Cabinet				

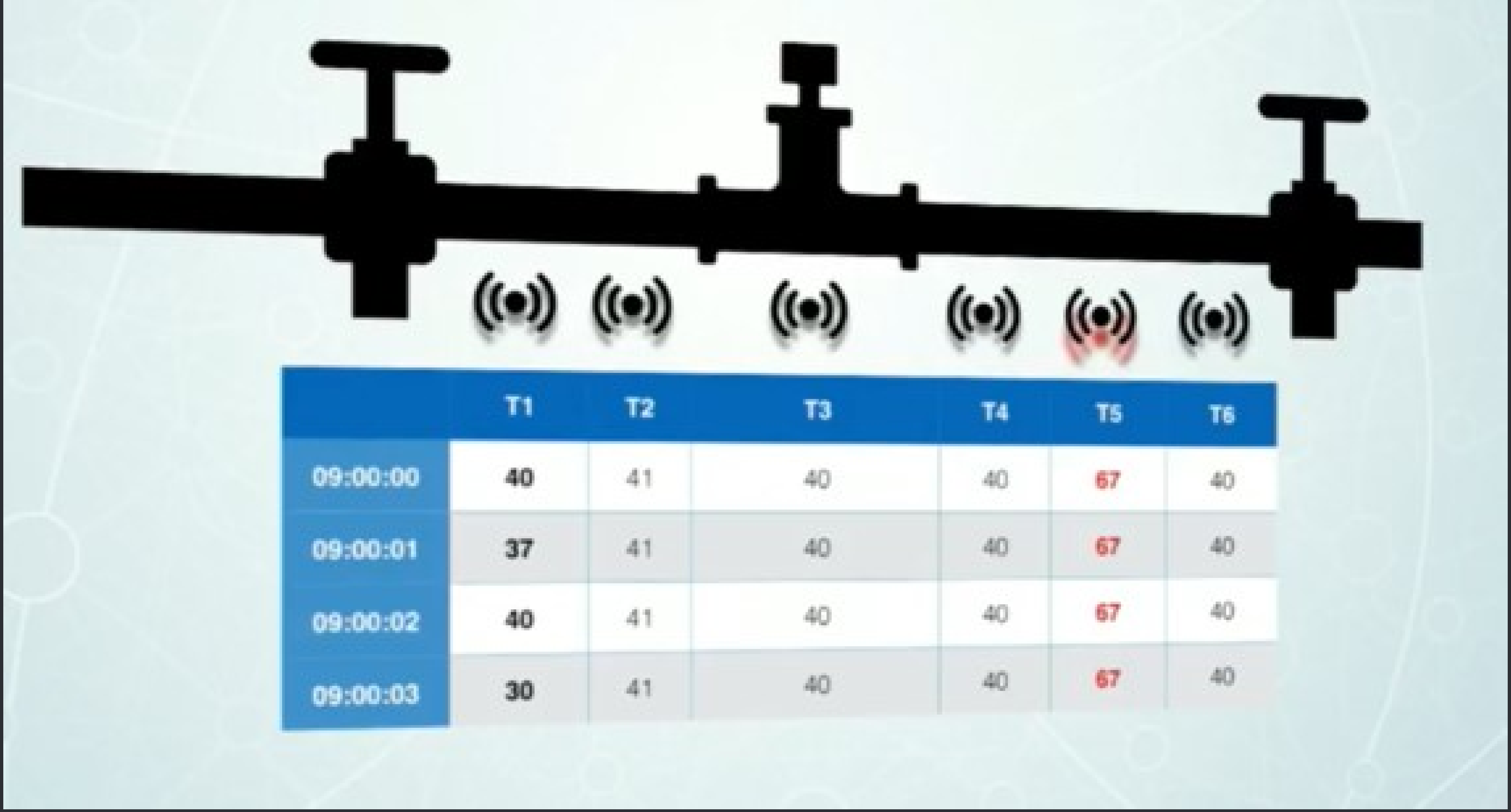
...there's an 87% chance
Linus Torvalds hates your code.

"Bad programmers **worry about the code**. Good programmers **worry about data structures** and their relationships."

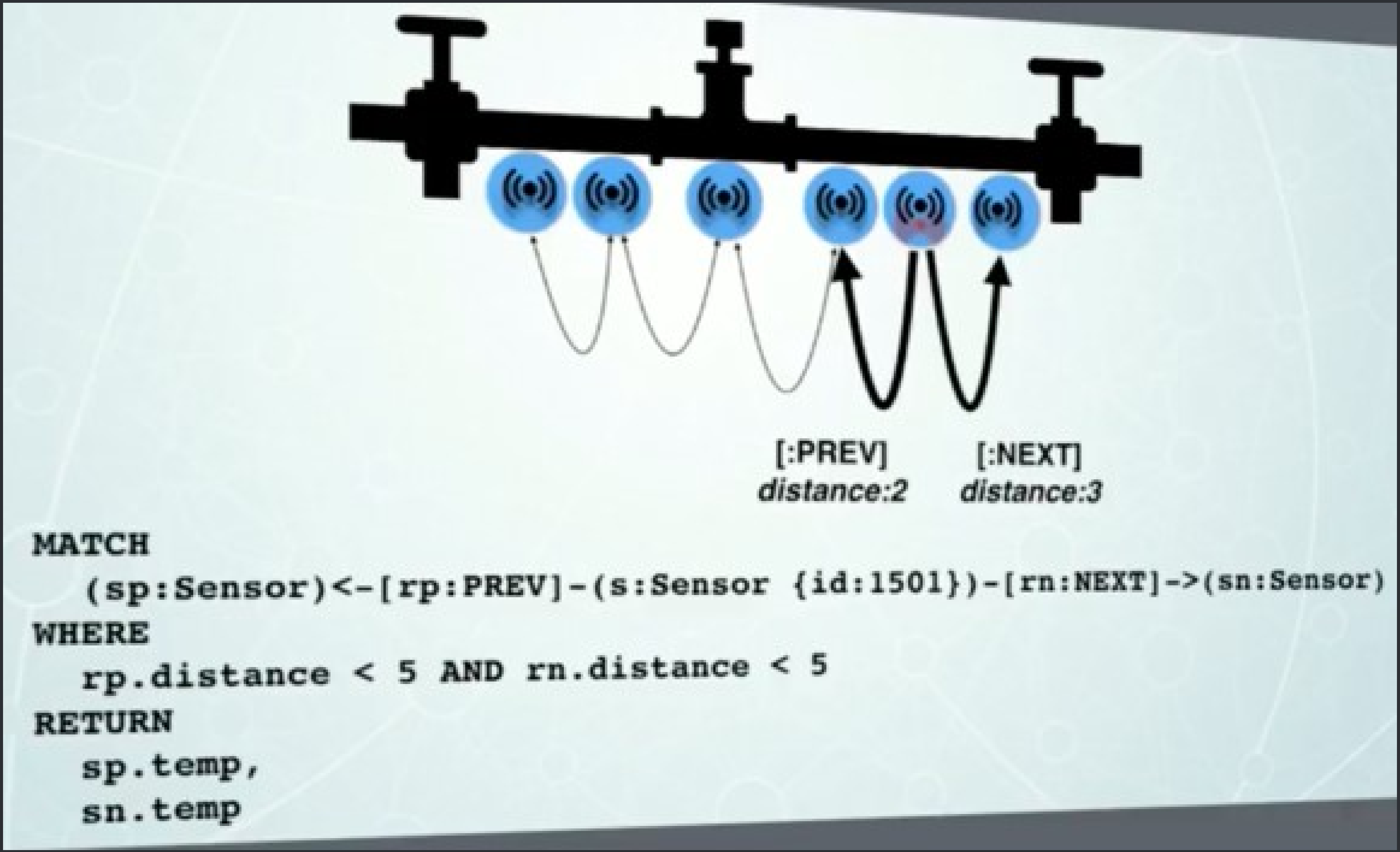


What about the Relationship Between Sensors?

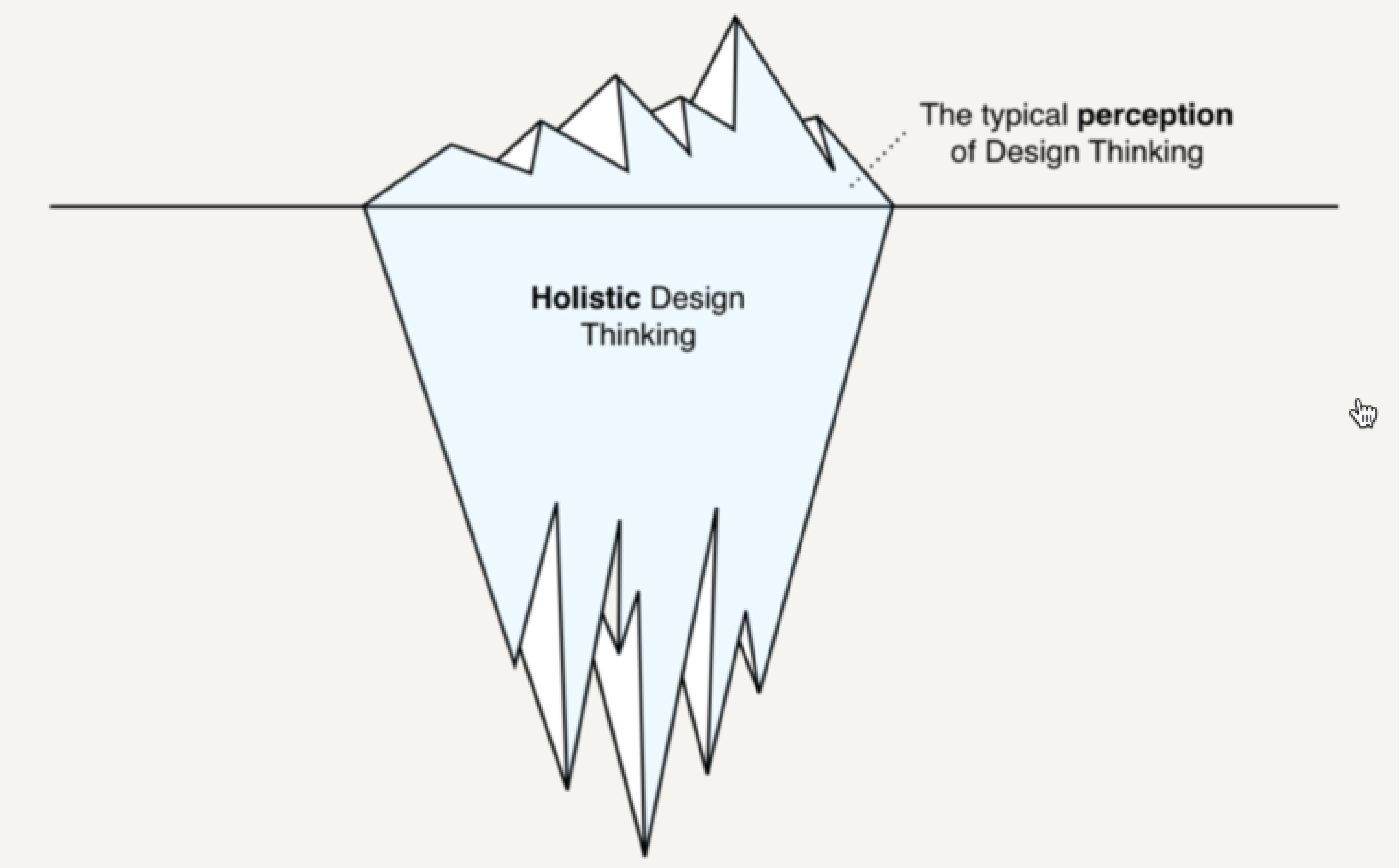




	T1	T2	T3	T4	T5	T6
09:00:00	40	41	40	40	67	40
09:00:01	37	41	40	40	67	40
09:00:02	40	41	40	40	67	40
09:00:03	30	41	40	40	67	40

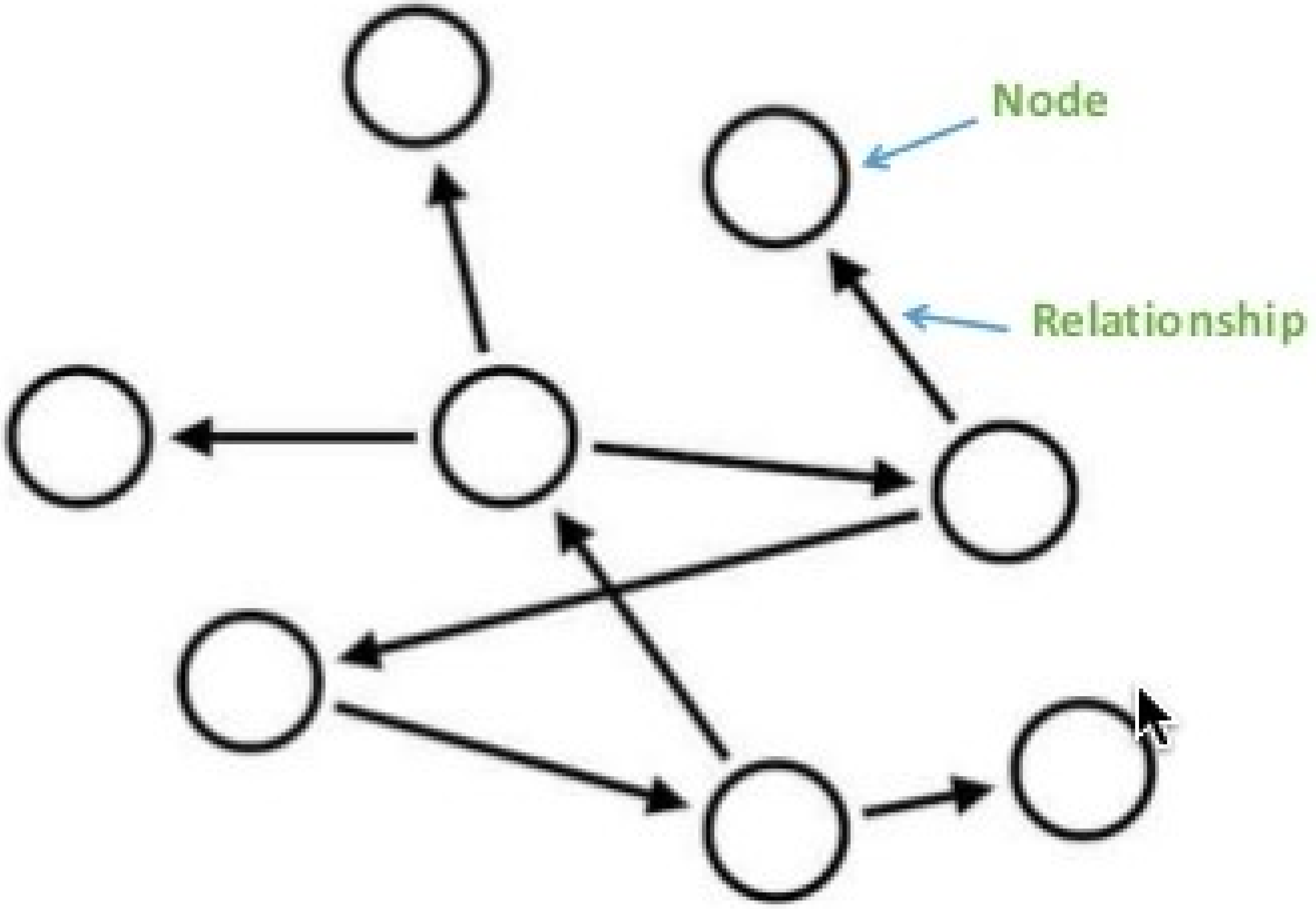


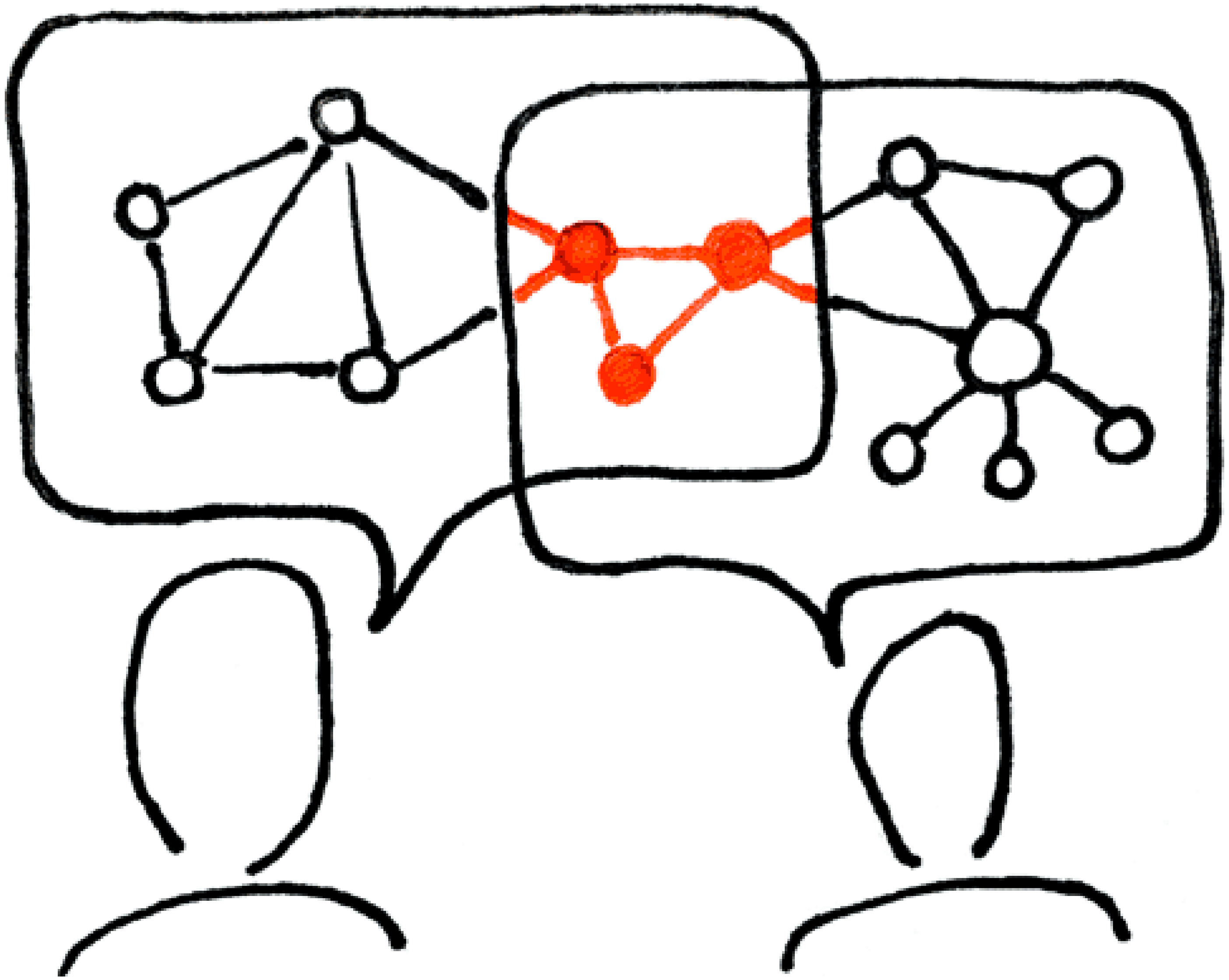
```
MATCH
  (sp:Sensor)<-[rp:PREV]-(s:Sensor {id:1501})-[rn:NEXT]->(sn:Sensor)
WHERE
  rp.distance < 5 AND rn.distance < 5
RETURN
  sp.temp,
  sn.temp
```


The typical **perception** of Design Thinking
Holistic Design Thinking

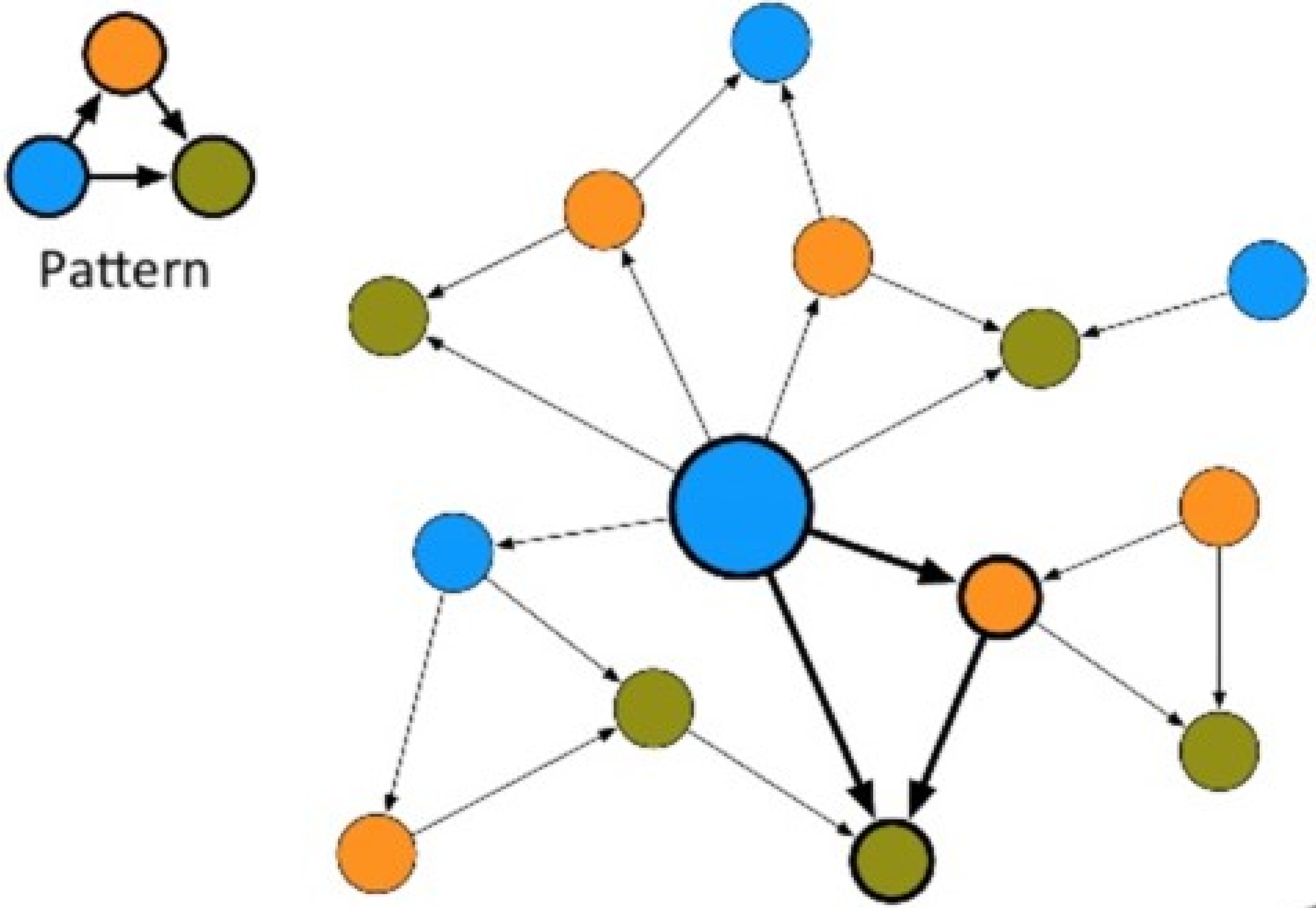
3DCuts.com

Engineer’s mental model

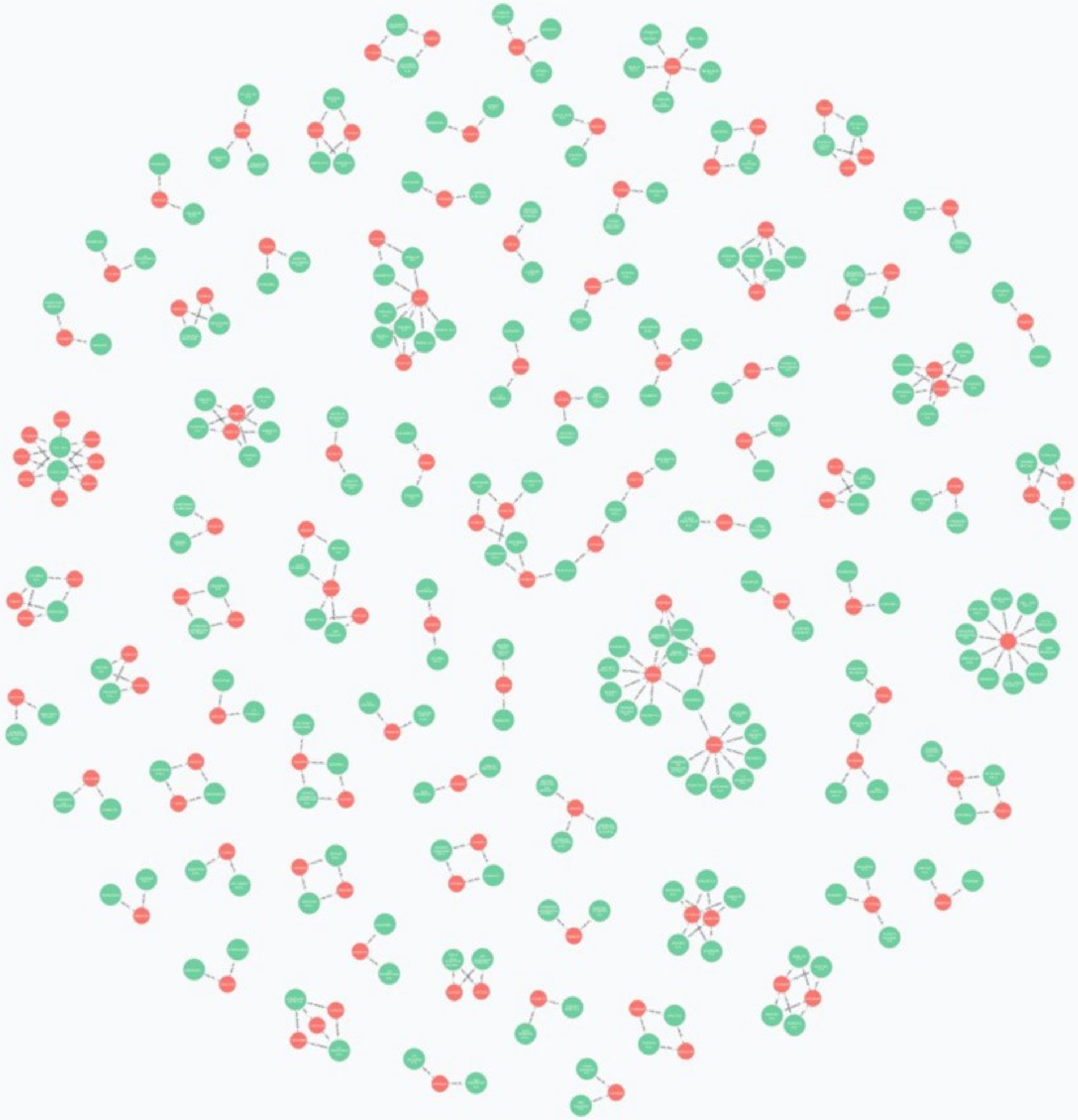

Node
Relationship

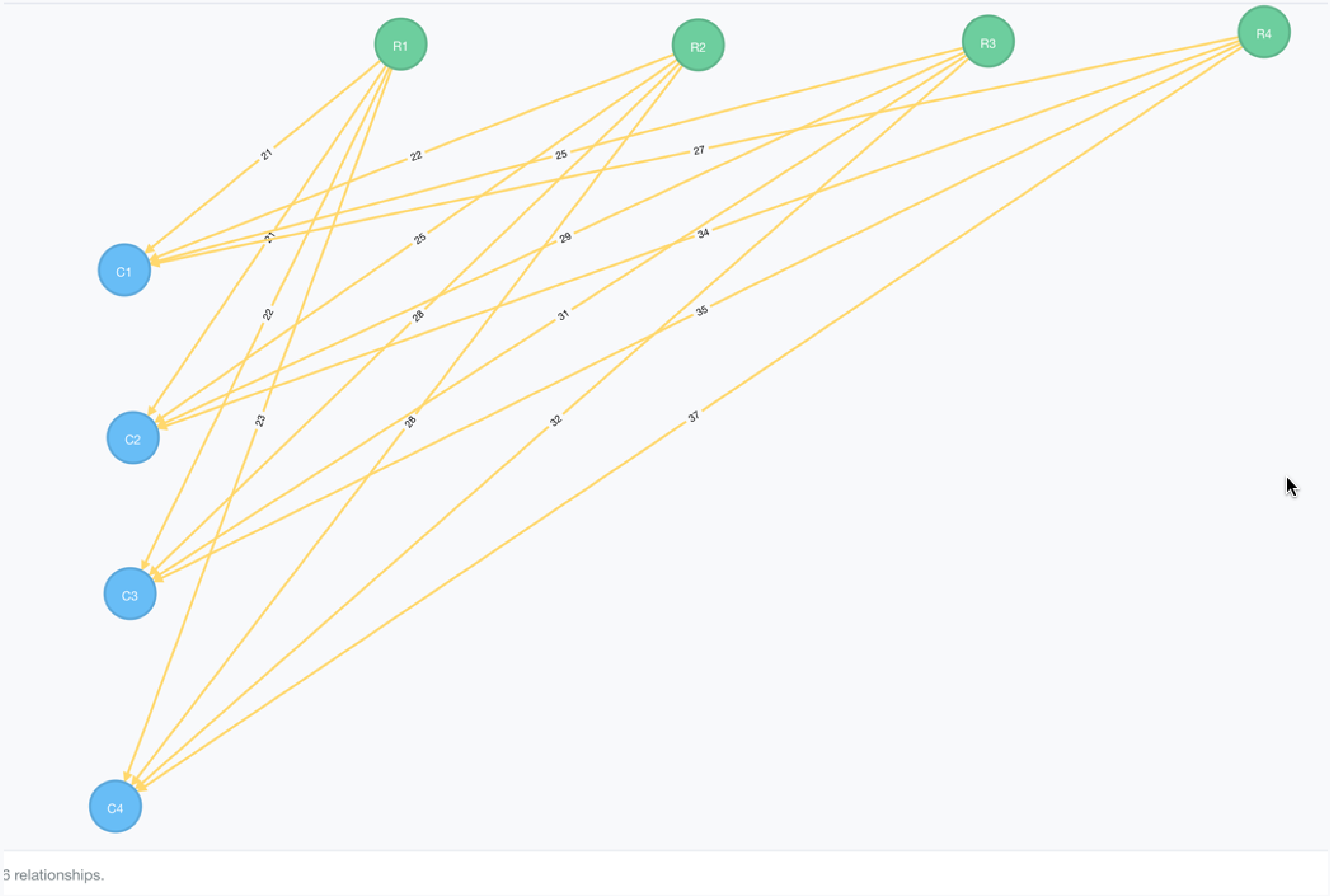


Match



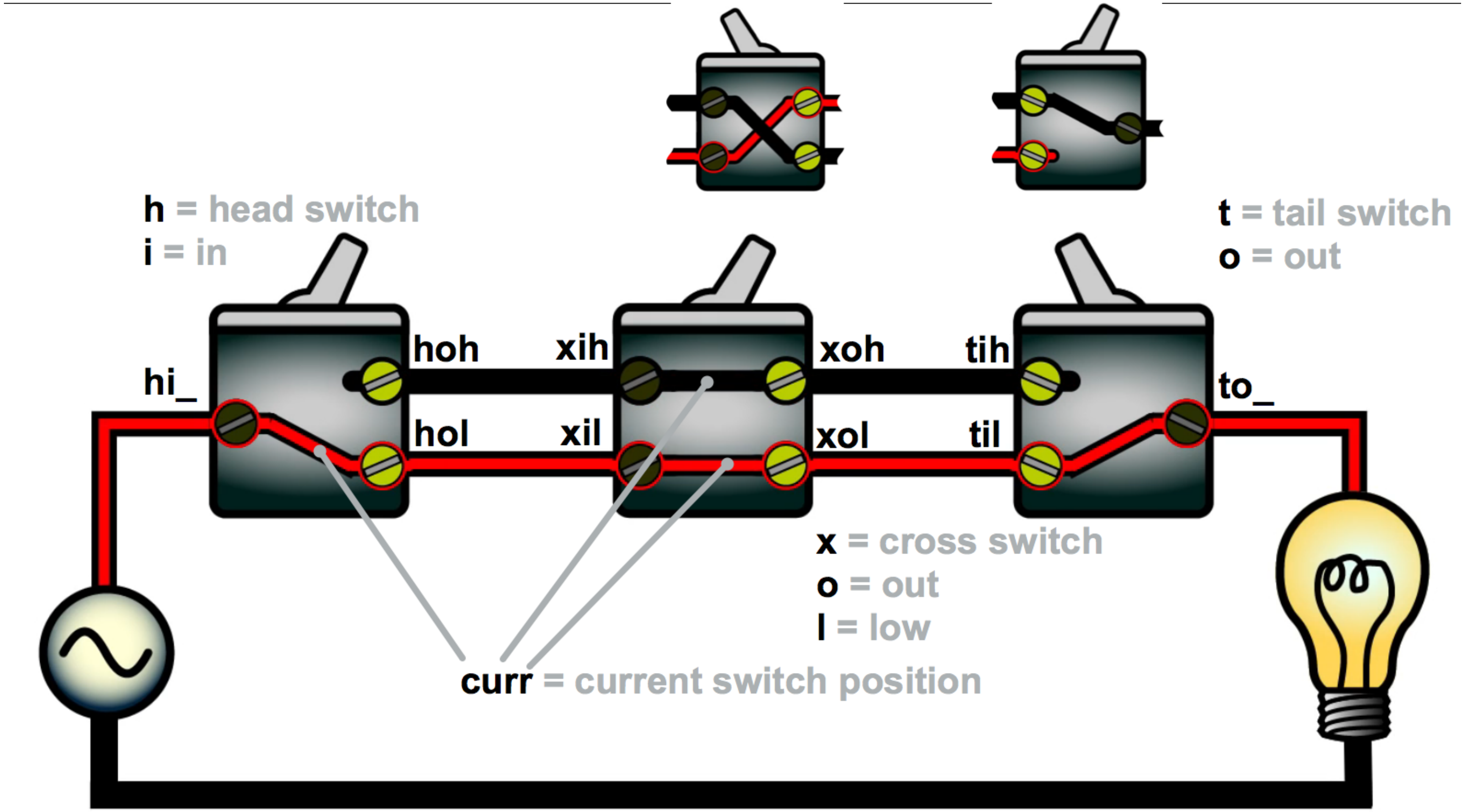
#neo4j





6 relationships.

A true story.

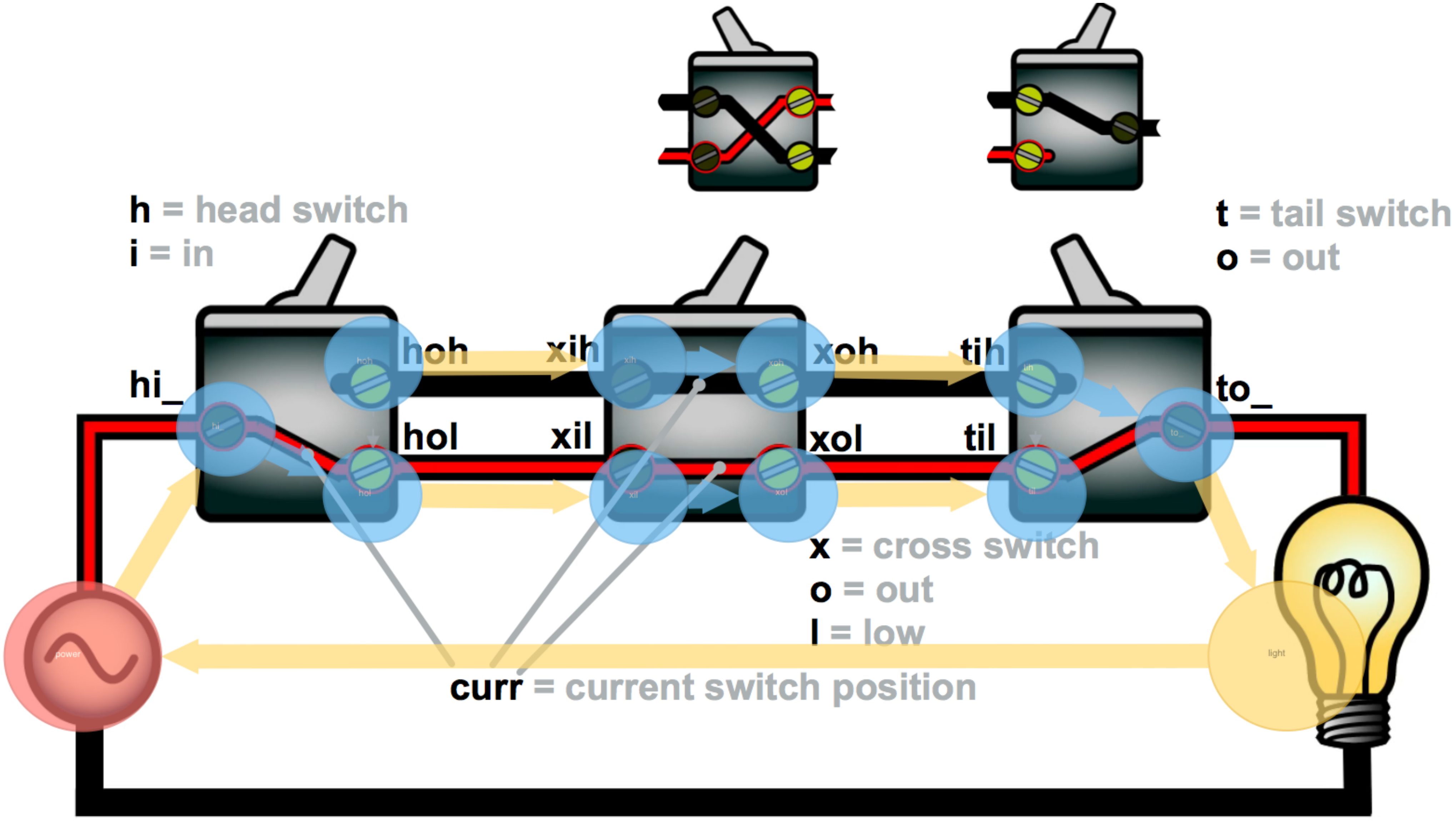


h = head switch
i = in

t = tail switch
o = out

x = cross switch
o = out
l = low

curr = current switch position

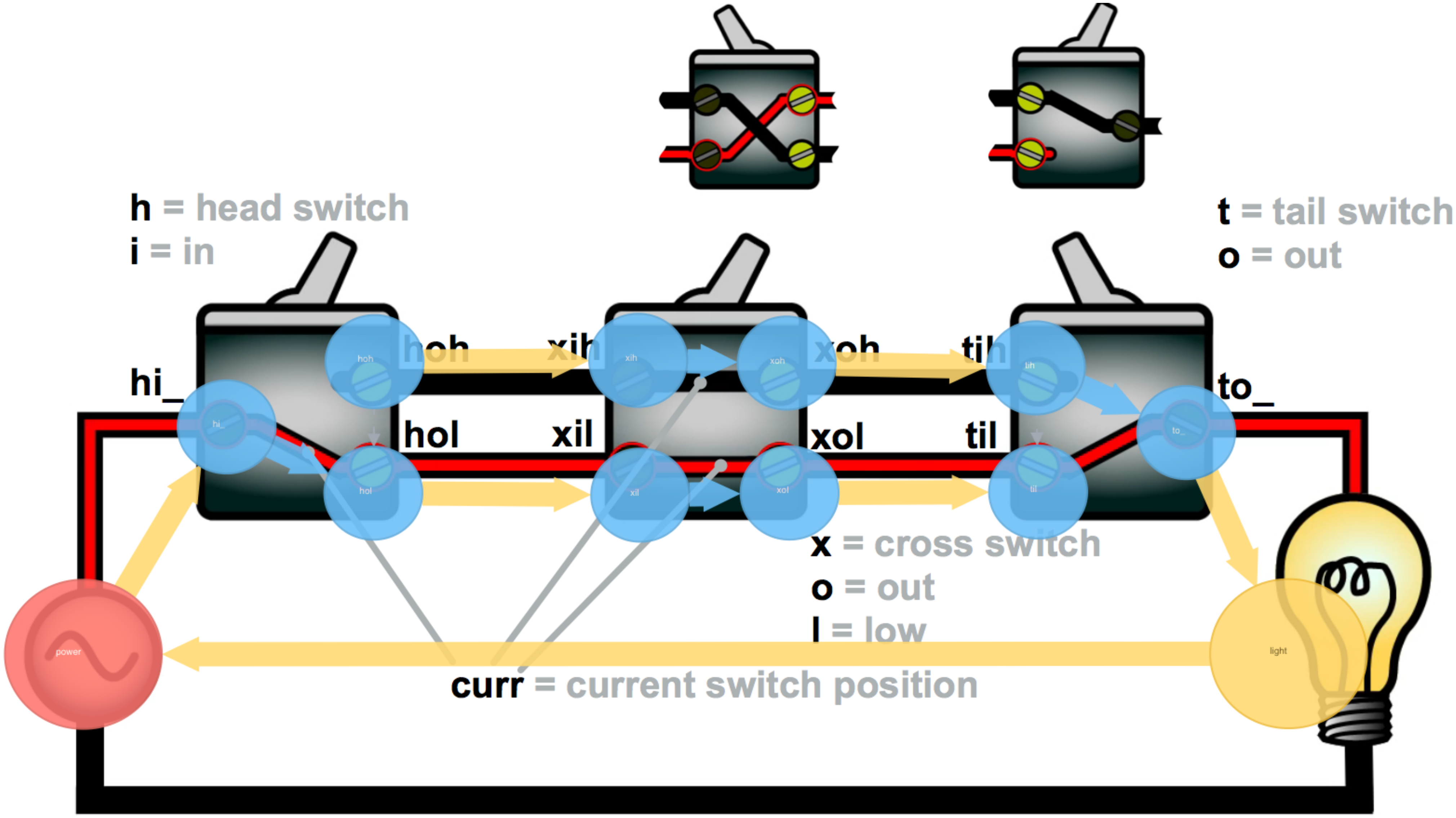


h = head switch
i = in

t = tail switch
o = out

x = cross switch
o = out
l = low

curr = current switch position

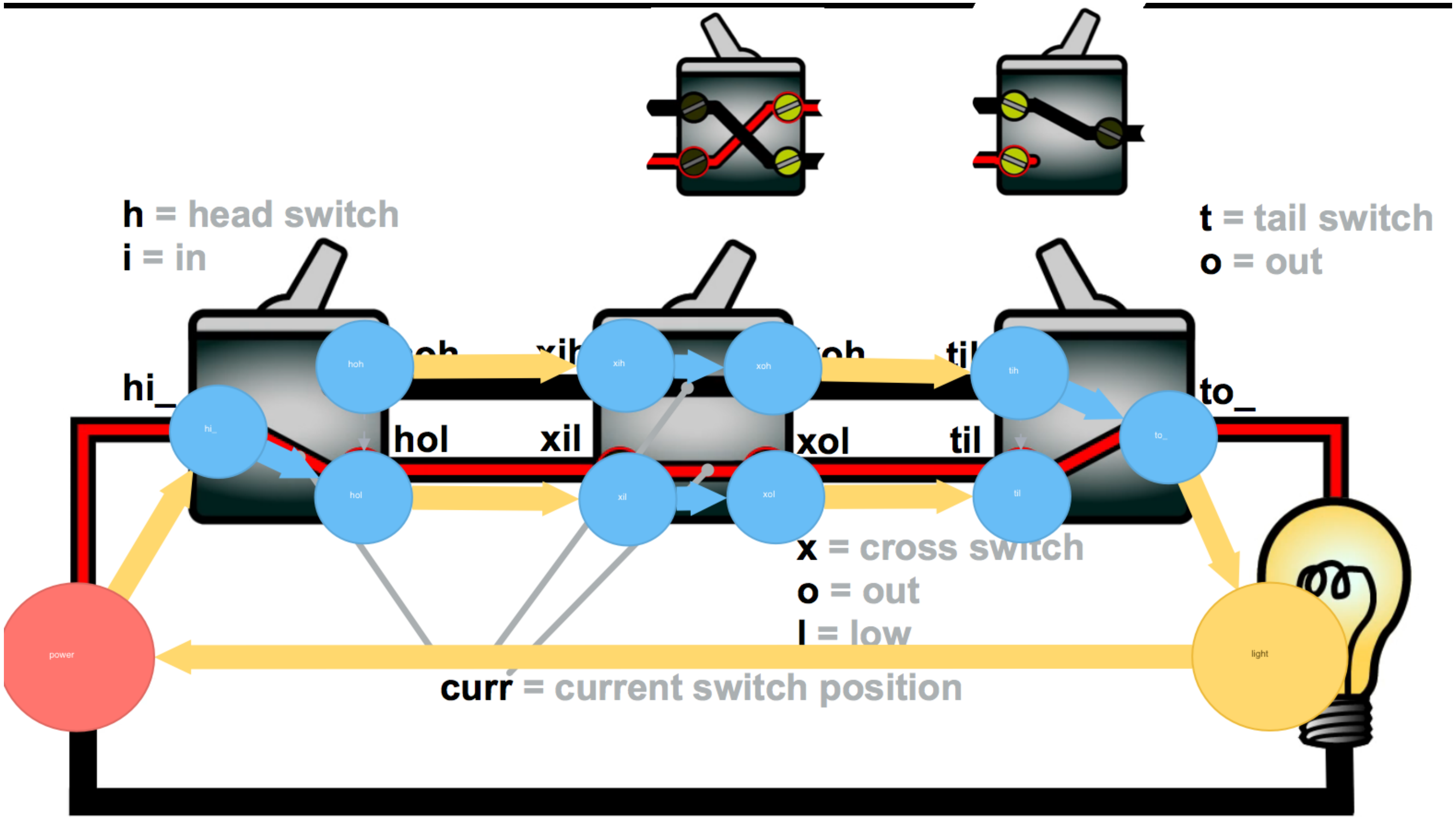


h = head switch
i = in

t = tail switch
o = out

x = cross switch
o = out
l = low

curr = current switch position

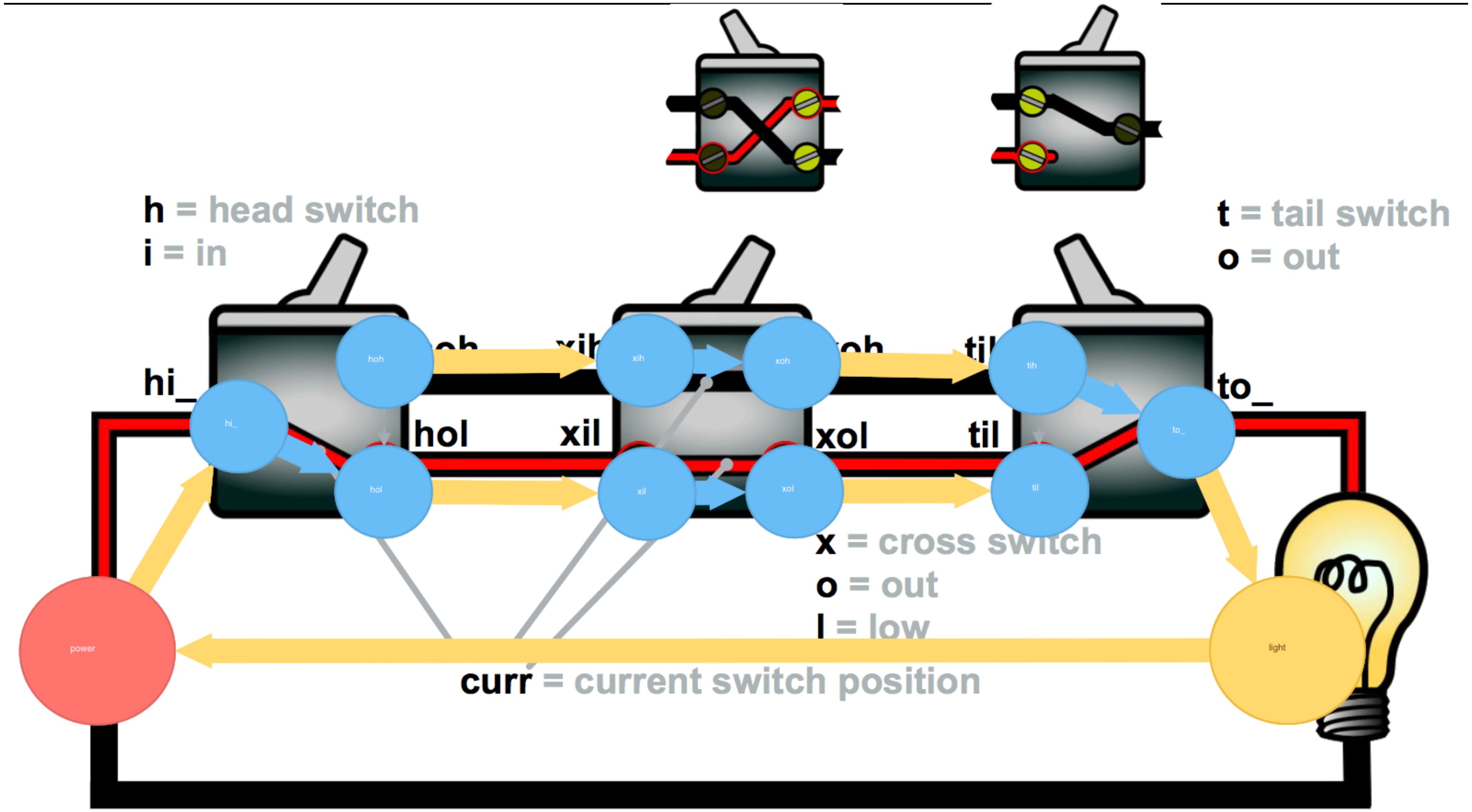


h = head switch
i = in

t = tail switch
o = out

x = cross switch
o = out
l = low

curr = current switch position

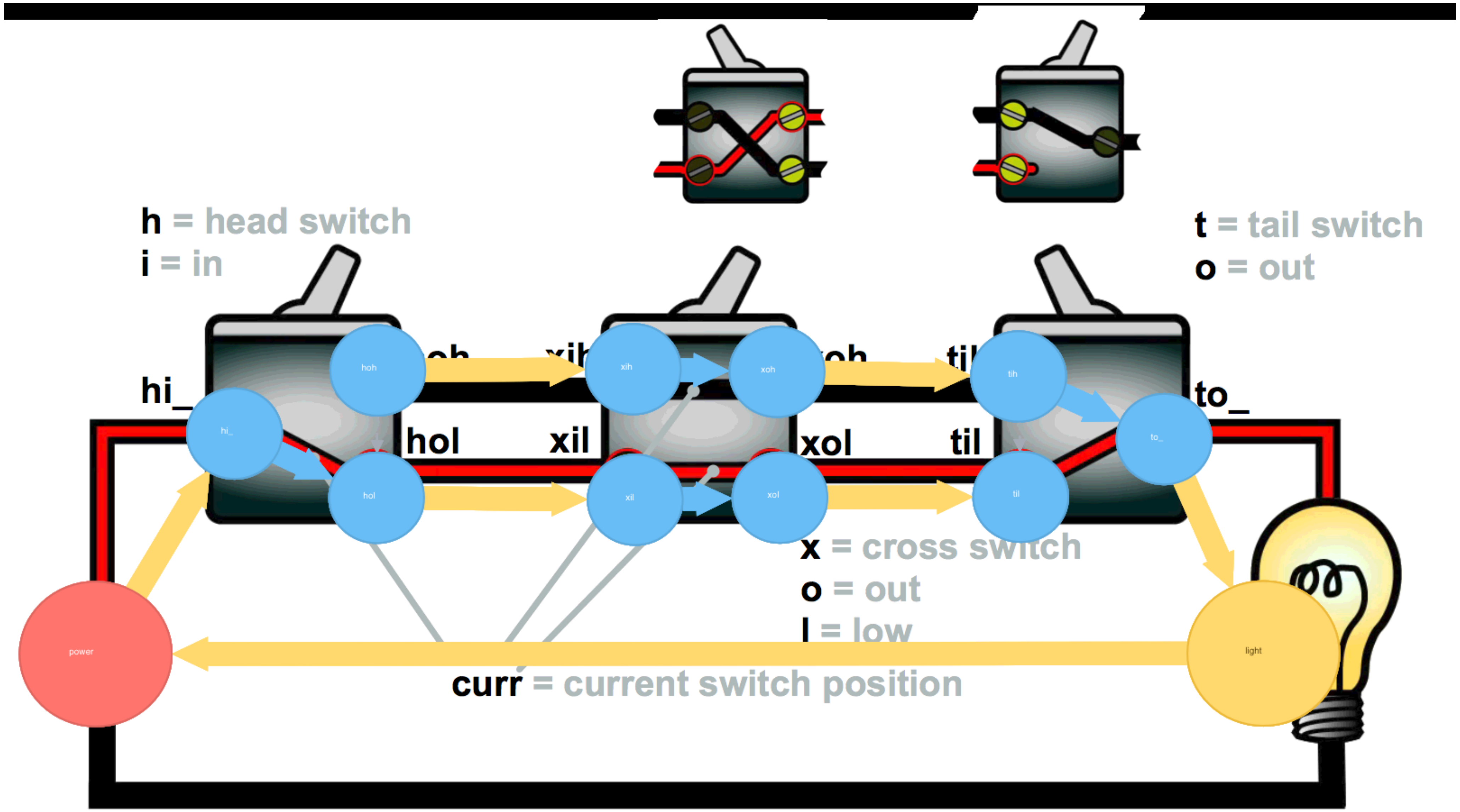


h = head switch
i = in

t = tail switch
o = out

x = cross switch
o = out
l = low

curr = current switch position

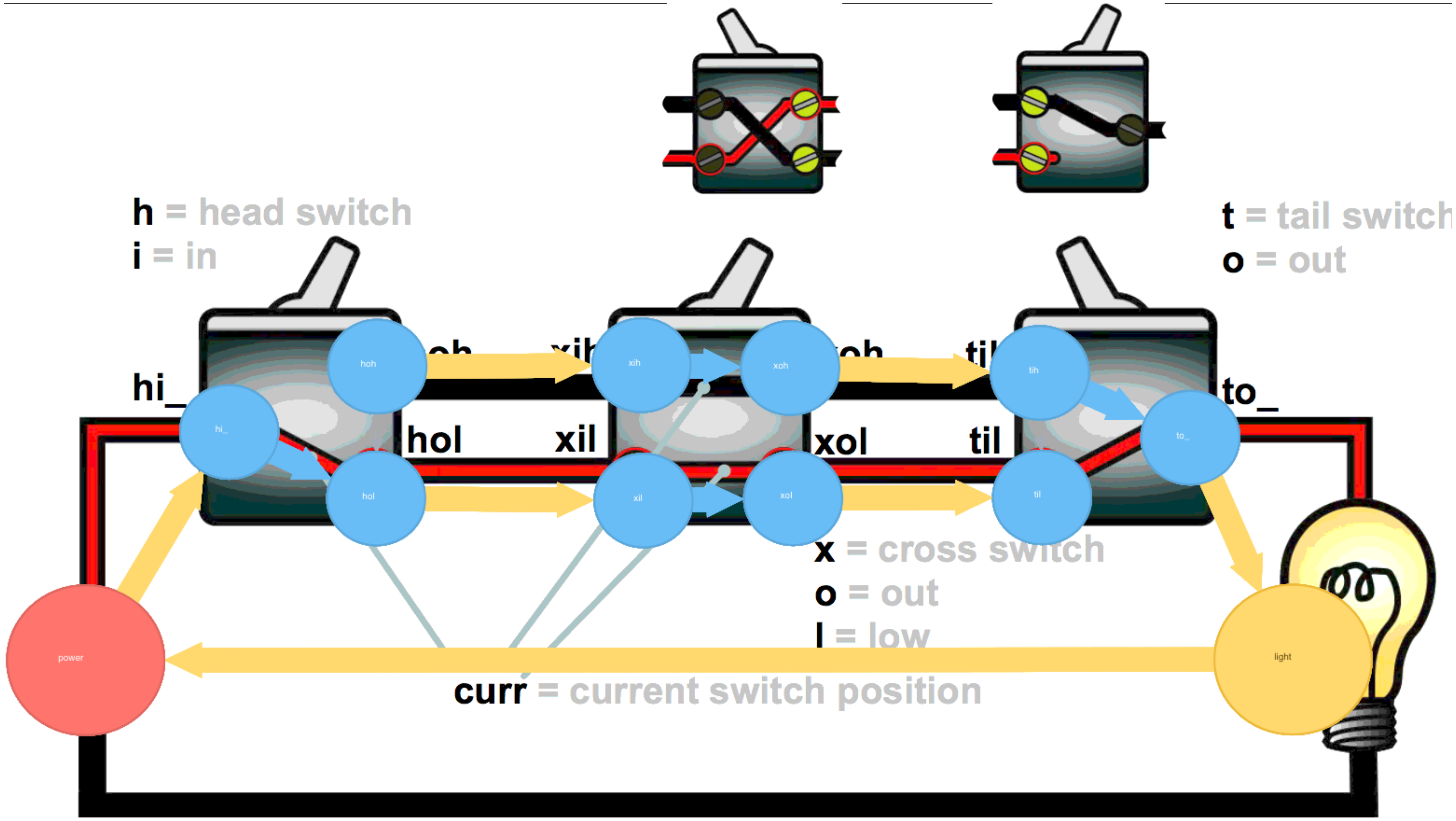


h = head switch
i = in

t = tail switch
o = out

x = cross switch
o = out
l = low

curr = current switch position

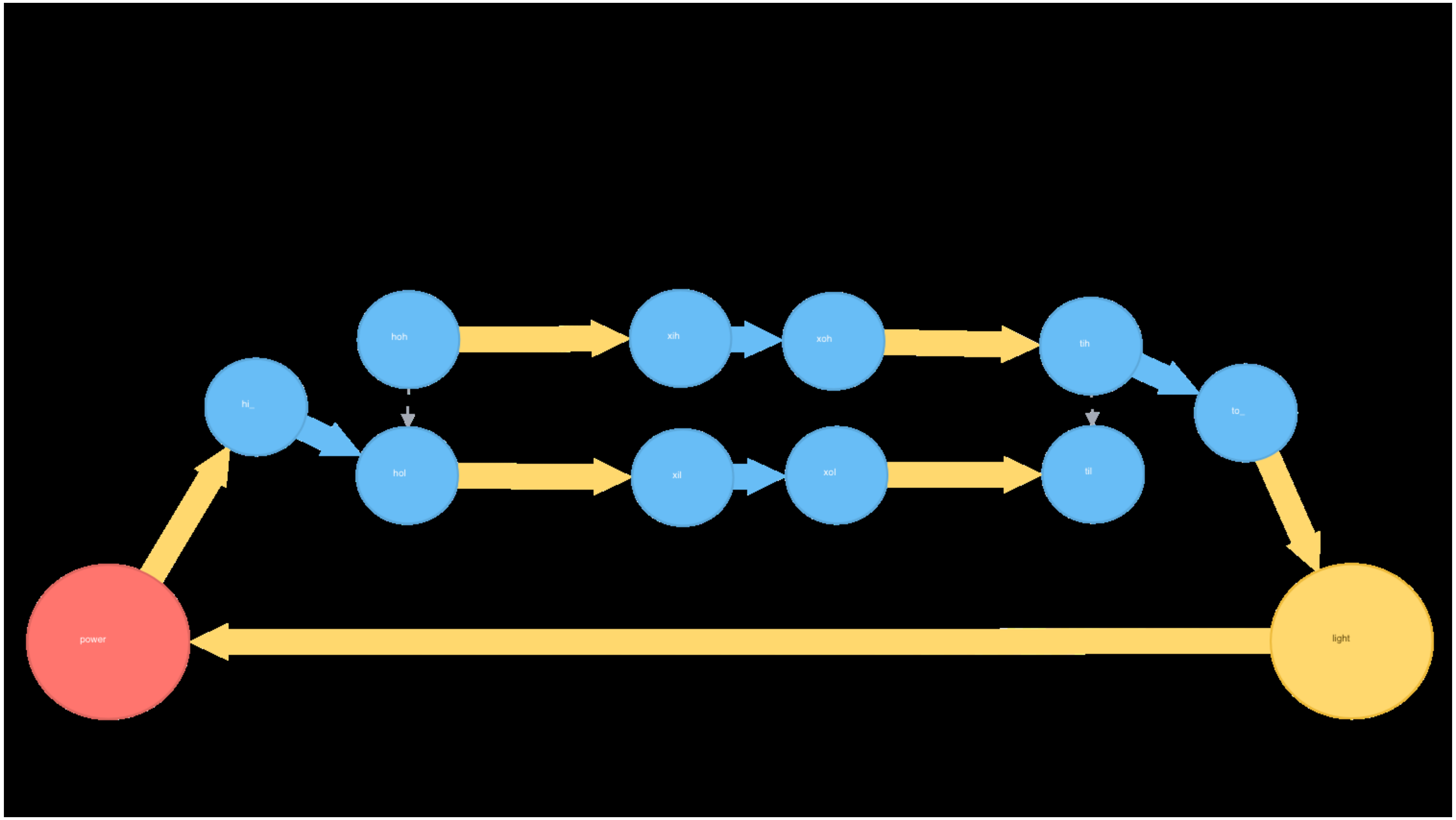


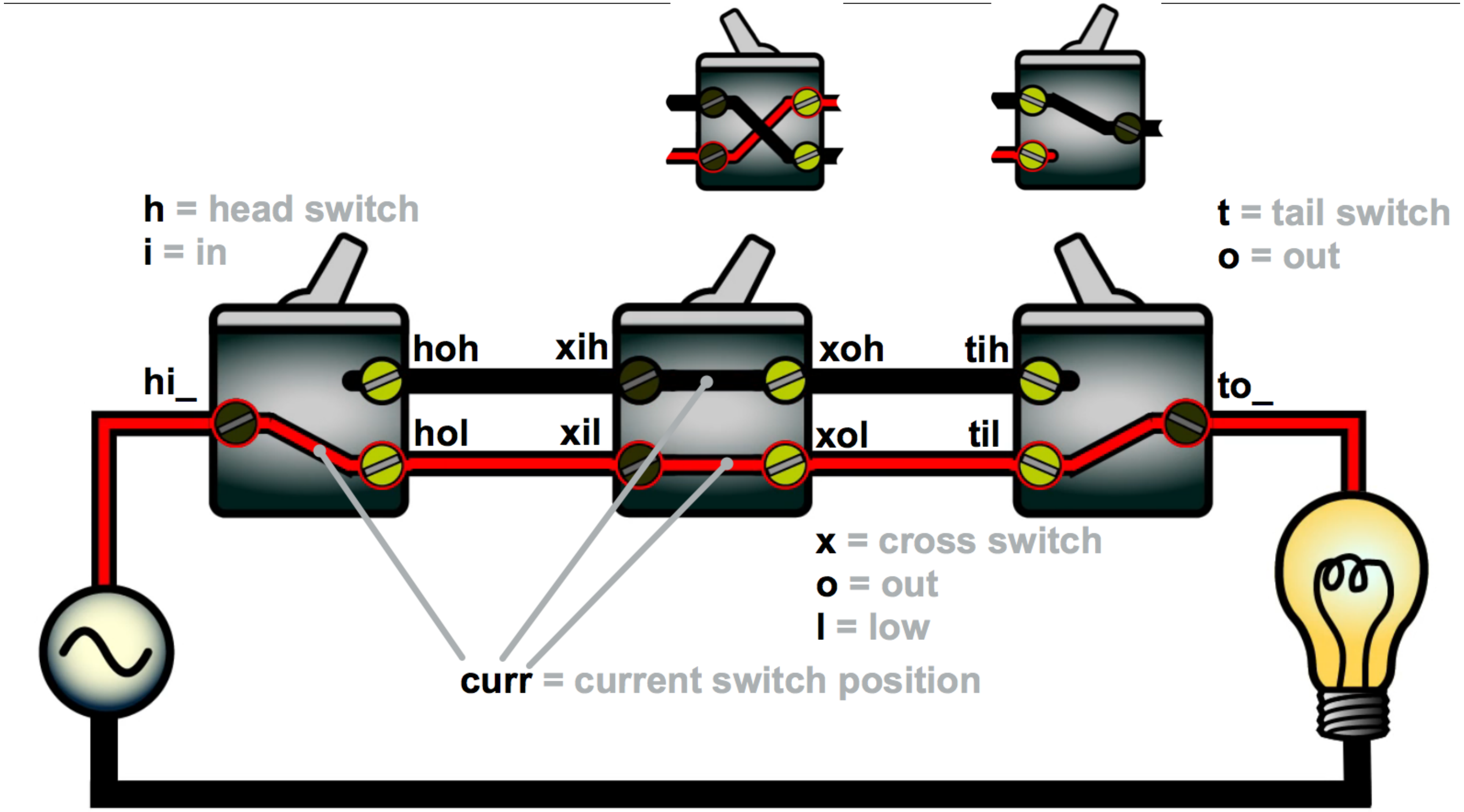
h = head switch
i = in

t = tail switch
o = out

x = cross switch
o = out
l = low

curr = current switch position


hoh
xih
xoh
tih
hi_
to_
hol
xil
xol
til
power
light

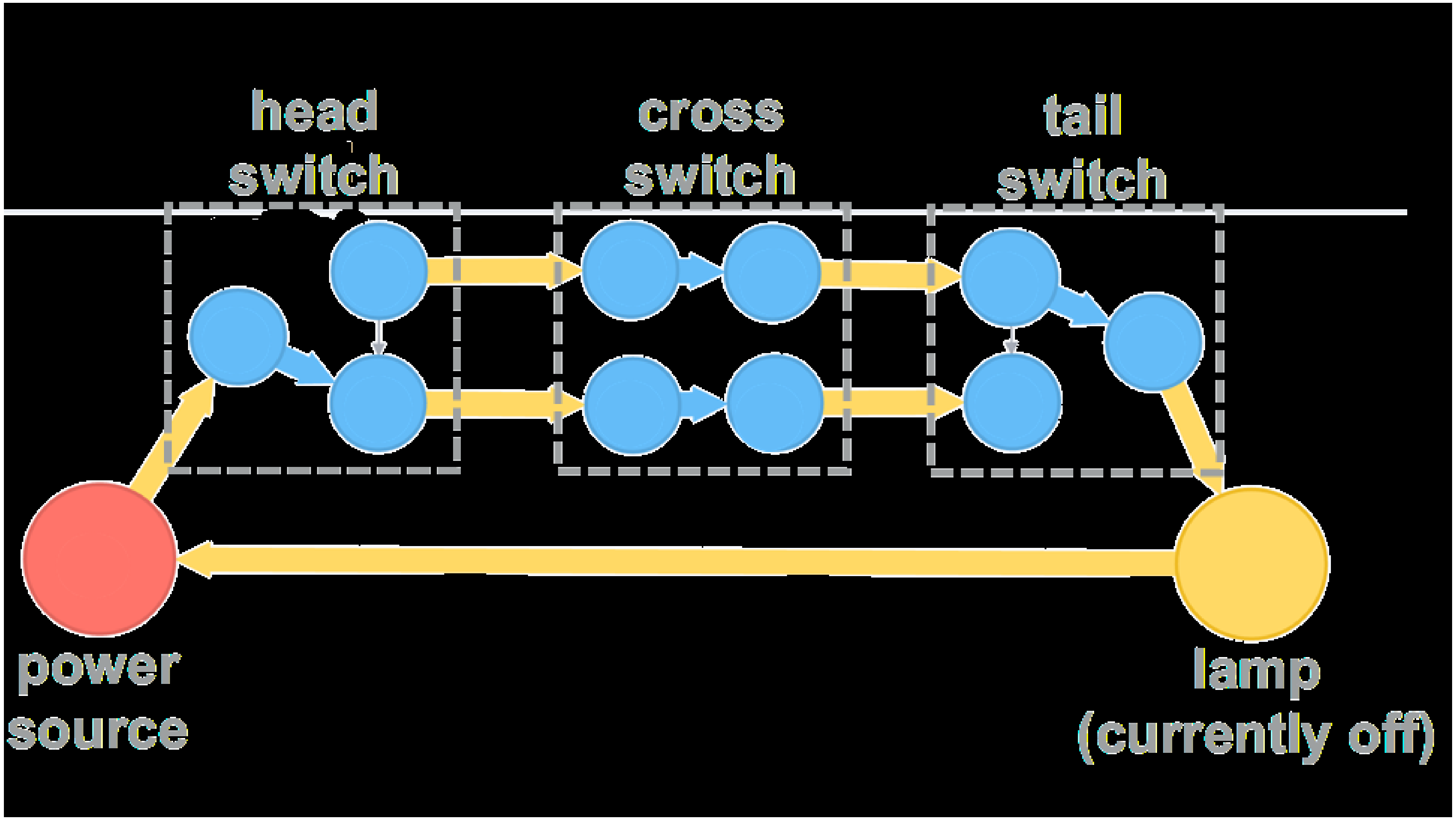


h = head switch
i = in

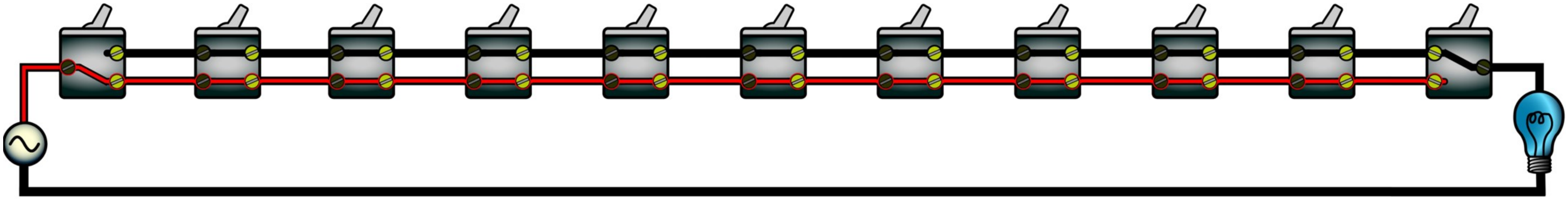
t = tail switch
o = out

x = cross switch
o = out
l = low

curr = current switch position


head switch
cross switch
tail switch
power source
lamp (currently off)

[redacted]

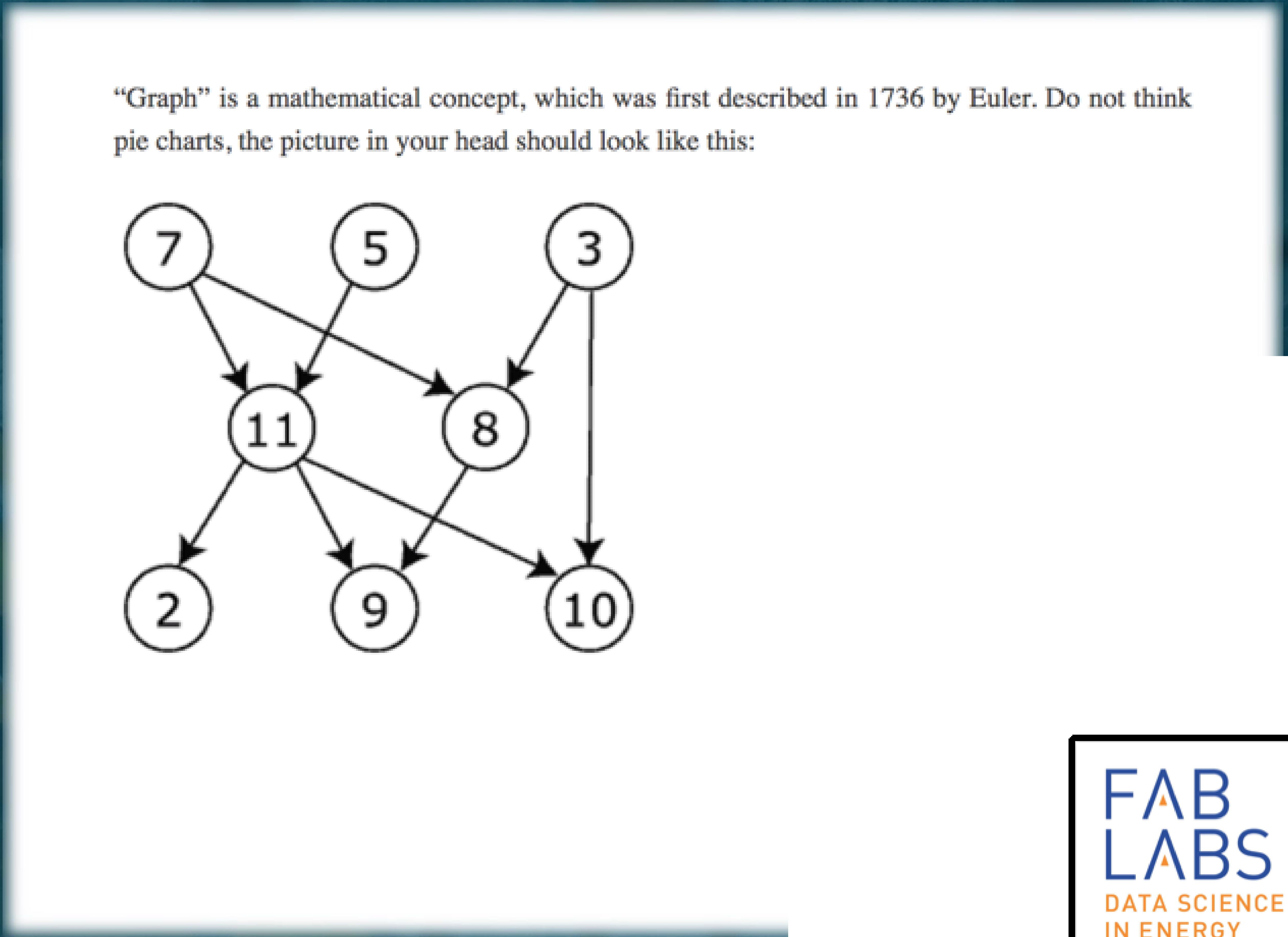


[redacted]

0
9
JUNE

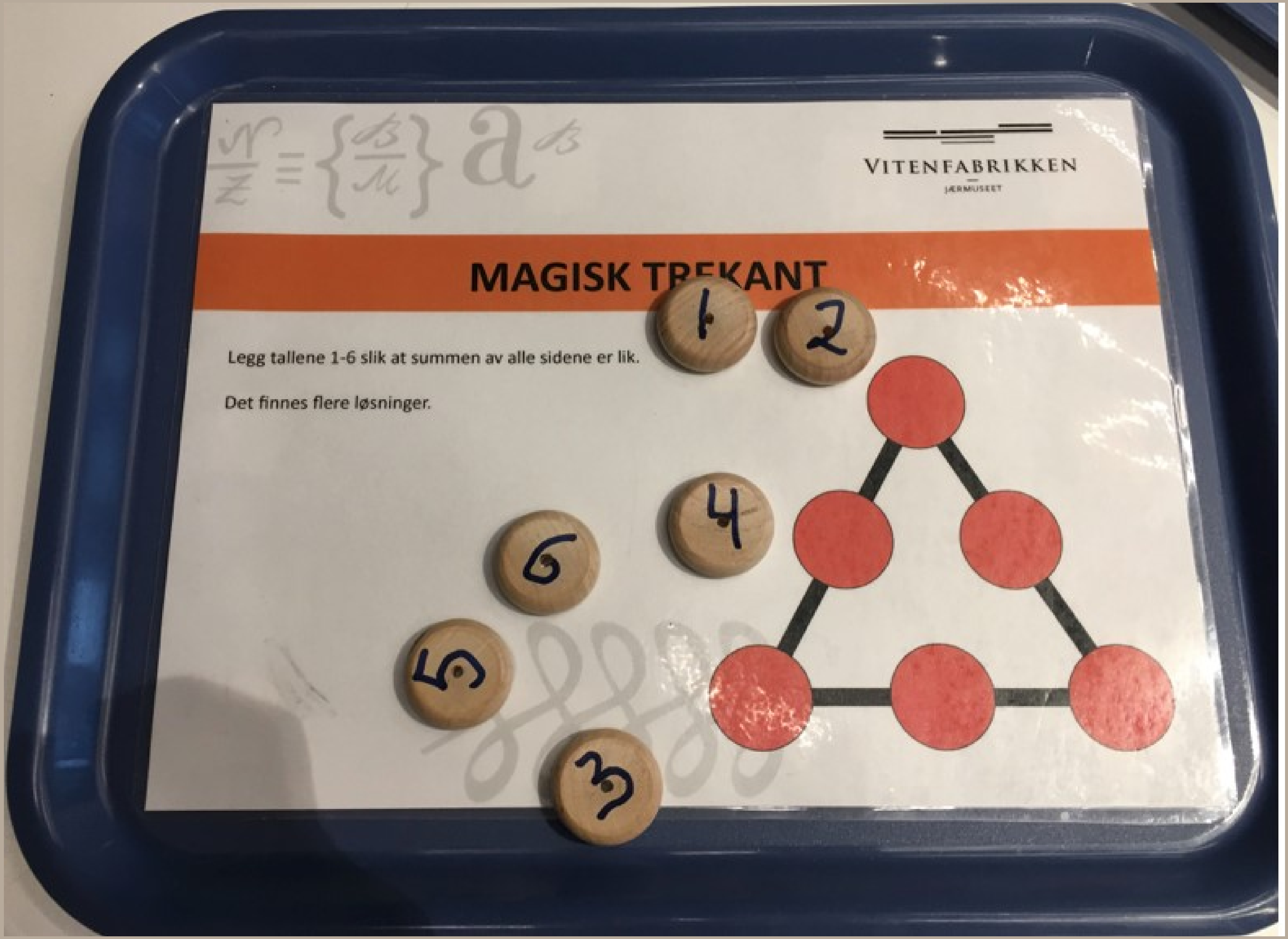
0
4
3
7

3
2
4
3
1
4



“Graph” is a mathematical concept, which was first described in 1736 by Euler. Do not think pie charts, the picture in your head should look like this:

FAB LABS
DATA SCIENCE
IN ENERGY



VITENFABRIKKEN

JÆRMUSEET

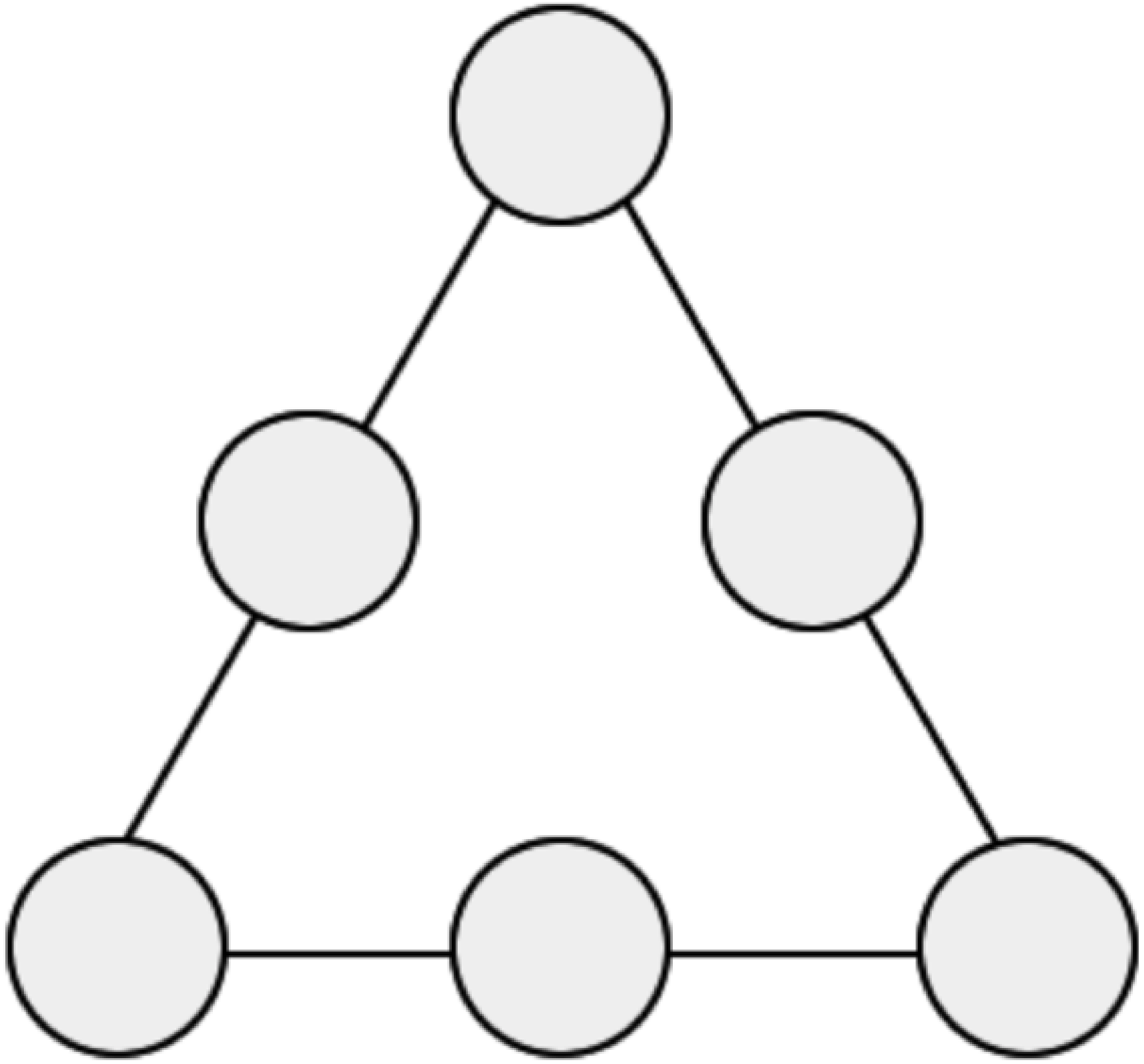
MAGISK TREKANT

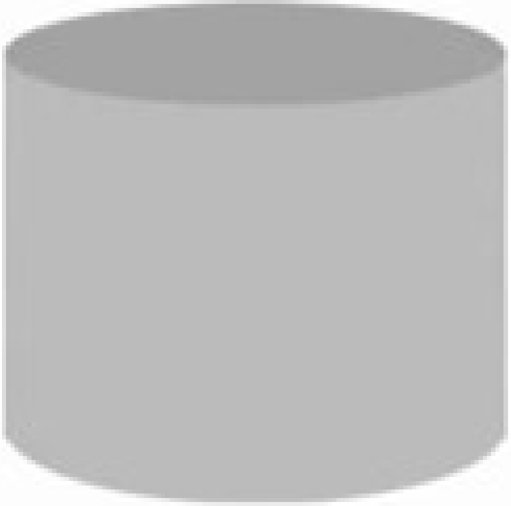
Legg tallene 1-6 slik at summen av alle sidene er lik.

Det finnes flere løsninger.

Magic Triangle

Directions: Arrange the numbers 1, 2, 3, 4, 5, 6 so each row has a sum of 9. Use all six numbers and each number once.

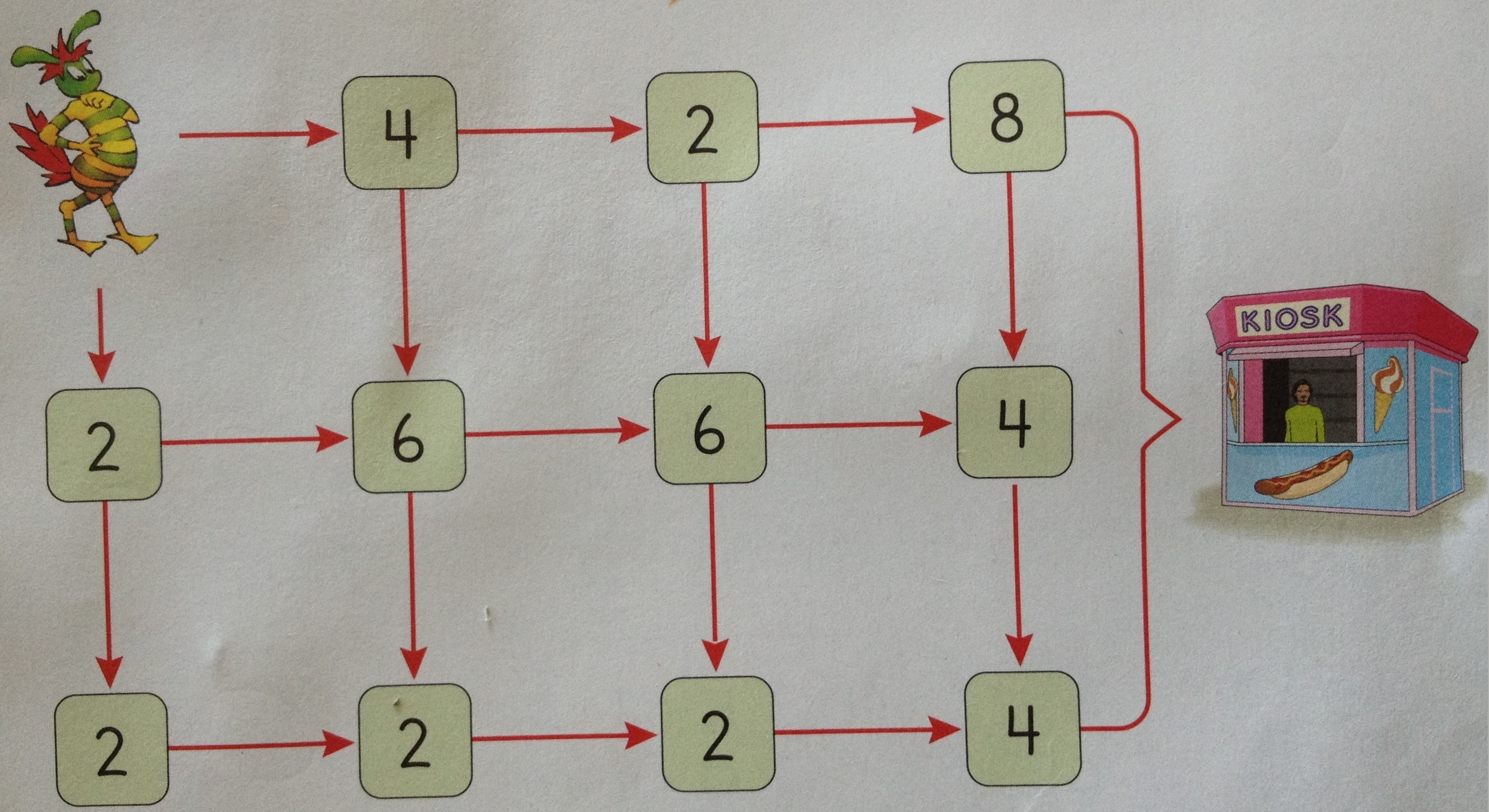


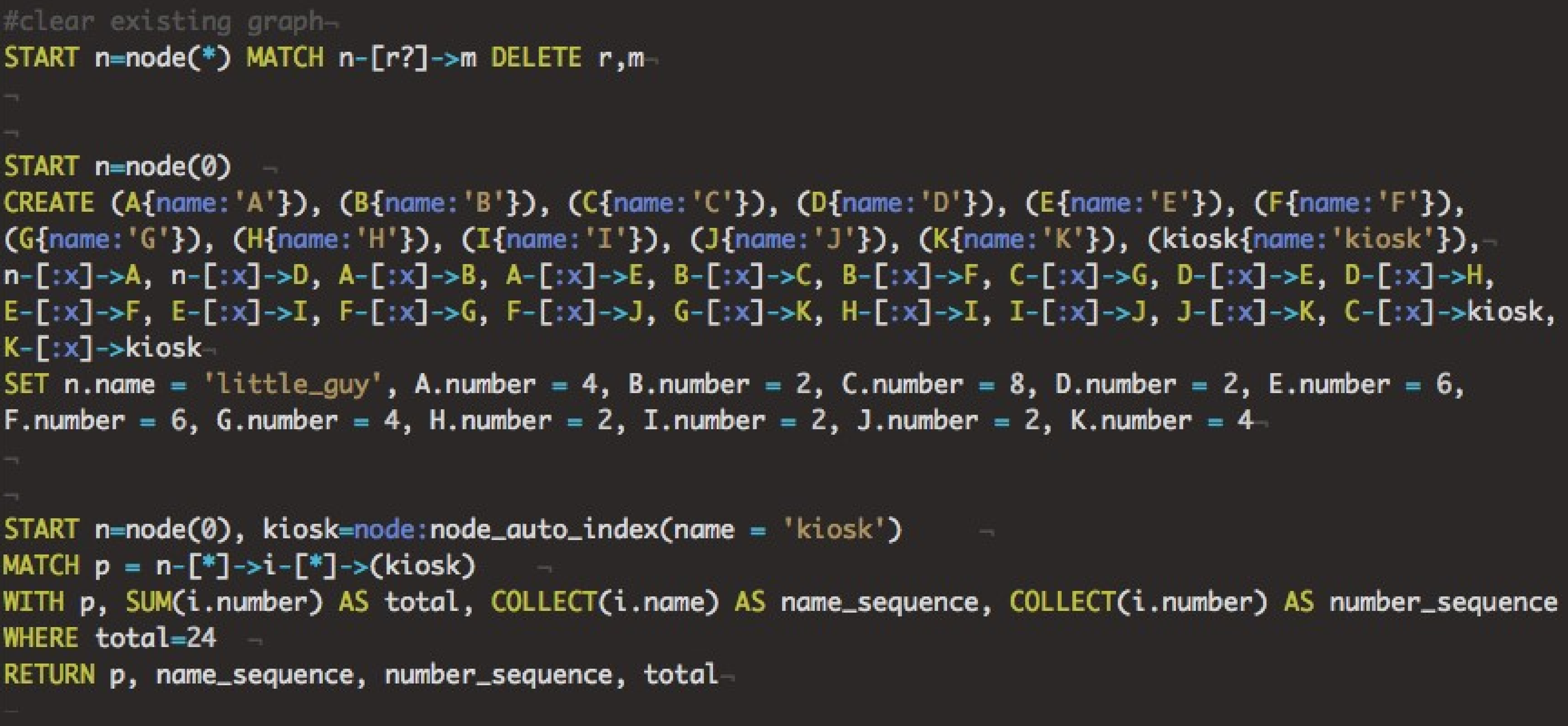


Store / Retrieve

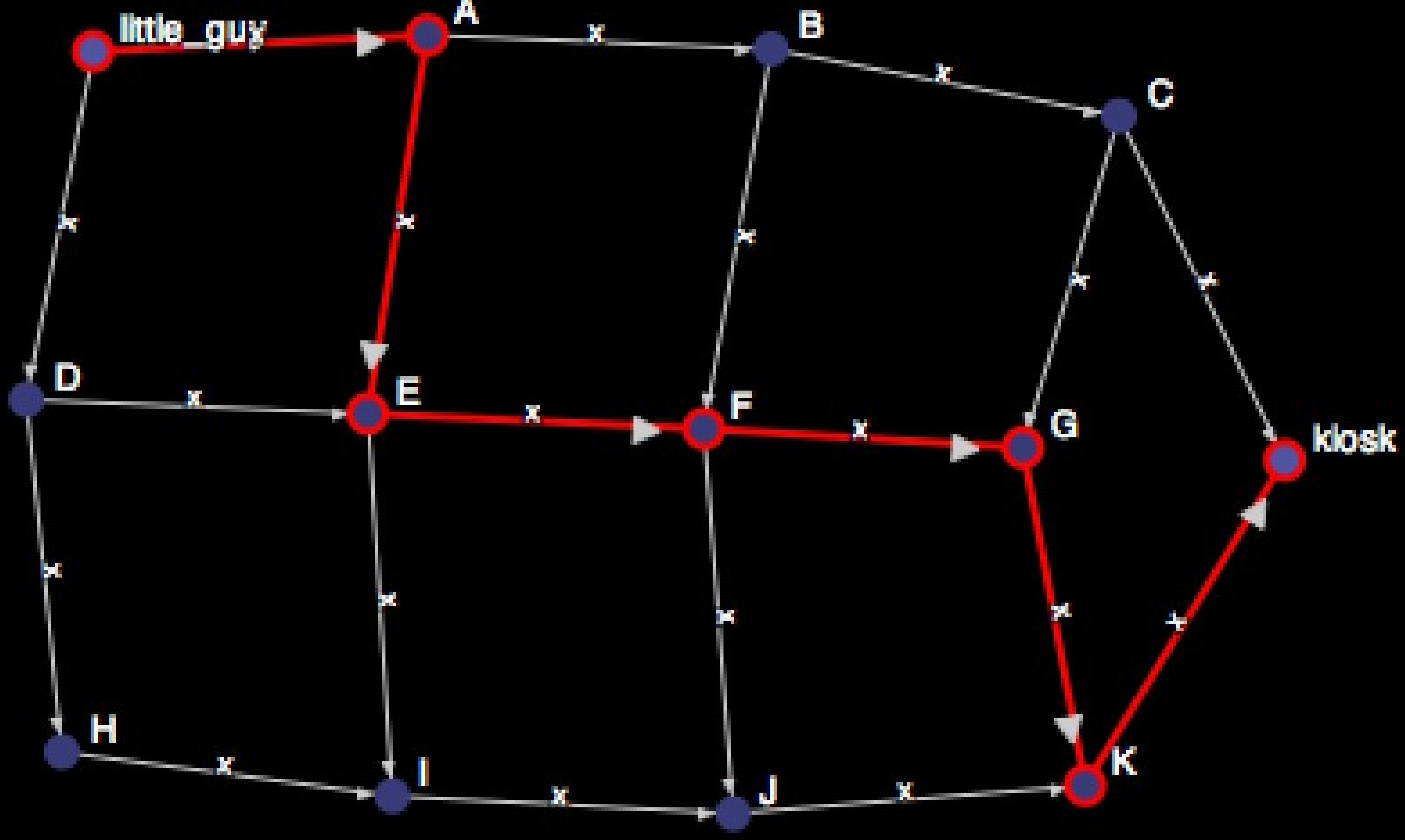
Actionable Insights

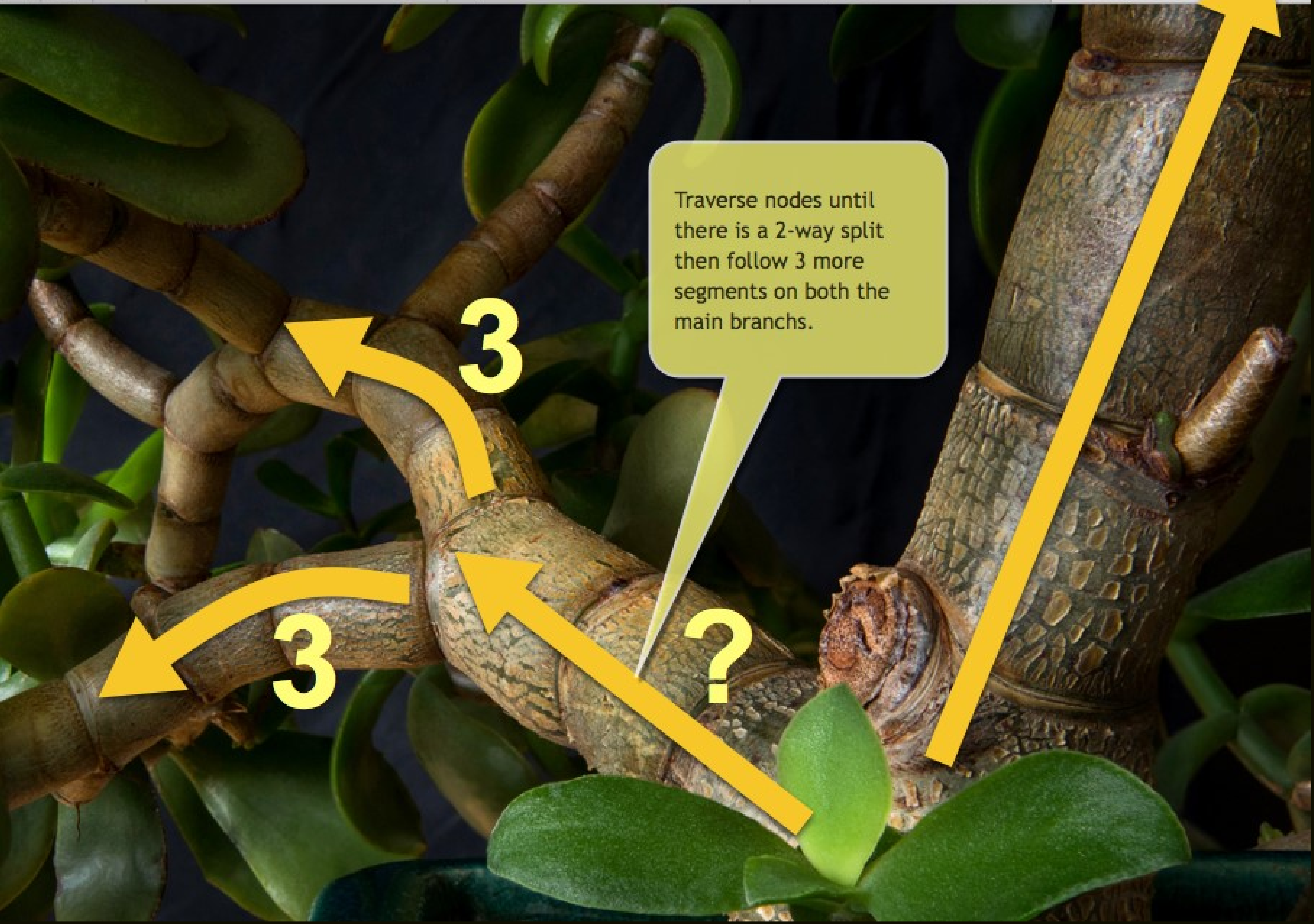
Følg pilene, og lag en vei til kiosken der summen blir 24.



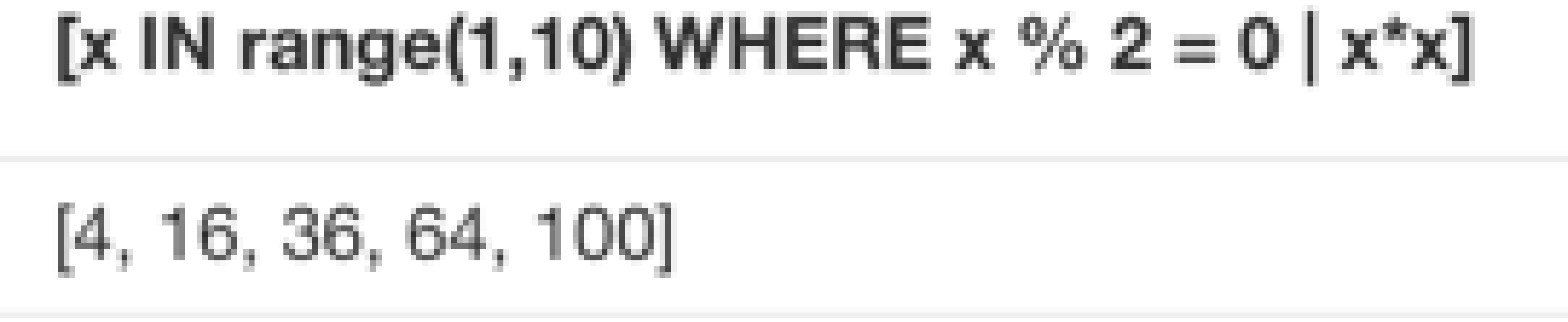


```
#clear existing graph
START n=node(*) MATCH n-[r?]->m DELETE r,m


START n=node(0)
CREATE (A{name:'A'}), (B{name:'B'}), (C{name:'C'}), (D{name:'D'}), (E{name:'E'}), (F{name:'F'}),
(G{name:'G'}), (H{name:'H'}), (I{name:'I'}), (J{name:'J'}), (K{name:'K'}), (kiosk{name:'kiosk'}),
n-[:x]->A, n-[:x]->D, A-[:x]->B, A-[:x]->E, B-[:x]->C, B-[:x]->F, C-[:x]->G, D-[:x]->E, D-[:x]->H,
E-[:x]->F, E-[:x]->I, F-[:x]->G, F-[:x]->J, G-[:x]->K, H-[:x]->I, I-[:x]->J, J-[:x]->K, C-[:x]->kiosk,
K-[:x]->kiosk
SET n.name = 'little_guy', A.number = 4, B.number = 2, C.number = 8, D.number = 2, E.number = 6,
F.number = 6, G.number = 4, H.number = 2, I.number = 2, J.number = 2, K.number = 4


START n=node(0), kiosk=node:node_auto_index(name = 'kiosk')
MATCH p = n-[*]->i-[*]->(kiosk)
WITH p, SUM(i.number) AS total, COLLECT(i.name) AS name_sequence, COLLECT(i.number) AS number_sequence
WHERE total=24
RETURN p, name_sequence, number_sequence, total
```


little_guy
A
B
C
D
E
F
G
kiosk
H
I
J
K
x



Traverse nodes until there is a 2-way split then follow 3 more segments on both the main branchs.



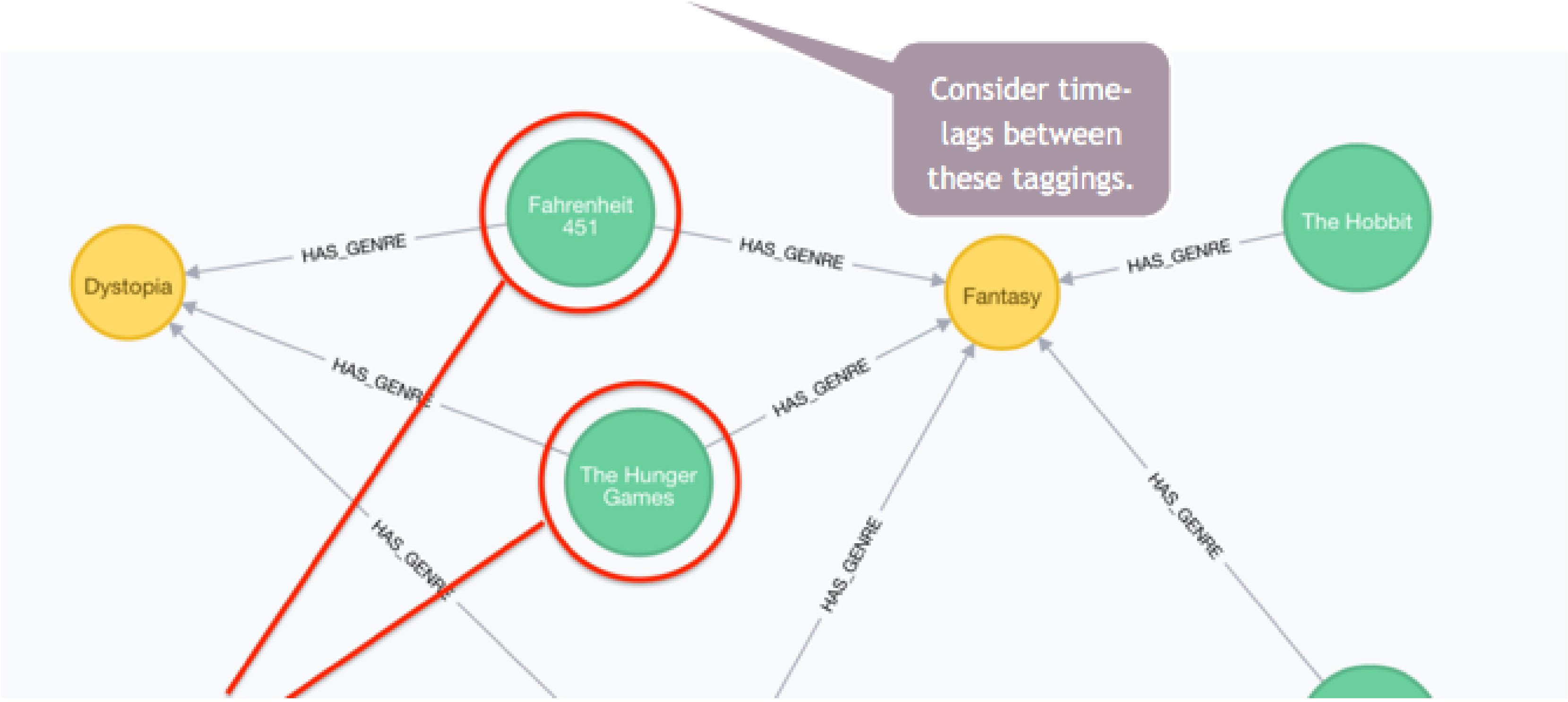
[x IN range(1,10) WHERE x % 2 = 0 | x*x]
[4, 16, 36, 64, 100]

STEP1: Compute co-occurrence

Co-occurrence is the basic building block for the algorithm and is in itself a quite useful relationship because it indicates some degree of overlap between tags and therefore a certain degree of similarity which is something that can be exploited for query expansion or recommendation.

The co-occurrence index between two categories A and B is computed as the portion of items in category A that are also in category B, this is a simple division of the number of items tagged as both A and B divided by the number of items in tagged as A.

COOC(A,B) = #items tagged as both A and B / #items tagged as A



Consider time-lags between these taggings.

what is connected data?

It’s almost as if we didn’t used to know even how to collect data, while preserving its full value. We collected data for a specific set of reports, and when all reports were made, we weren’t able to track details across those reports that came from the same data.

It cost money to store and sift through data. Like an old shirt. If you haven’t worn it in the past year, get rid of it.

Now, we hold the data, and if it’s not stored connected, at least we are capable of reconnecting it.

Questions

FAB LABS
DATA SCIENCE
IN ENERGY