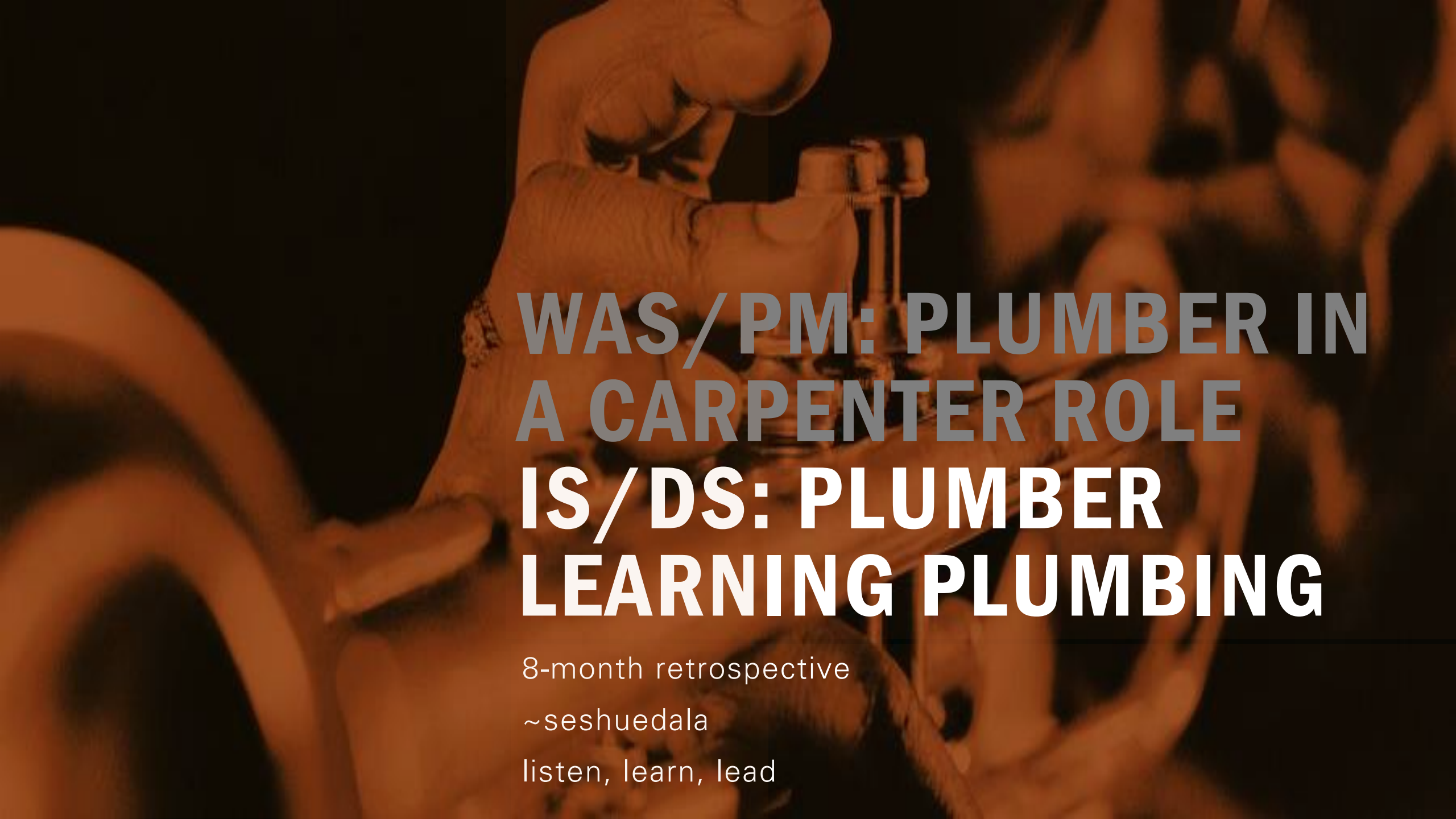


WAS/PM: PLUMBER IN A CARPENTER ROLE

IS/DS: PLUMBER LEARNING PLUMBING

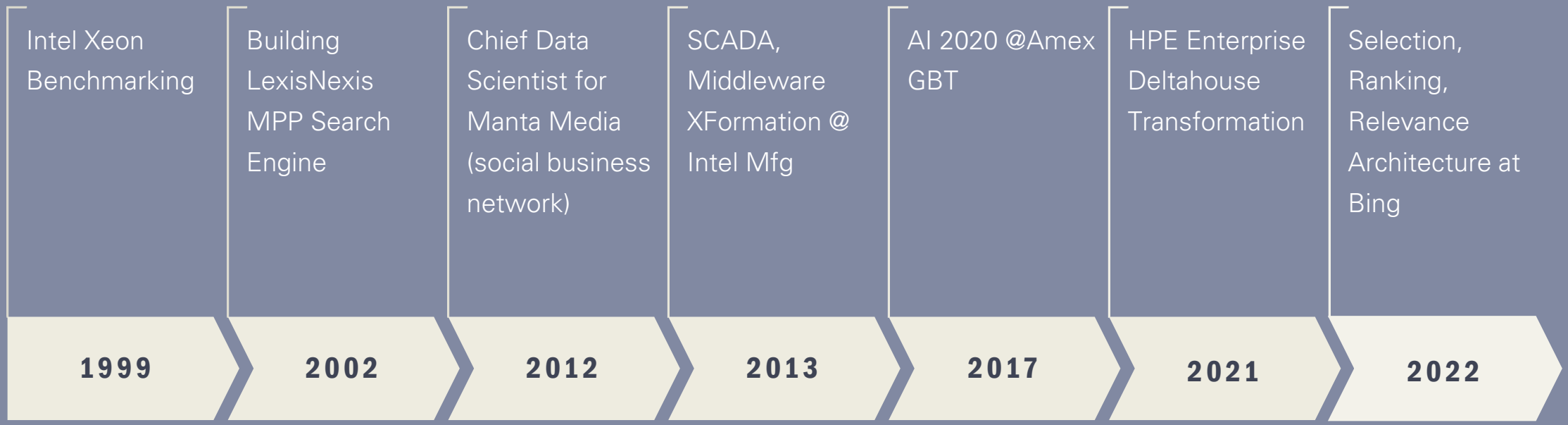
8-month retrospective

~seshuedala

listen, learn, lead

BACKGROUND

TIMELINE – HANDS-ON DELIVERY



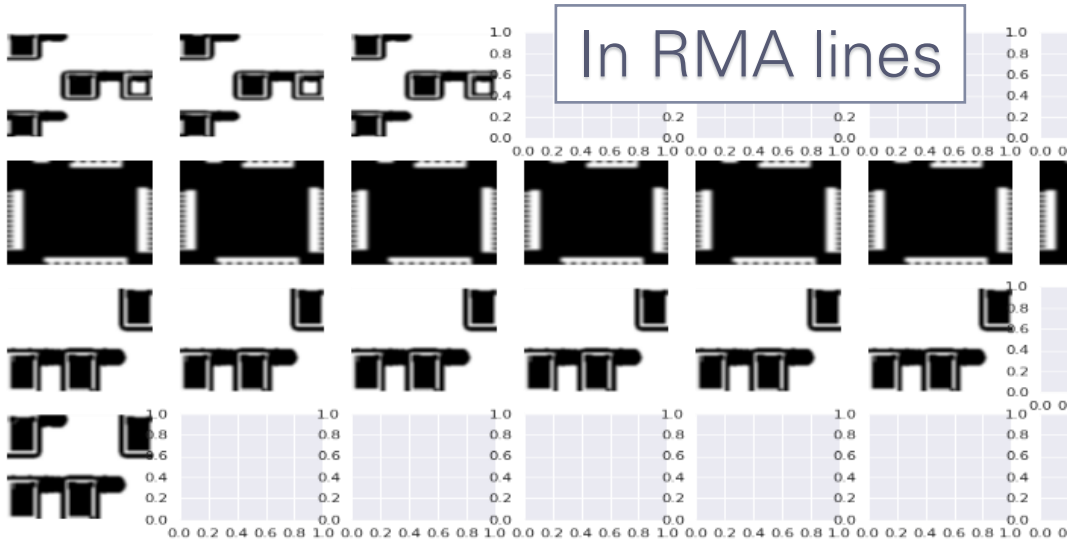
Defect Provenance @Intel

```
pipeline = Pipeline([('pca', pca)])
```

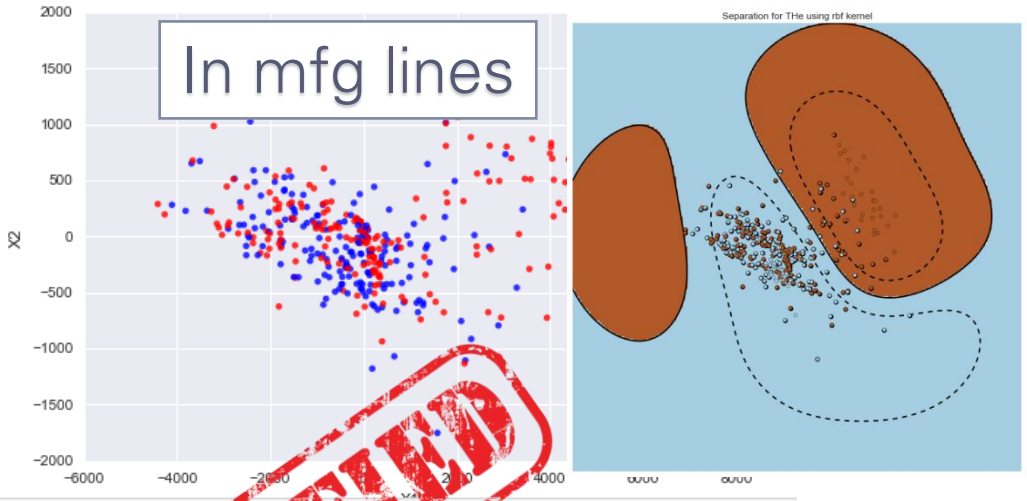
```
reduced_x = pipeline.fit_transform(numeric_x)
X_r2 = pd.DataFrame(reduced_x, columns=['X1', 'X2'])

# Percentage of variance explained for each components
print('Explained variance ratio (first two components): %s'
      % str(pca.explained_variance_ratio_))
ax = sns.regplot(data=X_r2, x='X1', y='X2', x_jitter=10**2, y_jitter=10
plt.show()
```

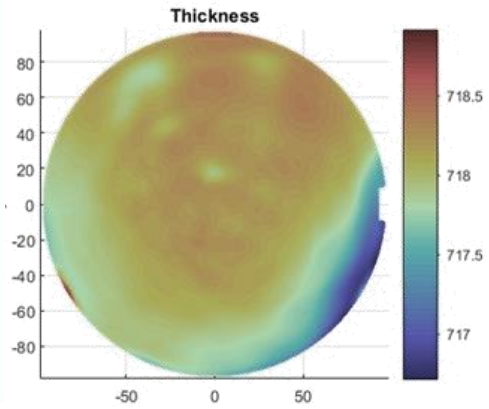
```
Explained variance ratio (first two components): [ 0.8605536   0.055053
```

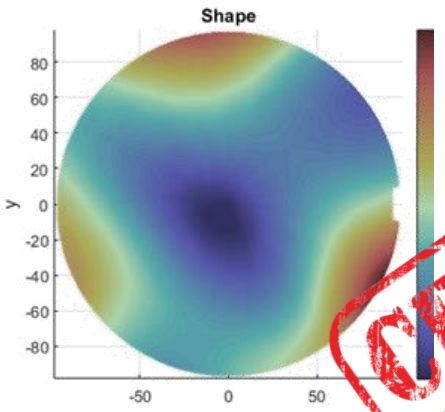
In RMA lines

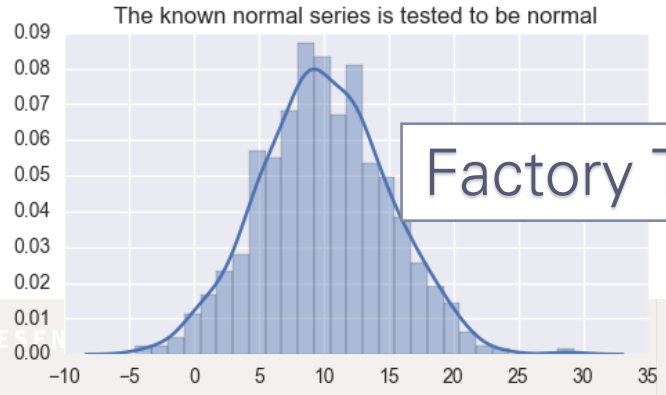


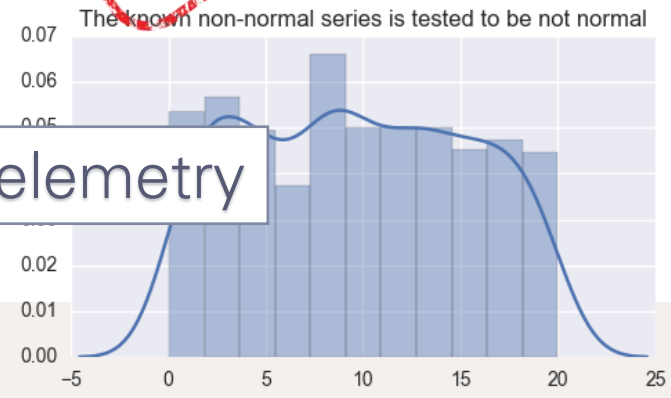
In mfg lines











Factory Telemetry

AI 2020 @ AMEX GLOBAL BUSINESS TRAVEL

Better outcomes for travelers, suppliers, clients, and counselors

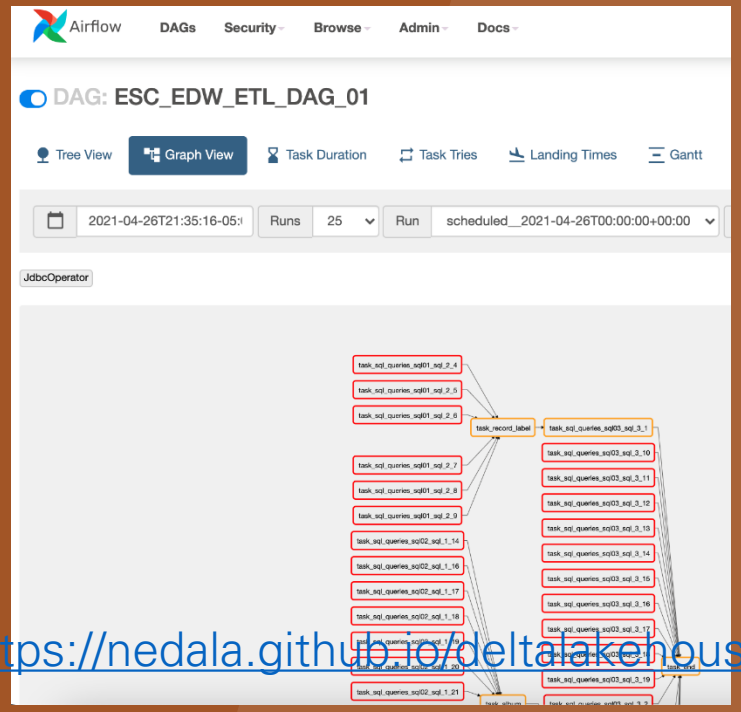
2019

WHAT WE HAVE DELIVERED?	0-6 MONTHS	6-15 MONTHS	15 MONTHS+
Hotel Recommender	Champion Challenger	TravelDash	Snap-n-shoot Expense
Benchmarking	Flight Recommender	Crisis Alerts Monitoring	Break the भाषा barrier
Forecasting	Customer Canary	Gilbert, Gaby, Gunther Digital Bots	Create Information Turks
Total Trip Identification	Best week/place to meet	Compass Dashboards	Self-Documenting Data Carts
Hotel Normalization	Dynamic Insights	What-If Savings PI	Alexa – Book a Hotel
Card Triangulation	Pre-sales FAQ Support	Real-time Customer NPS	Premier Insights by Gilbert
Outlier Control	Insights Automation	One-click Booking (+Rail/Car)	Opportunity Reminders
Book “Now”	24/7/365 DNA EOY	In-Trip Recommendations	Satisfaction Surveys
Prospect Harvesting	MM Client Prospecting	Trip Docket Visualizations	Dead Inventory Optimizations
Attrition Modeling	Expense Matching	Touch “Once”; Don’t “Reshop”	Gracias Road Warriors
Chatbot Automation	Preemptive Messaging	Meet your Best TC, Supplier	Omnichannel Monitoring

DATA-FIRST TRANSFORMATION

CLOUD/DIGITAL TRANSFORMATION

Pioneering cloud first PaaS Solutions at Fortune 50

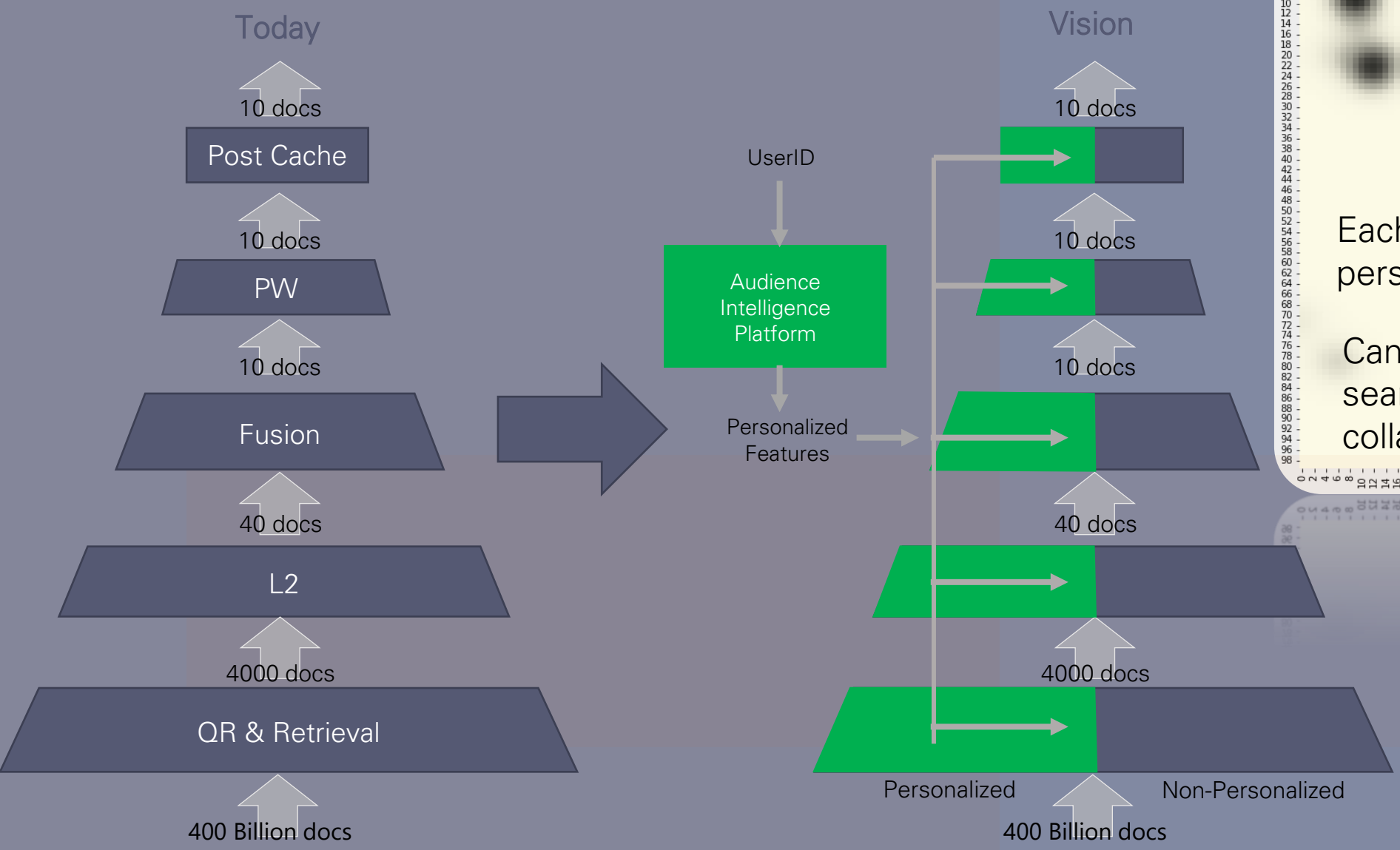


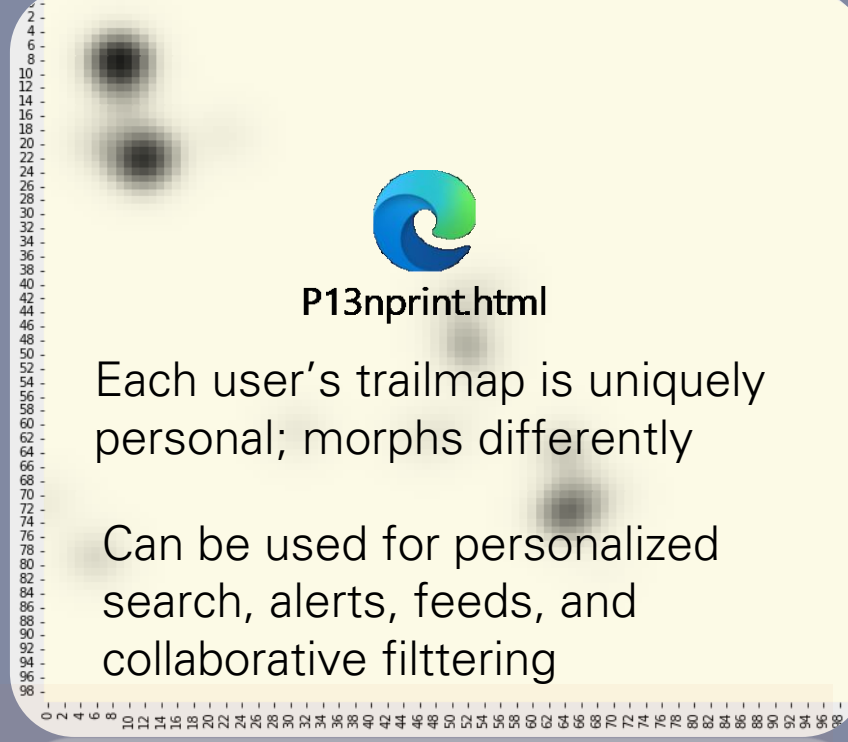
https://nedala.github.io/deltalakehouse/

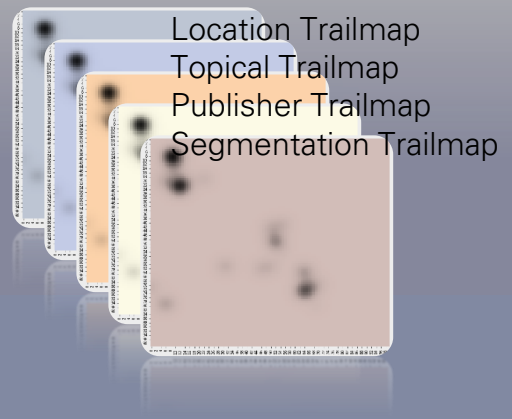
BI/MI/AI ACCELERATOR SERVICES

Removing FUD barriers for adoption of Big Data, Machine Learning and AI

https://data-bloom.com/







FIRST PROBLEM @ WEBXT/P13N

The Dewey Decimal Classification organizes library materials by discipline or field of study. A hierarchical notation of organizing knowledge

- 500 Natural sciences and mathematics
  - 510 Mathematics
    - 516 Geometry
      - 516.3 Analytic geometries
        - 516.37 Metric differential geometries
          - 516.375 Finsler geometry

SEARCH CHALLENGE

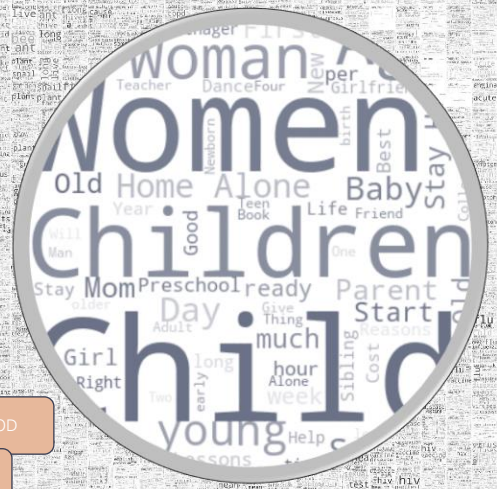
Organized catalog of information lets directed search thrive

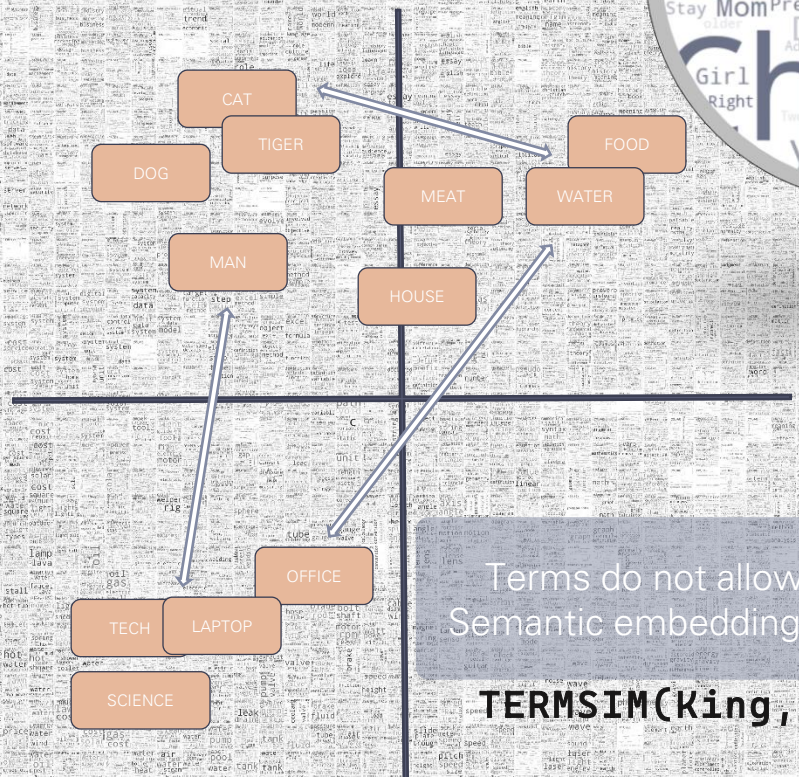
SEMANTIC EMBEDDINGS, NOT BAG-OF-WORDS

SEMANTIC ALGEBRA

```
import numpy as np
from scipy.spatial import distance
# Load Spacy Model
nlp_md = spacy.load('en_core_web_md')
# Get vector embedding
vec = lambda term: nlp_md(term).vector

print("Similarity between dog and canine is: ",
      (1 - distance.cosine(vec('dog'), vec('canine'))) * 100, "%")
print("Similarity between dog and tiger is: ",
      (1 - distance.cosine(vec('dog'), vec('tiger'))) * 100, "%")
print("Similarity between dog and quantum is: ",
      (1 - distance.cosine(vec('dog'), vec('quantum'))) * 100, "%")
print("Similarity between 123 and 345 is: ",
      (1 - distance.cosine(vec('123'), vec('345'))) * 100, "%")
print("Similarity between one and 1 is: ",
      (1 - distance.cosine(vec('one'), vec('1'))) * 100, "%")
print("King - Man + Woman = Queen? ",
    (1 -
     distance.cosine(vec('king') - vec('man') + vec('woman'), vec('queen'))) * 100, "%")
```

```
Similarity between dog and canine is:  100.0 %
Similarity between dog and tiger is:  43.654659390449524 %
Similarity between dog and quantum is:  4.363442584872246 %
Similarity between 123 and 345 is:  64.0597403049469 %
Similarity between one and 1 is:  26.636192202568054 %
King - Man + Woman = Queen?  78.80843877792358 %
```





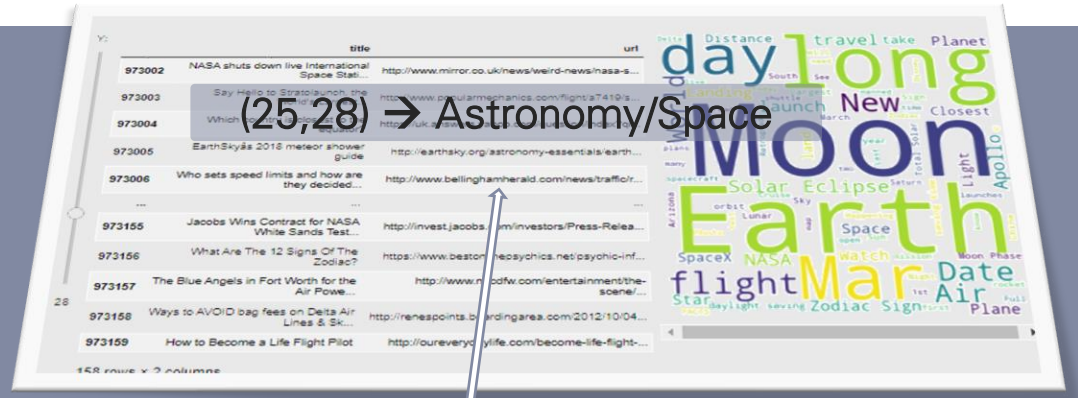
Notice the zipcode is locality preserving.

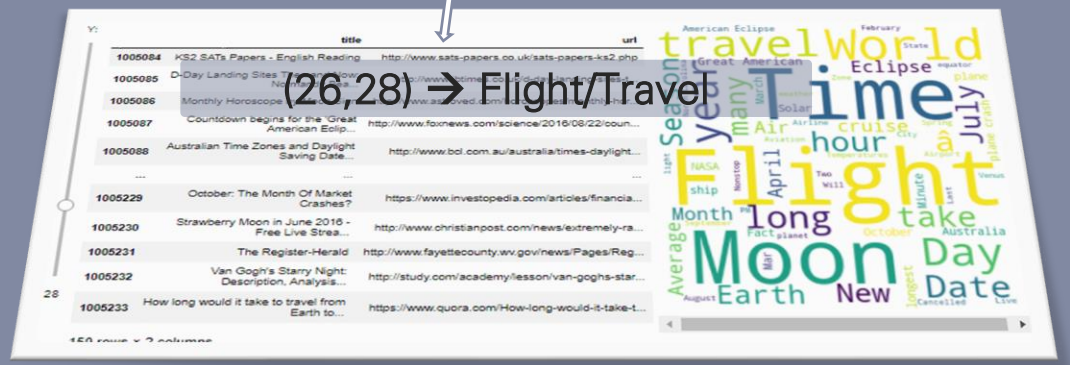
- 12601 – Poughkeepsie, NY
- 45435 – Dayton, OH
- 95630 – Folsom, CA
- 95661 – Roseville, CA

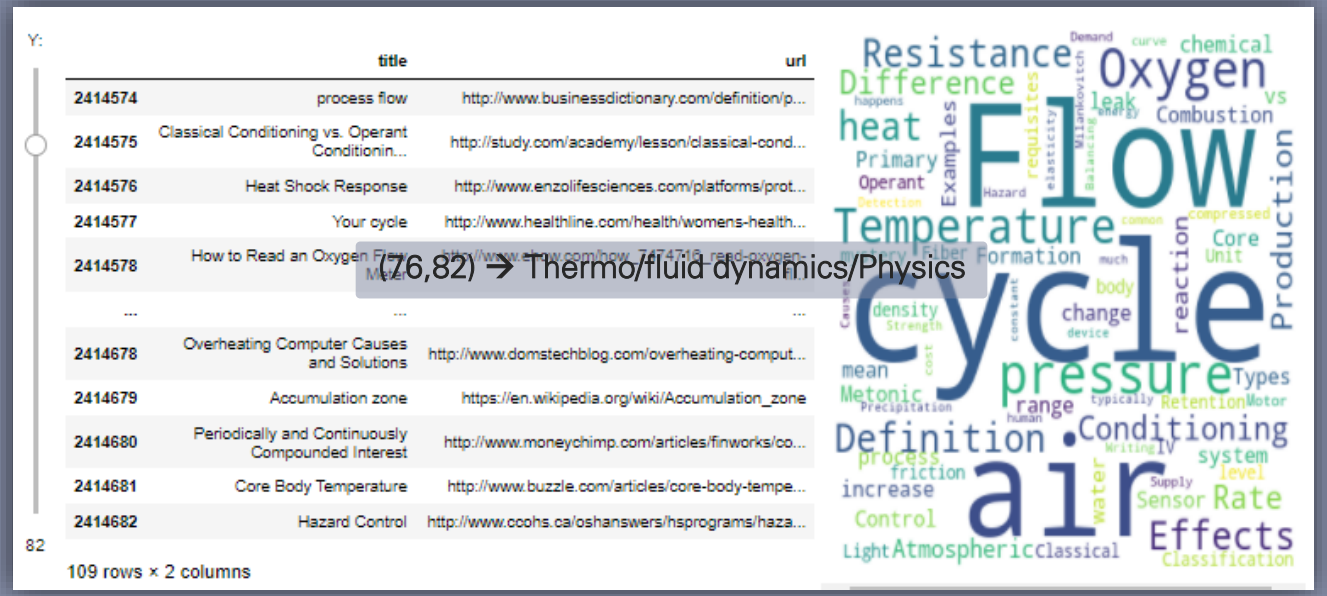
Terms do not allow math, spatial logic.
Semantic embeddings allow vector algebra

`TERMSIM(King, Queen) = 0`

“LOCALITY PRESERVING” 2D QUANTIZATION





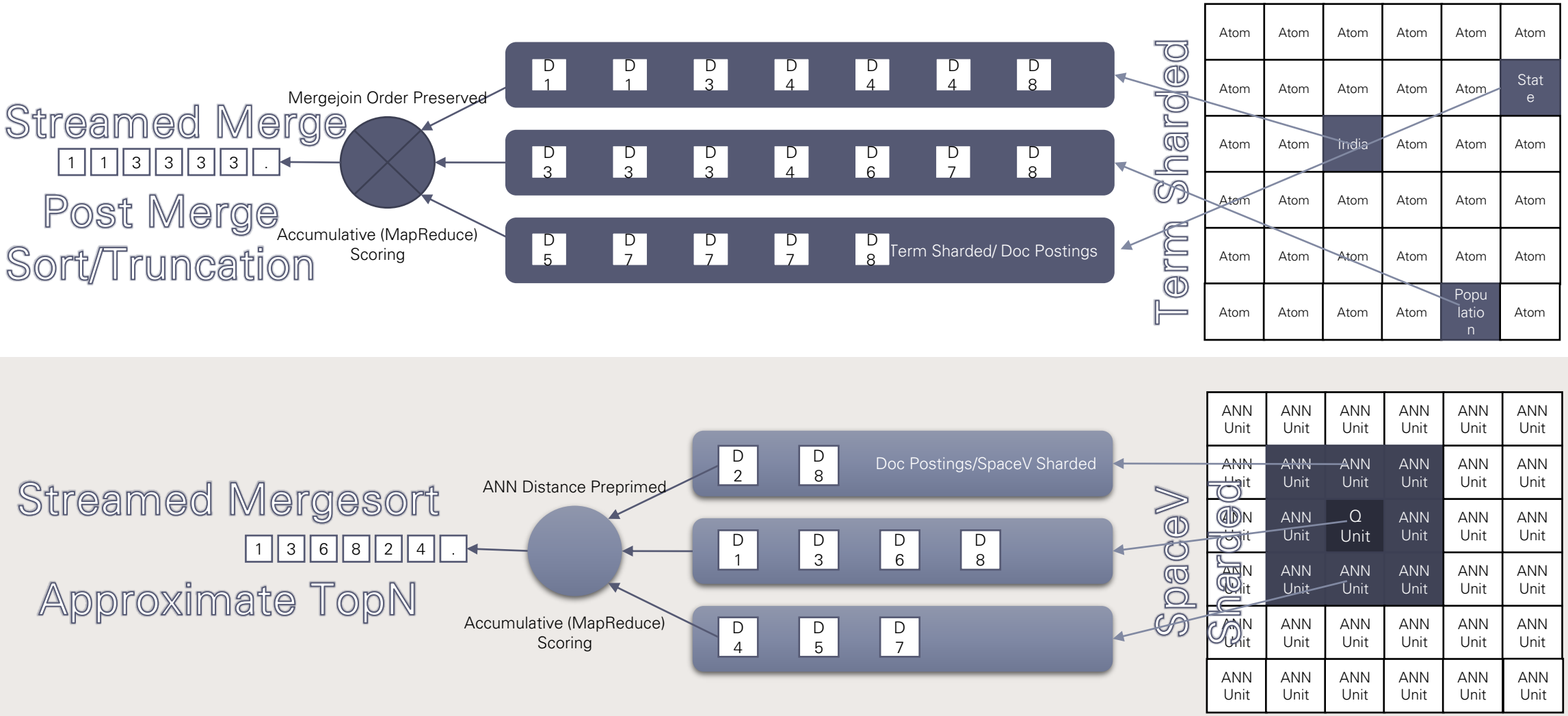


SOM BENEFITS FROM INCREMENTAL TRANSFER LEARNING + MPP SPARK CONFIGURABLE GRID SIZE/HIERARCHY *ANN1 → 1.A → 1.A.1*

TIGER?
FLUID?
SPACEV

Incepting semantic similarity (ANN, KNN) into selection layer

Using guided search with a Dewey Topic Map



NOTICE ANN SIMILARITY TO THE *FULL QUERY* IS PREBAKED VERY EARLY

terms King Man Woman

Query

	Term	TermID	Vector
0	king	0000000000000000AB3328BA7E060DEB	[0.31542, -0.35068, 0.42923, -0.53825, -0.1848...
1	man	00000000000000010CD976B896AD4F0	[-0.1731, 0.20663, 0.016543, -0.31026, 0.01971...
2	woman	000000000000000082000102200480C0	[0.025567, 0.27885, -0.16992, 0.27348, -0.0549...

Full Embedding PoV

Search Results for ['king', 'man', 'woman']

	Term	TermID	Vector	bmu	sem_vector
0	king	0000000000000000AB3328BA7E060DEB	[0.31542, -0.35068, 0.42923, -0.53825, -0.1848...	(54, 27)	bb7ea13e548cb3be09c4db3ec1ca09bf363c3dbecd239f...
1	man	00000000000000010CD976B896AD4F0	[-0.1731, 0.20663, 0.016543, -0.31026, 0.01971...	(68, 22)	2041...
2	woman	000000000000000082000102200480C0	[0.025567, 0.27885, -0.16992, 0.27348, -0.0549...	(73, 22)	e371d13c6dc58e3e82ff2dbe92058...

Prior Prescience into Potential Answers

Semantic Spatial Selection

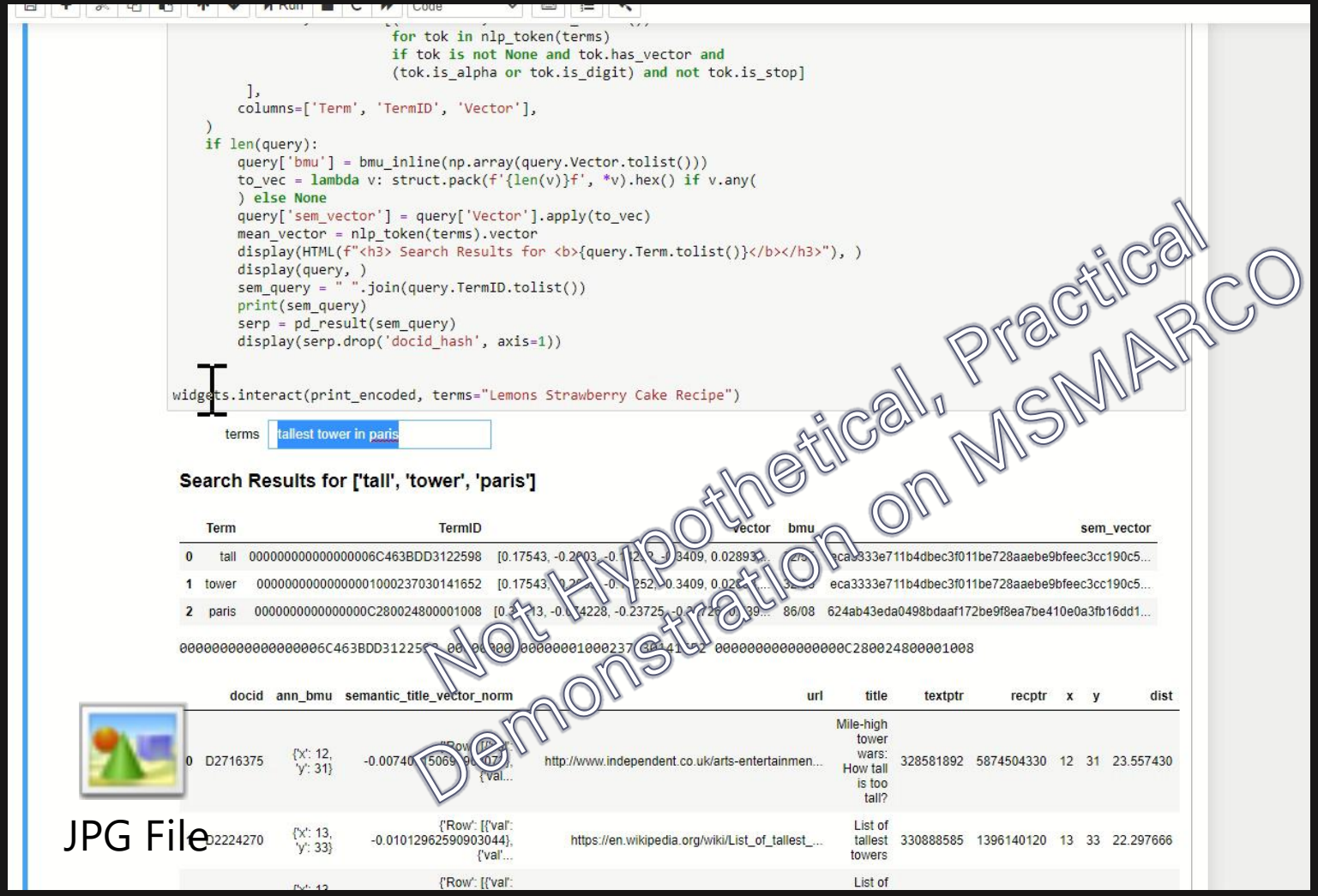
00000000010CD976B896AD4F0 000000000000000082000102200480C0 (98, 99)

	docid	ann	num_terms	url	title	textptr	recptr	x	y	dist	diversity
0	D894022	{'x': 68, 'y': 19}	4	http://www.kindredkreators.org/role1.php	The Man is the King The Woman is the Queen	9446137	136524099	68	19	33.691687	3
1	D2692469	{'x': 8, 'y': 0}	2038	http://www.1728.org/page8.htm	O S C A R SThe next Academy Awards (Osca...	33035198	4659786374	8	0	73.653734	3

It works

EXOCENTRIC & ENDOCENTRIC SCALES OF REFERENCE

SEARCHING QUICKLY WITH SEMANTIC ANN



```
                    for tok in nlp_token(terms)
                    if tok is not None and tok.has_vector and
                    (tok.is_alpha or tok.is_digit) and not tok.is_stop]
            ],
            columns=['Term', 'TermID', 'Vector'],
        )
        if len(query):
            query['bmu'] = bmu_inline(np.array(query.Vector.tolist()))
            to_vec = lambda v: struct.pack(f'{len(v)}f', *v).hex() if v.any(
            ) else None
            query['sem_vector'] = query['Vector'].apply(to_vec)
            mean_vector = nlp_token(terms).vector
            display(HTML(f"<h3> Search Results for <b>{query.Term.tolist()}</b></h3>"), )
            display(query, )
            sem_query = " ".join(query.TermID.tolist())
            print(sem_query)
            serp = pd_result(sem_query)
            display(serp.drop('docid_hash', axis=1))

widgets.interact(print_encoded, terms="Lemons Strawberry Cake Recipe")
```

terms: tallest tower in paris

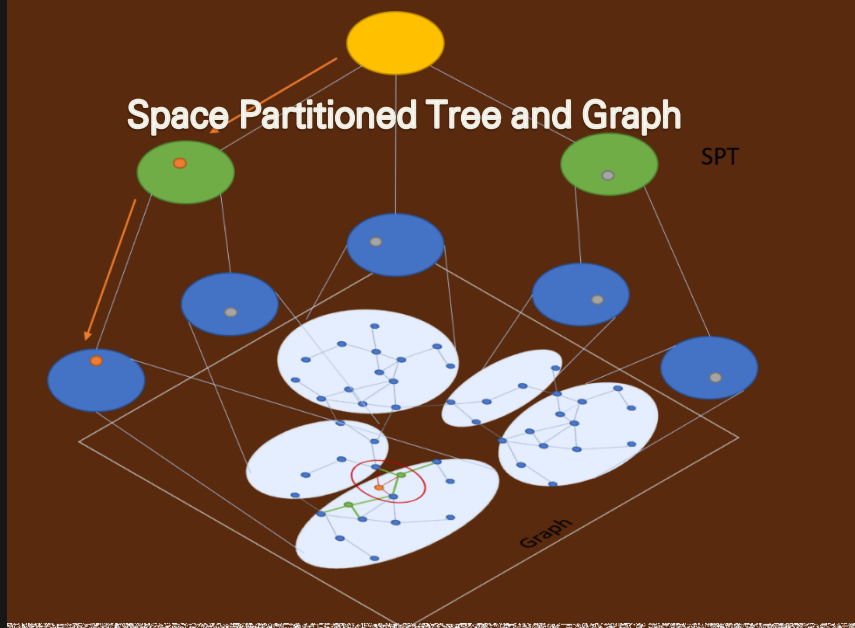
Search Results for ['tall', 'tower', 'paris']

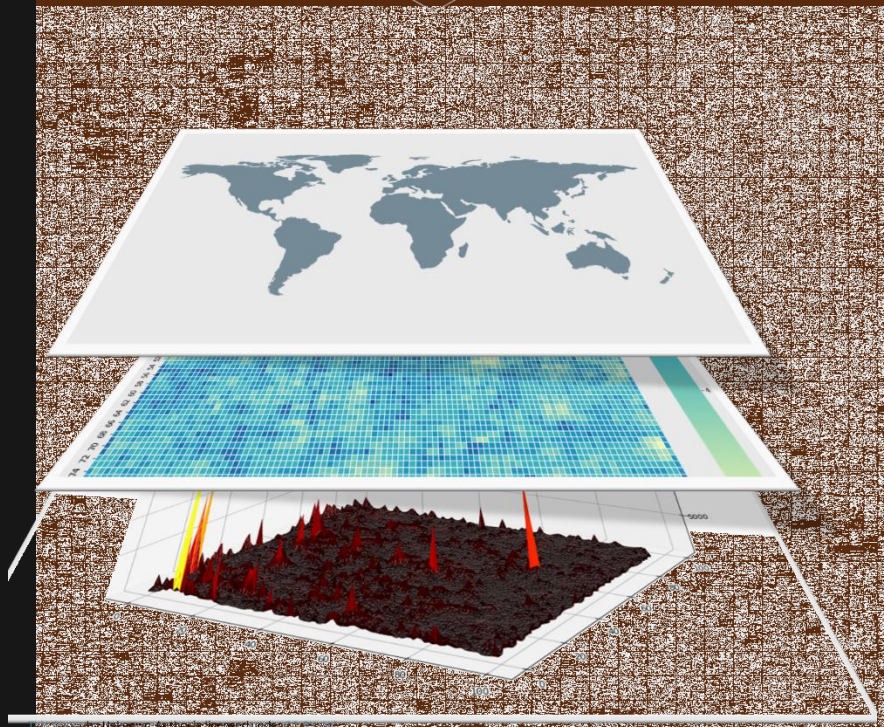
	Term	TermID	vector	bmu	sem_vector
0	tall	00000000000000006C463BDD3122598	[0.17543, -0.2003, -0.1[illegible]3409, 0.02893...	[illegible]	eca3333e711b4dbec3f011be728aaebe9bfeec3cc190c5...
1	tower	000000000000000010002370301416 52	[0.17543, [illegible] -0.[illegible]252, -0.3409, 0.02[illegible]	[illegible]	eca3333e711b4dbec3f011be728aaebe9bfeec3cc190c5...
2	paris	0000000000000000C280024800001008	[0.[illegible]13, -0.[illegible]4228, -0.23725, -0.[illegible]	86/08	624ab43eda0498bdaaf172be9f8ea7be410e0a3fb16dd1...

00000000000000006C463BDD3122598 00000000010002370[illegible] 0000000000000000C280024800001008

	docid	ann_bmu	semantic_title_vector_norm	url	title	textptr	recptr	x	y	dist
0	D2716375	{'x': 12, 'y': 31}	{'Row': [{'val': -0.00740[illegible]5069[illegible]}, {'val...	http://www.independent.co.uk/arts-entertainmen...	Mile-high tower wars: How tall is too tall?	328581892	5874504330	12	31	23.557430
[illegible]	D2224270	{'x': 13, 'y': 33}	{'Row': [{'val': -0.01012962590903044}, {'val'...	https://en.wikipedia.org/wiki/List_of_tallest_...	List of tallest towers	330888585	1396140120	13	33	22.297666
[illegible]			{'Row': [{'val':		List of					

JPG File





[GitHub - microsoft/SPTAG: A distributed approximate nearest neighborhood search (ANN) library which provides a high quality vector index build, search and distributed online serving toolkits for large scale vector search scenario](https://github.com/microsoft/SPTAG)

TAKEAWAYS

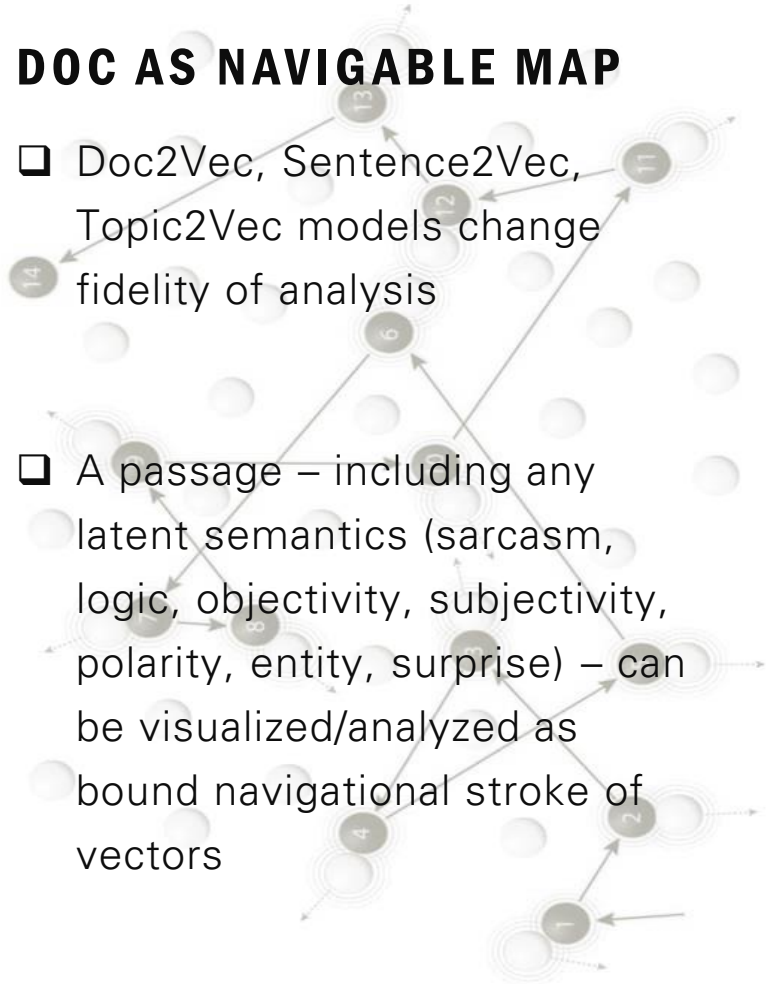
SPACEV SHARDING

- Distribute index by semantic/spatial partition ID – not by terms or docs that do not have any monotonicity
- The partitioned space clustering can be computed offline and inferenced for queries online performantly
- SPTAG – "Space" partitioned tree and graph search: make that "space" count

INCEPTION OF ANN SCORE

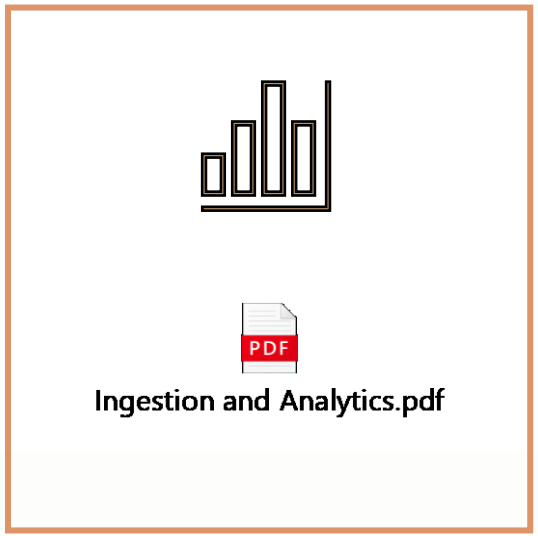
- High-grained ANN quantization into a 100x100 grid allows exocentric indexing/selection of content
- 100x100 (or even fine-grained 10000x10000) similarity matrix can be pre-computed and cached upfront so cosine similarities do not have to be constantly evaluated at runtime

DOC AS NAVIGABLE MAP



- Doc2Vec, Sentence2Vec, Topic2Vec models change fidelity of analysis
- A passage – including any latent semantics (sarcasm, logic, objectivity, subjectivity, polarity, entity, surprise) – can be visualized/analyzed as bound navigational stroke of vectors

PATH TRAVELED...

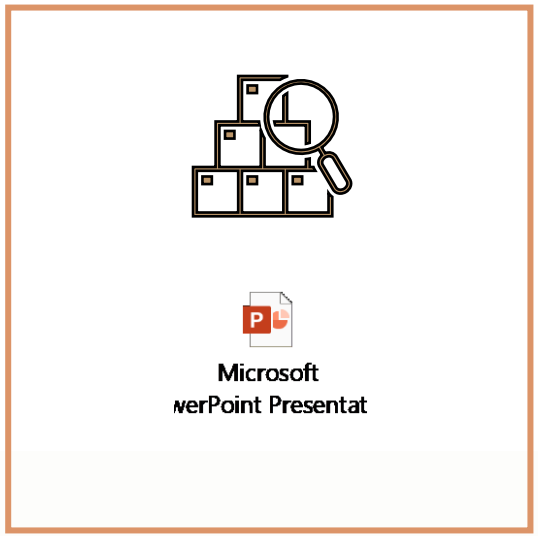


DATALAKE INTEGRATION

36 data sources for shared PoV

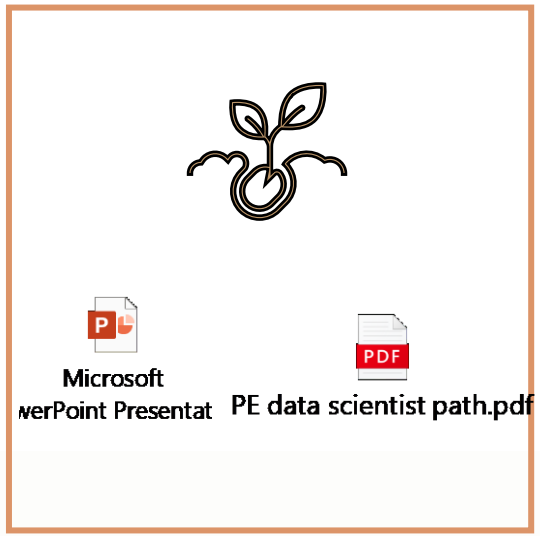
DATA WAREHOUSING

Out with the old, in with the new



DEFECT PROVENANCE

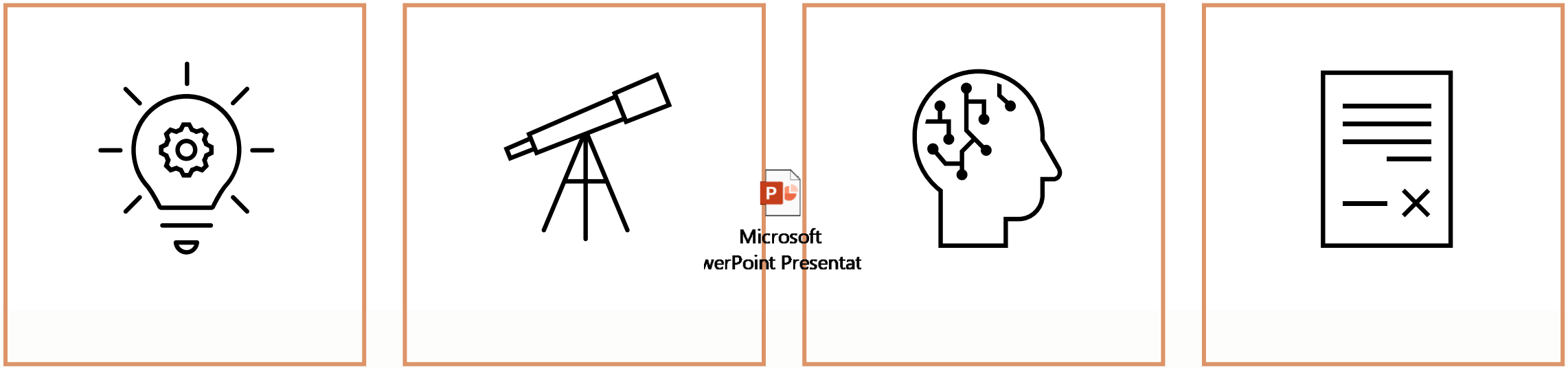
Improving yield with in-situ preemptive signaling



INVEST IN HUMAN ELEMENT

Aligning purpose with passion and prowess

PATH TRAVELED...



IDEATION
Scanning the industry and customer direction

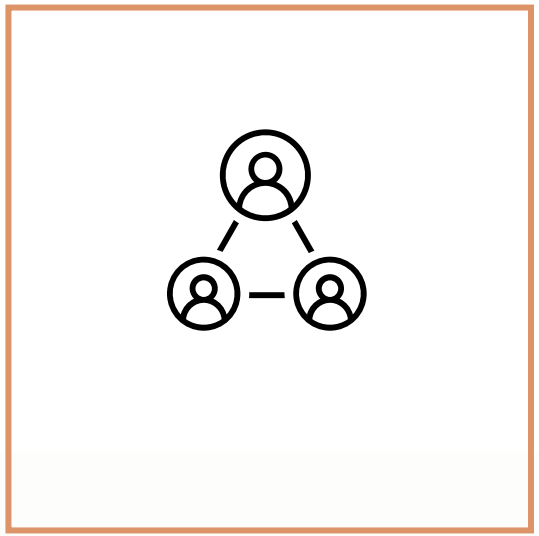
DATA OFFICE
Defining and executing BI/MI/AI strategy

AI ENGINEERING
Prove, nail, and scale science into solutions

CO-INNOVATION
Protecting and leading travel industry IP

PATH TRAVELED...

SEEDLE
Crowd-pledging reforestation platform



HBBL
News/Gossip feeds portal

SEATDASH
In-stadium food dash/delivery

CHWEF
Home food in your neighborhood; C2C Mobile Commerce