databricks

Any-Scale Spatial Analysis on Databricks

Michael Johns | Lead Geospatial Product Specialist

Agenda

1. Democratizing Data + AI
2. Spatial on Databricks
3. DGGS “Hybrid” Approaches
4. Unified Data Access + Sharing
5. What’s Next?

1

Democratizing Data + AI

databricks

The data and AI company

Gartner Magic Quadrant — LEADER 2023 Cloud Database Management Systems

Gartner Magic Quadrant — LEADER 2024 Data Science & Machine Learning

FORRESTER WAVE LEADER 2023 Cloud Data Pipelines

FORRESTER WAVE LEADER 2024 Data Lakehouses

FORRESTER WAVE LEADER 2024 AI Foundation Models For Language

IDC ANALYZE THE FUTURE LEADER Analytic Stream Processing

10,000+ global customers

$2.4B+ in annual revenue

4B+ in investment

Inventor of the **lakehouse** and pioneer of **generative AI**

DELTA LAKE

Unity Catalog

APACHE Spark

mlflow

DBRX

Every company wants to be a

Data + AI

company

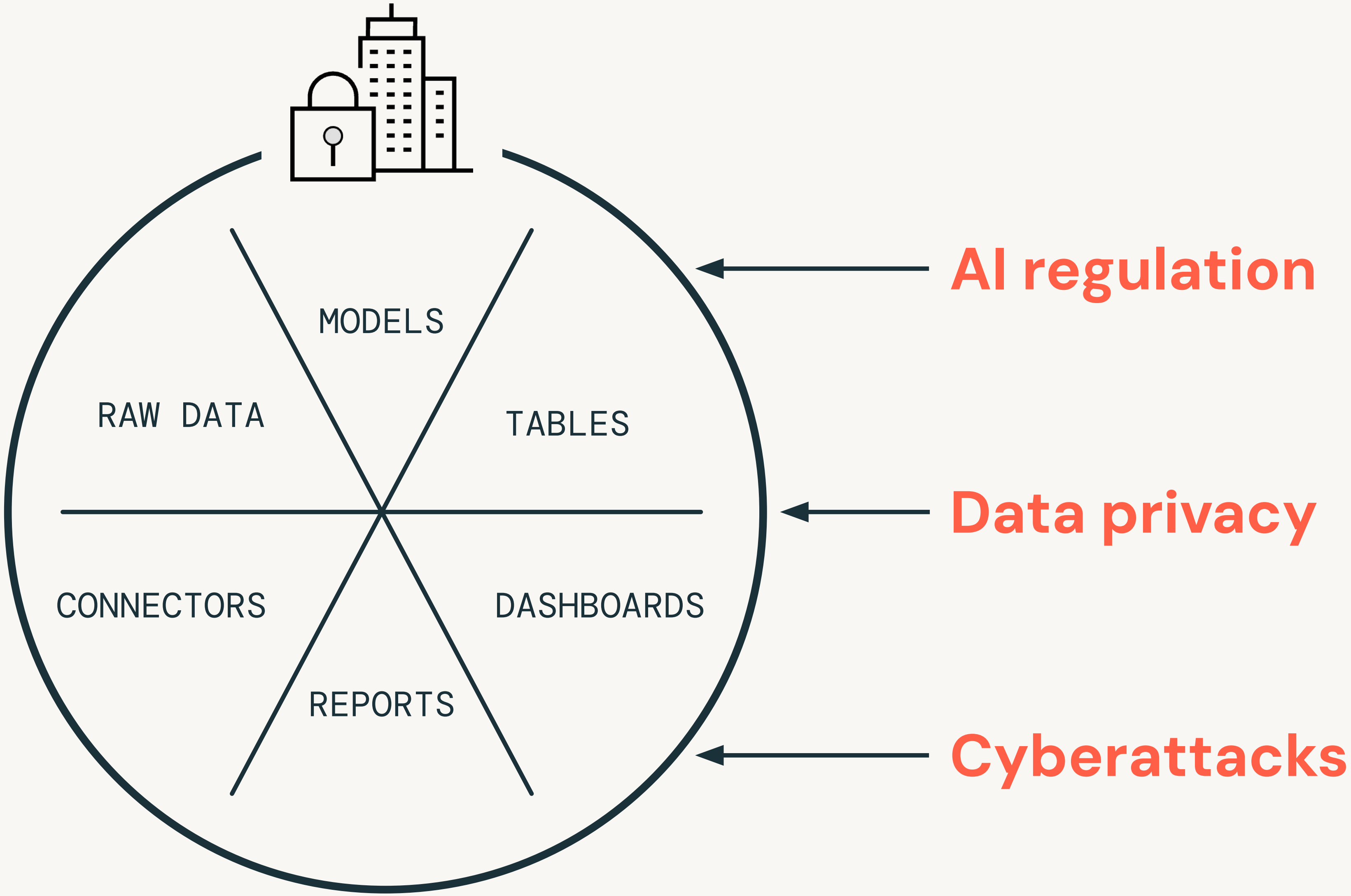
Data estate is fragmented

It's a complexity nightmare of high costs and lock-in

Data estate is fragmented

It's a complexity nightmare of high costs and lock-in

Governance of the entire data estate is hard



Own your data

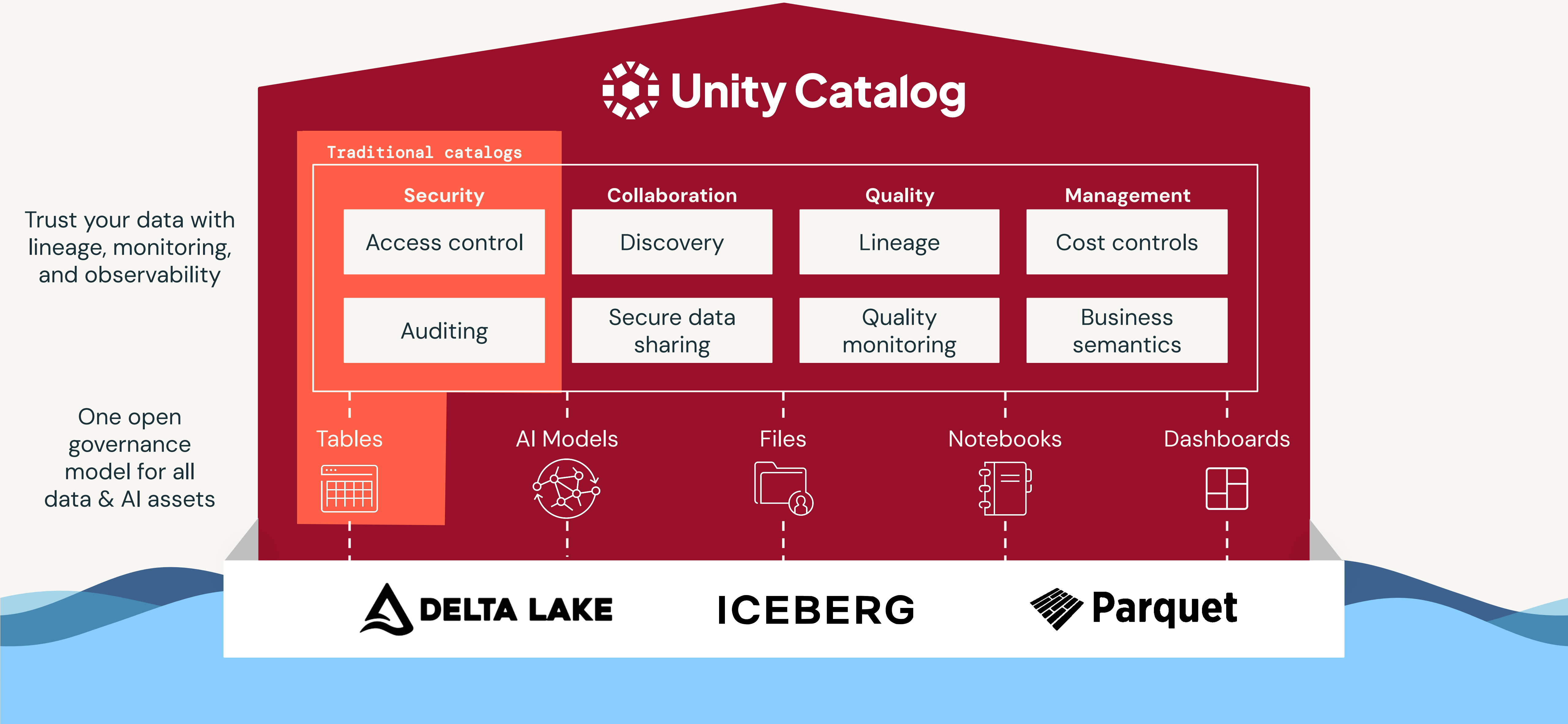
Eliminate silos

Read and write any data in any open format with full interoperability

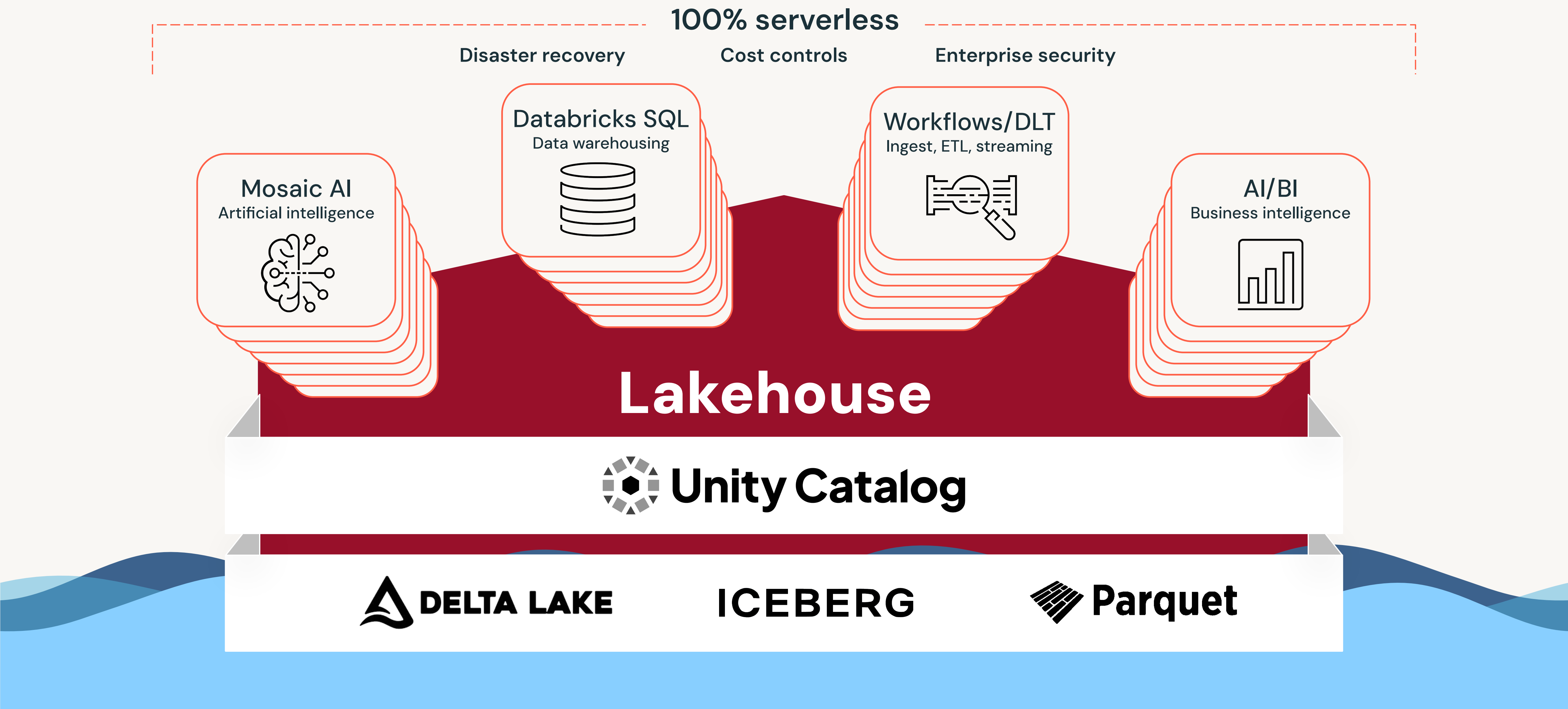
Eliminate unnecessary costs from multiple copies of data

DELTA LAKE | ICEBERG | Parquet

Unify governance for all data + AI



Databricks Data Intelligence Platform



databricks

Data Intelligence Platform

Lakehous
e

Unified data and governance

AI

AI tuned to your business

Production AI that reasons on your data is elusive

Experimentation

Quality

Cost

Privacy

Production

85% have not made it into production

Creating the Data Intelligence Platform

◆ WSJ NEWS EXCLUSIVE | CIO JOURNAL

Databricks Strikes $1.3 Billion Deal for Generative AI Startup MosaicML

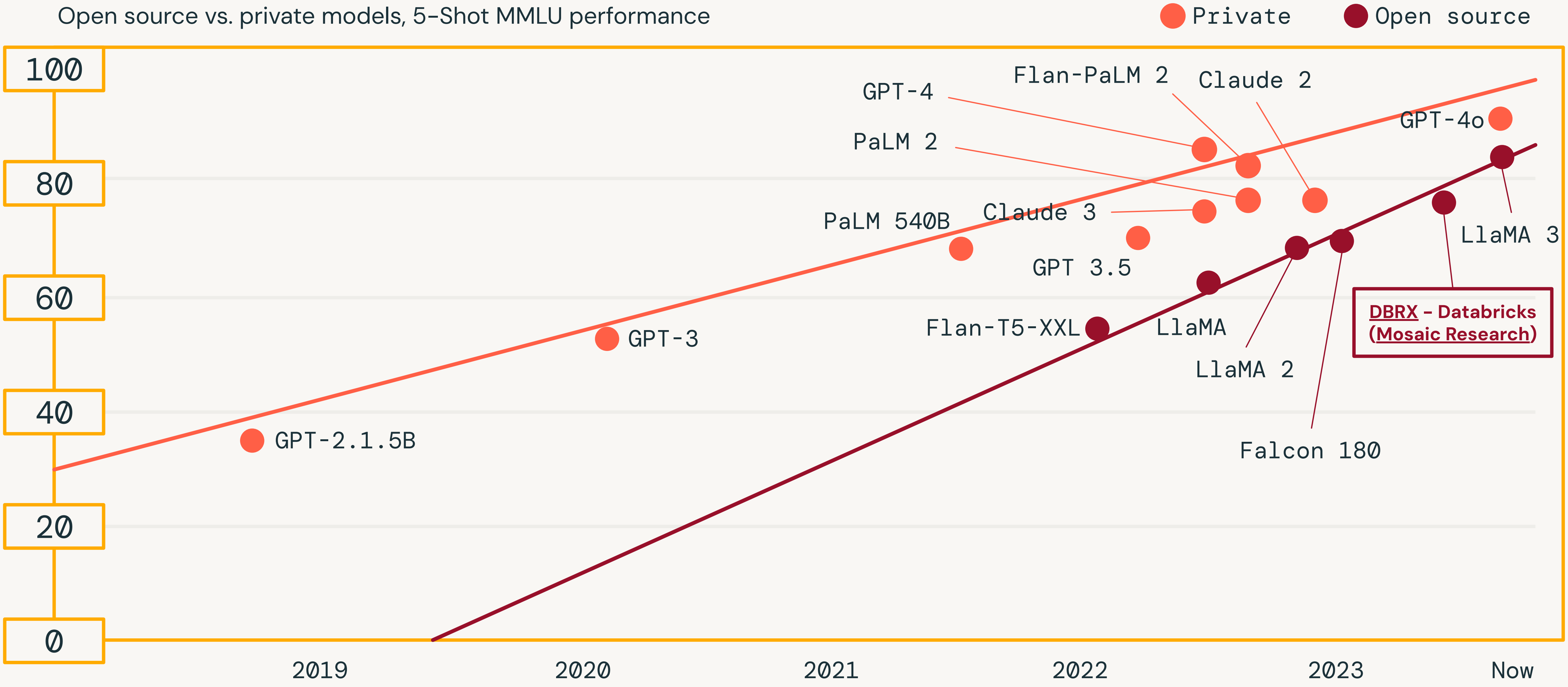
The deal aims at connecting businesses' data with services to help them build their own, cheaper language models, Databricks CEO says

By Angus Loten and Belle Lin

June 26, 2023 at 7:40 am ET

LLMs maxing out on general intelligence tests

Open source vs. private models, 5-Shot MMLU performance



General intelligence

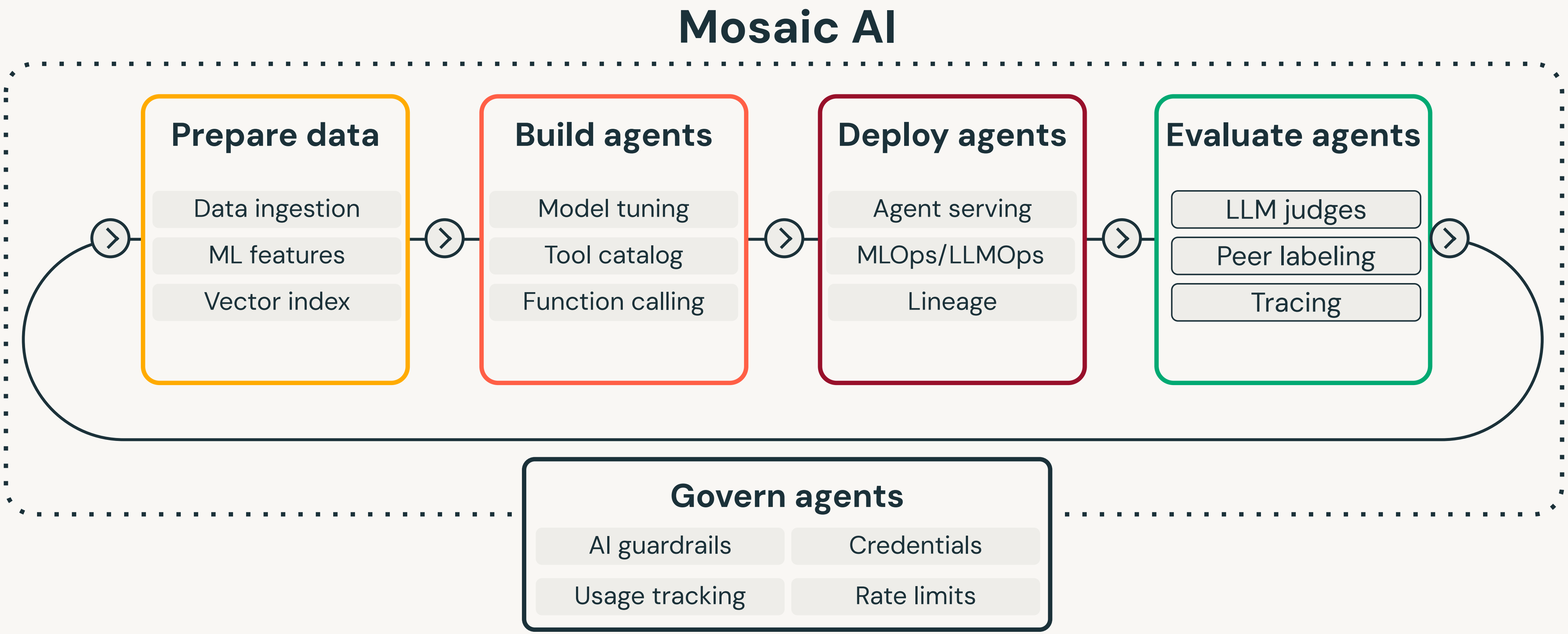
Consumer models trained on a broad dataset **disconnected** from your business data

VS

Data intelligence

AI **connected** to your customer data and able to solve domain-specific problems

Build agent systems with Databricks



How can Databricks help?

Get better answers from your data

Consolidate data, AI and governance silos with a unified platform, democratize data and allow everyone to get valuable insights rapidly.

Deliver generative AI

Easily combine generative AI models with your unique data to deliver a sustainable competitive advantage.

Optimize costs and improve ROI

The Databricks Data Intelligence Platform helps customers realize an average ROI of 482% with a payback period of as few as 4 months.

Minimize governance and security risks

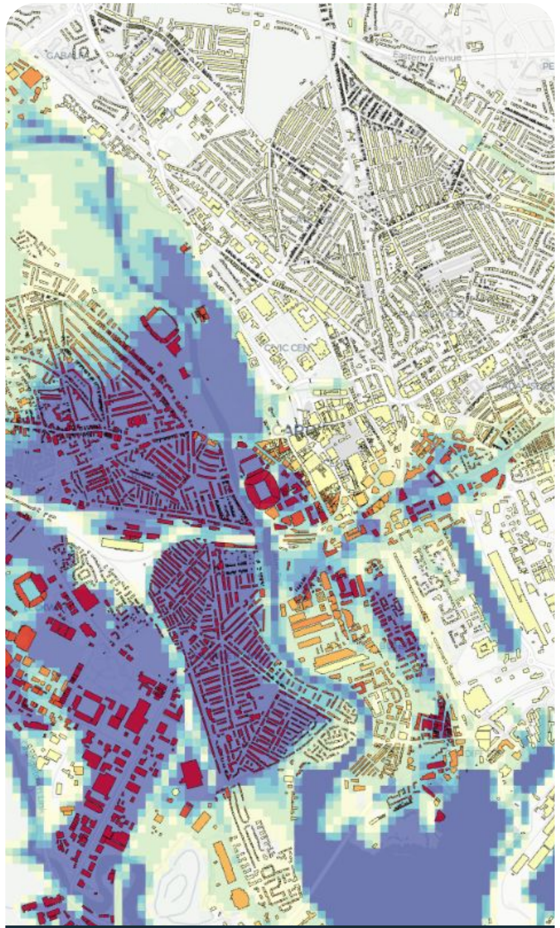
Apply consistent policies across all your data and workloads with fine-grained insight into data lineage so that you know where data came from, where it's going and how it's being used, from BI to AI.

2

Spatial on Databricks

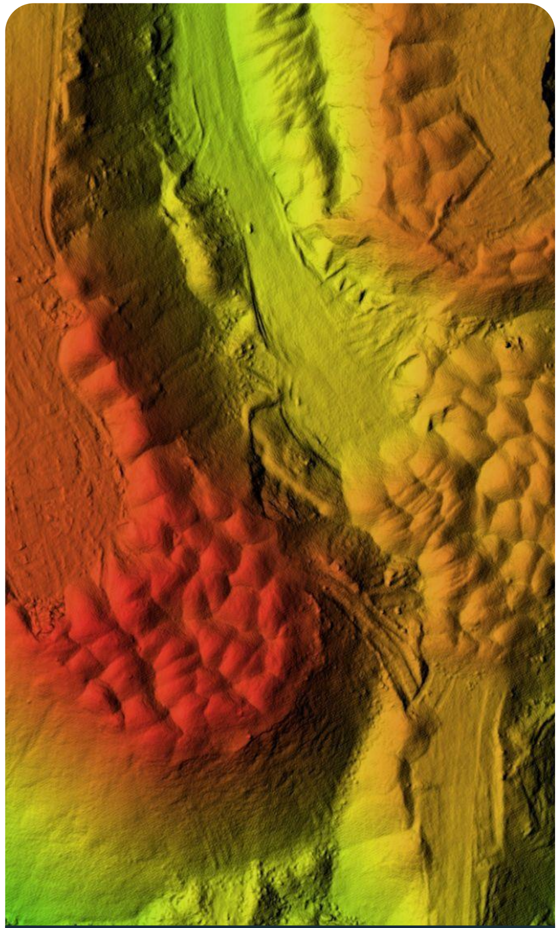
A	B
latitude	longitude
41.12887998	-73.7100635
41.13365032	-73.7099659
41.13009945	-73.70991433
41.13018385	-73.70989947
41.13043715	-73.70985008
41.1309958	-73.70976019
41.13354633	-73.70968264
41.13453456	-73.71234567
41.13987875	-73.71245671
41.13987988	-73.71267892
41.13987991	-73.714687912
41.14567889	-73.714522467

Anything with Lat/Long coordinates



Raster

Vector

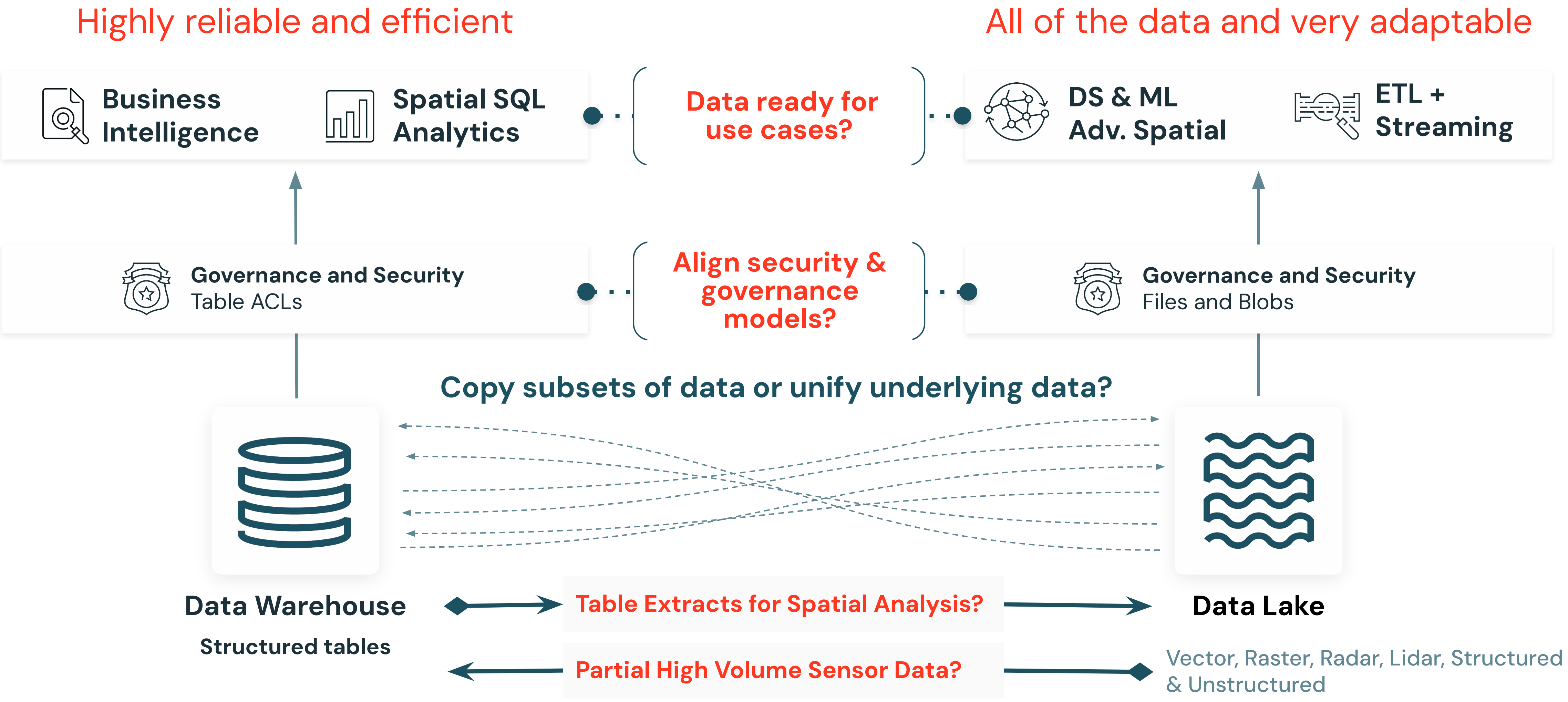


Lidar

Geospatial data is everywhere...

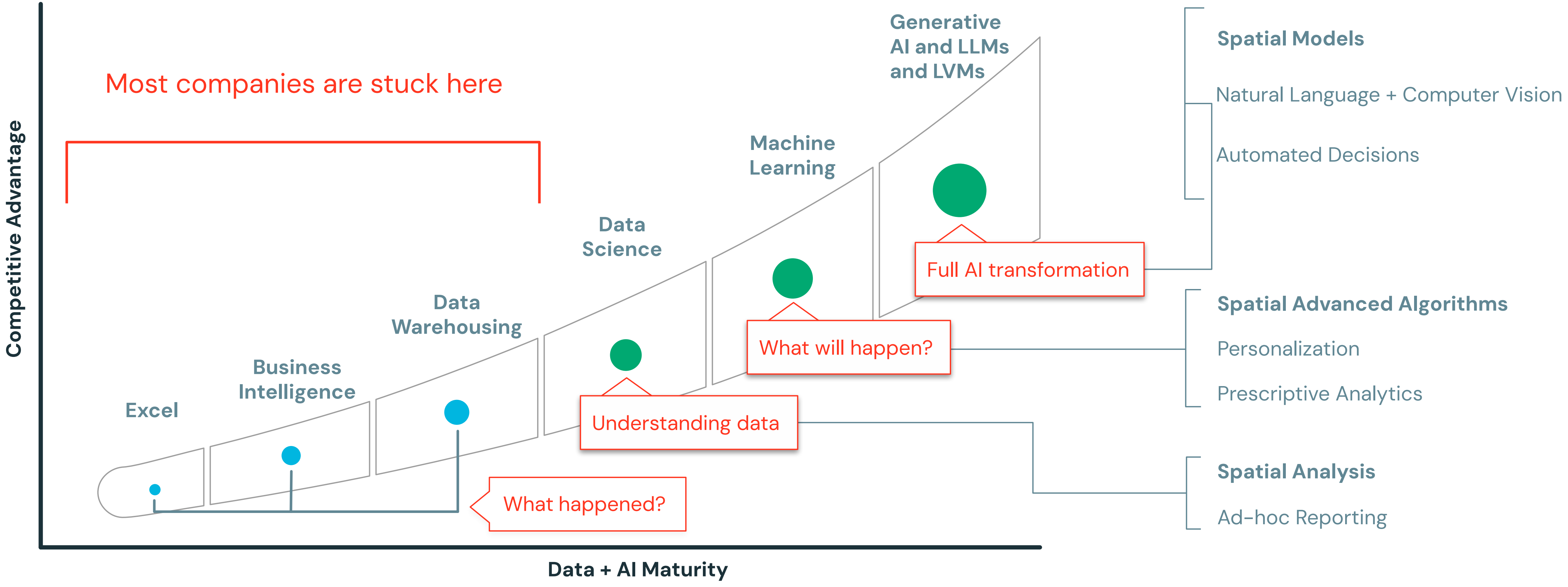
Lakehouse: Data Warehouse + Data Lake

For spatial workloads, how to best bridge the gap between?



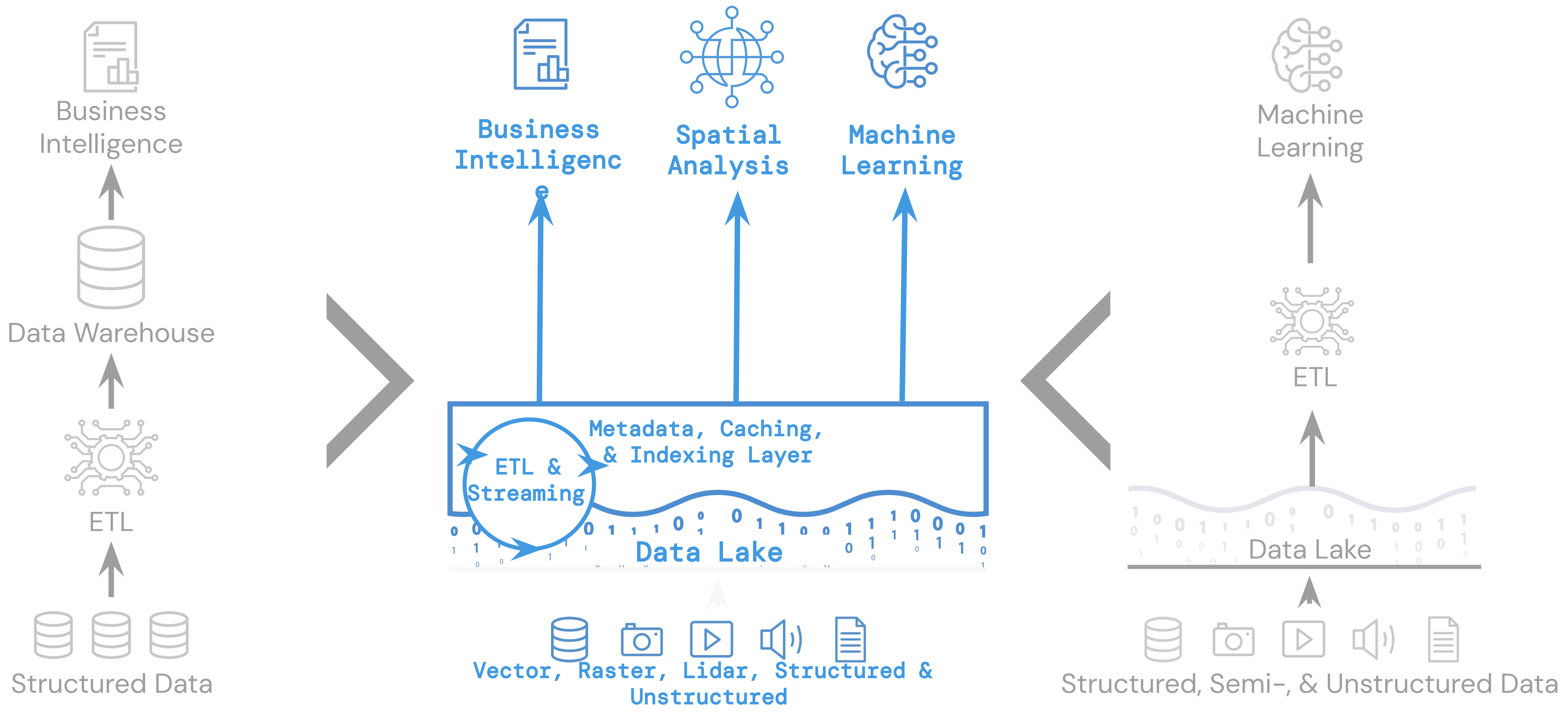
Data + AI maturity

Business + Technology driving organizational changes

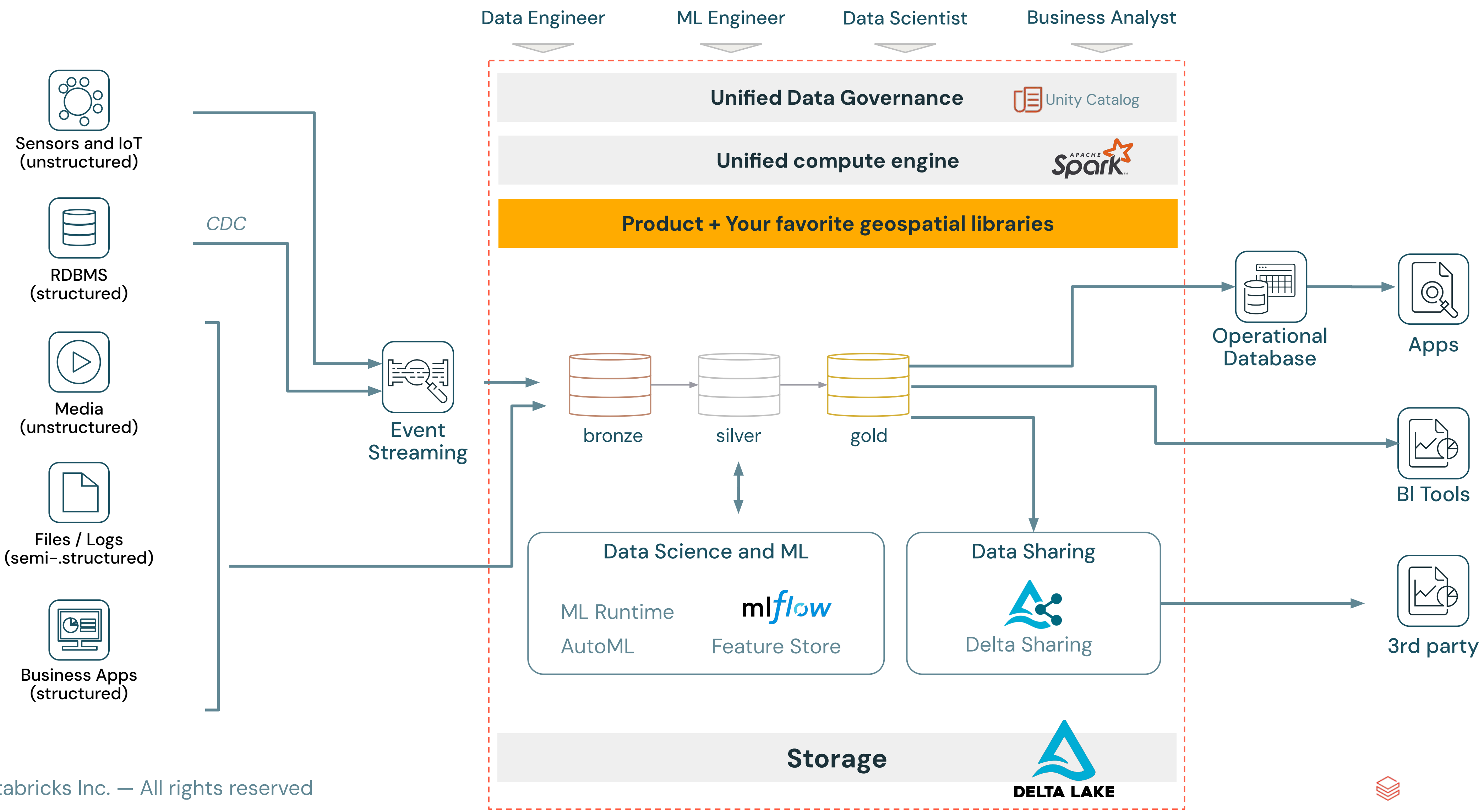


Why Geospatial on the Lakehouse?

Build a data asset strategy to enable multiple use cases



Geospatial + Lakehouse



Geospatial - Current State

1000+ Customers use Databricks for Geo

Built-in Geospatial Functions

- 30+ H3 APIs
- 60+ Spatial SQL [Preview]
- Initial GeoHash [Preview]
- Available Everywhere: Photon + Notebooks + DBSQL

Flexible Platform

- Open Source libraries
- Commercial libraries + Partners
- Vector, Raster, Lidar, +

Geospatial - Current State

Distributed Performance Journey

Single Node Libraries	**Spark Frameworks**	**Product Features**
e.g. GDAL, GeoPandas, RasterIO	e.g. Analytics Toolbox, Mosaic, Sedona, Esri GAE	H3 + Spatial SQL
(1) Can use only on Driver or Single Node Cluster (2) Can apply UDF patterns to distribute (3) Libraries can be complemented by use of Mosaic or avoided altogether	(1) Mosaic is aligned with Databricks Product, data engineering unified (2) Mosaic offers Spark APIs powered by GDAL + JTS (3) Works with other frameworks	(1) The fastest H3 API implementation through Photon Engine, also used by Mosaic (2) Support for Spatial SQL + in-memory geospatial types [Preview]
+ **Flexible, precise patterns** - Relatively slower at scale with UDFs	+ **Distributed, precise patterns** + **Relatively faster at scale through Spark optimizations (Mosaic best for Databricks)** - Each have tradeoffs in indexing decisions	+ **H3: Very fast at scale** + Spatial SQL collecting feedback from customers

Geospatial Reference Architecture

GeoJSON
Shapefile
GeoDB
OSM
GTiff
COG
CSV
Parquet
...

0101

ACID + Statistics + Indexing + Compaction + Columnar

Raw

- Historical
- Minimal ETL
- Initial Table Write [Delta Lake / UniForm]

Refined

- Validate
- SRID / Projection
- Base Geom Tables [Bounds Index]

Enriched

- "Hybrid" DE Pattern
- Compute Aggregates
- Global Grid Tables [H3 Index]

Analytics

- Spatial SQL APIs
- Global Grid APIs
- Non-Geo Blended
- Advanced Spatial
- AI/ML Models

CARTO
tableau
esri
kepler.gl

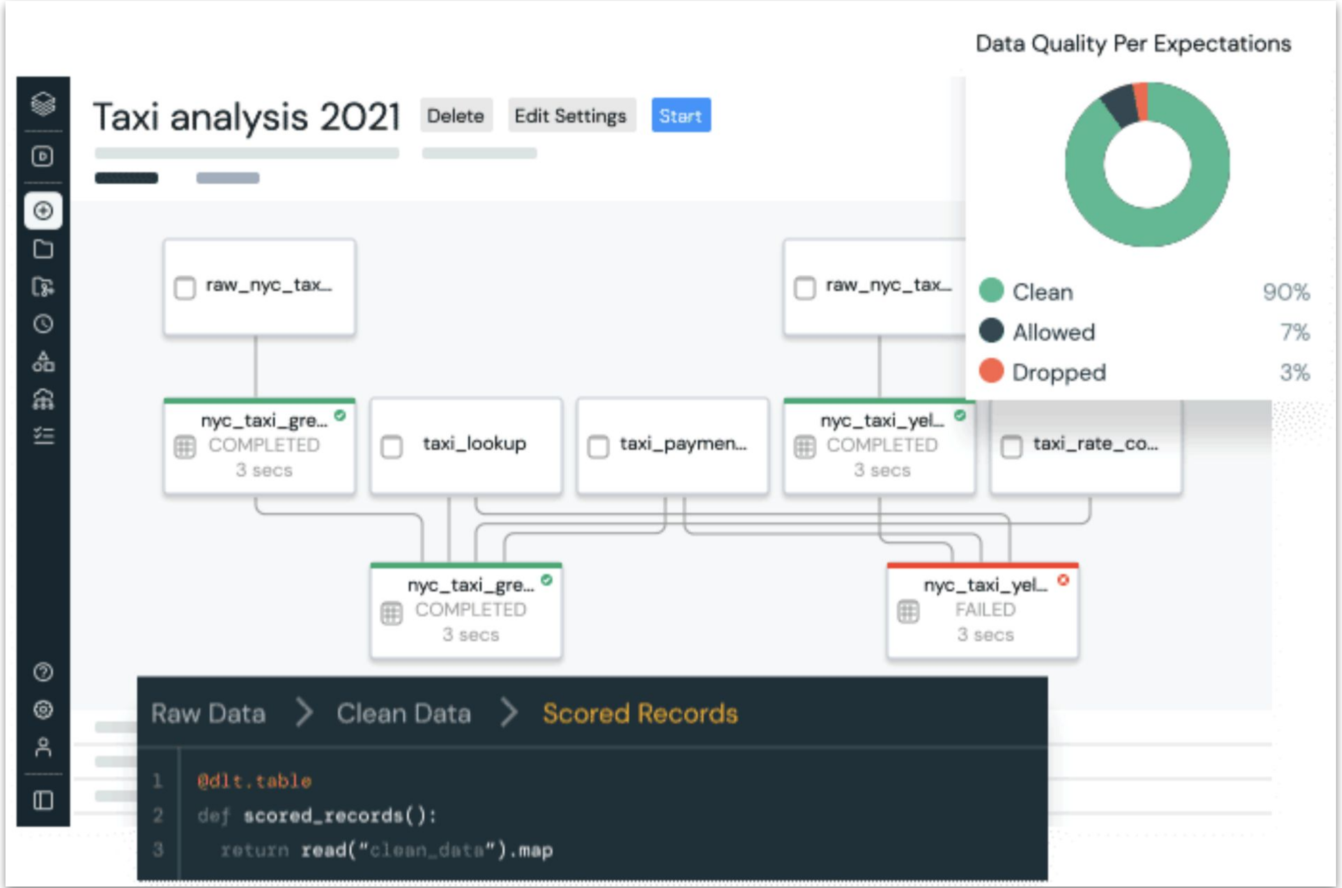
- Built-in readers [JSON, CSV, XML, Parquet]
- Various Library options to read formats
- Use of DataFrame [Spark / Pandas] APIs to abstract initial format

- Write to Delta Lake / UniForm + Liquid Clustering
- Data Engineering with "Medallion" Architecture
- Use of built-in APIs [H3, Spatial SQL] + Libraries

Spatial Data Engineering

Orchestrating production grade processing pipelines

- Data orchestration through Databricks Workflows
- Delta Live Tables manage your full data pipelines
- Simplifies data engineering with a curated data lake approach through Delta Lake



Databricks Solutions and more

Get your Geospatial Lakehouse journey started

Spatial Analytics at Any Scale With H3 and Photon

A Comparison of Discrete, Vector, and Hybrid Approaches

by Kent Marten, Michael Johns and Menelaos Karavelas

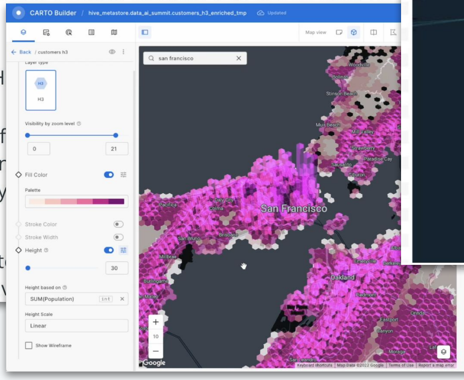
December 7, 2022 in Engineering Blog

Using H3 in Databricks with CARTO

Thanks to our Spatial Extension for Databricks, CARTO users can connect directly to their Databricks cluster to access data and perform massive-scale data visualization and analytics. The latest enhancements to the Databricks platform brings added H3 functionality to allow dynamic aggregation natively within Databricks. And in addition, our spatial data catalog opens up a wealth of H3-indexed datasets that can be used for highly efficient data enrichment workflows.

This Databricks release includes 28 native H

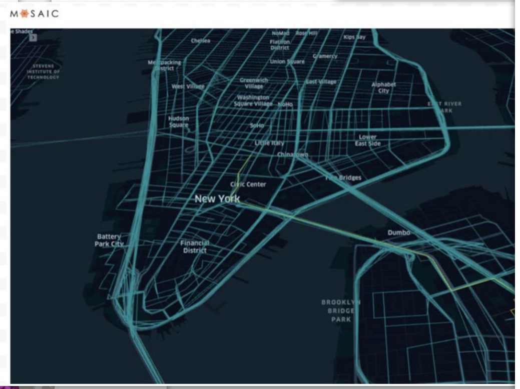
- Functions to generate H3 cells and grids f such as **h3_polyfillash3()** where users car to cover the extent of a polygon. Similarly a H3 cell at a defined coordinate.
- The reverse of this; functions which creat include **h3_boundaryaswkt()** - which conv

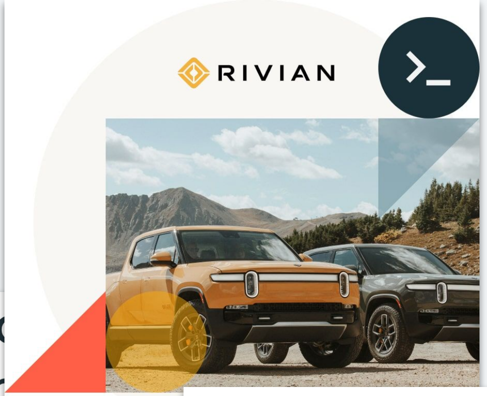


Scalable R
With Datab

by Rohit Nijhawan, Bryan

September 2, 2022 in Pr





Industrial AI: Increase Crop Yields

Autonomous and location-driven agriculture

Thasos: Customer Success

Overcoming scaling challenges using H3

Industrial AI: Increase Crop Yields

Tons of sensors, vision + guidance systems and wireless connectivity

- Trillions of records / several petabytes
- Dozens of layers of hundreds millions of unique acres
- Hundreds of ML features created and automatically maintained

"Cameras continually monitor images … We use machine learning to analyze grain quality and automatically adjust the operating parameters of the machine if any damage is detected to the grains."

Blog | Keynote from DAIS

Use Cases: Image + Raster Workloads

Geospatial cuts across every industry, blended with other sources

Oil & Gas

Pipeline predictive maintenance + monitoring, change detects

Agriculture

Crop yields, weed detects, climate

Autonomous Vehicles

Object detects, collision avoidance, navigation; also, offroad, e.g. self-driving large mining equipment

Financial Services

Cars in parking lots, ships in ports, construction, crop yields, natural catastrophes, fires, tracking renewable transition

National Security & Response

Orchestrating SAR image formation, pattern of life, situational awareness, manmade + natural disasters, warfighting

Image Segmentation with Databricks

From ingestion to prediction

End-to-end pipeline…

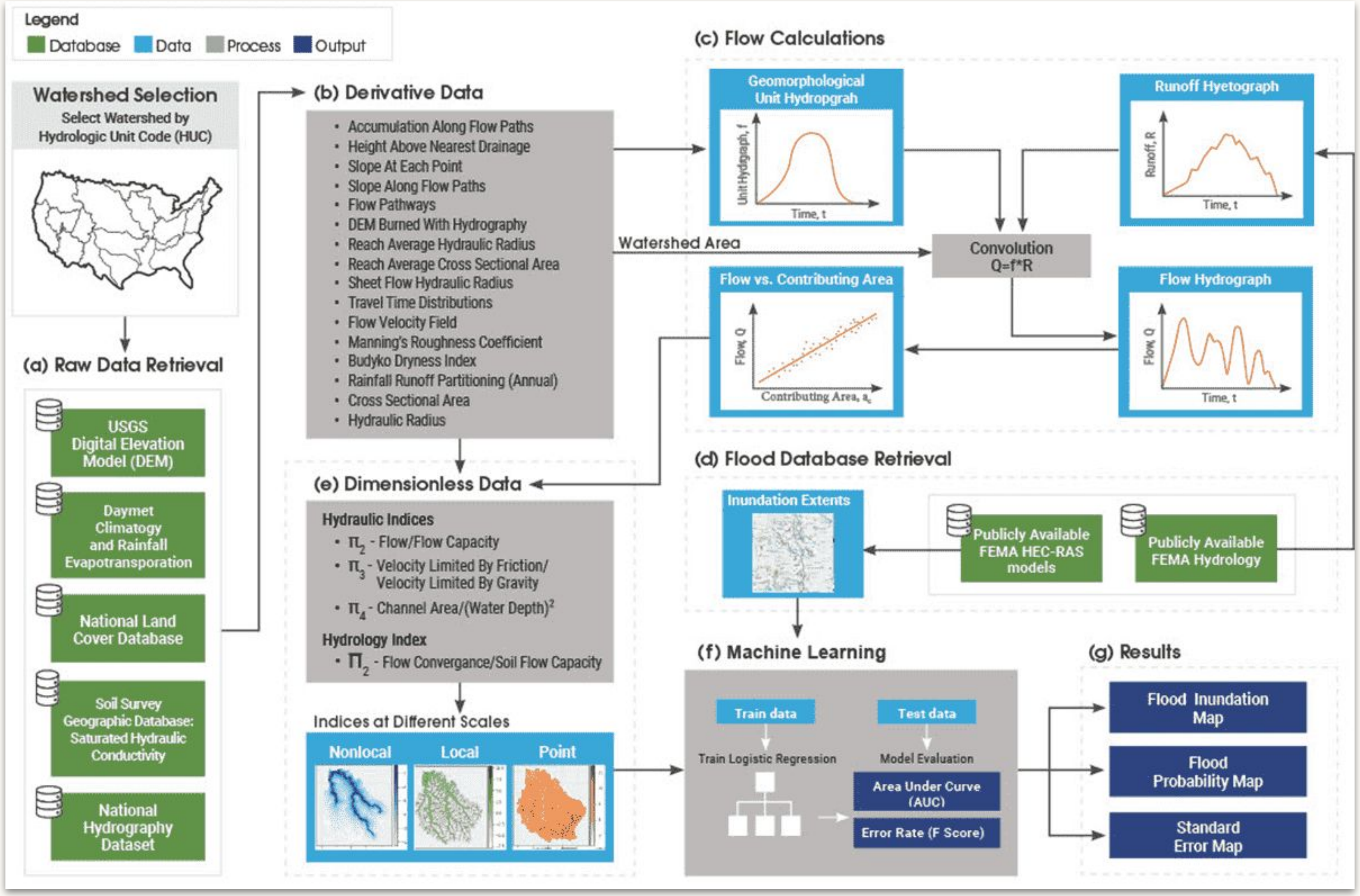
- Incrementally ingest datasets
- Clean, standardize, and join datasets
- Train Deep Learning model
- Deploy model for production

Blog Link

Stantec + Databricks: Flood Predictor

Secure, cloud-based solution to predict where and when flooding events will happen.

1. Ingests various open datasets
2. Performs geospatial computations
3. Publishes high-quality features to Feature Store
4. Spatial ML model trained on features



Blog | Stantec

DGGS "Hybrid" Approaches

3

Scalable Geospatial Analytics with H3

Supported natively in Databricks

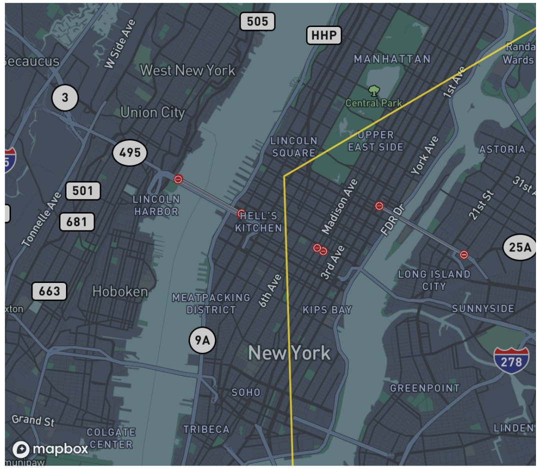
Grid indexing systems are ideally suited for scale

Hierarchical system offers flexibility

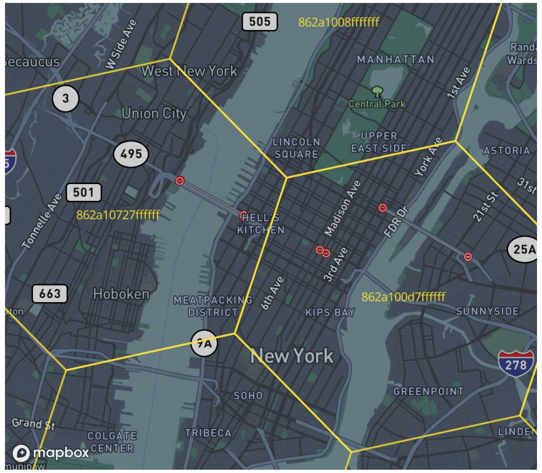
Easy and effective visualization

Precision Agriculture	Autonomous Vehicles	Retail Operations	Climate Risk/Modeling	Human Mobility
				

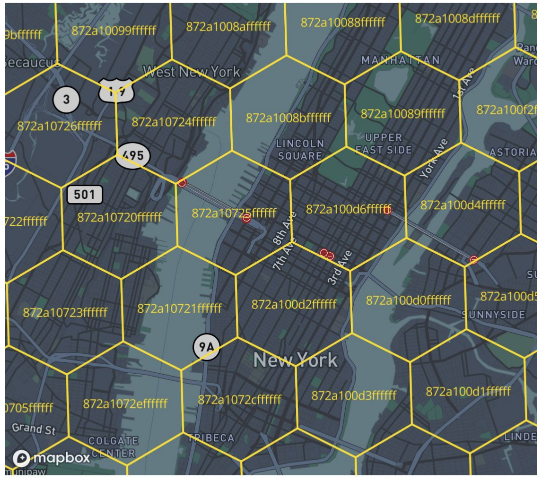
H3 Resolutions



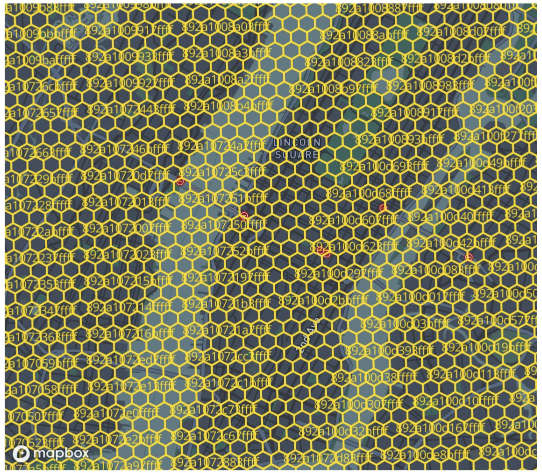
resolution = 5



resolution = 6



resolution = 7


resolution = 9

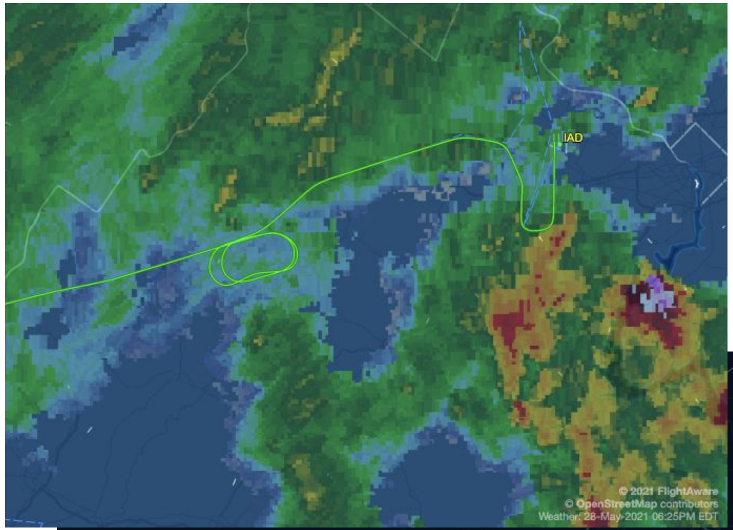
H3 Resolution	Average Hexagon Area (km^2)	Average Hexagon Edge Length (km)	Number of unique indexes
0	4,250,546.8477000	1,107.712591000	122
1	607,220.9782429	418.676005500	842
2	86,745.8540347	158.244655800	5,882
3	12,392.2648621	59.810857940	41,162
4	1,770.3235517	22.606379400	288,122
5	252.9033645	8.544408276	2,016,842
6	36.1290521	3.229482772	14,117,882
7	5.1612932	1.220629759	98,825,162
8	0.7373276	0.461354684	691,776,122
9	0.1053325	0.174375668	4,842,432,842
10	0.0150475	0.065907807	33,897,029,882
11	0.0021496	0.024910561	237,279,209,162
12	0.0003071	0.009415526	1,660,954,464,122
13	0.0000439	0.003559893	11,626,681,248,842
14	0.0000063	0.001348575	81,386,768,741,882
15	0.0000009	0.000509713	569,707,381,193,162

https://uber.github.io/h3/#/documentation/core-library/resolution-table

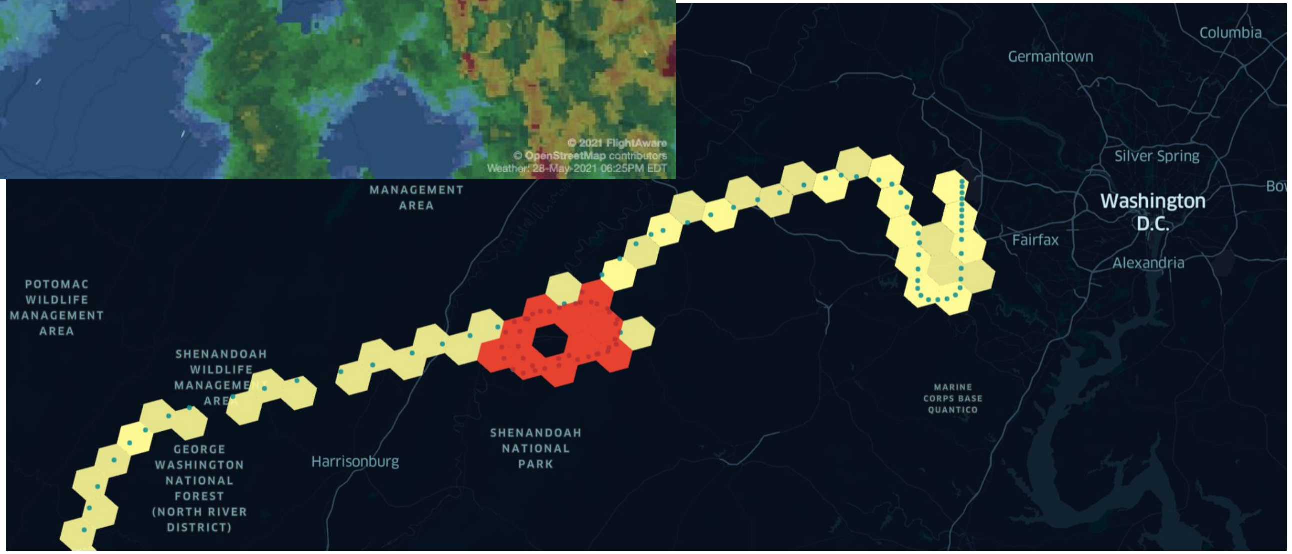
Res 15 is < $1m^2$ or ~ space of 1 person

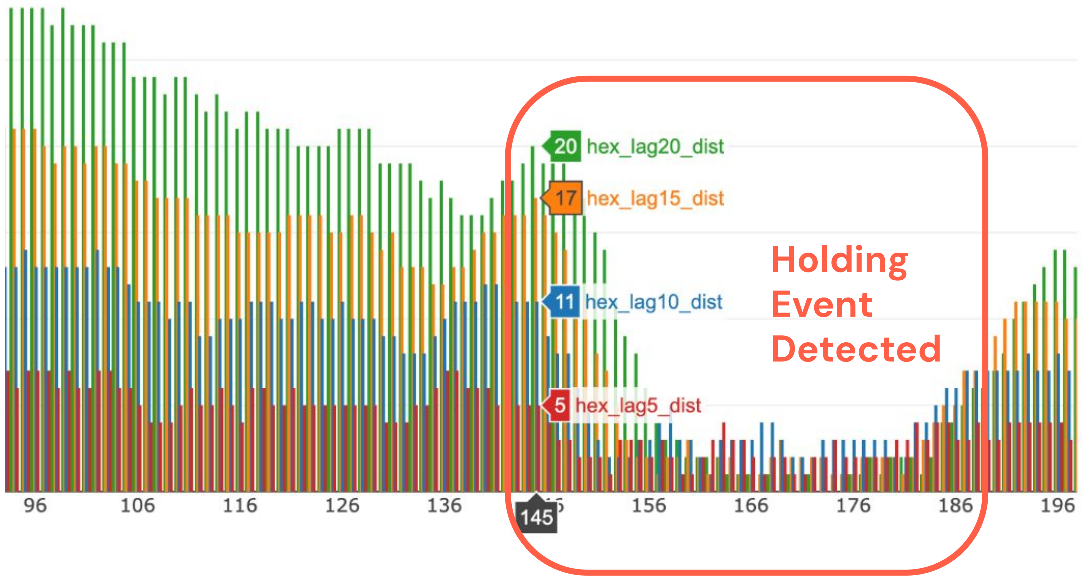
Example: Detect Flight Holding Patterns w/H3

Answer a number of spatial queries without Spatial SQL




```
df_hex6_dense = (
    df_hex6
        .filter("mph > 275")
        .groupBy("hex_id")
        .count()
        .filter("count > 3")
        .orderBy(F.desc("count"))
)
```





Combine with other lakehouse functions, e.g. lagging windows using **H3_Distance** .

Link to Notebook [Docs]

Creating safer roadways for U.S. drivers

Texas A&M Transportation Institute unlocks car sensor data on Databricks

100s

Of terabytes of spatial data unified for better roadways

Faster

Time-to-insights to support optimized transportation safety

- 150 terabytes or 1.2 trillion GPS points
- Seamless integrations with Power BI & Tableau
- Databricks H3 indexing for geospatial models

"The Databricks Data Intelligence Platform helps us easily ingest and collaborate on large datasets with trillions of GPS points. Now that we're effectively visualizing and analyzing connected vehicle and geospatial data, we can continue innovating on transportation safety without limitations."

Customer Story

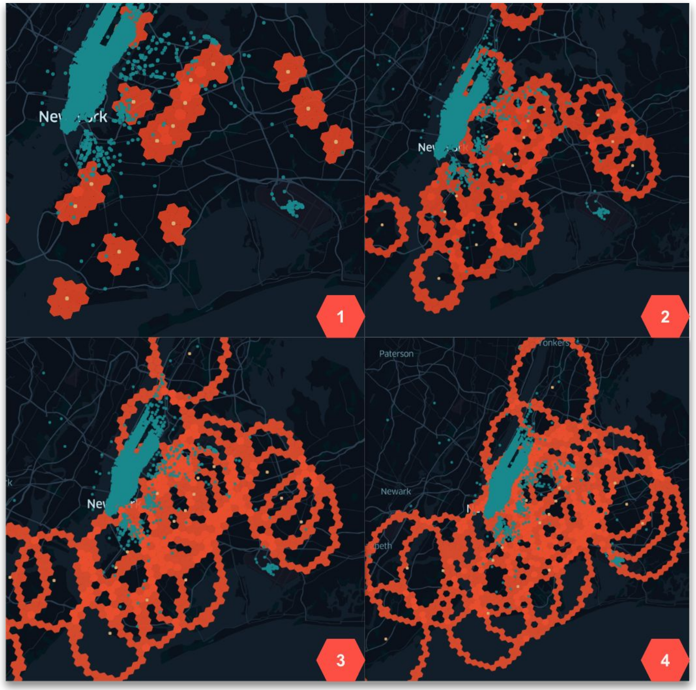
Solution Enabler: Scaled KNN

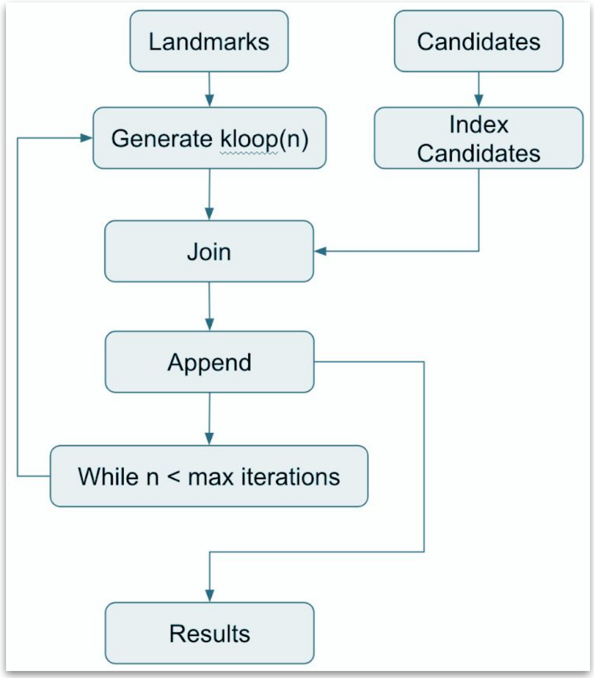
Spatial K-Nearest-Neighbors model w/ mlflow integration

Use cases covered in the blog:

1. Location-based advertising, personalized based on purchase history
2. Customer segmentation, based on location and similarities
3. Retail, e.g. store location
4. Real Estate, e.g. property searches
5. Traffic, e.g. roads with highest risk for car crash

I need nearest thing(s) of interest for up to every row of my data type of problems…





KNN Docs | Notebook | Blog

Vector data processing approaches

Pure vector	Hybrid (Tessellation)	Pure grid-based
Used in “classic” GIS processing. Good for OLTP and small datasets	“Cuts” the original polygon on grid cell boundaries and adds cell IDs.	Converts polygons & points to grid cell IDs, then operates only on cell IDs.
Databricks support is provided by third party libraries e.g. Geomesa, Sedona, and Mosaic.	Databricks provides **native support for H3 grid**. Vector support is provided by **Mosaic.**	Databricks provides **native support for H3 grid**. Other grids are worked on in **Mosaic**
+ Maintains full detail - Slow at scale	+ Maintains full detail [through chips] + Fast at scale - Requires tessellation	+ Very fast at scale - Loses some detail

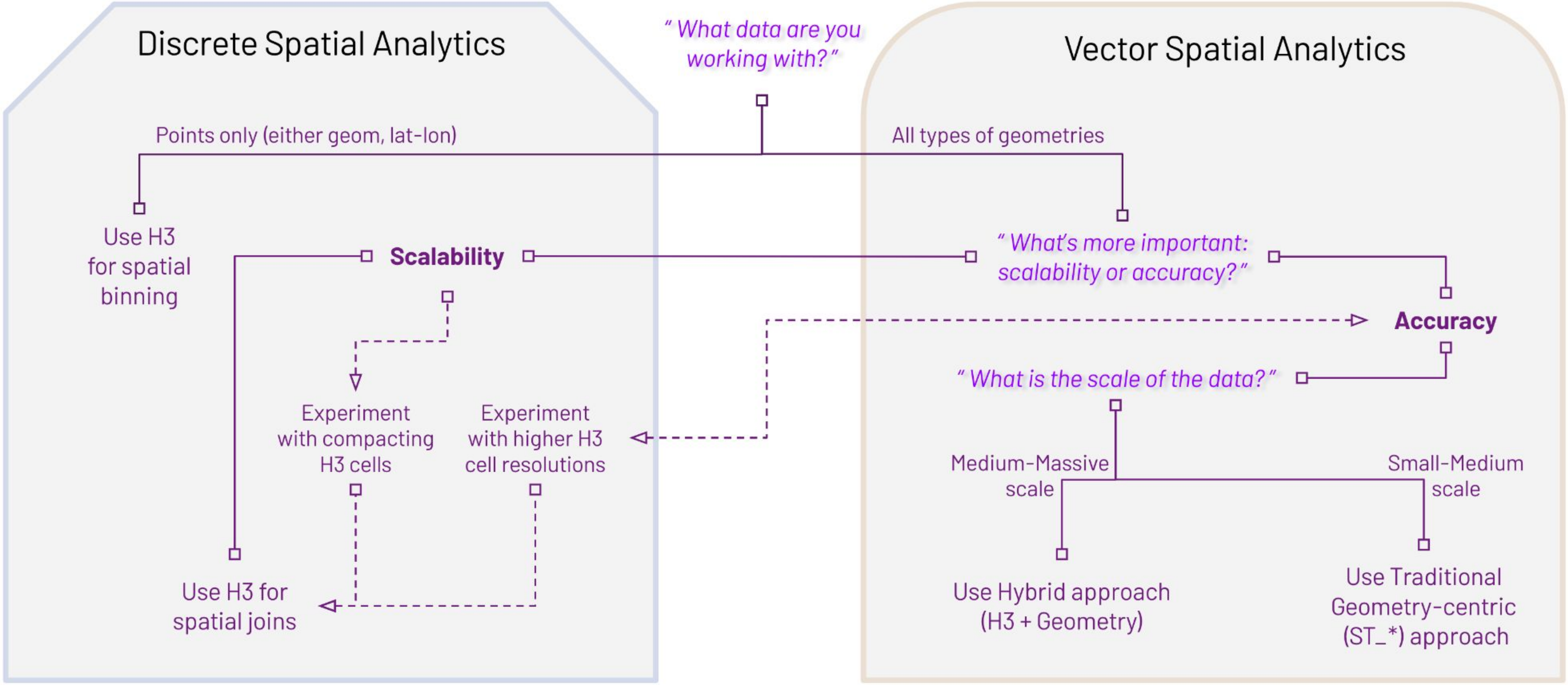
Most customers start their processing with this approach

Most flexible on Lakehouse!

Most performant on Lakehouse!

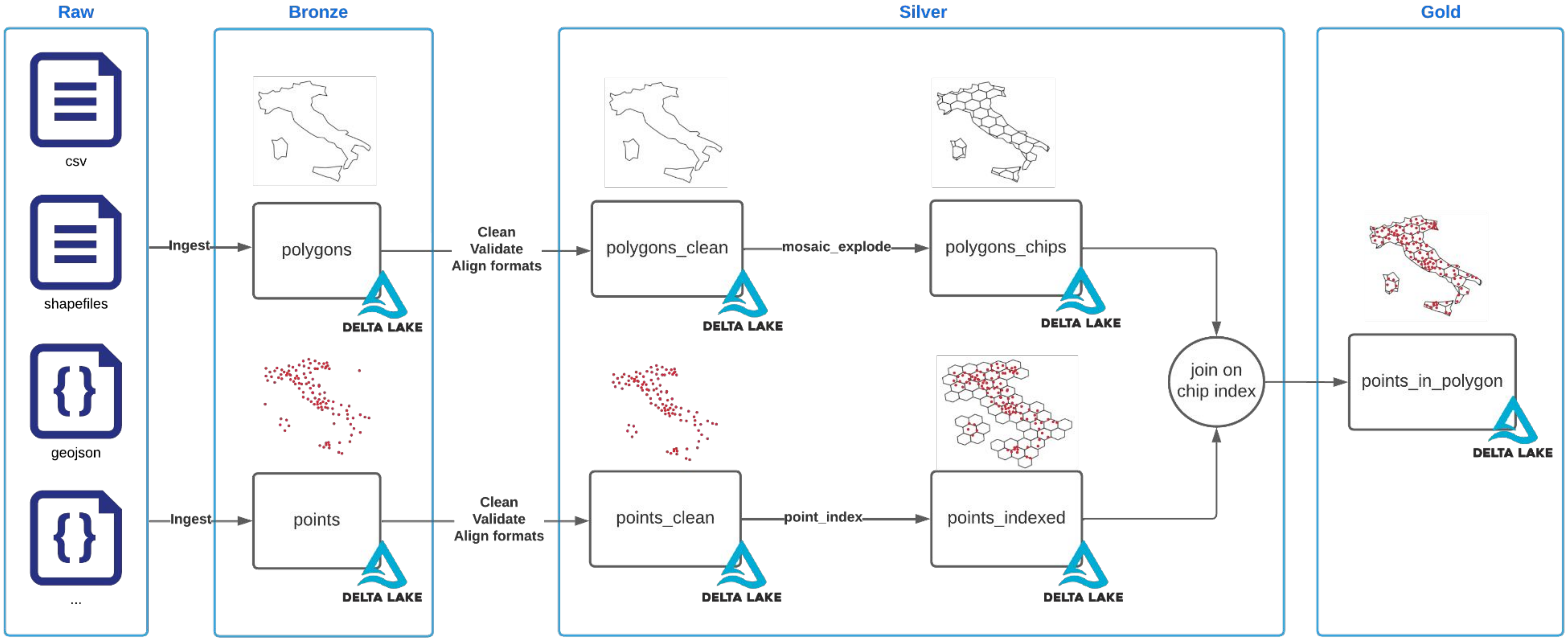
Optimization journey

Decision Tree for Spatial Joins



Link to Blog

Reference Join Pattern



databricks

[1] Coordinates -> H3

H3 API + Liquid Clustering to prepare lat / lon

- Uses function h3_longlatash3
- Liquid Clustering using cellid
- Numeric cellids much better for performance
- H3 resolution 6 used for this example

```sql
%sql
-- table will use liquid clustering
CREATE OR REPLACE TABLE trip_h3 CLUSTER BY (pickup_cellid) AS
(
  SELECT h3_longlatash3(pickup_longitude, pickup_latitude, 6) AS pickup_cellid, *
  FROM trip
);

-- execute liquid clustering
OPTIMIZE trip_h3;

SELECT * FROM trip_h3 LIMIT 10;
```

(66) Spark Jobs

_sqldf: pyspark.sql.connect.dataframe.DataFrame = [pickup_cellid: long, trip_row_id: long ... 18 more fields]

Table

	pickup_cellid	trip_row_id	vendor_id	pickup_datetime	dropoff_datetime	passenger_count	1.2 trip_distance
1	604222352397434879	-1928445750670314453	VTS	2009-06-18T21:25:00.000+00:00	2009-06-18T21:39:00.000+00:00	1	3.3
2	604222352263217151	-7362211067542464895	CMT	2014-05-04T13:19:06.000+00:00	2014-05-04T13:44:38.000+00:00	1	4.
3	604222352263217151	-4514160269029825361	CMT	2014-05-05T00:36:02.000+00:00	2014-05-05T00:44:17.000+00:00	1	1.
4	604222352263217151	-7968889011037105742	CMT	2014-05-04T15:28:15.000+00:00	2014-05-04T15:36:01.000+00:00	1	1.
5	604222352263217151	-6755337482988695663	CMT	2014-05-04T13:56:27.000+00:00	2014-05-04T14:03:33.000+00:00	1	1.
6	604222352263217151	1349114653060541076	CMT	2014-05-05T01:12:05.000+00:00	2014-05-05T01:27:48.000+00:00	1	4.
7	604222352263217151	2370329712272608276	CMT	2014-05-04T20:13:20.000+00:00	2014-05-04T20:31:33.000+00:00	1	3.
8	604222352263217151	2882698725654968184	CMT	2014-05-05T01:19:16.000+00:00	2014-05-05T01:29:45.000+00:00	1	3.
9	604222352263217151	8322486265112795975	CMT	2014-05-04T19:00:42.000+00:00	2014-05-04T19:28:29.000+00:00	2	4.
10	604222352263217151	-5027552516377834212	CMT	2014-05-04T21:37:31.000+00:00	2014-05-04T22:12:32.000+00:00	1	9.

10 rows | 2.04 minutes runtime   Refreshed 17 minutes ago

This result is stored as `_sqldf` and can be used in other Python cells.

[2] Areal Geometries -> H3

H3 API + Liquid Clustering to prepare polygons

- Uses function h3_tessellateaswkb
- Liquid Clustering using cellid
- Use of `INLINE` to shred the array result into `cellid`, `core`, and WKB `chip`
- Numeric cellids much better for performance
- H3 resolution 6 used for this example

```sql
%sql
-- table will use liquid clustering
-- The use of inline allows us to cluster by 'cellid'
-- NOTICE: we do not include the full geometry 'g' here;
--         rather, keep only vector chips per cellid
CREATE OR REPLACE TABLE state_h3 CLUSTER BY (cellid) AS
(
  SELECT state_row_id, INLINE(h3_tessellateaswkb(g, 6)), * EXCEPT(state_row_id, g)
  FROM state_poly
);

-- execute liquid clustering
OPTIMIZE state_h3;

SELECT * FROM state_h3 LIMIT 10;
```

(22) Spark Jobs

_sqldf: pyspark.sql.connect.dataframe.DataFrame = [state_row_id: long, cellid: long ... 3 more fields]

Table

	state_row_id	cellid	core	chip	stat
1	1	604222584191451135	true	AQMAAAABAAAABwAAAEyryLgnmlLAsXwZaYI7REBuW9j2/ZxSwAy2ks+LOkRAysw/2didUsDy2IXVjjZEQDZHXTjem1LAK0hF+Y...	New Je
2	1	604222540839124991	true	AQMAAAABAAAABwAAAA83vC0slFLAfLh81d7dQ0AP3St095ZSwGrM2lzr3ENAdfKM98+XUsC0zPxM+NhDQC+EJendlVLAleg...	New Je
3	1	604222345149677567	true	AQMAAAABAAAABwAAABbRyktLmlLAYtuBvt2QREAJaw3sK51SwJ0/jGHlj0RAllOm5wmeUsC9mEL034tEQJccJgMlnFLAQ0OU...	New Je
4	1	604233111693164543	false	AQMAAAABAAAABgAAAMDJ9HwMw1LAzKsfKXNsQ0BSs2ASicJSwPRu78TzbkNAbJmzDsy/UsANQeAR7W9DQEVXMpYmvlL...	New Je
5	1	604222508626870271	true	AQMAAAABAAAABwAAAB7aJXw/uVLAjMAWw+5vREBU5x4AG7xSwFj64ujwbkRAFP8ci/K8UsAu3AO+7GpEQG7We1LvulLAyP...	New Je
6	1	604222563119267839	false	AQMAAAABAAAACAAAANpyIXWDe1LAgBYjX40HRECRt4LEm3tSwAhr/sAeB0RArFYrtWx+UsD3Ajw0LwZEQDeiS51fgFLA+qS...	New Je
7	1	604222324077494271	true	AQMAAAABAAAABwAAANG3RK9YglLAiXtumTpsRECERe+bNYVSwllUar1Ha0RA3RRARRaGUsALZF96R2dEQHL26r0ahFLAsvo...	New Je
8	1	604222574259339263	true	AQMAAAABAAAABwAAAENFZVv3hVLAciONU7LXQ0CqP89awohSwLuZNLrB1kNA5TZM+pyJUsCnFz020NJDQJfaoU2th1LAX...	New Je
9	1	604233090620981247	true	AQMAAAABAAAABwAAALUI3Vxdx1LAJDPNKiLtQ0CJVayvKMpSwNB85Xkk7ENAWteJR/nKUsAu0iSuLOhDQC23w0X/yFLA43I...	New Je
10	1	604222530907013119	true	AQMAAAABAAAABwAAANwdm+KGo1LA1yjSmvLqQ0DEphwzU6ZSwLjrquL76UNASJzbpSmnUsCM5NeBBuZDQIc0oX40pVL...	New Je

databricks

[3] Tessellation technique: H3 + Geometry Harmonized

Point-in-Polygon join using already prepared tables

H3-Centric → **H3-Geometry Hybrid** ← Geometry-Centric

```
%sql
CREATE OR REPLACE TABLE pickup_state_h3 CLUSTER BY (cellid) AS (
SELECT
  state_h3.*,
  t.*
from
  (select /*+ SKEW('pickup_cellid') */ * from trip_h3) as t
  join state_h3
  on cellid == pickup_cellid
  where
    core or
    st_contains(st_geomfromwkb(chip), st_point(pickup_longitude, pickup_latitude))
);
```

Approximate results (when no "where" clause)

Take advantage of fully contained cellids

Take advantage of vector chips

4

Unified Data Access + Sharing

Delta Sharing: An open approach to data sharing

Data Provider

databricks

or any other platform that supports Delta Lake

Delta Sharing

No replication
Easy to manage
Secure

Data Recipient

Any tool	Any use case	Any location
Power BI	BI	aws
tableau	Data Science	
Apache Spark	Analytics	
pandas		
Java		
And many more		On-premises

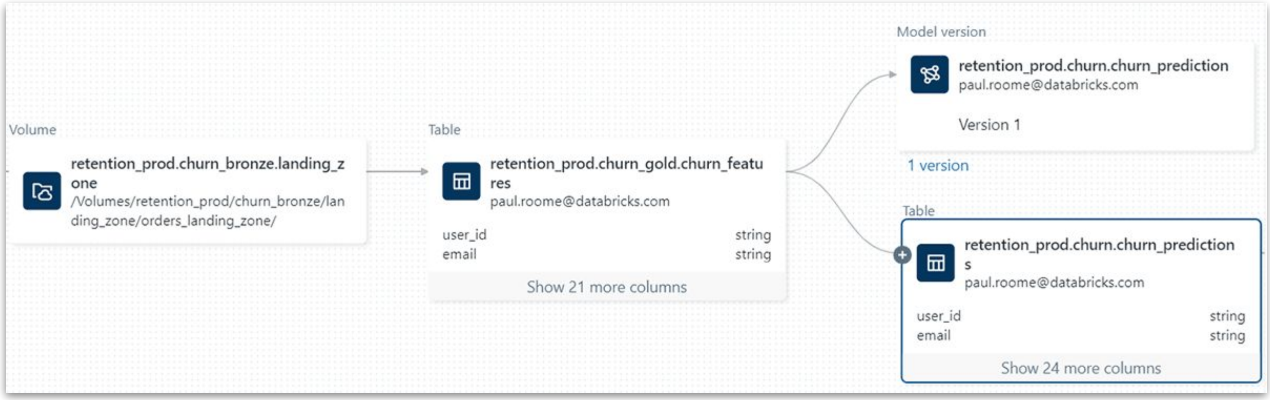
[Delta Sharing]

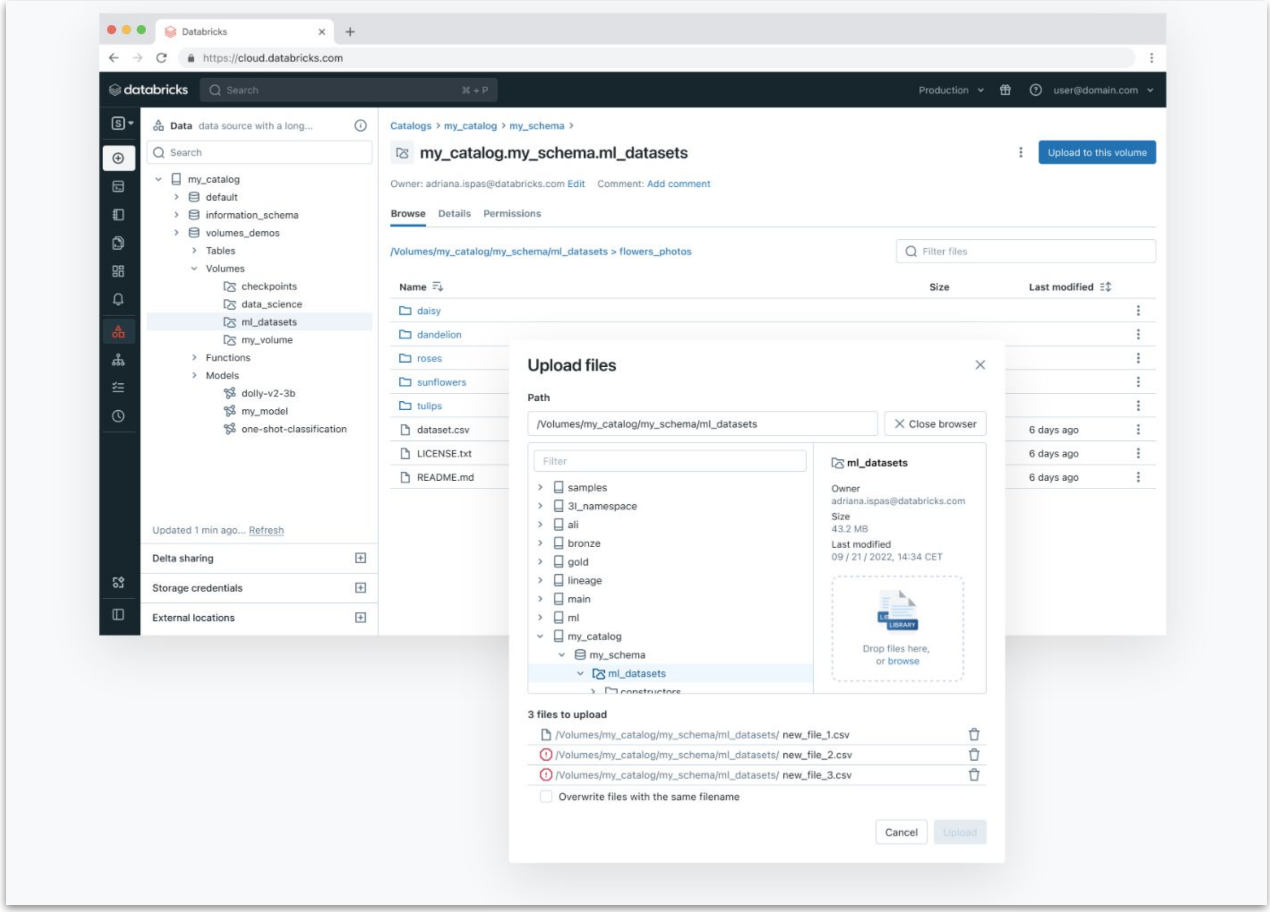
Volumes: Unity Catalog

Manage non-tabular data

- Arbitrary files [text, image, audio, video, PDF, or XML] governed inside schemas in UC alongside tables, models, + functions
- **Metadata Patterns Emerge** (heavy file payload with meta tables)

“[We thought that] we'll have to build some kind of abstractions around tables to deal with regular non-tabular files. With Volumes, we can rely on Unity Catalog for everything going forward.” - **Nike**





Announcement Blog

databricks

E.g. Raster Metadata Architecture

Bronze (Raw)

- Landing zone for **‘raw’** raster data.
- Little to no data architecture investment.

Silver (Base)

- Raster tiles stored using **GeoTIFF** format.
- Accompanying **Delta metadata table** providing acquisition date and time, resolution, bands etc.

Gold (Indexed)

- **Indexed raster tiles** subdivided by grid index strategy [e.g. h3, maybe also quadbin].
- Raster represented using **binary type** (or other universal type).

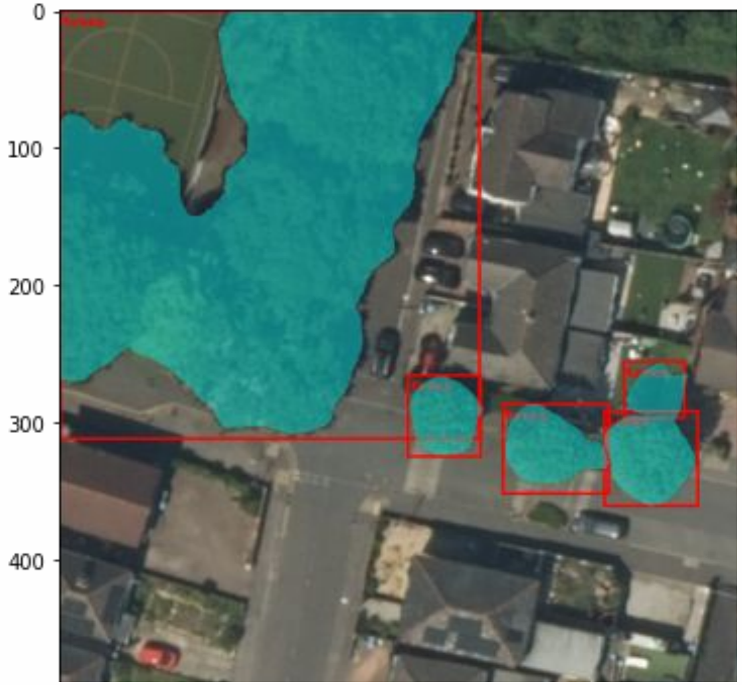
Platinum (Derive and Enrich)

- Rasters clipped by the **‘highest value’** vector geometries.
- Derived from indexed raster tiles via collocate, clip, and merge steps.

DELTA LAKE

Lakehouse + Spatial AI

"Give me Trees"



LLM + LVM

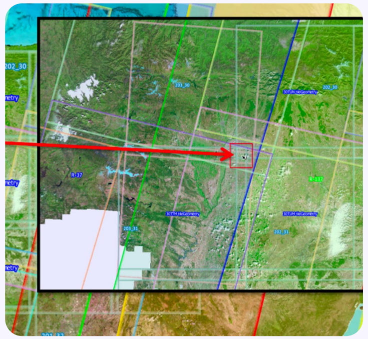
Generate Geotagged vectors

Spatial Framework

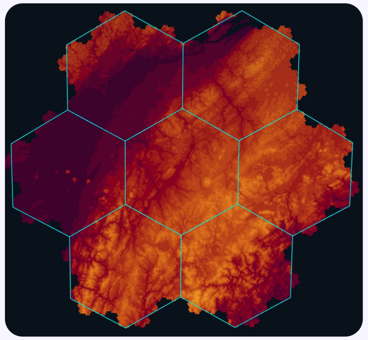
Unified Data Model for all GIS Data

Vector + raster for “all spatial” analysis

Earth Observation Data / Planetary Computer

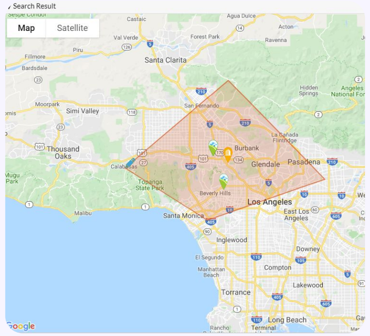


Earth Observation Data



Raster Grid Tiles

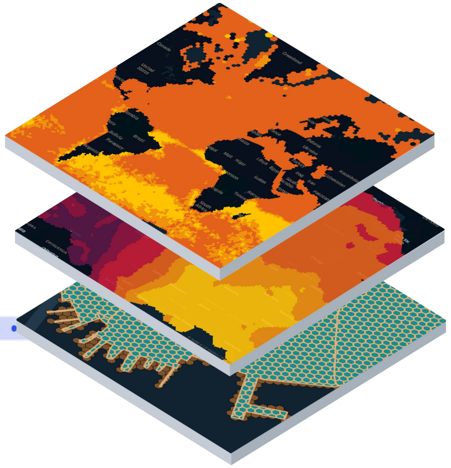
Areas of interest



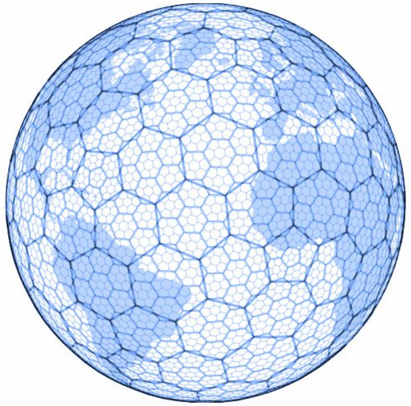
Areas of Interest as
Open Spatial Formats

Vector Grid Tiles

Unified Data Model



Observable Earth Model



- All data uses same grid w/CRS abstracted
- Layer joining + stacking is just idx ==
- Stored as Delta; metadata + data in the same format; Delta Sharing ready
- Marketplace ready
- AI ready

5

What's Next?

Spatial SQL Roadmap

Supercharge geospatial analysis

30+ H3-based indexing [Already GA] makes it easy to see spatial patterns, combine disparate data, visualize and integrate with ML

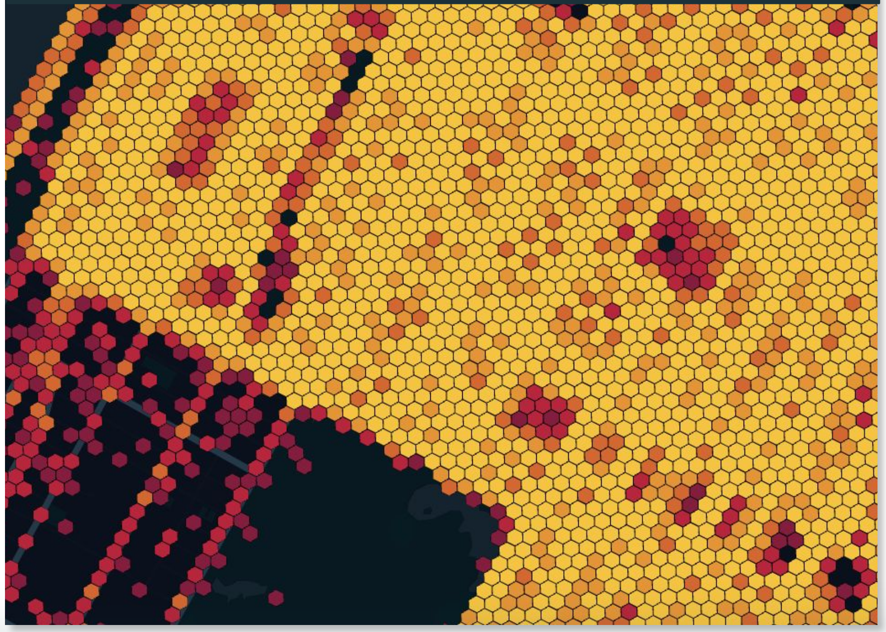
60+ Spatial Functions [Private Preview] – broad set of ST_ expressions provide flexibility for working with Vector data

Geometry / Geography Types [2024+] – read and write spatial data to native types, easily convert between WKT, WKB, GeoJson

Fast Spatial Joins [2024+] – efficient spatial query execution, builds on types + *will be ongoing improvements after initial release*

Additional ST_ Functions [2024+] – prioritizing additional customer and partner inputs

Rideshare pick-up locations in New York City visualized in a Databricks Notebook using Kepler.gl

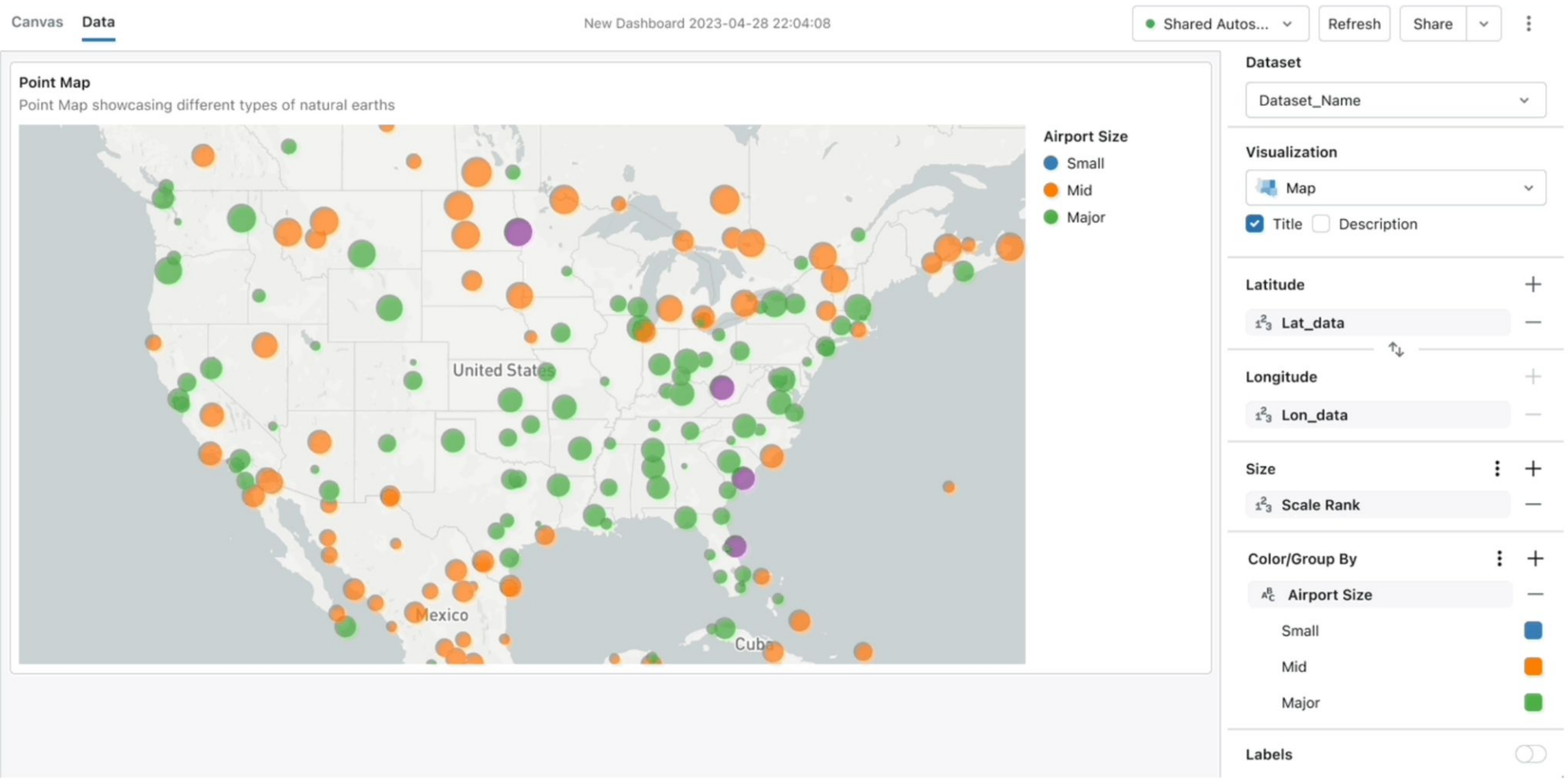


```
with wkt_poly as ( select
   'POLYGON((-115.42 32.57, -115.42 32.57,-115.42
      32.57, -115.42 32.57, -115.42 32.57))' as g )
select
 st_geogarea(g) as dbx_area_meters,
 st_area(g) as dbx_area_units
from wkt_poly
```

Map Visuals

Full refresh on visualizations– including Maps

- Point maps for lat/lon **[Public Preview OCT]**
- Automatic geocoding for choropleth maps **[2024+]**
- Geometry/Geography data (Point, Line, or Polygon) **[2025]**
- H3 Maps (hex bins) **[2025]**



Geospatial: Key Partners

Best-in-class technology + mindshare

CARTO	esri	SAFE SOFTWARE
Spatial BI Platform	GIS Platform | ArcGIS	Spatial ETL | FME
▪ Native integration with Databricks H3 ▪ Analytics Toolbox [Download] ▪ Builder ▪ Workflows **Future integration initiatives:** ○ Iterate on existing ○ GeoParquet + RasQuet	▪ Existing Spark extension: ○ GeoAnalytics Engine **Future integration initiatives:** ○ Integrate with ArcGIS CDW connections [requires public preview spatial types + perf]	▪ Spatial ETL ▪ Read/Write to all common spatial formats ▪ Databricks Connector [Preview] **Future integration initiatives:** ○ Native integration with Databricks Spatial SQL

Thank You!