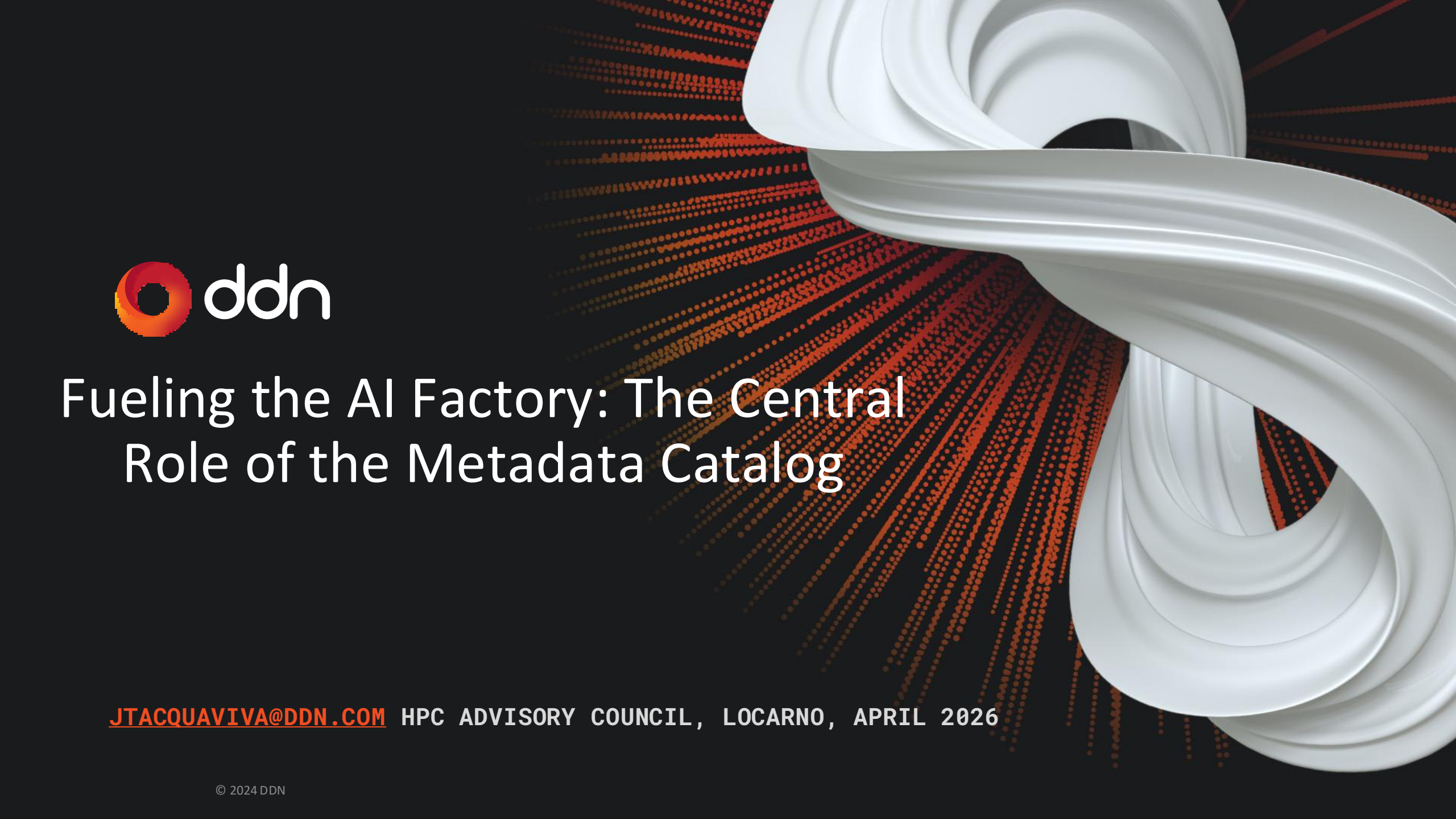


Fueling the AI Factory: The Central Role of the Metadata Catalog

JTACQUAVIVA@DDN.COM HPC ADVISORY COUNCIL, LOCARNO, APRIL 2026

If data is new oil, what is metadata?

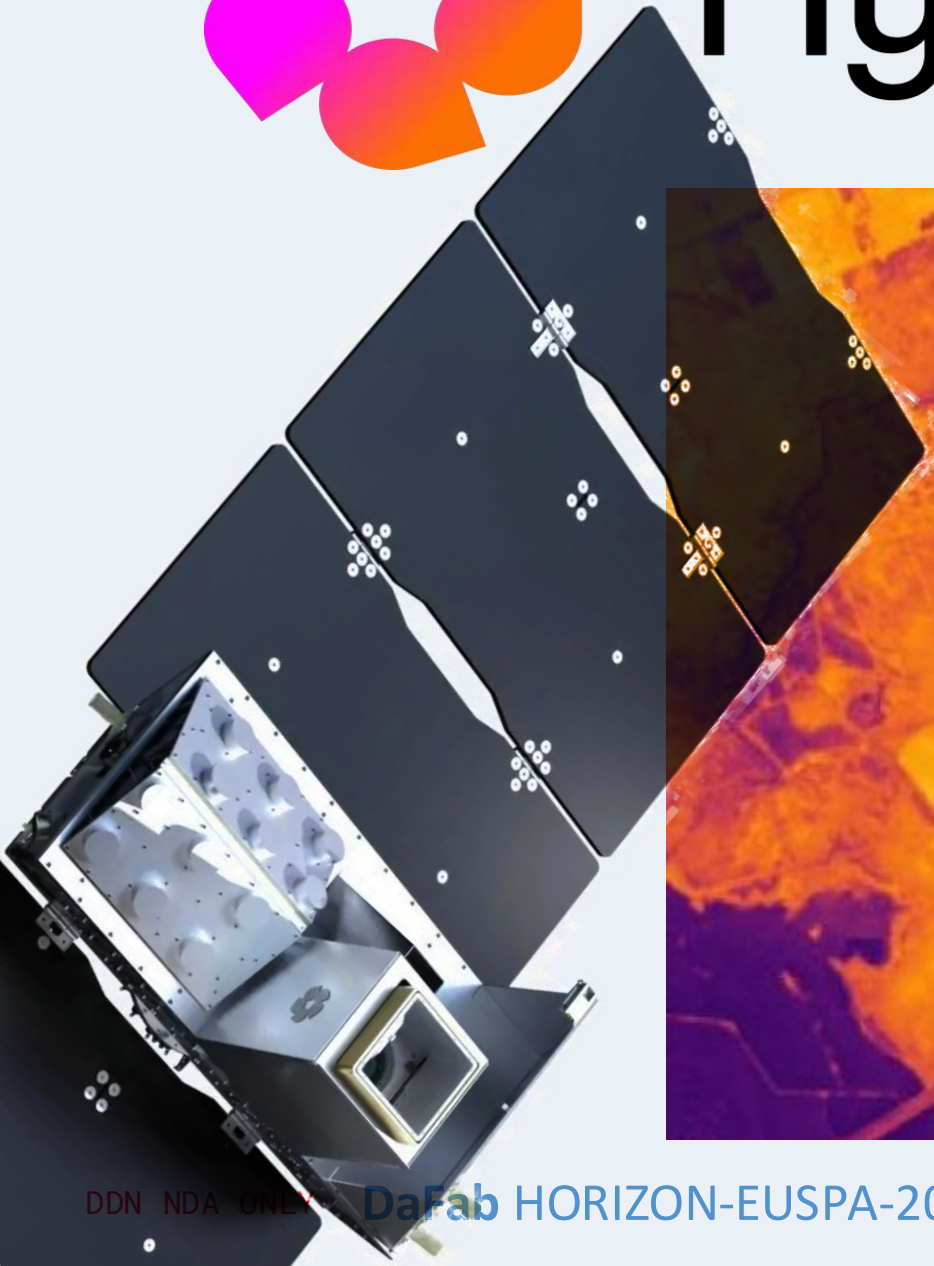
Infinia has been built by DDN **from scratch** to address Modern Data and Application Demands

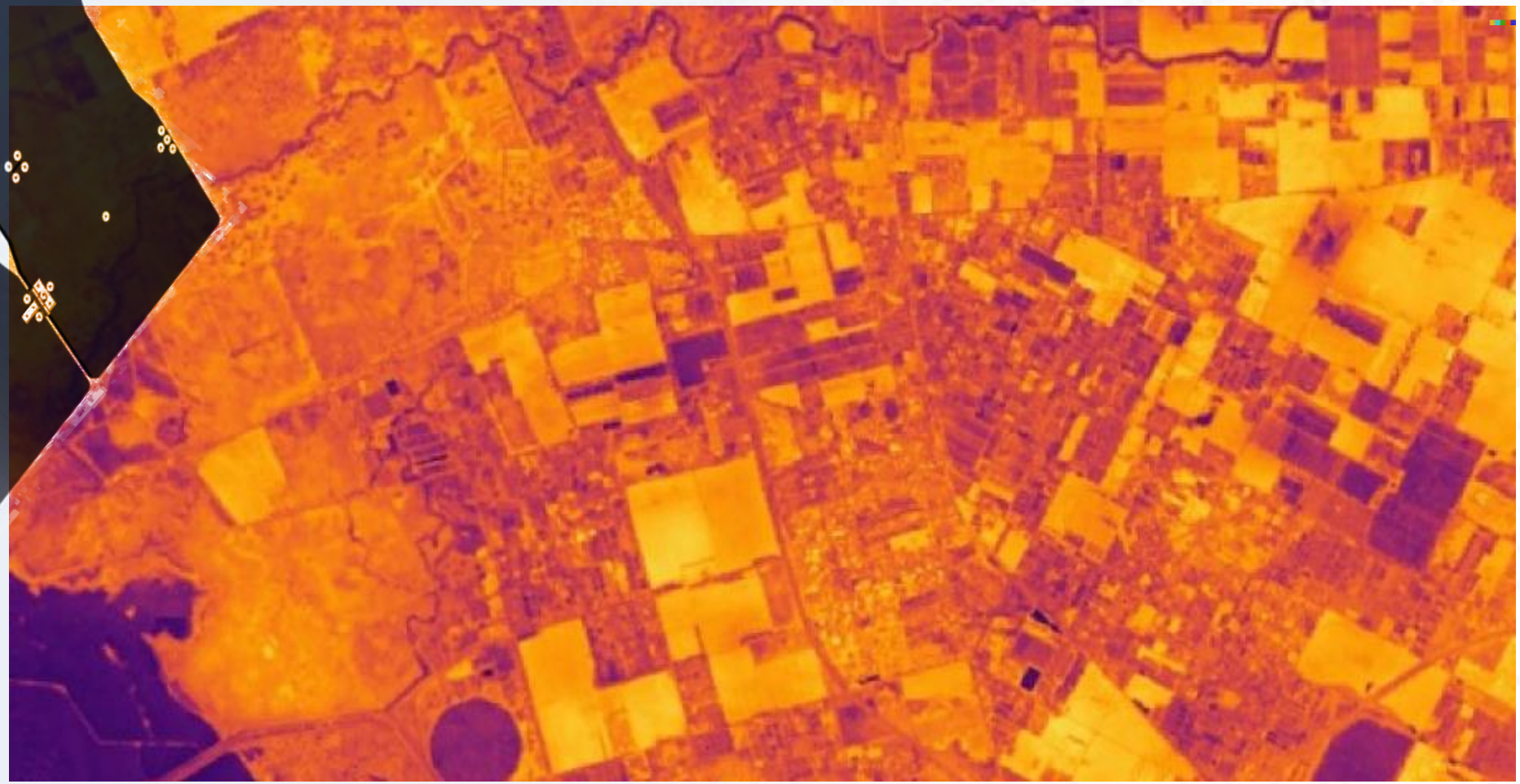
DaFab

AI Factory for Copernicus Data at Scale

1. Applying AI to Sentinel-2 images
2. Increase data informational density
3. Foster EU AI startup competitiveness

Hydrosat

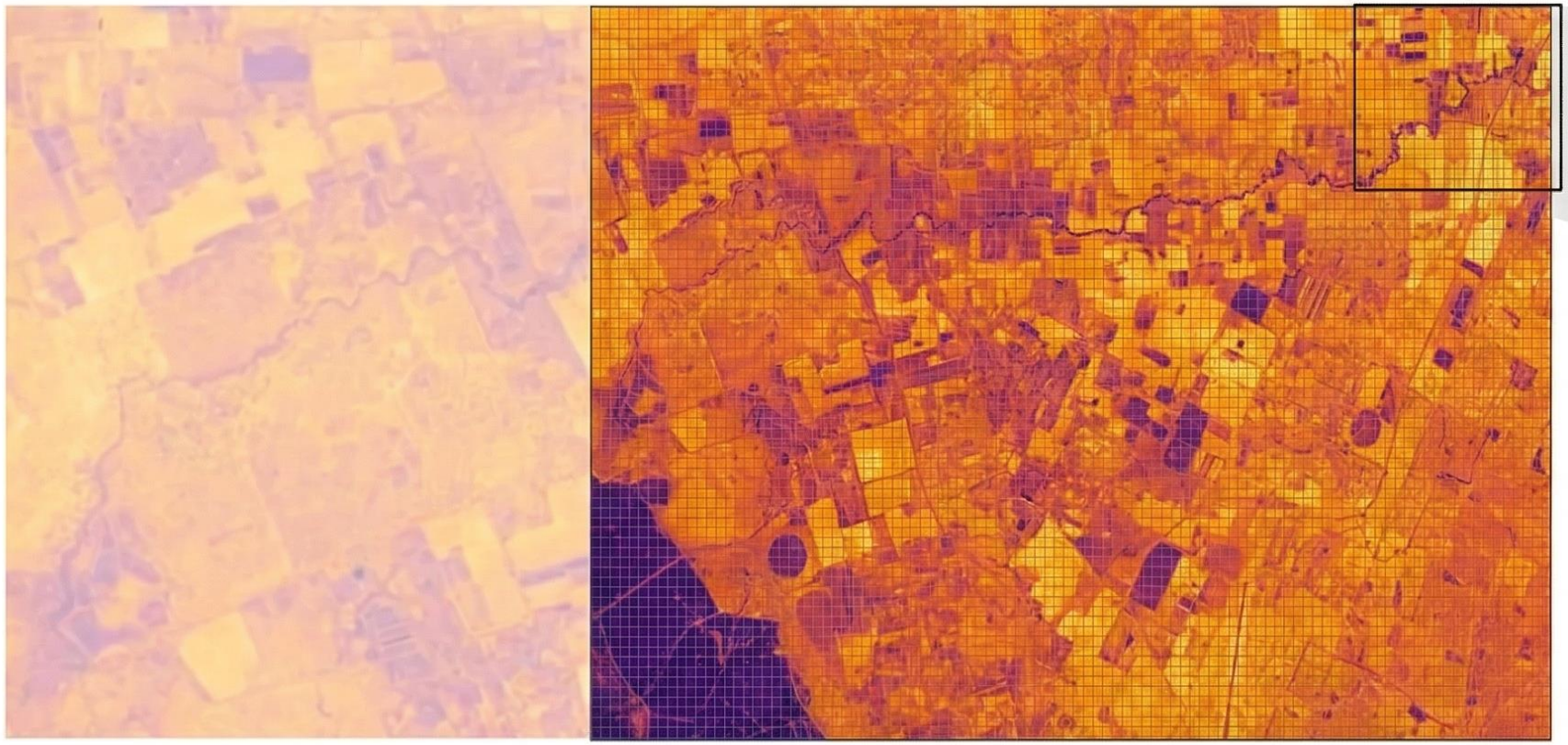


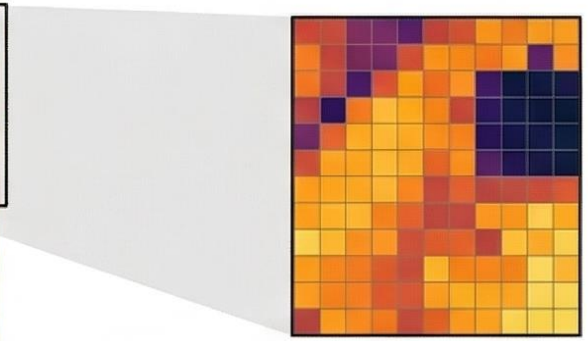


Hydrosat

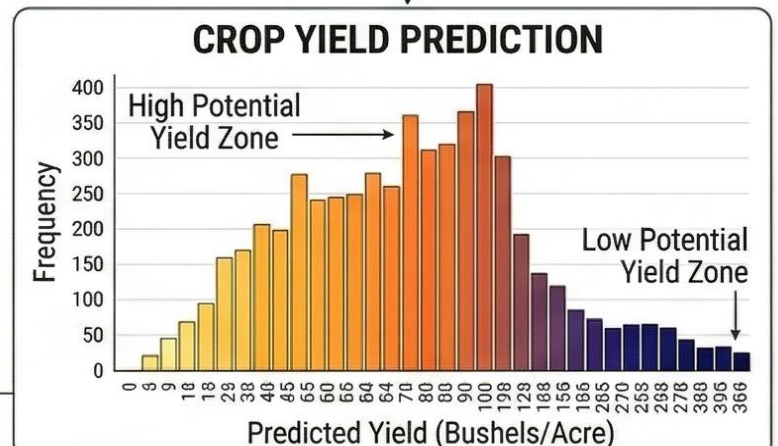
PIXEL-BY-PIXEL GEOSPATIAL ANALYSIS

ANALYSIS: Each pixel (approx. [Calculated Pixel Count]) is processed individually.





Grid Magnification (e.g., 200x)
e.g., [Example Row x Col Pixel Area]



HydroSat + DaFab Metadata

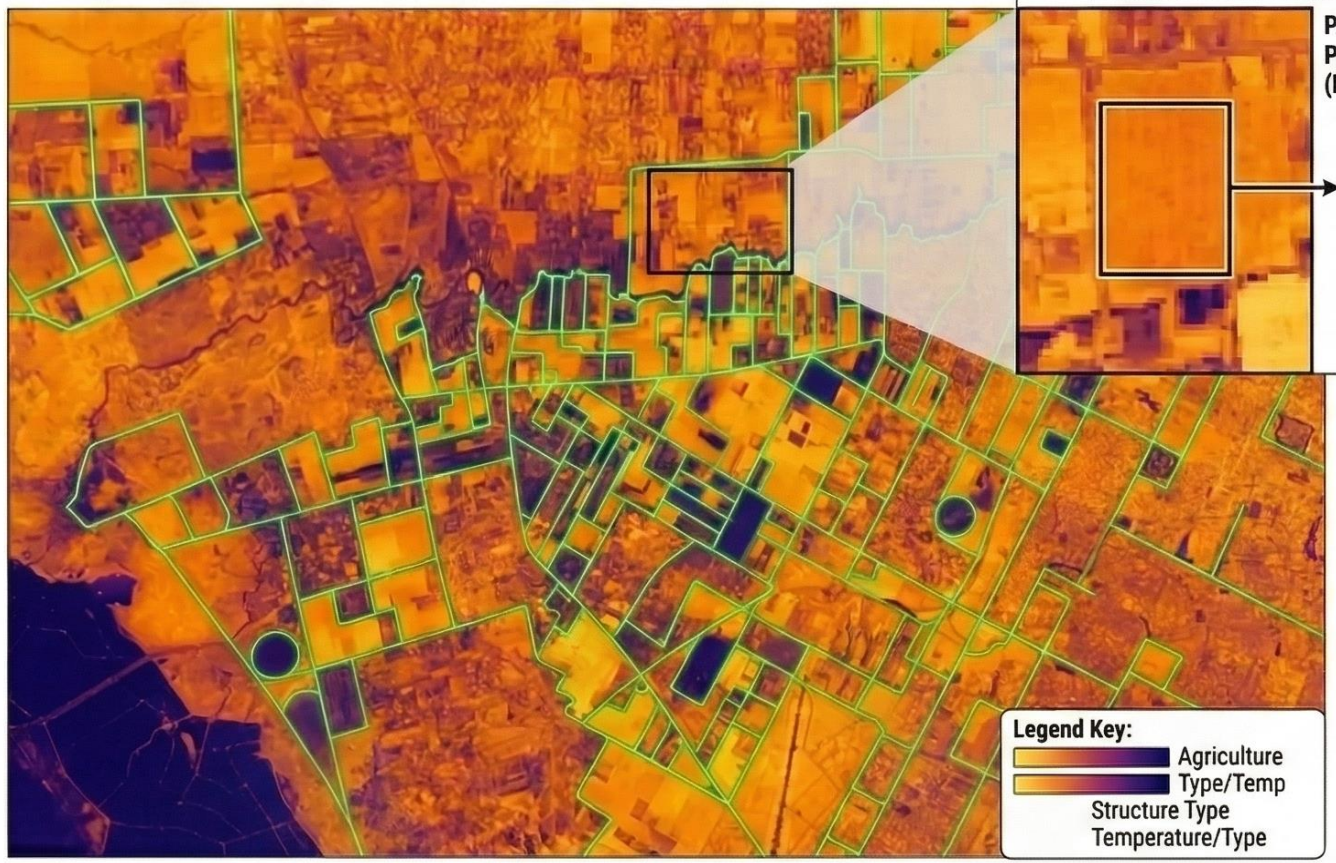
Border detection
DaFab fine-tuned
NIVA open-source AI model

GeoJSON files compliant with STAC EO standard added as metadata to all images

Hydrosat

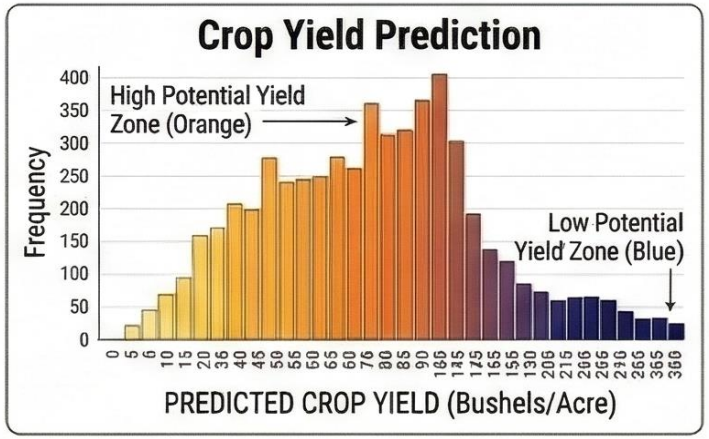
PARCEL-PER-PARCEL GEOSPATIAL ANALYSIS: CROP YIELD PREDICTION

INPUT: MULTI-SPECTRAL PARCEL MAP



Zoomed-In Fields Analysis

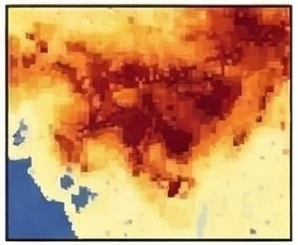
Crop Yield Prediction



SCIENTIFIC DATA MANAGEMENT

RUCIO

Yield Trend Map



Outcome: infer 416x faster

Parcel-based approach → much lower runtime, memory, and storage, enabling affordable, large-scale deployment on HPC. Pixel-based is still useful when per-pixel detection sensitivity is required, but it is far costlier to operate.

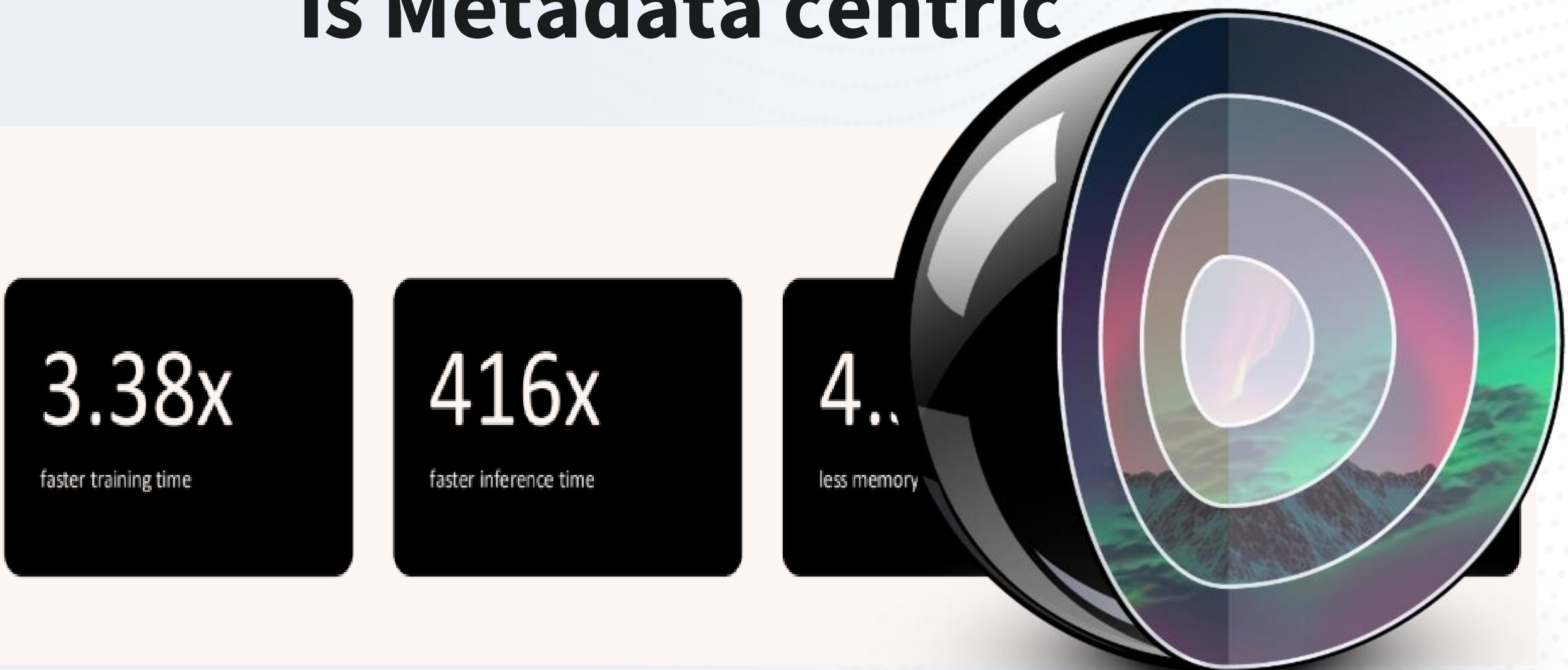
3.38x
faster training time

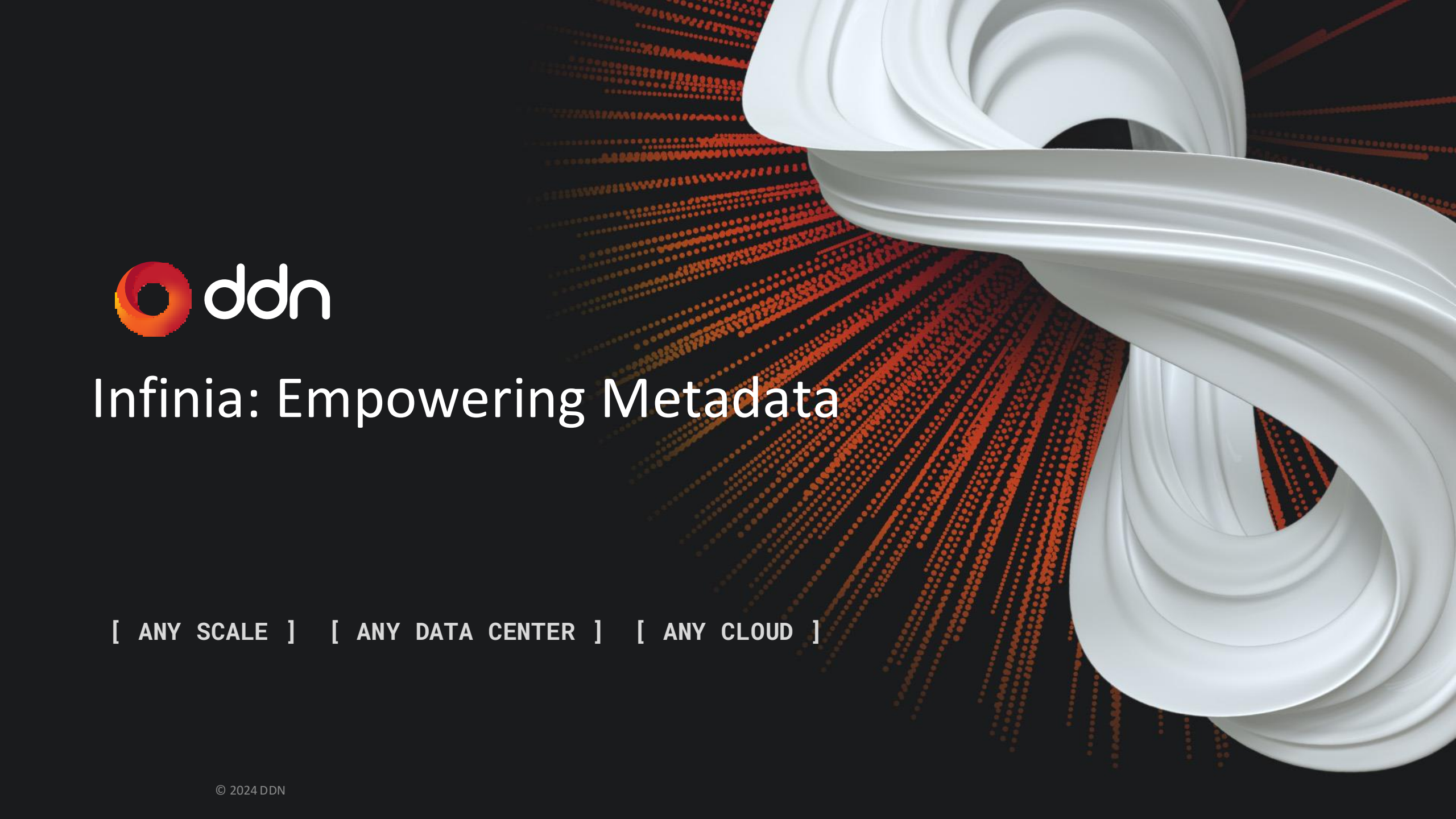
416x
faster inference time

4.36x
less memory used

4x
less storage required

An Intelligent Data Platform is Metadata centric





Infinia: Empowering Metadata

[ANY SCALE] [ANY DATA CENTER] [ANY CLOUD]

DDN Infinia is a high performance, SW-defined, distributed data platform designed for AI and Cloud

Infinia has been built by DDN **from scratch** to address Modern Data and Application Demands

Unstructured AND Structured data

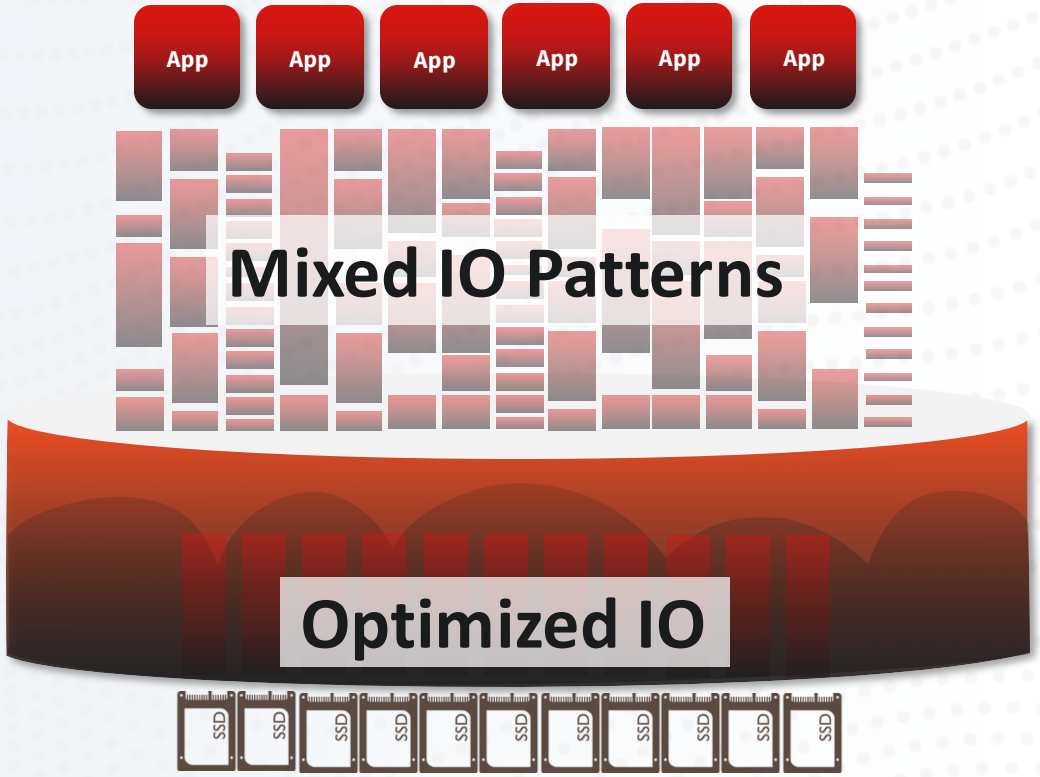
Data Platform Delivered as a Service

Hyper Simple Operation

Infinia Built for Performance

Data Managed to Flash for Highest Performance and Optimal Longevity

- Autonomous optimization for low latency and throughput workloads
- No read/modify/write overhead due to log structured operations
- Byte addressable I/O with wide striping for large i/o blocks
- Client-side coalescing of small I/O's



Under the hood!

- Infinia is a cluster

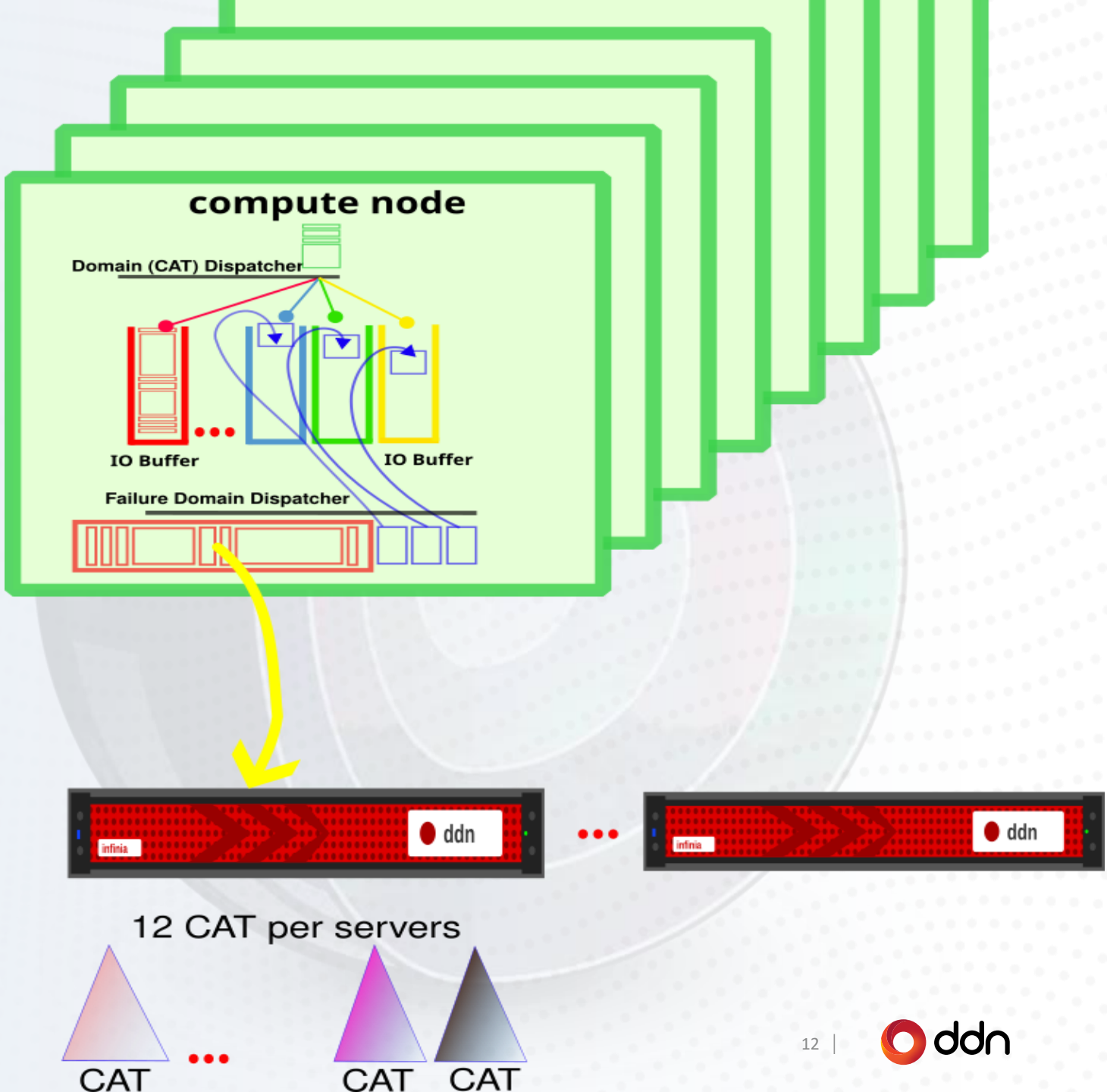
  Core Affine Threads - CAT

- Failure domain

  Topology aware

- Small IO Acceleration

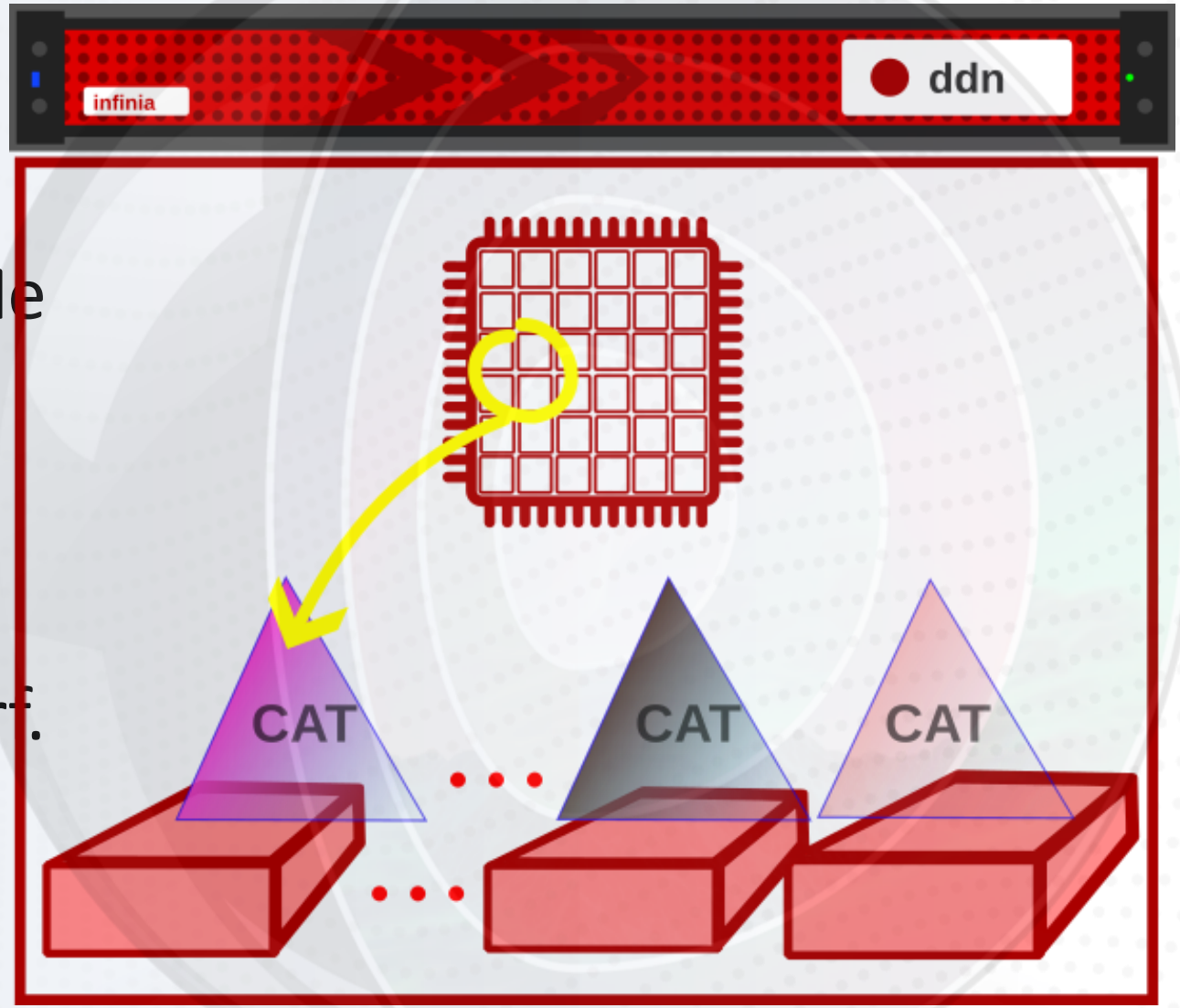
  Bypass EC stage



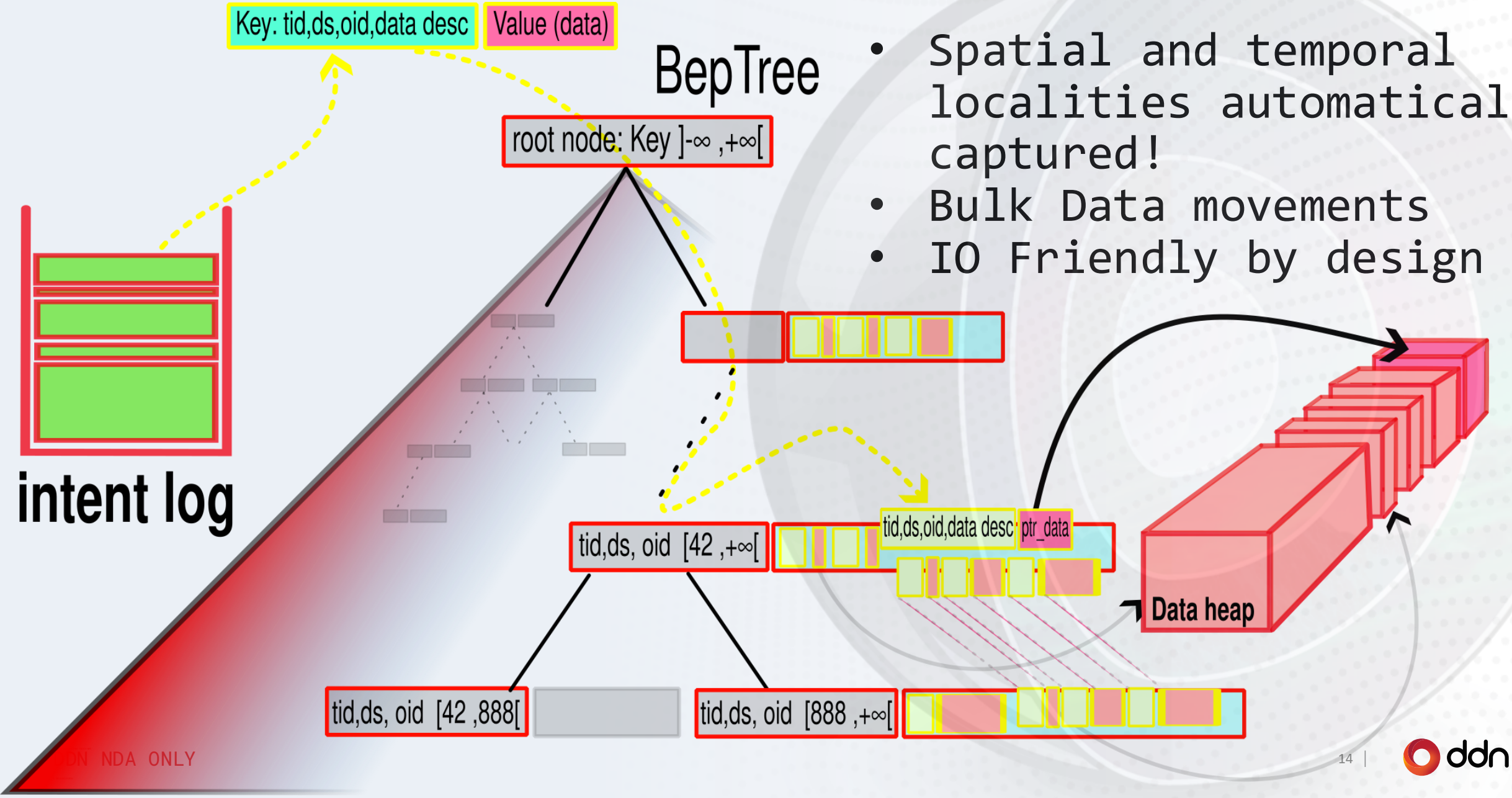
Infinia Server: Efficiency

Set of Cores exclusive to each NVMe

- Minimize threads migration
- No context switch
- Adjustment of CPU/NVMe perf.



Infinia: BepTree spilling

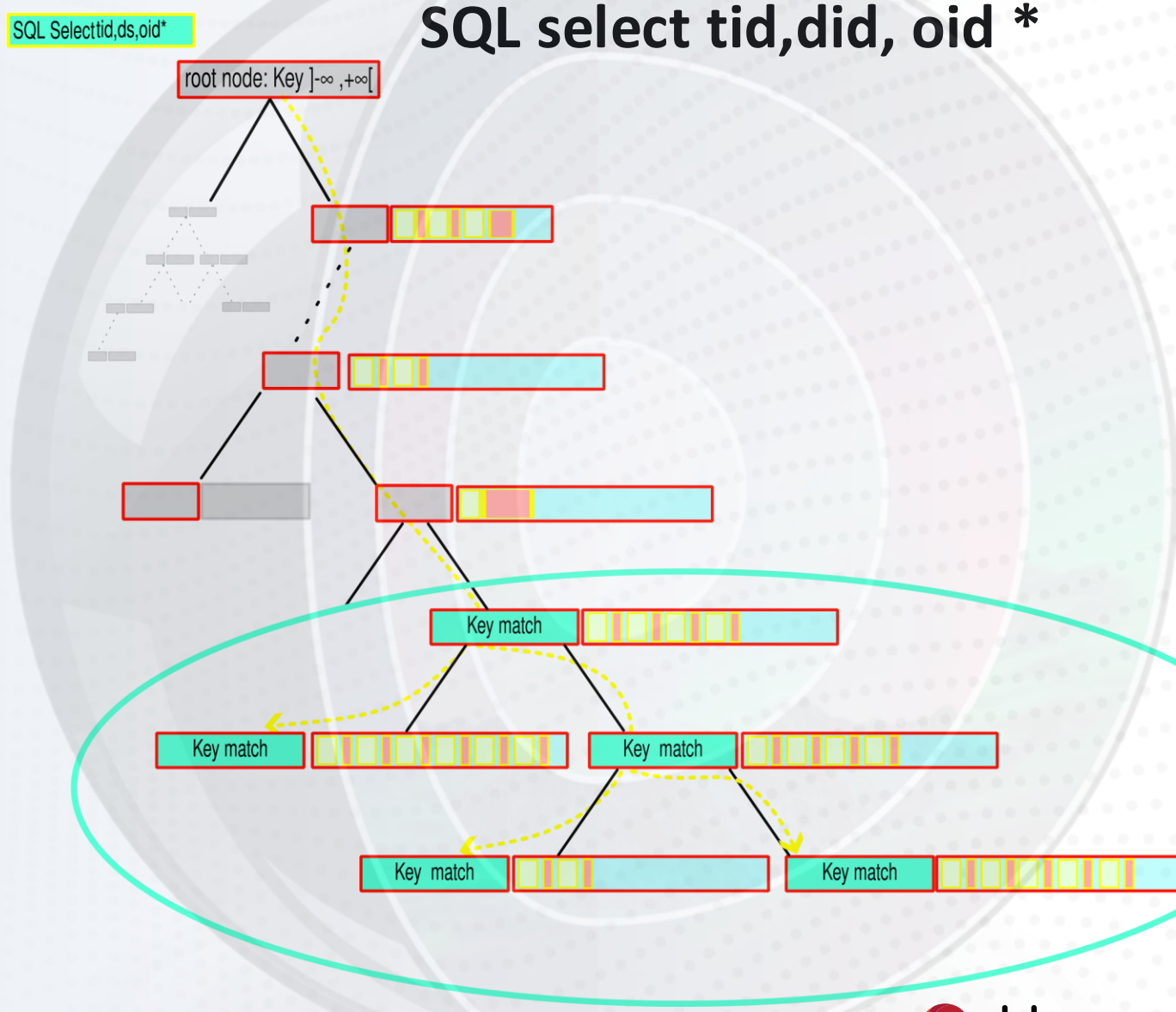


- Spatial and temporal localities automatically captured!
- Bulk Data movements
- IO Friendly by design

ddn

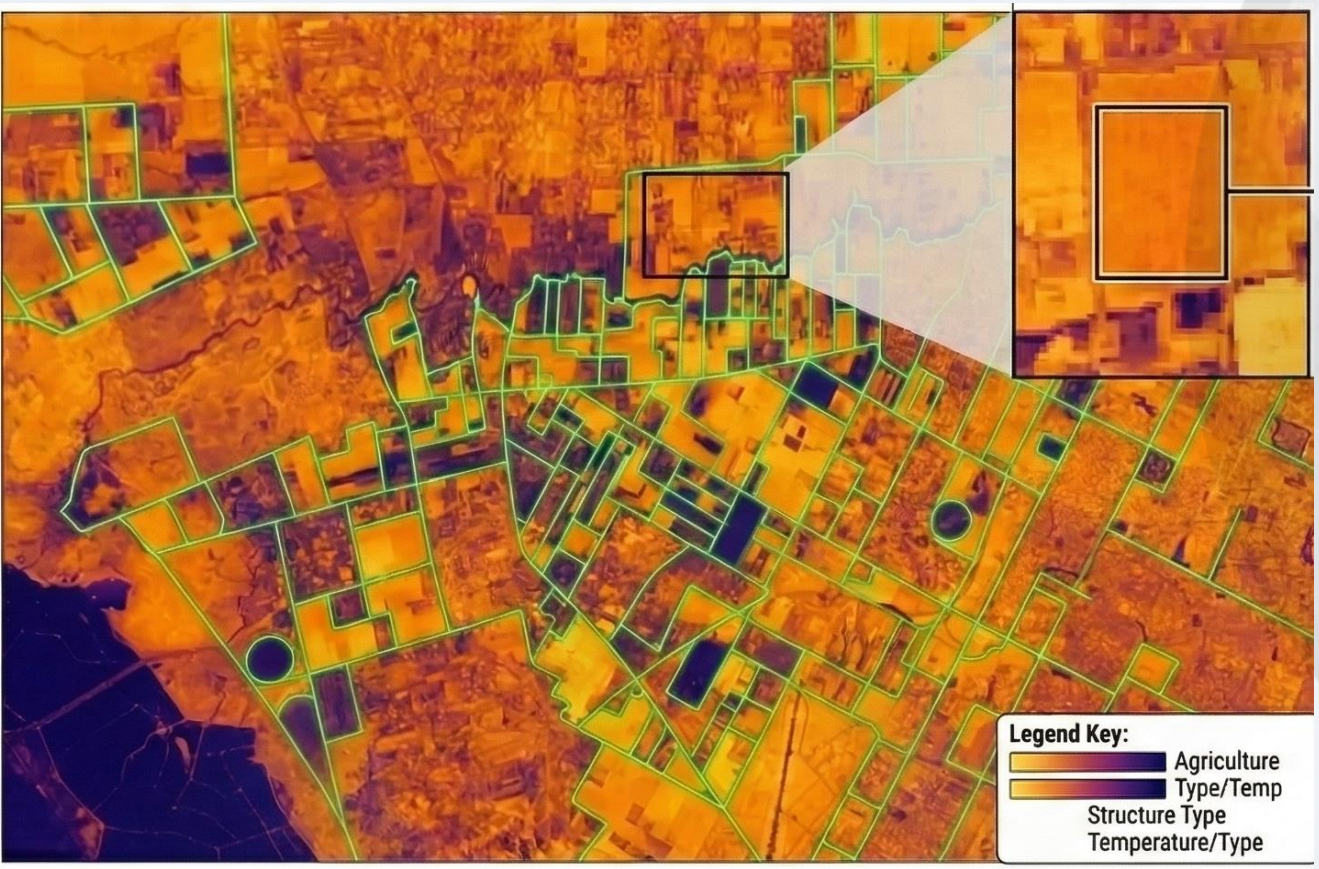
Infinia metadata integration

- Data and metadata in the tree
  - Indexed by key
- Key search -> tree traversal
  - Data and metadata related to a specific key are in a neighboring subtree
  - Catch temporal locality
- Leaf parsing = spatial locality

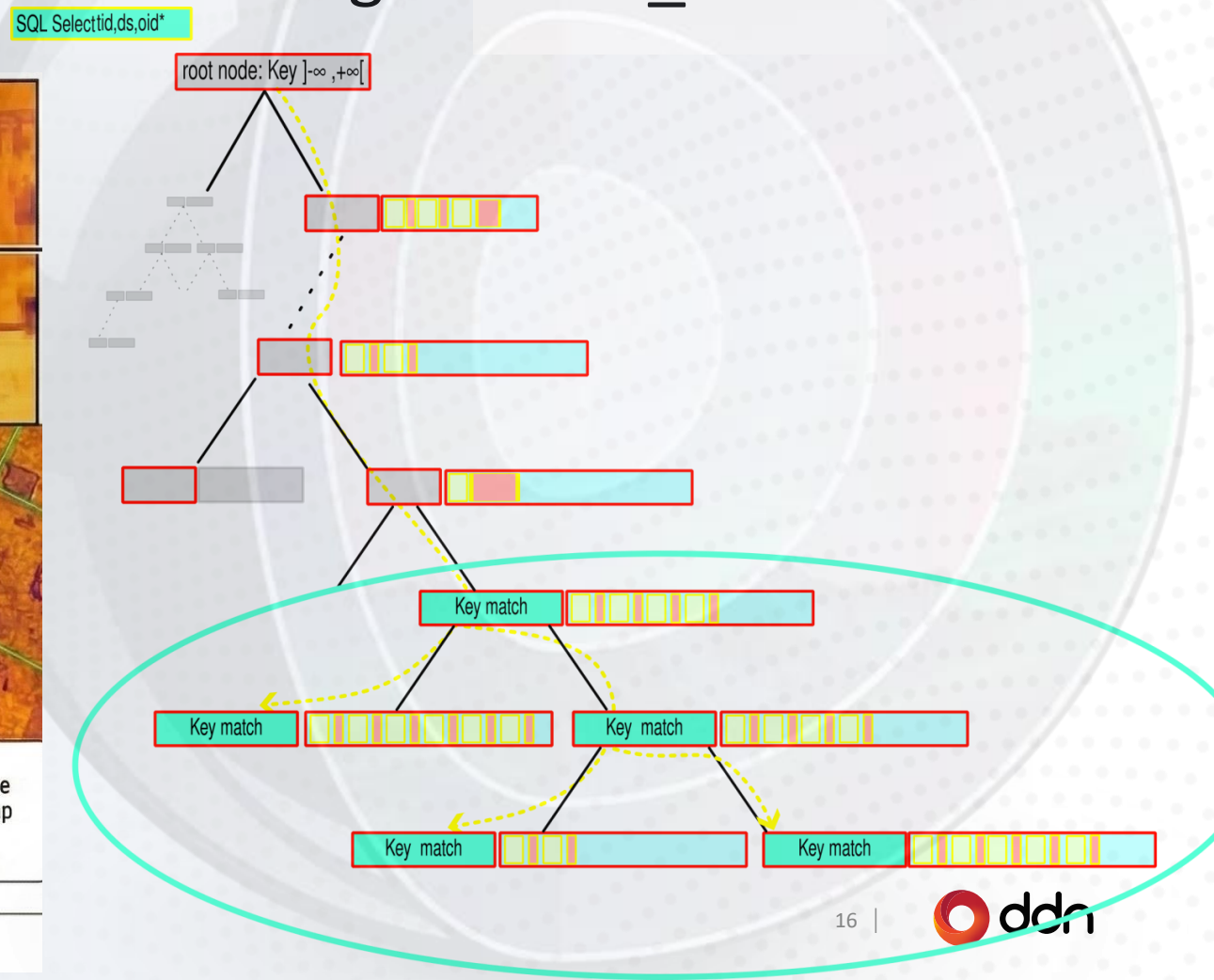


More Infinia metadata acceleration: indexes

SQL CREATE INDEX crop ON field.color



Insert subtree for the index linking to all ob_id as value



PREDICTABLE PERFORMANCE

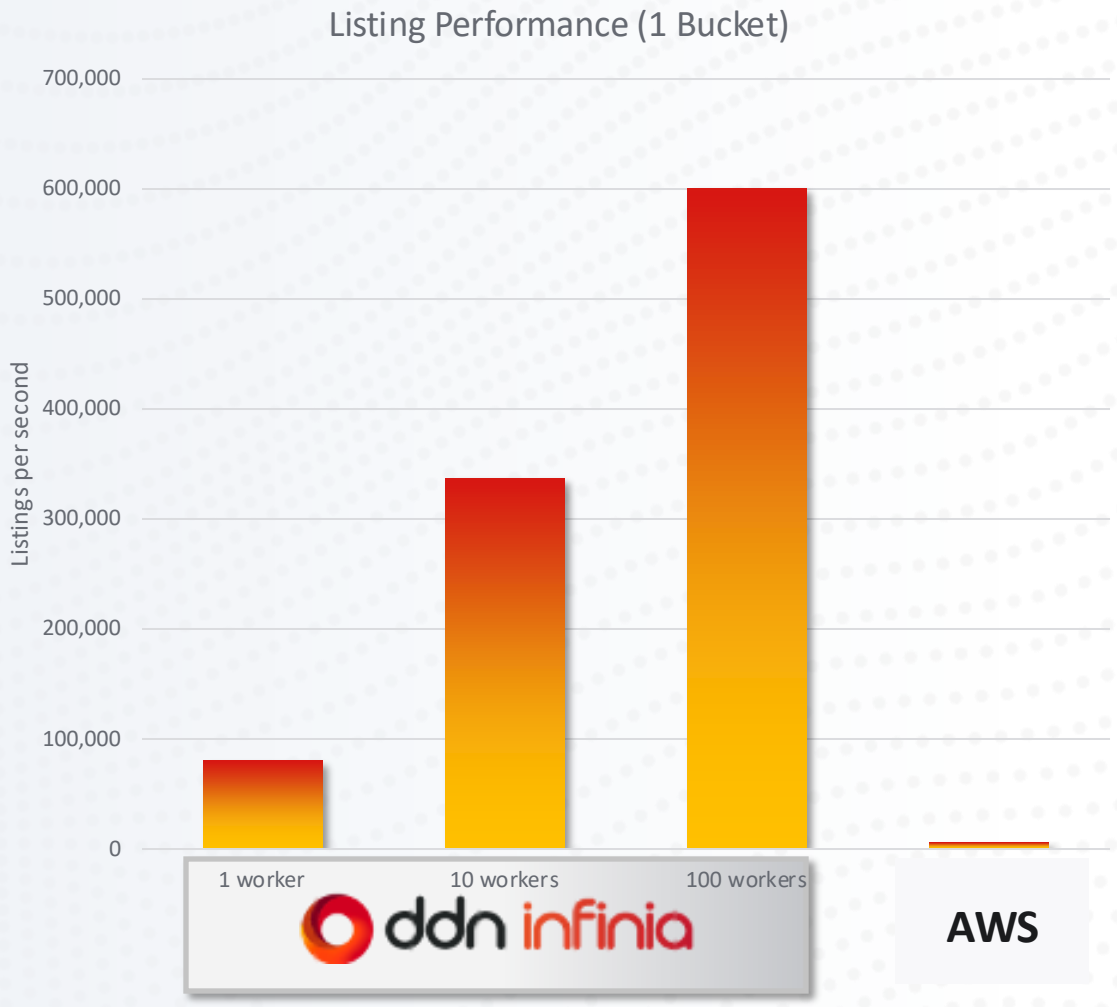
Infinia built to deliver business outcome

Fast Object Listing + SQL = Dataset selection

```
-- Satellite images by time range
Sql> SELECT * FROM Copernicus WHERE date >= '2025-01-01' AND
date < '2025-02-02';

-- Satellite images for a given area and time range
Sql> SELECT * FROM Copernicus WHERE date >= '2025-01-01' AND
date < '2025-02-02' AND latitude BETWEEN min_latitude AND
max_latitude AND longitude BETWEEN min_longitude AND
max_longitude;

-- Satellite images for a given area and time range
and a specific feature
Sql> SELECT * FROM Copernicus WHERE date >= '2025-01-01' AND
date < '2025-02-02' AND latitude BETWEEN min_latitude AND
max_latitude AND longitude BETWEEN min_longitude AND
max_longitude AND water_body_presence > 0.85;
```



Infinia Metadata Engine

Research shows x400 inference acceleration

- Requires external database
- Consistency and time gap

- **Infinia** supports **semantic metadata**
  - Scale with capacity
  - Advanced SQL support
  - Index for semantic acceleration

Infinia from research to production

- **Available Q3 26**