ANADARKO PETROLEUM CORPORATION

KINETICA INC

FROM DEEP LEARNING TO NEXT-GEN VISUALIZATION: A GPU-POWERED DIGITAL TRANSFORMATION

Ingrid Tobar

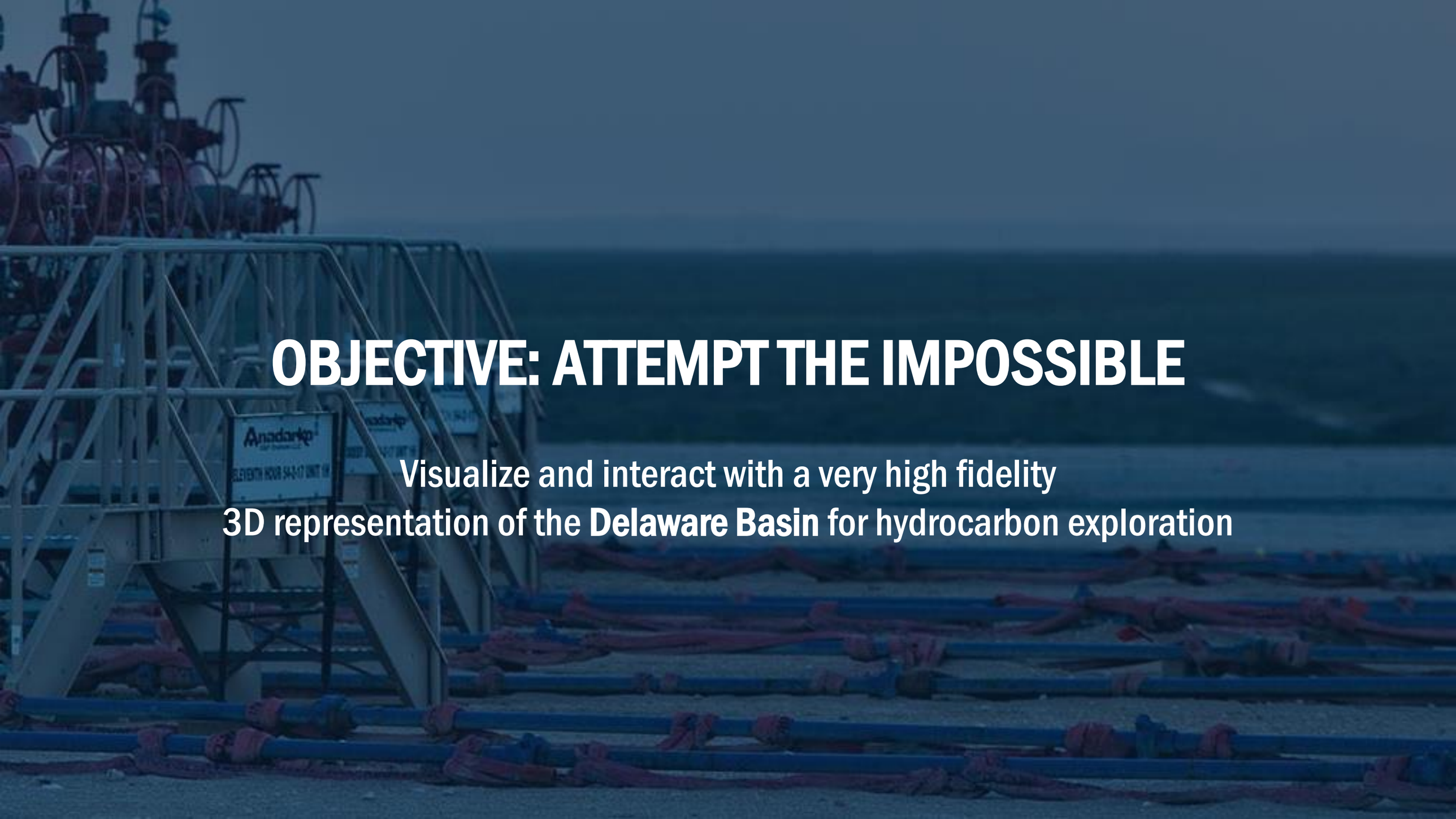
Senior Data Scientist
Anadarko

Amit Vij

President & Co-Founder
Kinetica

San Jose, California

March 18, 2019



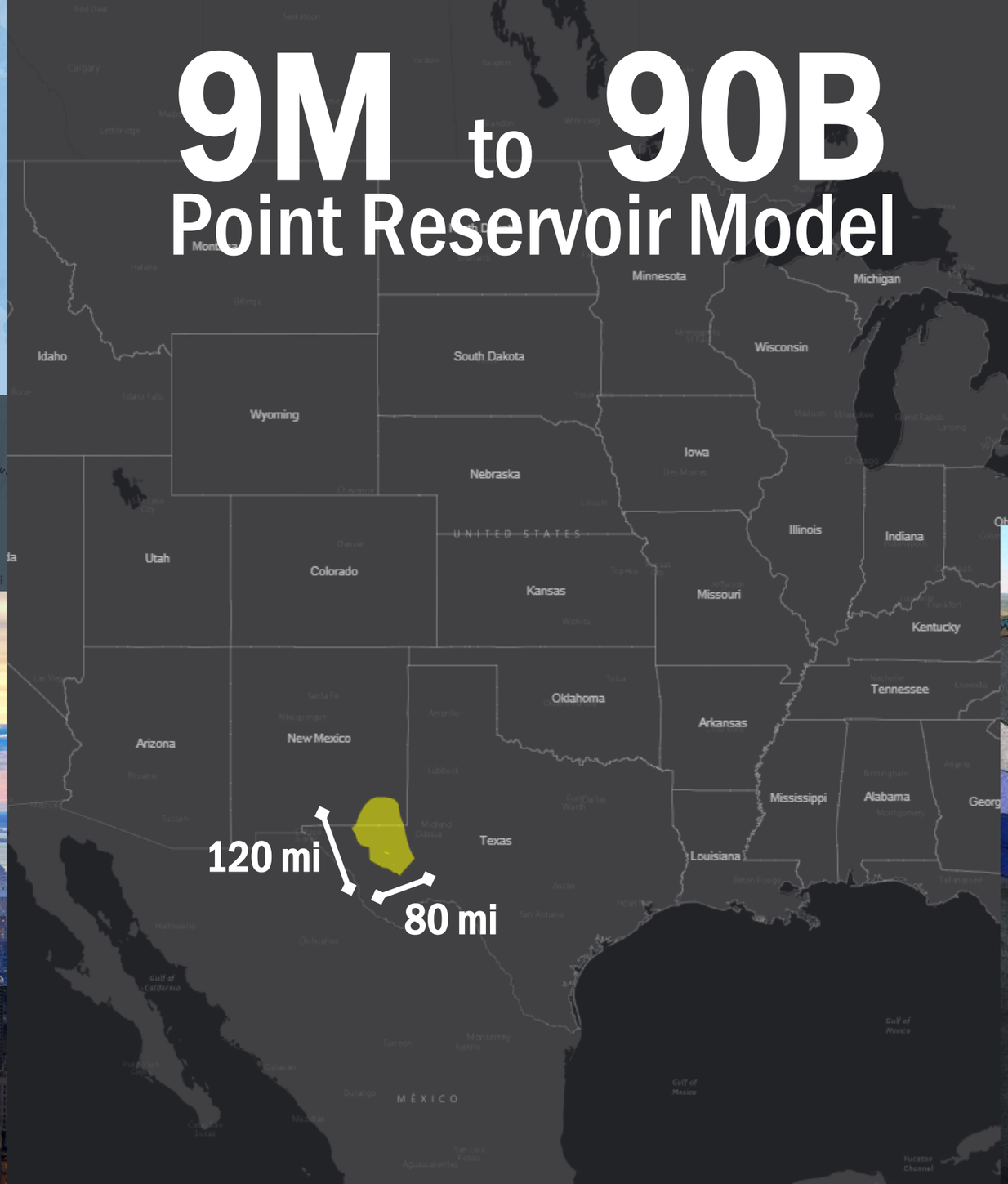
OBJECTIVE: ATTEMPT THE IMPOSSIBLE

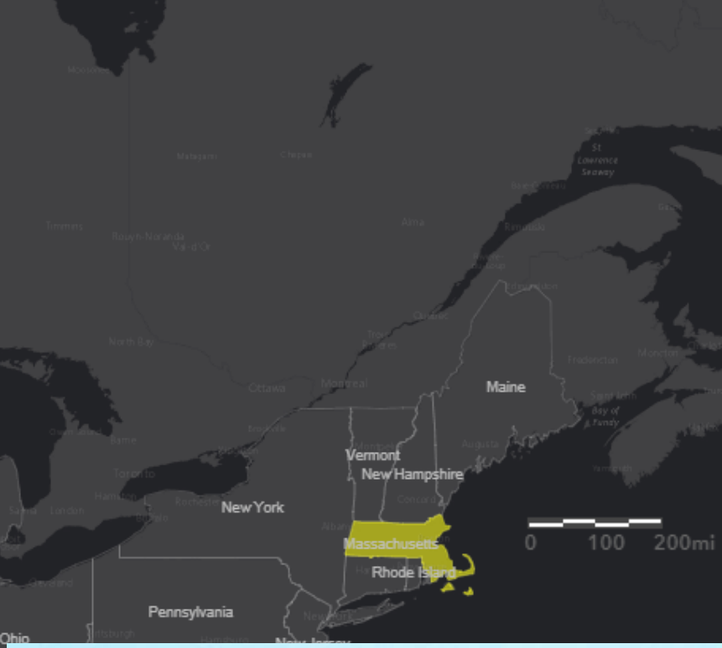
Visualize and interact with a very high fidelity
3D representation of the **Delaware Basin** for hydrocarbon exploration

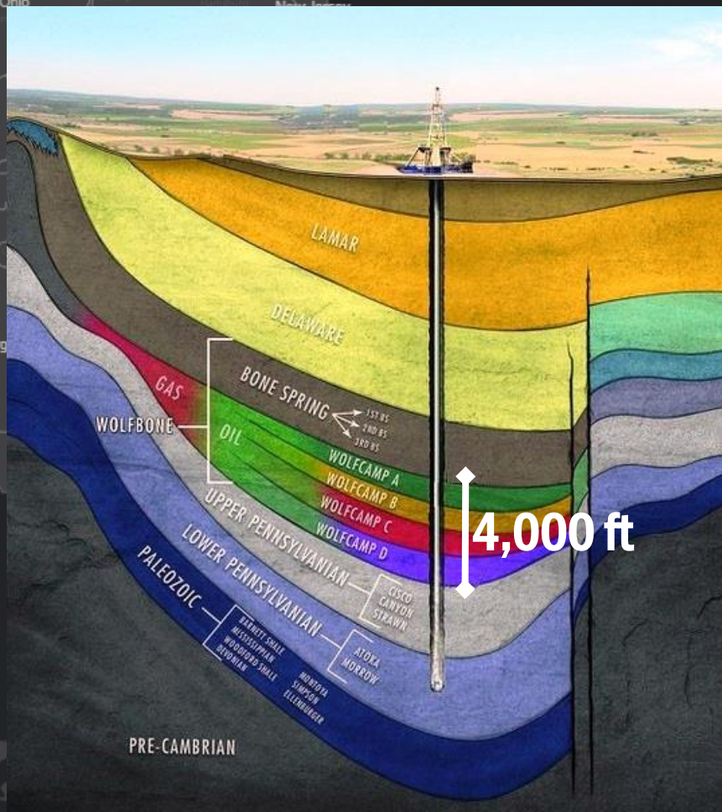
9M to 90B
Point Reservoir Model

The Delaware Basin is roughly the size of Massachusetts...

and 3x the height of the Empire State Building







Agenda

- About Anadarko
- GPU-Enabled Tech and Projects
- About Kinetica
- GPU-Accelerated Visualization

About Anadarko

Corporate Strategy

LEGEND

FOCUS AREAS
DJ Basin, Delaware Basin, DWGOM

CASH GENERATION
DWGOM & International Oil

FUTURE VALUE
Exploration & LNG

ANADARKO STRATEGIC FOCUS AREAS

- DJ Basin
- Delaware Basin
- Deepwater GOM

TOTAL VOLUME
700
MBOE/D

CAPEX
$4.8B

ENHANCING SUCCESS IN DWGOM

Advanced geophysical analytics to enable exploration with tiebacks to existing infrastructure

EXPANDING LOWER 48 FOOTPRINT

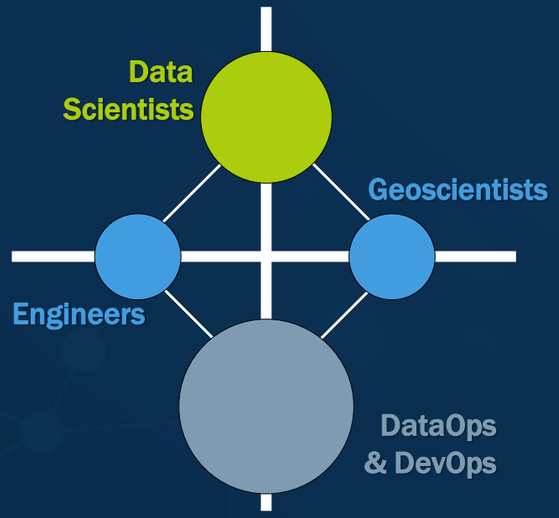
High density Lower 48 subsurface characterization to provide optionality

ENABLING DIGITAL OPERATIONS

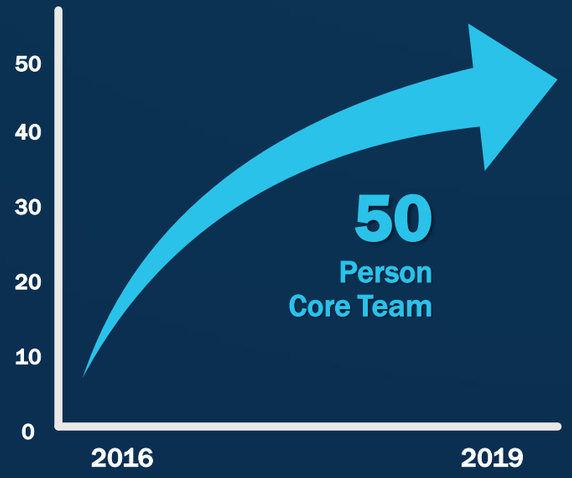
"Intelligent" control and edge computing in Drilling, Completions and Production

AAET: Advanced Analytics and Emerging Technology

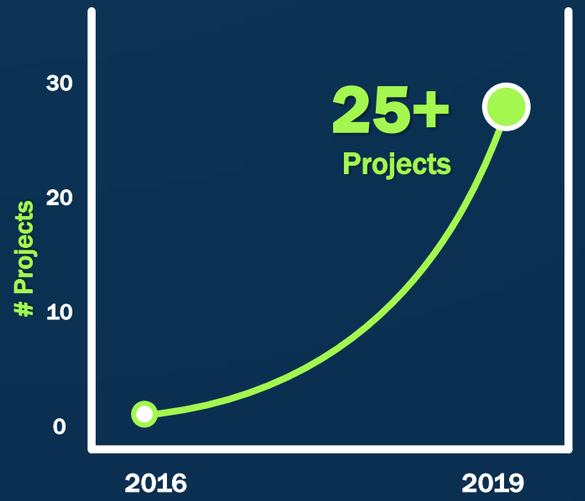
TEAM DEMOGRAPHIC



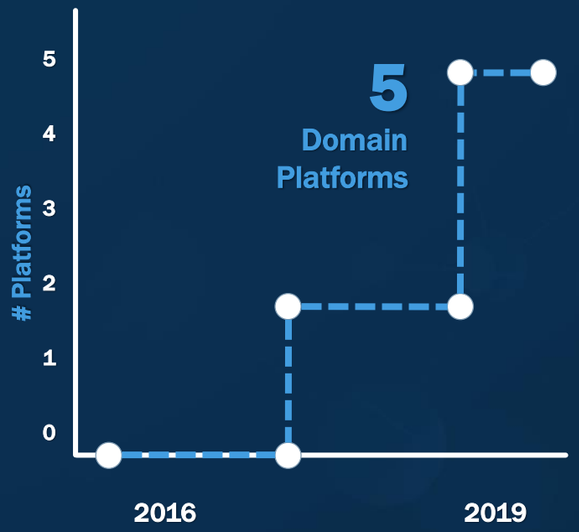
TEAM MEMBERS



PROJECT PORTFOLIO



PLATFORM DEPLOYMENT



2016 — INCEPTION OF **DATA SCIENCE SKILLS** IN APC
- Dr. Sean Gourley appointed to Board

2017 — **PRODUCTIZATION** STRATEGY DEVELOPED
- AAET Team Formed
- Strategic Alliance with RE Energy Group

2018 — **STAKEHOLDER ENGAGEMENT** FOCUS
- Kinetica Visualization Project Kick-Off
- APC Announces Google Partnership

2019 — **DEPLOYMENT** AT SCALE THROUGH **PLATFORMS**
- NVIDIA GPU Technology Conference

Operationalizing Digital

Exploration

*Identifying **sweet spots** where well performance is high and land entry costs are low can generate significant value to the company*

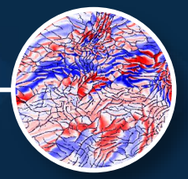
Development

*Selecting the **optimal well design** – which involves choices in numerous areas such as completion size and well spacing – requires predicting the performance for each candidate design*

Operations

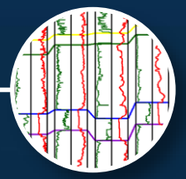
*Monitoring and understanding **asset behavior** through the life-cycle of well construction (drilling) to extraction of underground resources (production)*

ENHANCING SUCCESS IN DWGOM



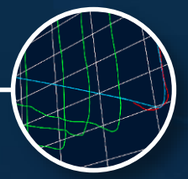
Seismic Interpretation

EXPANDING LOWER 48 FOOTPRINT

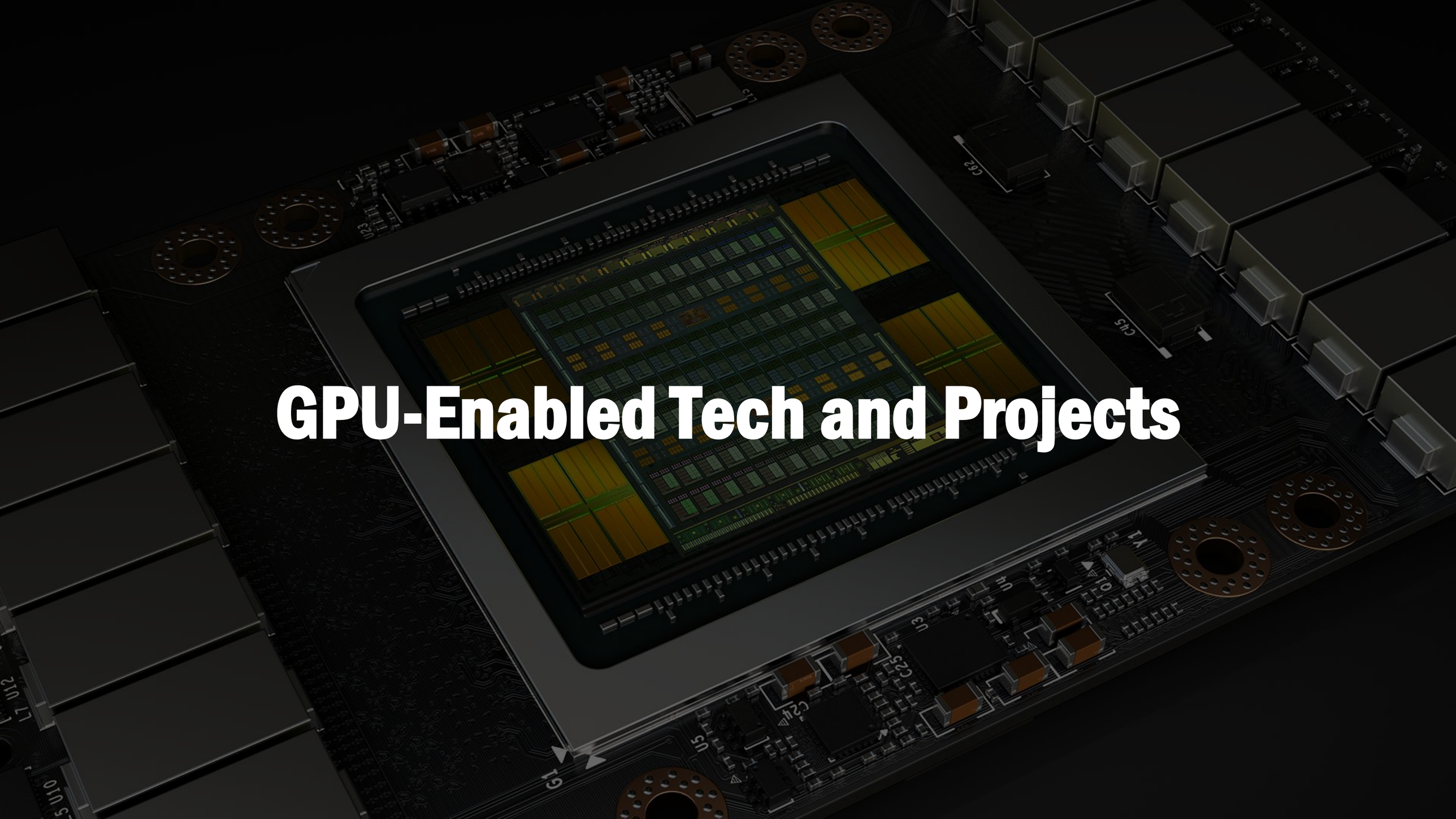


Stratigraphic Top Correlation

ENABLING DIGITAL OPERATIONS



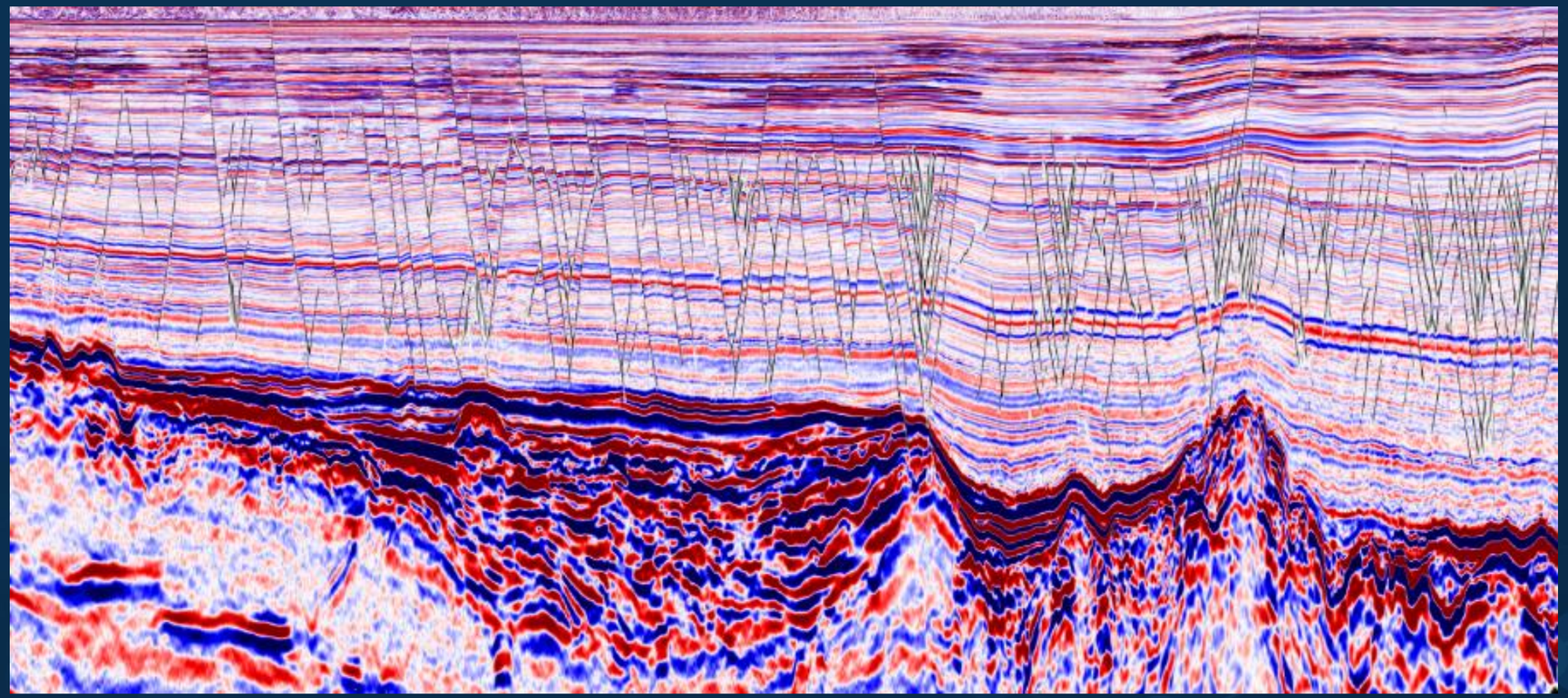
Real-Time Drilling



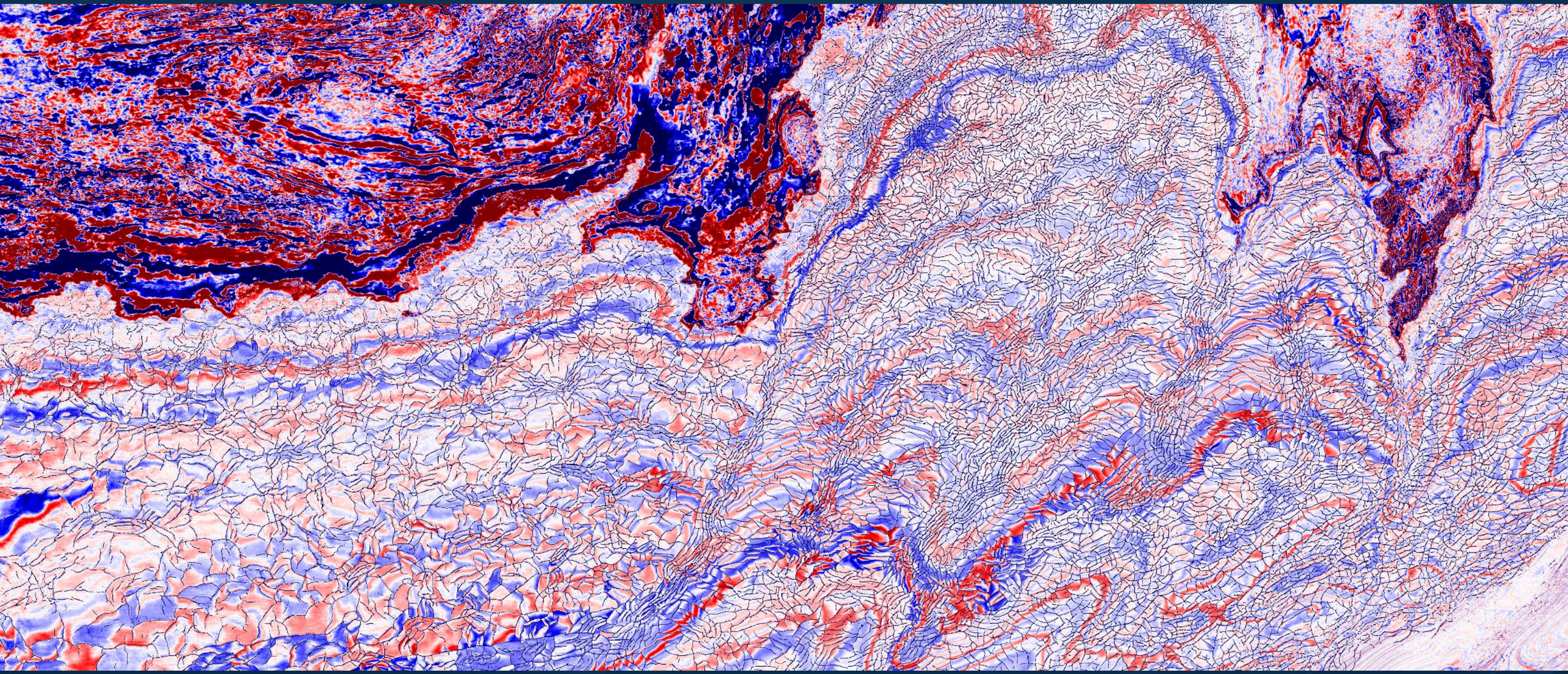
GPU-Enabled Tech and Projects

Seismic Interpretation

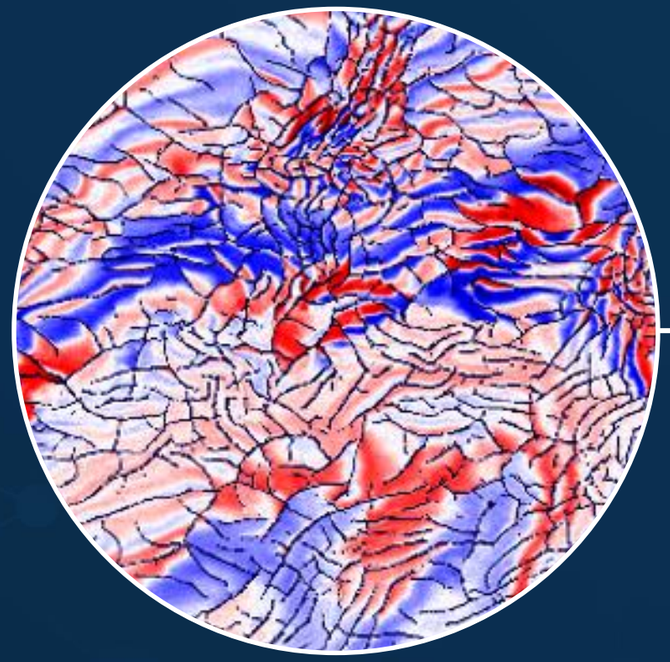
- **Project Scope**
  - Seismic interpretation deep neural network model for image processing
- **Data Volume**
  - 100s GB – several TB
  - 1000s images/attributes
  - Training on 1% data
  - Inference across full image
- **Framework**
  - TensorFlow/PyTorch
  - 2 concurrent fault prediction models
- **Environment**
  - 1.5 yr. ago: **DGX-1 8x Tesla P-100 GPUs**
  - Today:
    - **DGX-1 8x Tesla V-100 GPUs &**
    - **DGX-2 16x Tesla V-100 GPUs**



Seismic Interpretation



Seismic Interpretation



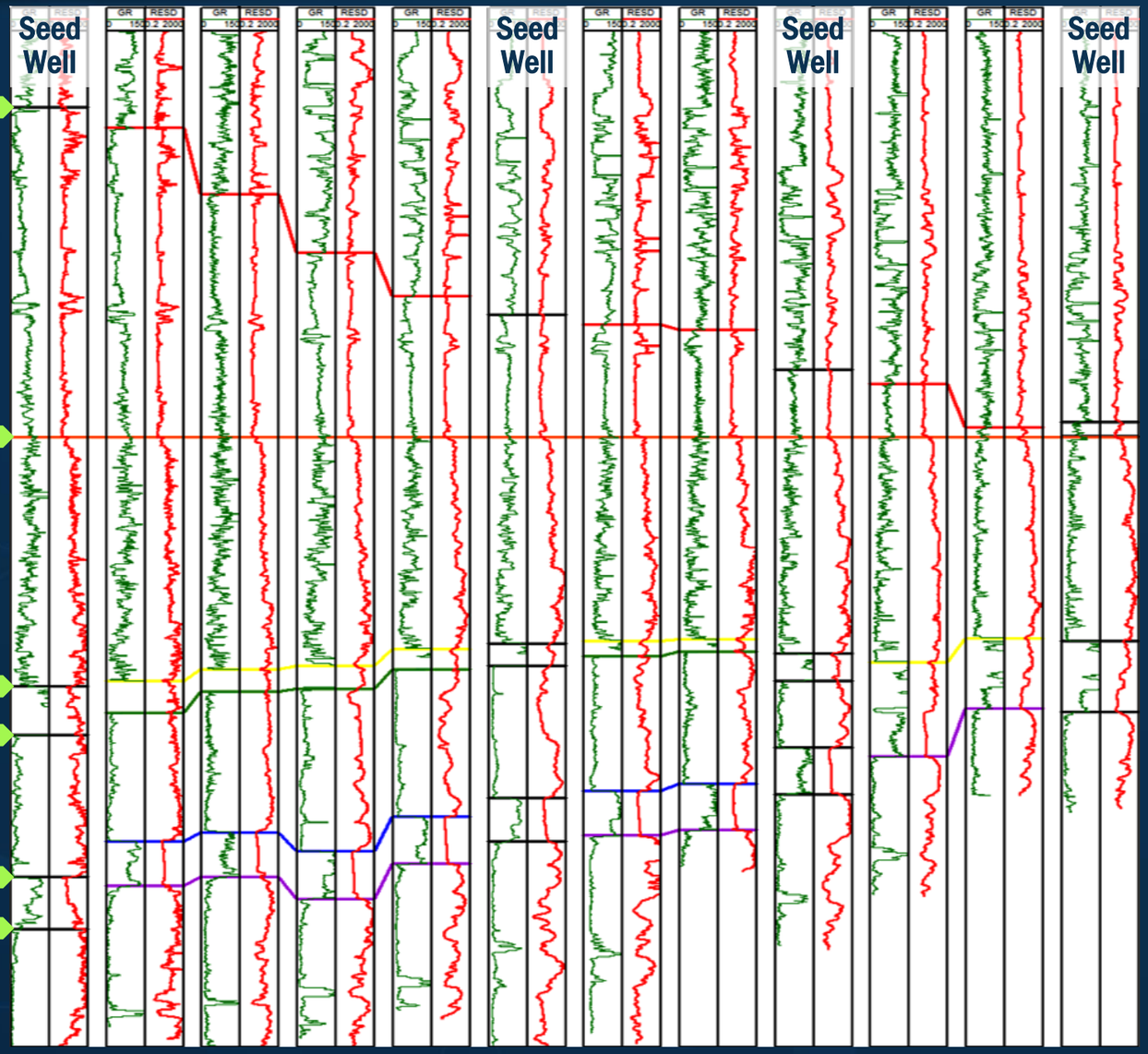
- **Benefits**
  - Training and inference

    1.5 yr. ago: **~20 hours** Today: **<10 hours**
- **Challenges**
  - Time intensive training process
  - Loading data into GPU memory
- **Next Steps**
  - Network enhancements
  - Workflow improvements
  - New DGX-2 box:
    - **16x V-100 GPUs + 512 GB GPU Memory**
  - Future Environment:
    - Google Cloud Platform (GCP)

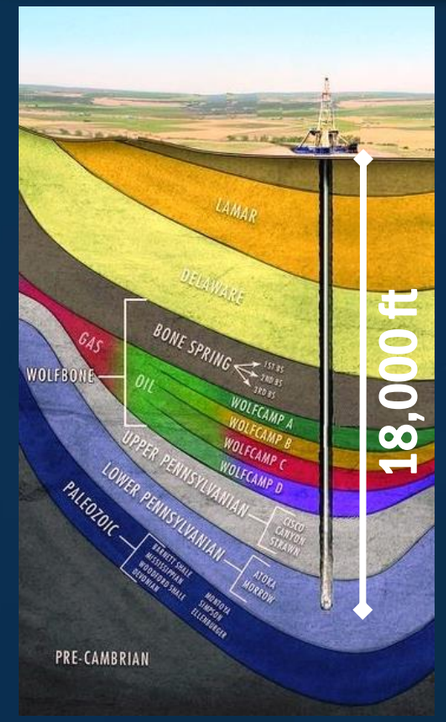
Currently, the models only use 1% of the data for training, however, inference is performed on the entire image.

This means that we need to dedicate significant amounts of time to training in order to deliver good inferences.

Stratigraphic Top Correlation

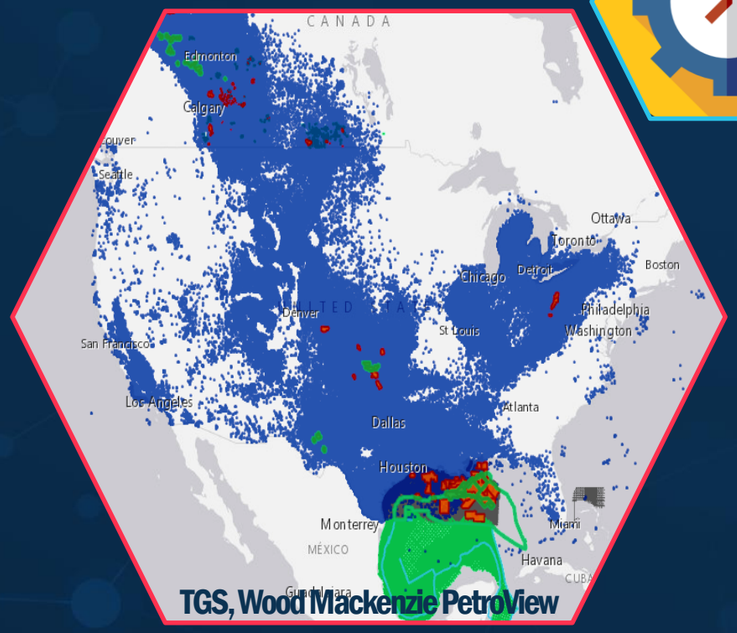


- **Project Scope**
  - Learn from identified **tops** propagate at basin scale
- **Data Volume**
  - Training ~25GB
  - Inference: Size varies, on the fly
- **Framework**
  - CNN in TensorFlow
- **Environment**
  - Dev/Train (on prem.):
    - ***DGX-1 8x Tesla V-100 GPUs***
  - Inference/UI (on cloud):
    - ***GCP V-100 and T4 GPUs***



Stratigraphic Top Correlation

- **Massive data volumes**
  - Rapidly growing geo data
  - Picked and inference wells
  - Expert picking (labeling)



- **Time intensive training process**
  - CNN training
  - GPU tech advances

- **Faster training with new GPU chips**
  - GPU Quadro P6000: **couple weeks**
  - DGX-1 8x Tesla P-100: **1.5 – 2 days**
  - DGX-1 8x Tesla V-100: **< 24 hours**
- **Accelerated basin evaluation process**

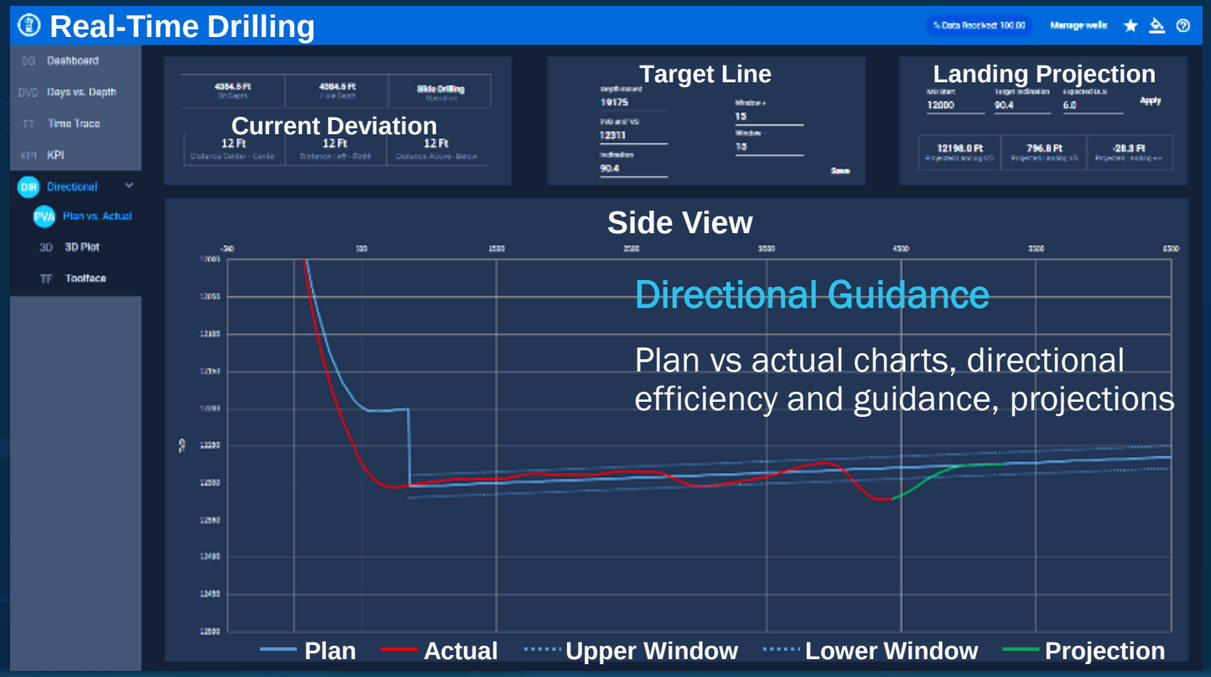
- **Short term**
  - Better networks as CNN runs
  - ‘Self-tuning’ mechanism
- **Long term**
  - Move workflow to cloud
  - New **T4 GPU** for inference
    - In GCP since Jan 2019

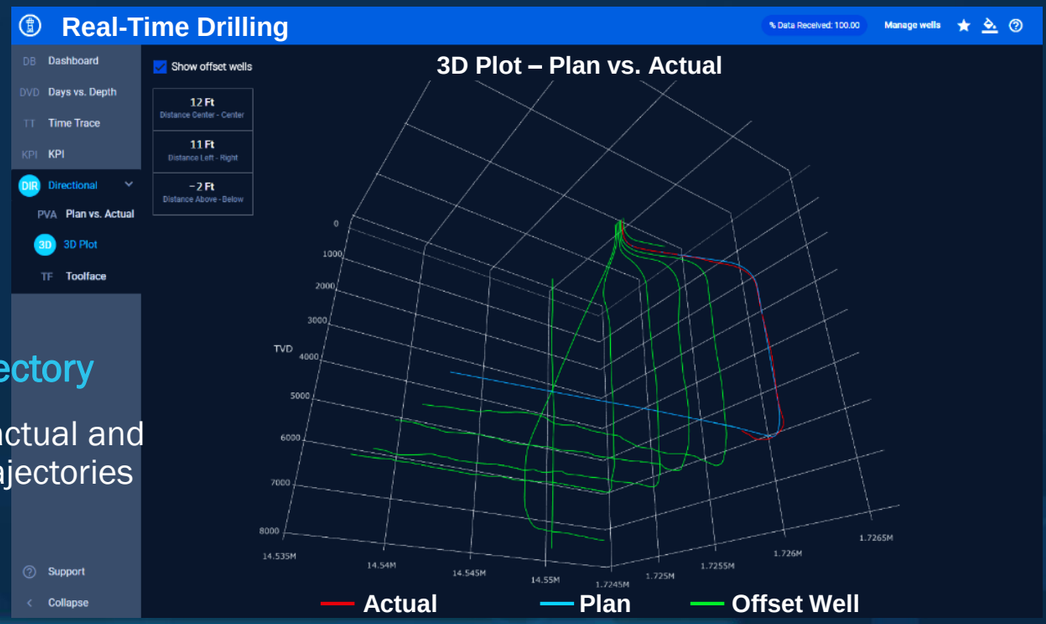
Real-Time Drilling

- **Project Scope**
  - Drilling Ops: $M decisions
  - **Analytics and DL models** process real-time streaming log data & other non-streaming data
  - Rig states → Derive **operational KPIs** of drilling ops at very high resolution

Operational KPIs at high resolution

Footage, Time, ROP, connection statistics





Real-Time Drilling

- **Data Volume**
  - Training dataset ~5GB
  - Inference: Real-time streaming
    - *Sliding window partition data stream*
    - *Runs online 24/7*
- **Framework**
  - RNN: > training, no parallelization
  - CNN: Current model, TensorFlow
- **Environment**
  - Dev/Train (on prem.):
    - ***DGX-1 8x Tesla V-100 GPUs***
  - Inference (on cloud):
    - ***Google Cloud ML Engine***

Benefits

- Very light for inference
- High res. KPIs to evaluate drilling performance and correct trajectories

Challenges

- Real-time inf. requires fast response
  - **<100 millisecond** for each inference
  - **>1 sec** ⟶ heavy traffic, potential jams

Next Steps

- Short-term
  - Offline model using historical data
  - **Divergence Detect:** Real-time vs offline
- Long-term
  - More complex models, more data

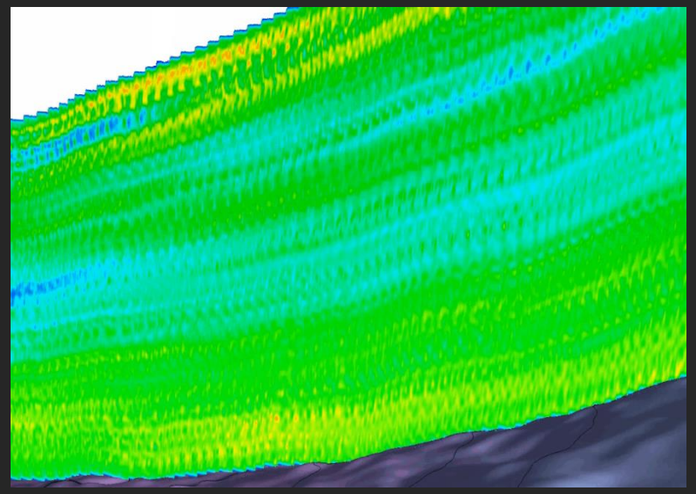
Amit Vij
President & Co-Founder
Kinetica

“NVIDIA and Kinetica have enabled us to do the impossible — render a high fidelity, 3D view of an oil basin using 100 billion data points at scale.”

Sanjay Paranji, CTO at Anadarko Petroleum Corporation

PASSIVE ANALYTICS

BEFORE GPUS

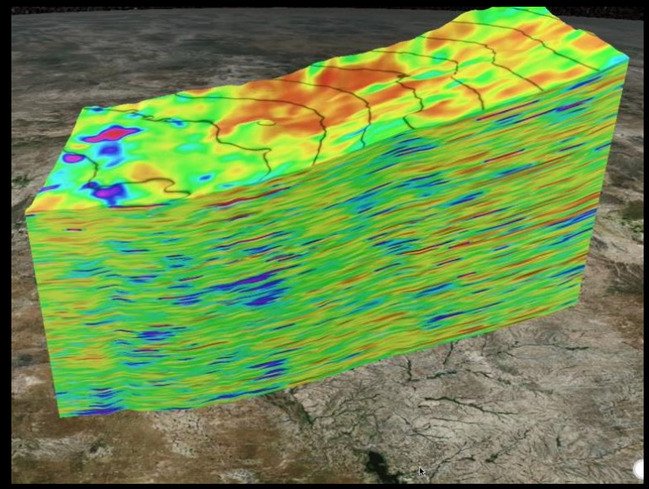


9M

Low Fidelity / Small Sample Set

ACTIVE ANALYTICS

AFTER GPUS

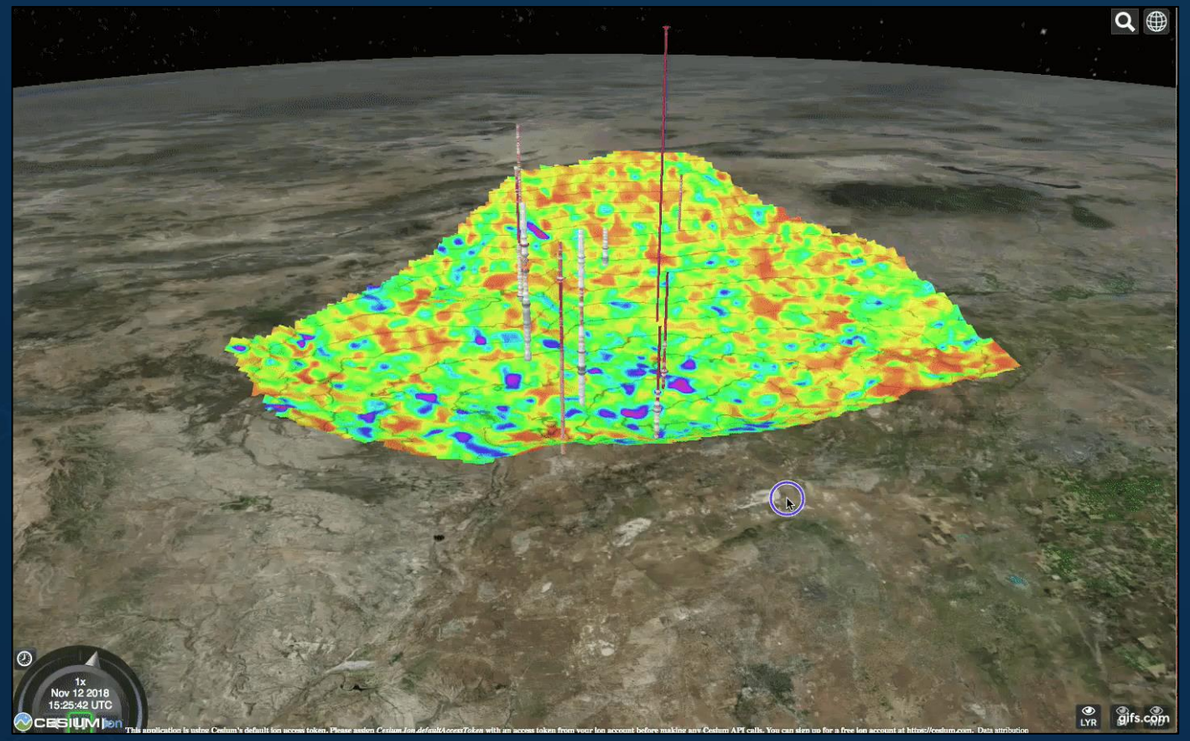


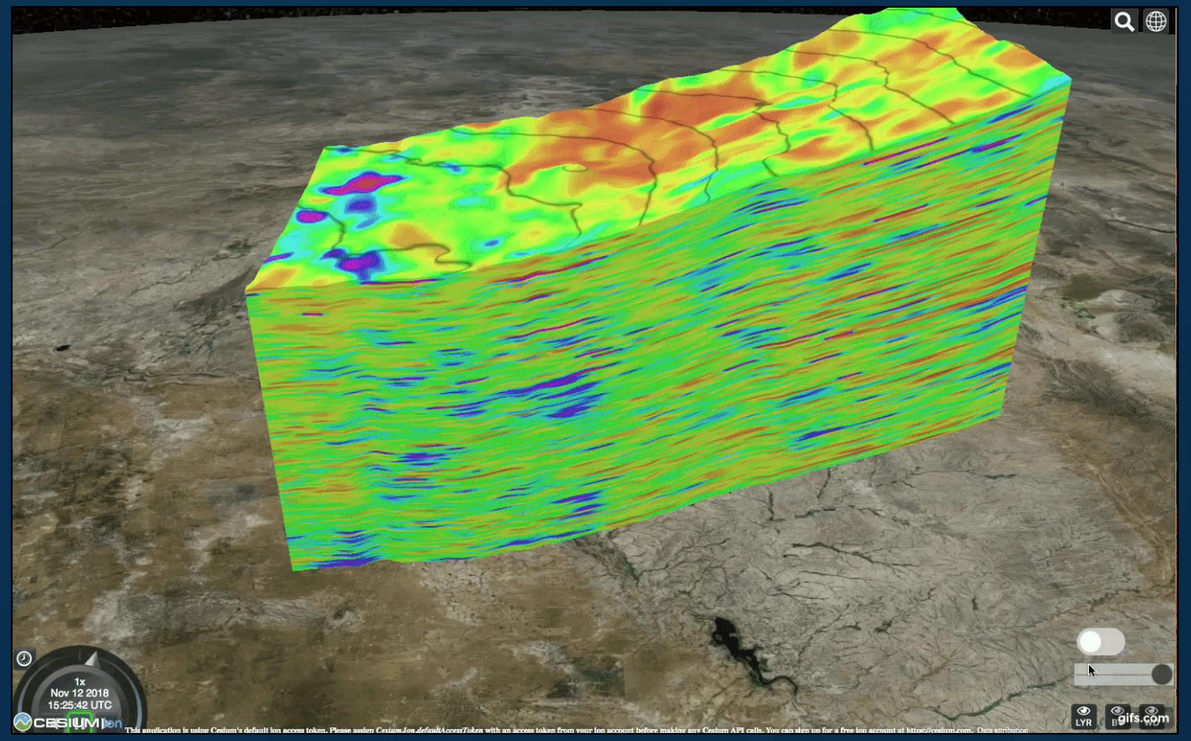
90B

High Fidelity / Full Data Set + Streaming

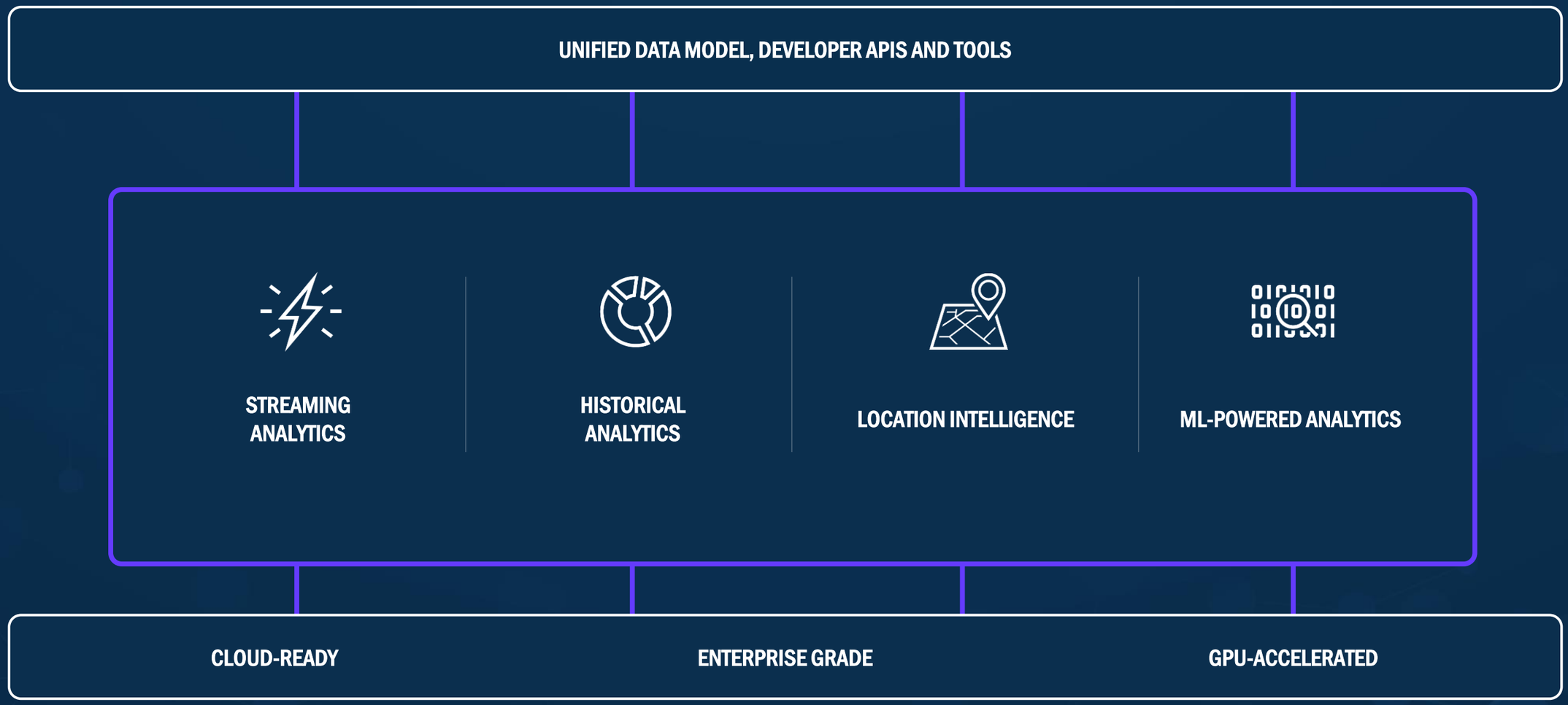
Anadarko

High Fidelity 3D Visualization



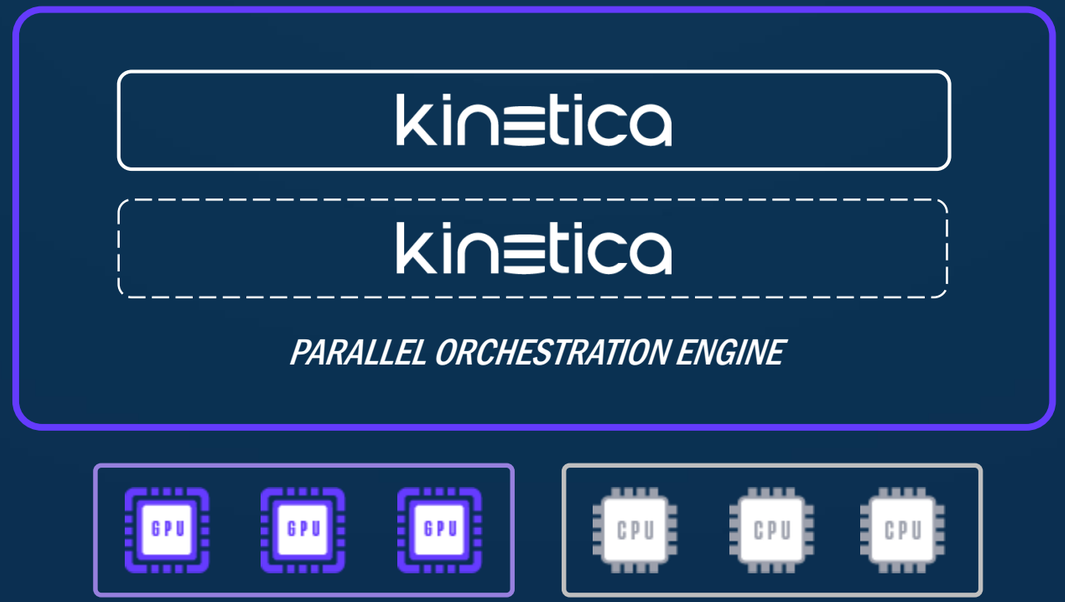


Kinetica Active Analytics Platform



In Your Ecosystem

GPU-Accelerated Database



In-Memory OLAP Database

- Distributed
- Columnar
- Vectorized
- SQL 92 Compliant

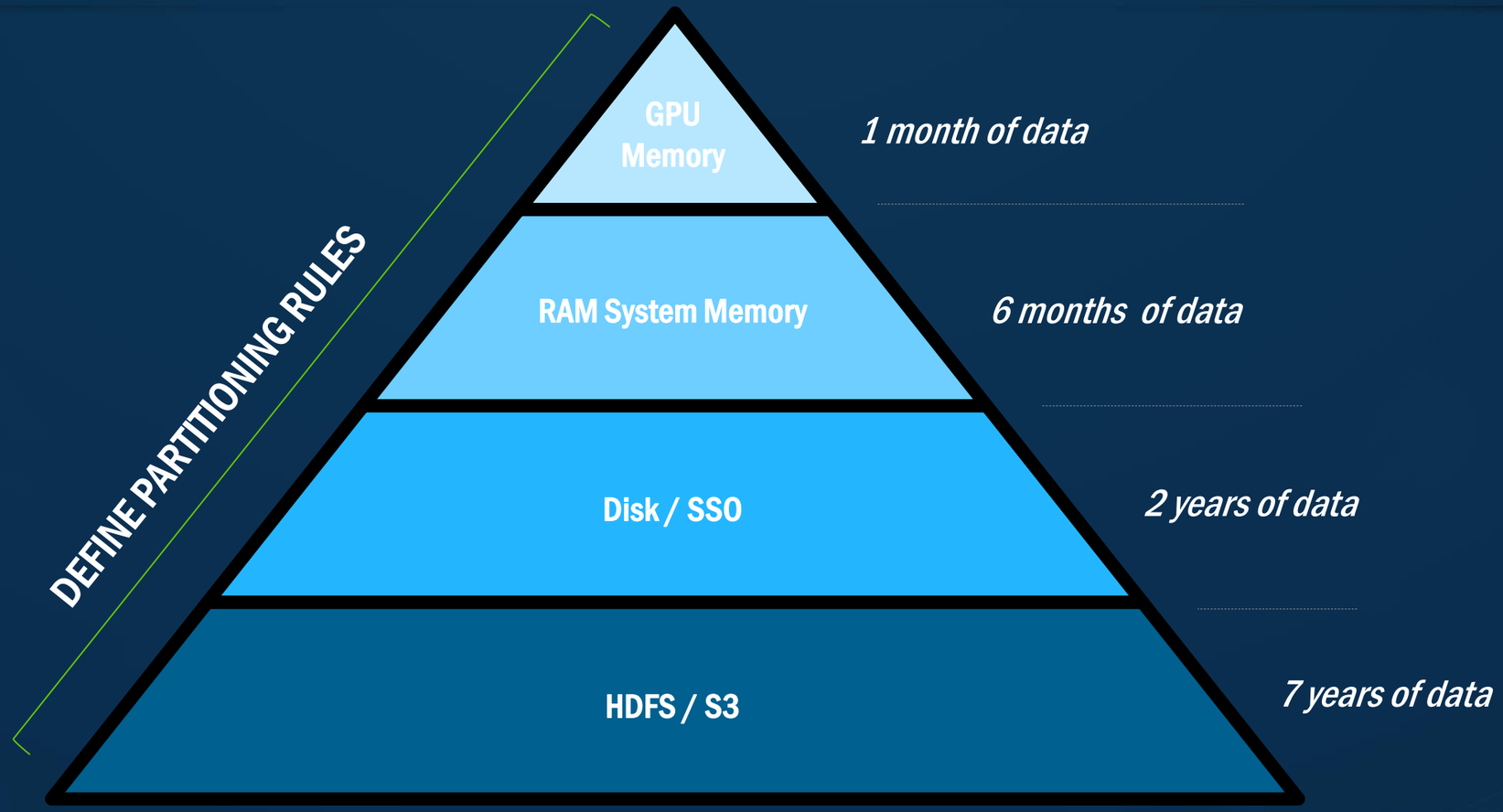
Wide Range of Analytical Techniques

- Full text search
- Time series analysis
- Location intelligence
- Graph analytics

Enterprise Scale Tiered Storage

- Access entire data corpus
- Optimized performance

Kinetica TIERED STORAGE



Data movement across tiers automatically managed by Kinetica

NEW IN 7.0

Location Intelligence and Visualization

kinetica

DISTRIBUTED GPU PIPELINE

3D TILES

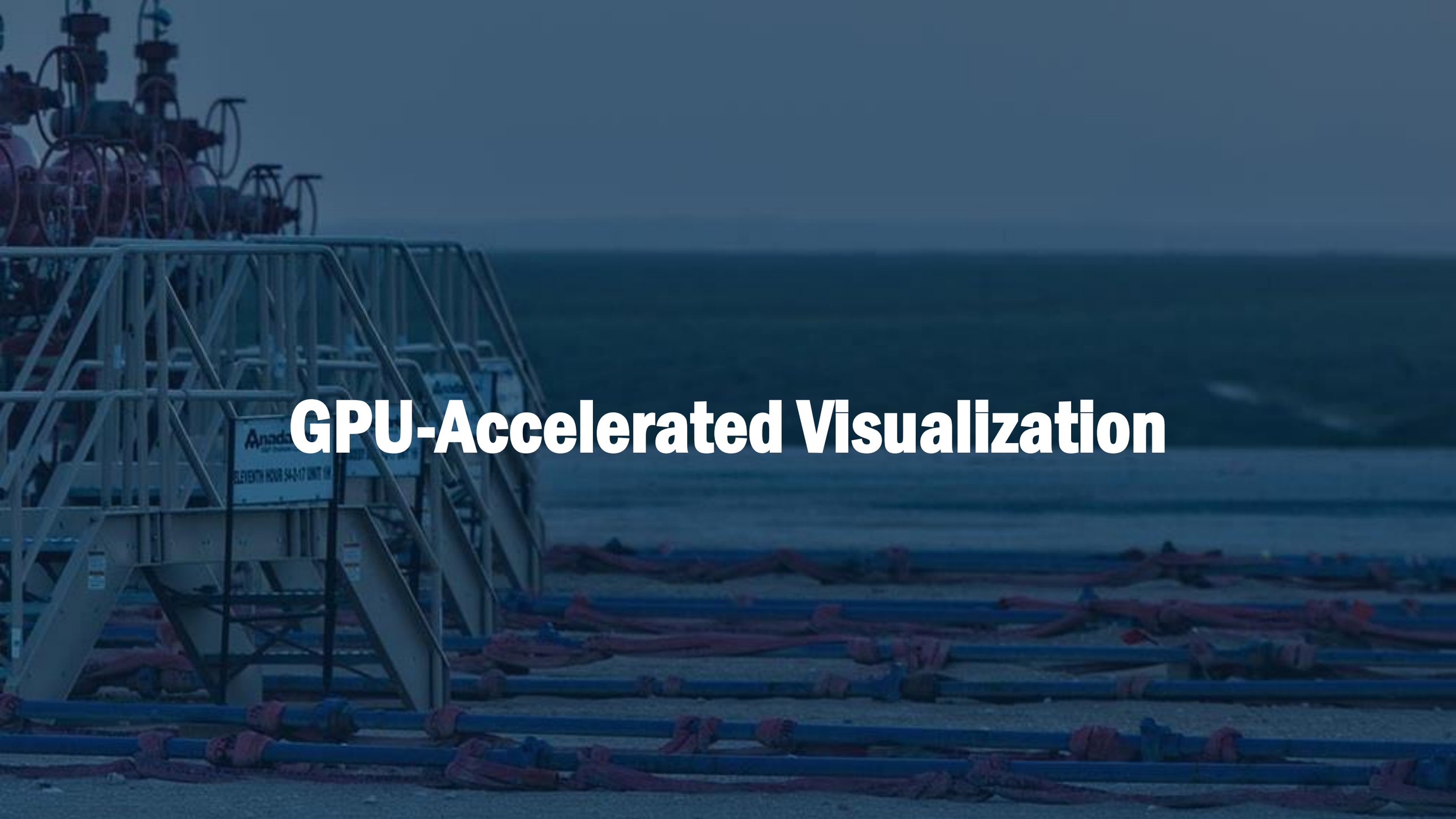
GEOSPATIAL APIS

GEOSPATIAL SERVER

GRAPH ANALYTICS

COMPLEX GEOSPATIAL OPERATIONS

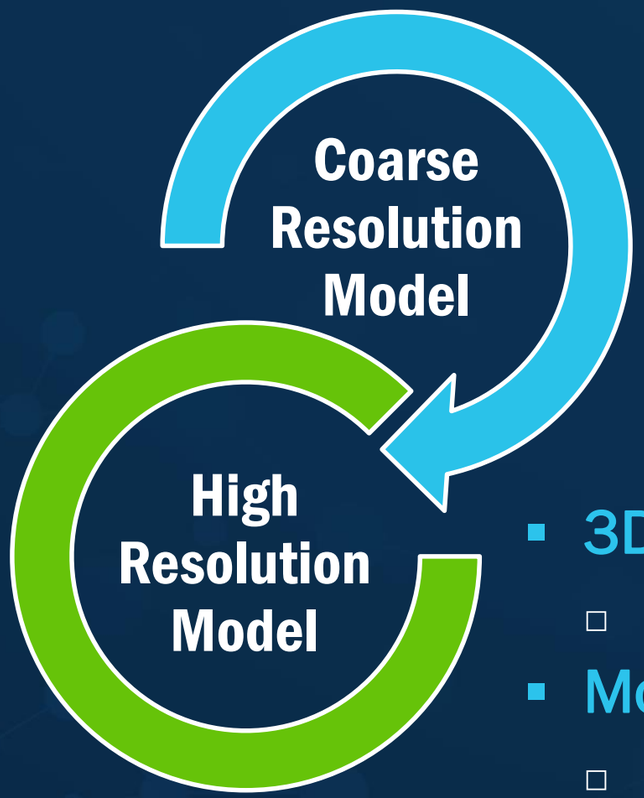
What We're Doing Today



GPU-Accelerated Visualization

Motivation

- **Two-Stage Reservoir Modeling Strategy**



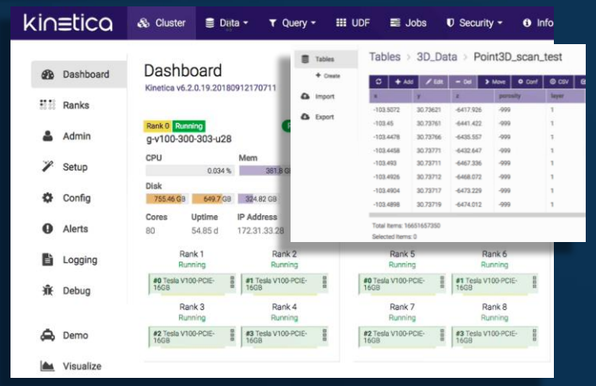
  - Large AOI
    - Basin extent
  - Subset Data
    - Reduced attributes
  - Rendering & processing on CPU
  - 3D Region of Interest
    - Reduced coverage
  - Model at Scale
    - Repeat process in neighboring areas

- **Challenges**

  - Time-consuming
    - Geologists & reservoir modelers performing tasks in sequence
  - Incompatibility between models
    - Conditional to different inputs (defined by AOI/ROI)
  - Incl. relevant info. in decision-making
    - New workflows require viz & rendering of massive data volumes

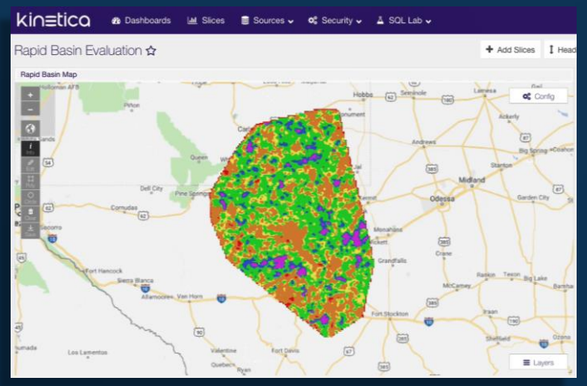
Approach

Build High-Res Model



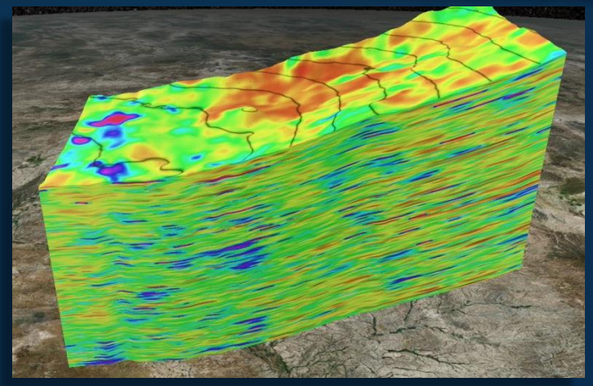
- **Size & Dimensions**
  - 90 billion points
    - ~4k slices (layers)
    - ~24 mill. pts/layer
  - > 10,000 sq. mi.
  - 100ft XY (spatial) resolution
  - 1ft Z (depth) resolution

Visualize Data in 2D

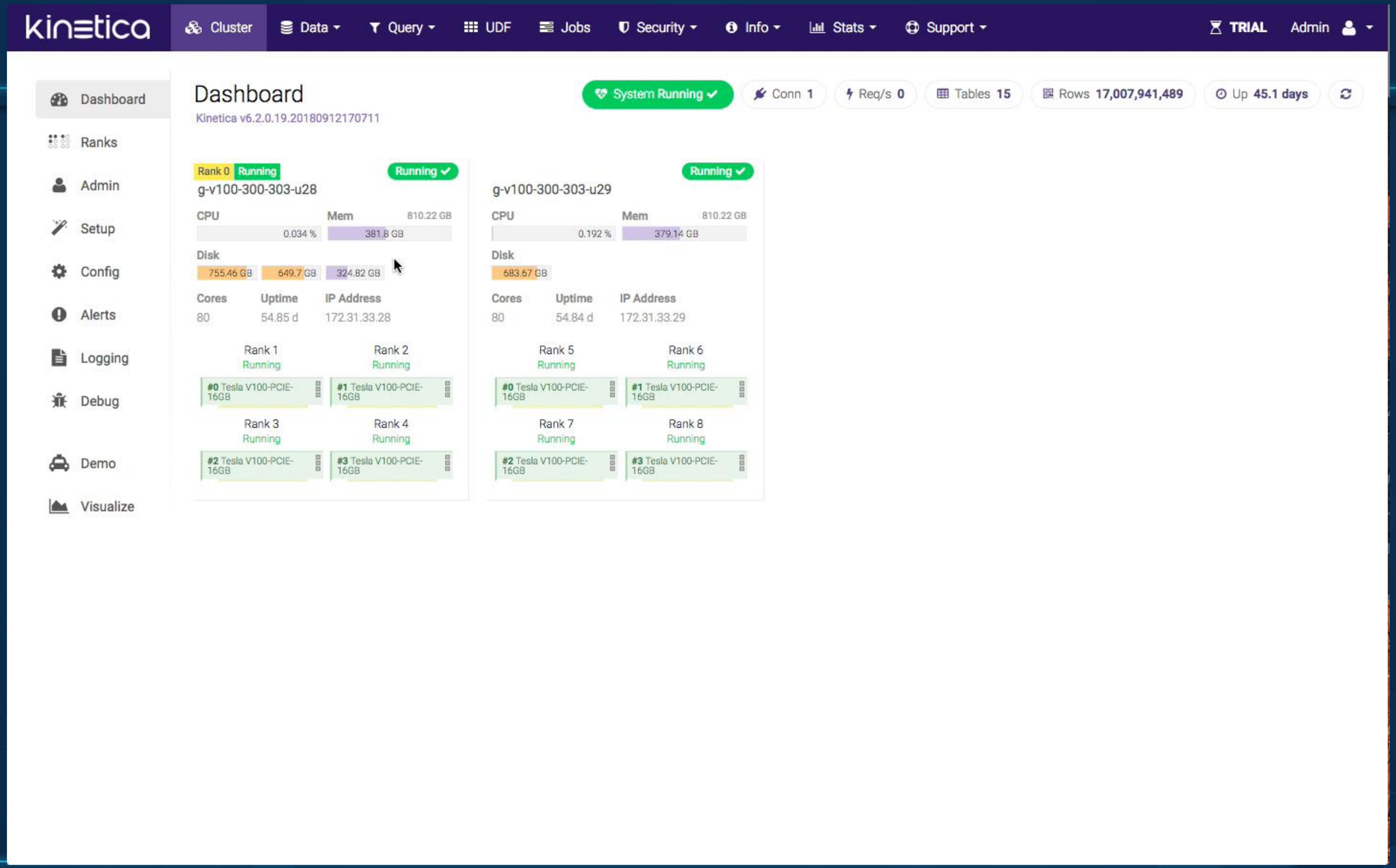


- **Data Ingestion**
  - 6TB of point data
- **Reveal**
  - Kinetica's viz framework as end-user web client
  - Layers draped over base map

Visualize Data in 3D

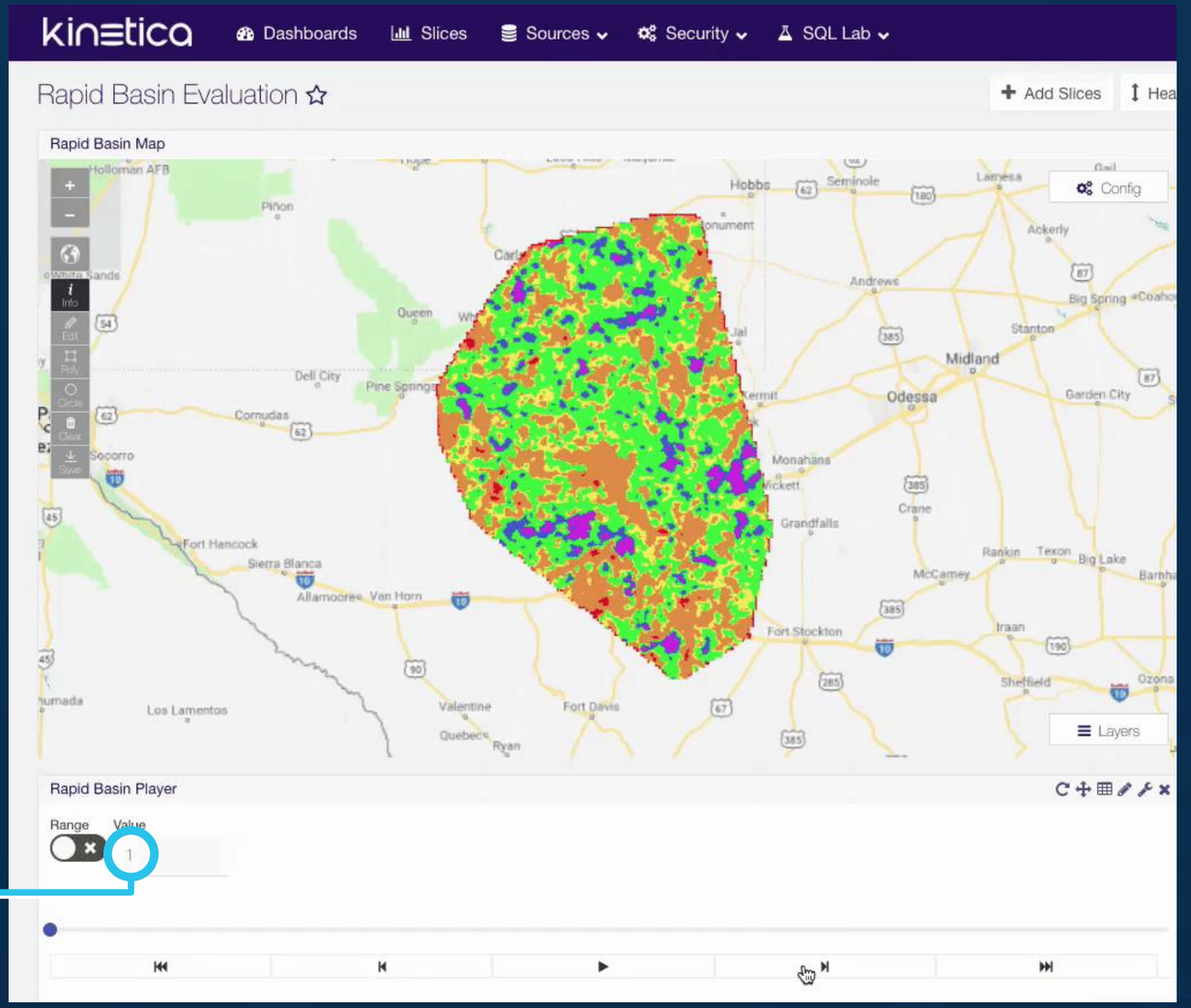


- **User-Defined Functions (UDF) Framework**
  - 3D tiles on the fly
  - Representation of reservoir model
- **CesiumJS**
  - Geospatial 3D mapping platform to render tiles


kinetica
Cluster
Data
Query
UDF
Jobs
Security
Info
Stats
Support
TRIAL
Admin
Dashboard
Ranks
Admin
Setup
Config
Alerts
Logging
Debug
Demo
Visualize
Dashboard
Kinetica v6.2.0.19.20180912170711
System Running
Conn 1
Req/s 0
Tables 15
Rows 17,007,941,489
Up 45.1 days
Rank 0
Running
Running
g-v100-300-303-u28
CPU
0.034 %
Mem
810.22 GB
381.8 GB
Disk
755.46 GB
649.7 GB
324.82 GB
Cores
80
Uptime
54.85 d
IP Address
172.31.33.28
Rank 1
Running
#0 Tesla V100-PCIE-16GB
Rank 2
Running
#1 Tesla V100-PCIE-16GB
Rank 3
Running
#2 Tesla V100-PCIE-16GB
Rank 4
Running
#3 Tesla V100-PCIE-16GB
Running
g-v100-300-303-u29
CPU
0.192 %
Mem
810.22 GB
379.14 GB
Disk
683.67 GB
Cores
80
Uptime
54.84 d
IP Address
172.31.33.29
Rank 5
Running
#0 Tesla V100-PCIE-16GB
Rank 6
Running
#1 Tesla V100-PCIE-16GB
Rank 7
Running
#2 Tesla V100-PCIE-16GB
Rank 8
Running
#3 Tesla V100-PCIE-16GB

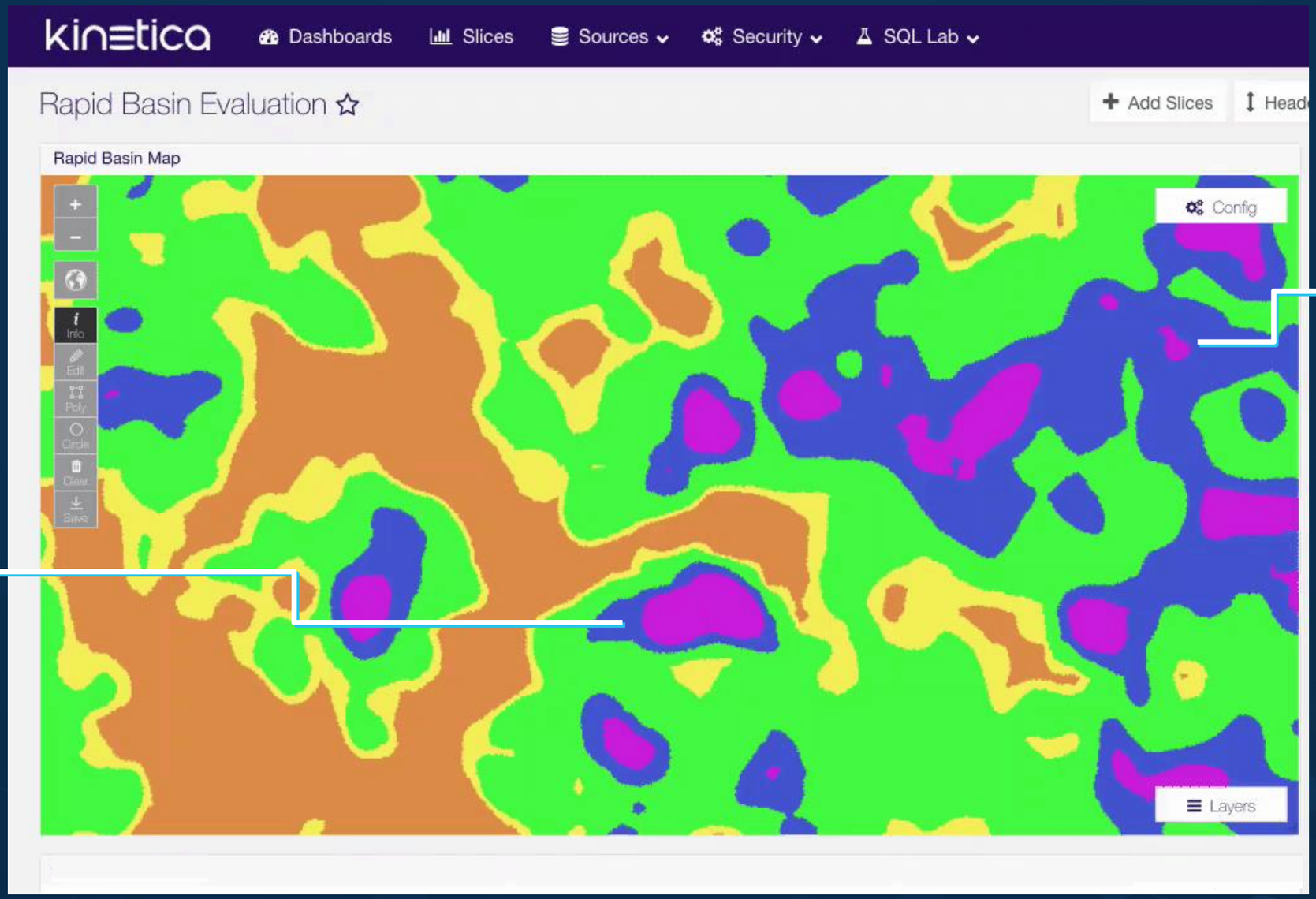
Results

- **Reveal dashboard with model loaded and points geospatially referenced**



Rapid navigation through ~4k model layers

Results

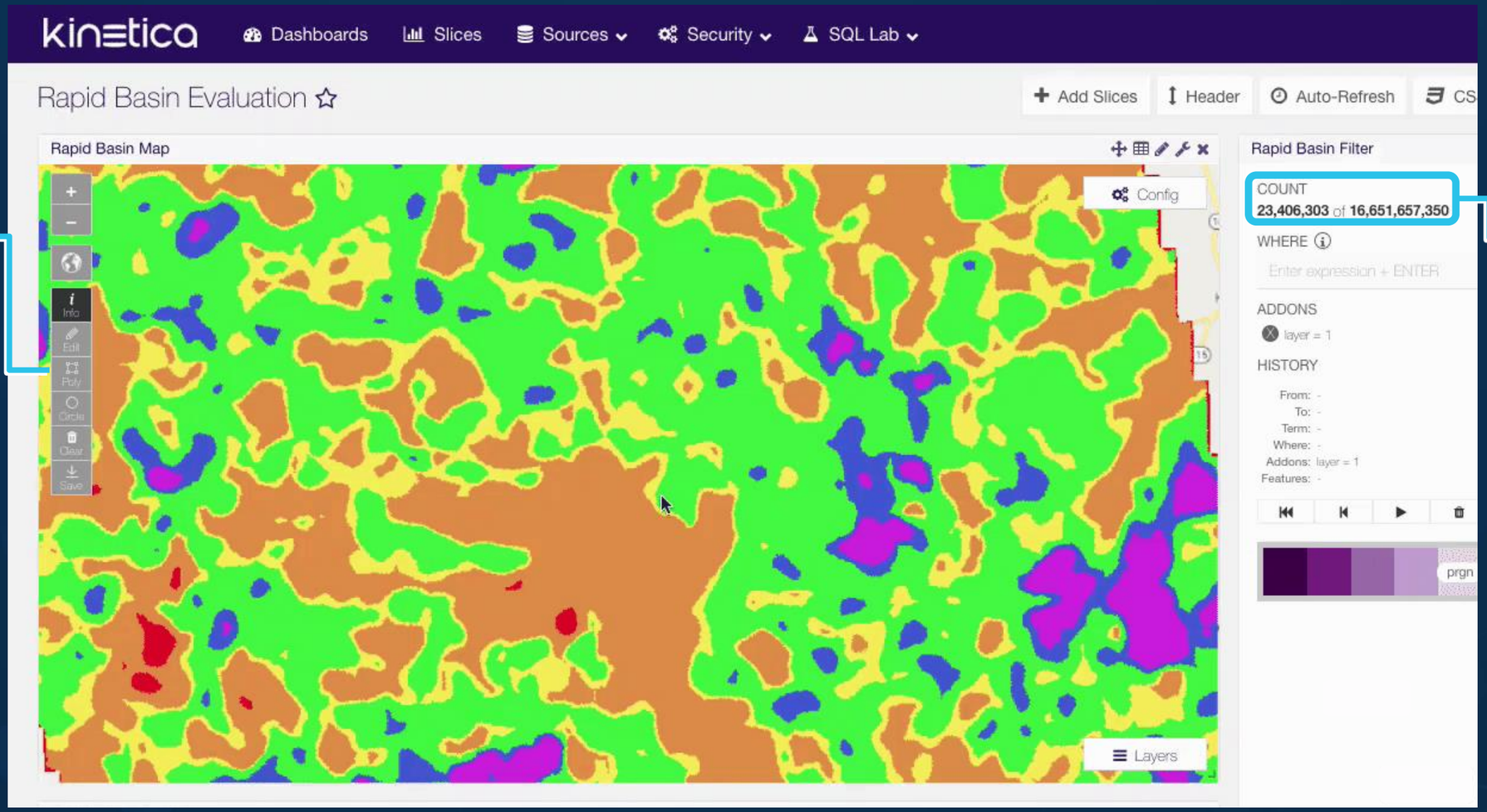


Point resolution varies according to zoom level

Map property (porosity) can be interrogated at any point location

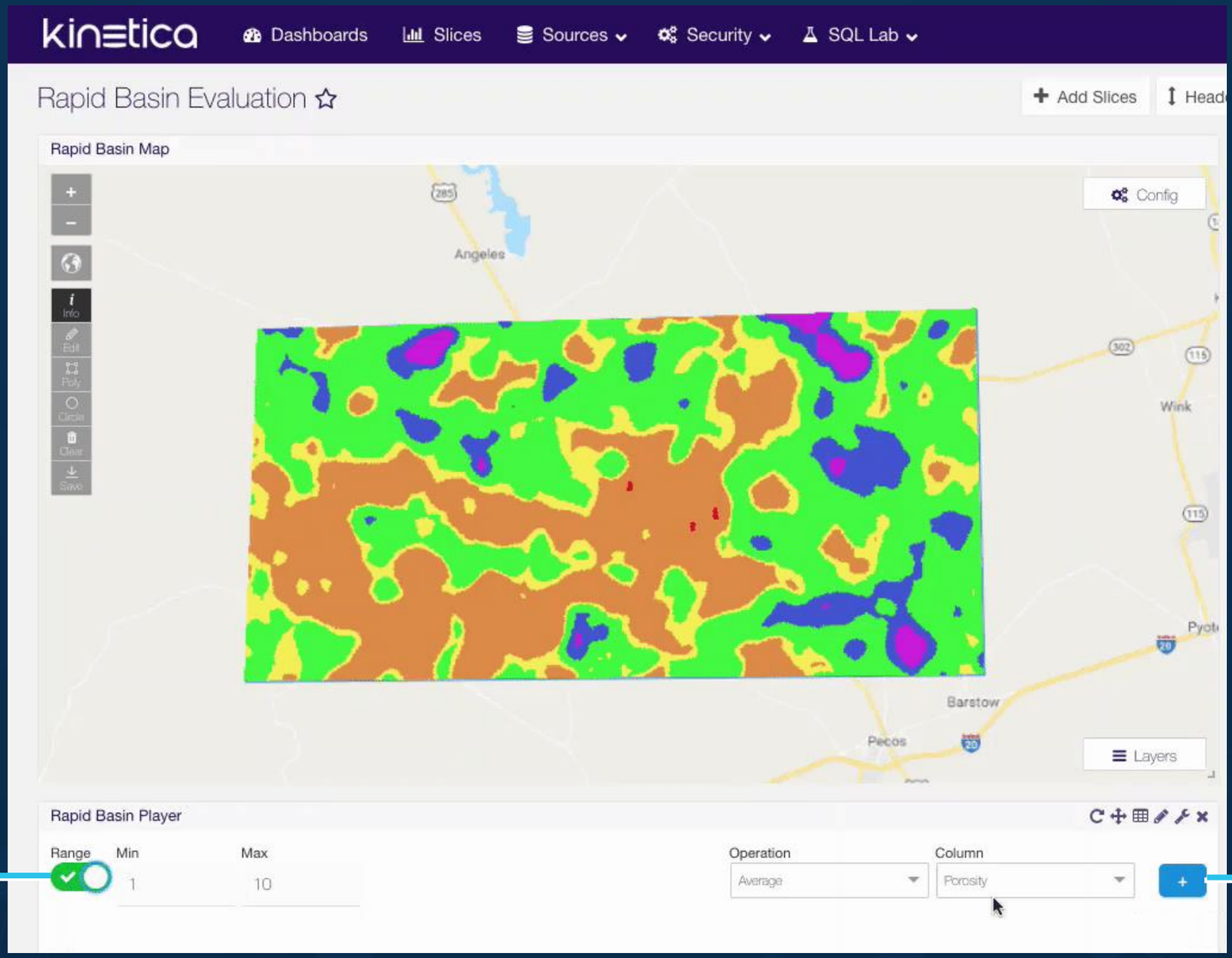
Results

Select and filter records by spatial extent (drawing a shape)



Point count updates when spatial filter is applied

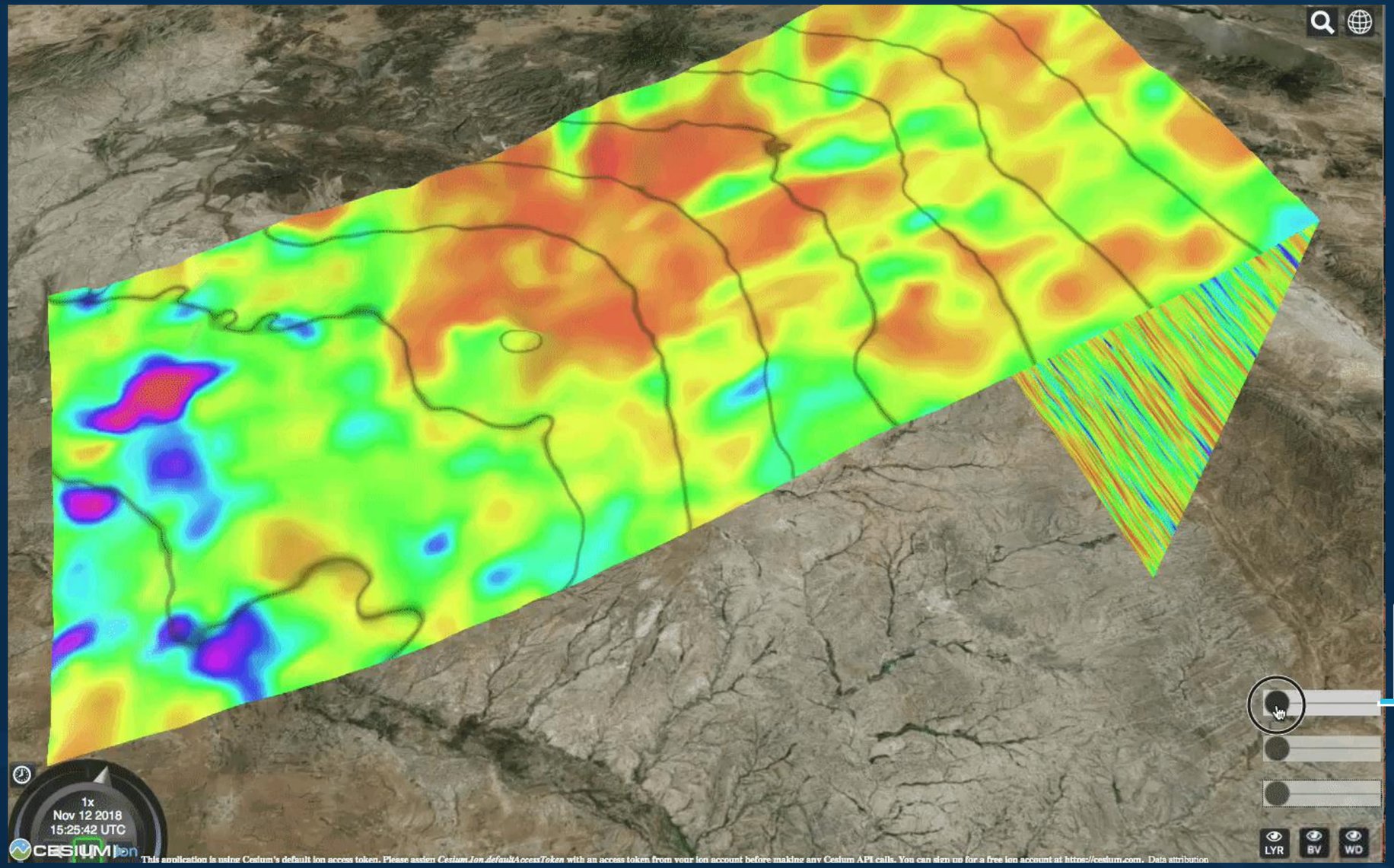
Results



Select a range of layers (1-10)

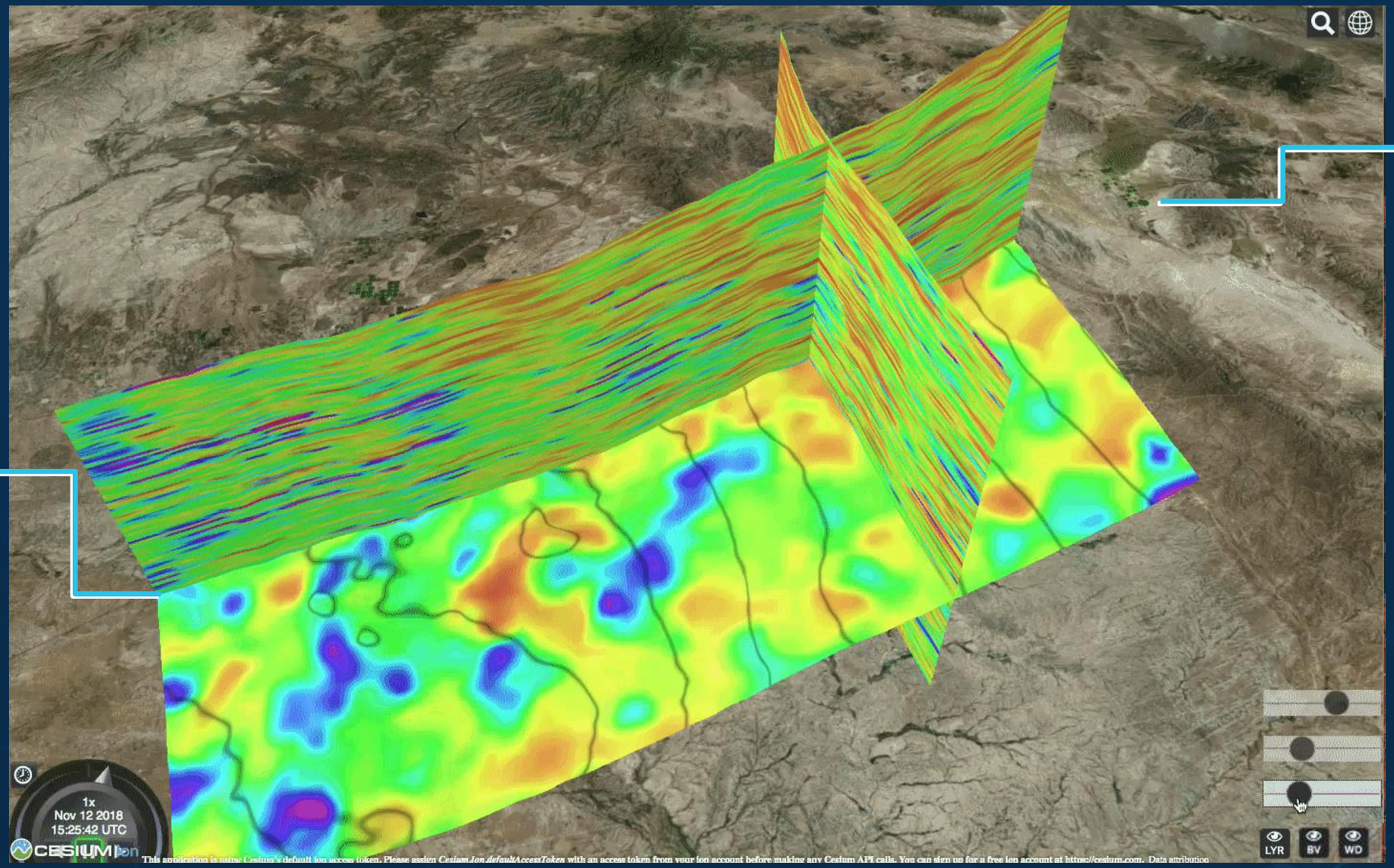
Quick execution of calculations on layer range and generation of derivative data (e.g. Calculate average porosity)

Results



Slider bars enable smooth movement through horizontal slices of the model

Results

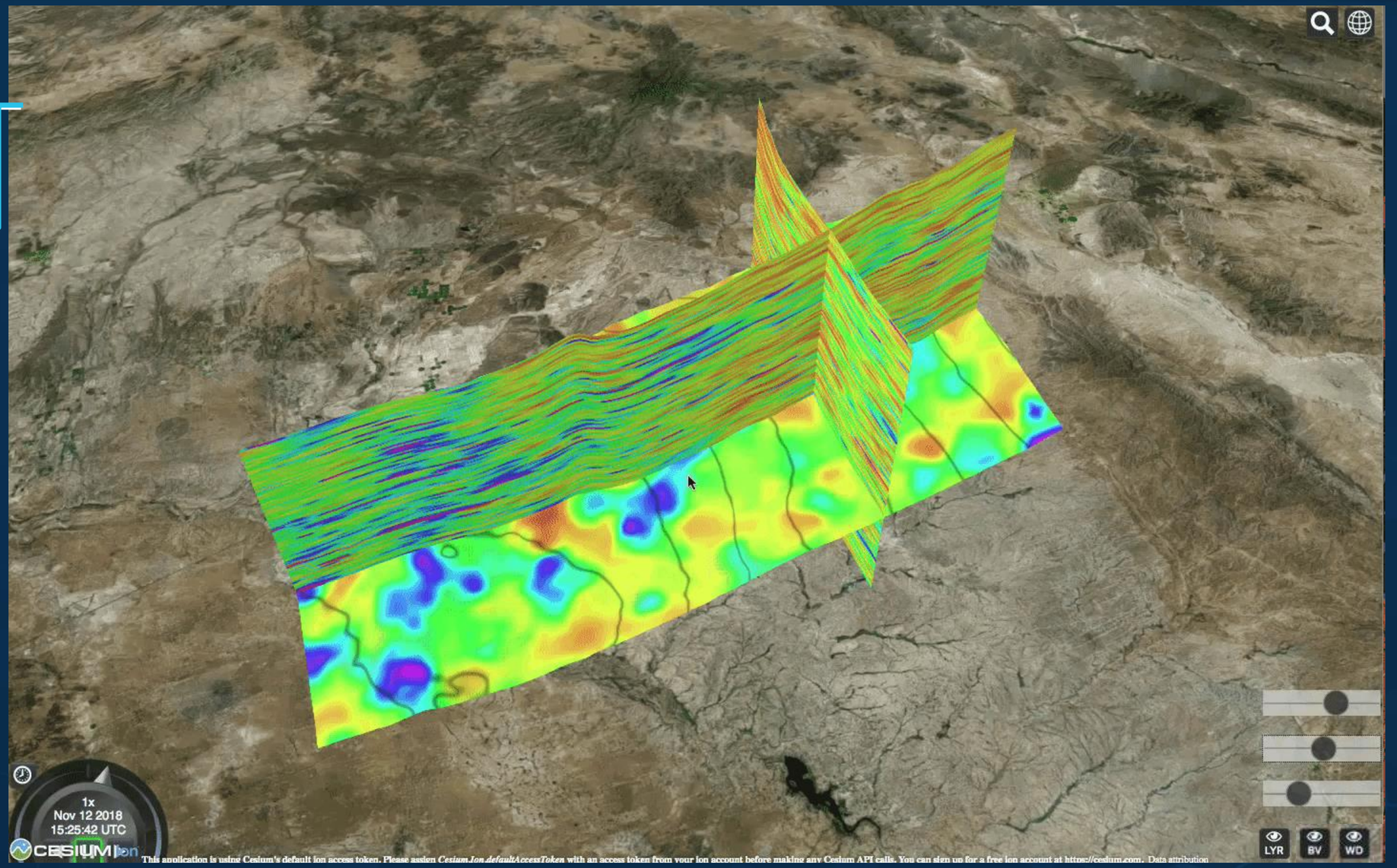


Full model volume is cut by intersecting cross-sectional slices

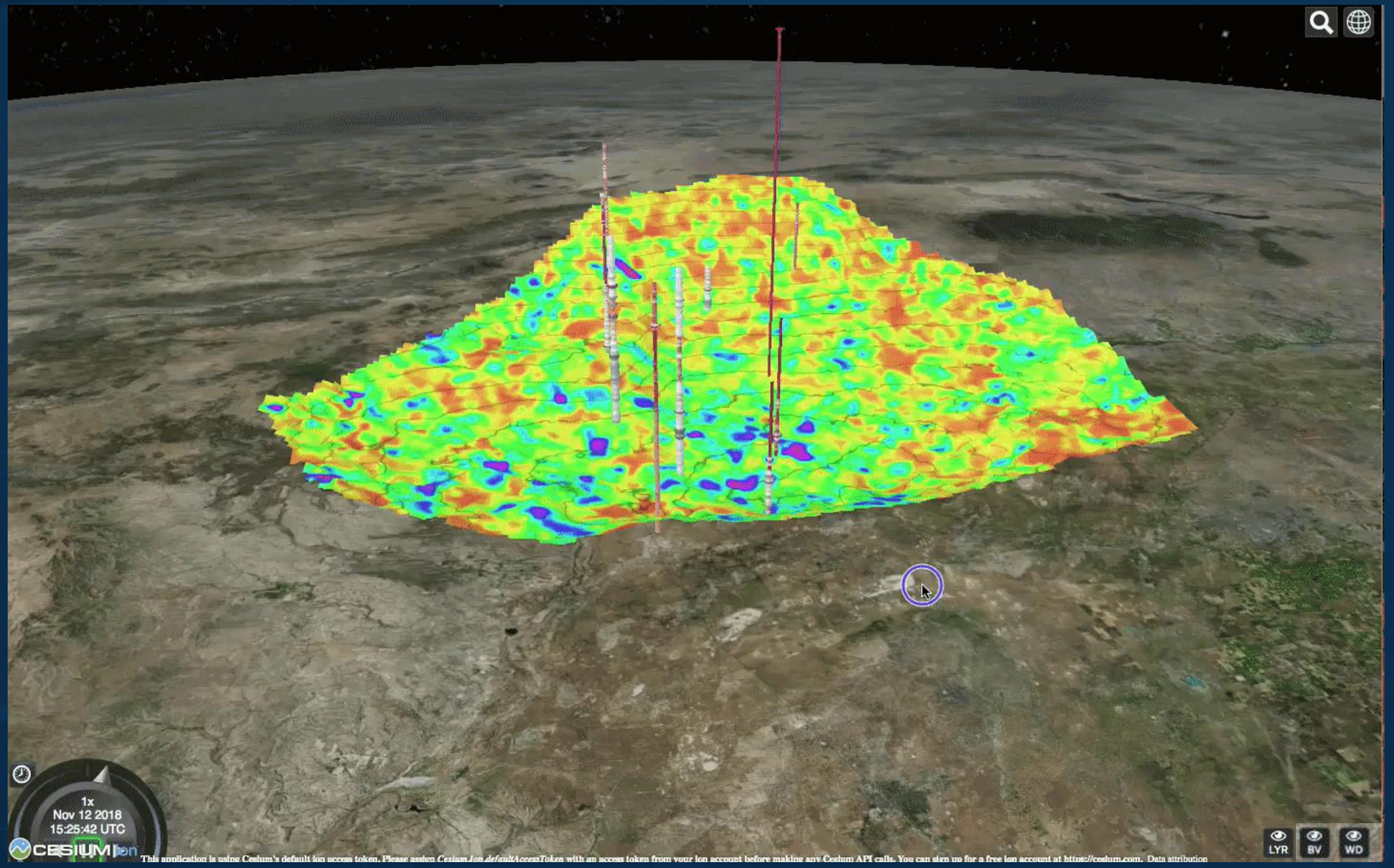
High resolution of model maintained in 3D view

Results

Map property can be interrogated by selecting a point of interest



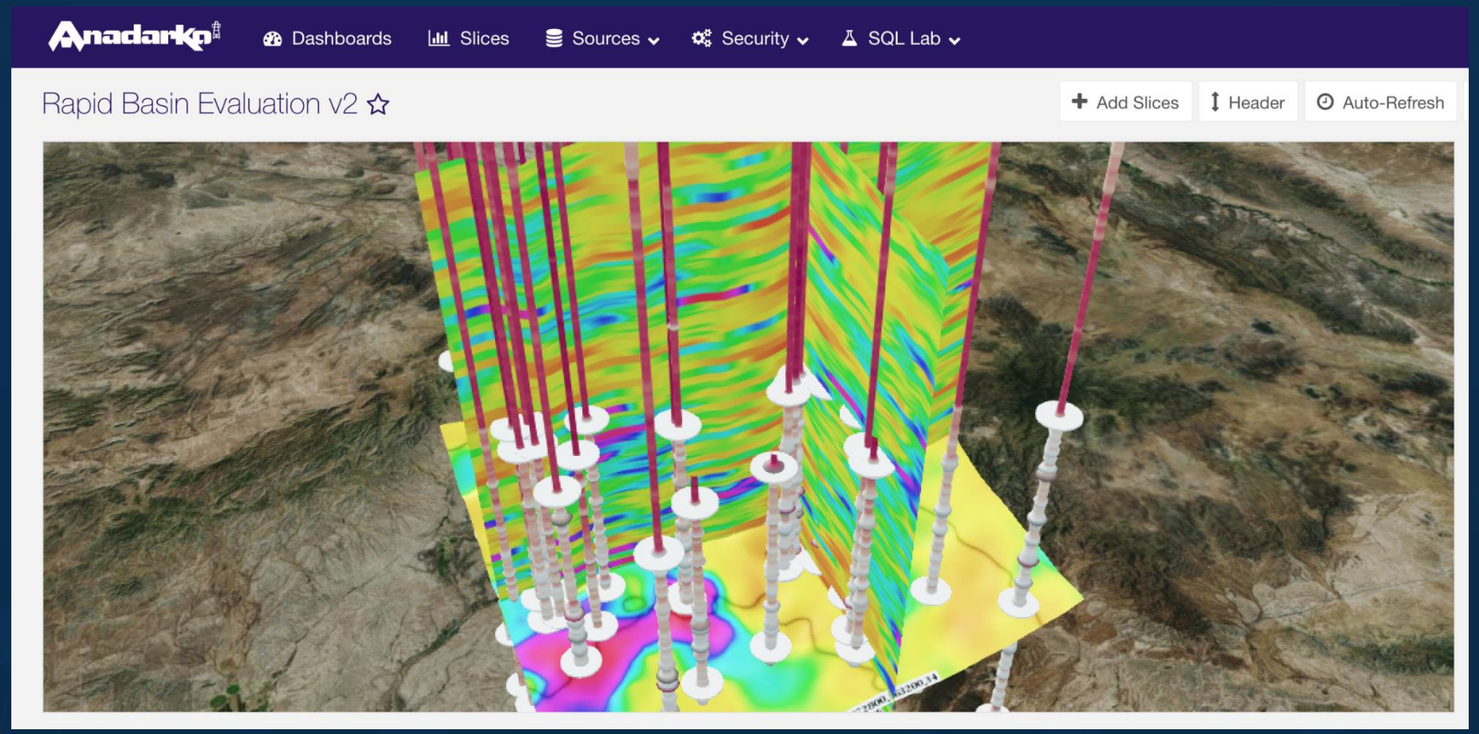
Next Steps



- **Interaction**
  - 2D & 3D well logs
  - More models & petroph. attributes
  - Complex and stored calculations
- **Visualization**
  - High-res rendering
- **UI Enhancements**
  - Integrated views
  - Well log display and histogram
  - Cross-sections

Closing

1. **Digital transformation:** Leverage GPU technology to derive commercial outcomes and support our corporate strategy
2. Cloud and GPU technology improvements will allow us to make decisions with more accuracy, faster
3. Kinetica's GPU-Accelerated Database technology gives us a new way to visualize and interact with massive data



Embracing GPU-Accelerated Visualization at Ultra-High Resolution to Support Next-Generation Geologic Modeling Capabilities

[1] Anadarko Petroleum Corporation

Ingrid Tobar[1], Senior Data Scientist; Richard Sech[1], Senior Staff Geologist; Amit Vij[2], President & Co-Founder; Chad Juliano[2], Sr. Solutions Architect; Rydel Pereira[2], Director of Professional Services

[2] Kinetica DB, Inc

ABSTRACT

A 3D numerical model representing the subsurface volume of a hydrocarbon reservoir was constructed at extremely high resolution. Composed of 90 billion points that characterize the spatial distribution of geologic properties, this dataset exceeded the rendering and processing capability of industry-standard software. Using the GPU-accelerated, in-memory, distributed database technology of Kinetica, along with its GPU-powered analytics platform, we were able to load and access the full extent of this subsurface model.

MOTIVATION

Our reservoir model consists of 90 billion points covering over 10,000 square miles at 100ft spatial resolution in the X and Y dimensions, and 1ft resolution in the Z dimension. Each of the ~4k slices (layers) in the model contains ~24 million points. Facing a demand to incorporate all relevant information in decision-making, 3D visualization and rendering strategies that support large data volumes are a requisite in our workflows and should become targets for next-gen geology software.

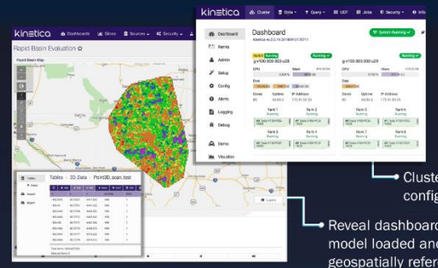
APPROACH

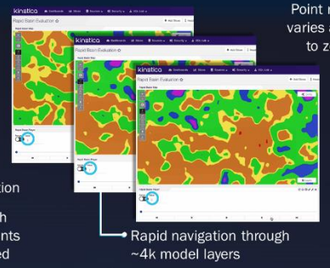
Using GPU-accelerated database technology and on-the-fly 3D tile generation via Kinetica we built an application that enables rapid visualization, interaction (e.g. data selection and filtering), and computation at native resolution. Kinetica's visualization framework, Reveal, served as the end-user web client to render layers of the model draped over a base map. Further, using Kinetica's GPU acceleration and its distributed User-Defined Functions (UDF) framework, we generated 3D tiles on the fly to build a 3D representation of the model.

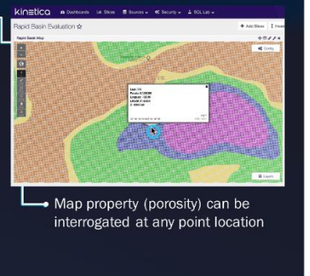
RESULTS

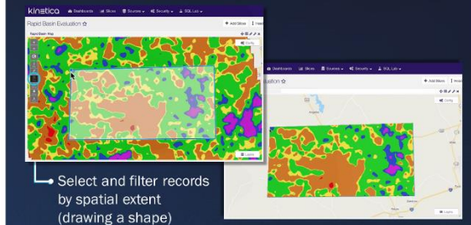
Through this test, we were able to leverage next-generation rendering and visualization techniques to demonstrate:

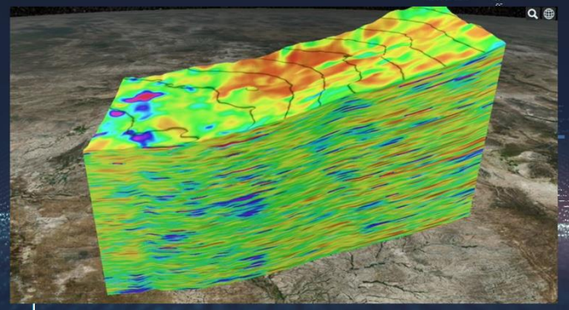
1. The ability to load a pre-built, basin-scale reservoir model and fully render it in its native resolution, and
2. The ability to interact with the entire model, interrogate data points, filter data by spatial extent or by attribute, and quickly perform basic calculations across multiple layers.

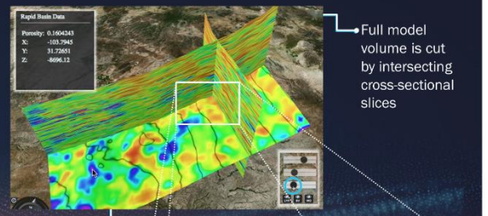


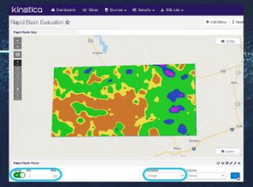


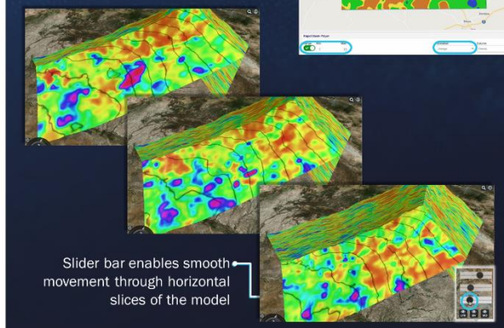


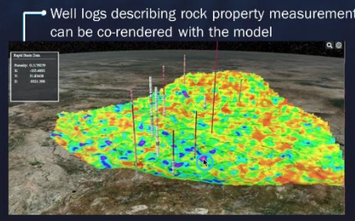


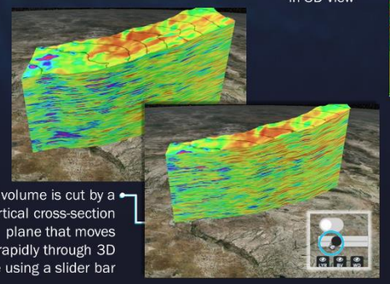


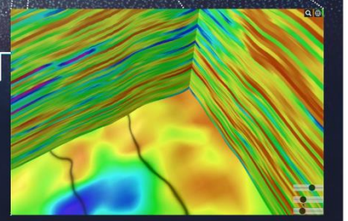












NEXT STEPS

Ongoing studies aim to challenge the platform by scaling up the data volume and the number of petrophysical attributes modeled maintaining high-resolution. In addition, we seek to implement complex calculations, well log attribute display and enhancements to the user interface.

POSTER SESSION

- Monday, Mar 18, 6:00 PM - 08:00 PM
- SJCC Upper Concourse

Acknowledgments

THANK YOU

QUESTIONS?