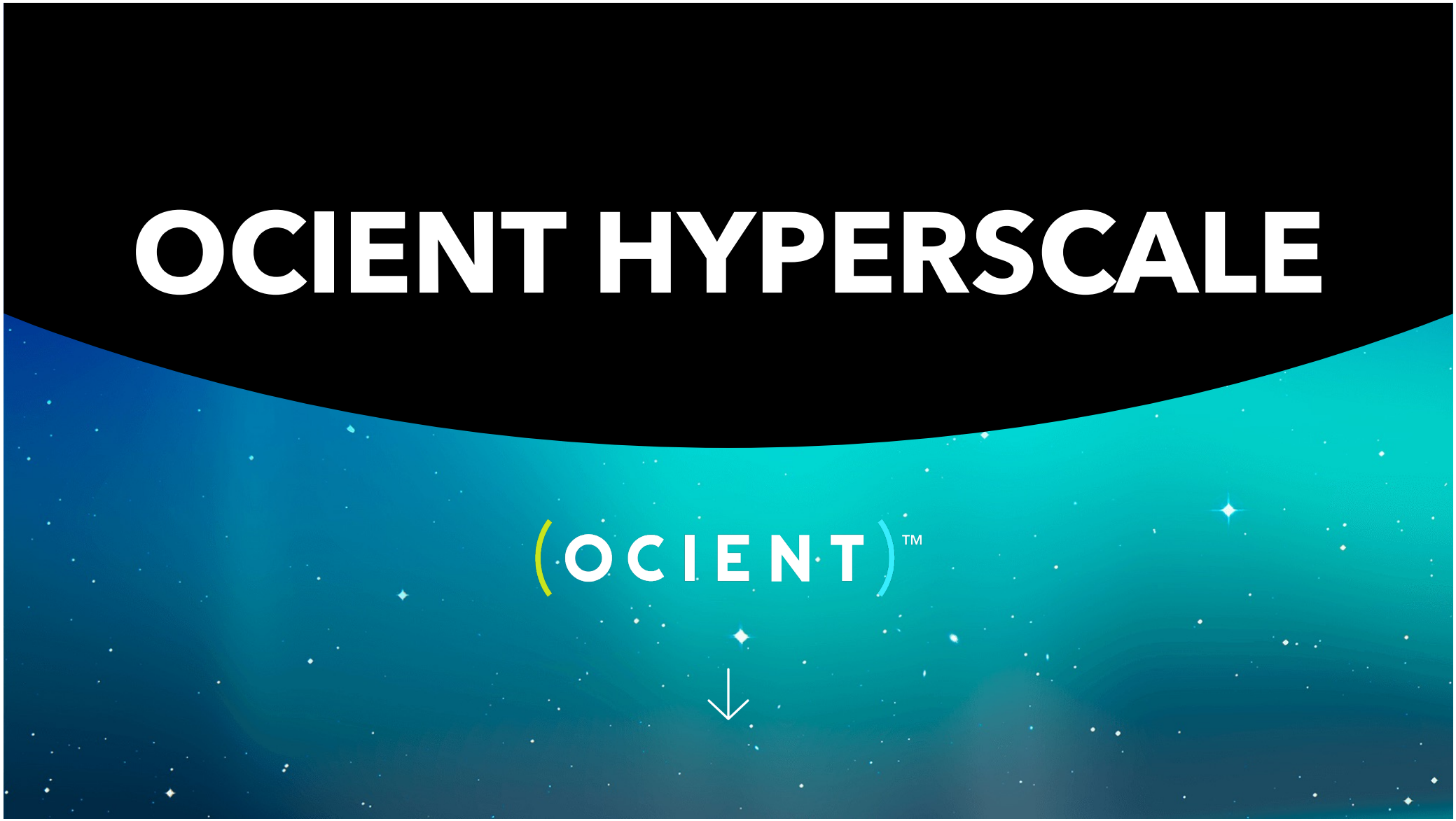
OCIENT HYPERSCALE



(OCIENT)™

BACKGROUND

2004-2015

Led industry transformation of data storage

Cleversafe

2016

Founded by request of the world's largest data analyzing enterprises

2016-2024

Over $175M of financing

OCA VENTURES
GREYCROFT
iqt IN-Q-TEL
PSP PARTNERS
V CAPITAL
Pritzker Group VENTURE CAPITAL
VALOR EQUITY PARTNERS
Levy Family Partners
Riverwalk Capital, LLC
GREAT POINT CAPITAL LLC
BUOYANT VENTURES
WOLF CAPITAL MANAGEMENT LLC

2021

Product Launch after 300+ R&D person years

Over 70 issued patents; Over 300 documented inventions

2024

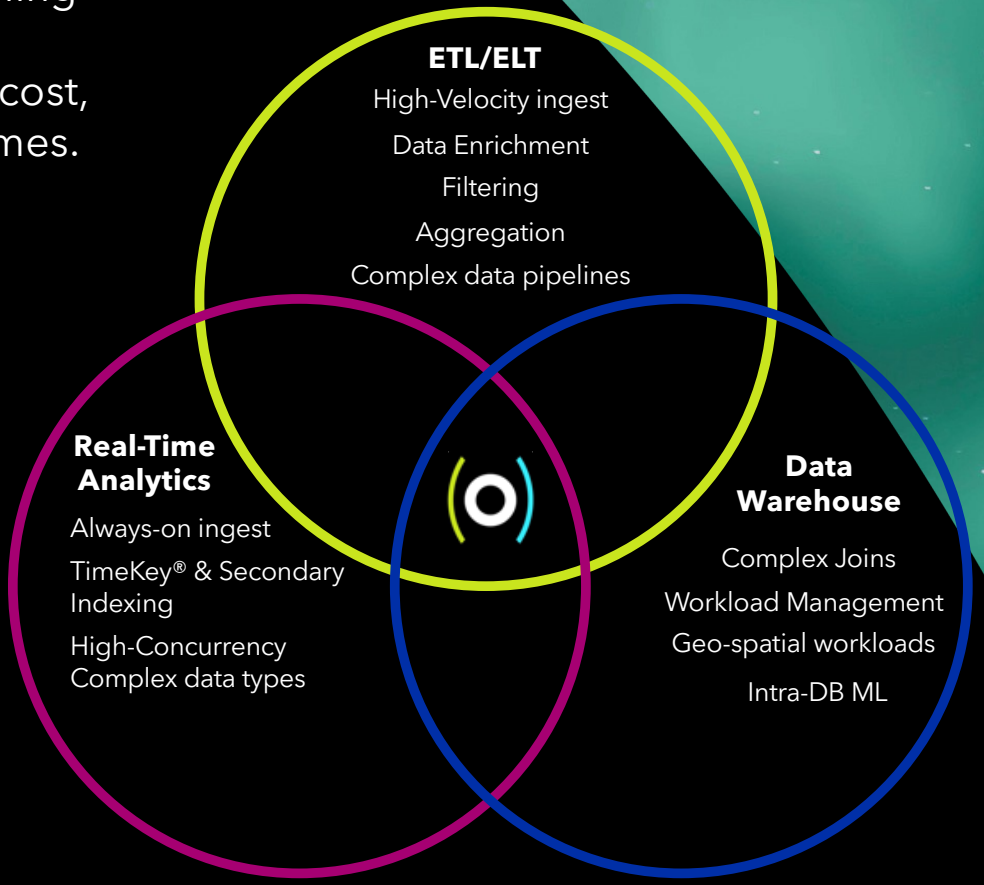
Team of 160+ employees with 100+ engineers

WHAT IS OCIENT

A data analytics solutions company, enabling high-performance data-driven business transformation, with the lowest operational cost, while processing very large-scale data volumes.

Ocient provides a single solution that can flex across…

- Real-Time Analytics
- Data Warehousing
- Complex Data Loading & Transformations
- Machine Learning & Geospatial Analytics

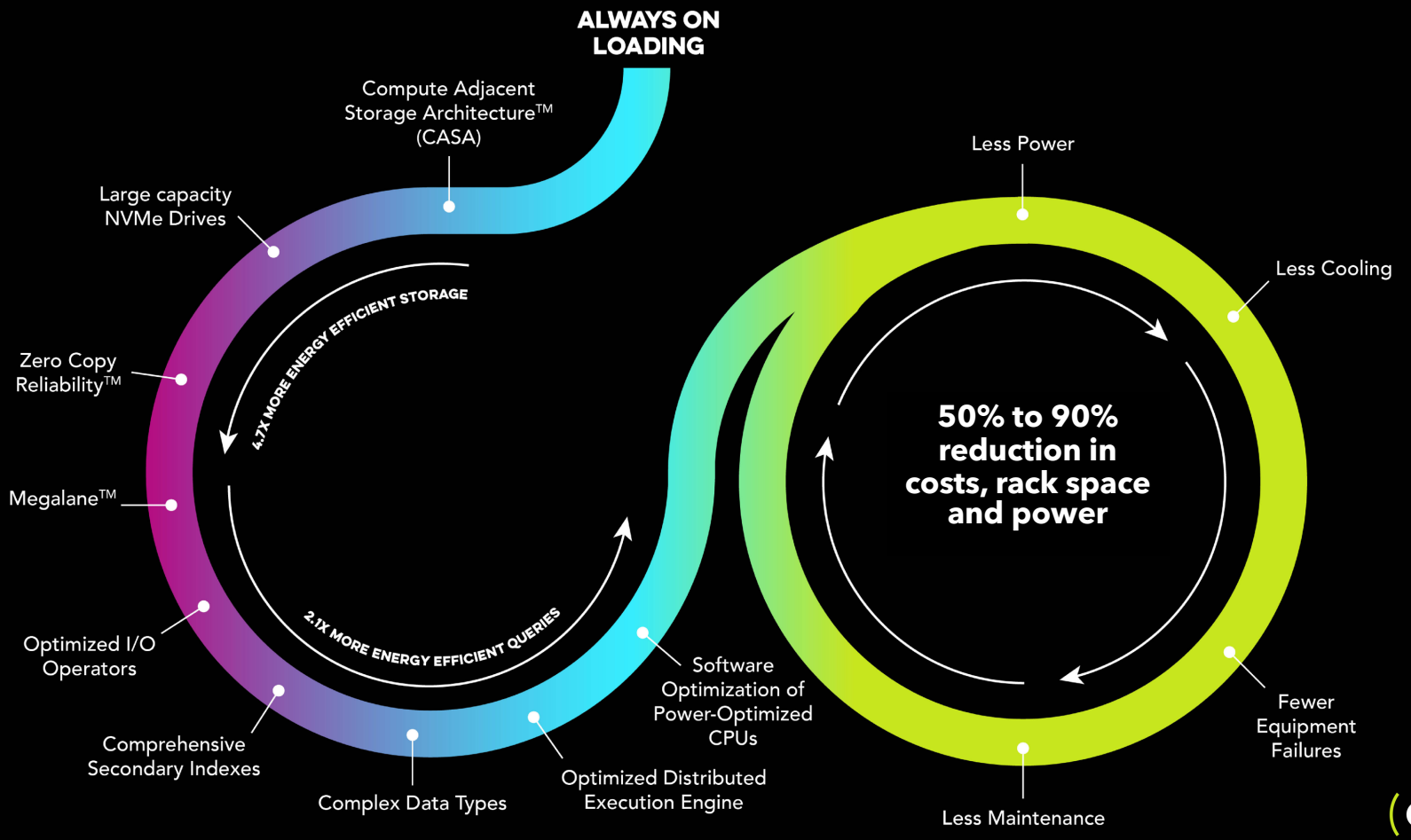


Ocient Advantage

10X–100X Price Performance on trillion+ record workloads

50-80% Lower TCO vs incumbent solutions

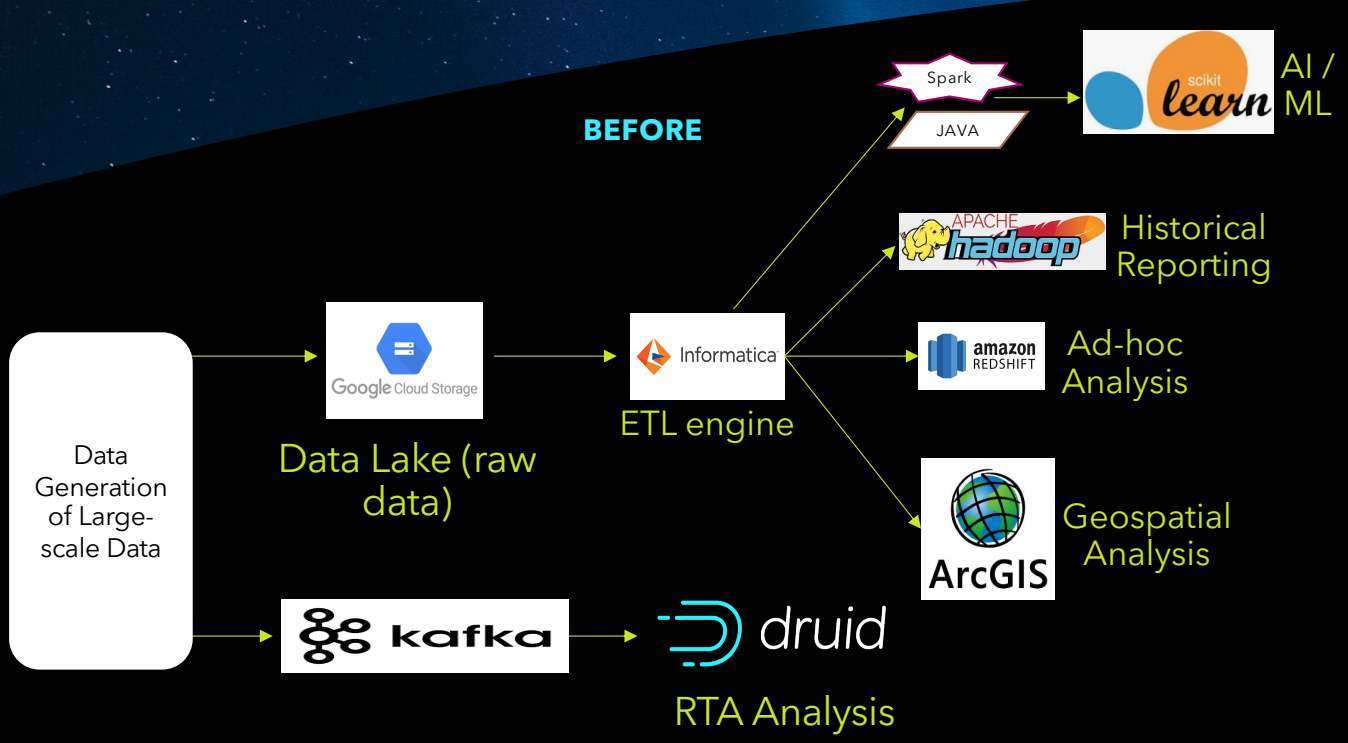
DELIVERING 50%-90% REDUCTION IN SYSTEM COSTS, RACKSPACE AND ENERGY CONSUMPTION

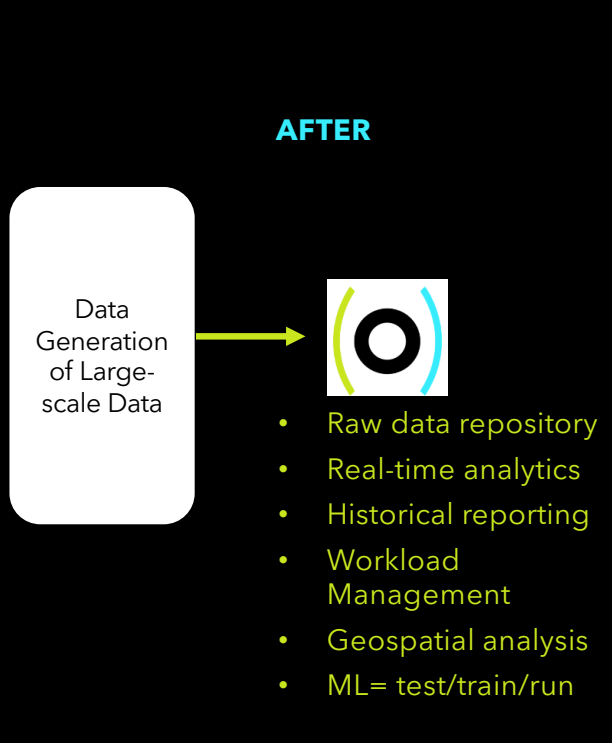


(OCIENT)™

UNIFYING WORK ON LARGE DATASETS

BEFORE / AFTER





(OCIENT)

OPEN SOURCE DATA MOVEMENT
THE PROS AND CONS

Open source frameworks are powerful and flexible but building and running a hyperscale data movement stack with them comes with tradeoffs in cost, complexity, and reliability.

Pros

- No license fee
- Highly customizable
- Large community and ecosystem
- Deploy anywhere

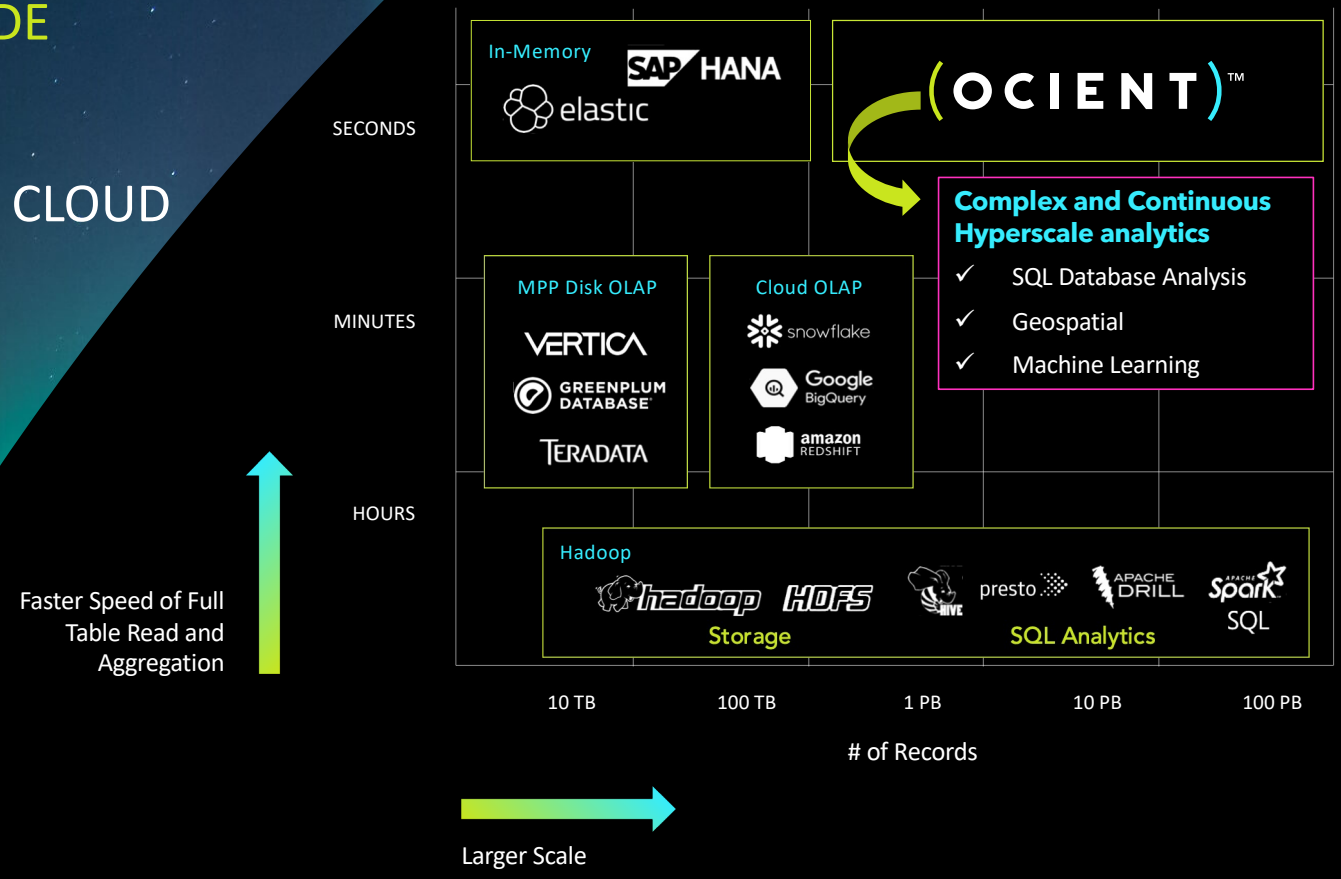
Cons

- High engineering DevOps cost
- Performance at scale requires deep expertise
- Multiple moving parts
- Upgrade and dependency management is complex

Ocient Advantage

- Integrated stack
- Predictable pricing
- Minimal ops overhead
- Consistent performance
- Faster time to value

ORDERS OF MAGNITUDE FASTER THAN LEGACY SOLUTIONS, MORE COST EFFECTIVE THAN CLOUD SYSTEMS AT SCALE



OCIENT™

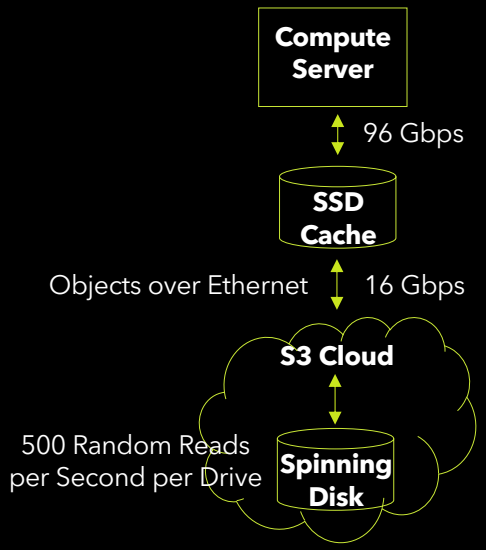
DISRUPTIVE TECHNOLOGY CHANGE*

CUSTOMER BENCHMARK ON 7 TRILLION ROWS AT SAME COST

snowflake · Google BigQuery · amazon REDSHIFT · Azure Synapse Analytics · databricks

Remote Object Storage Architectures (ROSA)

Optimized For: Elastic OLAP workloads



420x Data Bandwidth → 6,720 Gbps

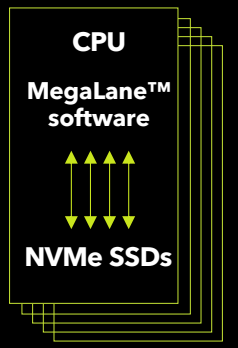
2,000x more IOPS → 1,000,000 Random Reads/Second per Drive

Analytics run 10x – 1,000 faster

All Ocient data essentially "in cache"

OCIENT™ Compute *Adjacent* Storage Architecture™ (CASA)

Optimized For:
1,000,000x larger average Active Data and Ad Hoc queries

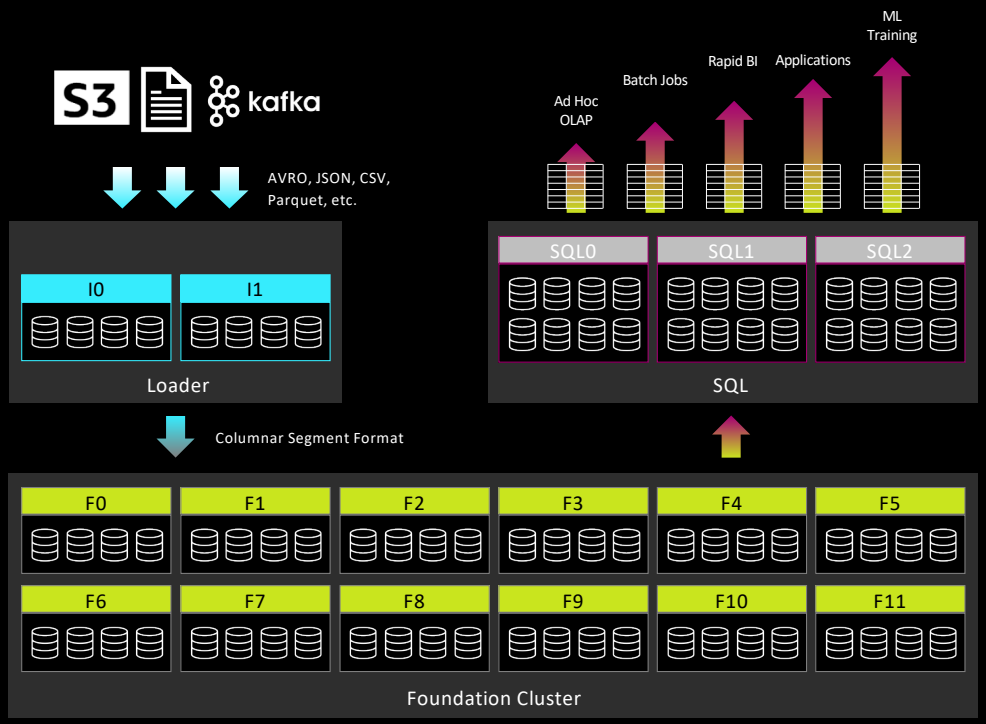


Parallel PCIe 4.0 Lanes

* as defined in *The Innovators Dilemma*

EXAMPLE SYSTEM

NOTIONAL SYSTEM ARCHITECTURE



Queuing / Kafka	Receive, protect, queue
Load and Transform	Accepting multiple, simultaneous streams of data; parse and transform from semi-structured to relational schema, protect, deduplicate and index - made available to query within seconds
Foundation	Erasure storage protection and processing of data
SQL	Provides query results to JDBC / ODBC connections as well as performs admin / metadata duties

Node	CoresSQL	Memory	Storage	Qty
Queuing	32	192GB	12 x 10TB HDD	2
Loader	56	1TB	8 x 2TB NVMe	2
Foundation	56	1TB	12 x 15.3TB NVMe	12
SQL	56	1TB	4 x 2TB NVMe	3
Total Core Count			896	
Raw Disk Storage			2,200TB	
Protection			8 Data + 2 Parity + 2	

System design is dependent on complexity of transformation, sampling rate, scale, number of mixed workloads, concurrency etc. Workshops will be used to determine an appropriate BOM and build the financial and technical business case based on solution requirements both now and in the future.

(OCIENT)

OPEN SOURCE DATA MOVEMENT
THE PROS AND CONS

Open source frameworks are powerful and flexible but building and running a hyperscale data movement stack with them comes with tradeoffs in cost, complexity, and reliability.

Pros

- No license fee
- Highly customizable
- Large community and ecosystem
- Deploy anywhere

Cons

- High engineering DevOps cost
- Performance at scale requires deep expertise
- Multiple moving parts
- Upgrade and dependency management is complex
- No guaranteed backward compatibility

Ocient Advantage

- Integrated stack
- Predictable pricing
- Minimal ops overhead
- Consistent performance
- Faster time to value

DDL PIPELINES

EXAMPLE CAPABILITIES

- **Declarative Loading:** Define and Preview pipelines like a SQL query
- **Observability:** Full catalog support
- **Troubleshooting:** events and errors in catalog and bad data to Kafka topics
- **Control:** Pipeline Controls via DDL
- Complete User Access Controls

Transforms

- Rich SQL transforms
- EXPLODE
- Geospatial data types
- User defined “pipeline functions” in groovy

Formats

- JSON, CSV, Binary, Parquet, ASN.1, etc

Sources

- Kafka, S3, Filesystem, HDFS, etc

Commands/Modes

- Pipeline Preview
- Continuous File Loading
- Create, Start, Stop, Drop Pipeline

CUTTING-EDGE GEOSPATIAL ANALYTICS

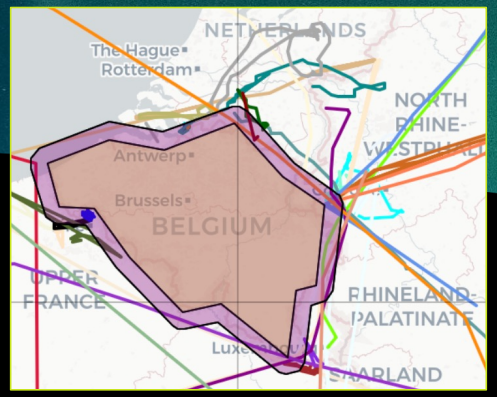
> 120 UNIQUE FUNCTIONS DRIVEN BY THE NEED FOR VECTOR DATA ANALYSIS

NATIVE GEOSPATIAL AND COMPLEX COLUMN TYPES

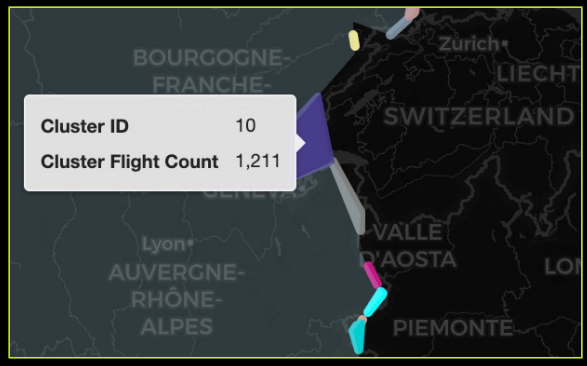
- Supports Complex Types
  - Arrays, Tuples
- Based on Non-Euclidian Spherical Geometry
  - Leads to minimal distortion vs projections
- Complex functions at *scale*
  - Buffer, Envelope
  - Convexhull, Covers
  - Project
  - Distance

ADVANCED GEOSPATIAL FUNCTION SUPPORT

- Unique Spatiotemporal functions answer when *and* where
  - Lat/lon/timestamp(s) of exact intersection of 2 paths/objects
  - Duration of intersection
- ML clustering algorithm returns physically clustered objects based on min number and radius of proximity



Return list of flights which encroached on a country's border but did not actually enter the airspace



Single function returns clusters of border crossings (x crossings in y distance of each other)

SCALING GEOSPATIAL DATA LOADING AND ANALYSIS FOR THE GOVERNMENT

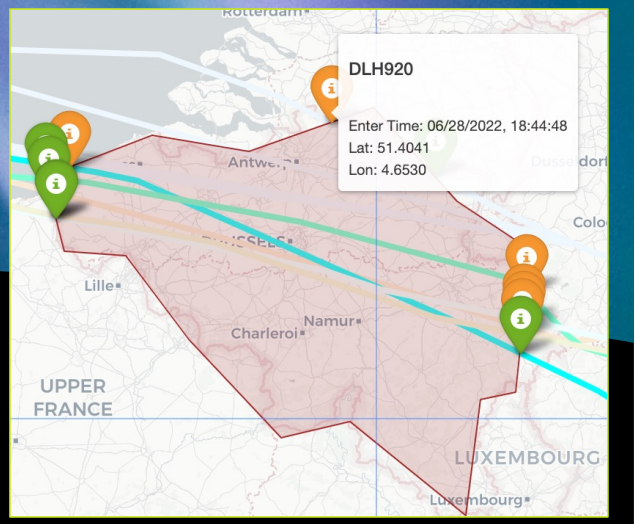
OCIENT IS AN IN-Q-TEL PORTFOLIO COMPANY

CUSTOMER CHALLENGE:

A potential customer was challenged to accurately predict and track objects over the earth's surface with interactive response times on < 30TB of geospatial data. Testing with Hbase, Databricks, and Amazon Aurora failed to meet the scale and performance criteria to execute always-on data ingest alongside concurrent queries.

OCIENT SOLUTION:

OcientGeo's ability to index spatiotemporal data over land and sea accelerated query performance 323x while providing 100x higher resolution insights (versus alternative solutions). Ocient also enables aggregated data across multiple columns and billions of rows in seconds, and tracks aircrafts across disparate areas of interest in seconds, both previously infeasible capabilities that are essential for their mission critical workloads.



KEY RESULTS:

- 323X query acceleration on a 1.2 petabyte geospatial dataset
- ~20,000 queries per minute
- 100X resolution insights
- ~2.5TB of data ingested continuously per hour per loader node
- Previously infeasible insight into air traffic patterns with predictive modeling on future flight paths

(OCIENT)™

MACHINE LEARNING IN PLACE

ENABLING HYPERSCALE MACHINE LEARNING

TRAIN IN PLACE

Train ML models without moving data, enabling larger training sets and better data security.

ANALYZE AND ITERATE

Choose and customize data analysis using a large toolbox of ML algorithms in multiple categories:

Regression	Simple Linear Regression, Multiple Linear Regression, Polynomial Regression, Linear Combination Regression, Nonlinear Regression, Vector Autoregression
Classification	K Nearest Neighbors, Naïve Bayes, Decision Tree, Logistic Regression, Support Vector Machine, Random Forest
Clustering / Dimension Reduction	Principal Component Analysis (U), K-Means Clustering (U), Gaussian Mixture, Linear Discriminant Analysis
Other	Association Rules, Feedforward Neural Network

EVALUATE AT LOAD (FUTURE)

Evaluate trained ML models at load time as part of an ETL pipeline. Classify, filter, and/or enrich streaming or batch loaded data in one step.

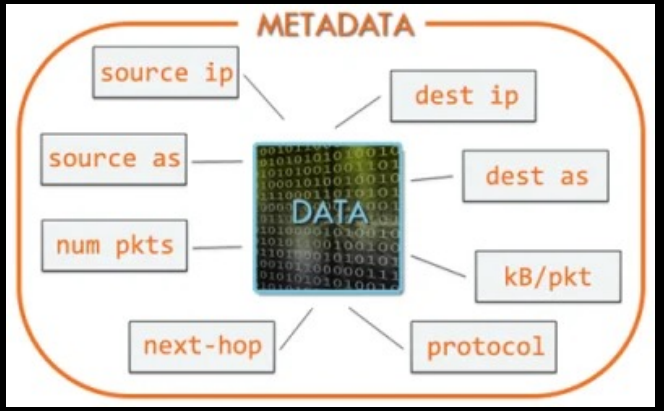
UNIFY OLAP SQL AND ML

ML as a first-class citizen in the database enables full expressivity of SQL for data analysis with ML and linear algebra mixed in - *no need to code 13k case statements for a Decision Tree model*

NETWORK TRAFFIC ANALYSIS

SEARCH AND ANALYZE YEARS OF NETWORK METADATA

- Built-in support for IP address data types
- Support for Netflow/IPFIX records
- Existing integrations with top probe vendors such as Gigamon
- Scalable ingest to handle millions of records/sec
- Interactive-time query processing can handle comparisons across years of cyber data
- Built-in ML enables rapid model iteration and full-resolution training to enable accurate predictive analytics



(OCIENT)

CASE STUDY: TIER 1 TELCO

Avoided millions in fines while modernizing a multi-petabyte on premises legacy solution

SITUATION

Needed to reduce exposure to fines due to lack of SLA Compliance around lawful disclosure. Other data warehouse providers couldn't handle billions of new records per day and scale to hold 10PB+ data overall with always-on queries.

Business requirements indicated a **cost-effective, on-premises deployment** with geographically dispersed **disaster recovery**.

RESULTS

- **Query times reduced** from **10+ minutes to seconds**, while **loading 700k records per second** and supporting **10,000+ queries per day**
- **3x increase in data capacity, 5x+ cost reduction**
- **Minimal footprint** with entire solution in one 42U cabinet
- **Easy integration and maintenance** via Ocient's high-touch delivery model and ongoing management services

OCIENT IN ACTION – ON PREM ENERGY SAVINGS

LEADING TELCO CUSTOMER 5 PB EXAMPLE

OpenSearch elastic

= 5+ racks

100+ Nodes
40TB/node
5+ Racks & DC Floor Tiles
$5K per DC rack per month
$25K per month KWhr

OCIENT™

= Single rack

16 Nodes
200TB/node @ 85% compression
Single rack / data center floor tile
$2.5K per DC rack per month
$2.5K per month KWhr

(OCIENT)™

PRICE PERFORMANCE

7x less than other solutions

100 TBs Per Day 30 days of retention – 3PB	Ocient	Druid Imply Polaris	Snowflake 4XL	Elastic Cloud
Capacity of Hot/Cached Data 4X Compression	**3,146 TB**	307 TB	102 TB	600 TB
Second Tier Storage (less performant)	**None**	1500TB	750TB	5,025TB
Processing Cores	**448**	480	512	240
Deployed	**On-Prem**	AWS	AWS	AWS
Capex/Opex	**Capex**	Opex	Opex	Opex
Cost Per Month	**$94,867**	$787,536	$654,880	$488,016
Cost Per 36 Months	**$3,415,194**	$28,351,296	$23,575,680	$17,568,583
Percentage Increase	**1.0X**	**8.3X**	**6.9X**	**5.1X**

The Ocient solution listed above included hardware, perpetual software, (3) years of support and remote Management Services, as well as predictable pricing which allows for peaks

Adrian Burke
Solutions Architect
aburke@ocient.com
www.ocient.com

Bill Minarchi
Dir Federal Sales
bminarchi@ocient.com
www.ocient.com