

# Cybage helped in big data validation of next generation machine data processing

## About the Client

Is a global full lifecycle API management SaaS solution provider for internal APIs, B2B APIs, and public API programs

Has headquarters in Palo Alto, California, and offices in North America, Europe, Asia, the Middle East, Africa and South America

Its platform seamlessly connects any application or data source; intelligently unifies data for greater access, trust, and control; and confidently predicts outcomes in real time and at scale

## Business Needs

- Validating next-generation machine data processing solution for a rapidly extending business (250,000+ Events per second & over 10 TB per day)
- Ensuring near real-time analytics processing pipeline with 1-minute aggregation window and making it available in less than 2 minutes for the report engine
- Creating configurable Test Cluster mimicking massive, multi structured voluminous data ingestion test pipeline
- Identifying and tuning auto-scalable cloud solution for low cost with honoring service performance and resilience SLAs
- Ensuring processing of high velocity batch pipeline with each record being processed within 30 min window
- Confirming multi-cluster deployment against chaos engineering practices for multi-failure cases (Partial, Full, Gradual, Sudden etc.)

## Solutions

- Designed an end-to-end QA strategy and test cluster targeting the 4Vs of Big Data
- Validated efficient deployment considering reliability & fault tolerance across Multi-zone cluster topology
- Ensured data quality and integrity validation at Data ingestion layer & Data processing layer
- Derived and benchmarked every layer of the platform to achieve optimized cost to performance cluster provisioning
- Performed data analytics and reports validation against multiple data Dimensions for both batch as well as real time data processing

## Technology Stack

### Test Engineering Stack



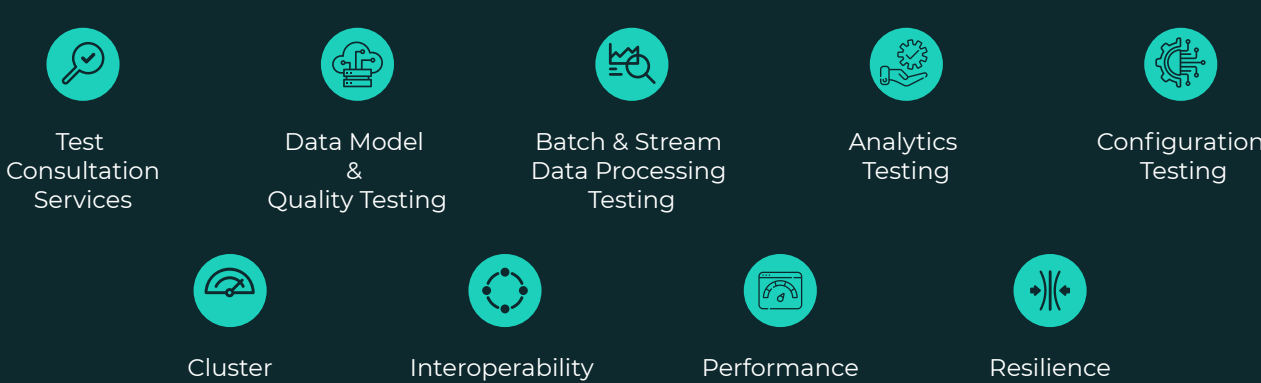
### Application Technology Stack



## Business Impact

- Enhanced** processed data quality with comprehensive test covering Volume, Variety, Value and Velocity
- Derived** best practices for tracking application and cluster health
- Enriched** the relationship between throughput, required processing, and infrastructure capacity
- Ensured** valuable insights into Op-Ex and future scaling needs
- Increased** ROI by fine-tuning configurations for optimized services utilization
- Improved** application scalability, availability, and readiness for anticipated failover cases

## Cybage Bigdata Testing & QA Capabilities



## Assuring Quality for Unique Big Data Solutions