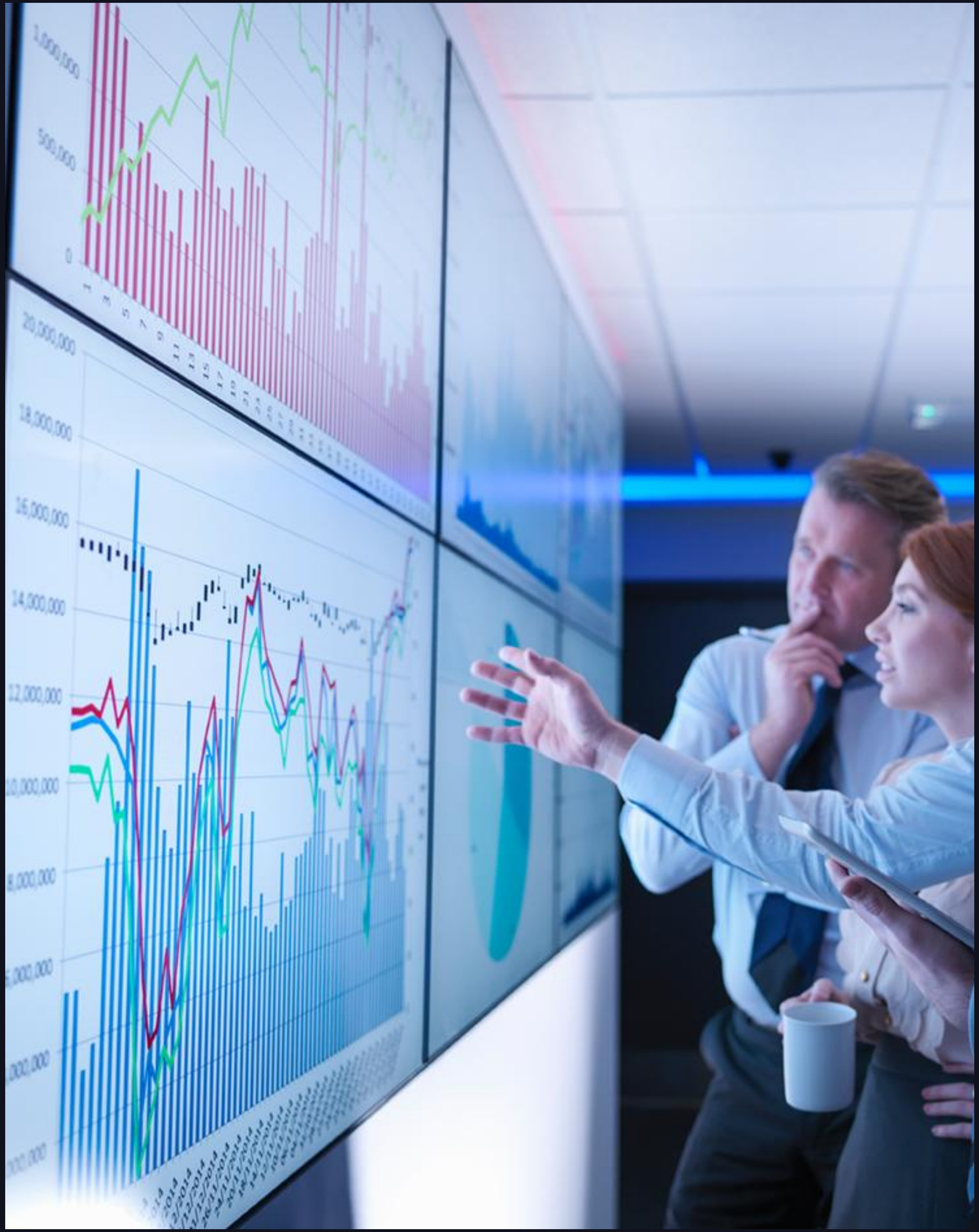
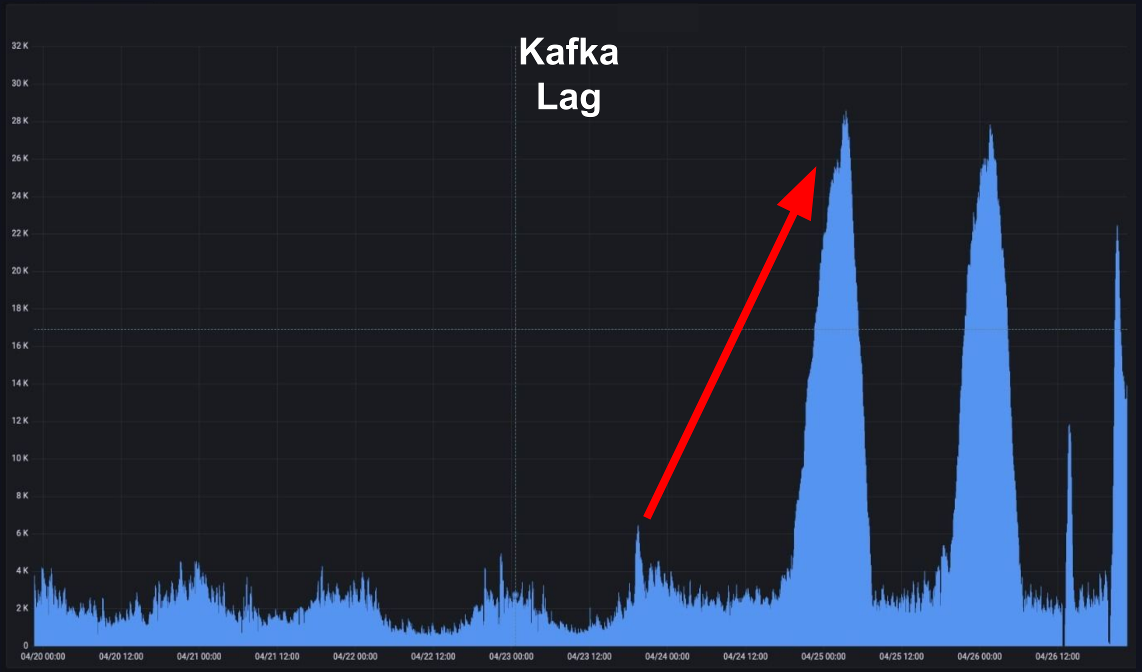
When Working with Big Data…



… You’ve Probably Encountered This

```
ERROR …StorageException: Status code 503,
"{"error":{
    "code":"ServerBusy",
    "message":"Operations per second is
                over the account limit."
}}
```

… Or This…


Kafka Lag
32 K
30 K
28 K
26 K
24 K
22 K
20 K
18 K
16 K
14 K
12 K
10 K
8 K
6 K
4 K
2 K
0
04/20 00:00
04/20 12:00
04/21 00:00
04/21 12:00
04/22 00:00
04/22 12:00
04/23 00:00
04/23 12:00
04/24 00:00
04/24 12:00
04/25 00:00
04/25 12:00
04/26 00:00
04/26 12:00

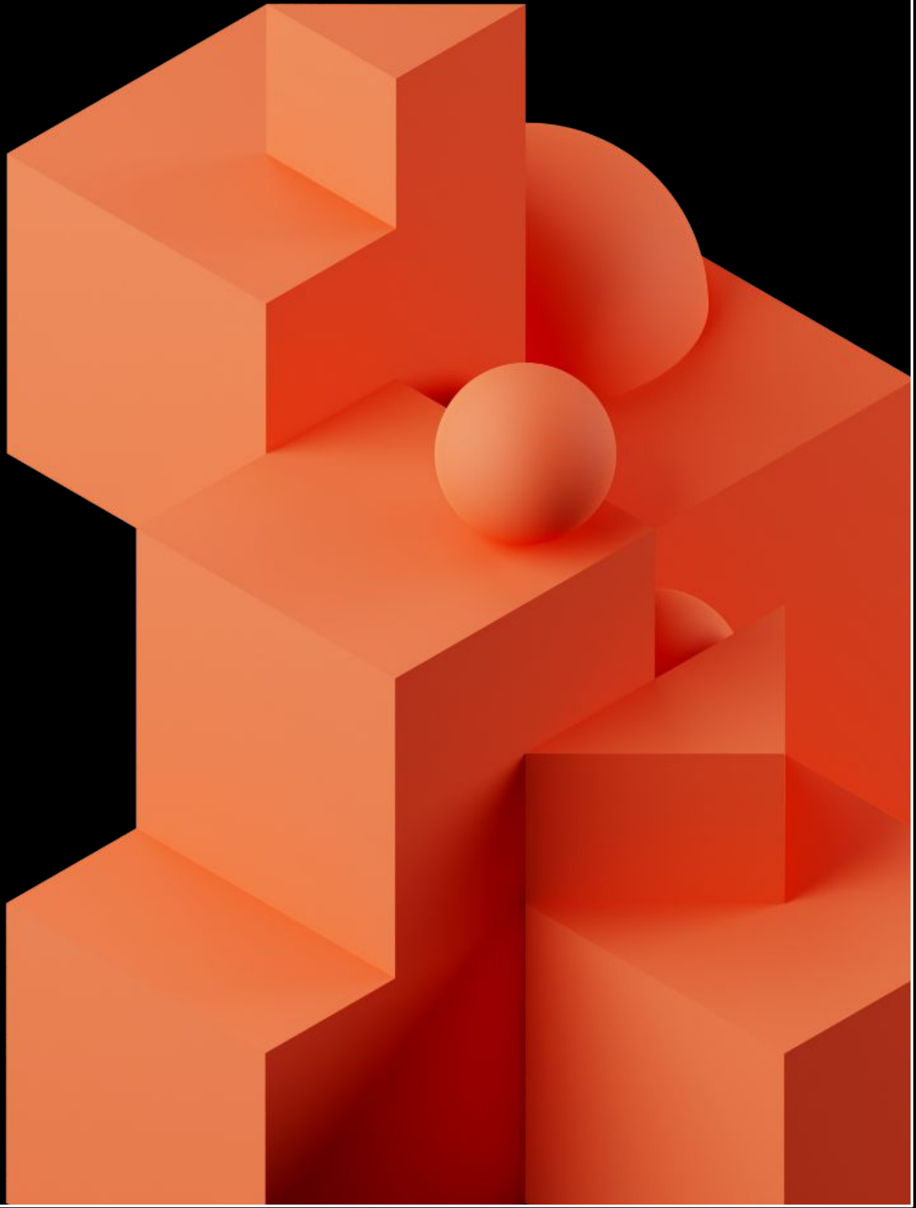
… Or This…

This is the Talk for You!

DATA+AI SUMMIT
BY databricks

Taking Your Cloud Vendor To The Next Level

Solving Complex Challenges With Azure Databricks



Introduction

Tomer Patel

Engineering Manager @ Akamai
Prev. Team Lead @ Clarizen
Tomer Patel @tomer_patel

Itai Yaffe

Senior Big Data Architect @ Akamai
Prev. Sr. Solutions Architect @ Databricks
Dealing with Big Data challenges since 2012
Itai Yaffe @ItaiYaffe

What Will You Learn?

- Understanding the main **challenges** of a cloud-based massive-scale data infrastructure

What Will You Learn?

- Understanding the main **challenges** of a cloud-based massive-scale data infrastructure
- How to iteratively architect such an infrastructure to **mitigate** those challenges

What Will You Learn?

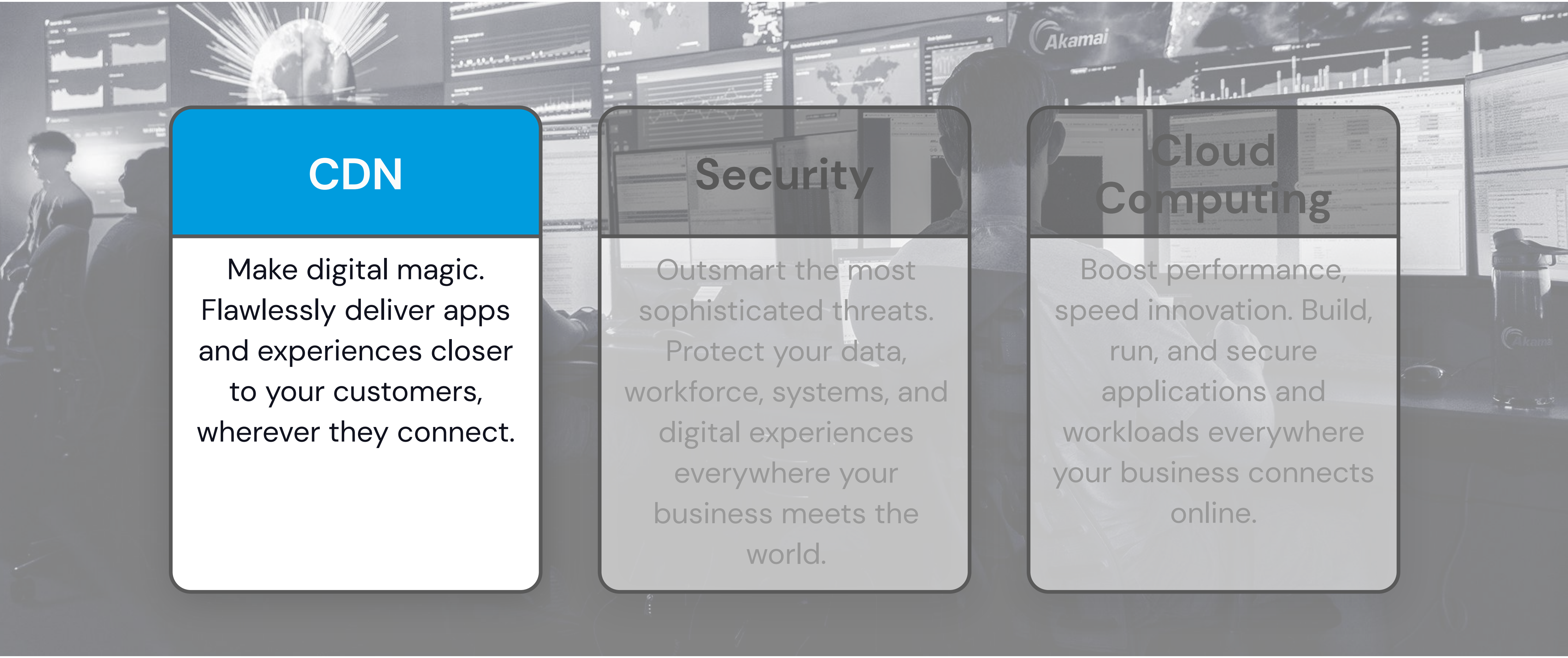
- Understanding the main **challenges** of a cloud-based massive-scale data infrastructure
- How to iteratively architect such an infrastructure to **mitigate** those challenges
- Tips for **optimizing** a massive-scale data infrastructure

About Akamai

Power and Protect Life Online

Over 20 years ago, we set out to solve the toughest challenge of the early internet

Akamai's 3 Pillars



CDN

Make digital magic. Flawlessly deliver apps and experiences closer to your customers, wherever they connect.

Security

Outsmart the most sophisticated threats. Protect your data, workforce, systems, and digital experiences everywhere your business meets the world.

Cloud Computing

Boost performance, speed innovation. Build, run, and secure applications and workloads everywhere your business connects online.

Akamai’s 3 Pillars

CDN

Make digital magic. Flawlessly deliver apps and experiences closer to your customers, wherever they connect.

Security

Outsmart the most sophisticated threats. Protect your data, workforce, systems, and digital experiences everywhere your business meets the world.

Cloud Computing

Boost performance, speed innovation. Build, run, and secure applications and workloads everywhere your business connects online.

Akamai’s 3 Pillars

CDN

Make digital magic. Flawlessly deliver apps and experiences closer to your customers, wherever they connect.

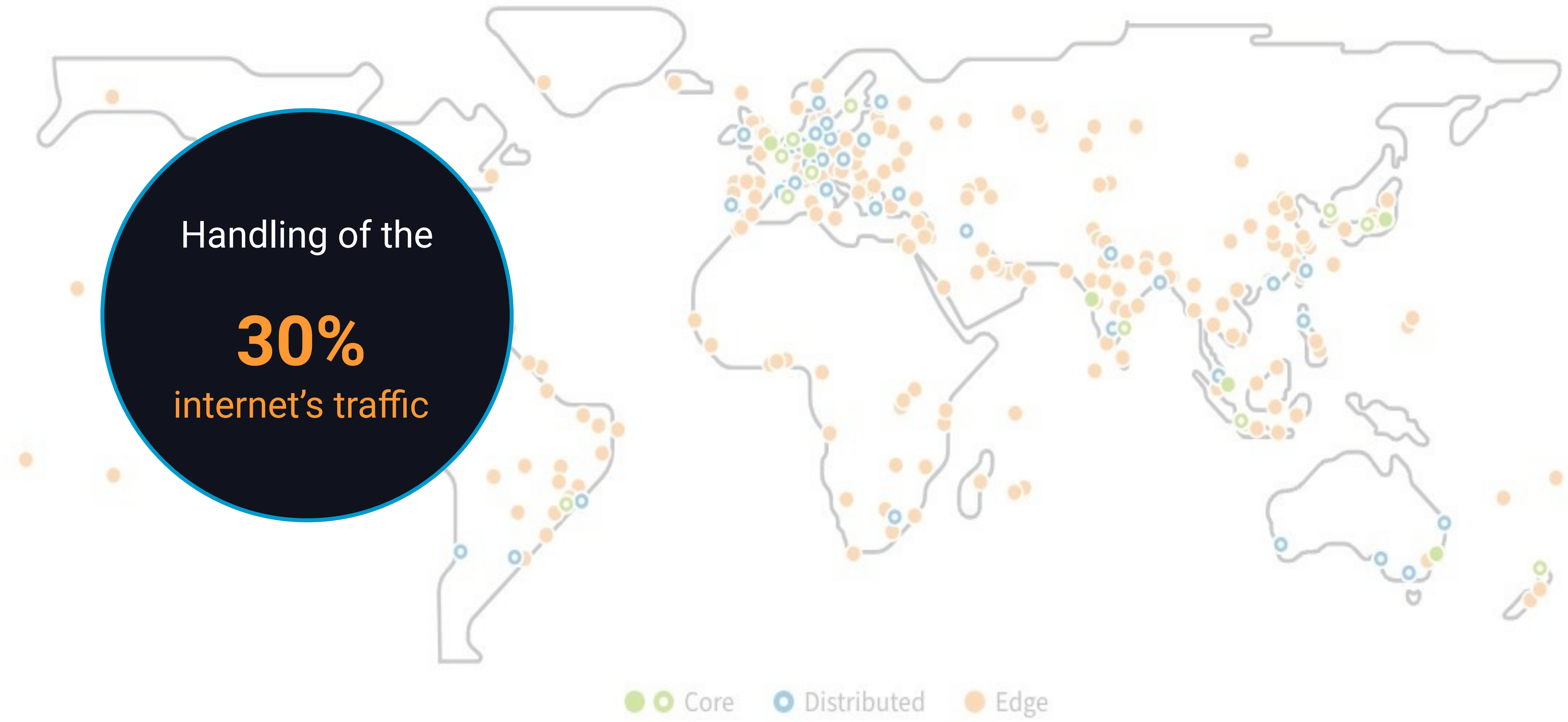
Security

Outsmart the most sophisticated threats. Protect your data, workforce, systems, and digital experiences everywhere your business meets the world.

Cloud Computing

Boost performance, speed innovation. Build, run, and secure applications and workloads everywhere your business connects online.

Akamai in Numbers



Handling of the

30%

internet's traffic

Akamai in Numbers

Handling of the

30%

internet's traffic

Employees

> 10,000

Akamai in Numbers

Handling of the

30%

internet's traffic

Employees

> 10,000

Data processed using Databricks

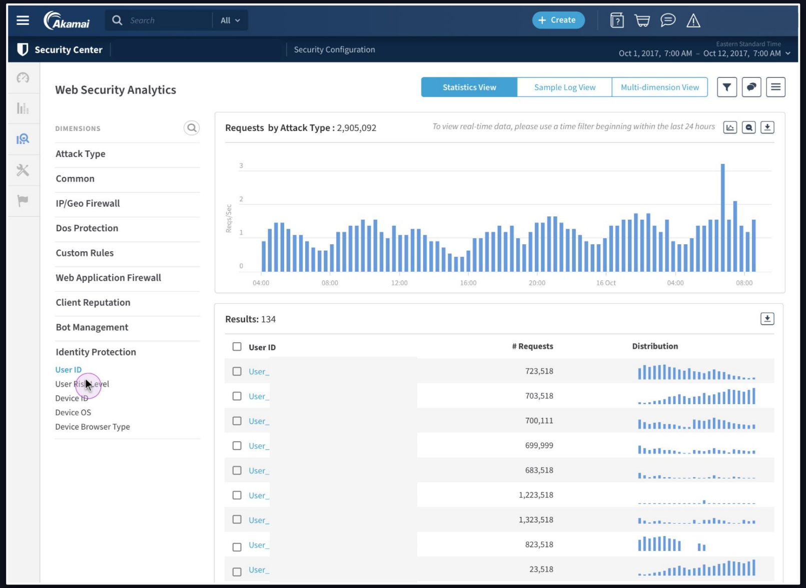
~ 50

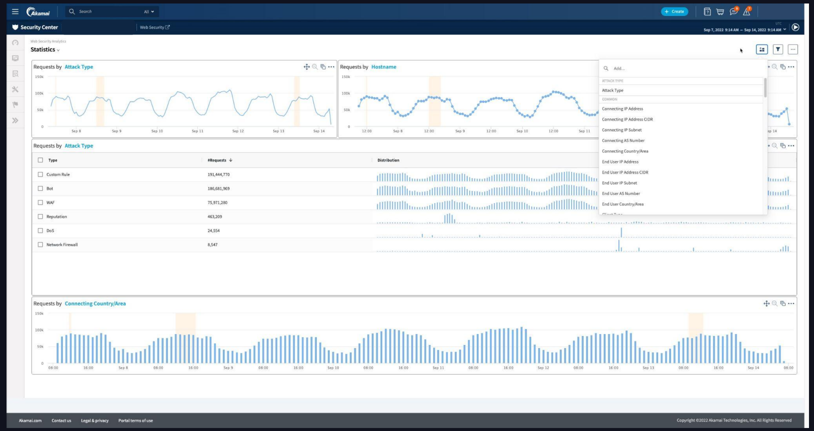
Exabytes

What is WSA?

Web Security Analytics

A **unified** and efficient **platform**, that enables **Akamai's customers** to **assess** a wide range of **streaming security events**, and perform **analysis** of those events, so they can take **informed actions** in **real-time**

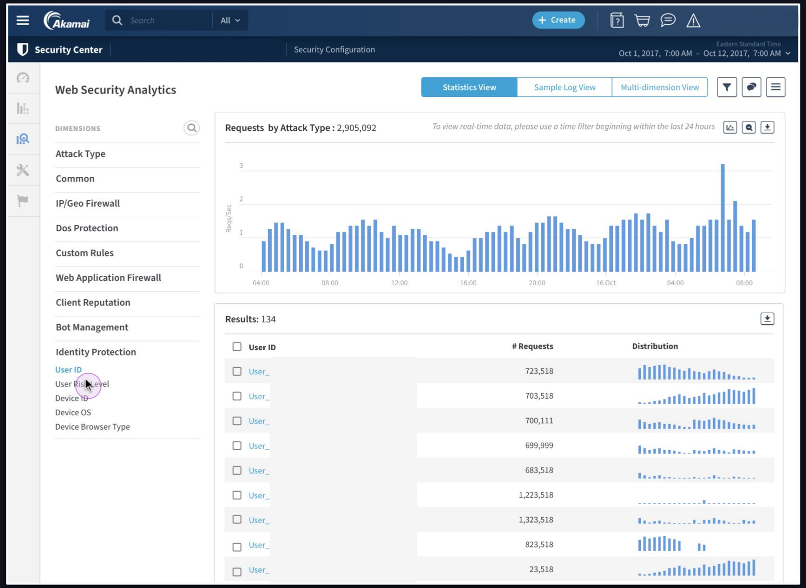


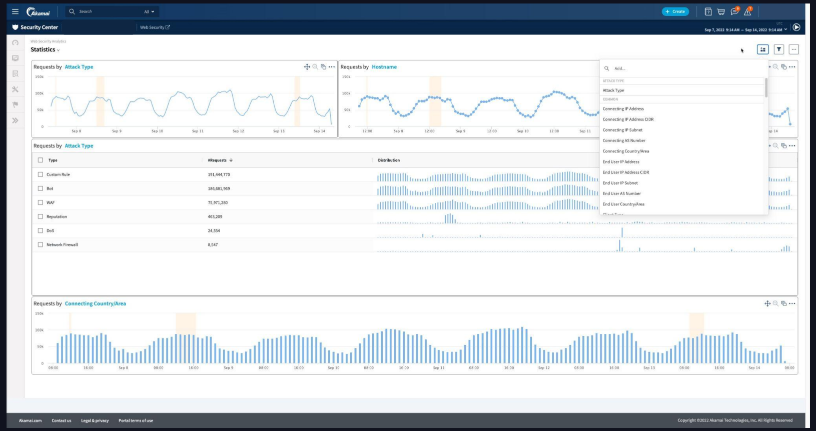


What is WSA?

Web Security Analytics

A Massive-Scale Data Infrastructure

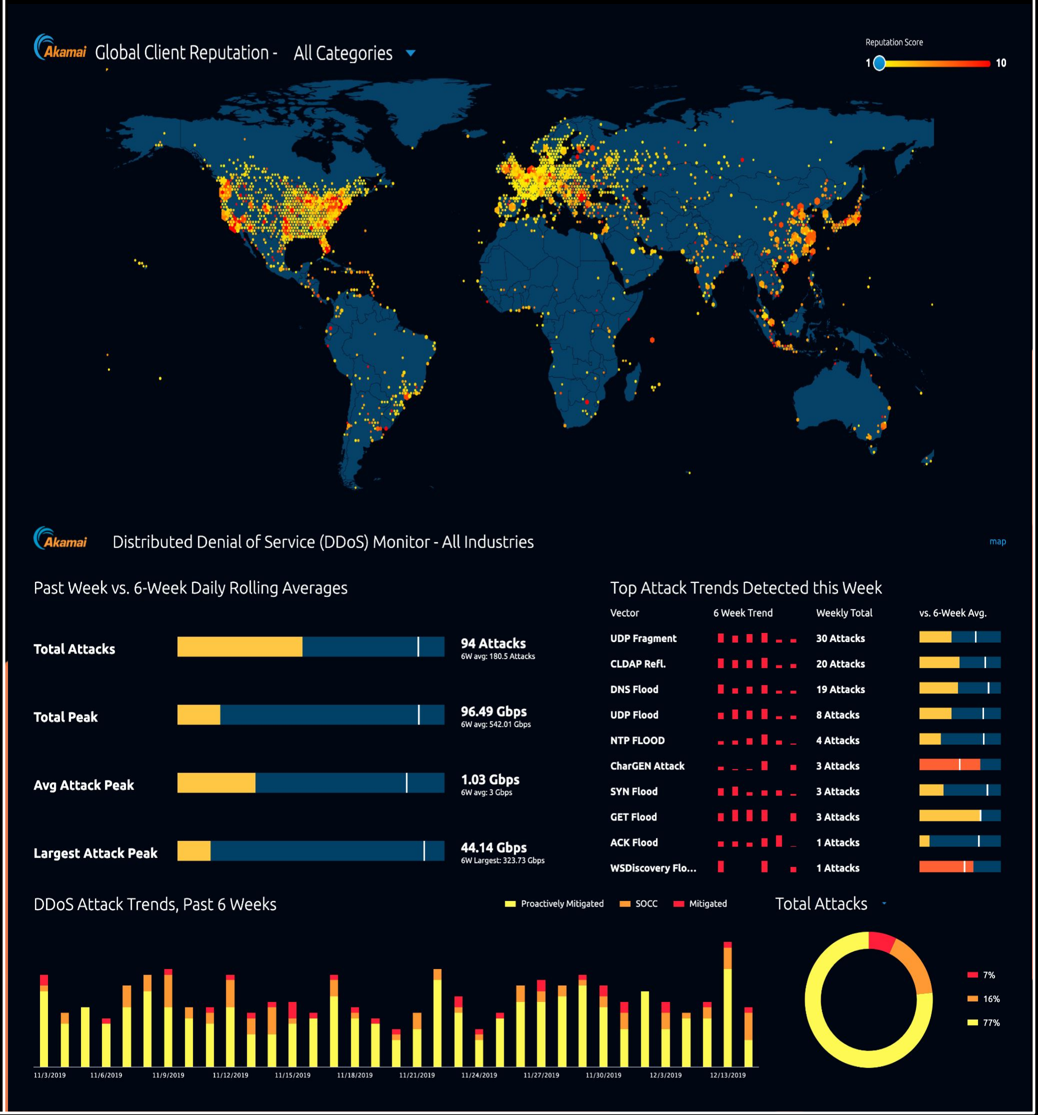




What does Massive-Scale Data Infrastructure Mean?

Generally speaking, it's about efficiently **handling massive amounts of data at scale**

Main Challenges of a Cloud-Based Massive-Scale Data Infrastructure



3 Main Challenges

Processing

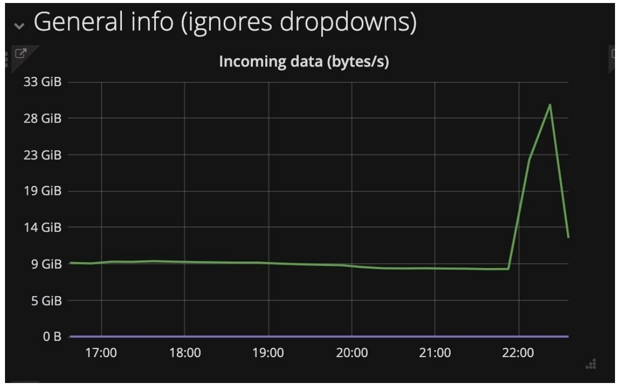
Storing

Analyzing

WSA Main Challenges

Processing

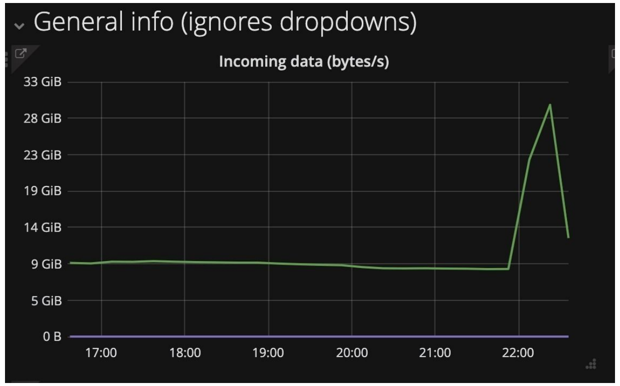
- Volume – 10–14 Gbps (and increasing)
- SLA – 5 minutes from our Edge servers to our Data Lake



WSA Main Challenges

Processing

- Volume – 10-14 Gbps (and increasing)
- SLA – 5 minutes from our Edge servers to our Data Lake



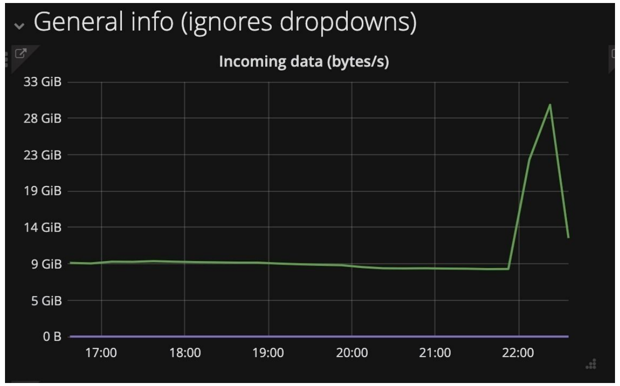
Storing

- Storage capacity – over 6PB
- Retention period – 31 days

WSA Main Challenges

Processing

- Volume – 10-14 Gbps (and increasing)
- SLA – 5 minutes from our Edge servers to our Data Lake



Storing

- Storage capacity – over 6PB
- Retention period – 31 days

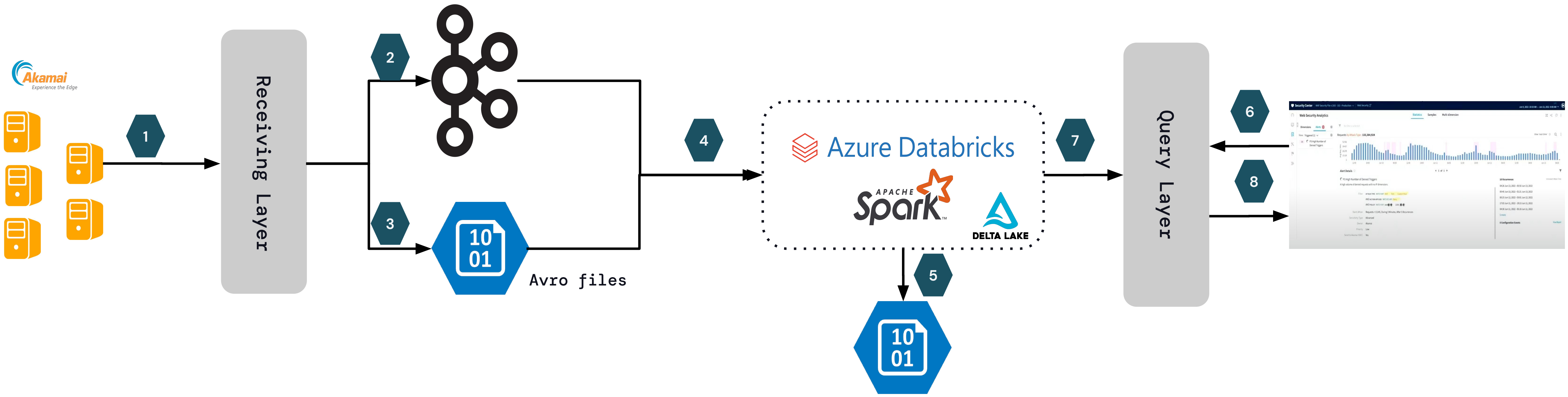
Analyzing

- # of queries – 100s of queries/minute
- SLA – 10s for 99% of the queries
    - Each query can scan 100s of TBs
    - 60+ dimensions, (almost) infinite number of filter combinations

Akamai

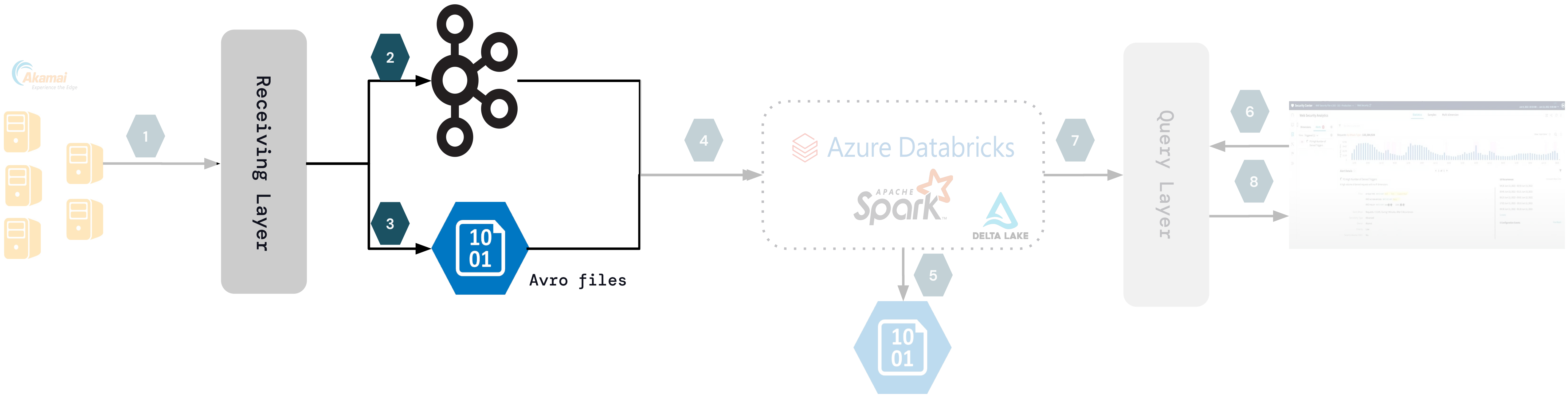
Architecting and Re-Architecting to Mitigate the Main Challenges

CSI High-level Architecture



Akamai

CSI High-level Architecture



Akamai

Receiving Layer – Raw Data

kafka

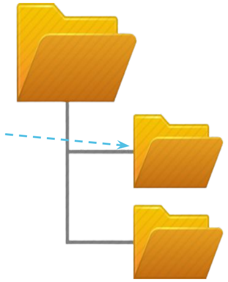
Queue with “**pointers**” to the blob storage, plus **metadata**.

For example:

```
{
  "Path": "/2023-05-16/…/…avro.deflate",
  "size": 7526435,
  "recordsCount": 9686
}
```

Microsoft Azure
Blob Storage

The actual avro files



Akamai

Receiving Layer – Storage Types

3 Types of Storages In Use

1. Azure **Standard** Blob Storage
   a. Relatively cheap and write-performant

Receiving Layer – Storage Types

3 Types of Storages In Use

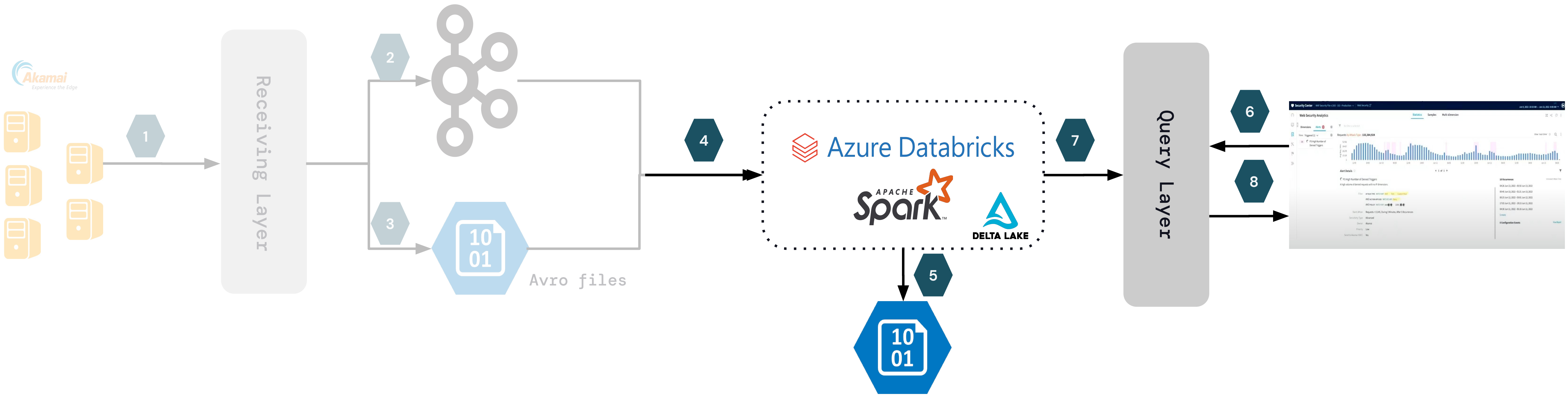
1. Azure **Standard** Blob Storage
    a. Relatively cheap and write-performant
2. Azure **Premium** Blob Storage
    a. Where we need minimal write latency
    b. More expensive than Standard (~10x)

Receiving Layer – Storage Types

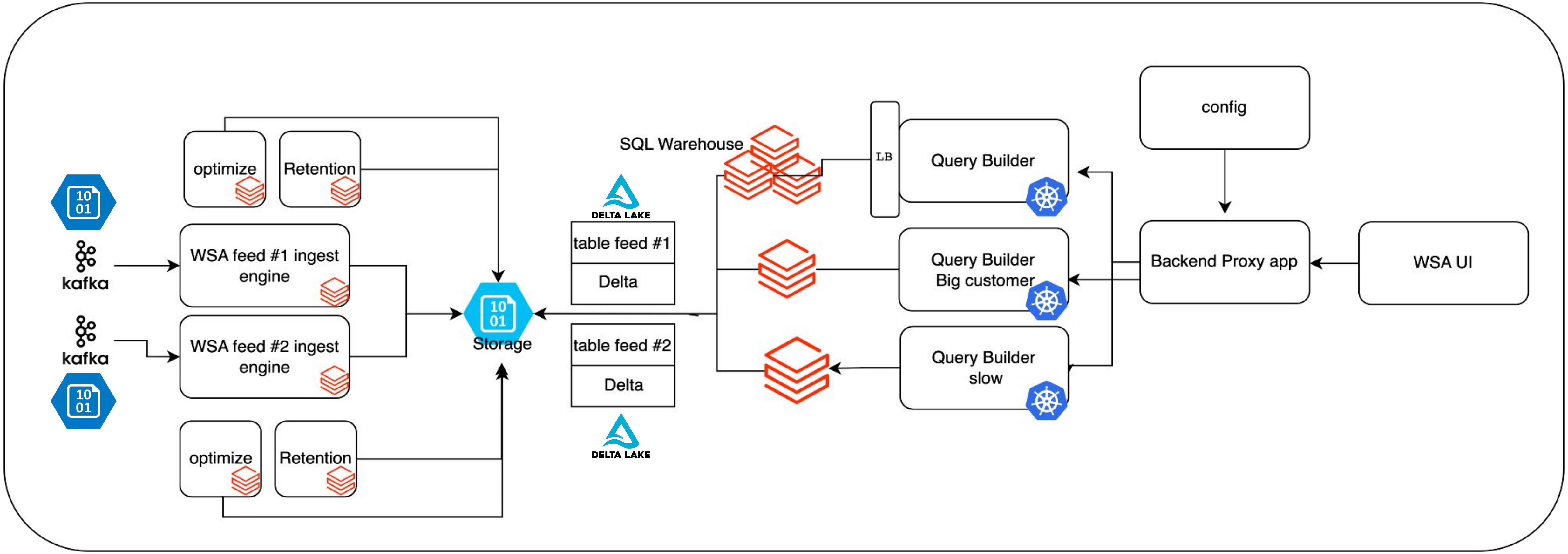
3 Types of Storages In Use

1. Azure **Standard** Blob Storage
   a. Relatively cheap and write-performant
2. Azure **Premium** Blob Storage
   a. Where we need minimal write latency
   b. More expensive than Standard (~10x)
3. Azure **Data Lake Storage Gen2**
   a. Provides additional capabilities such as
      i. Hadoop-compatible access
      ii. Hierarchical directory structure for high-performance data access

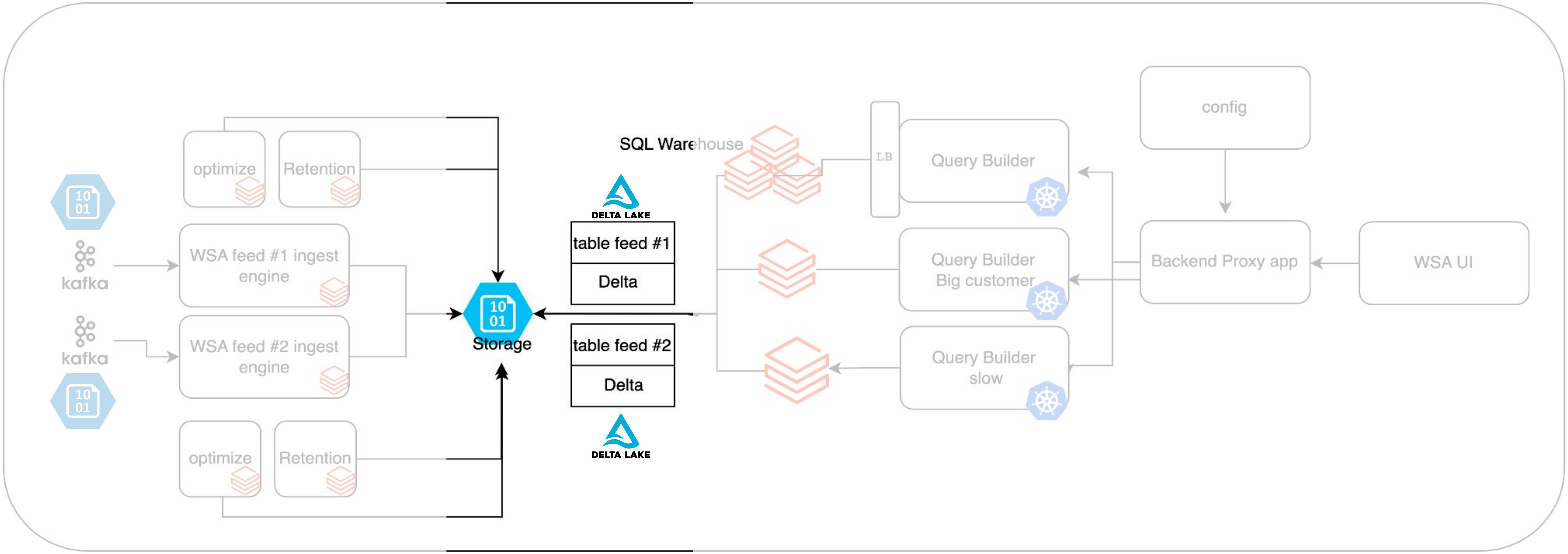
CSI High-level Architecture



WSA Architecture



WSA Architecture



Akamai

WSA Tables

- **Huge** tables
  - Over 6PB in total

WSA Tables

- **Huge** tables
  - Over 6PB in total
- Table format is **Delta Lake**
  - 1 of 3 leading Open Table Formats
    - Alongside Apache Hudi, Apache Iceberg
  - Brings **reliability** to data lakes (e.g ACID transactions)
  - Uses versioned **Parquet** files to store the data
  - Also **stores a transaction log**
    - To keep track of all the commits made to the table or blob store directory
  - Has a large ecosystem

WSA Tables

- **Huge** tables
  - Over 6PB in total
- Table format is **Delta Lake**
  - 1 of 3 leading Open Table Formats
    - Alongside Apache Hudi, Apache Iceberg
  - Brings **reliability** to data lakes (e.g ACID transactions)
  - Uses versioned **Parquet** files to store the data
  - Also **stores a transaction log**
    - To keep track of all the commits made to the table or blob store directory
  - Has a large ecosystem
- Storage type is **Azure Data Lake Storage Gen2**

Storage Limits

Facts

Cloud storage has capacity limits - ingress, egress, TPS

Storage Limits

Problem

- We started seeing a lot of **throttling & server busy errors** from Azure storage
    - ~300K/day
- That had a **negative impact** on ingest, optimize and query time

Storage Limits

Solution #1 – Regional Storage

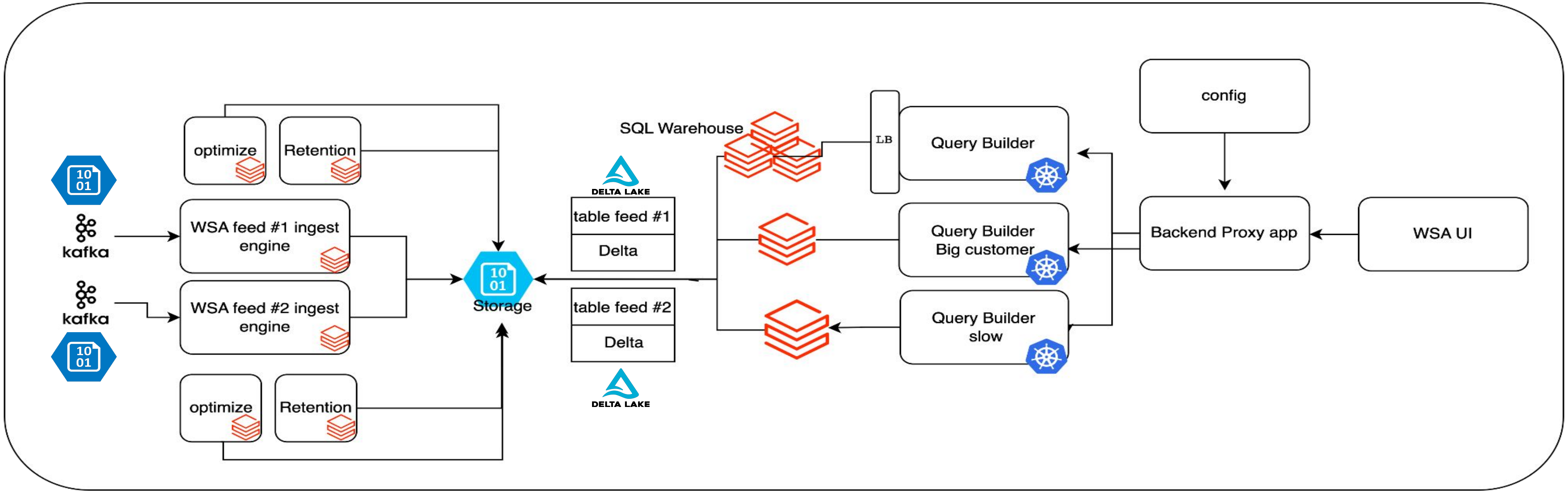
A preview feature (hidden feature) – a multi-cluster storage

Account name	Current capacity	Ingress (Gbps)	Egress (Gbps)	TPS
Input storage – 6-7 clusters	242.76 TiB	430	860	50k
Output storage – 9-10 clusters	5.93 PiB	540	1080	50k

Storage Limits

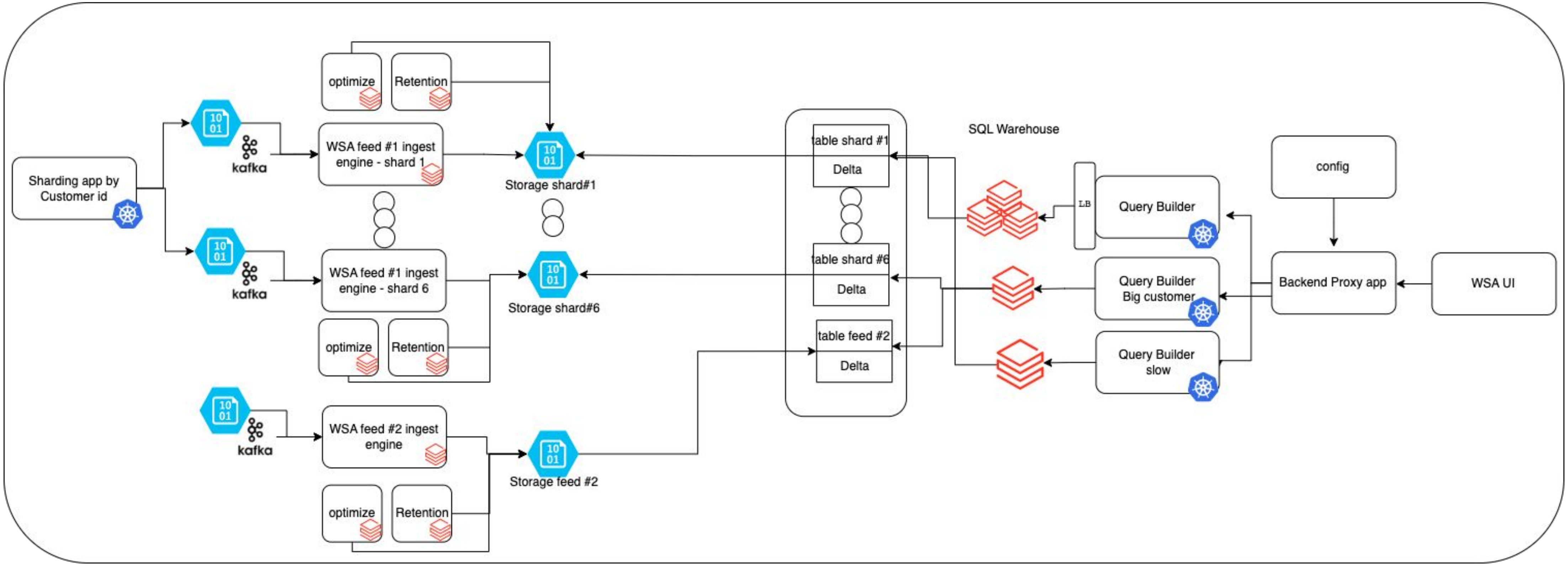
Solution #2 – Sharding

So… What did we have until now?



Storage Limits

Solution #2 – Sharding



Storage APIs

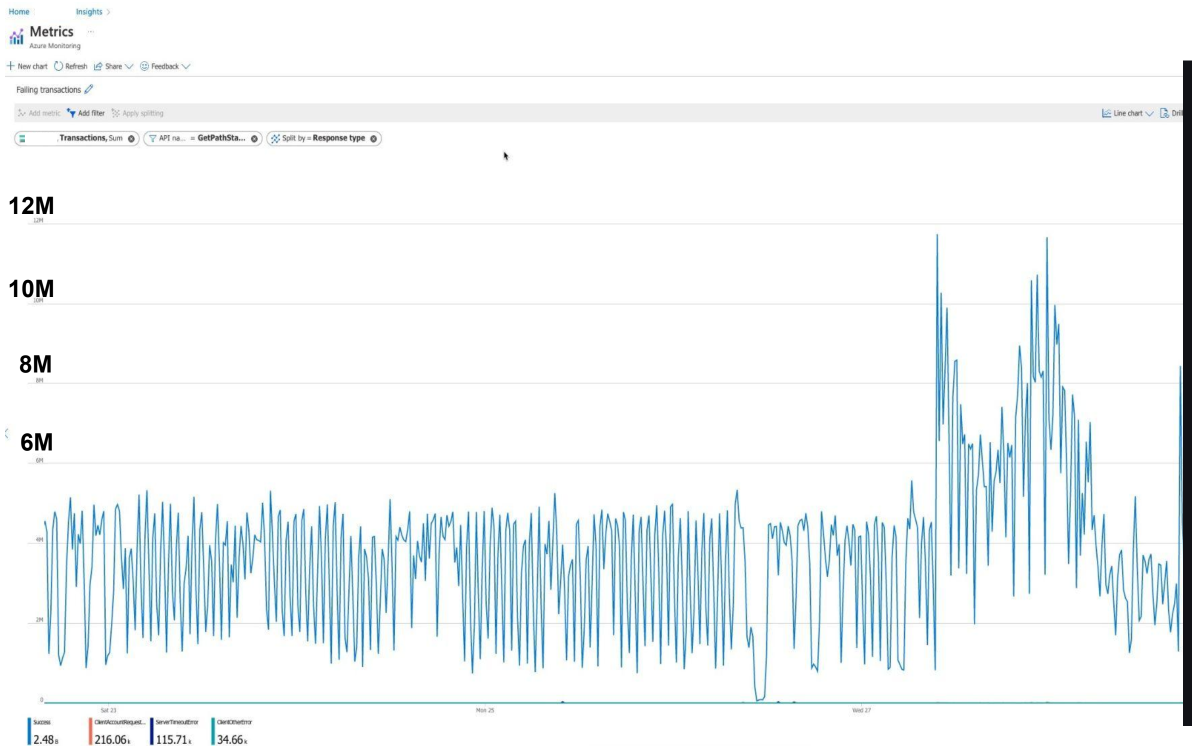
Facts

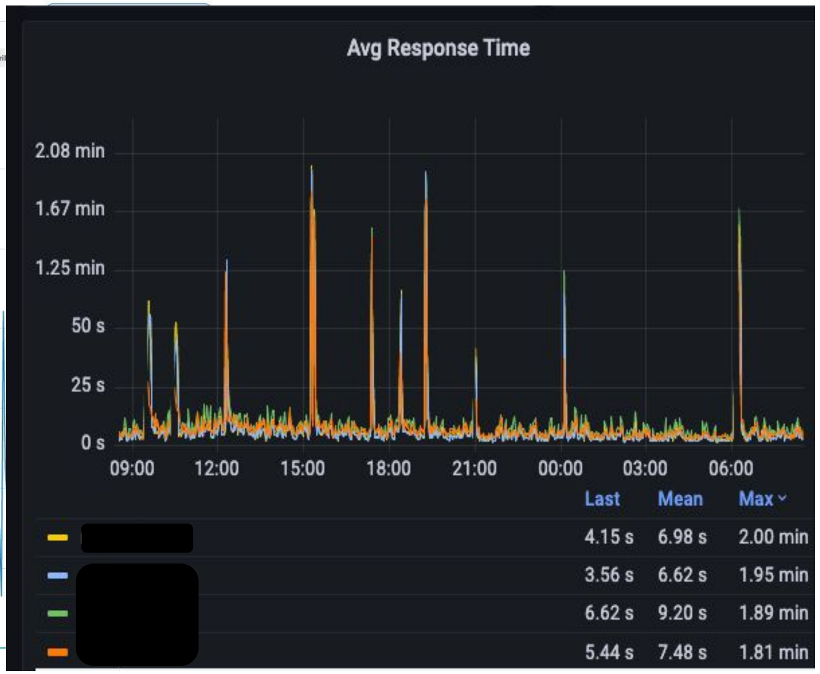
Excessive number of invocations of the `GetPathStatus` storage API

Storage APIs

Problem

Negative impact on performance – query and ingest delays



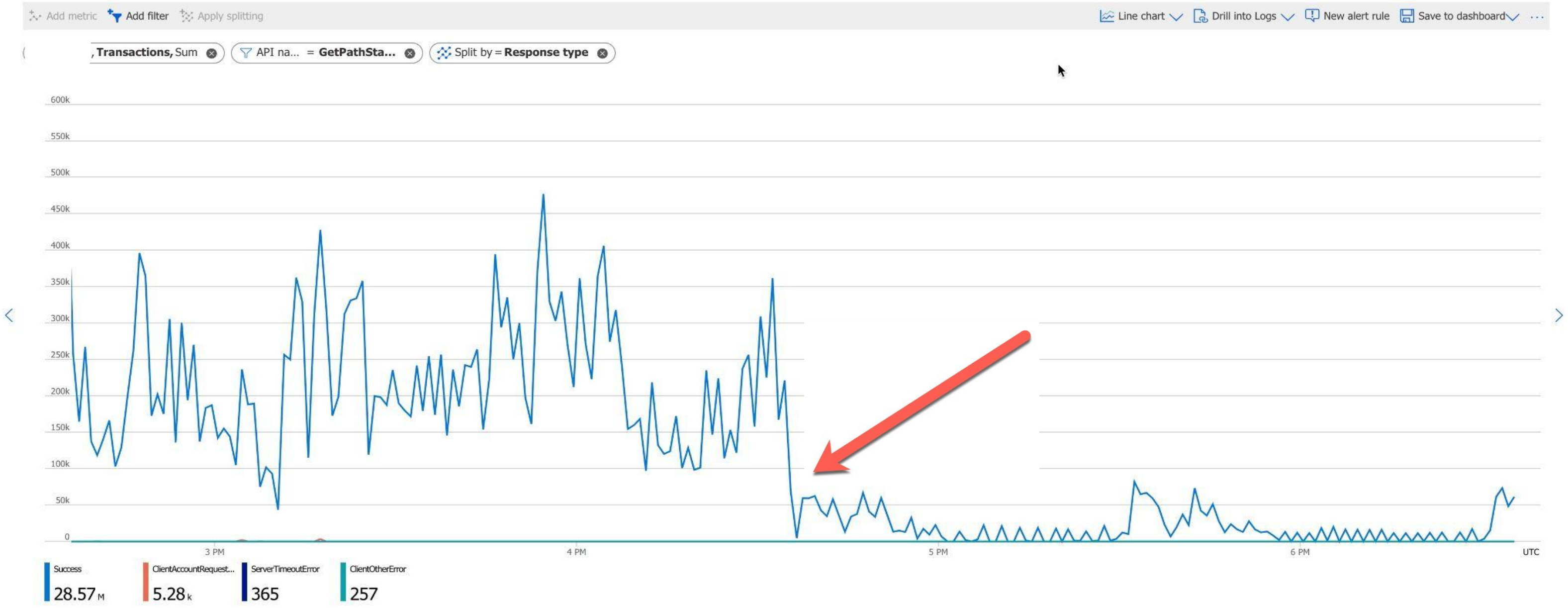


Akamai

Storage APIs

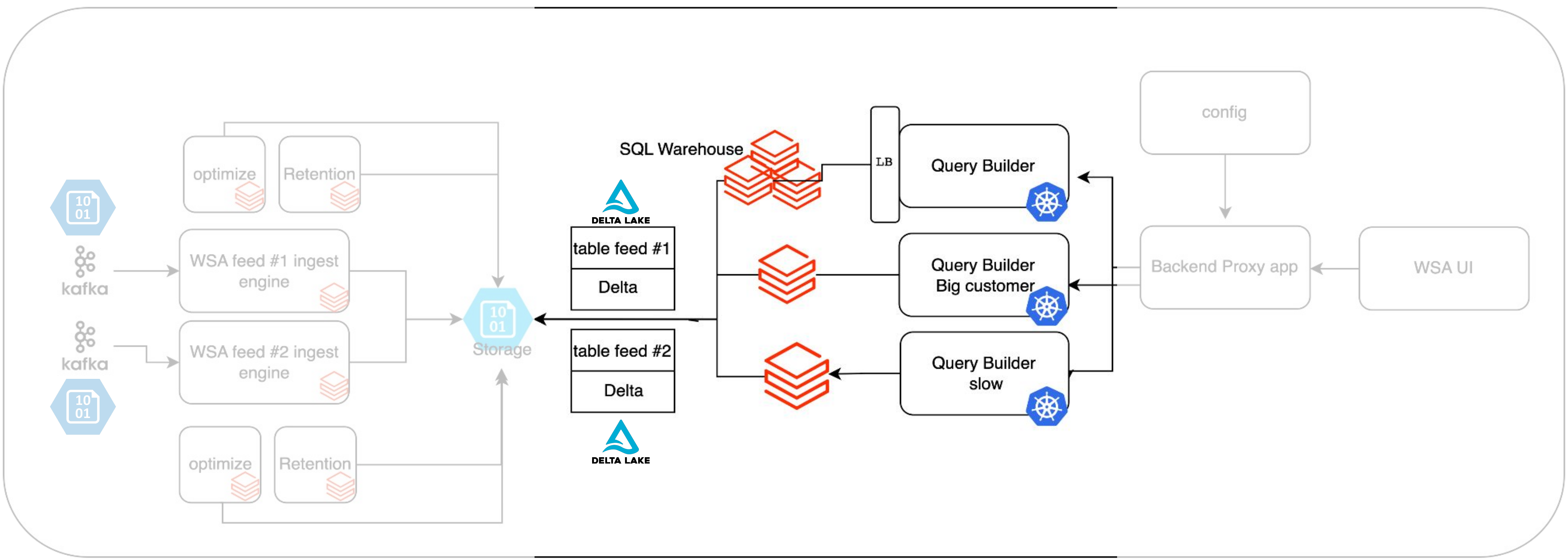
Solution and Recommendations

- Databricks updated Azure Storage APIs to newer ones in DBR
- **Upgrade DBR** version to +11.2



Akamai

WSA Architecture



Strict Query SLA

Facts

WSA needs to execute queries with strict SLA

- For different use cases - e.g aggregated, raw data
- On up to the last 31 days of data

Strict Query SLA

Problem

Most queries were taking
10s of seconds or even minutes
– even after OPTIMIZE

Strict Query SLA

Problem

Most queries were taking **10s of seconds or even minutes** – even after OPTIMIZE

Solution and Recommendations

- Combining **Regional Storage and Sharding**
  - Allowed us to support significantly more egress and TPS
- Using Databricks Photon
- Building an **in-house Load Balancer** on All-Purpose/SQL Warehouse
- **Sampling**

Sampling

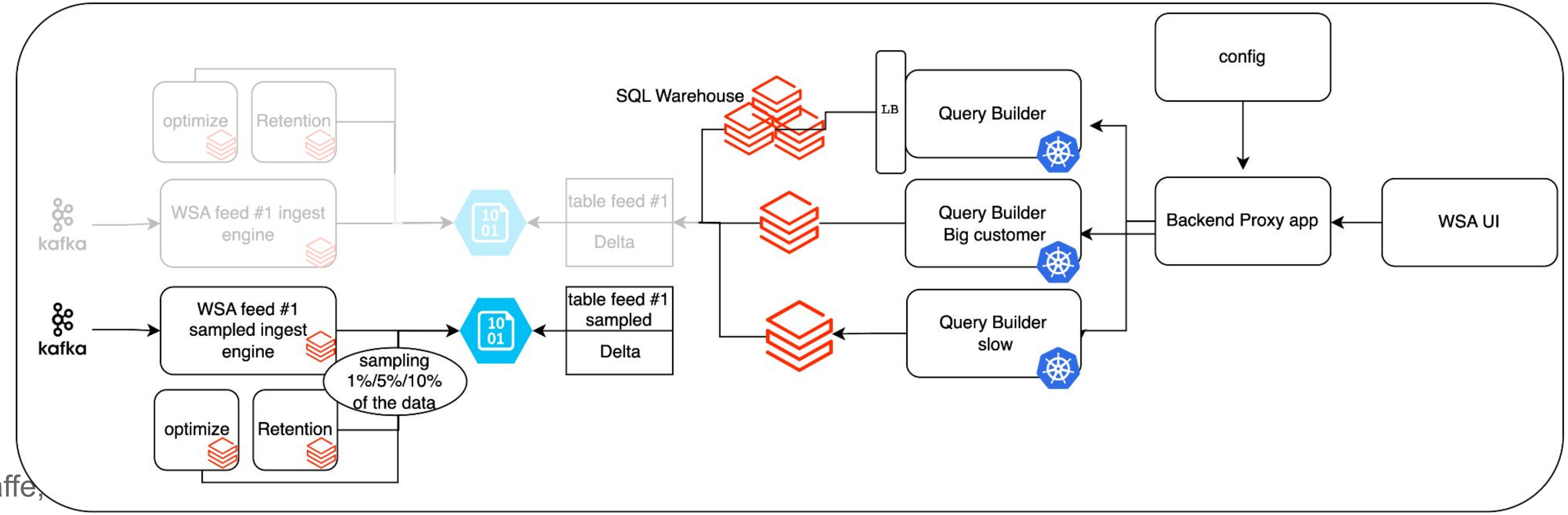
Main Goals

- Better query response time
- Cost reduction
  - By reducing the number and size of query clusters
- Reduce issues in Storage

Sampling

Creating a Sampled Dataset – 1%/5%/10% of the Data

- Redirection to fast query dataset based on decision tree
  - Specific APIs, specific filters and etc.
- By default, the user will query the fast query dataset
  - Users will still able to query the full dataset
- Currently based on a statistical model
  - In the future, based on an ML model



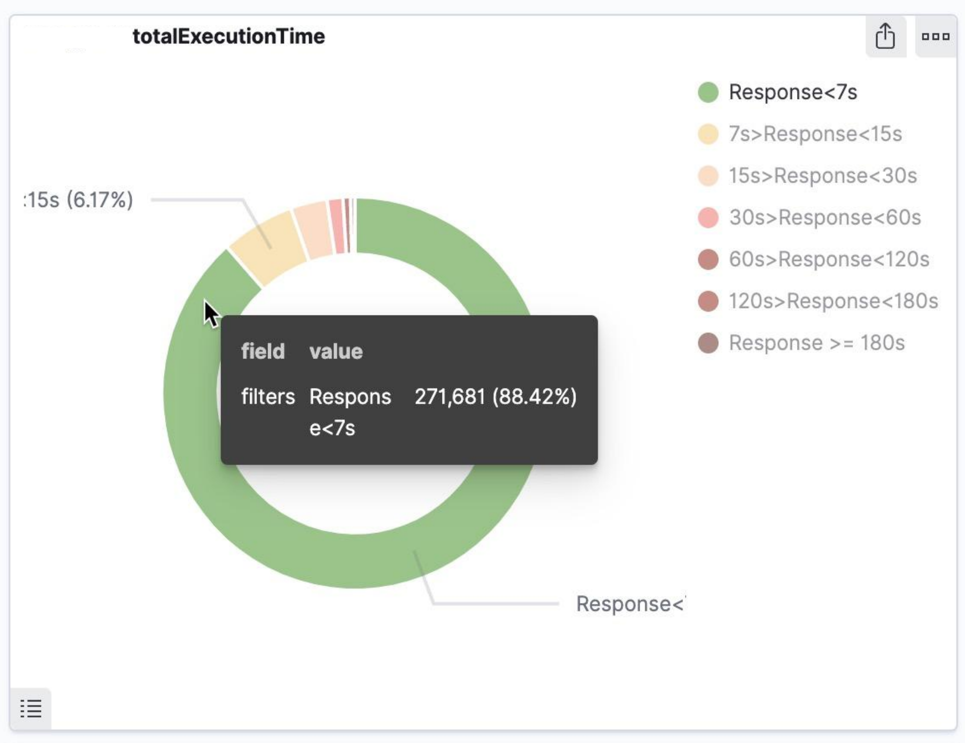
Akamai

Strict Query SLA

Results

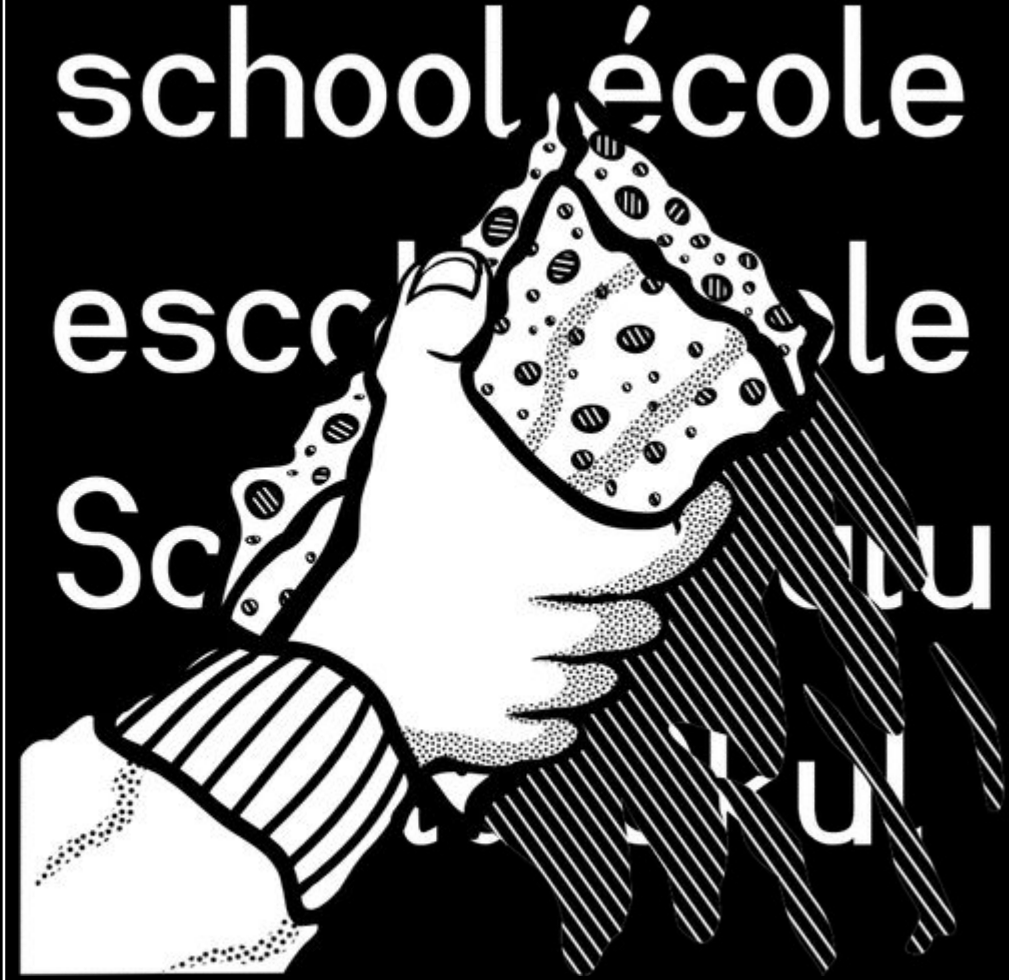
Significantly improved query response times

- From 10s of seconds (or even minutes) to **less than 7 seconds** for **~85%** of the queries



Tips for Optimizing a Massive-Scale Data Infrastructure

- Data Retention
- Compression Formats



Data Retention

Deleting Old Records

- Deleting data older than a defined threshold (a.k.a TTL) is very common
- Delta Lake does **not** support
  - `ALTER TABLE table_name DROP PARTITION`
  - `TRUNCATE TABLE table_name PARTITION clause`
- Instead, it provides a **`DELETE FROM`** statement, e.g
  - `DELETE FROM table_name WHERE event_time < (now() - INTERVAL '31' DAY)`
  - Where `event_time` is `TIMESTAMP`

Akamai

Data Retention

Potential Impact of `DELETE FROM`

- But… This can actually create **new files** in your Delta table!
  - `DESCRIBE HISTORY table_name`

```
operation || operationParameters || operationMetrics
========||===================||================
DELETE || {"predicate":"["(my_table.event_time < TIMESTAMP
'2023-04-03 12:45:34.813')"]"} ||
{"executionTimeMs":"2354",...,"numAddedFiles":"3","numCopiedRow
s":"1321","numDeletedRows":"60654",...,"numRemovedFiles":"155",
"rewriteTimeMs":"1438","scanTimeMs":"916"}
```

Data Retention

Why?

- Parquet files are **immutable**
- Hence, Delta Lake has to
  - Read the existing Parquet file(s)
  - Filter out the records to be deleted
  - Write new Parquet file(s) with the remaining records

Data Retention

How to Avoid Creating New Files in this Use-case?

- E.g `table_name` includes
  - `event_time` - `TIMESTAMP`
  - `event_day` - `DATE`
- Re-write your `DELETE FROM` statement, to match the partition columns
  - `DELETE FROM table_name WHERE event_time < (now() - INTERVAL '31' DAY)`

Data Retention

How to Avoid Creating New Files in this Use-case?

- E.g `table_name` includes
  - `event_time` - `TIMESTAMP`
  - `event_day` - `DATE`
- Re-write your `DELETE FROM` statement, to match the partition columns
  - ~~`DELETE FROM table_name WHERE event_time < (now() - INTERVAL '31' DAY)`~~
  - `DELETE FROM table_name WHERE event_day < to_date(now() - INTERVAL '31' DAY, 'YYYY-MM-DD')`

Akamai

Data Retention

Rewrite Impact

- Let’s check the table history now
  - DESCRIBE HISTORY table_name

```
operation || operationParameters || operationMetrics
========||===================||================
DELETE || {"predicate":"["(my_table.event_day < DATE
'2023-04-03')"]"} ||
{"executionTimeMs":"26",...,"numAddedFiles":"0","numCopiedRows"
:"0","numDeletedRows":"8745",...,"numRemovedFiles":"78","rewrit
eTimeMs":"0","scanTimeMs":"25"}
```

Data Retention

Rewrite Impact - Full Scale

- For our petabytes Delta Lake tables
  - Partitioned by `<customer ID, date>`
  - `DELETE` job is executed on a daily basis
- Achieved:
  - Execution time (per job)
    - 4-5 hours -> ~20 minutes
  - Costs (in total)
    - ~$500/day (max.) -> ~$10/day
  - Significantly less IOPS on storage

Tips for Optimizing a Massive-Scale Data Infrastructure

- Data Retention
- Compression Formats

Compression Formats

Facts

- WSA writes and reads TBs of data to/from Delta Lake tables stored in ADLS
- **Snappy** is the **default** compression format for **Parquet** files written by **Spark**
  - Spark supports other compression formats, e.g LZ4, zstd, gzip, etc.

Compression Formats

Main Goal

- Reduce the amount of data written to/read from ADLS
  - Reduces IOPS and costs

Compression Formats

ZSTD

- A **modern** compression format developed by Meta
- Has a **promising compression ratio**
  - Supports 22 compression levels, the default is 3
- Databricks **Photon** has a **built-in support** of optimized execution for zstd

Compression Formats

ZSTD Setup

Setup is **easy**

- `spark.conf.set("spark.sql.parquet.compression.codec", "zstd")`
  OR
  `spark.sql.parquet.compression.codec zstd`
  in the Spark Config of the Databricks cluster
  OR
  `df.write.mode("overwrite").format("delta")`
  `.option("compression", "zstd")``.saveAsTable("my_table")`

Compression Formats

ZSTD Setup

Controlling the **specific zstd level**

- Requires the **addition** of
  `parquet.compression.codec.zstd.level 19`
  in the Spark Config of the Databricks cluster

Compression Formats

ZSTD Setup

It's **very important** to set it up on **all jobs that manipulate the data**

- Remember - even Delta Lake's **`DELETE FROM`** **statement can potentially create new files!**

Compression Formats

ZSTD Benchmark

- Benchmarked **snappy vs 3 levels of zstd** in **pre-production**
- Results:

Comparison aspect vs Snappy	**Zstd (level 3 –** default**)**	**Zstd (level 11)**	**Zstd (level 19)**
Used storage	~50%	>50%	>50%
Ingest performance – micro-batch mean duration	~1.1X	~1.3X	~2X
Query performance	Roughly the same	Roughly the same	N/A

Compression Formats

ZSTD Actual Results

Snappy vs zstd (default level) in **production:**

Comparison aspect vs Snappy	**Zstd (level 3)**
Used storage	~35%
Ingest performance - micro-batch mean duration	Roughly the same
Query performance	Roughly the same

One Last Re-Architecture (For Now...)

- Akamai recently announced it's new offering, Akamai Connected Cloud (formerly Linode)
- As part of our ongoing efforts to **optimize efficiency**, and the "drinking your own champagne" mindset, we're in the process of moving some workloads to Akamai's cloud.
  We are applying the lessons learned from our Azure Databricks journey, e.g
  - Using zstd compression format where applicable
  - Sharding our ingest pipelines to avoid throttling

Summary

Processing

- Using Kafka to store only **"pointers"** to raw data files
- Splitting ingest pipeline to overcome storage limitations – a.k.a **Sharding**
- Avoid excessive Storage API invocations where possible

Summary

Processing

- Using Kafka to store only **“pointers”** to raw data files
- Splitting ingest pipeline to overcome storage limitations – a.k.a **Sharding**
- Avoid excessive Storage API invocations where possible

Storing

- Choosing the right **storage type** for each workload
- Using an **Open Table Format** (e.g Delta Lake)
- Leveraging advanced, preview features such as Regional Storage
- Properly **deleting** old data
- Using the appropriate **compression format**

Summary

Processing

- Using Kafka to store only **"pointers"** to raw data files
- Splitting ingest pipeline to overcome storage limitations – a.k.a **Sharding**
- Avoid excessive Storage API invocations where possible

Storing

- Choosing the right **storage type** for each workload
- Using an **Open Table Format** (e.g Delta Lake)
- Leveraging advanced, preview features such as Regional Storage
- Properly **deleting** old data
- Using the appropriate **compression format**

Analyzing

- Sampling can improve query performance with a little impact on results' accuracy

Want To Know More?

- **Women in Big Data**
  - A world-wide program that aims:
    to inspire, connect, grow and champion success of women in all data domains
  - 50+ chapters and 20,000+ members world-wide
  - Everyone can join (regardless of gender), so find a chapter near you – tinyurl.com/mv4668sy
  - Women in Data+AI panel and luncheon (Thursday, 11:30AM) – tinyurl.com/yc7drfpy

- **Upcoming talks tomorrow**
  - "From Snowflake to Enterprise-Scale Apache Spark™" by **Nic Jansma & Amir Skovronik** (12:30PM) – tinyurl.com/bdhs7dcv
  - "Unleashing the Power of Interactive Analytics at Scale with Databricks & Delta Lake" by **Tomer & myself** (1:30PM) – tinyurl.com/4wzkv6mb
  - "Internet-Scale Analytics: Migrating a Mission Critical Product to the Cloud" by **Yaniv Kunda** (2:30PM) – tinyurl.com/bdp48a43

Thank You!

Your feedback is important to us!
Feel free to reach out 🙂

in [Tomer Patel] [@tomer_patel]

in [Itai Yaffe] [@ItaiYaffe]