STORAGE DEVELOPER CONFERENCE

SDC 23

BY Developers FOR Developers

An AI Inference Engine for Object Storage Systems

Presented by Dan Pollack and Jessica Bresnahan

Outline

- What are we building? – A plugin AI Engine for Object Storage Systems
- Why is it interesting? – Metadata and search
- What does it do? – Automatically inspects file content
- How does it work? – The magic of software
- What's next? – Features to do list
- Q & A

What are we building?

A plugin AI Engine for Object Storage Systems

- Object storage systems store unstructured data
- We know when objects are stored and where but not much else
- The storage system is often used by a different group than the one that manages it
- More awareness of the content is better
- The Engine is still a work in progress

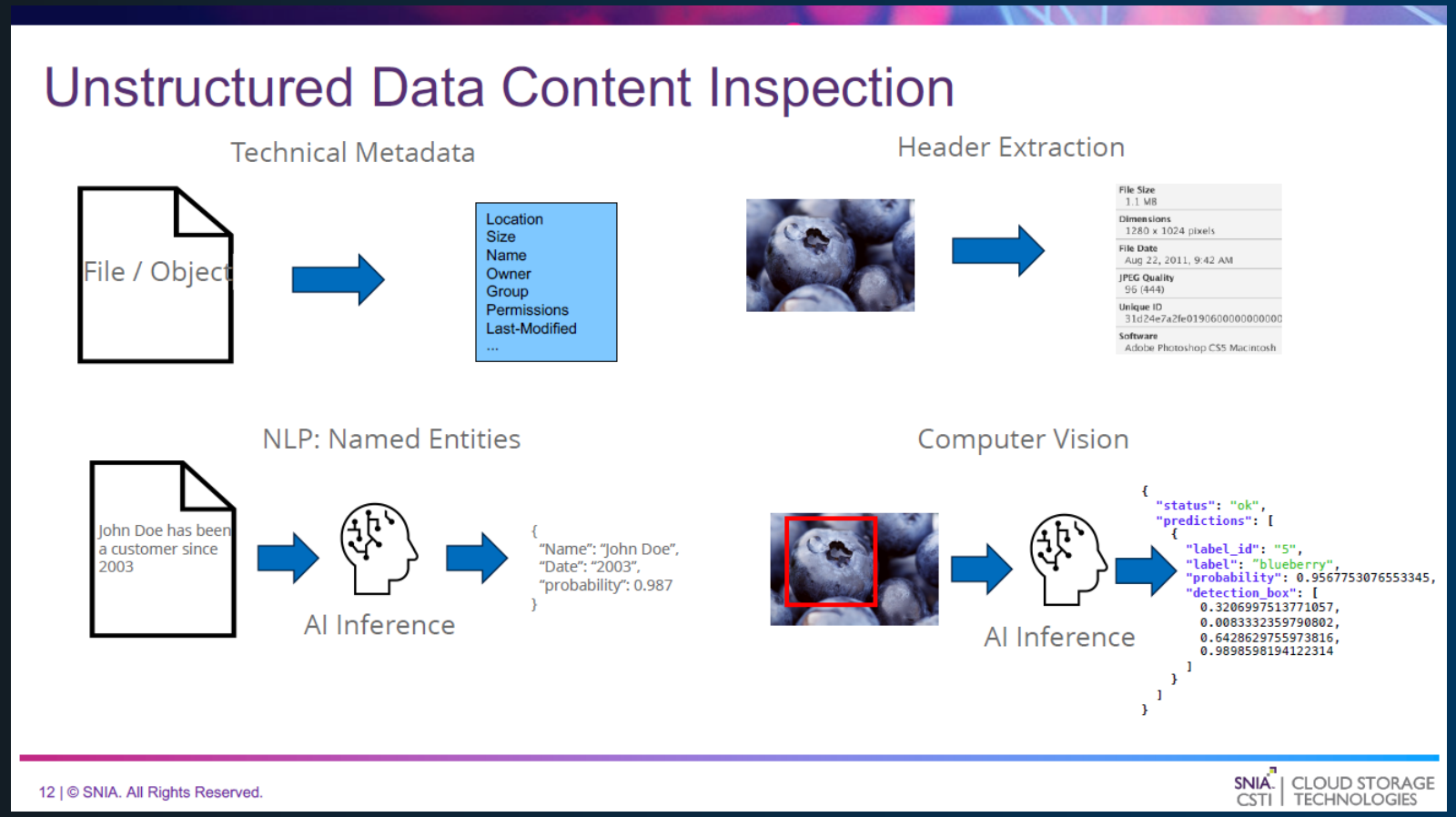
Why is it interesting?

- Content awareness enables better utilization and data management
- Discover multiple data types in objects (e.g text in images)
- Search files for content in addition to more traditional metadata
- Apply policy to data based on content

Why is it interesting?

SNIA CSTI | CLOUD STORAGE TECHNOLOGIES

June 28 2023
Data Fabric: Connecting the Dots between Structured and Unstructured Data



SDC 23

What does it do?

- Watch for object creation events
- Examine object type
- Perform multiple AI inference operations based on expected content
- Perform secondary AI inference operations based on extracted content
- Index the extracted content

How does it work?

- Use native event notification of Object Storage System
  - PoC use webhooks
  - Queues and databases are also options depending on the Object storage system
- On object creation event detect the file type
- Run inference operations based on the file type
- Report the inference results as text
- Flag the operation as completed

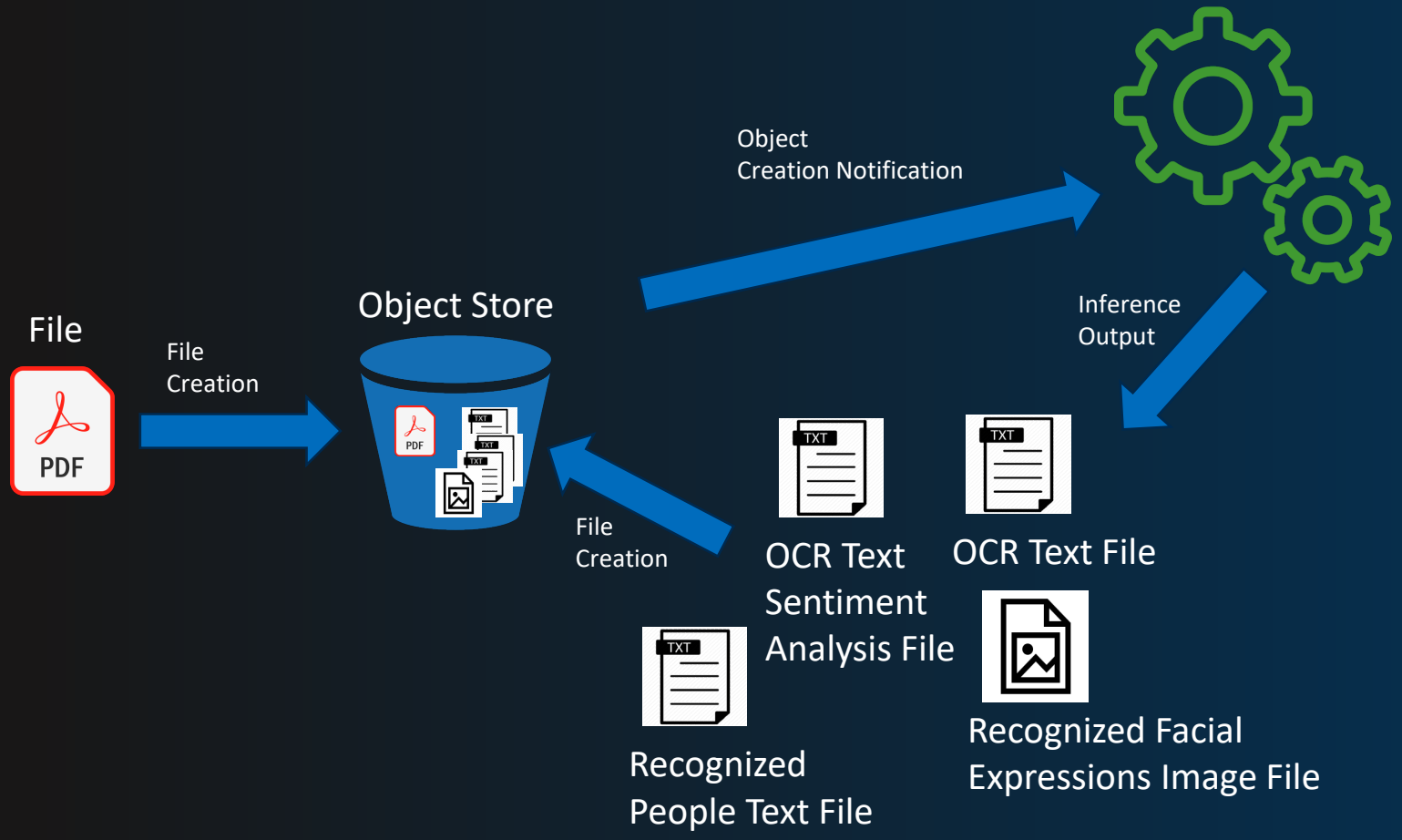
How does it work?

- Use native event notification of Object Storage System
  - MinIO PoC use webhooks
  - Queues and databases are also options depending on the Object storage system – AWS S3,
- On an object creation event detect the file type stored
  - Images
  - Text
  - Audio
  - Video
- Run inference operations based on the file type – local and aaS based operations
  - OCR
  - Object recognition – images and video
  - Object detection – images and video
  - Text in images
  - Audio speech to text
  - Audio from video
  - Images from video – presentation slides
  - Forms recognition
- Report the inference results as text
  - Bounding boxes
  - Object types
  - Text
- Flag the operation as completed

SDC 23

How does it work?

- OCR example



How does it work?

- OCR example

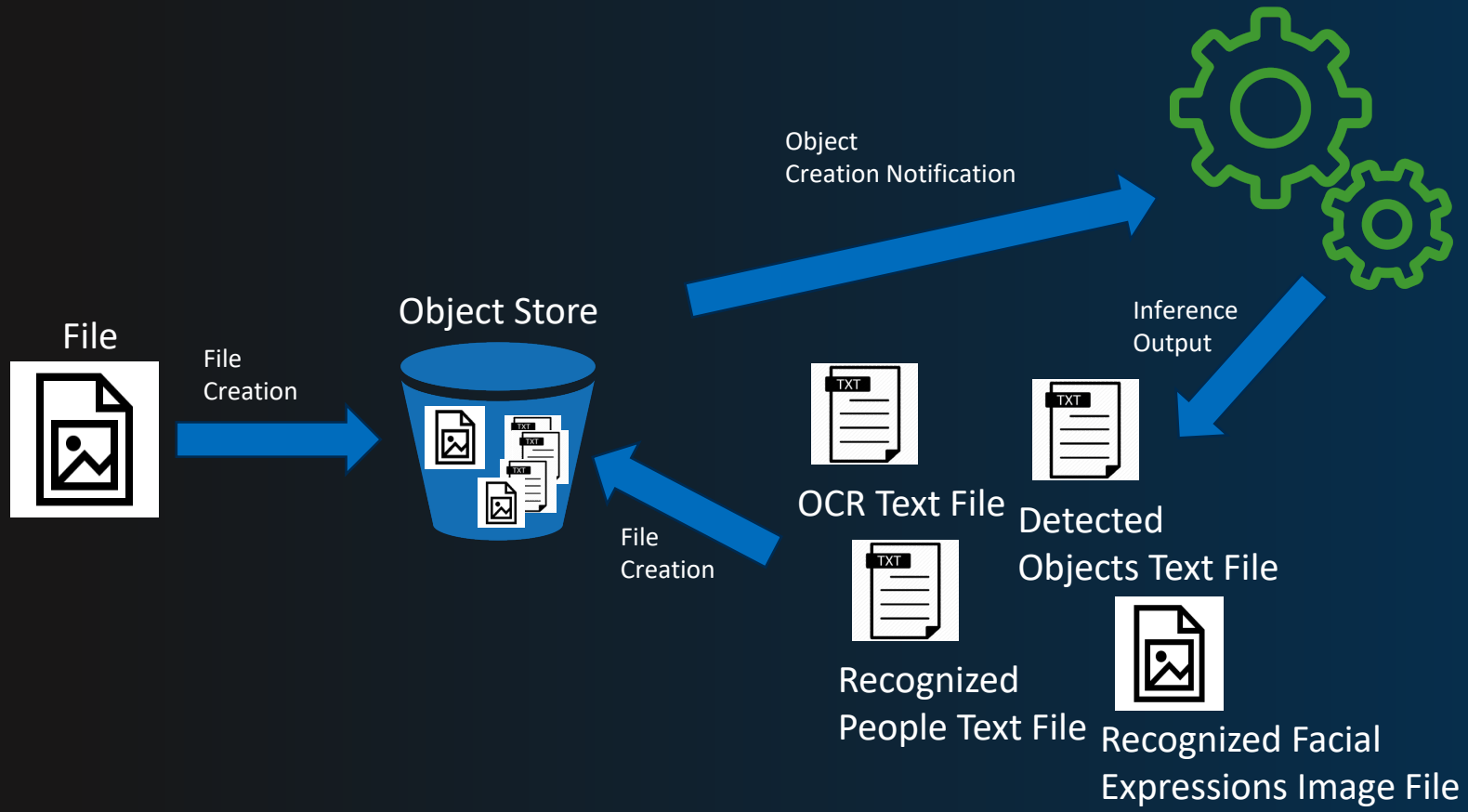
1 faces detected.
0: 442 118 64 89
458 154 488 152
473 170 463 183
489 181

Face | Bounding Box | Emotion
Face | 0: 442 118 505 206 | happy.

Foreword by Jim Bridenstine Pushing the boundaries of space exploration, science, and technology once again, America is on the verge of exploring more of the Moon than ever before. This new era of lunar exploration is called Artemis. Named after the twin sister of Apollo, she is the Goddess of the Moon, and we are the Artemis Generation. NASA is building a coalition of partnerships with industry, nations and academia that will help us get to the Moon quickly and sustainably, together. Our work to catalyze the U.S. space economy with public-private partnerships has made it possible to accomplish more than ever before. The budget we need to achieve everything laid out in this plan represents bipartisan support from the Congress. Today, NASA is delivering more missions, more science, and more innovation at a better value for the American taxpayer than at any point in the agency's history and with half the buying power than 1964, when Apollo development was at its peak. We thank the President, the Congress, the American taxpayers, and the emerging space industry for the combined efforts to strengthen our nation's space program. Under the Artemis program, humanity will explore regions of the Moon never visited before, uniting people around the unknown, the never seen, and the once impossible. We will return to the Moon robotically beginning next year, send astronauts to the surface within four years, and build a long-term presence on the Moon by the end of the decade. I am proud to share NASA's Artemis Plan—this is how we will go to the Moon once again. And how we will use the Moon as the stepping stone for our next greatest leap—human exploration of Mars. We are going, and we go together. Ad lunam! NASA Administrator Jim Bridenstine

How does it work?

- Image example



How does it work?

- Image example

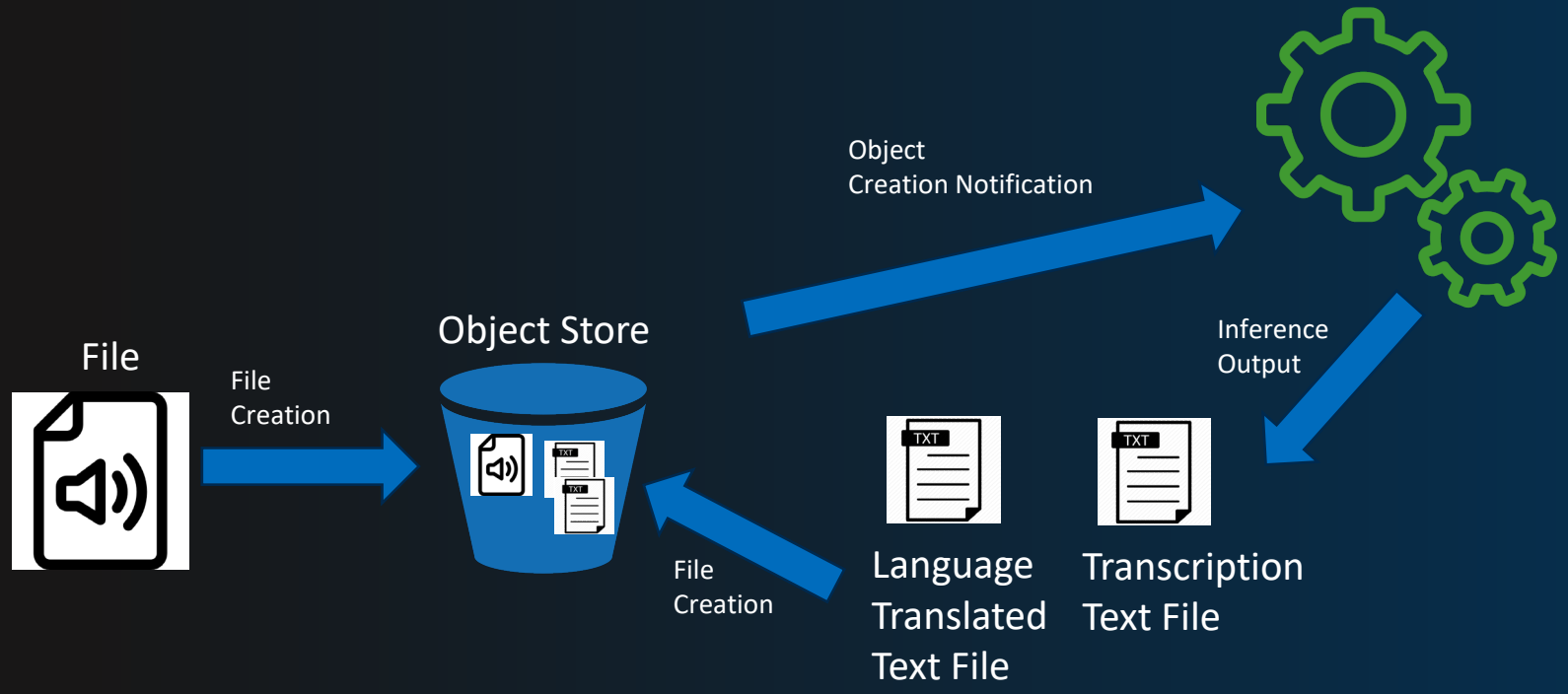
1 faces detected.
0: 273 85 118 142
289 148 332 143 300
176 301 201 337 196

Face | Bounding Box | Emotion
Face | 0: 273 85 391 227 | neutral.

Detected Text: your: [630 200] [630 181] [675 181][675 200]party: [832 200] [832 179] [885 179][885 200]extraordinary: [720 198] [720 178] [833 179][833 199]own: [675 199] [675 180] [723 180][723 199]to: [555 197] [555 181] [580 181][580 197]throw: [576 199] [576 178] [632 178][632 199]and: [554 178] [554 161] [596 161][596 178]prizes: [833 178] [833 159] [886 159][886 178]taylor: [746 179] [746 159] [793 159][793 179]cool: [588 177] [588 160] [630 160][630 177]aand: [704 178] [704 159] [747 159][747 178]tcoke: [657 177] [657 157] [713 159][713 178]diet: [626 178] [626 158] [667 158][667 178]swilt: [790 178] [790 156] [840 156][840 178]wina: [706 157] [706 138] [756 138][756 157]andy: [594 158] [594 138] [636 138][636 158]giltcard: [813 159] [813 136] [889 136][889 159]lycucouldy: [624 159] [624 136] [714 136][714 159]enter: [549 159] [551 135] [602 137][600 160]s25000: [751 159] [751 135] [821 133][825 157]taylorswift: [533 126] [533 44] [898 43][898 126]coke: [0 143] [0 32] [276 32][276 143]diet: [174 66] [174 33] [258 33][258 66]in: [699 52] [699 37] [722 37][722 52]entertain: [616 56] [616 33] [706 30][706 55]style: [718 56] [718 31] [770 31][770 56]with: [764 54] [764 30] [810 30][810 54]extraordinary: [46 41] [46 27] [165 28][165 42]stayy: [19 41] [19 27] [52 27][52 41]

How does it work?

- Audio example

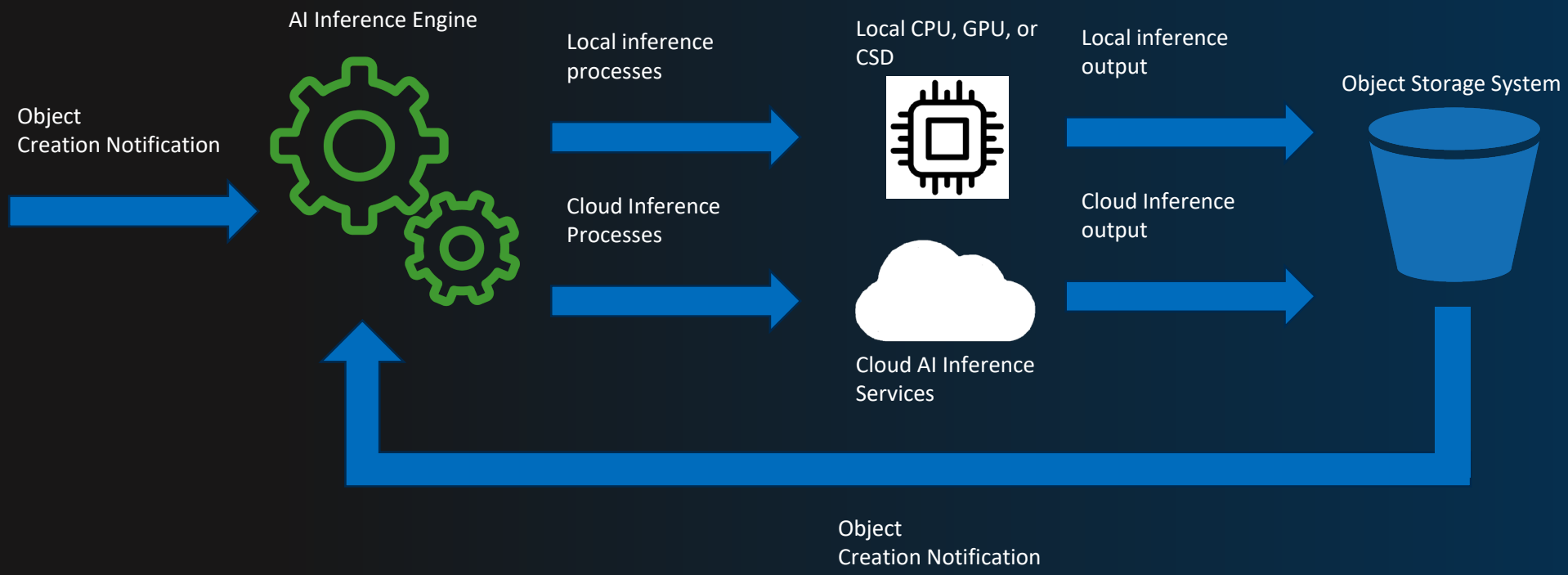


How does it work?

- Video presentation example

How does it work?

- AI processing

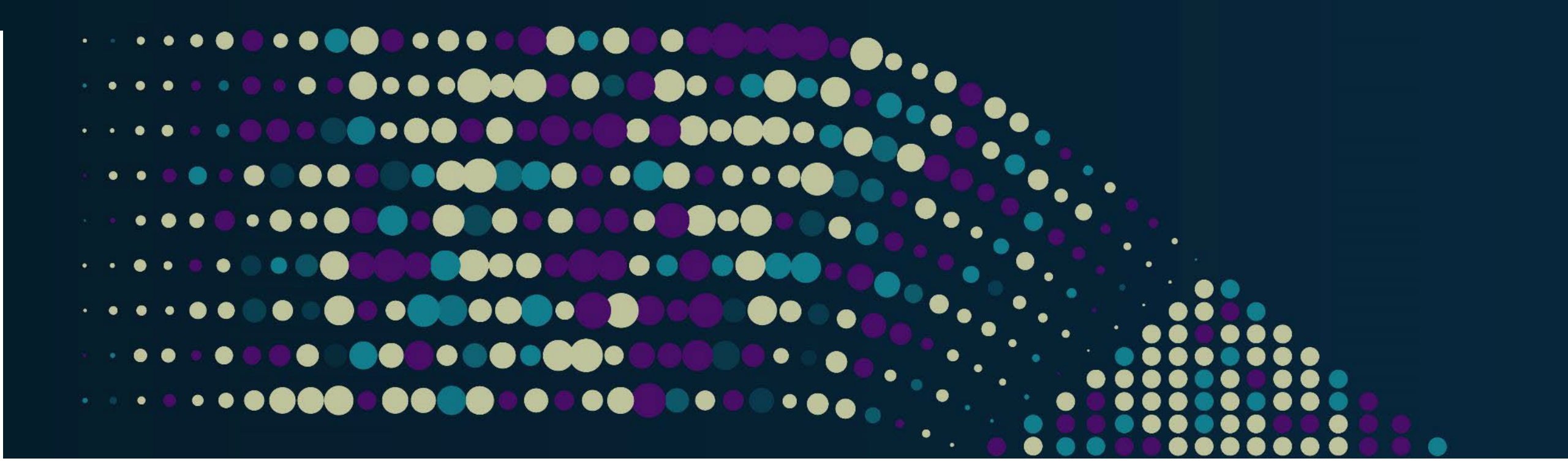


What's next?

- Background process to check for objects that failed to complete or were missed at creation time
- Tool to run new inference operations on some or all old objects on demand
- Access to perform inference operations on encrypted content if authorized
- Auto-scale server backend as load increases
- Load balancer for incoming inference operations
- Run AI Engine on alternative compute - Computational Storage Devices, External CPU/GPU complexes

Questions?

Questions?



Please take a moment to rate this session.

Your feedback is important to us.

