Expanso: The **Compute** Over **Data Platform**

Expanso is the leading solution for distributed computing across cloud, on-premise, and edge environments.

David Aronchick, CEO
aronchick@expanso.io

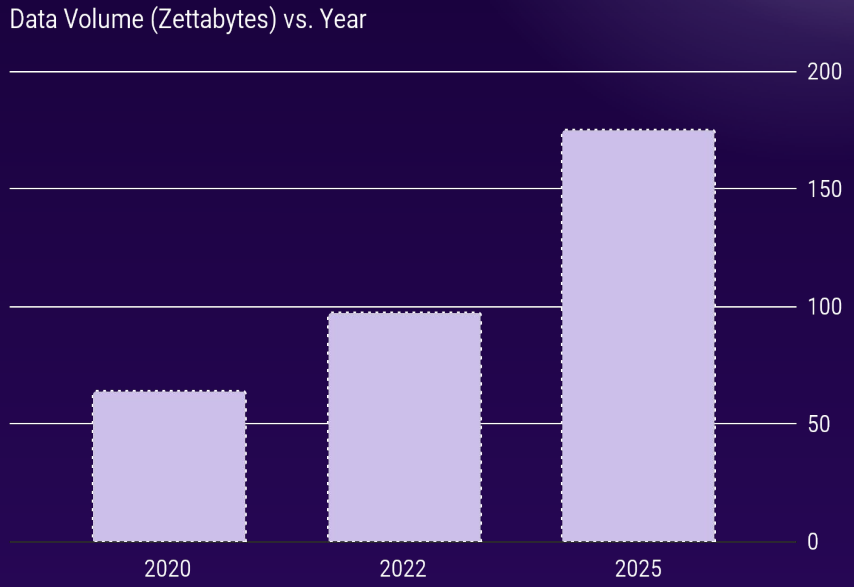
expanso

Data is exploding

and mostly goes unused.

- Global data reaches **175 zettabytes** this year (IDC).
- **60%+** of enterprise data is unstructured and is generated outside of traditional data centers (IDC).
- **73%** of organizations can't keep up with data processing demand (Gartner).

Traditional, centralized data approaches can't keep up

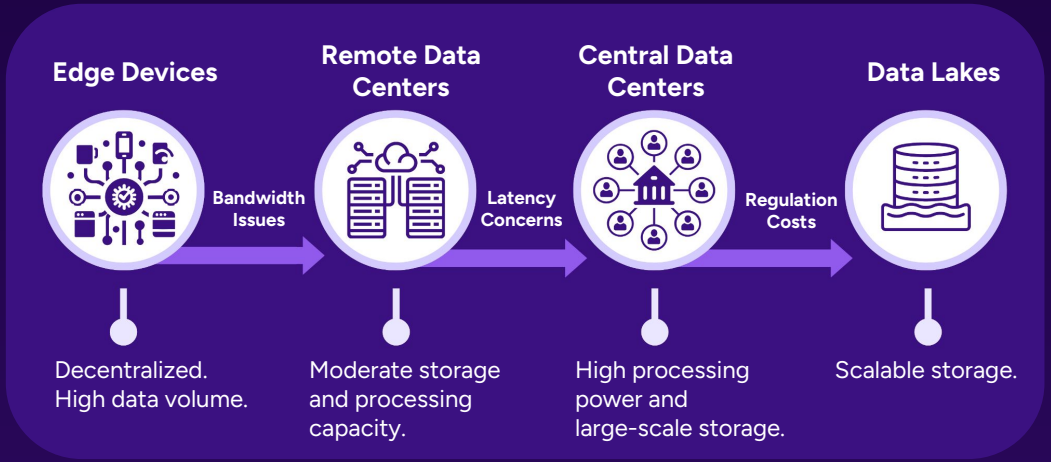


expanso

Centralizing data adds risks and bottlenecks

- **Slow & Costly:** Distributed data pipelines result in slower answers and updating large datasets is a logistical nightmare.
- **Security:** Centralizing data creates a vulnerable attack surface.
- **Regulations:** GDPR, CCPA, and other regulations favor localized data processing.

Organizations become slower and more at risk by **EXCLUSIVELY** relying on a data lake/warehouse



expanso

Modern **path for data processing?**

Raw Data	Ingestion	Storage	Transform	Load	Integrate	Visualize

Multi-Region

Multi-Cloud

Edge

IoT

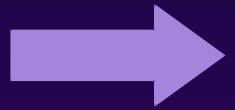
On-Prem

Embedded

???

???

???



snowflake

amazon REDSHIFT

databricks

mongoDB.

ORACLE

Azure

Google Cloud

dremio

Observable

Power BI

+ableau

Looker

Qlik

sigma

expanso

Meet Expanso's Bacalhau platform

1

Faster:
Compute Occurs on Location

- Reduce data transit and storage costs
- Decode/Encode
- Real-time insights

2

Cheaper:
Move Only Results

- Leverage unused hardware
- Reduce data transport & data center costs

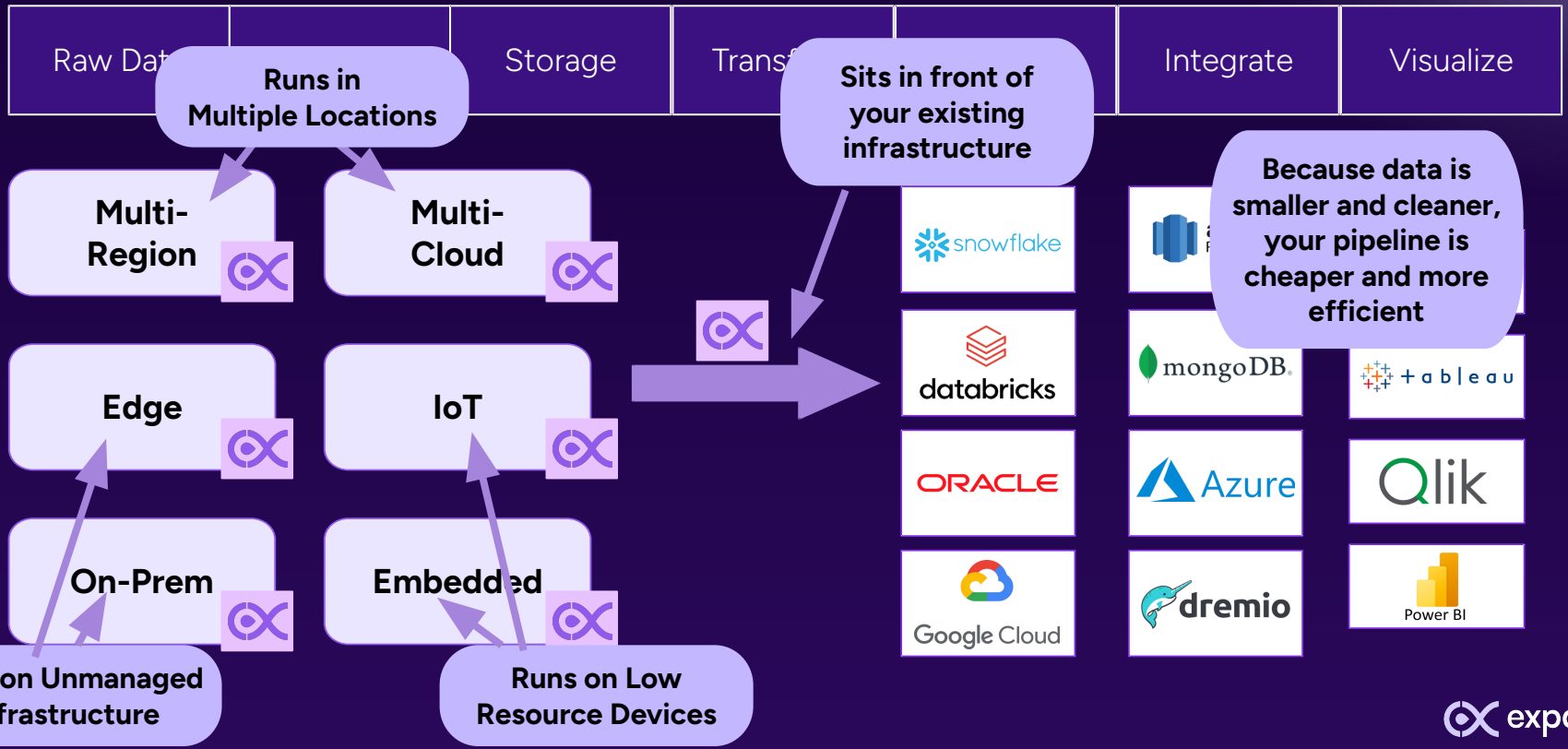
3

More Secure:
Access and Limited Data Movement

- Limit regulatory issues around data movement
- Improve visibility & enable auditing

expanso

Bacalhau runs where your data is



expanso

Use Case:

The Distributed Data Warehouse

expanso

Use Case: The Distributed Data Warehouse

The Problem with Traditional Data Warehouses

- Data Sources: Data comes from everywhere (applications, logs, IoT, etc.).
- ETL Pipelines: Complex and costly
- Centralized Warehouse: All data is stored centrally, regardless of usage frequency.

Key Pain Points

- **Costs**: Storing and managing ALL data centrally = High egress, storage, and security costs.
- **Performance**: Slow queries due to resource contention, bottlenecks in processing and analysis.
- **Complexity**: Growing data volumes increase operational overhead; Even “easy” answers require significant challenges to answer

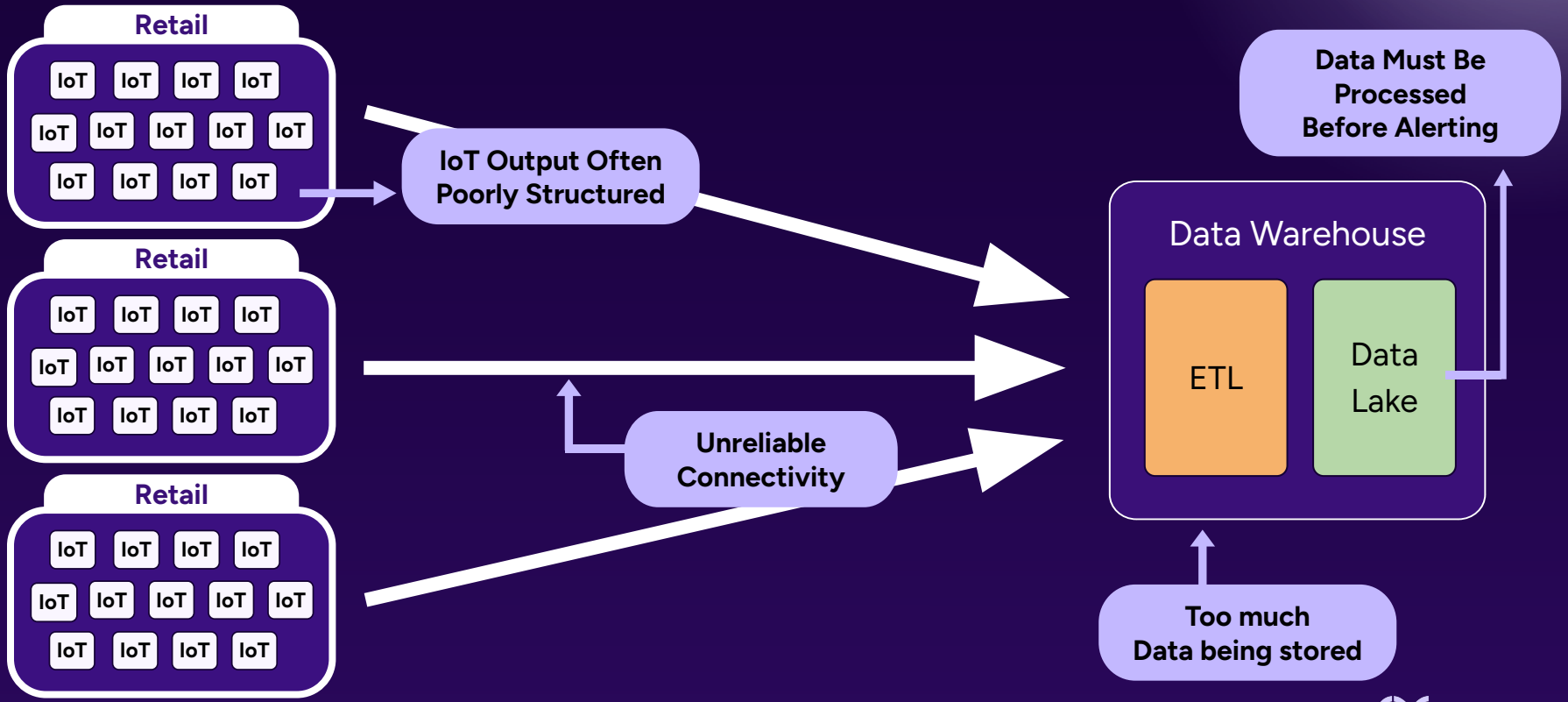
The Impact

- Inefficient Spending: Paying for storage and compute you don’t need.
- Slower Insights: Delayed decision-making.
- Operational Overhead: Managing ETL pipelines and warehouse scaling becomes a nightmare.

Before Expanso

expanso

Use Case: The Distributed Data Warehouse

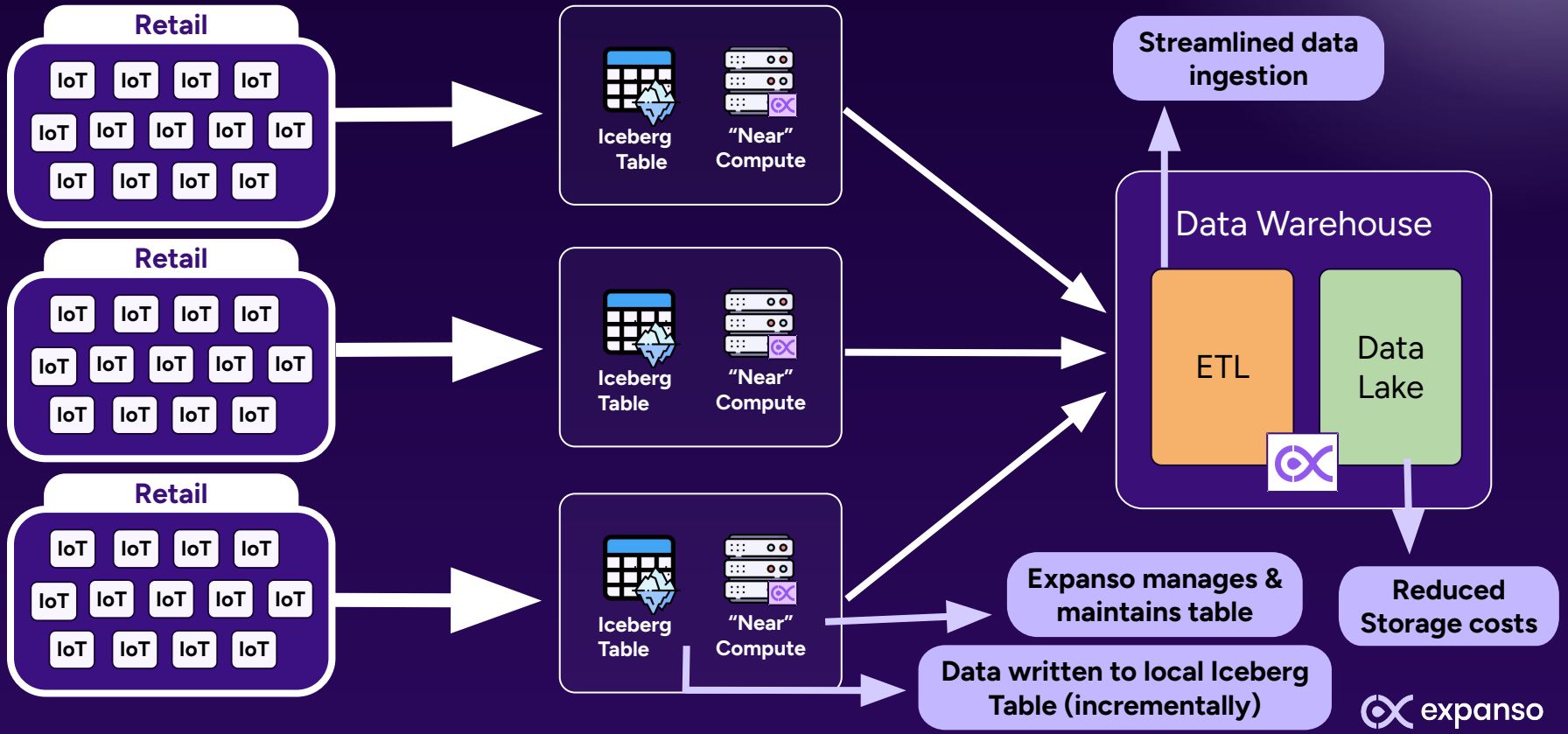


expanso

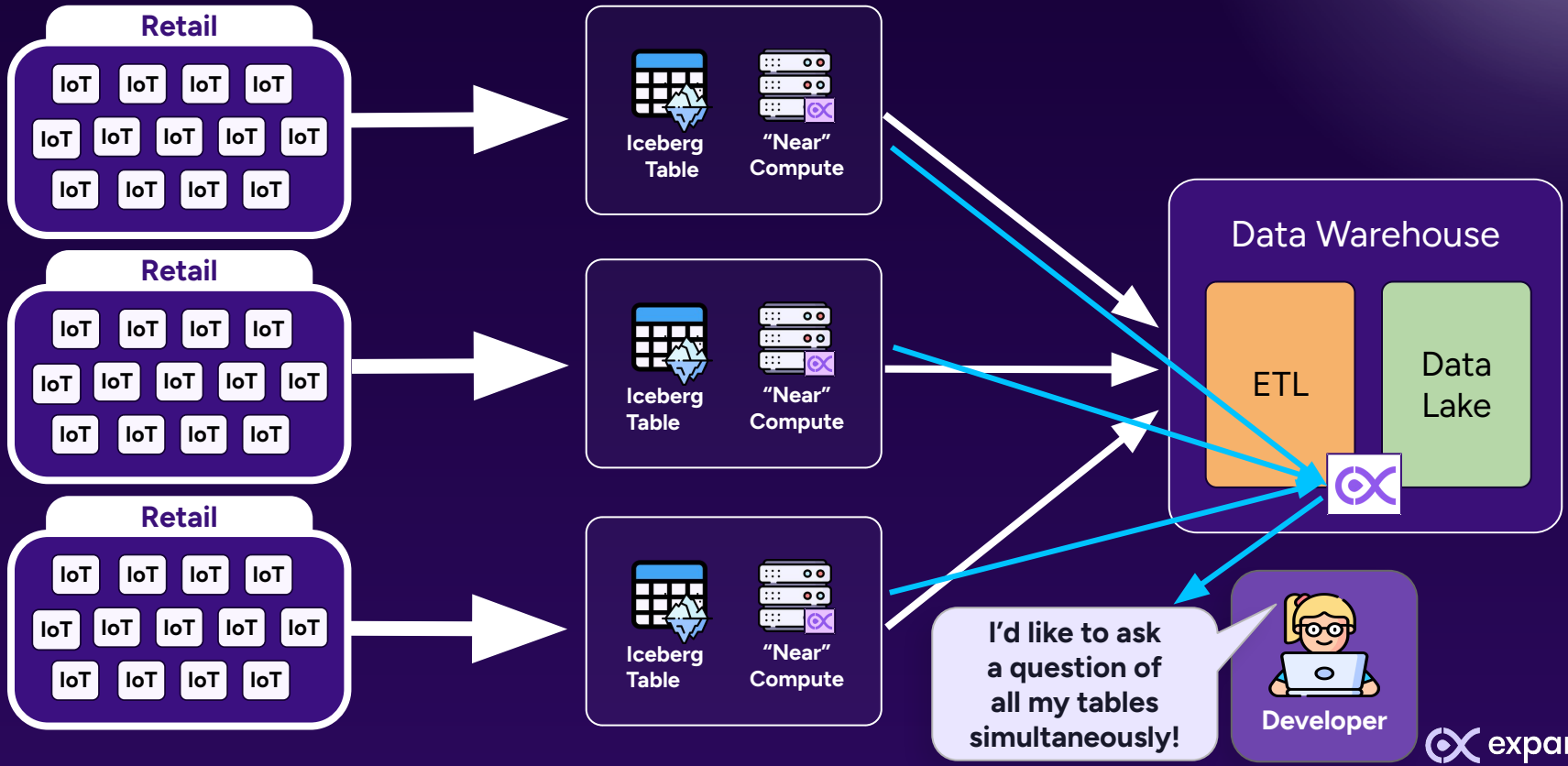
After Expanso

expanso

Use Case: The Distributed Data Warehouse



Use Case: The Distributed Data Warehouse



Use Case:

Streamlined Data Pipelines

expanso

Use Case: Streamlined Data Pipelines

Situation:

- Servers produces +1GB/day - even small deployments quickly overwhelm naive processing
- Mid-to-large enterprises rely on traditional ETL pipelines
- Clusters for processing data require centralized deployments

Challenges:

- Networks will always delay the flow of information due to geography
- Centralized clusters can be over-provisioned, slow to spin up, or both

Centralized Log Processing

RAW LOG DATA

databricks

splunk>

DATADOG

And all the others.

Massive amounts of raw log data are transported to a centralized location for processing, resulting in significant **transfer and storage costs and delays.**

expanso

Distributed Log Processing

databricks

splunk>

DATADOG

And all the others.

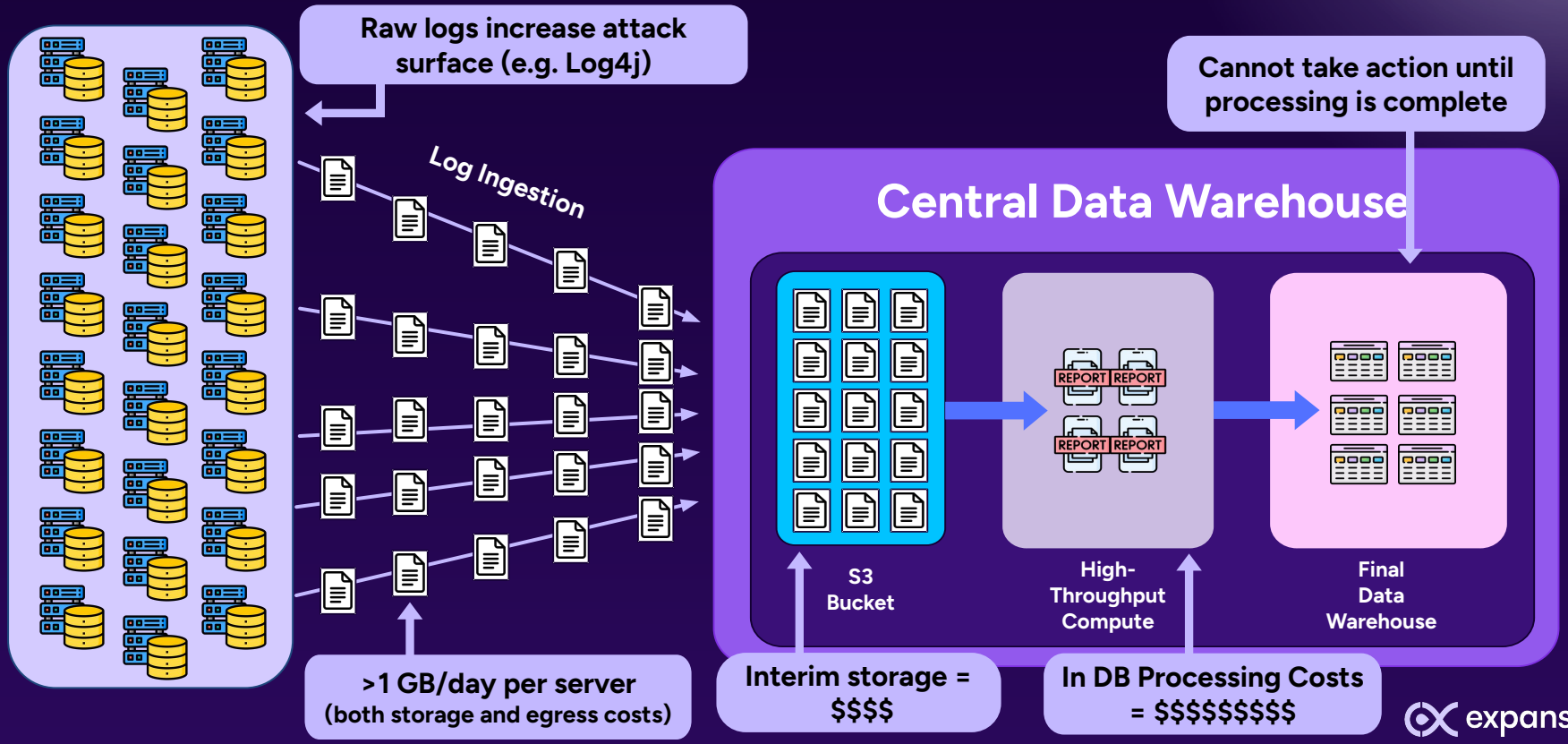
Bacalhau operates on each machine, creating a network that connects all data points. This setup enables users to **run local filtering, analysis, and compression tools on their data.**

expanso

Before Expanso

expanso

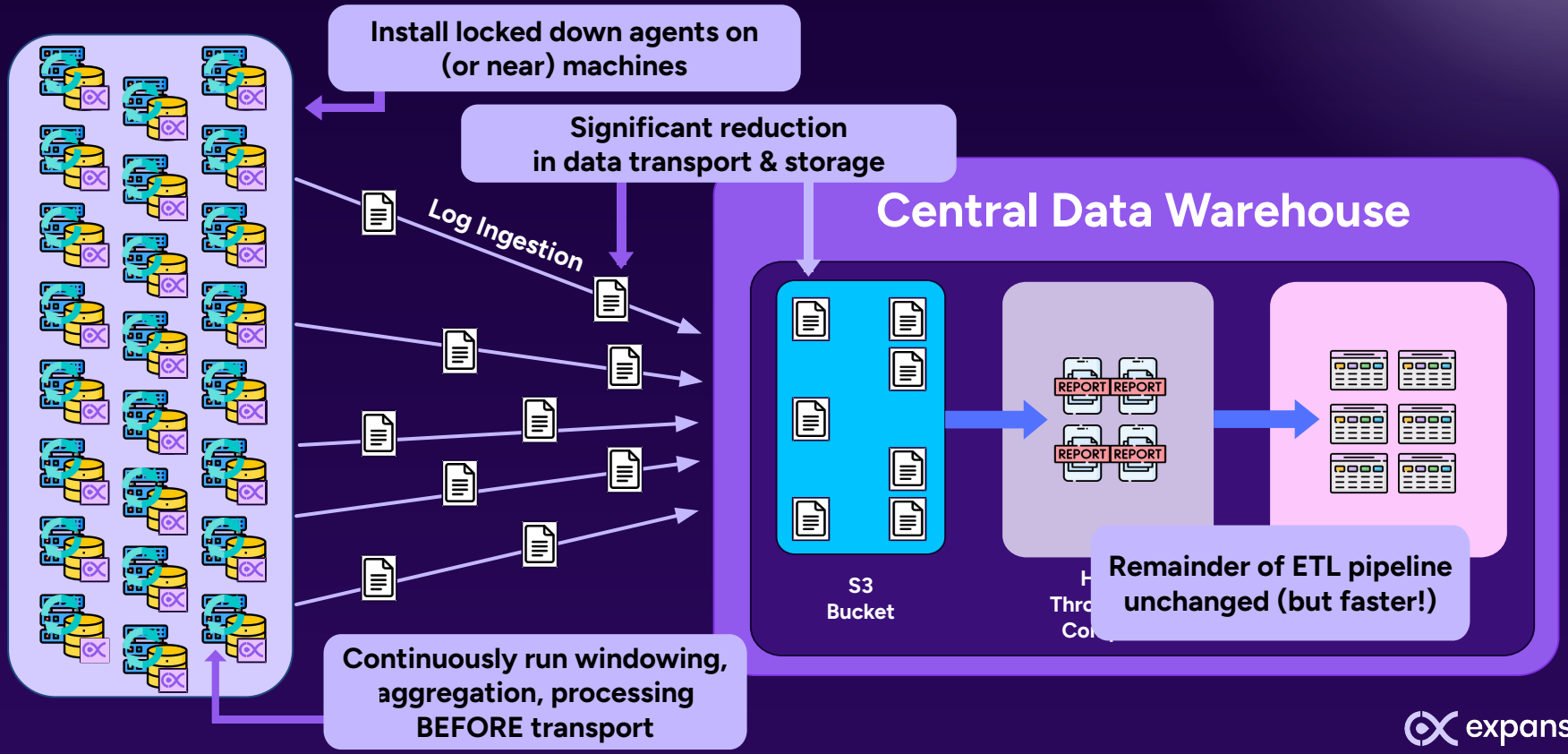
Use Case: Streamlined Data Pipelines



With Expanso

expanso

Use Case: Streamlined Data Pipelines



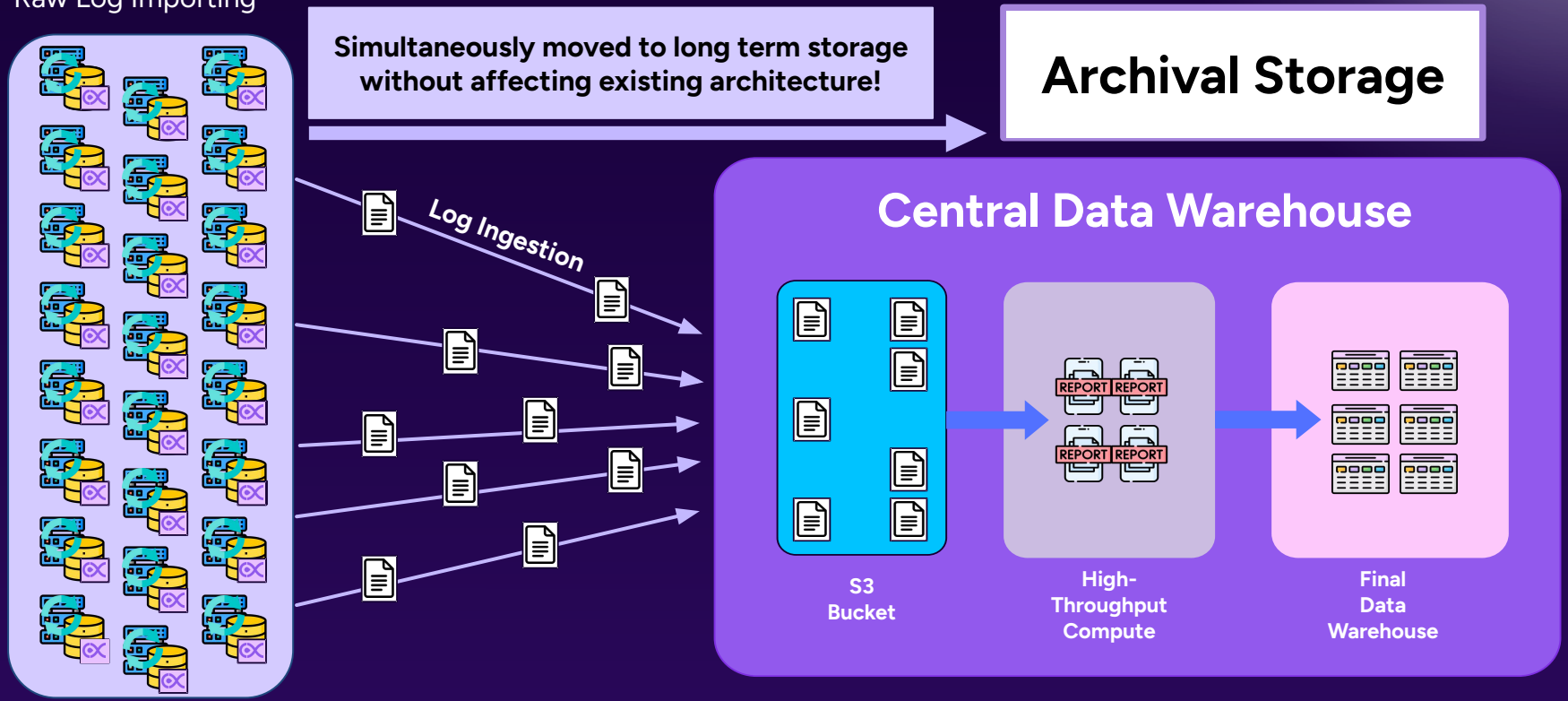
expanso

And...

expanso

Use Case: Optimized Data Pipelines

Raw Log Importing



expanso

Streamlined **Data Pipelines**

Local (Pre-) Processing

Smart and efficient

- Reliable and auditable
- Processing at the log source
- Optimized scheduling; data and network awareness

Immediate Actions

Run tools in-situ

- Filtering
- Analytics
- Compression
- Alerting

Enhances Your Setup

Reuse existing tools

- Reuse the tools you love
- Multi-region
- Fault tolerant

expanso

Benchmarked Cost Savings

Orchestrate global log processing to reduce log processing costs and accelerate insights*

Name	Log Collector	expanso
Storage Annual Cost	-	$1,400
Compute Annual Cost	-	$4,000
Pub/Sub Annual Cost	-	$500
Log Collector Annual Cost	$2,500,000	$12,000
Subtotal	**$2,500,000**	**$18,000***

*Benchmarking details available here. Expanso pricing not included.

98%+ Cost Savings*

expanso

Use Cases:

Securing Customer Data Made Easy

expanso

Use Case: SOCs Made Easy

Situation:

- Organization has multiple locations, with customers in each location
- Want to perform analysis over their customers (e.g. do risk analysis vs. leverage)
- Requires merging many sources of data in order to assess

Challenges:

- Data needs to be updated regularly (e.g. market close)
- Moving data across country/region boundaries triggers compliance issues
- Once the data is moved, need clean audit logs to show regulators

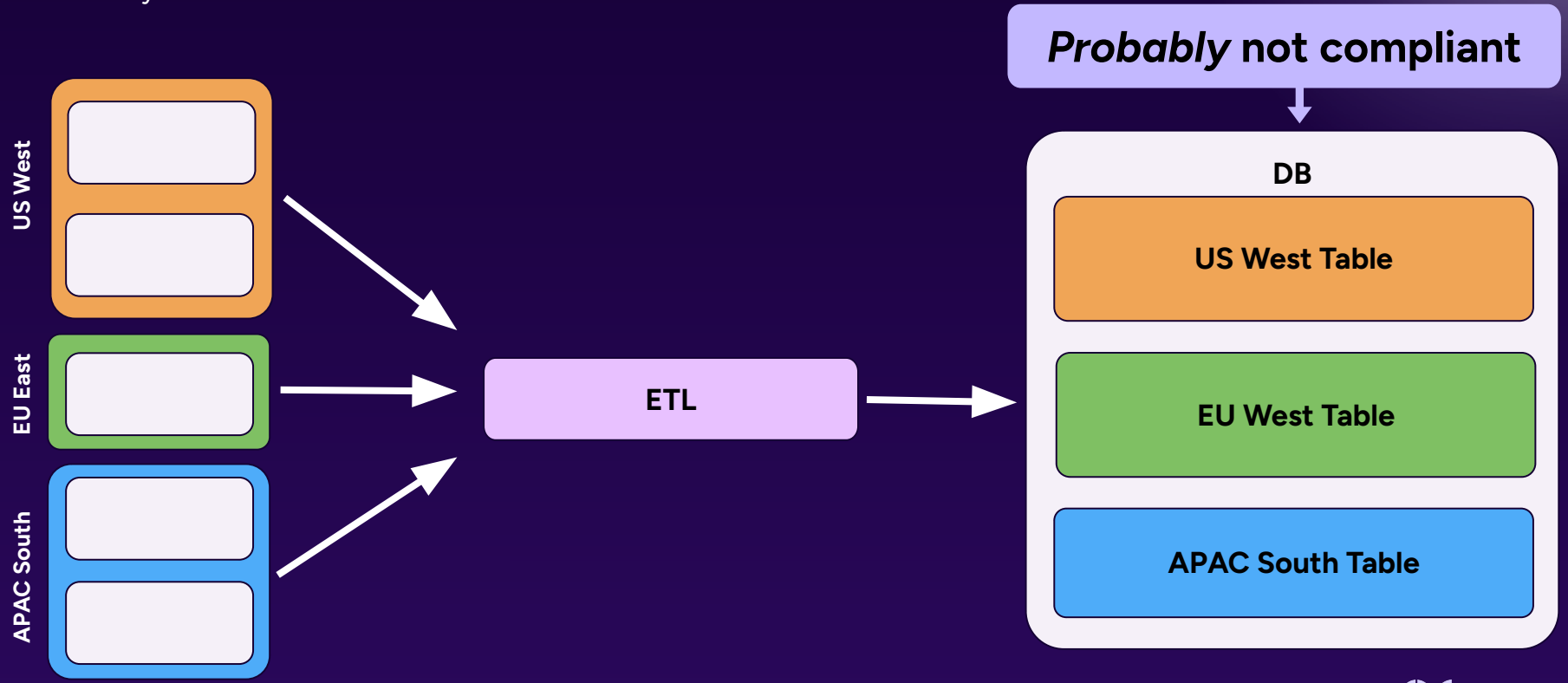
expanso

Before Expanso

expanso

Use Case: SOC Logs Meet Modern Data Stack

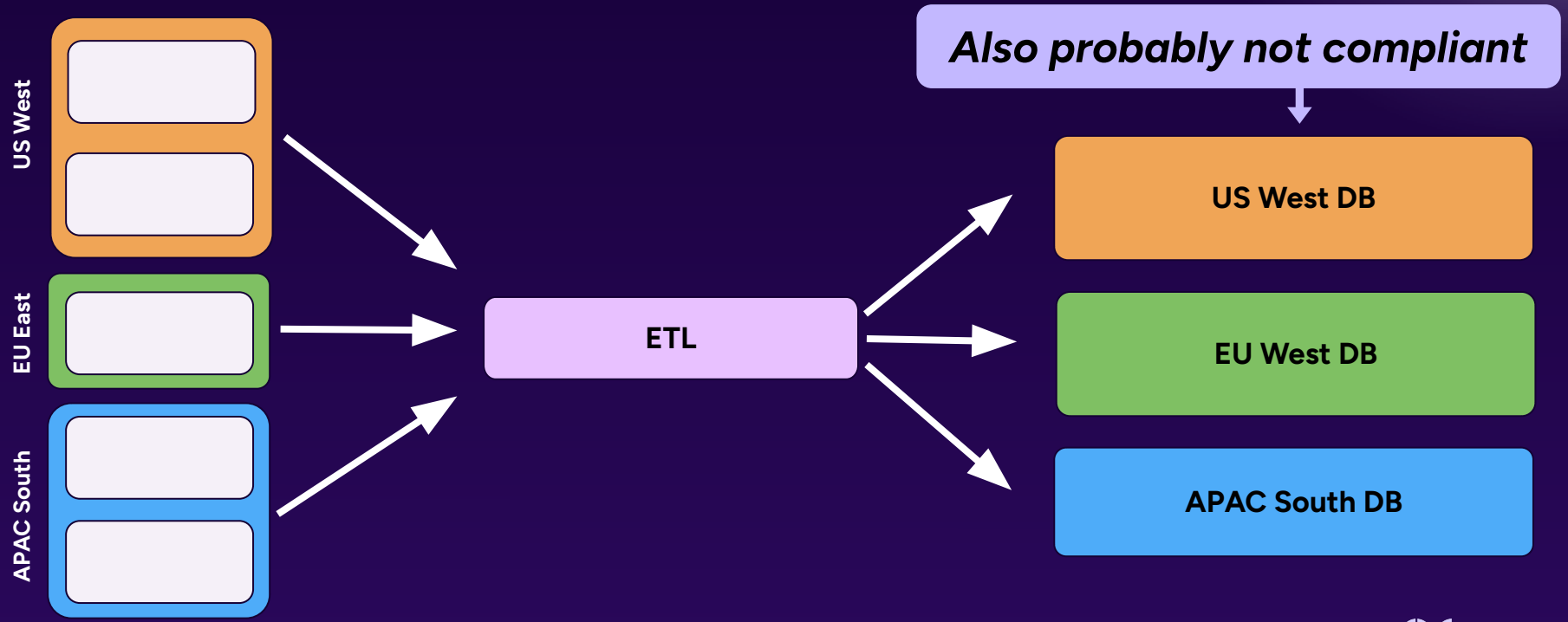
Security Without the Overhead



expanso

Use Case: SOC Logs Meet Modern Data Stack

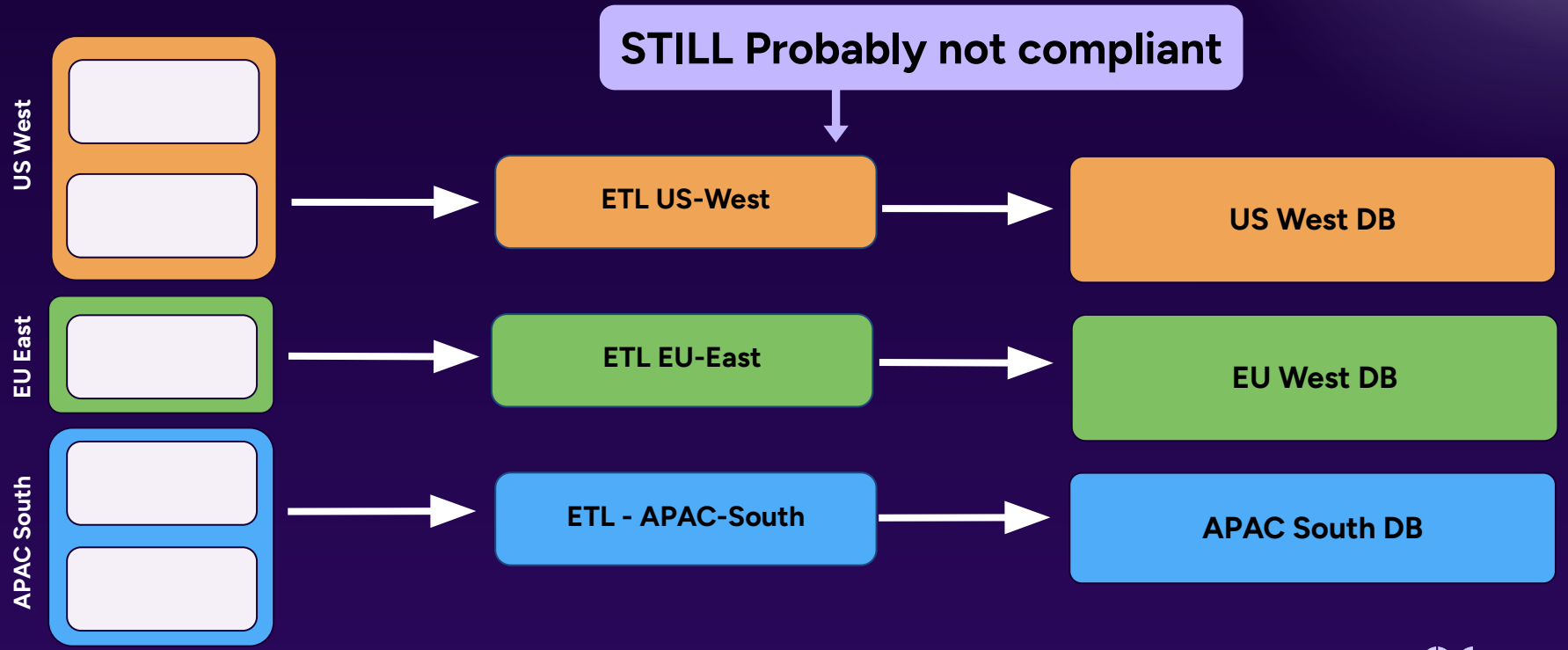
Security Without the Overhead



expanso

Use Case: SOC Logs Meet Modern Data Stack

Security Without the Overhead



expanso

???

expanso

Use Case: SOCs Made Easy

- “But Isn’t There a Playbook?”
- SOC (and other regulatory compliance) issues:
  - Data Deluge: Volume of raw data to transfer and secure
  - Privacy Pitfalls: Risk of exposing sensitive customer information
  - Compliance Chaos: Difficulty in tracking data across borders
  - Audit Nightmares: Complex process to prove data integrity
  - Resource Drain: Higher costs for storage, transfer, and security
  - Performance Lag: Slower due to processing unoptimized data
  - Regulatory Roulette: Greater chance of non-compliance and potential fines when data is merged together
  - Insight Delays: Longer time-to-value for critical business intelligence

expanso

Use Case: SOCs Made Easy

- "But Isn't There a Playbook?"
- SOC (and other regulatory compliance) issues:
  - Data Deluge: Volume of raw data to transfer and secure
  - Privacy Pitfalls: Risk of exposing sensitive customer information
  - Compliance Chaos: Difficulty in tracking data across borders
  - Audit Nightmares: Complex process to prove data integrity
  - Resource Drain: Higher costs for storage, transfer, and security
  - Performance Lag: Slower due to processing unoptimized data
  - Regulatory Roulette: Greater chance of non-compliance and potential fines when data is merged together
  - Insight Delays: Longer time-to-value for critical business intelligence

expanso

Use Case: SOCs Made Easy

- "But Isn't There a Playbook?"
- SOC (and other regulatory compliance) issues:
  - Data Deluge: Volume of raw data to transfer and secure
  - Privacy Pitfalls: Risk of exposing sensitive customer information
  - Compliance Chaos: Difficulty in tracking data across borders
  - Audit Nightmares: Complex process to prove data integrity
  - Resource Drain: Higher costs for storage, transfer, and security
  - Performance Lag: Slower due to processing unoptimized data
  - Regulatory Roulette: Greater chance of non-compliance and potential fines when data is merged together
  - Insight Delays: Longer time-to-value for critical business intelligence

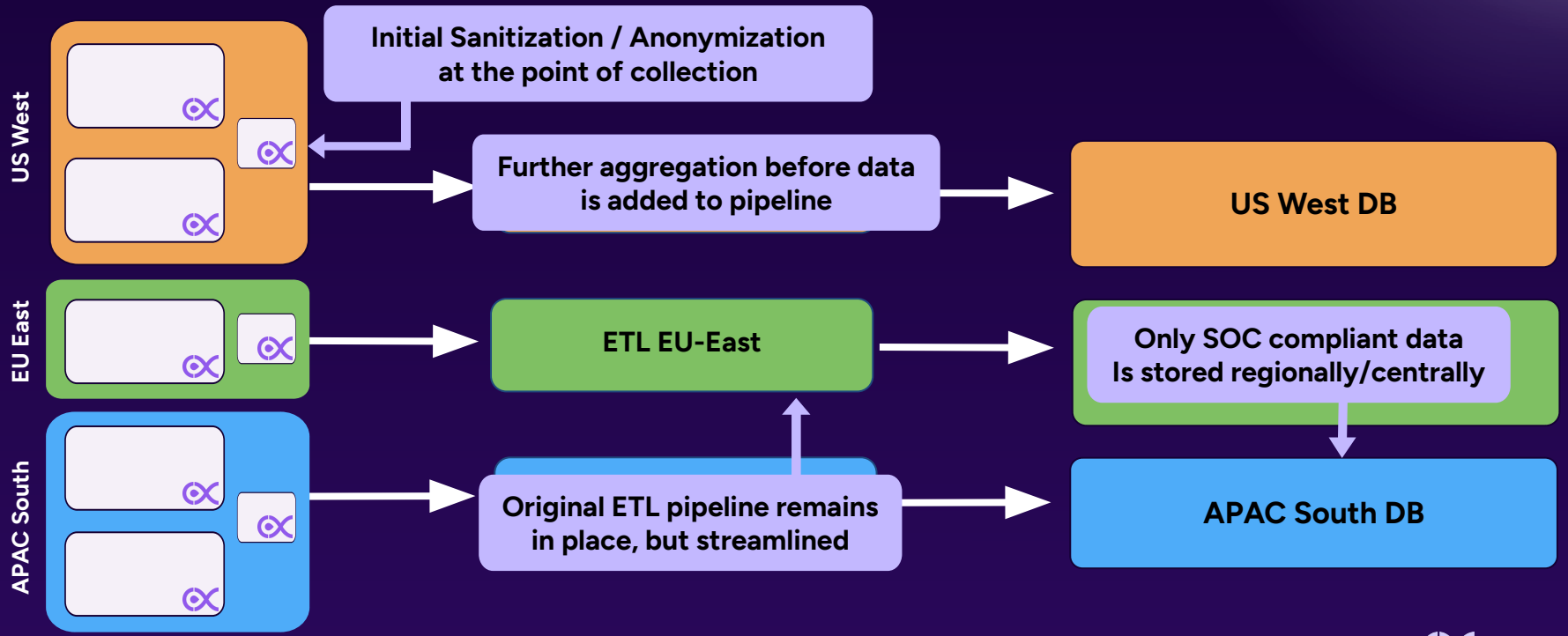
expanso

With Expanso

expanso

Use Case: SOC Logs Meet Modern Data Stack

Security Without the Overhead



expanso

Use Case: SOCs Made Easy

Situation:

- Organization has multiple locations, with customers in each location
- Want to perform analysis over their customers (e.g. do risk analysis vs. leverage)
- Requires merging many sources of data in order to assess

Challenges:

- Data needs to be updated regularly (e.g. market close)
- Moving data across country/region boundaries triggers compliance issues
- Once the data is moved, need clean audit logs to show regulators

Expanso Benefit:

- More efficient pipeline
- Cost and time savings
- Build in isolation

AND AUDIT LOGGING FOR FREE!

expanso

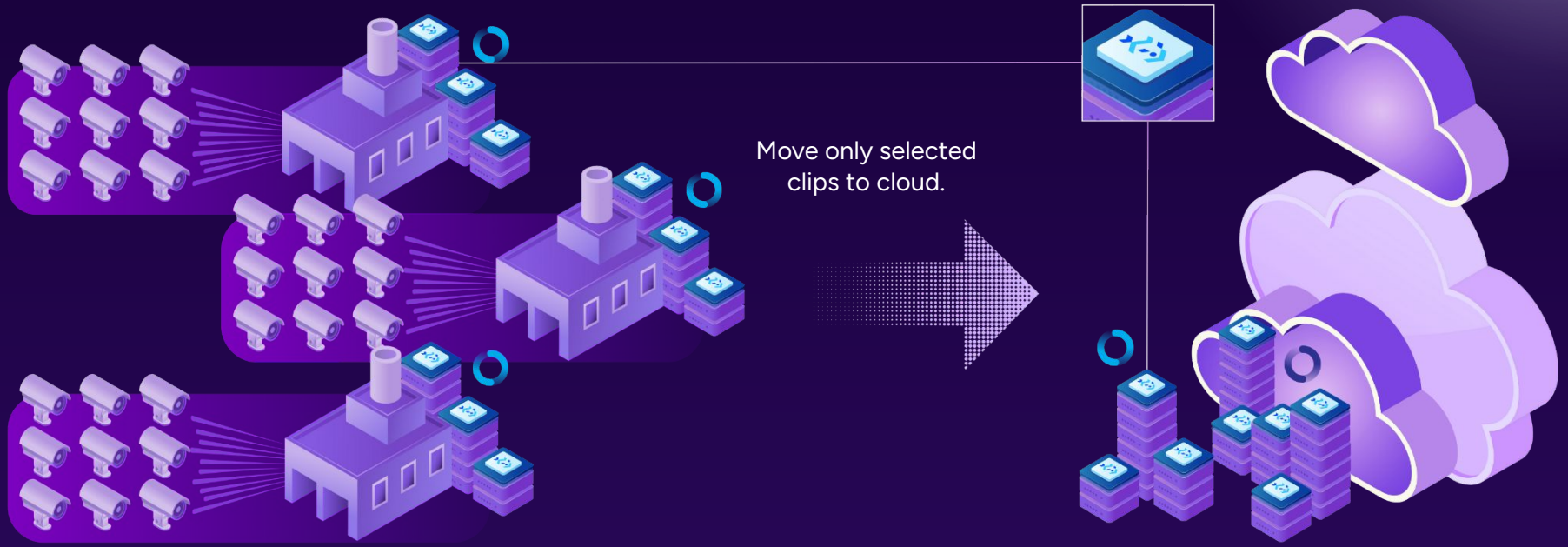
Use Case:

Scale Machine Learning Deployments

expanso

Scale Machine Learning Deployments

Multi-region inference model deploy and execution.



Each factory has many cameras; collecting video from each location.

Now the cloud deployment is more efficient and more cost effective

expanso

Scale Machine Learning Deployments

Device + Edge Inference

Efficient Processing

- Faster and lower latency
- Reduce storage and transit costs
- Managed queuing, poor network and job failover

Global Model Orchestration

GitOps for Models

- Simplify network wide ML model deploys
- Coordinate jobs across fleet for cost and availability
- Supports execution and training workloads

Traceable Execution Log

Permanent Audit Trail

- All model and artifact deployments are tracked
- Inference and data sets traced end-to-end
- Debuggability and security built-in to the system

expanso

Scale Machine Learning Deployments

Data Bottlenecks	High Costs	Difficult Regulations
On-Device AI	Cost & Energy Savings	Data Governance
Train and execute **ML at the edge**. Near **real-time** insights.	Save up to **90%** compared to central processing. Reduced bandwidth ~**93%**.	Real-time **controls** and **alerts**. Enhanced **pipeline control** and **governance**.

expanso

Cost Savings for Machine Learning

Move ML processing to the edge to reduce processing costs and accelerate insights

Name	Hyperscale Cloud	Expanso
Storage Annual Cost	$210,500	**$21,050**
Compute Annual Cost	-	**$8,000**
Network Annual Cost	$750	**$500**
AI Services Annual Cost	$47,200	**$4,720**
Subtotal	**$258,450**	**$34,270**

Assumes 100 Cameras / Location * 10 Locations

87% Cost Savings*

* Does not include Expanso pricing

expanso

Summary

expanso

Customer Benefits

- **Simplified Data Management:** Eliminate the need for a centralized data lake, reduce data movement, and simplify data governance.
- **Improved Performance & Scalability:** Analyze data in near real-time, scale compute resources on demand, and handle growing data volumes.
- **Reduced Costs:** Lower infrastructure costs, optimize data storage, and reduce data transfer expenses.
- **Increased Agility & Innovation:** Faster time to insights, enable new use cases, and foster data-driven decision-making.

Get to Know Us

bacalhau-project

www.expanso.io

@Bacalhauproject

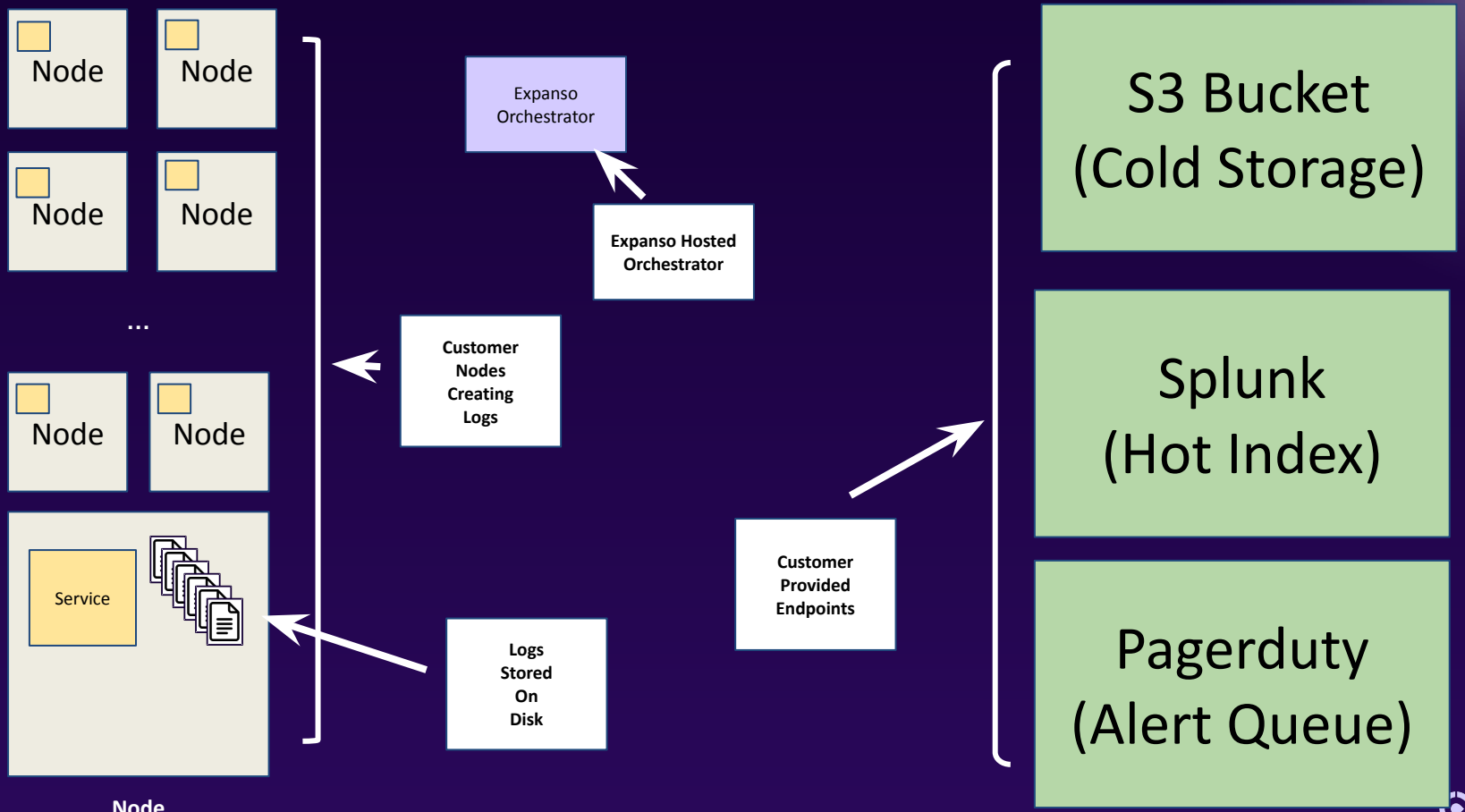
@BacalhauProject

@ExpansoIO

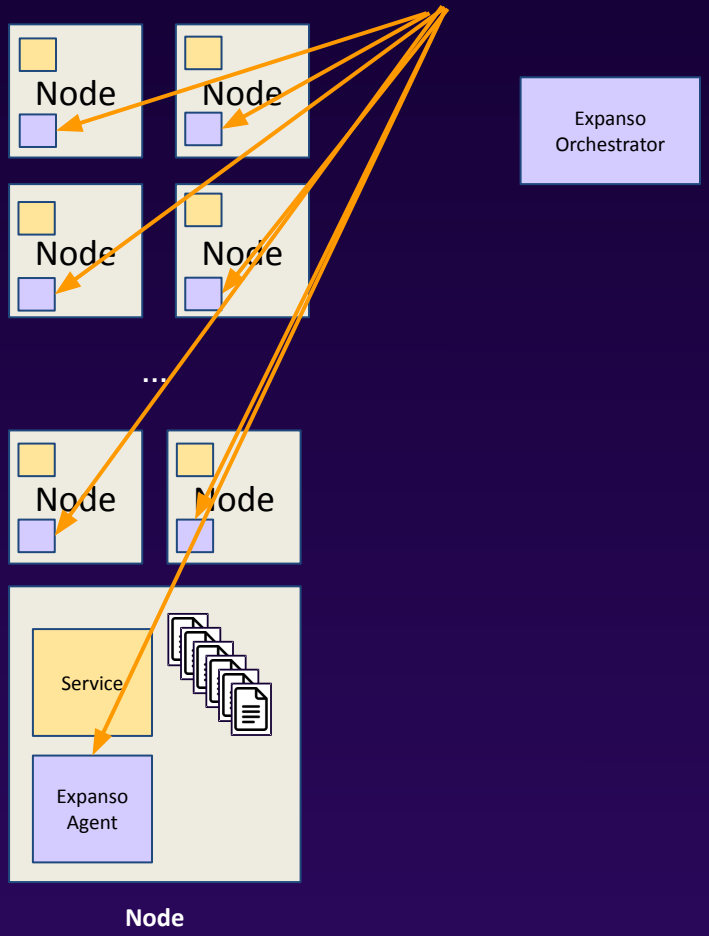
@BacalhauProject

@company/expanso-io

expanso


Node
Node
Node
Node
...
Node
Node
Service
Node
Expanso Orchestrator
Expanso Hosted Orchestrator
Customer Nodes Creating Logs
Logs Stored On Disk
Customer Provided Endpoints
S3 Bucket (Cold Storage)
Splunk (Hot Index)
Pagerduty (Alert Queue)

Step 1: Customer Puts Expanso Agent on Nodes (One Time Setup)



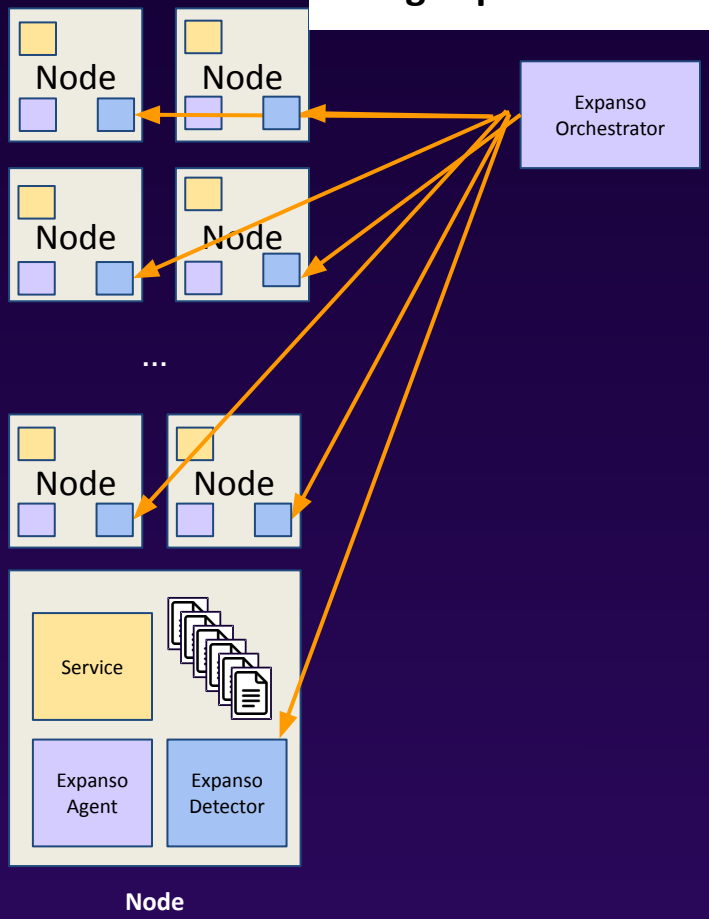
S3 Bucket
(Cold Storage)

Splunk
(Hot Index)

Pagerduty
(Alert Queue)

expanso

Step 2: Customer Places Expanso Detector on Selected Nodes Using Expanso Orchestrator



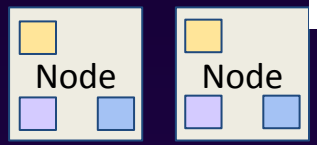
S3 Bucket
(Cold Storage)

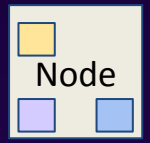
Splunk
(Hot Index)

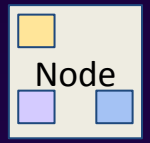
Pagerduty
(Alert Queue)

expanso

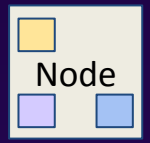
Step 3: Customer Configures Expanso Detector to Observe and Route Logs To Correct Location via Expanso Orchestrator

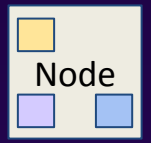


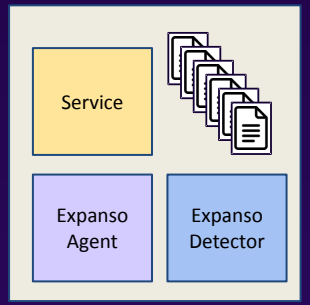




...







Node

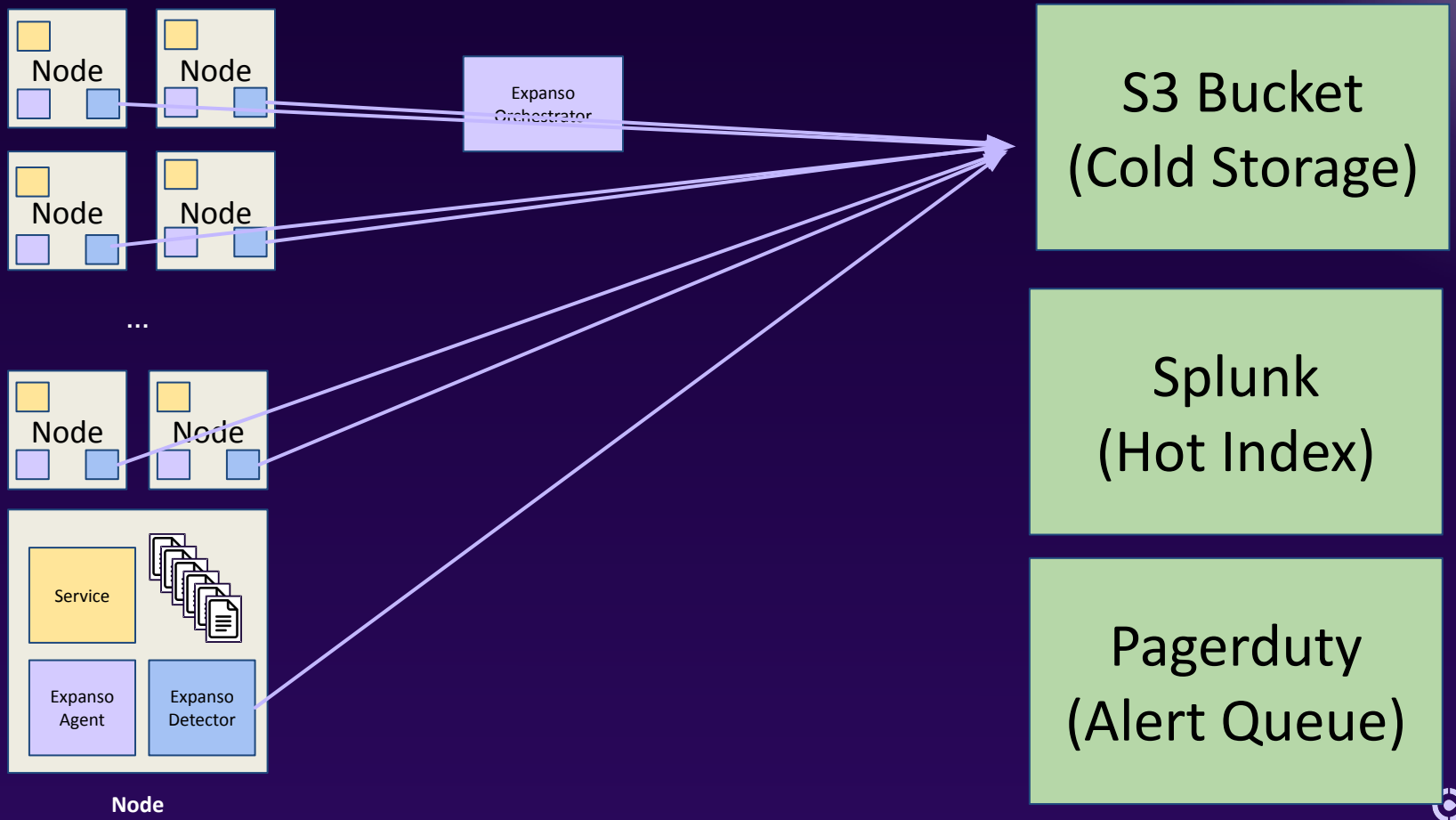
S3 Bucket (Cold Storage)

Splunk (Hot Index)

Pagerduty (Alert Queue)

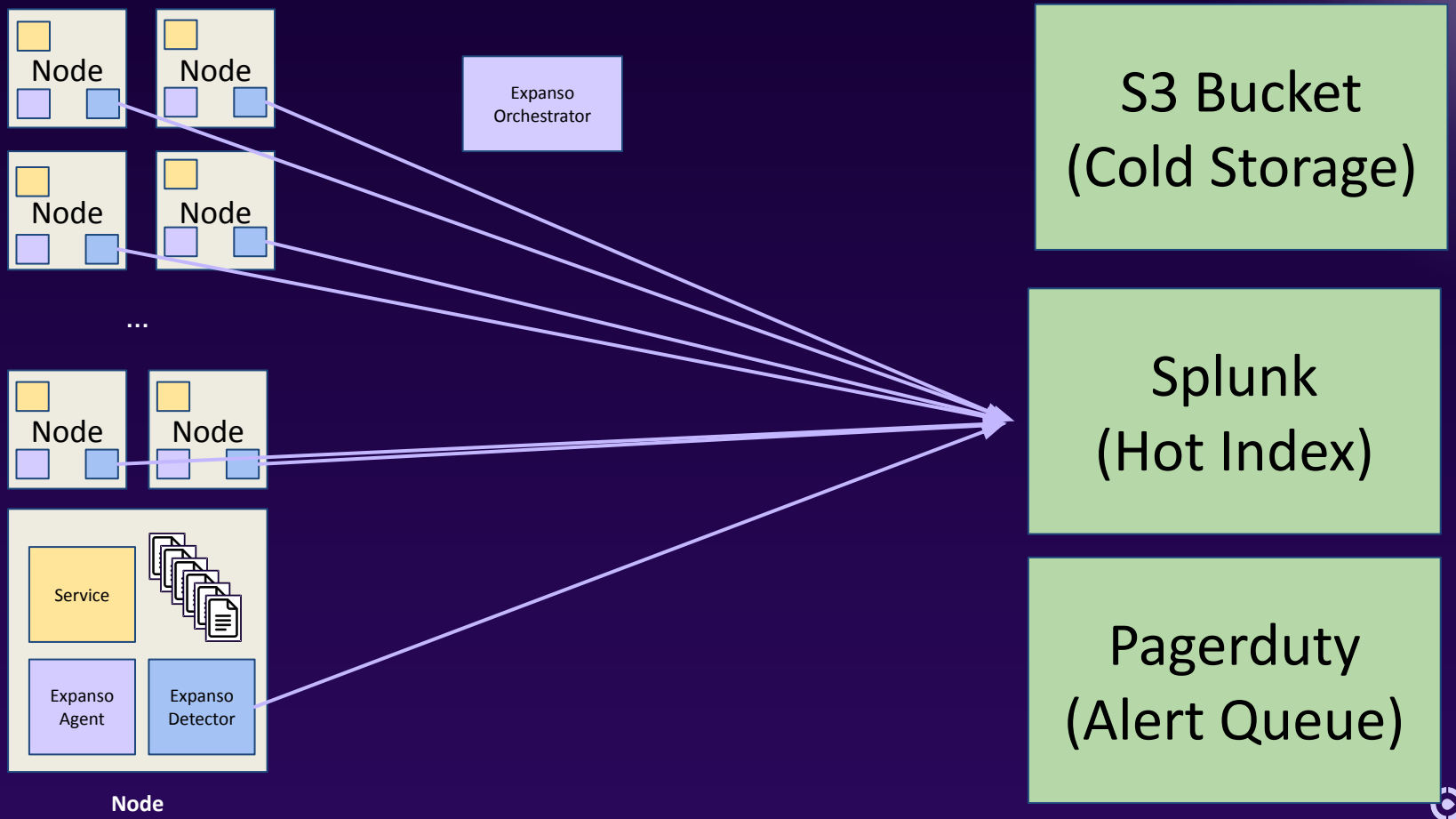
expanso

Step 4a: Detector Collects Raw Logs and Pushes to S3 Bucket

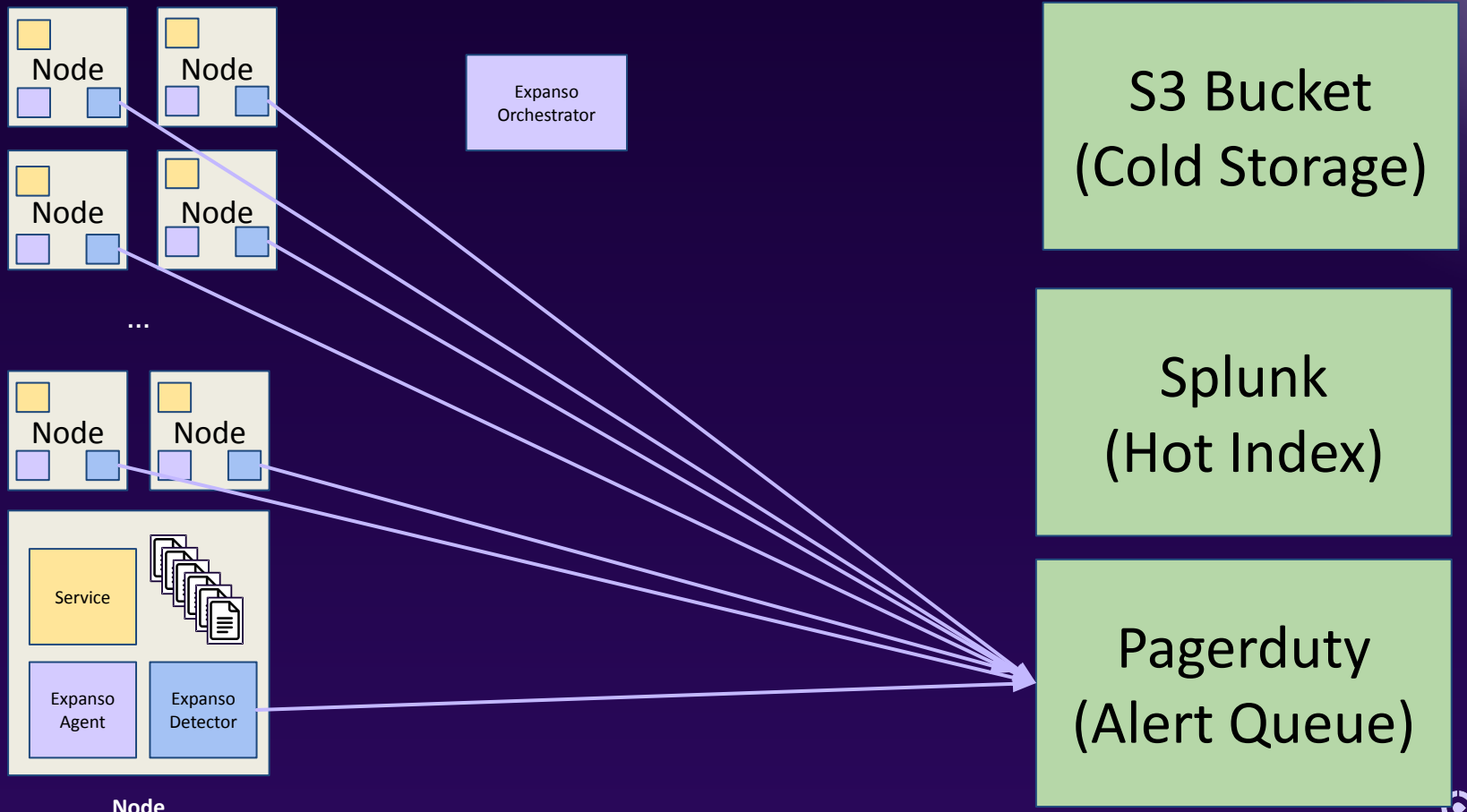


expanso

Step 4b: Detector Aggregates Log Data and Pushes to Splunk



Step 4c: Detector Finds Alert Situations and Sends to Queue



expanso