
AWS
re:Invent
NOV. 28 – DEC. 2, 2022 | LAS VEGAS, NV

CMP331-R

High-throughput computing for financial risk management and modeling

Mark Norton

Principal EC2 & HPC Specialist
AWS

Kirill Bogdanov

Senior Solutions Architect
AWS

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

What are the challenges, and how are financial services customers addressing them?

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

What is driving this transformation?



Customer/
market signals

Regulatory requirements – FRTB/RWA/ESG

Increasing market volatility

Desire for faster intra-day results and on-demand analytics

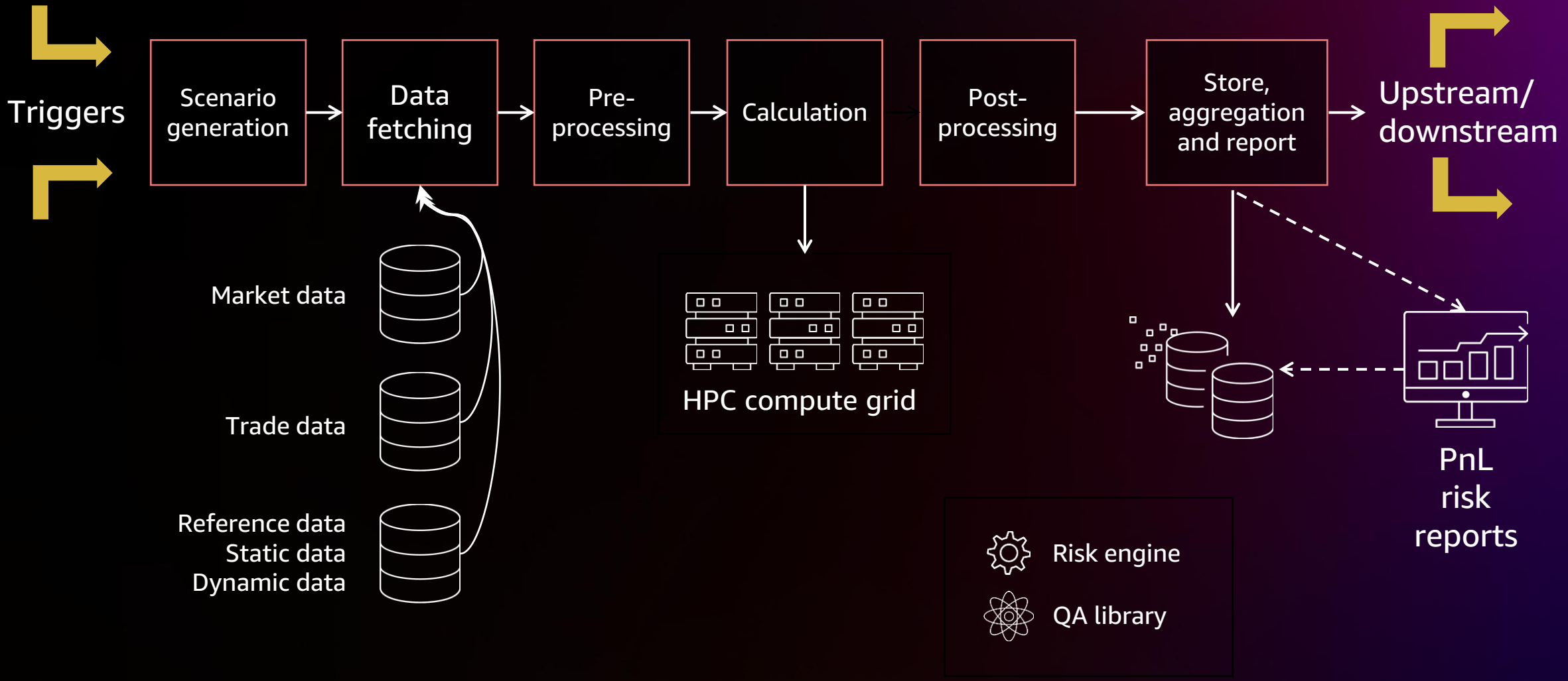
Skill set shortage

Siloed on-premises centralized grid resource

Cost of hosting/replacing aging physical compute

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Risk pipeline



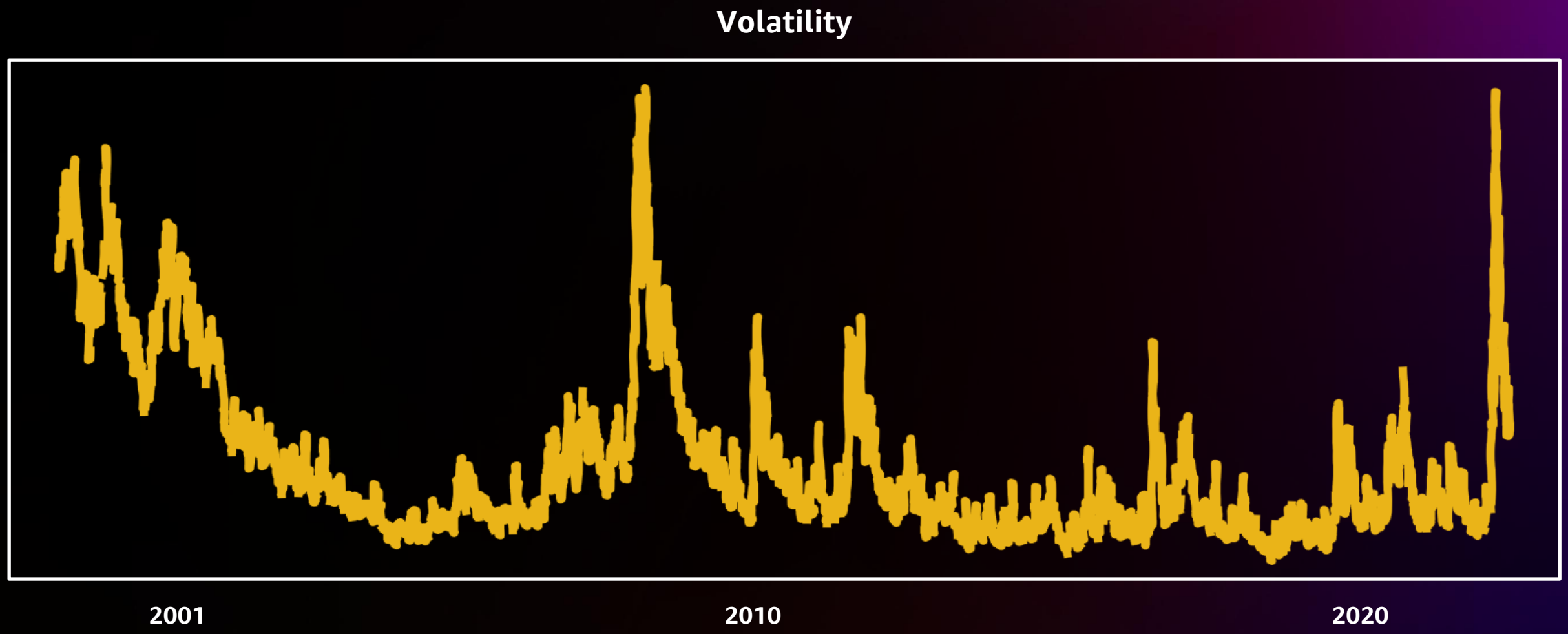
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

A typical day for an HPC cluster at an investment bank



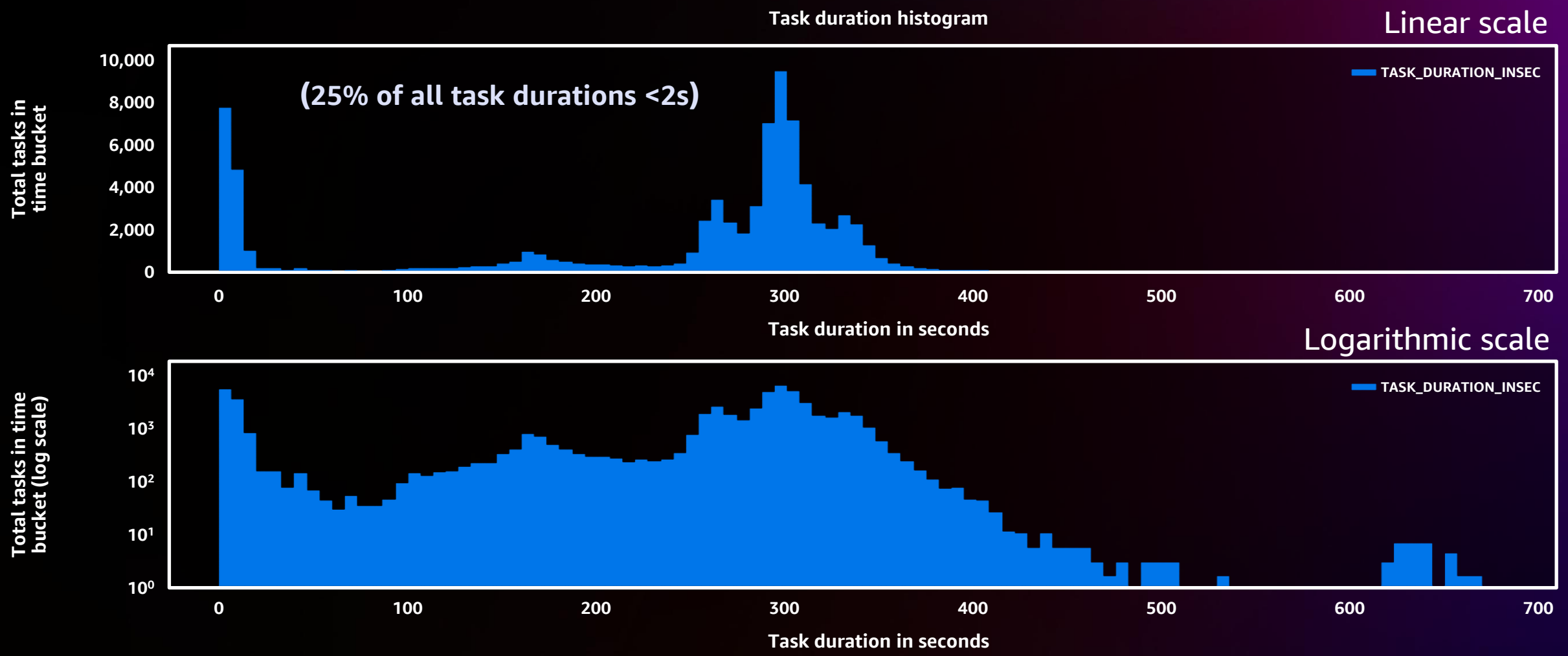
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Market volatility over 2 decades



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Task completion time histogram



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

How our customers are approaching grid on AWS

Flexible configurations and virtually unlimited scalability to grow and shrink your infrastructure as your financial risk management and financial modeling workloads dictate



Lift and shift

- Move your existing high-performance computing (HPC) platform, algorithms, and execution environments



Replatform and re-engineer

- Migrate to cloud-native high-throughput computing (HTC)
- Modernize your risk algorithms and replatform to HTC

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Replatform to HTC (high-throughput computing project)

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

European bank accelerates capital markets risk processing on AWS



"We have reduced the execution time from 5 hours to 1 hour; additionally, we have operation metrics that we didn't have before, that can now correlate to our business metrics to drive the optimization of the environment."

Grid Developer, Major European Bank

To prepare for new FSI regulations and in response to market conditions, a major European bank needed to accelerate risk processing capabilities while reducing associated operational costs

Seeing the potential of replatforming Murex from an on-premises TIBCO GridServer environment to cloud-native HTC with AWS, the bank engaged the AWS Professional Services and prototyping teams

Replatforming to HTC (a 6-month project) enabled the bank to cut risk batch times from 5 hours to 1 hour while realizing significant financial savings, including removing third-party grid licenses and associated operational overhead

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Introducing HTC-Grid



HTC-Grid

A cloud-native “blueprint” for high-scale-out, high-throughput compute

Benefits

- Algorithms are unchanged
- Cost savings (removal of third-party grid license costs and grid infrastructure management/engineering costs)
- Flexible pricing models (Spot Instances, On-Demand, workload)
- Access to a diverse range of modern CPUs – Intel, AMD, and Graviton

Open source AWS Labs project: https://github.com/awslabs/aws-htc-grid

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

HTC-Grid design tenets

Custom and on-demand

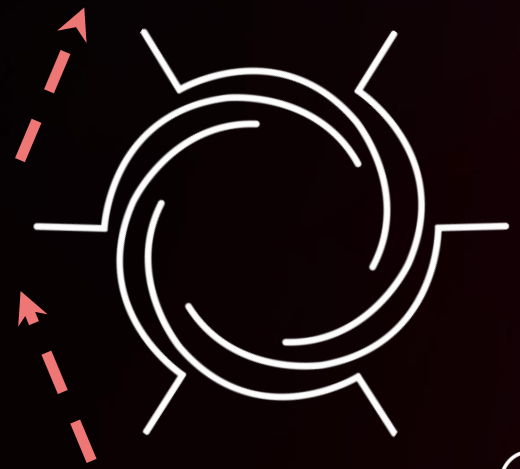
Rapidly create workload-optimized/dedicated HTC-Grid services

Cloud-native

Use mature, robust, and performant AWS services already familiar to AWS customers

Modular

Interchangeable data and compute subsystems via infrastructure as code



Scale and low latency

Rapid scale-out compute, a low infrastructure latency control plane, and optional high data IOPS/throughput

All compute is an AWS Lambda function

Tasks APIs are Lambda functions – allowing interchangeability of compute plane implementations

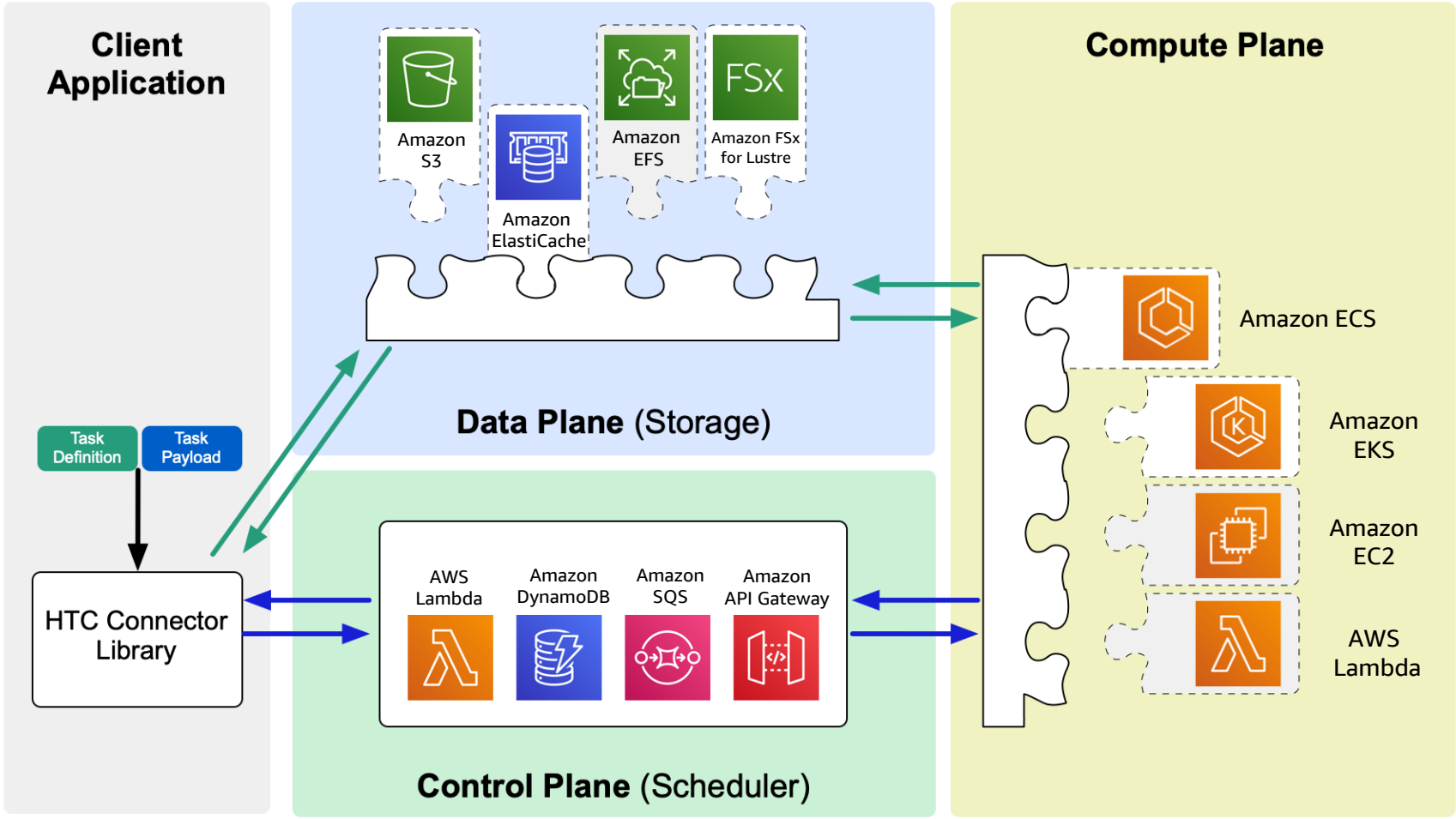
Simplify replatforming

Provide a client API that is designed and familiar to our FSI users

Open source AWS Labs project: https://github.com/awslabs/aws-htc-grid

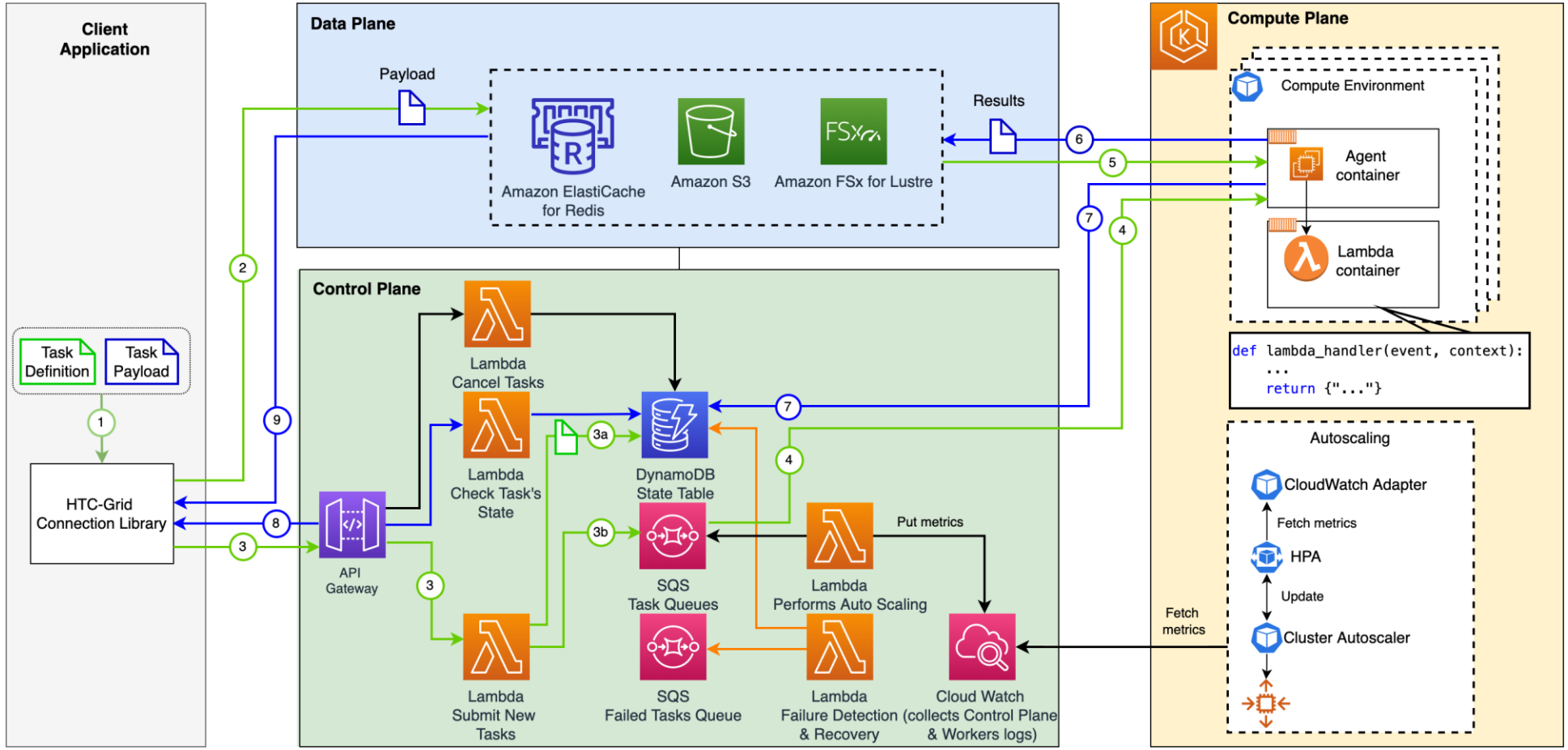
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Architecture overview



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

HTC – Amazon EKS compute plane variant



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Current customer measured performance

AMAZON ECS AND AMAZON EKS FOR COMPUTE PLANES; ELASTICACHE FOR REDIS AS DATA PLANE

Scalability/elasticity

(ECS) Tested up to 80,000 workers running on 3,000+ EC2 instances; scaling time 15 minutes

(EKS) Tested up to 20,000 pods running on 700+ nodes; scaling time 20 minutes

Throughput for high volume of short running tasks

10,000 tasks per second using ~3,000 pods, which is ~36 million tasks per hour or ~300 million tasks per night batch (tested with zero job tasks, 0 size payload)

Infrastructure latency

~0.3 seconds per task (excluding task execution time)

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Resources



FSI HPC whitepaper
https://d1.awsstatic.com/whitepapers/aws-financial-services-grid-computing.pdf

FSI blog
https://aws.amazon.com/blogs/industries/how-cloud-increases-flexibility-of-trading-risk-infrastructure-for-frtb-compliance/

HTC-Grid blog
https://aws.amazon.com/blogs/hpc/cloud-native-high-throughput-computing-with-aws-htc-grid/

AWS Labs HTC-Grid project and documentation
https://github.com/awslabs/aws-htc-grid

HTC-Grid hands-on workshop
https://main.d5fll76yf0v34.amplifyapp.com/

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Thank you!

Mark Norton
marknort@amazon.com

Kirill Bogdanov
kirillb@amazon.com



Please complete the session survey in the **mobile app**

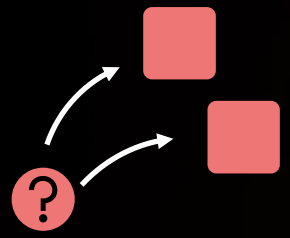
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Appendices

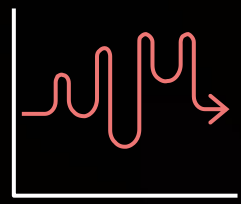
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon EC2 purchase options

On-Demand Instances

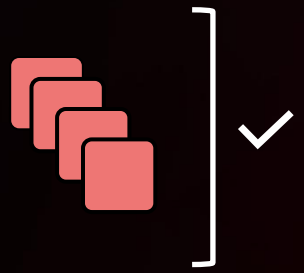


Pay for compute capacity by **the second** with no long-term commitments



Spiky **stateful** workloads, to define needs

Reserved Instances/ Savings Plans

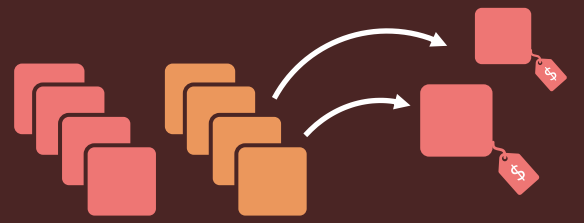


Make a 1- or 3-year commitment and receive a **significant discount** off On-Demand prices



Committed and steady-state usage

Spot Instances



Spare Amazon EC2 capacity at **savings of up to 90%** off On-Demand prices

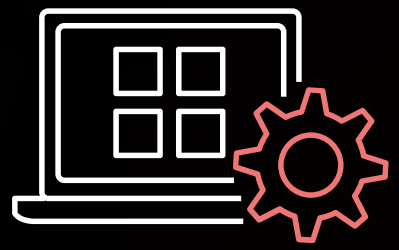


Fault-tolerant, flexible, **stateless** workloads

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Re-engineering is the final destination



Re-engineer

Algorithm modernization
+ replatforming

Benefits

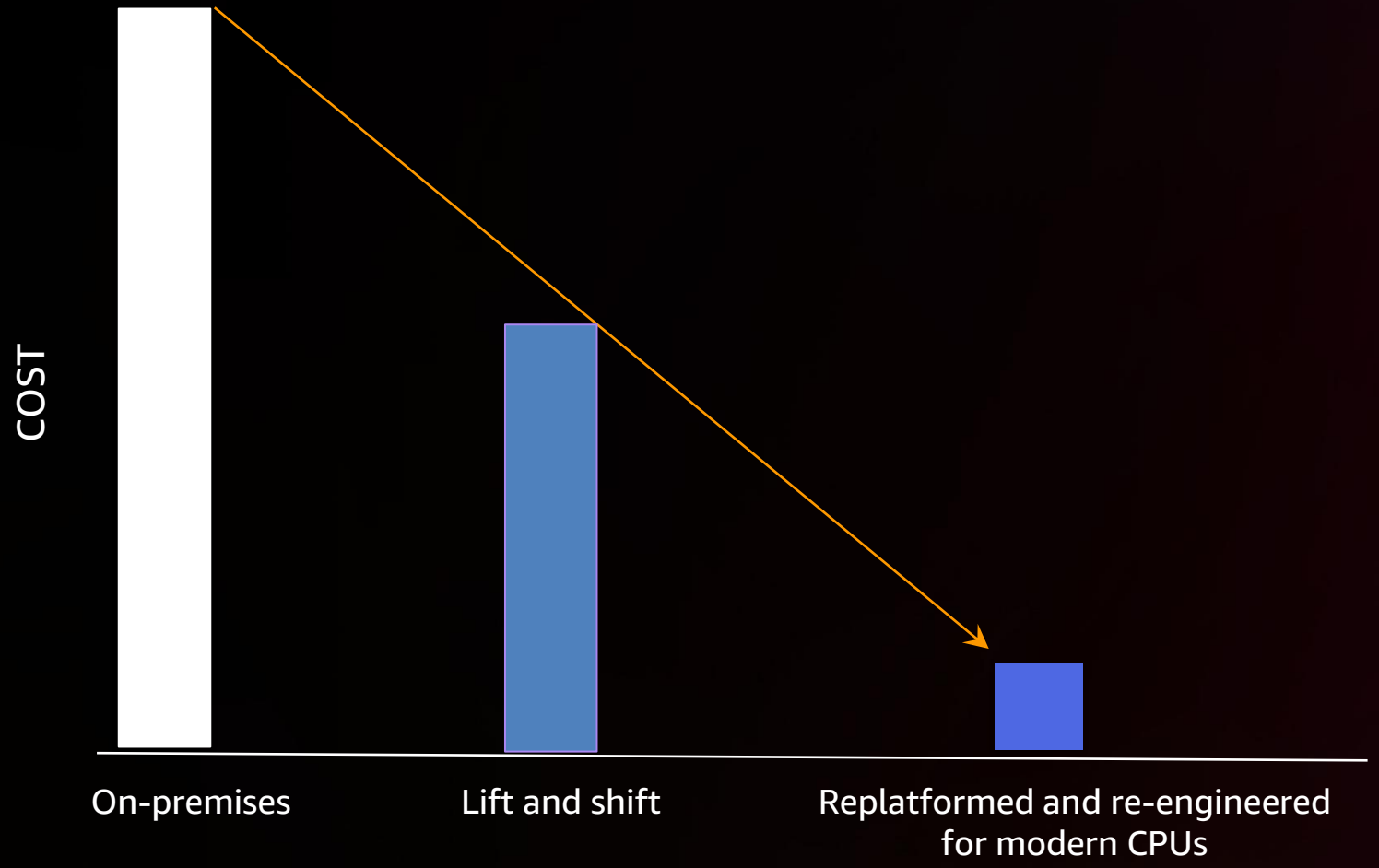
- Further removes technical debt and reduces business/regulatory risk
- Cost savings by removing dependency on legacy OS licenses
- Ability to optimize using a diverse set of processor architectures
- Ability to leverage modern algorithmic techniques and different architectures, and explore hybrid HPC
- Plus all of the benefits of replatforming

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Re-engineering – Algorithm modernization

Re-engineering of risk algorithms may be pursued alongside or as a follow-on activity to replatforming



Benefits:

- **Leverage modern CPU designs**
  Migrate to AWS Graviton II/III, FPGA, GPU, Quantum
- **Remove legacy software infrastructure**
  Removal of .COM, .DCOM, .NET Windows taxes
- **Overnight to intra-day/real-time risk reporting**
- **Reduce CO2 footprint**
  Environmental, social, and governance (ESG)

aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.