AWS re:Invent

NOV. 29 – DEC. 3, 2021 | LAS VEGAS, NV

Goldman Sachs

FSI204

Goldman Sachs: Reinventing financial data analytics

Andrew Phillips
Managing Director
Goldman Sachs

Francis Giannaros
Vice President
Goldman Sachs

Balaji Gopalan
Principal Solutions Architect
AWS

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Agenda

1 Intro

Introduction to Goldman Sachs and a typical quantitative analytics problem

2 Technical deep dive

How we re-architected our data platform on AWS

- Real-time data infrastructure
- Data modeling and integration
- Creating and analytics ecosystem

3 What's next?

What we have learned and what we are taking on next

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.



What does Goldman Sachs do?

- A leading global investment bank and financial services company
- Customers include institutions, corporations, consumers
- Technology is fundamental to everything we do

$1.4T
in assets

10,000
engineers

100
locations

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Who you will find on a trading floor

Sales

Generate ideas
Cover clients
Manage relationships

Traders

Analyze markets
Execute transactions
Manage risk

Engineers

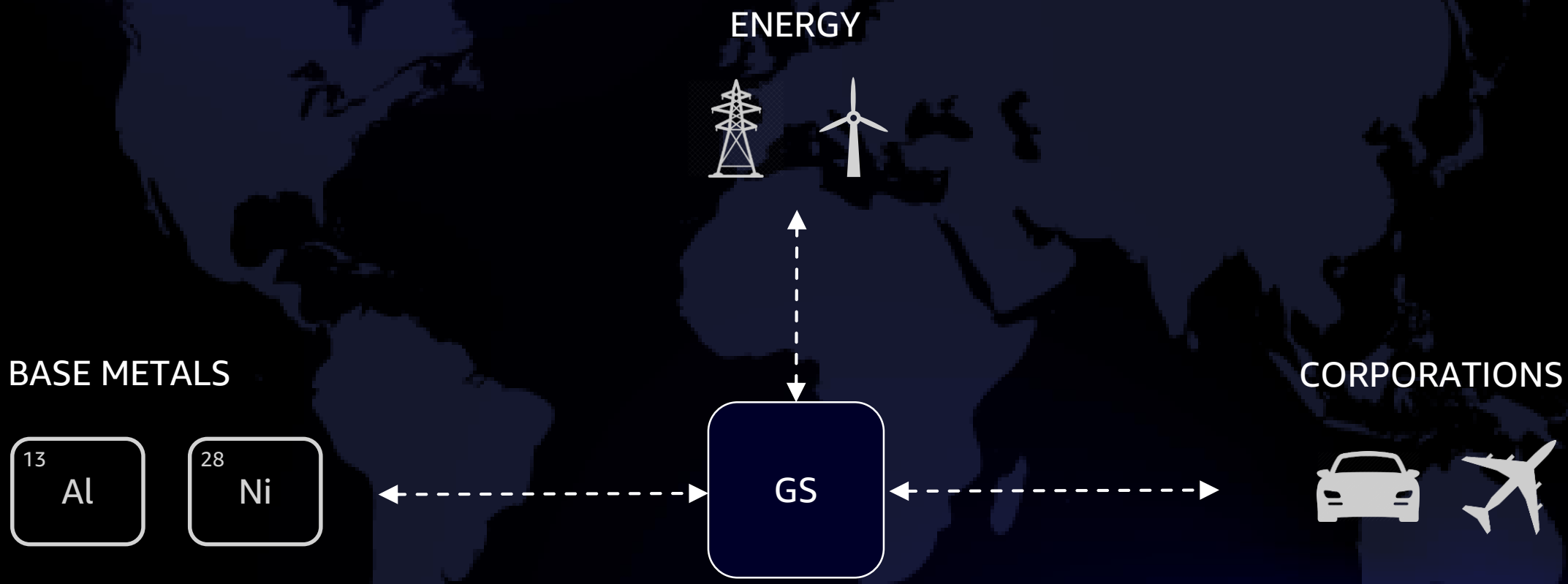
Architect solutions
Develop applications
Manage infrastructure

Quants

Analyze data
Develop strategies
Manage models

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

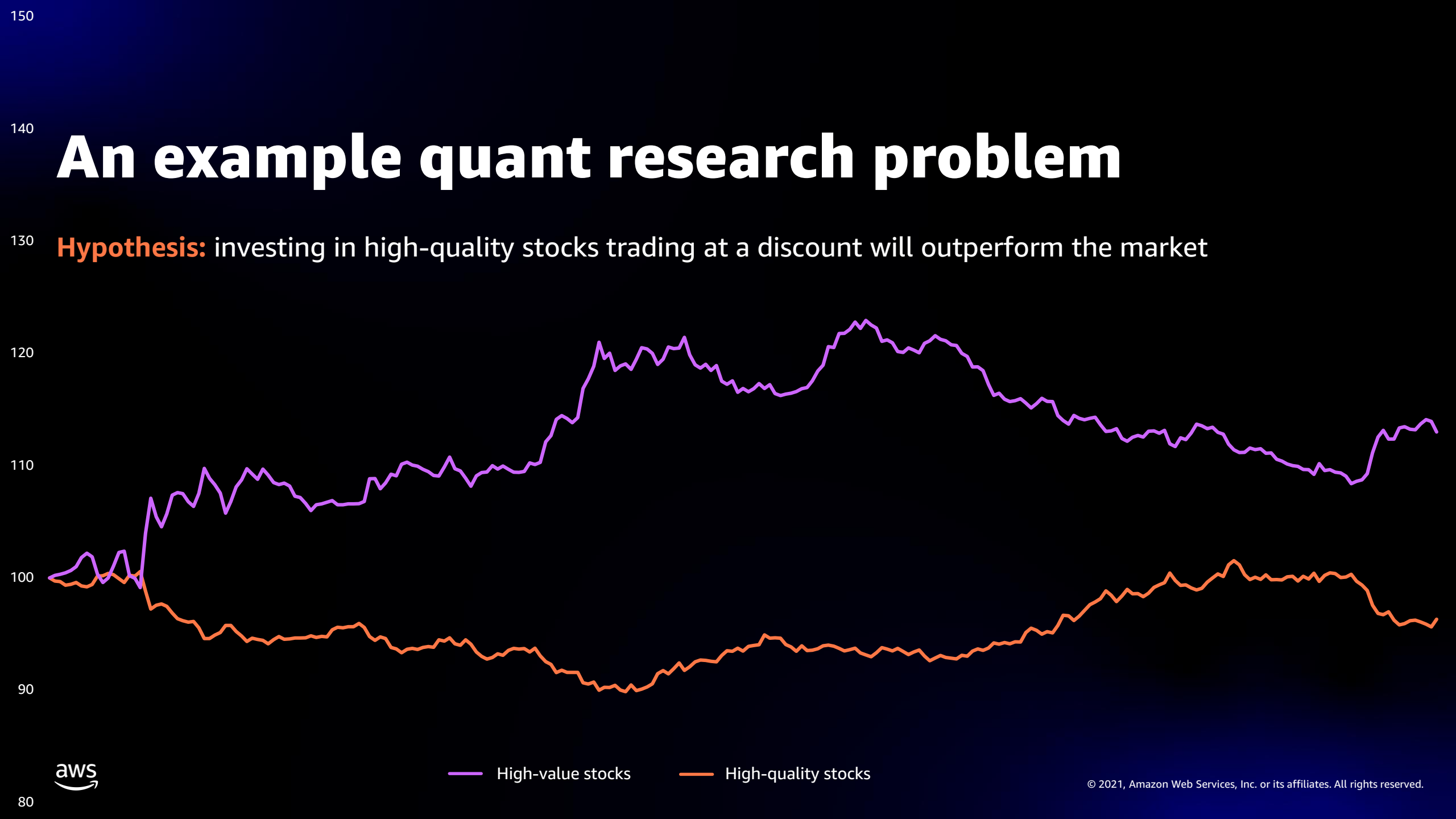
Global markets



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

An example quant research problem

Hypothesis: investing in high-quality stocks trading at a discount will outperform the market



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Methodology



Setup

Deploy infrastructure

Source data



Develop

Create strategy

Optimize backtest



Analyze

Compare metrics

Run scenarios

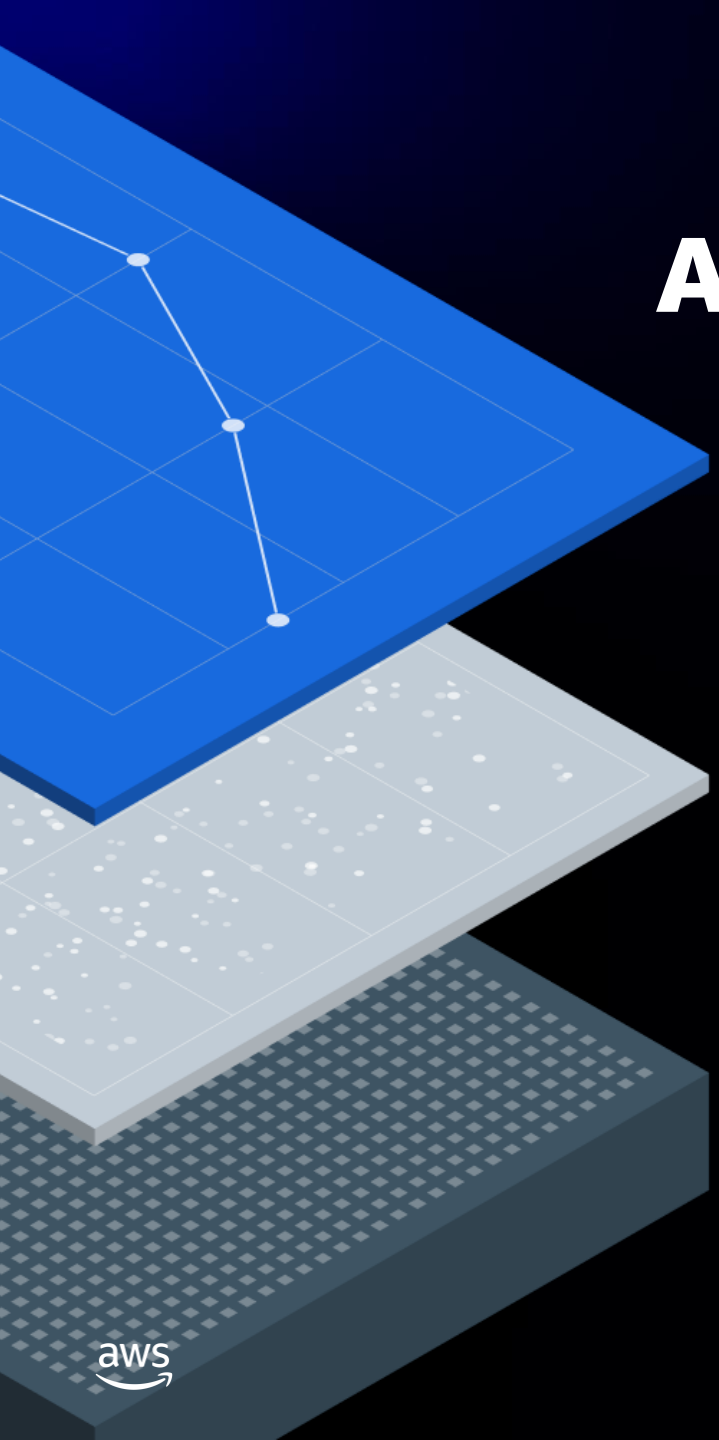


Productionize

Execute trades

Monitor risk

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.



Architecture of a data platform

Analytics

Web applications running parallelized quantitative analytics jobs, with the ability to do further research directly in a notebook environment

Data modeling & integration

Consistent data models for securities and derivatives, exposed through a centralized data catalog on top of both relational and object databases

Tick database

Horizontally scalable in-memory, time-series database handles thousands of updates per second across millions of securities at millisecond latency

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

The cloud opportunity

Re-architect our data platform on AWS, enabling us to do what we already do. . . but better

Velocity	Spin up infrastructure on demand; accelerate data acquisition and modeling
Elasticity	Dynamically adjust compute and storage to manage peak volumes efficiently
Innovation	Access latest scientific compute ecosystem natively in the cloud
Security	Isolate infrastructure and leverage cloud-native security protocols
Collaboration	Redefine interactions between our developers and our clients

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

PART 1:
TICK DATABASE

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Real-time market data

Intra-day data reflecting changes to asset prices in real time

- Sparse time tables at nanosecond resolution
- Most recent data must be accessible with low latency
- Bi-temporal data requiring idempotent updates
- Out-of-sequence backfill and corrections in real time

10M updates/sec

25M+ assets

150 markets

TIME	BID	ASK	TRD
2021-01-26 14:30:32.573275266			88.25
2021-01-26 14:30:32.628910783	88.24	88.32	
2021-01-26 14:30:32.628911913			88.24
2021-01-26 14:30:32.628912153	88.24	88.32	
2021-01-26 14:30:32.628912373	88.24	88.32	
2021-01-26 14:30:32.628914563	88.24	88.50	
2021-01-26 14:30:32.628988674	88.20	88.50	
2021-01-26 14:30:32.629008244	88.15	88.50	
2021-01-26 14:30:32.629288140			88.20
2021-01-26 14:30:32.629288637	88.06	88.50	
2021-01-26 14:30:32.629289584	88.14	88.50	
2021-01-26 14:30:32.629291960	88.15	88.50	
2021-01-26 14:30:32.629292657			88.47
2021-01-26 14:30:32.629295212			88.47
2021-01-26 14:30:32.629965881			88.47
2021-01-26 14:30:32.629969803			88.47
2021-01-26 14:30:32.630215817	88.04	88.50	
2021-01-26 14:30:32.630218046	88.06	88.50	
2021-01-26 14:30:32.630218492	88.48	88.50	
2021-01-26 14:30:32.630218957	88.48	88.50	
2021-01-26 14:30:32.630220185	88.48	88.50	
2021-01-26 14:30:32.630591421	88.48	88.50	
2021-01-26 14:30:32.630595227			88.50
2021-01-26 14:30:32.630595699			88.51
2021-01-26 14:30:32.630598022			88.60
2021-01-26 14:30:32.630947063	89.00	89.03	
2021-01-26 14:30:32.630947771			88.68
2021-01-26 14:30:32.630951768			88.80
2021-01-26 14:30:32.630952680			88.85
2021-01-26 14:30:32.630953171			88.87
2021-01-26 14:30:32.631230068			88.88
2021-01-26 14:30:32.631230536			88.89
2021-01-26 14:30:32.631231023			88.90
2021-01-26 14:30:32.631231258	89.00	89.03	
2021-01-26 14:30:32.631235856			88.91
2021-01-26 14:30:32.631236635	89.00	89.03	
2021-01-26 14:30:32.631237799			88.96
2021-01-26 14:30:32.631238254			88.98
2021-01-26 14:30:32.631238719			88.99
2021-01-26 14:30:32.631238930	89.00	89.03	
2021-01-26 14:30:32.631239380			89.00

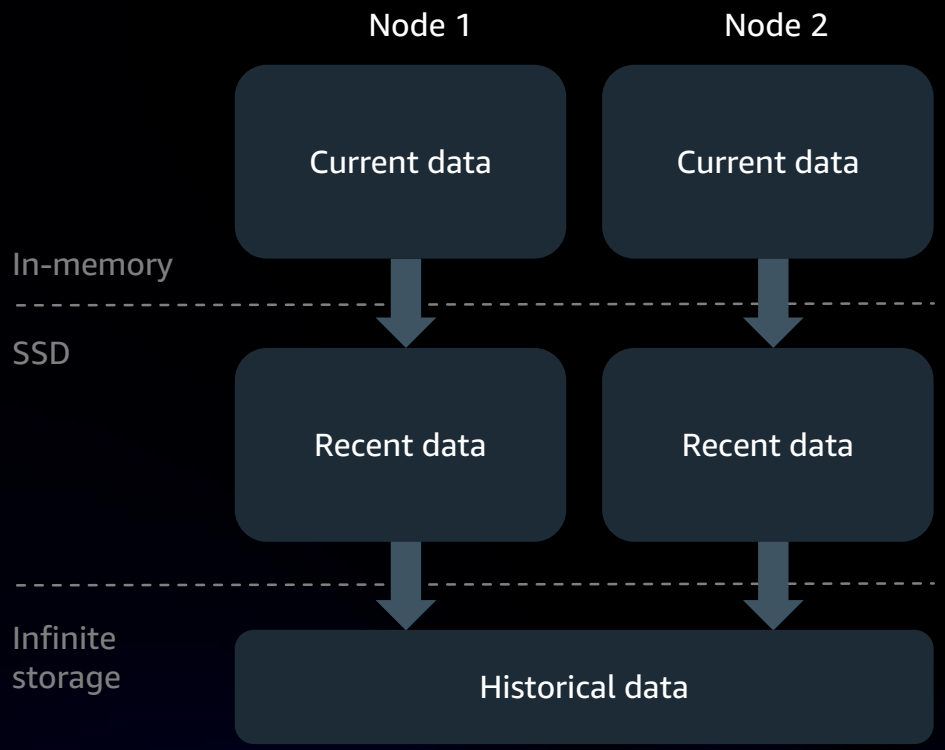
aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Inside a tick database

Tick database \ˈtik-ˌdā-tə-ˌbās\
noun: A time-series store optimized for low latency and real-time data access patterns commonly employed in financial markets

- Three-tier architecture for millisecond read times
- High-throughput writes (500K/sec/symbol)
- Horizontally scalable to handle petabytes of data
- Schema-less and fault-tolerant under live trading
- User-defined server-side computations
- ACID transactions for regulatory compliance

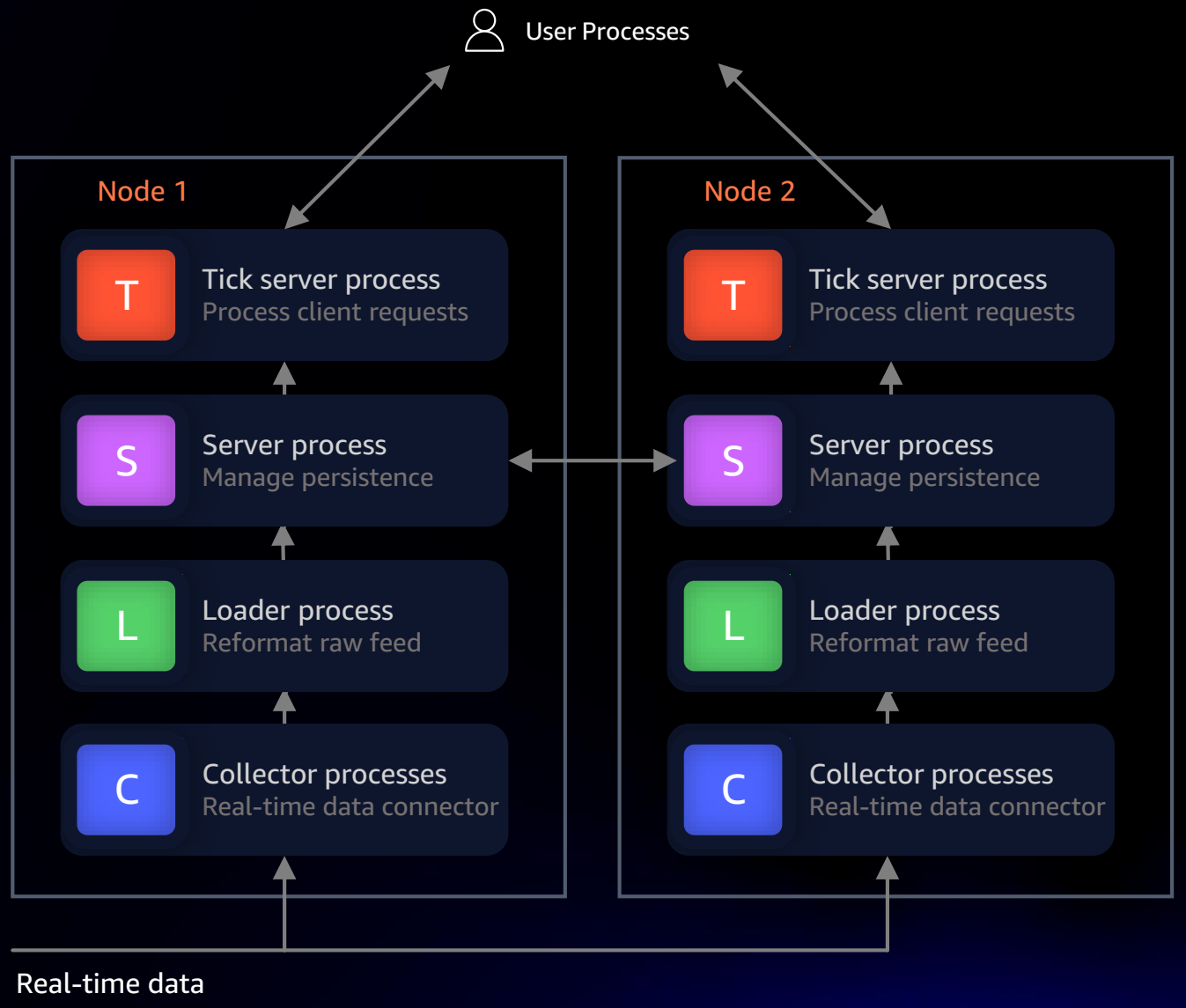


© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Distributed setup

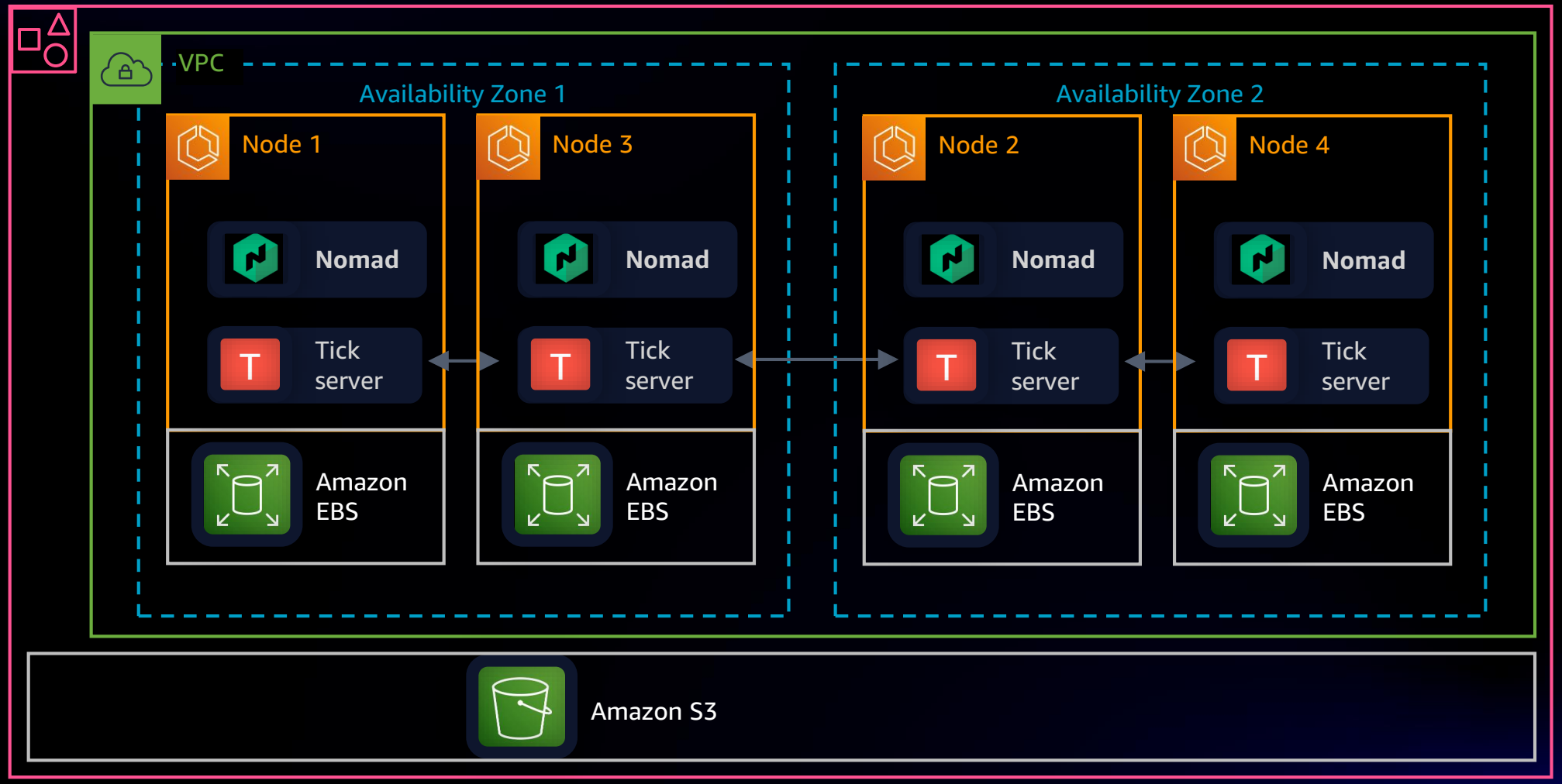
Scales horizontally with throughput increasing linearly with node count

- Clients access via **tick server process**
- **Server process** handles persistence to each data store as well as replication
- If a node fails, the full dataset is available through other processes
- Data collected by **loader process**, which hands off to dedicated servers
- Collection of raw data feeds can be parallelized across **collector processes**



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Cloud-native architecture



Amazon Managed Service for Prometheus

AWS Secrets Manager

Amazon CloudWatch

AWS KMS

AWS CloudFormation

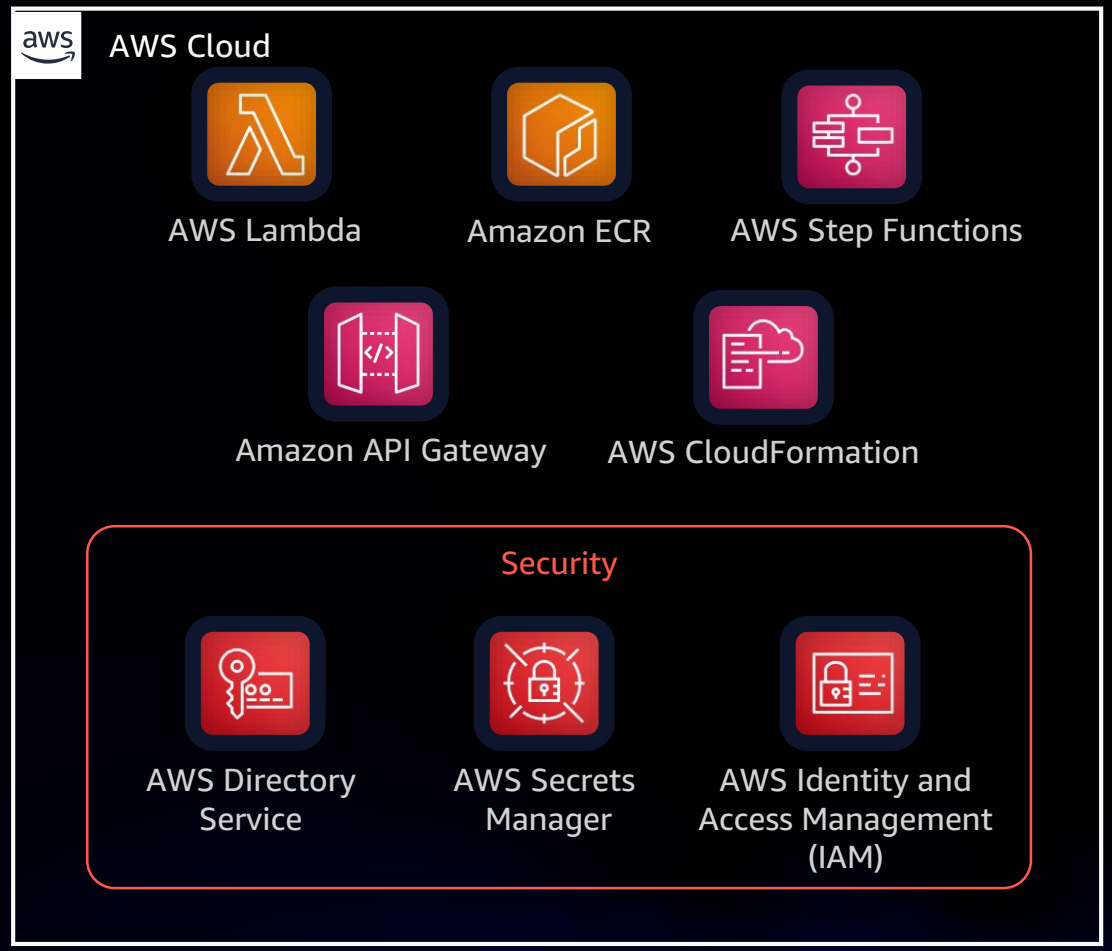
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Cluster provisioning

How do we deploy a tick database on the cloud?

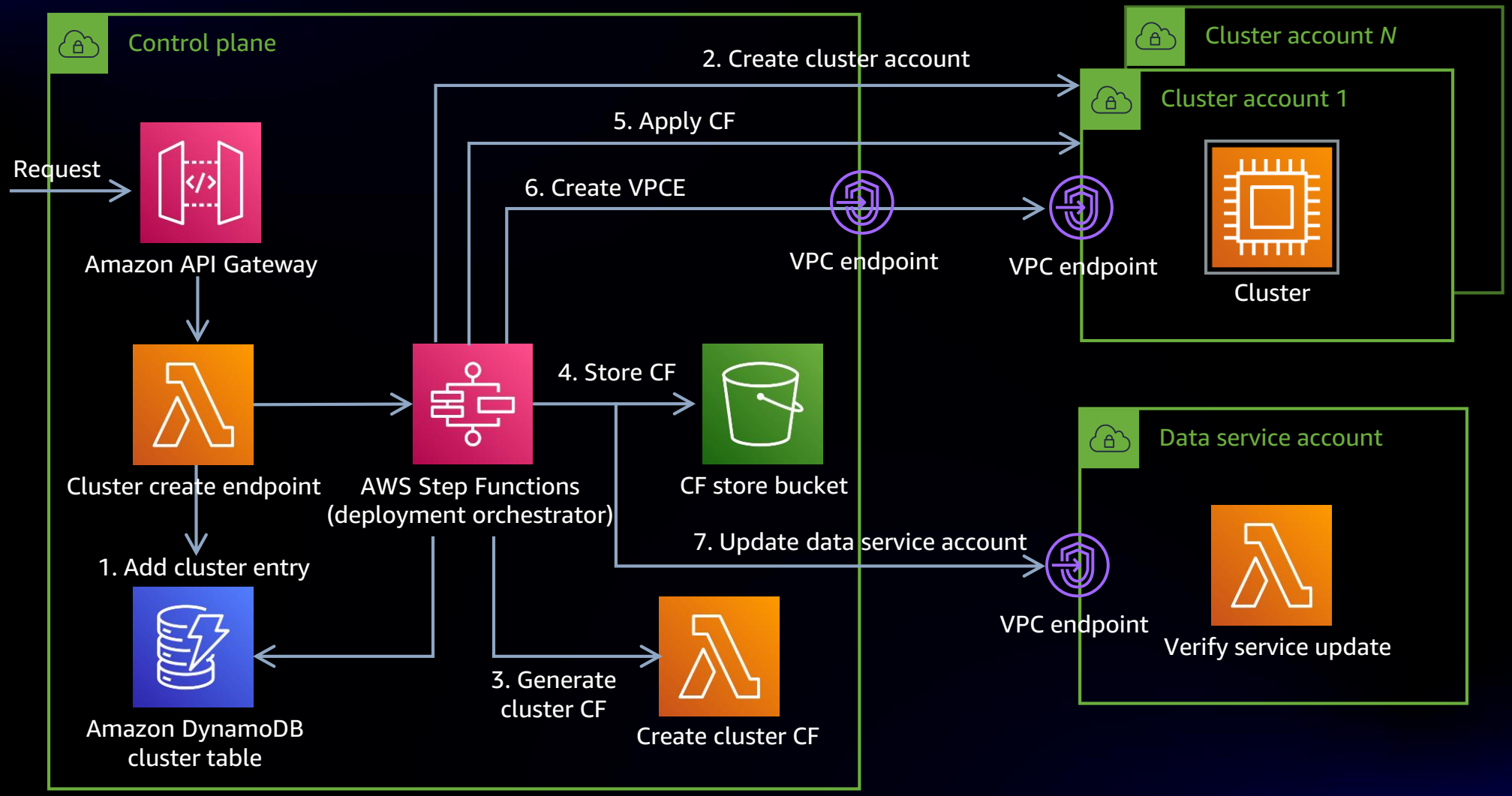
- AWS Lambda for on-demand deployments
- Amazon CloudWatch for logging
- Nomad for process orchestration
- Amazon DynamoDB for cluster metadata

End result: 2-week deployment → 5 minutes



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Cluster provisioning architecture



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

PART 2: DATA MODELING & INTEGRATION

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

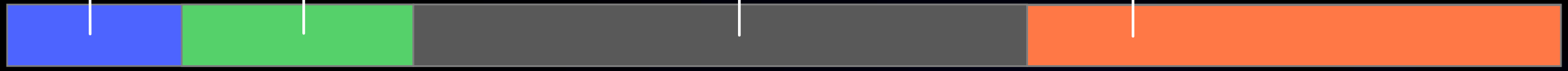
Why is this so hard?

01
Scour available dataset options
Interpret esoteric licensing models
Evaluate using limited samples

02
Ingest and parse hundreds of files
Navigate inconsistent data dictionary
Remap historical format changes

03
Interpret PDF-based manual
Remap to standard data models
Join to other data sources

04
Run technical and semantic checks
Iterate with market experts
Set up ongoing quality feedback loop

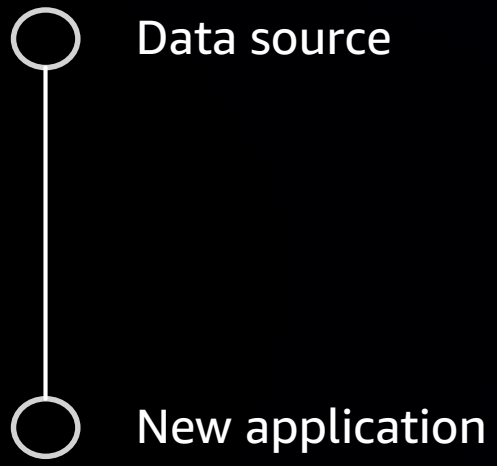


Source | Ingest | Understand & model | Validate & productionize

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

An expensive mistake

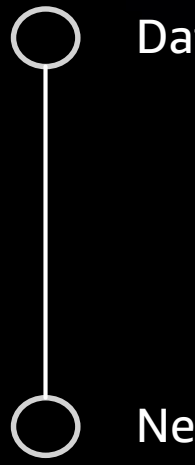
“We’ll just quickly onboard this new feed.” – Eager and entrepreneurial engineer



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

An expensive mistake

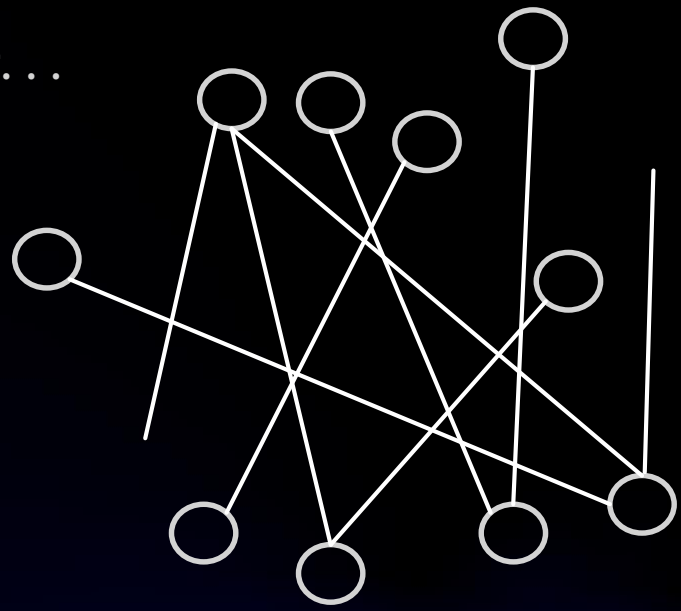
“We’ll just quickly onboard this new feed.” – Eager and entrepreneurial engineer



Data source

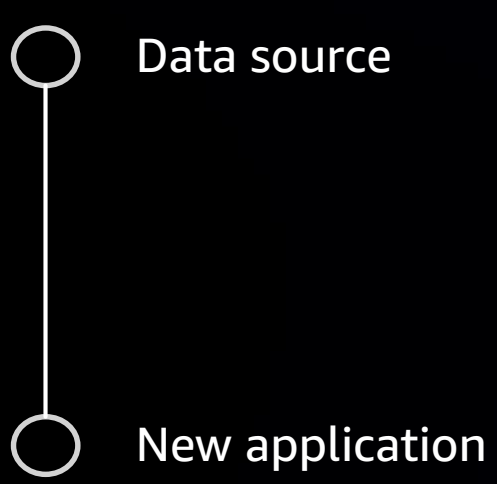
New application

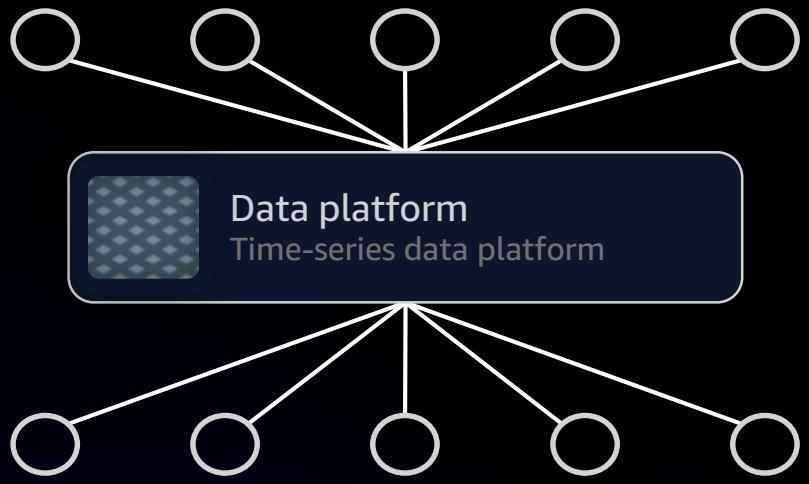
Some time later. . .



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

A better plan



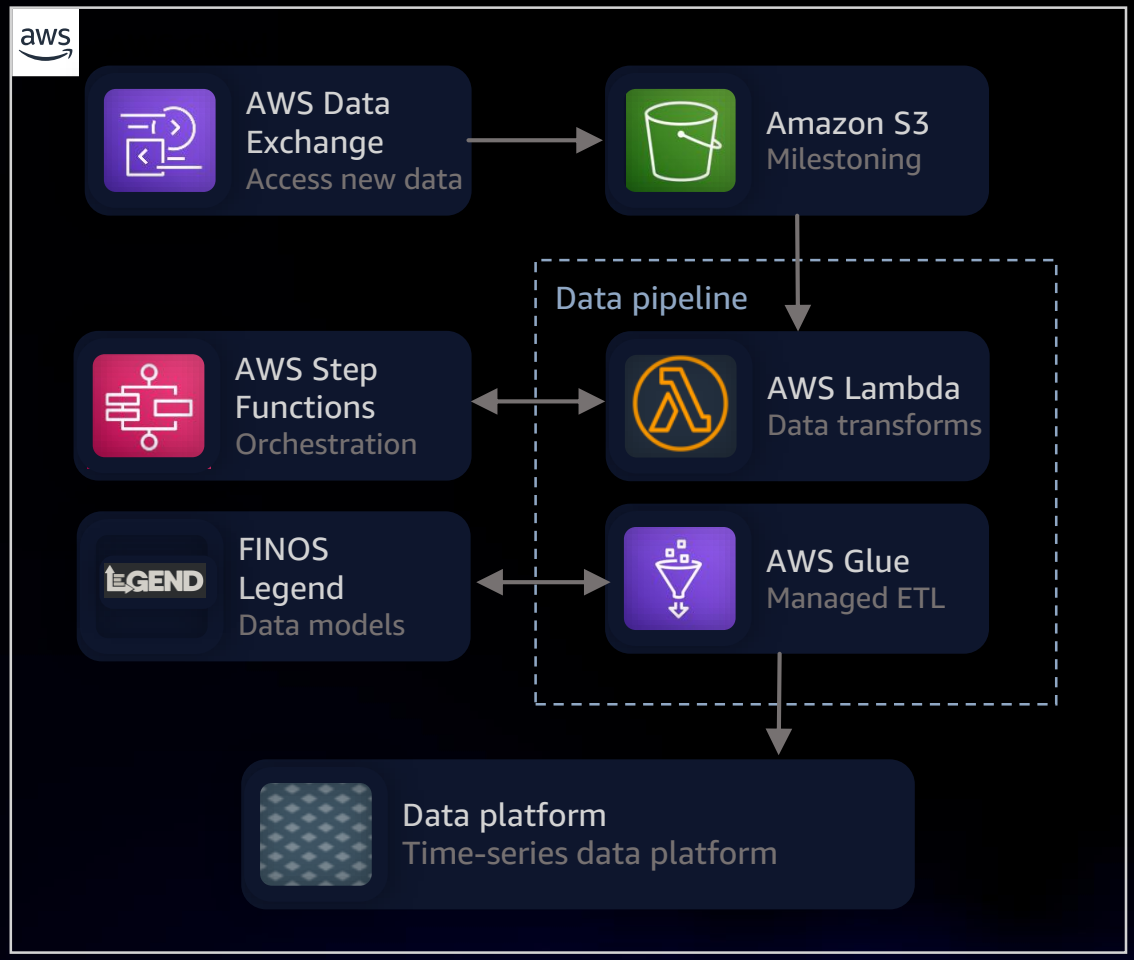


© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Desired state

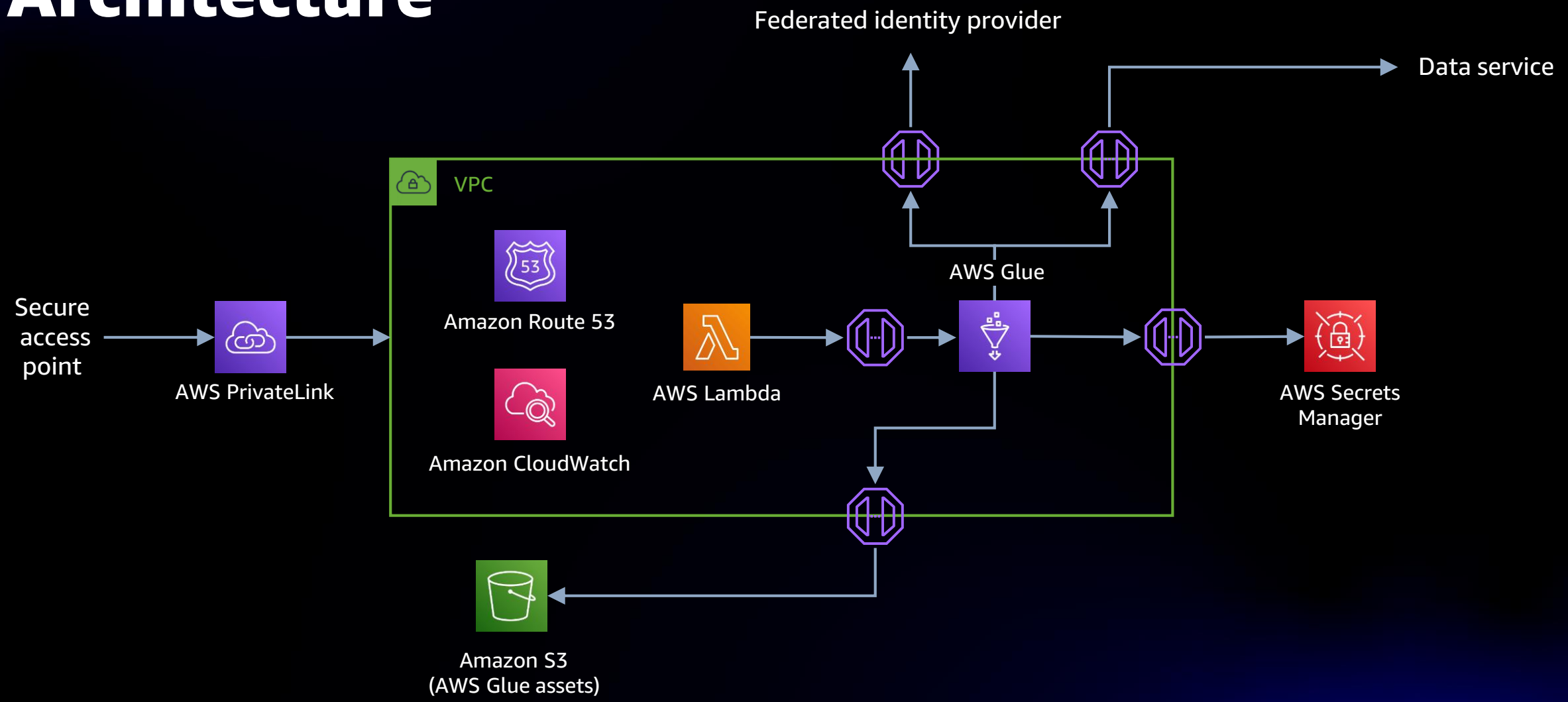
Getting existing and new data into the ecosystem efficiently, including

- Onboard once, distribute to all
- Seamless sourcing of new data feeds from AWS Data Exchange
- Combine with existing data by integrating other AWS data sources using AWS Glue
- Reuse and open-source data models through FINOS Legend



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Architecture



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Distribution & catalog

GET /v1/data/catalog/{dataSetId}

Request samples

curl | Python | Node.js | JS | Java | C#

Copy

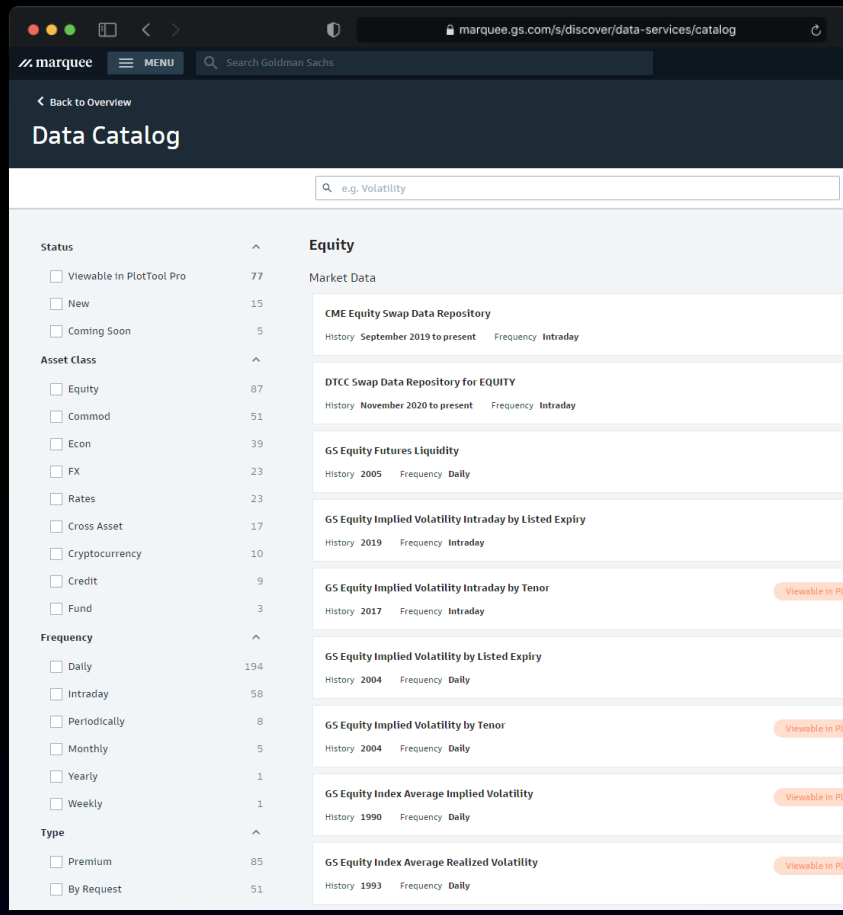
```
# Requires curl >= 7.2

curl -i -X POST \
  https://idfs.gs.com/as/token.oauth2 \
  -H 'Accept: application/json' \
  -H 'Content-Type: application/x-www-form-
urlencoded' \
  -d grant_type=client_credentials \
  -d client_id=YOUR_CLIENT_ID \
  -d client_secret=YOUR_CLIENT_SECRET

curl -i -X GET \

https://api.marquee.gs.com/v1/data/catalog/:data
SetId \
  -H 'Authorization: DashAuthz
<YOUR_TOKEN_HERE>'
```

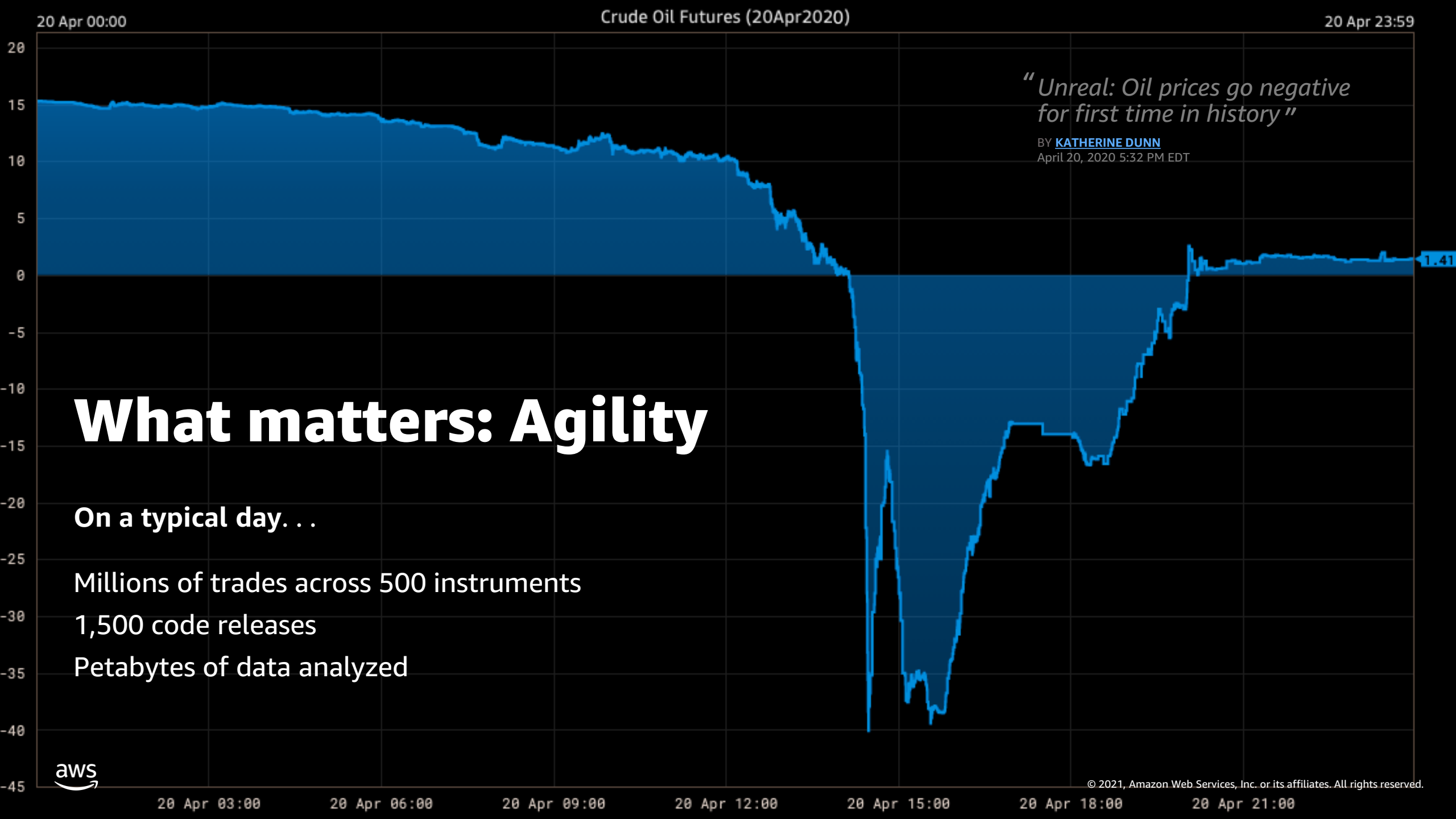
- Datasets modelled in a single catalog available programmatically or via UI
- Schema can be altered dynamically using common field definitions
- Data entitlements enforced at the dataset, row, and field level
- Data can be accessed through a single API, irrespective of database location



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

PART 3: ANALYTICS

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.



"Unreal: Oil prices go negative for first time in history"

BY KATHERINE DUNN
April 20, 2020 5:32 PM EDT

What matters: Agility

On a typical day. . .

Millions of trades across 500 instruments

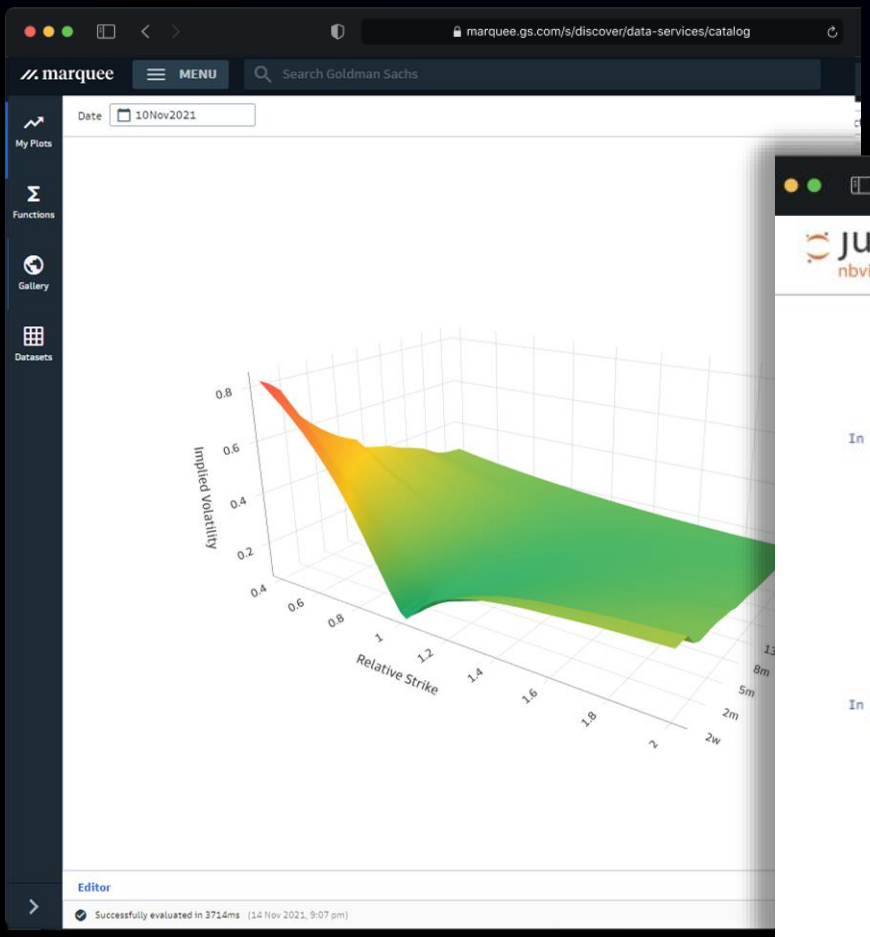
1,500 code releases

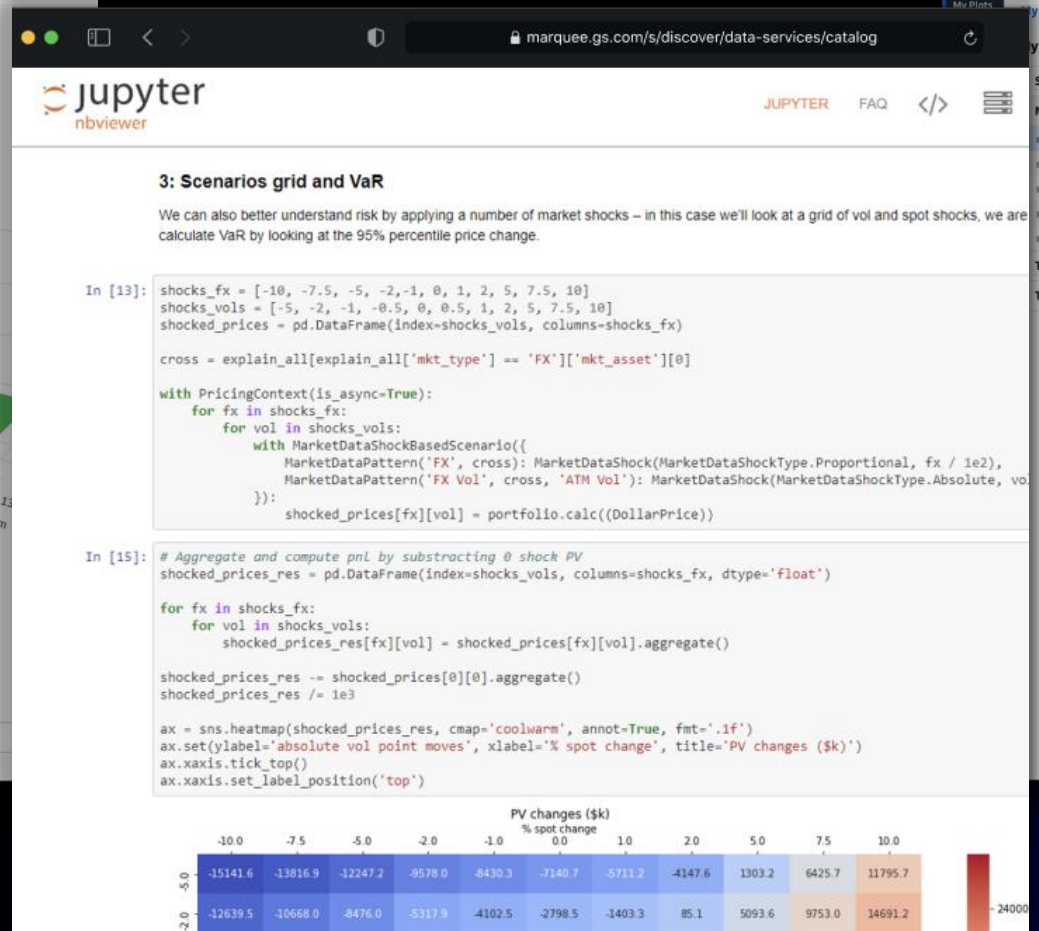
Petabytes of data analyzed

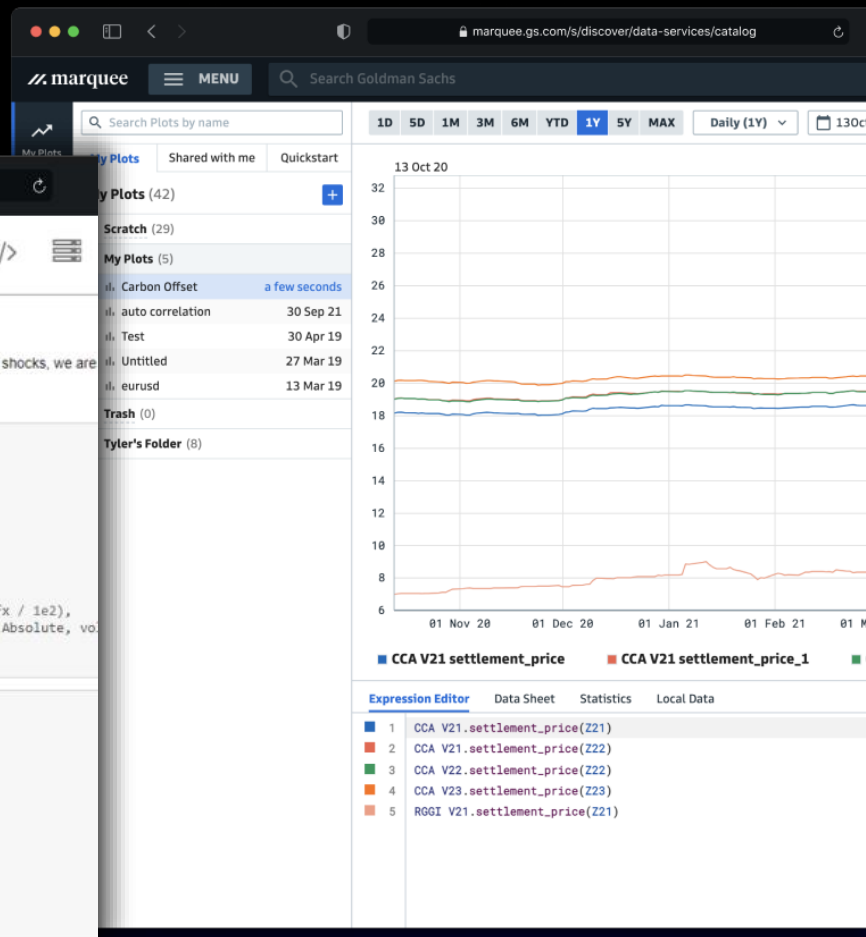
aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Quantitative analytics



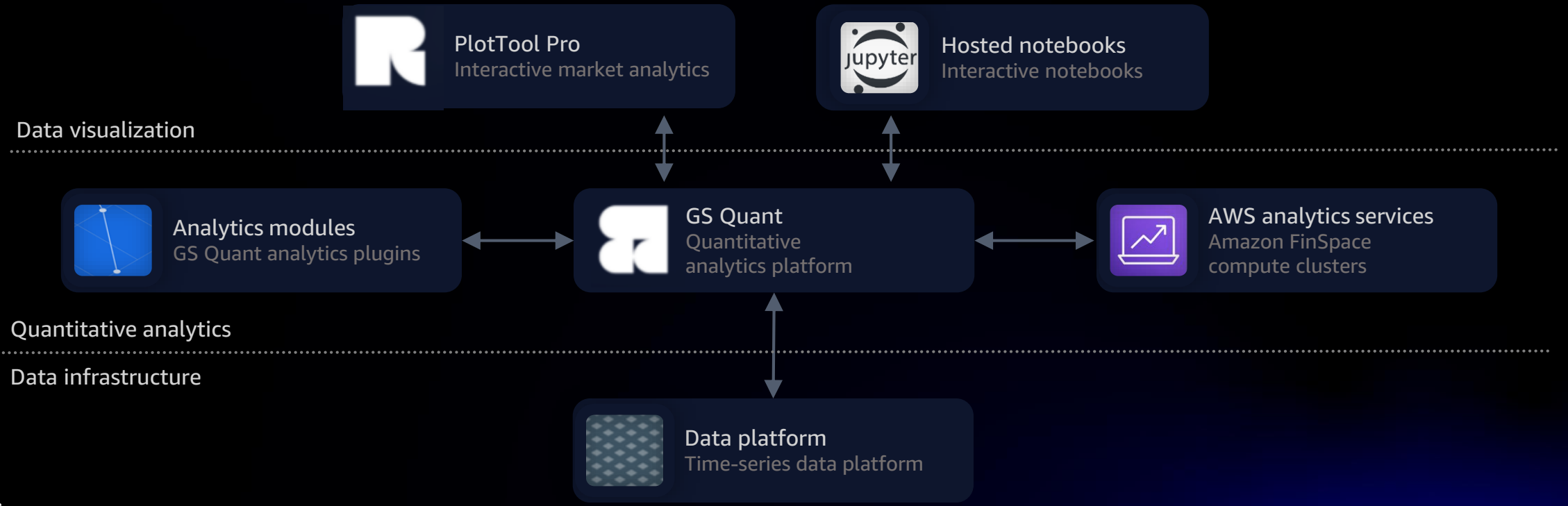




© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Analytics overview

Data analysis and visualization allowing quants to contribute their own analytics and run at scale



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

import gs_quant

- **GS Quant:** Open-source Python quantitative toolkit
- Access to 500+ datasets in pandas
- 100+ time-series analytics functions
- SDK for all core APIs (data, pricing, risk, etc.)
- Expose financial market model (securities, derivatives)
- Internal == external

github.com/goldmansachs/gs-quant

```
from gs_quant.session import GsSession
from gs_quant.markets.securities import SecurityMaster, AssetIdentifier
from gs_quant.data import DataFrequency, DataMeasure, DataContext
from datetime import datetime

GsSession.use()


asset = SecurityMaster.get_asset('AMZN', AssetIdentifier.TICKER)

measures = [
    DataMeasure.BID_PRICE,
    DataMeasure.ASK_PRICE,
    DataMeasure.TRADE_PRICE
]

with DataContext(start=datetime(2021,1,26,14,30,0), end=datetime(2021, 1, 26, 21,30)):

    table= asset.get_data_frame(measures, frequency=DataFrequency.REAL_TIME)

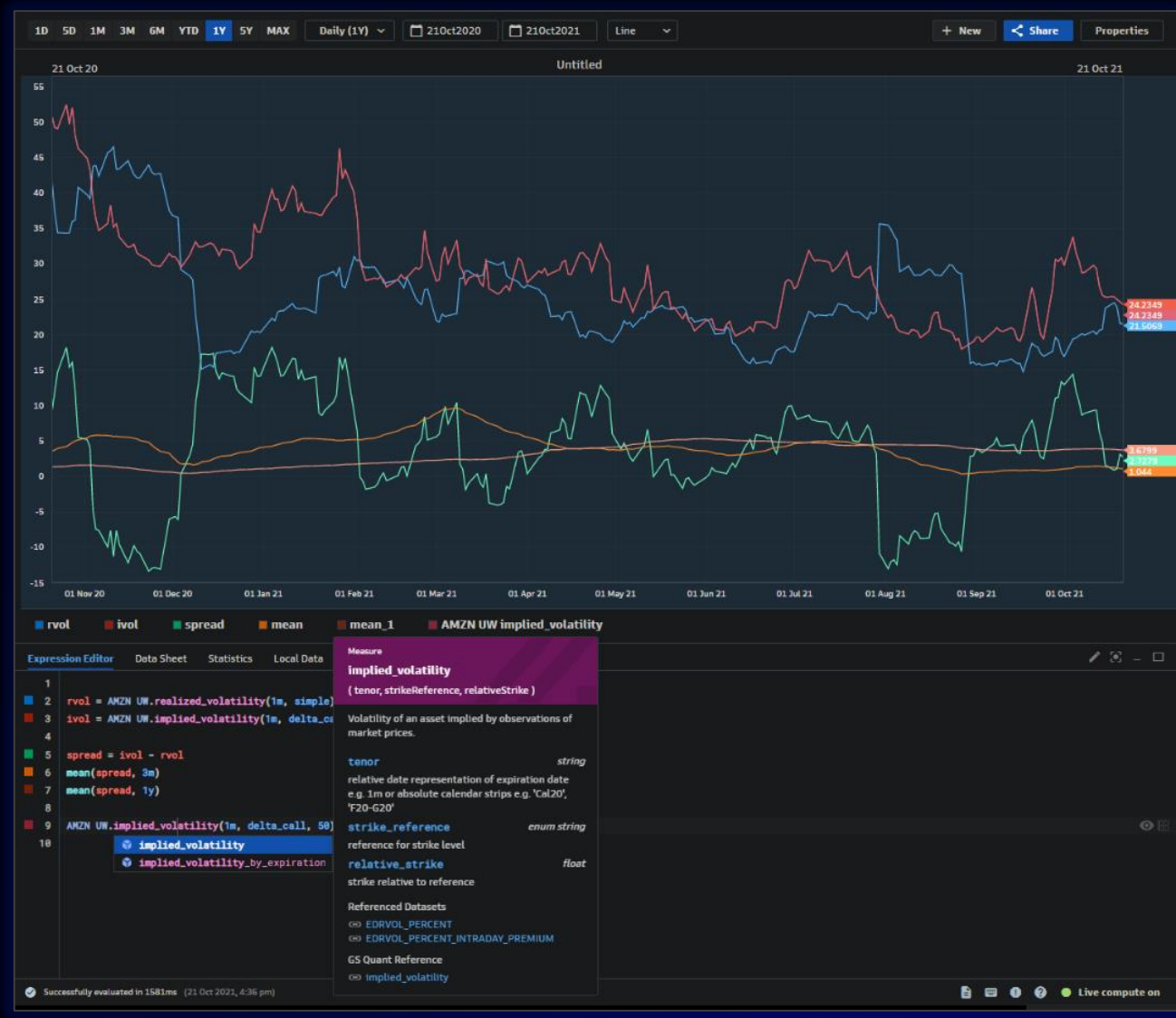
table.head(50)
```

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Interactive analytics

- Point-and-click solution? No thanks!
- DSL for financial time-series analytics
- Expose semantic data and object model
- Written with `lark` and `gs_quant`
- Enable low-code access to quant functions
- Automatic down sampling for UI
- Elasticsearch for type-aheads



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

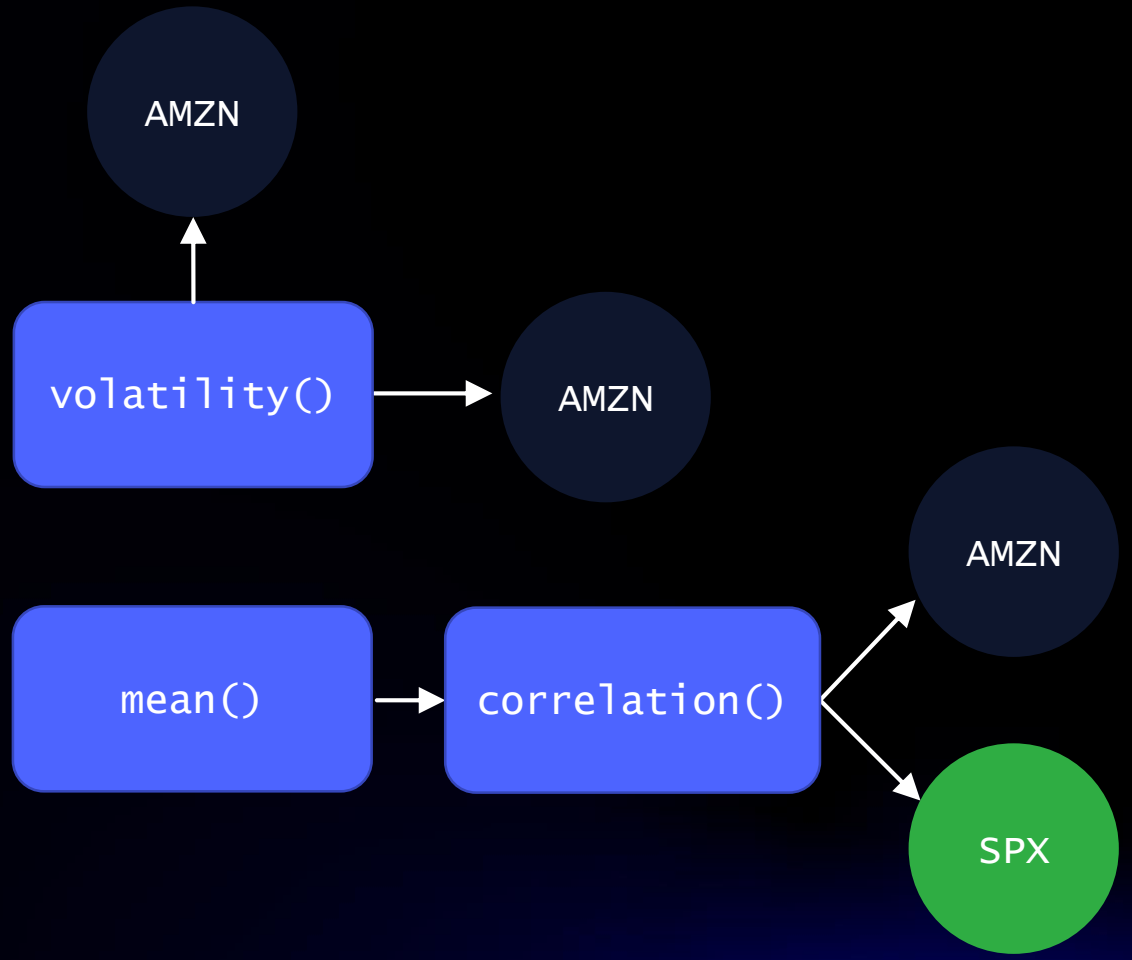
Graph calculation & caching

- Functional with parallel execution
- Data caching for each node

```
1  AMZN UW.close()
2  volatility(AMZN UW.close())
3  mean(correlation(AMZN UW.close(), SPX.close()))
```

```
from lark import Lark, Transformer, Token


GRAMMAR = r"""
    ?start: assign | expr
    expr: comparison | call | math | datetime | nested | number
        | STRING | null | general_plus -> access_plus
        | method | _coordinate -> call | array | boolean | N_REL_DATE
    comparison: expr COMPARE_OP expr
    nested: "(" expr ")"
    array: "[" expr ("," expr)* "]"
    call: general "(" [expr ("," expr)*] ")"
```

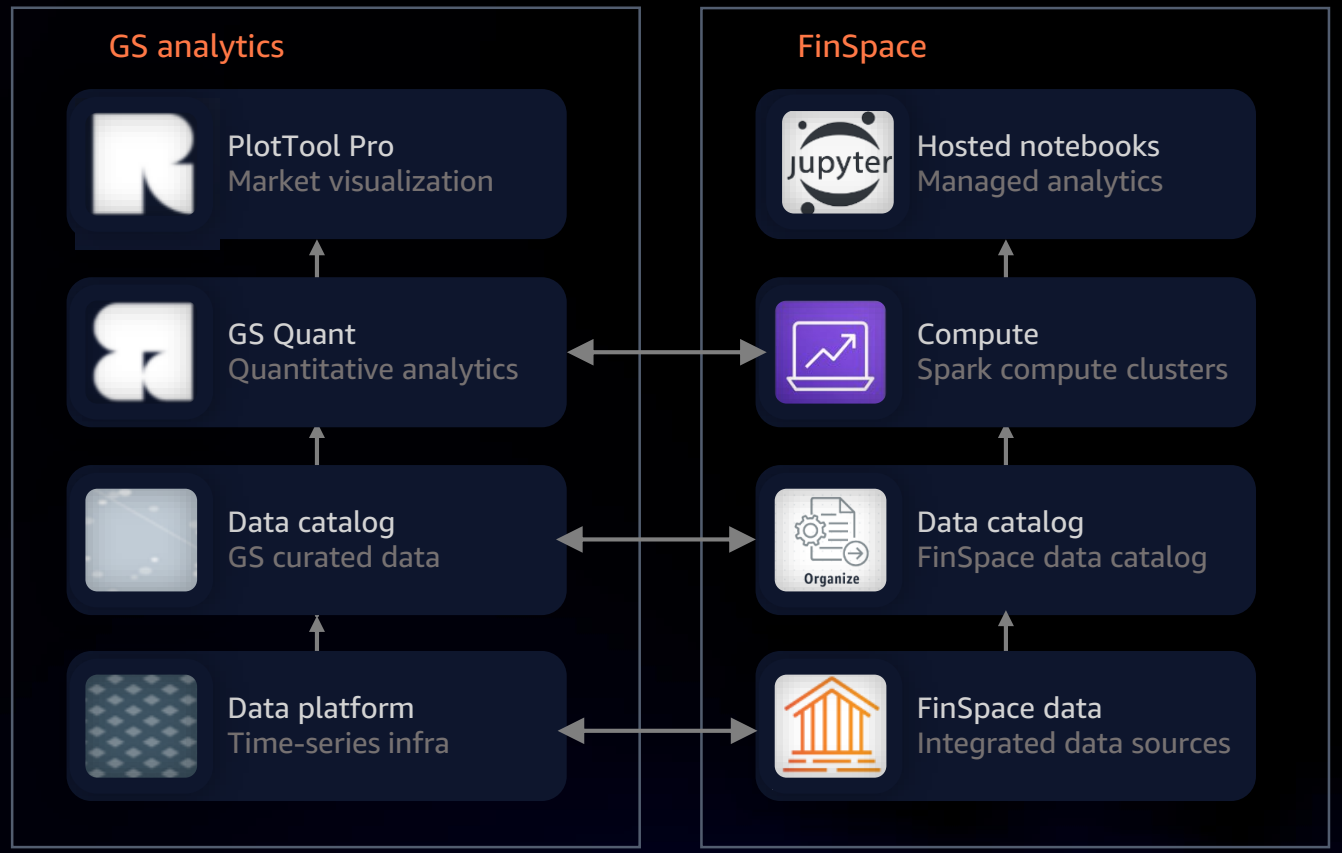


© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

GS + Amazon FinSpace

Combine GS curated data and quantitative analytics with FinSpace-hosted notebooks and compute

- Leverage FinSpace-managed Spark compute clusters on top of GS Quant
- Integrate GS curated data with other AWS services
- Run FinSpace notebook environment preconfigured with GS analytics



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

WHAT'S NEXT?

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Lessons learned

Usability	Build with operator simplicity in mind
Risk management	Running in hybrid mode is challenging but usually necessary
Performance	Consider network performance characteristics up front
Security	Design security from the ground up – avoid trying to retrofit

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

GS Financial Cloud for Data

GS Financial Cloud lowers the barrier to entry to sophisticated financial data management and analytics through a **one-stop financial data management and analytics solution on the cloud**

1. Accelerate time to market with market-tested data and sophisticated real-time analytics tools, freeing up tech and quant resources to focus on alpha-generation

2. Future-proof data strategy with cloud-native services that are designed to scale horizontally and are proven across a global trading business

3. Retain flexibility and cost control through modular and secure services charged on a consumption basis, which can be composed independently

4. Integrate natively with other AWS services to build solutions tailored to any environment to retain full control of any cloud data strategy

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Learn how you can get access to this technology at the Goldman Sachs booth:

#540

Join GS at the Financial Services Lounge from 3:45–4:45 for more info!

developer.gs.com

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Goldman Sachs

Thank you!

Descriptions and images of the products and services provided in this presentation are for informational purposes only and do not reflect all information that may be relevant in determining whether use of any such product or service is suitable for your circumstances. Nothing in these materials constitutes an offer, or solicitation to make an offer, to sell or buy any security or other product. Any information is for indicative purposes only, correspond to the date this material was prepared, and may change from time to time. Goldman Sachs does not make any representation or warranty as to the accuracy or completeness of the material, nor is it under any obligation, express or implied, to update the information to correct any inaccuracies or to reflect any changes that may occur in the future. The Goldman Sachs products and services discussed are for institutional and professional clients only and your use of any products and services will be subject to the applicable terms and conditions. Some of the services and products described in these materials may not yet be available or may not be available in certain jurisdictions or to certain types of client. Please contact your Goldman Sachs sales representative with any questions. These materials are given for purely indicative purposes and do not create any contractual relationship between you and Goldman Sachs. These materials are confidential and may not be disclosed to any third party (with the exception of external advisers on the condition that they themselves are bound by this confidentiality undertaking), or copied in whole or in part, without the prior written consent of Goldman Sachs. Please see https://www.goldmansachs.com/disclaimer/sec-div-disclaimers-for-electronic-comms.html for additional information.

2021 Goldman Sachs & Co. All Rights reserved.

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.