

Public Cloud-Based Remote Access Infrastructure for Neutron Scattering Experiments at MLF, J-PARC

J-PARC Materials and Life Science
Experimental Facility

Computing group

Kentaro Moriyama (CROSS)

Takeshi Nakatani (JAEA)



CROSS



Outline

1. Introduction of J-PARC MLF

2. Project Status

3. Requirements for Remote Access

4. Public Clouds & Development Approach

- AWS, Service selection, Microservices

5. Development of Remote Access Infrastructure

- Experimental monitoring, Data distribution, Data analysis

6. Summary & Future plans

J-PARC MLF

Materials and Life Sciences Experimental Facility

- **User facility for scattering experiments**

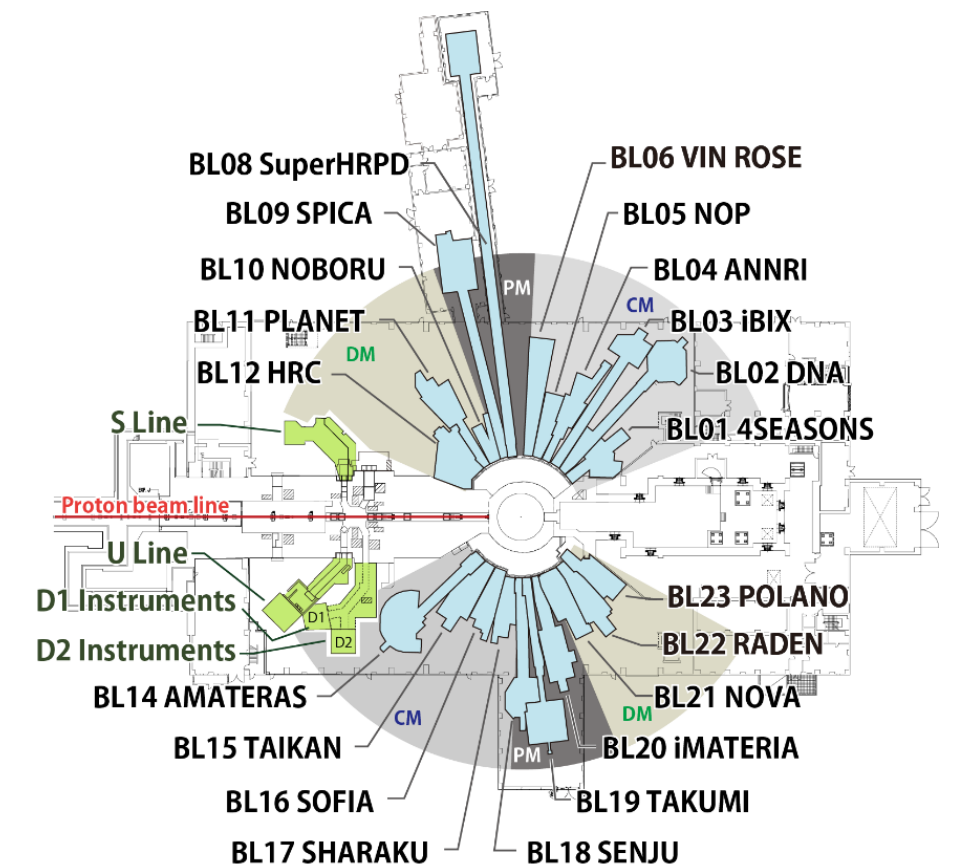
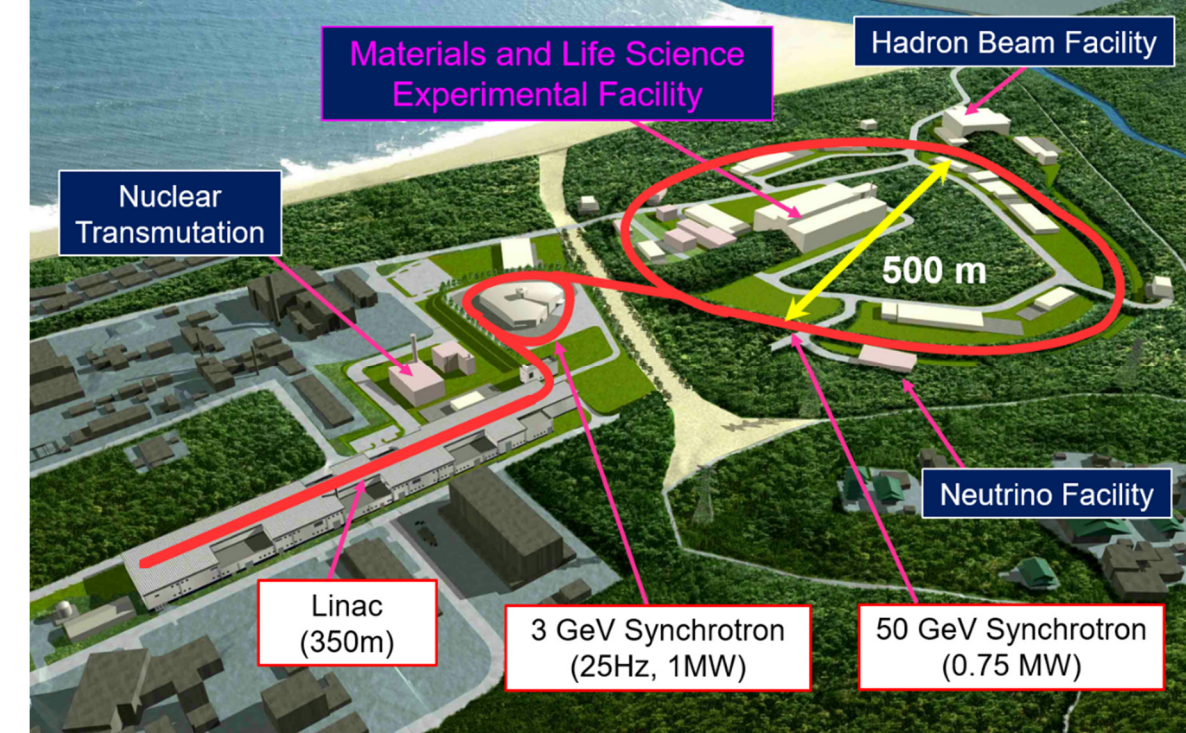
- 21 neutron beamlines
- 4 muon beamlines

- **Operation status**

- User program from 2008
- Beam power: currently 500kW (\rightarrow 1MW)
- Operating day: about 200 days /year

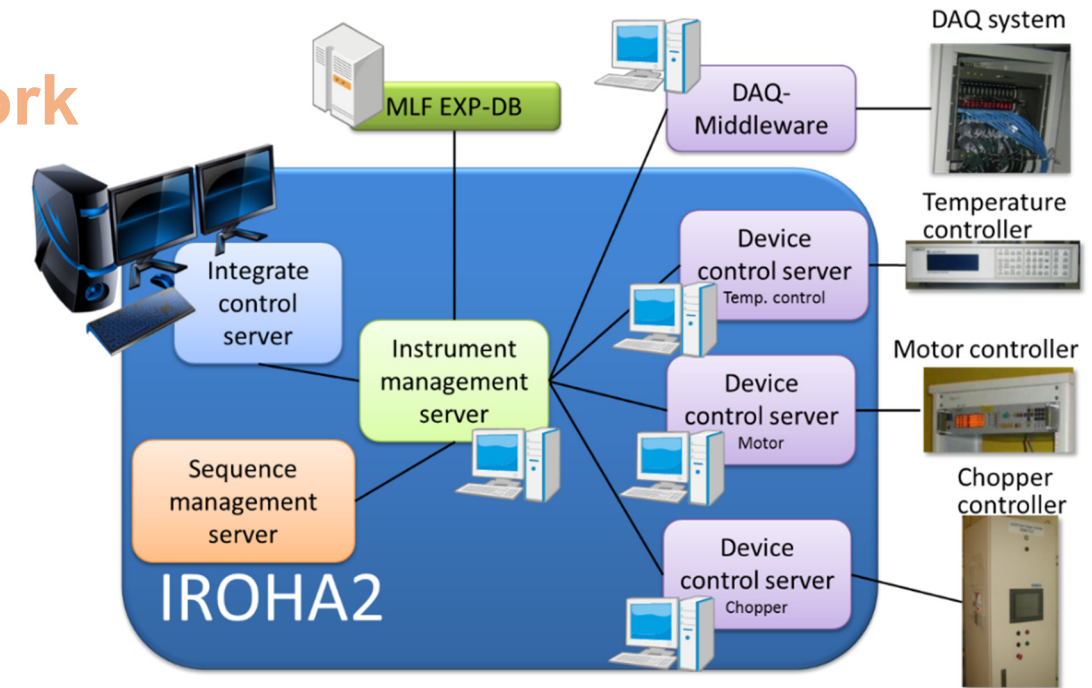
- **Usage status**

- About 1000 users / year (\rightarrow several 1000)
- About 450 experiments / year (\rightarrow 1000)



Experiments at beamline

- **IROHA2** : an integrated instrument control framework
 - four web-based software components:
 - Automatic measurements
 - Web monitoring for internal LAN
- **Data volume**
 - several hundred MB to several TB per experiment
- **Data analysis**
 - Linux-based software for data reduction and analysis
- **Data distribution**
 - Bring back data stored in the portable storage
- **Distribution of Data analysis software**
 - RPM package
 - Virtual Machine already installed analysis software.



Experiments at beamline

- **IROHA2** : an integrated instrument control framework

- four web-based software components:
- Automatic measurements
- Web monitoring for internal LAN

- **Data volume**

- several hundred MB to several TB per experiment

- **Data analysis**

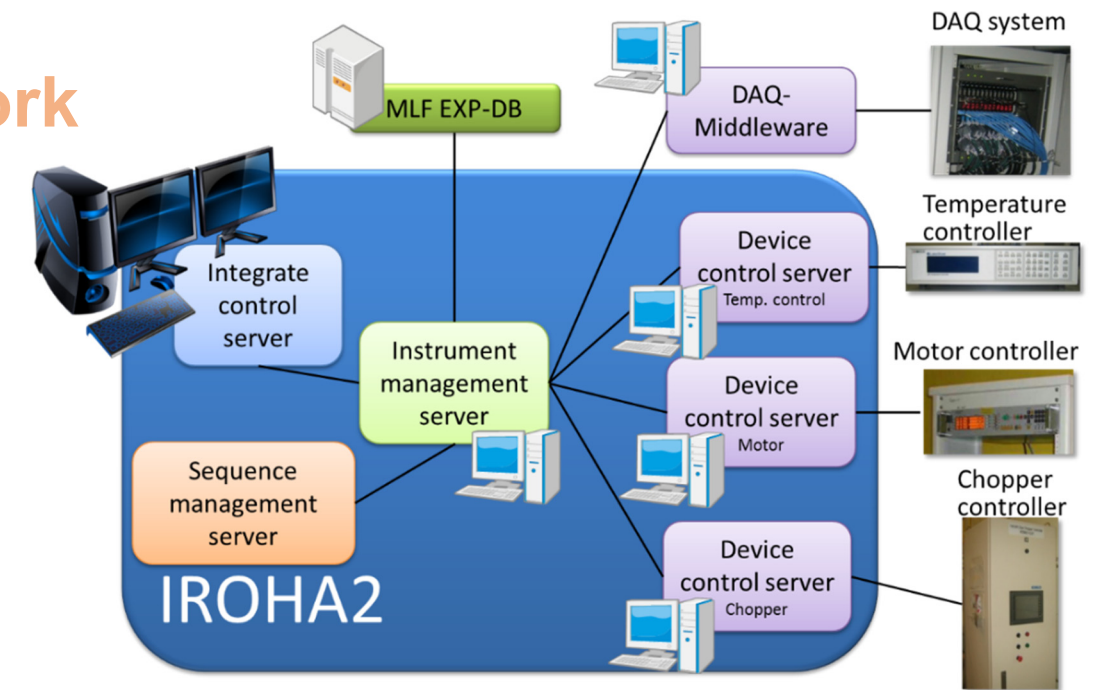
- Linux-based software for data reduction and analysis

- **Data distribution**

- Bring back data stored in the potable storage

- **Distribution of Data analysis software**

However, many facility users are unfamiliar with Linux environment setup and VM handling...



Project status

- **Started development using public cloud in FY2017.**

Since 2018, launched a simple data distribution service at MLF BL18 using Amazon S3.

- **Currently still in development & trial operation stage.**

Newly developed services:

- Experimental monitoring
- Data distribution (Web-based)
- Data analysis



Requirements for remote access

1. Monitoring measurement status

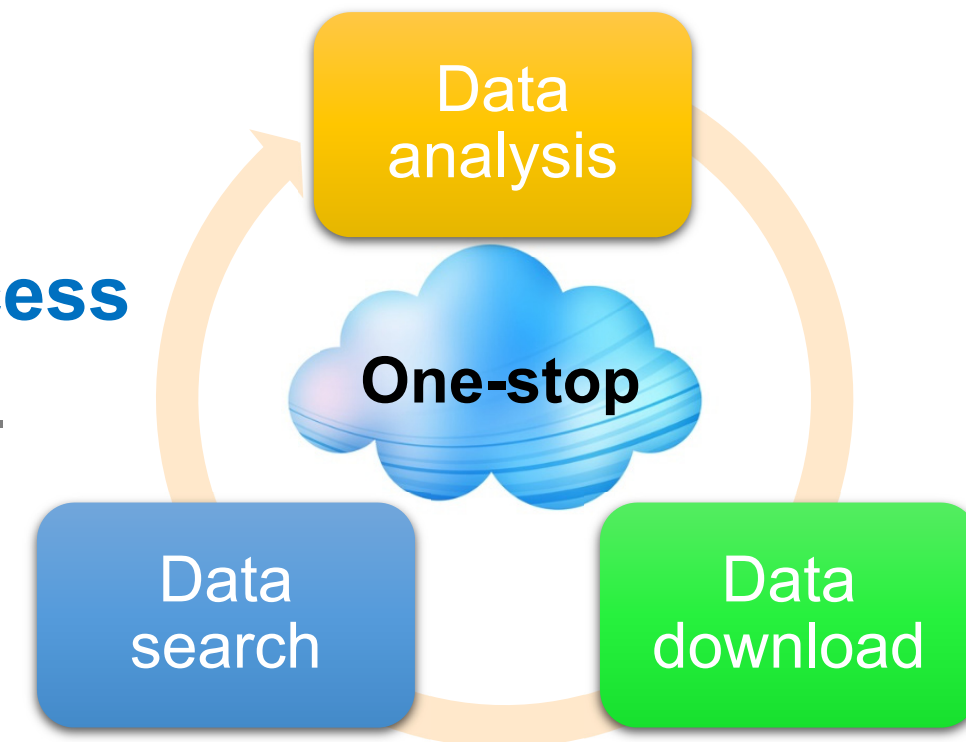
- Access from outside the facility via the internet.
- Current system operates only for staff use.

2. One-stop data search, analysis and downloading

- No need to bring back large amount of raw data to home institution.
- No need to build an analysis environment

3. Access independent of location and number of access

- High-speed access from domestic and foreign users.
- Applicability to massive simultaneous access.



Why use public cloud ?



- **No hardware management or installation cost are required.**
- **System development and operation can be started small.**
 - Operating costs can be optimized.
- **Advanced services and function provided in public cloud can be used.**
 - Effective development is possible.
 - In Particular, the **proactive use of advanced AI-related service** can be expected.
- **Security of service infrastructure is ensured at higher-level than the of on-premises systems.**
- **Global presence of public clouds allows facility users access with low latency, whether domestic or foreign.**

Development approach on public cloud

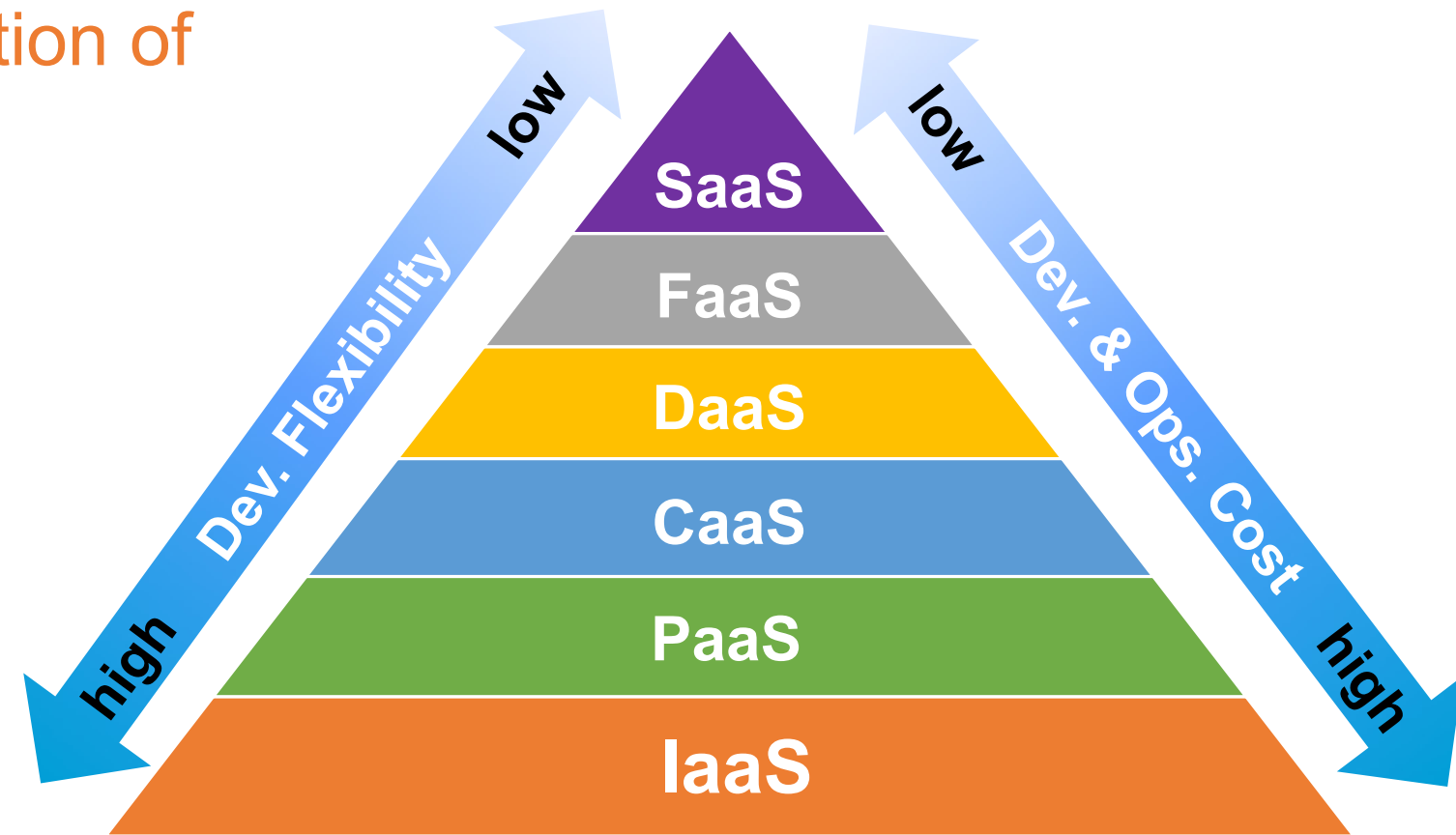
Selection policy of cloud services:

- Wide range of services from physical level to software.
- Importance of selection in terms of development & operation.

➔ Cost & flexibility depend on the location of service used.

Layer	Flexibility	Costs
Higher	Low	Low
Lower	High	High

Service selection should be take into account the trade-off between flexibility & costs.



Cloud computing service model

Development approach on public cloud

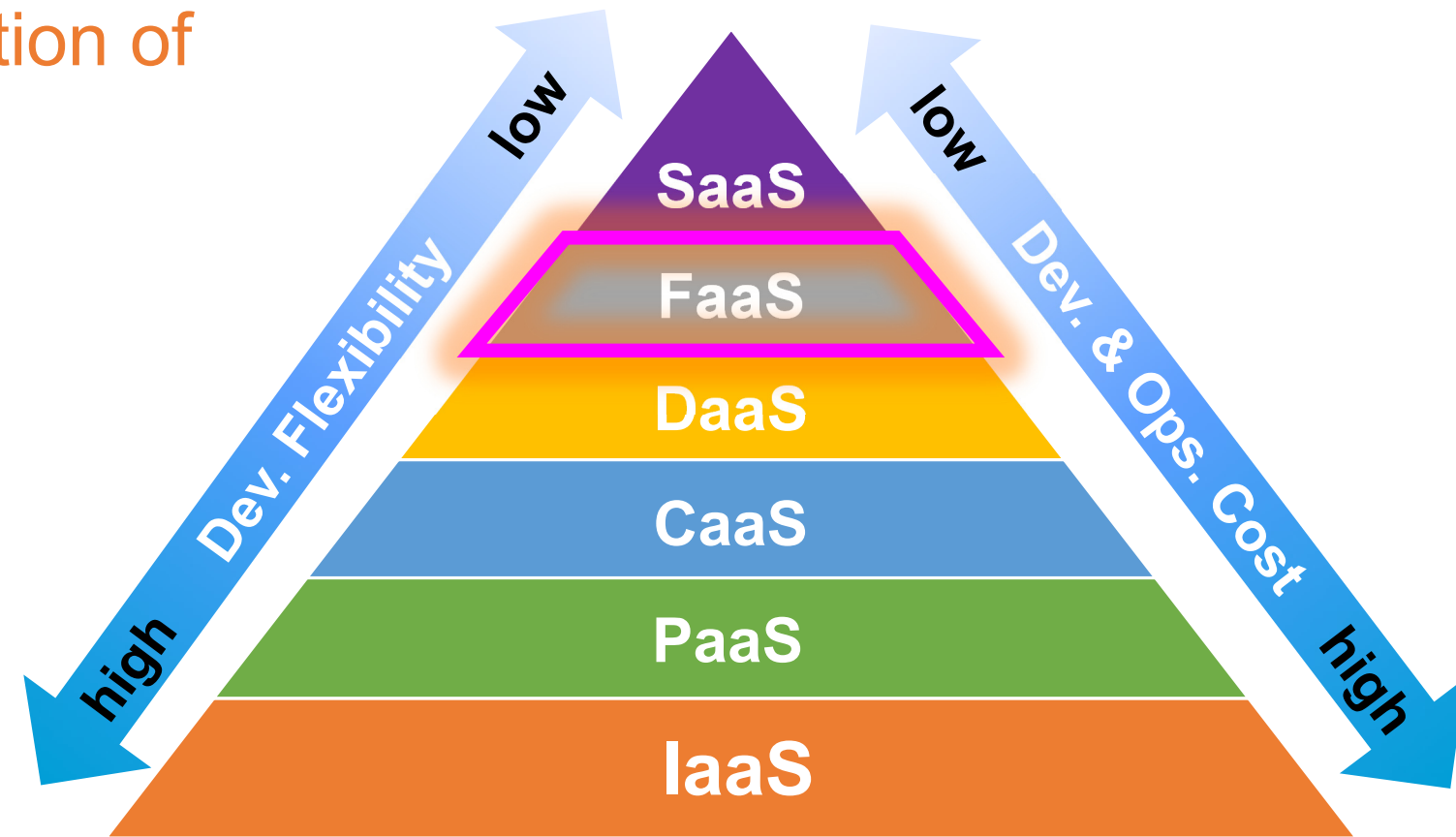
Selection policy of cloud services:

- Wide range of services from physical level to software.
- Importance of selection in terms of development & operation.

➔ Cost & flexibility depend on the location of service used.

Layer	Flexibility	Costs
Higher	Low	Low
Lower	High	High

Service selection should be take into account the trade-off between flexibility & costs.



Cloud computing service model

Monolithic vs Microservices Approach

Monolithic system

- Run all components as a single process within the system.
- Performance degradation due to failure or resource shortage can affect the entire system .

⇒ Not easy to change the system.

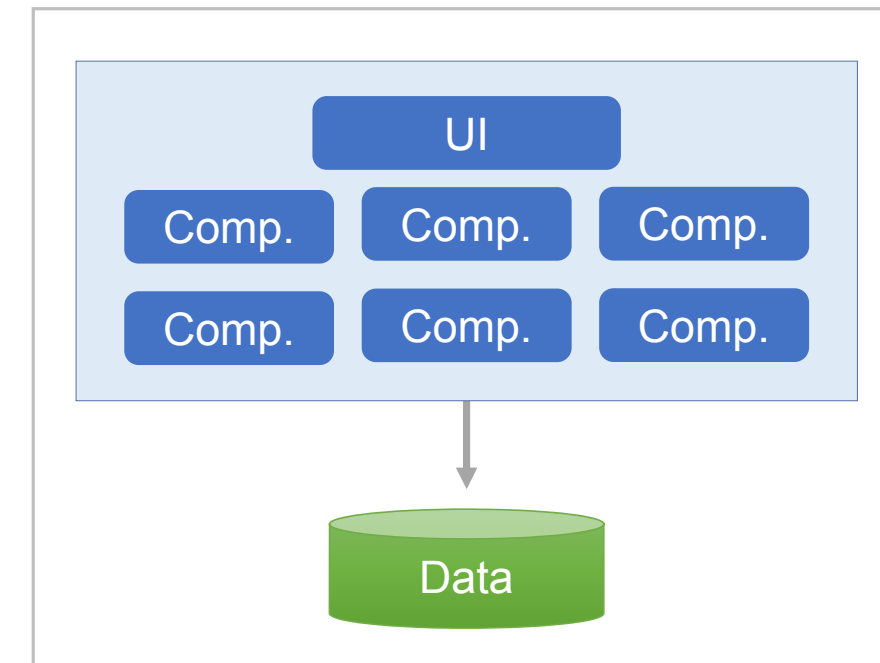
Microservices

- Divide functions into small independent services.
- Communication between services via API.
- Loosely coupled services.

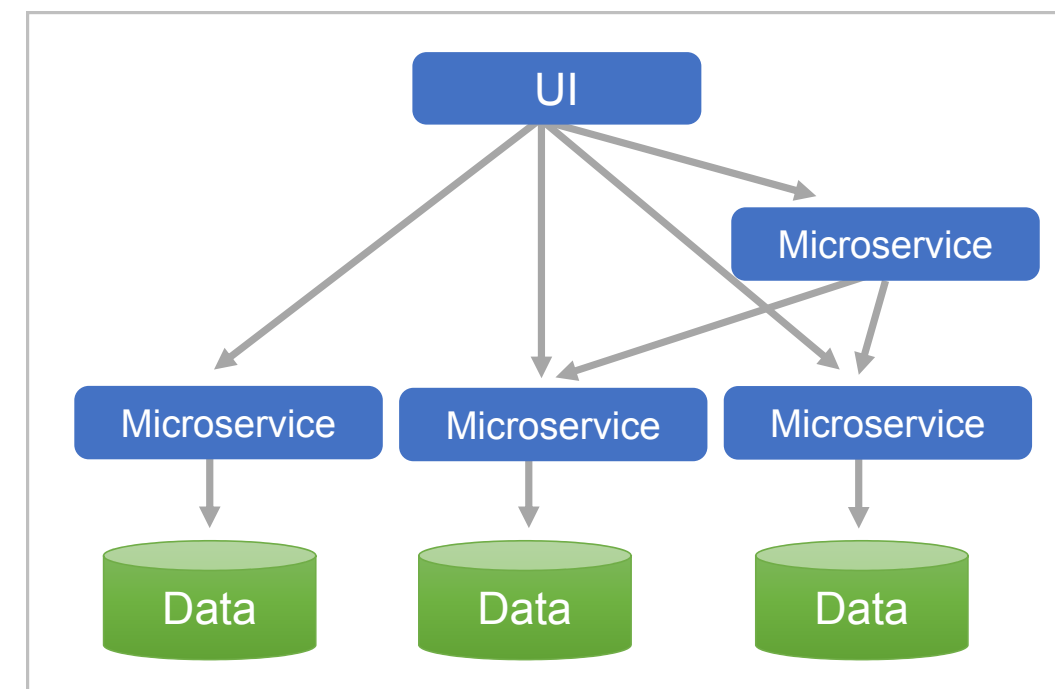
⇒

- Possible to adopt the most suitable architecture for each service.
- Possible to change, deploy, and scale individually.

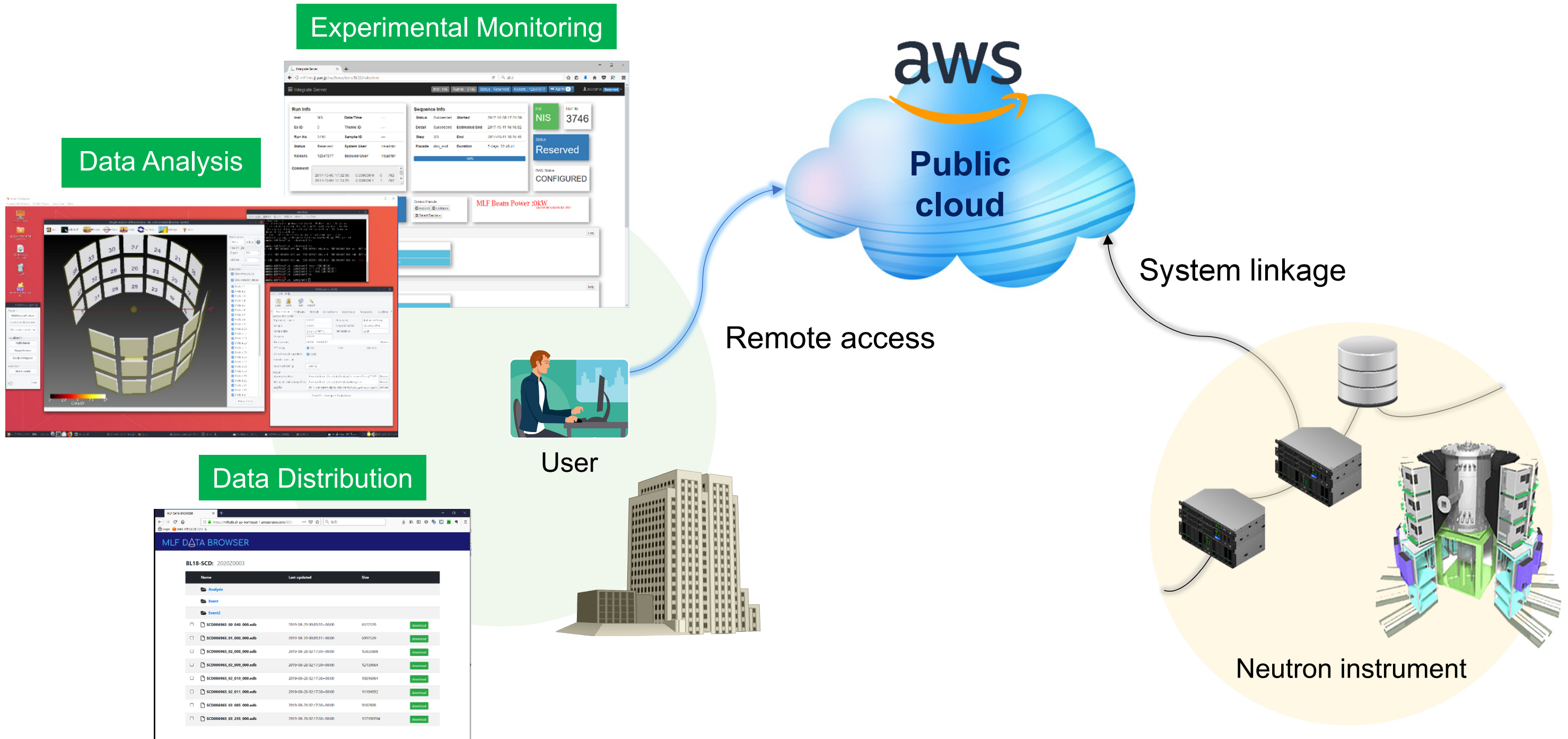
Monolithic











Microservices











Remote access infrastructure on public cloud (AWS)



Cloud services on AWS

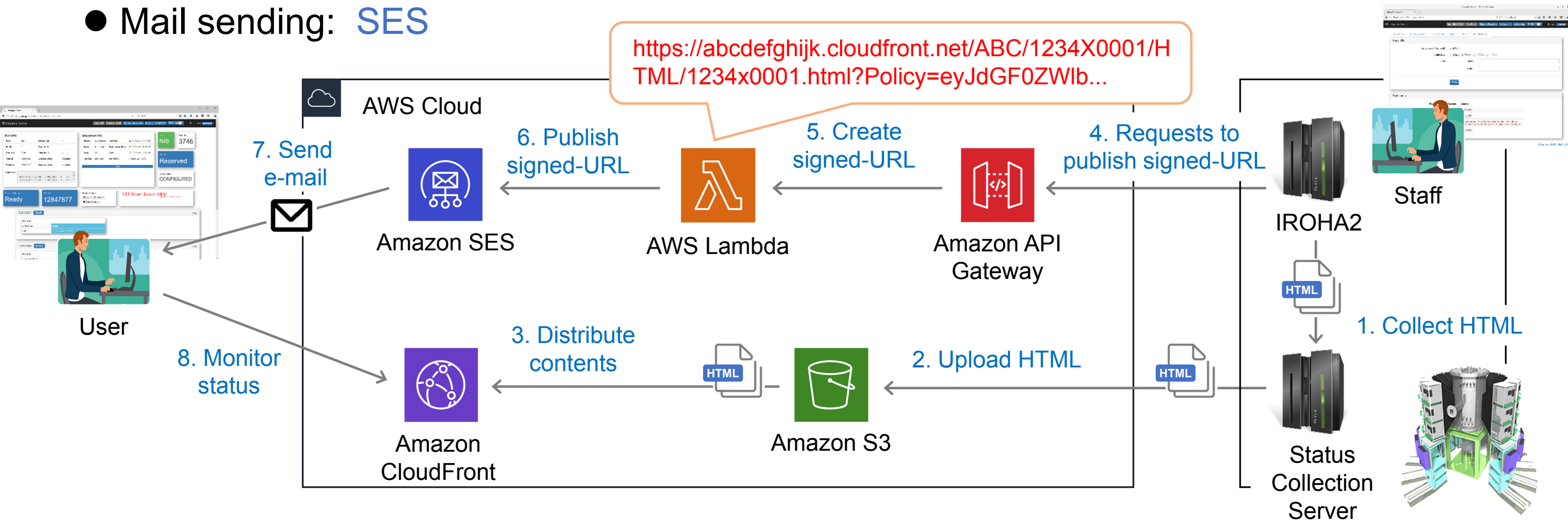
Object Storage	API Proxy	VDI	Directory Service
 Amazon Simple Storage Service	 Amazon API Gateway	 Amazon WorkSpaces	 AWS Directory Service
Serverless Computing	Content Delivery Network	Mail Sending	Storage Gateway
 AWS Lambda	 Amazon CloudFront	 Amazon Simple Email Service	 AWS Storage Gateway

Cloud services on AWS

Object Storage	API Proxy	VDI	Directory Service
 Amazon Simple Storage Service	 Amazon API Gateway	 Amazon WorkSpaces	 AWS Directory Service
Serverless Computing	Content Delivery Network	Mail Sending	Storage Gateway
 AWS Lambda	 Amazon CloudFront	 Amazon Simple Email Service	 AWS Storage Gateway

Service 1: Experimental monitoring

- Link existing on-site systems with public cloud services.
- Web distribution: S3 and CloudFront.
- Access control: signed URL of CloudFront (encrypted & signed with RSA)
- Issue of signed URL: Lambda serverless computing
- Executing Lambda function: via API Gateway
- Mail sending: SES

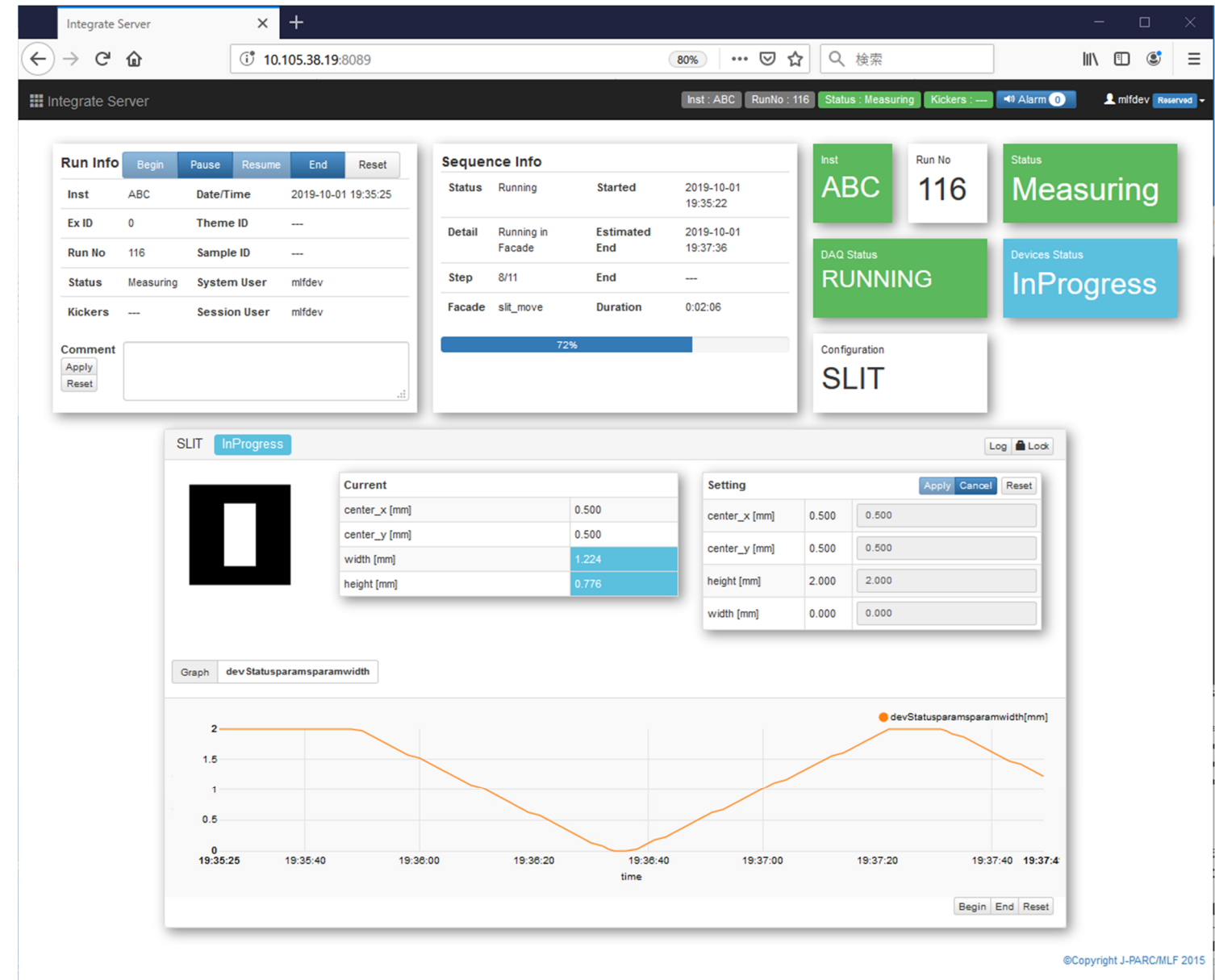


Web interface for experimental monitoring

- Measurement status

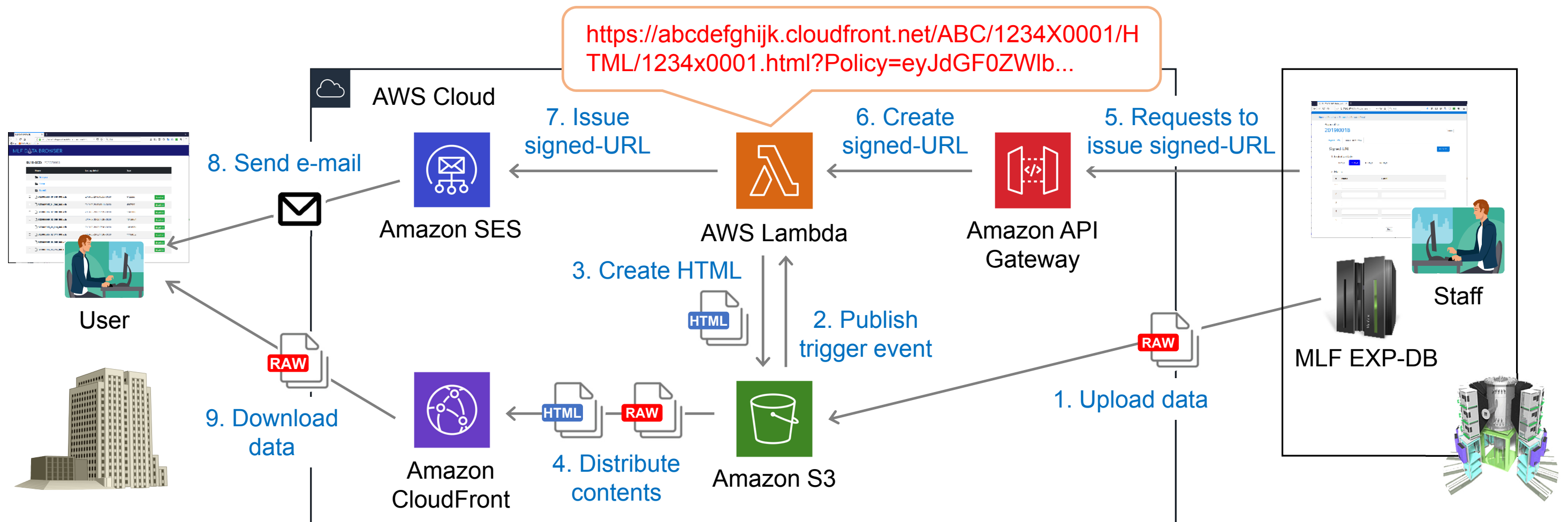
- Run No.
- Measurement status
- Status of DAQ and Device servers
- Status of each device:
 - physical parameters
 - time series graph
- Progress of automatic measurement

- Automatic updating



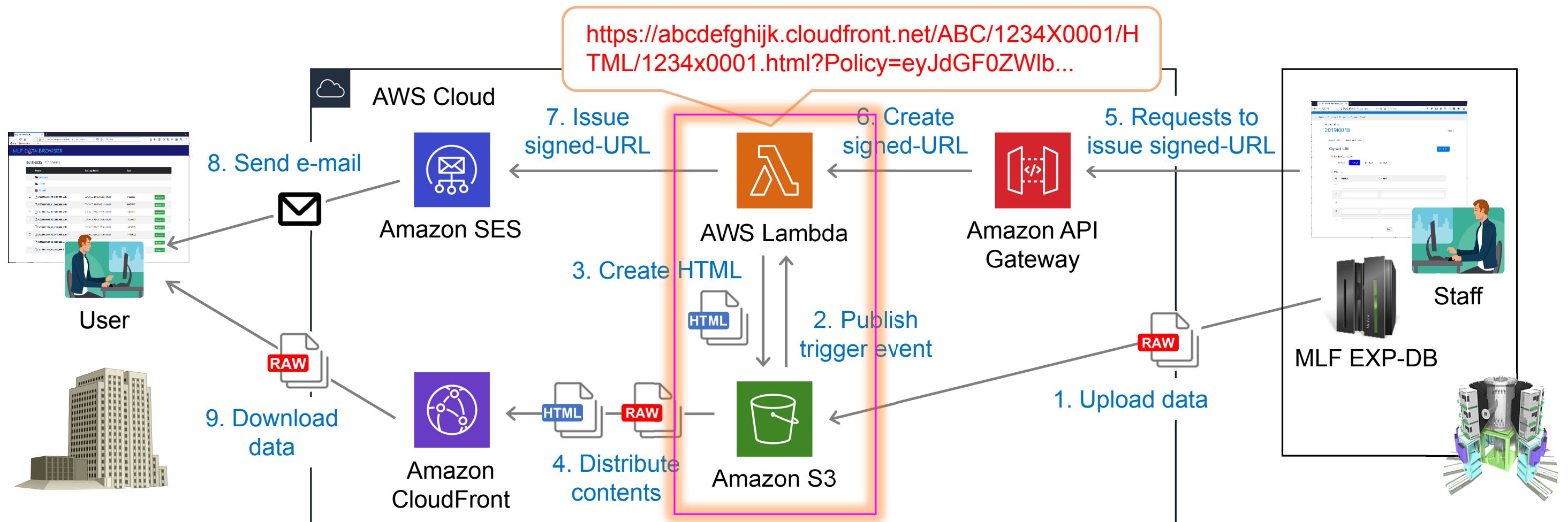
Service 2: Data distribution

- Adopting the same architecture as experiment monitoring.
 - Web distribution: S3 + CloudFront
 - Access control: signed URL
 - Issue of signed URL: Serverless computing
- Execution of Lambda function: via API Gateway
- Mail sending: SES
- HTML generation: Event-driven serverless computing



Service 2: Data distribution

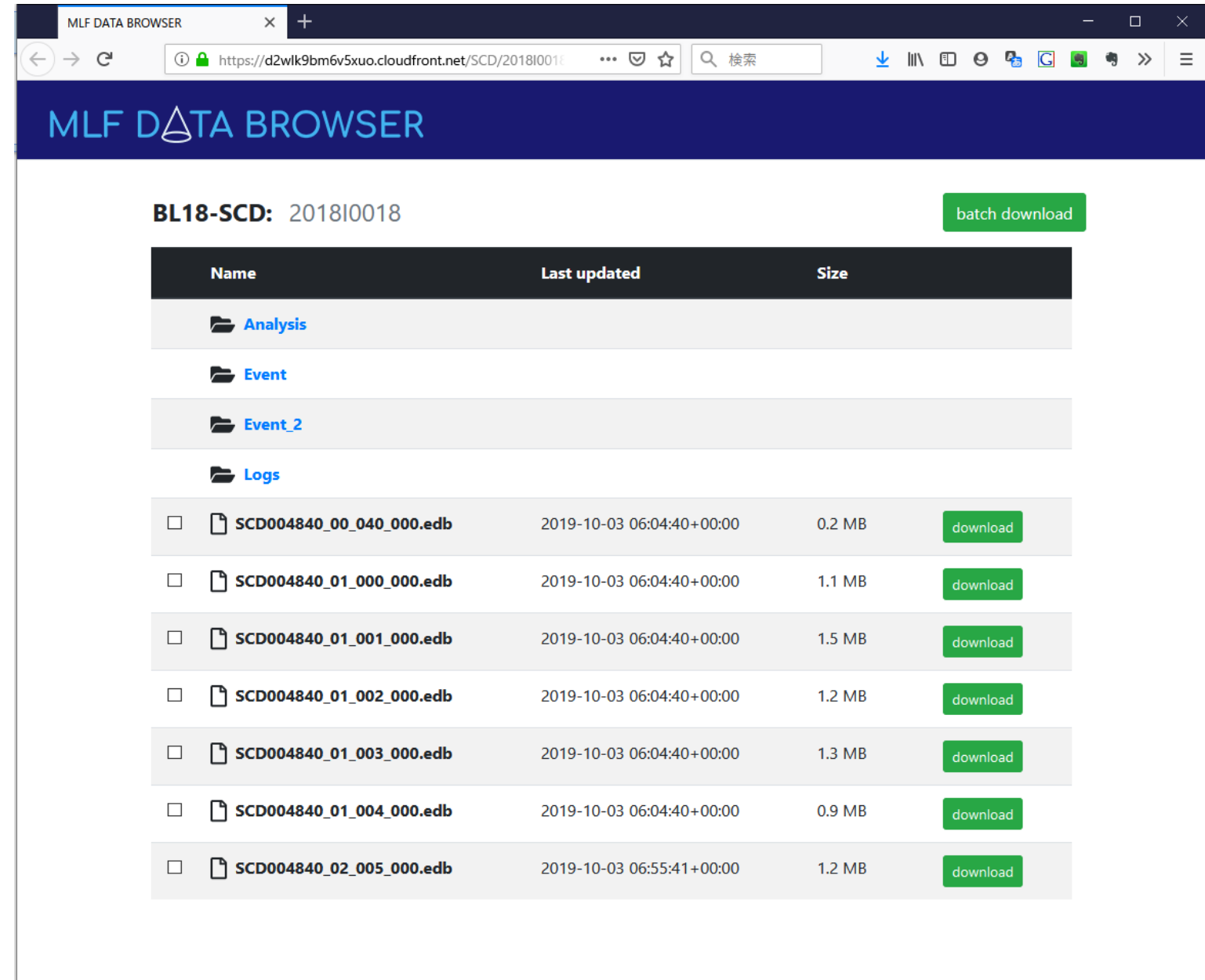
- Adopting the same architecture as experiment monitoring.
 - Web distribution: S3 + CloudFront
 - Access control: signed URL
 - Issue of signed URL: Serverless computing
 - Execution of Lambda function: via API Gateway
 - Mail sending: SES
- HTML generation: Event-driven serverless computing



Web interface for data distribution

MLF DATA BROWSER

- Data list generated by event-driven serverless computing.
- Simple web access and authentication with signed URL.
- High-speed download of large amounts of data via CloudFront (CDN).



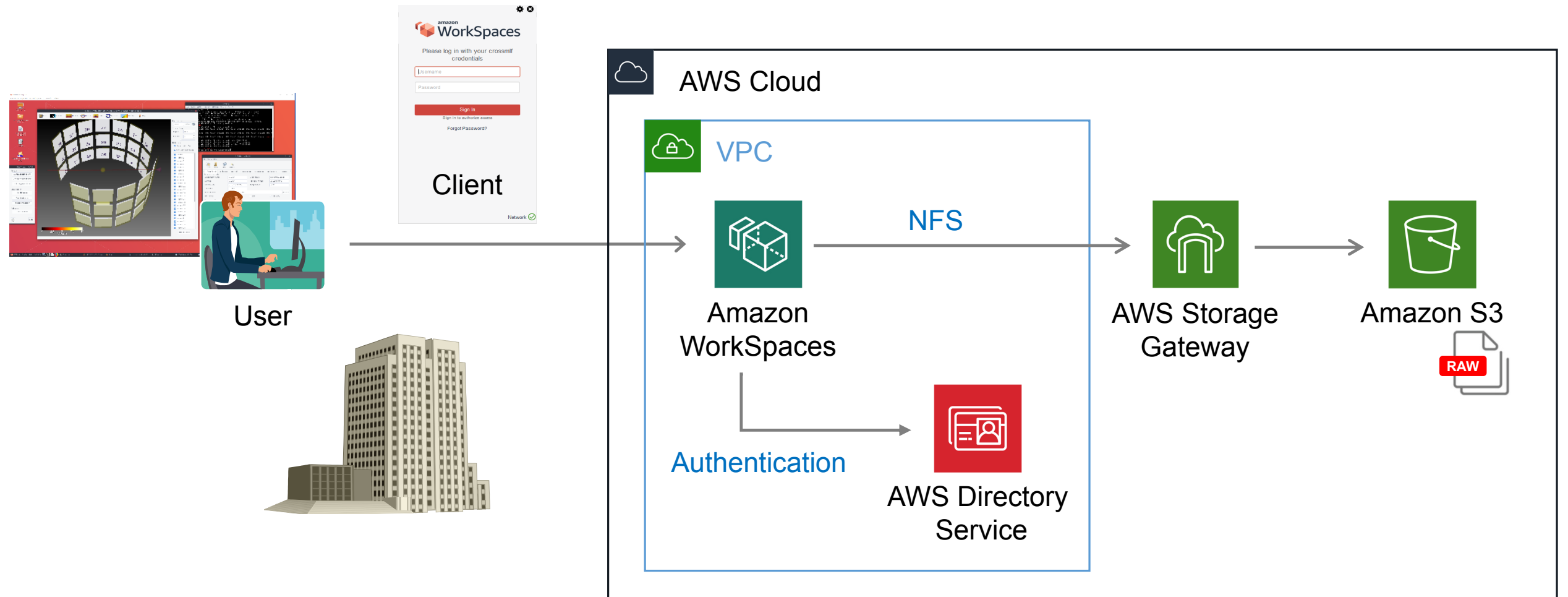
The screenshot displays the MLF DATA BROWSER web interface. The browser's address bar shows the URL: <https://d2wk9bm6v5xuo.cloudfront.net/SCD/2018I0018>. The page title is "MLF DATA BROWSER". The main content area shows the identifier "BL18-SCD: 2018I0018" and a "batch download" button. Below this is a table with columns for "Name", "Last updated", and "Size". The table lists several folders and files:

Name	Last updated	Size
Analysis		
Event		
Event_2		
Logs		
<input type="checkbox"/> SCD004840_00_040_000.edb	2019-10-03 06:04:40+00:00	0.2 MB
<input type="checkbox"/> SCD004840_01_000_000.edb	2019-10-03 06:04:40+00:00	1.1 MB
<input type="checkbox"/> SCD004840_01_001_000.edb	2019-10-03 06:04:40+00:00	1.5 MB
<input type="checkbox"/> SCD004840_01_002_000.edb	2019-10-03 06:04:40+00:00	1.2 MB
<input type="checkbox"/> SCD004840_01_003_000.edb	2019-10-03 06:04:40+00:00	1.3 MB
<input type="checkbox"/> SCD004840_01_004_000.edb	2019-10-03 06:04:40+00:00	0.9 MB
<input type="checkbox"/> SCD004840_02_005_000.edb	2019-10-03 06:55:41+00:00	1.2 MB

Each file row includes a checkbox on the left and a "download" button on the right.

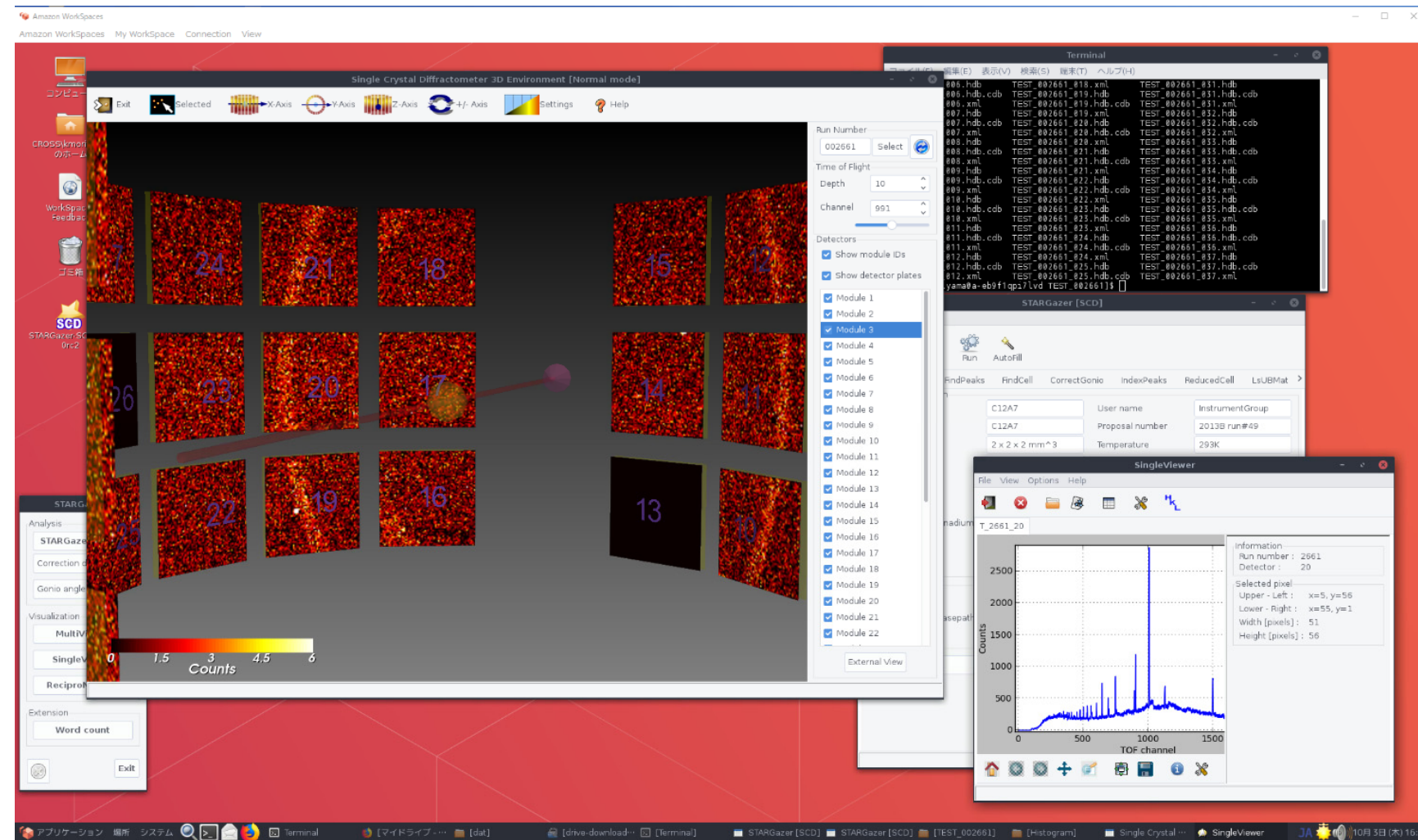
Service 3: Data analysis

- VDI (DaaS): WorkSpaces (OS: Amazon Linux 2)
- Access control: Simple AD (Directory Service)
- Data access: NFS (S3+ Storage Gateway)
- Client: WorkSpaces client (cross platform)



Remote desktop of WorkSpaces

- Analysis environment for a single crystal diffraction experiment (MLF BL18)
- Data analysis software: STARGazer
- Same environment can be used from anywhere.



Summary and future plans

- We have developed services for remote access using AWS cloud services.
 - Experimental monitoring
 - Data distribution
 - Data analysis
- We have adopted a microservice approach for system development.
- The developed services are planned to be introduced to beamlines gradually from FY2019.
- We have also plan to introduce other cloud services such as Google Cloud Platform and Microsoft Azure.



Google Cloud



Windows Azure