

# Integrity Data Hub (IDH) Overview Presentation

*July 26, 2018  
1:00-2:00 PM EST*



<http://integrity.naswa.org>

# Webinar Administration/Ground Rules

- Session will be recorded and posted
- We will respond to questions on a preliminary basis
  - Please hold question until the end – raise hand function
  - Official responses will be posted to questions received via email
- Questions should be submitted electronically to [DataHubRFP@naswa.org](mailto:DataHubRFP@naswa.org)
- All questions and official responses will be posted on the RFP webpage at:  
<http://www.itsc.org/Pages/datahubrfp.aspx>

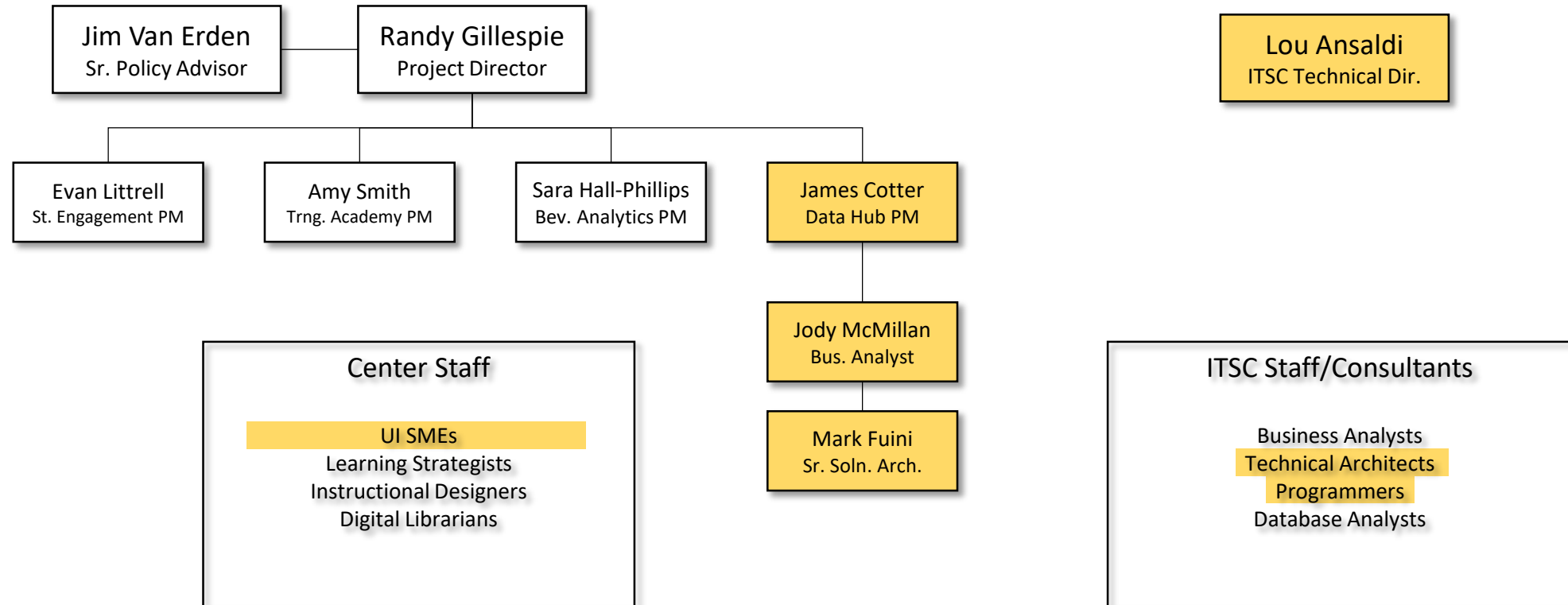
# Agenda

- To provide background information
  - UI Integrity Center (the Center)
  - Integrity Data Hub (IDH) project
- To review the IDH Request for Proposal
  - Purpose
  - Response requirements
  - Solicit feedback/questions
- To review the Data Hub technical architecture

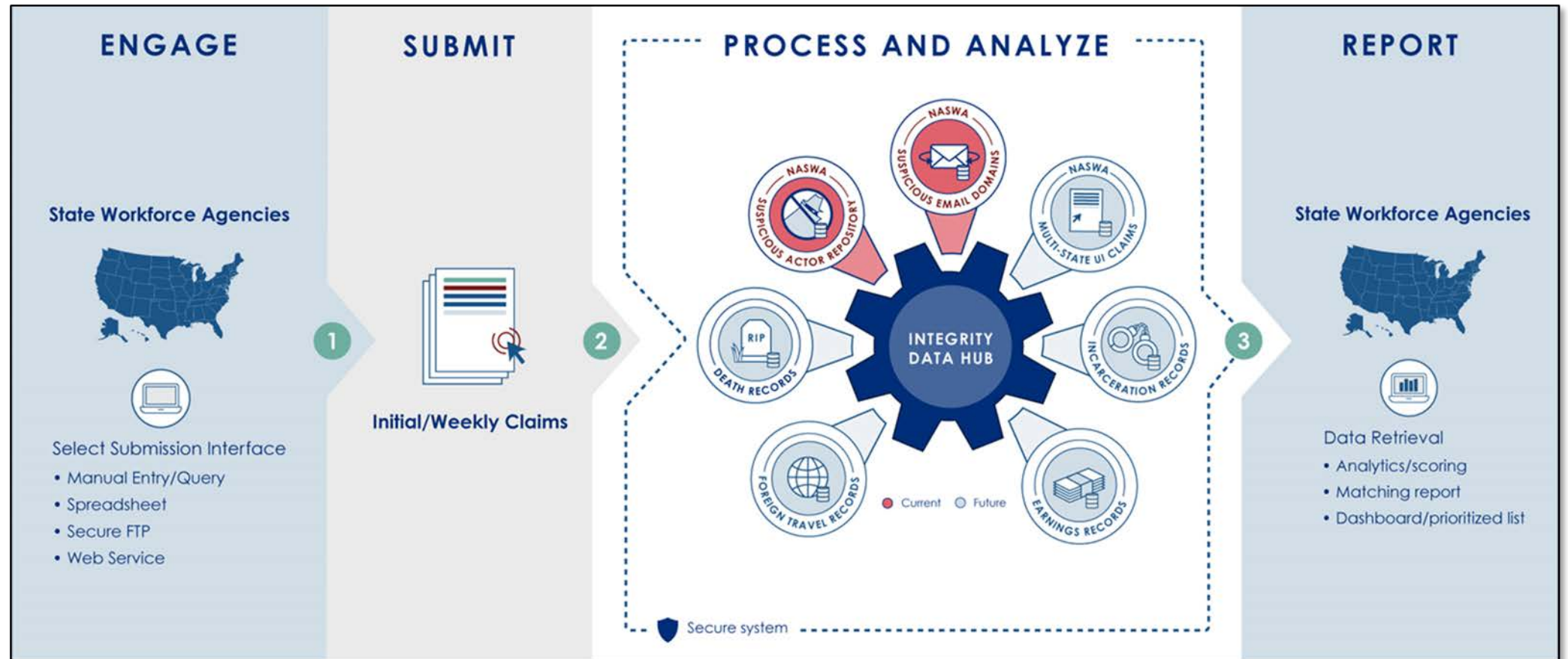
# UI Integrity Center

- In 2012, USDOL selected New York State Dept. of Labor (NYSDOL) to lead the development of the UI Integrity Center
- NYSDOL subsequently passed leadership responsibilities for the Center to NASWA/CESER
- The Center is charged with developing:
  - “innovative UI program strategies to reduce improper payments, prevent and detect fraud and recover any improper payments made” (UIPL 28-12).
- Additional information about the Center
  - <http://naswa.org/integrity/>

# NASWA Data Hub Project Team

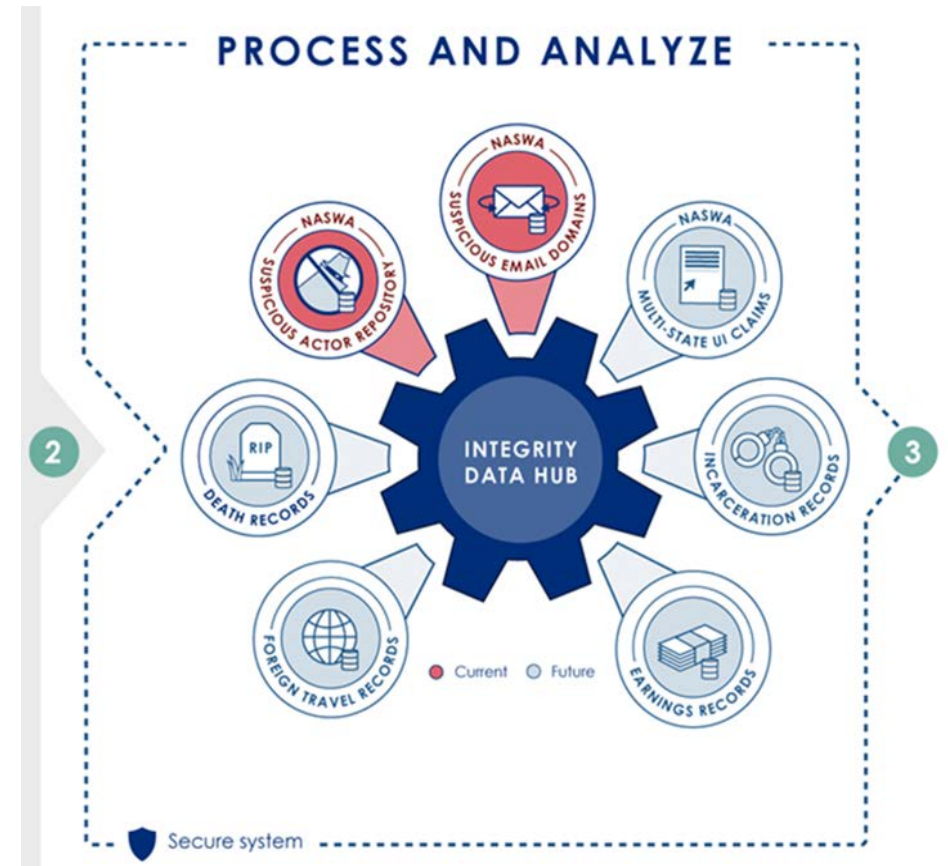


# Integrity Data Hub (IDH)



# Expanding the Data Hub

- Future enhancements
  - Filtering function
  - Partial IP addresses
- Future capabilities
  - Suspicious green-dot card accounts
  - Fraud alerting
  - Multi-state database
- Participation
  - Engage all State Workforce Agencies
  - Expanded use benefits all participants

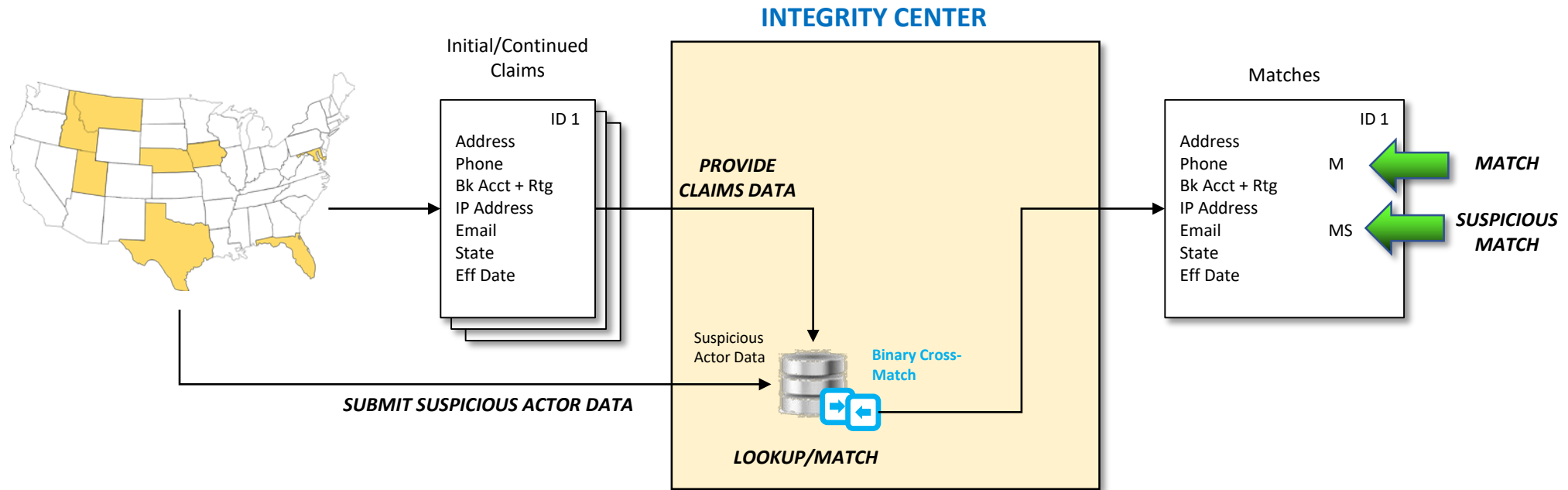


# Background

- Many states collecting information on suspicious actors
  - IP addresses
  - Email addresses/domains
- Emergent need to focus on prevalence of ID theft
- Build a repository to allow states to submit and retrieve information on suspicious actors
  - Allow participation of states with varying levels of technology and volume
  - Minimize potential issues with data sharing
  - Develop a foundation on which to achieve larger IDH vision



# Data Hub Operation

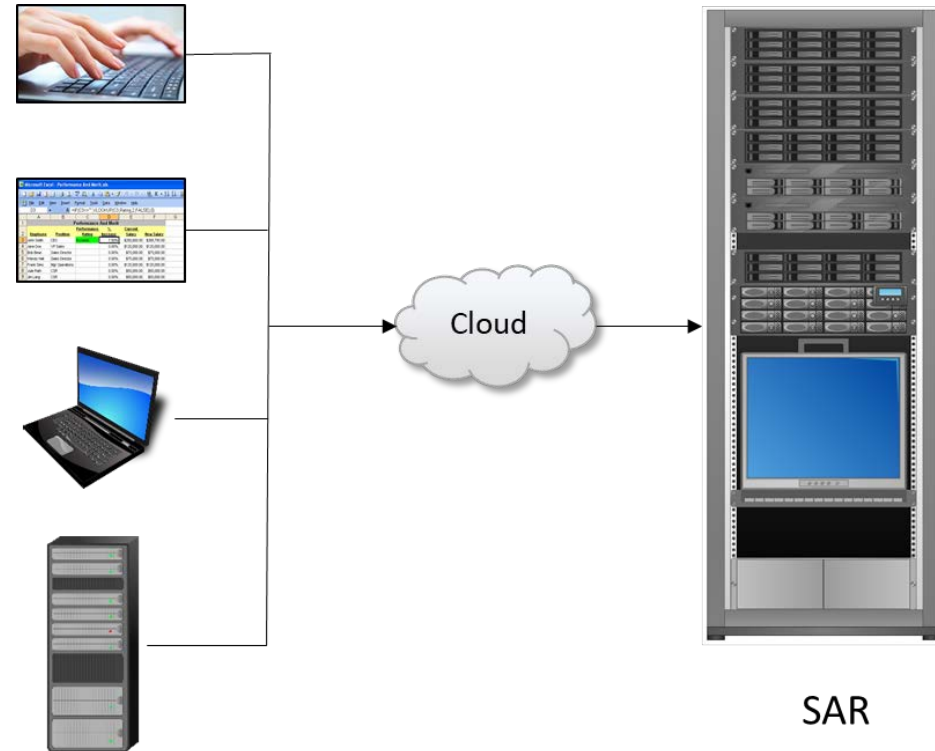


1. Pilot states submit information on suspicious actors
2. Pilot states provide similar information initial/weekly claims
3. System compares claims data fields to suspicious actors to identify matches and source
4. System highlights matches in report/output
5. System deletes states' claims data after matching



# Submit/Lookup Processes

- SAR accommodates different methods for collecting data
  - Manual
    - Data entry
    - Upload
  - System-to-System
    - Secure FTP (sFTP)
    - Web service



# Data Hub Key Functionality

- Submit
- Lookup
- Modify
- Reporting
  - Administrative
  - Analytical
- Administration
  - User access/roles

Home Report Lookup Search Submit Modify Admin

SAR File Upload Templates

- SAR Lookup File Upload Template
- SAR Submit File Upload Template

Submit Suspicious Actor

State: Arizona Unique ID: 000000000000004

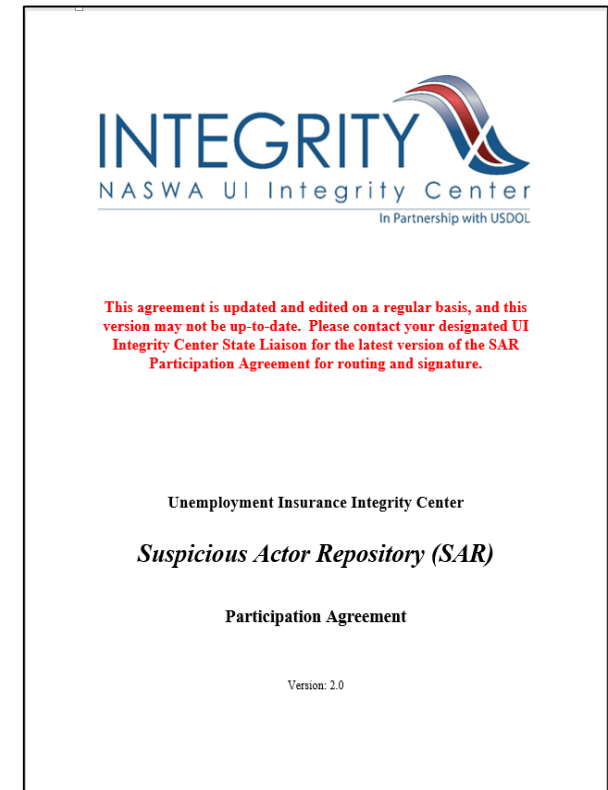
Effective Date \* Occurrence Date \*

Suspicious Data Elements	Suspicious Element?
IP Address: #####/0xfh::0xfh::0xfh::0xfh::0xfh:	No
Email: xxxxxx@xxxxxxx.xxx	No
Address 1: Street Address 1, Street Address 2, City, Select State, Zip	No
Address 2: Street Address 2 1, Street Address 2 2, City 2, Select State, Zip 2	No
Phone 1: (999) 999-9999	No
Phone 2: (999) 999-9999	No
Phone 3: (999) 999-9999	No
Direct Deposit Routing Number: 999999999	No
Direct Deposit Account Number: Bank Account Number	No

Review Cancel

# Data Hub Documentation

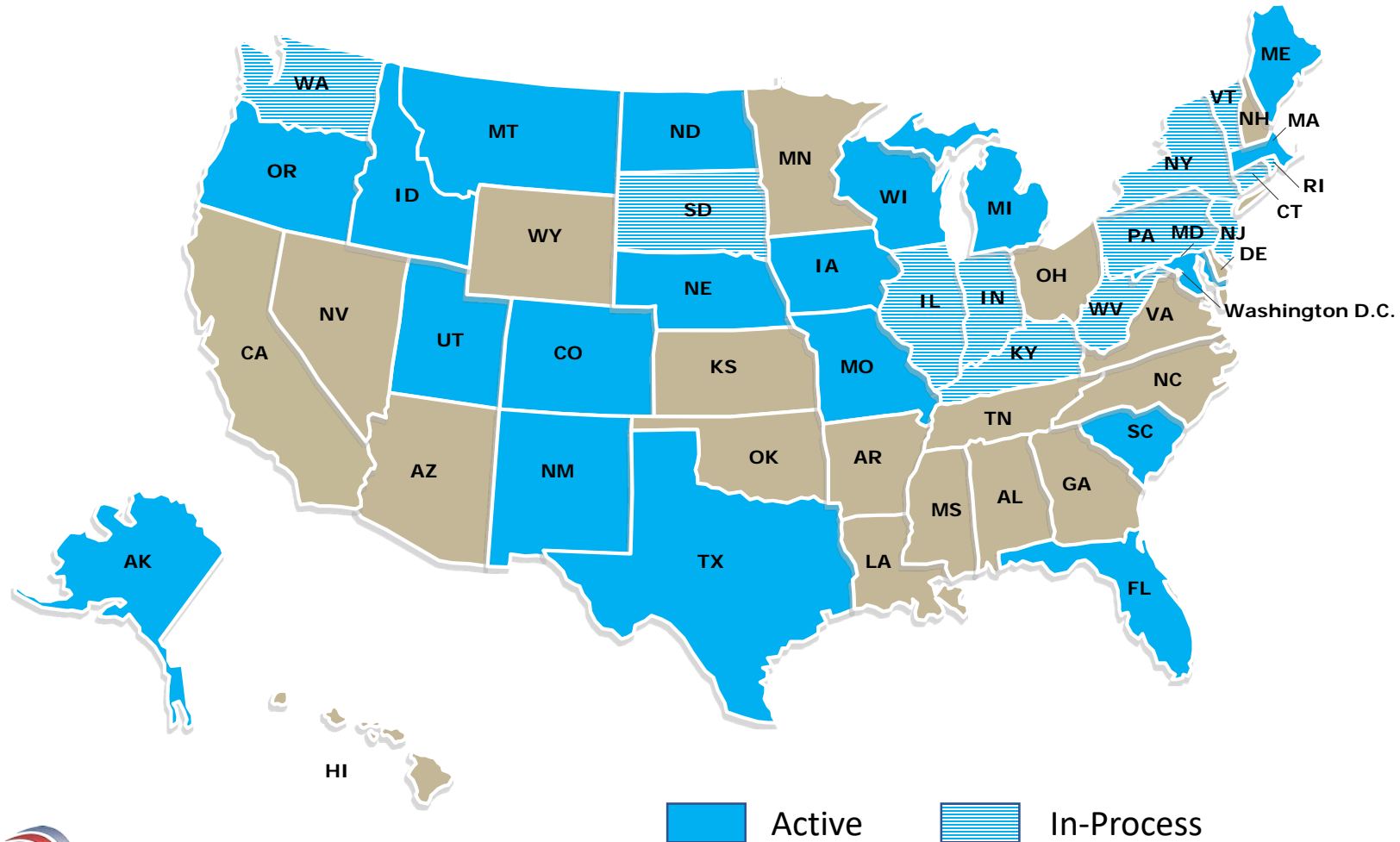
- Participation agreement
  - Not to be used to auto-block claims
  - Specify SAR state administrator
- Fact sheet
- User guide
- Technical info and products
  - SAR FTP and web service integration guides
  - Clients
    - Java client/.NET client
- Data transmission and security summary



# Data Hub Status

- Data Hub Phase 1 is complete
  - SAR launched Fall 2017
    - SAR data has been loaded into database (over 13,000 suspicious actors)
    - Working with ~26 states to integrate in UI systems and processes
- Data Hub Phase 2 initiatives
  - SAR enhancements
  - Ongoing security assessment/testing
  - SAR application and database monitoring/optimization
  - Expanded capabilities

# Data Hub Participation



# IDH RFP Purpose

- To identifying industry partners to augment the Center project team in the ongoing development of Data Hub Phase 2 capabilities
- The Center is seeking assistance in the following areas:
  - Project Strategy/Management
  - Software Engineering/Development
  - IT Security
  - System Monitoring/Administration
  - Requirements Development
  - Data Architecture/Management
  - Database Architecture/Management
  - ETL Management
  - Program Analysis
  - System Architecture

# RFP Structure/Process

Strategic Support



1 Award

Dev. Env. Support



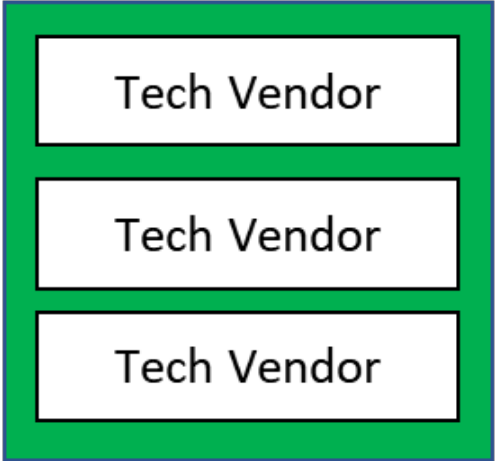
1 Award

Rqmts. Mgmt. Support



1 Award

Development Support



Multiple Award



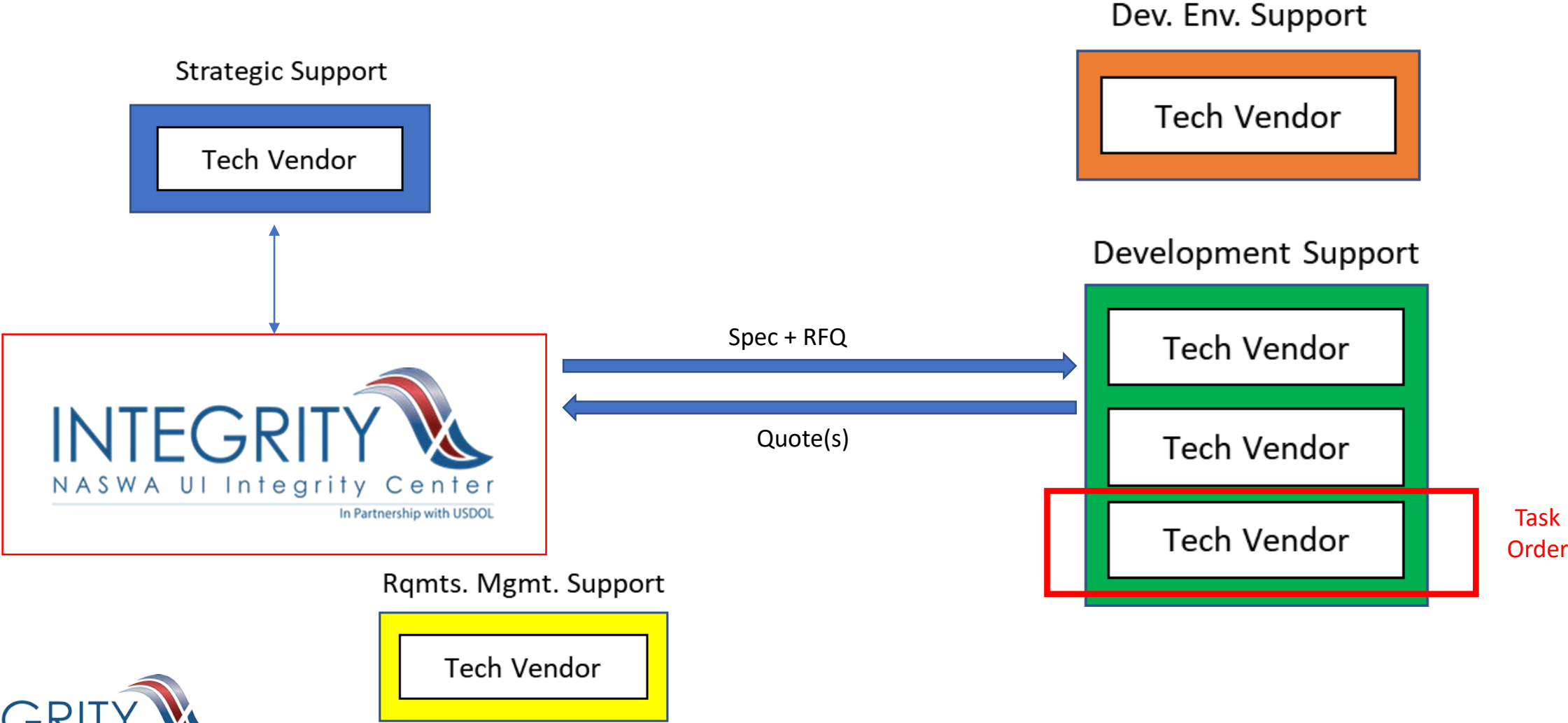
# Phase 2 RFP Elements

- Strategic Support
  - Weekly project status meetings
  - Periodic virtual/on-site meetings and working sessions
  - Quarterly strategic planning sessions
  - Preparing, reviewing and maintaining project plans and documents
- Development Environment
  - Assessment
  - Implementation Support
    - Tools (emphasis on open-source)
    - Processes and Methods
    - Standards
    - Project Governance

# Phase 2 RFP Elements

- Business Analysis/Requirements Management
  - Requirements gathering sessions
  - Interviewing state end users and other stakeholders
  - Use cases and design documentation
  - Test plans and procedures
  - Planning and coordinating testing
- Software Development
  - Software engineering/development
  - System architecture review/assessment
  - Data architecture/management
  - IT and data security
  - System monitoring/administration
  - Database architecture/management
  - Extract/Transform/Load (ETL) management

# RFP Structure/Process



# RFP Timeline

Project Activity	Timeline
Data Hub RFP Webinar	26 Jul 18
Final Clarification Questions	2 Aug 18
Questions and Responses Posted	9 Aug 18
Proposals Due	7 Sep 18
Offeror Presentations	Week of 17 Sep 18
Best and Final Offer Pricing (optional)	28 Sep 18
Award (anticipated)	12 Oct 18

# RFP Response (1 response per organization)

- Company Overview – 1 page
  - Brief description of the company, products/services, size and Point of Contact (PoC) for communication
- Project Summary Citations – Up to 3 citations – 3 pages total each citation
  - Experience in providing strategic/technical support for projects of similar content, size and scope to the Data Hub
  - Include: summary, size/scope, initial/final budget, agency, agency PoC

# RFP Response (per SOW area)

- Technical/Management Approach – 10 pages per SOW area
  - How will personnel from your organization be selected to provide selected services?
  - How will you organize and manage providing strategic services?
  - How will you ensure availability of key staff?
  - What deliverables will you develop and/or maintain?
- Key Personnel Resumes – 3 resumes, 2 pages, per SOW area
  - Provide three resumes (two pages maximum per resume) for key personnel to be assigned to the project for each SOW area
  - The same resume may be provided for multiple SOW areas
  - Include: name, proposed labor category, percentage of time allocated to the Data Hub project, and relevant work experience

# RFP Response

- Cost
  - Complete cost template sections as appropriate

**Part A: Program Strategic Services – Firm Fixed-Price**

Provide annual firm-fixed pricing for up to 30 hours per month for program strategic services.

Base Year	Option Yr. 1	Option Yr. 2	Option Yr. 3	Total

# RFP Response

## **Part B: Program Strategic Services – Firm Fixed-Price/Time & Materials**

Provide a firm fixed-price for conducting a comprehensive assessment and preparation of a detailed report providing recommendations for establishing an independent, stand-alone distributed development environment within the Center. Also, please provide T&M pricing for the labor categories listed to support implementation.

### Assessment Report and Recommendations

<b>Deliverable</b>	<b>Price</b>
Comprehensive assessment report and recommendations	

### Implementation Support

<b>Labor Category</b>	<b>Education &amp; Experience</b>	<b>Base Yr. Labor Rate</b>	<b>Option Yr. 1 Labor Rate</b>	<b>Option Yr. 2 Labor Rate</b>	<b>Option Yr. 3 Labor Rate</b>
Sr. Implementation Specialist	MS + 5				
Implementation Specialist	BS + 5, MS				
Jr. Implementation Specialist	BS + 2				



# RFP Response

**Part C: Requirements Development/Business Analysis Services – Firm Fixed-Price/Time & Materials**

Provide T&M hourly rates for the following labor categories for base period and all option periods. Please use a 3% escalation.

Labor Category	Education & Experience	Base Yr. Labor Rate	Option Yr. 1 Labor Rate	Option Yr. 2 Labor Rate	Option Yr. 3 Labor Rate
Sr. Business Analyst	MS + 5				
Business Analyst	BS + 5, MS				
Jr. Business Analyst	BS + 2				

# RFP Response

**Part D: Technical Support – Firm Fixed-Price/Time & Materials**

Provide T&M hourly rates for the following labor categories for base period and all option periods.

Labor Category	Education & Experience	Base Year Labor Rate	Option Yr. 1 Labor Rate	Option Yr. 2 Labor Rate	Option Yr. 3 Labor Rate
Sr. Project Manager	BS +10, MS +5				
Project Manager	BS +5				
Sr. Software Developer	MS +5				
Software Developer	BS +5, MS				
Jr. Software Developer	BS +2				
Cybersecurity Analyst	BS + 8				
Sr. Database Administrator	MS +5				
Database Administrator	BS +5, MS				
Jr. Database Administrator	BS +2				
Sr. Solution Architect	MS +5				
Solution Architect	BS +5, MS				
Jr. Solution Architect	BS +2				
Data Architect	BS +10, MS +5				
Sr. Data Scientist	BS +10, MS +5				
Data Scientist	MS +5				
ETL Specialist	MS +3				
Sr. System Administrator	MS +5				
System Administrator	BS +5, MS				
Jr. System Administrator	BS +2				

# RFP Administrative Items and Guidance

- Confidentiality
  - Mark all confidential information as such
- Format
  - MS Word or Adobe Acrobat
  - 8.5" x 11", 12-point font, ½ inch margins, page numbers
- Questions
  - Submit to: [DataHubRFP@naswa.org](mailto:DataHubRFP@naswa.org)
  - Please put "RFP QUESTION" in the subject line
- RFP Responses
  - Submit to: [DataHubRFP@naswa.org](mailto:DataHubRFP@naswa.org)
  - Due: 5:00pm ET September 7, 2018

# RFP Administrative Items and Guidance

- RFP Information
  - <http://www.itsc.org/pages/DataHubRFP.aspx>
  - Webinar recording
  - Webinar presentation
  - Questions and responses
  - Updates (as needed)

**ITSC** DataHub RFP ⓘ

ITSC Public Site | About ITSC | ITSC Products | UI IT Modernization | UI Learning Center | Workforce Connect | UI SIDES | NASWA | UI Integrity

### Members-Only Section

**Members Login**  
To sign up for access to the Members-Only site:  
Contact the ITSC Webmaster

### About ITSC

Learn About the ITSC  
Contact Us

### Vendor Opportunities

Request for Proposals (RFPs)  
Request for Information (RFIs)  
UI 101 Training for Vendors

### Why Sign Up for ITSC?

State UI Members can...

- SHARE UI information technology information with the UI community.
- LEARN and train on a wide range of UI program and technical topics through online and in-person courses.
- RECEIVE updates on such topics as RFP opportunities, upcoming events, and new product information.
- PARTICIPATE in collaborative projects online and view project portals for multi-state projects.
- DOWNLOAD source code, documentation and maintenance support agreements for products such as SDDS and OccuCoder.

## Request for Proposal: Program and Technical Support for Ongoing Development of the UI Integrity Data Hub

The UI Integrity Center (Center) is interested in identifying potential industry partners to augment the Center project team in the ongoing development of Data Hub Phase 2 capabilities. Interested parties are asked to provide information on their personnel, corporate experience, and related labor category and pricing information. The Center is seeking assistance in the following areas:

- Project Strategy/Management
- Software Engineering/Development
- IT Security
- System Monitoring/Administration
- Requirements Development
- Data Architecture/Management
- Database Architecture/Management
- ETL Management
- Program Analysis
- System Architecture

[RFP Data Hub Program and Technical Support \(.pdf\)](#)  
[RFP Data Hub Program and Technical Support \(.doc\)](#)

### UI Integrity Data Hub Webinar

July 26, 2018  
1:00 pm – 2:00 pm EST

Join: <https://naswa.zoom.us/j/833507745>  
Meeting ID: 833 507 745

Telephone:  
+1 669 900 6833 (US Toll) (preferred)  
+1 646 558 8656 (US Toll)

[Recording \(link\)](#)  
[Presentation \(.pdf\) \(link\)](#)  
[Questions \(.pdf\) \(link\)](#)

The purpose of the Webinar is to provide information on the Center, review the RFP, provide information on support requirements, and to document and respond to questions.

Important Dates	
Project Activity	Timeline

# Integrity Data Hub (IDH)

## Applications, Architecture & Technology Stack



# IDH Applications

- SAR Portal
  - Lookup & Submit SAR Records
  - User Administration
  - Reporting
- SAR Web Services
  - Real-time
  - Lookup and Submit SAR Records
- SAR FTP Processing
  - Batch
  - Lookup and Submit SAR Records
- SSO Portal

**SAR** **INTEGRITY** **Suspicious Actor Repository**  
NASWA UI Integrity Center

Home Report NASWA **Lookup** Search Submit Modify

**Lookup Suspicious Actor**

IP Address: #####-###-#### / 0xfh::0xfh::0xfh::0xfh::

Email: xxxxxx@xxxxxxxx.xxx

Address 1: Street Address 1, Street Address 2, City, Select State, Zip

Address 2: Street Address 2 1, Street Address 2 2, City 2, Select State, Zip 2

Phone 1: (999) 999-9999

Phone 2: (999) 999-9999

Phone 3: (999) 999-9999

Direct Deposit Routing Number: 999999999

Direct Deposit Account Number: Bank Account Number

Lookup Upload File

**Submit** - putting suspicious actor data into the system.  
**Lookup** - searching suspicious actor data.

# IDH Applications



- Java Programming Language
  - Java Server Faces (JSF)
  - Spring Framework
  - CXF Web Services
  - Web Service Clients

# IDH Architecture Key Features



- Open Source
- AWS Cloud Based Environment
  - Flexibility, Availability, Scalability
  - Redundancy - separate servers in different availability zones
  - Platform tools allow small team to manage many servers
- Security – Developed with NIST Based Best Practices
- Interface Channel Flexibility
  - State Integration SFTP, Lookup File, Web Services



# Open Source Ecosystem



**cassandra**



**FORGEROCK**

Linux



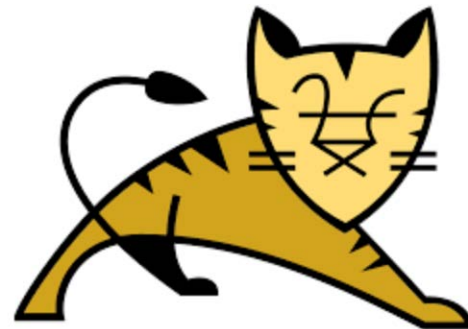
**APACHE**

HTTP SERVER PROJECT

DATASTAX

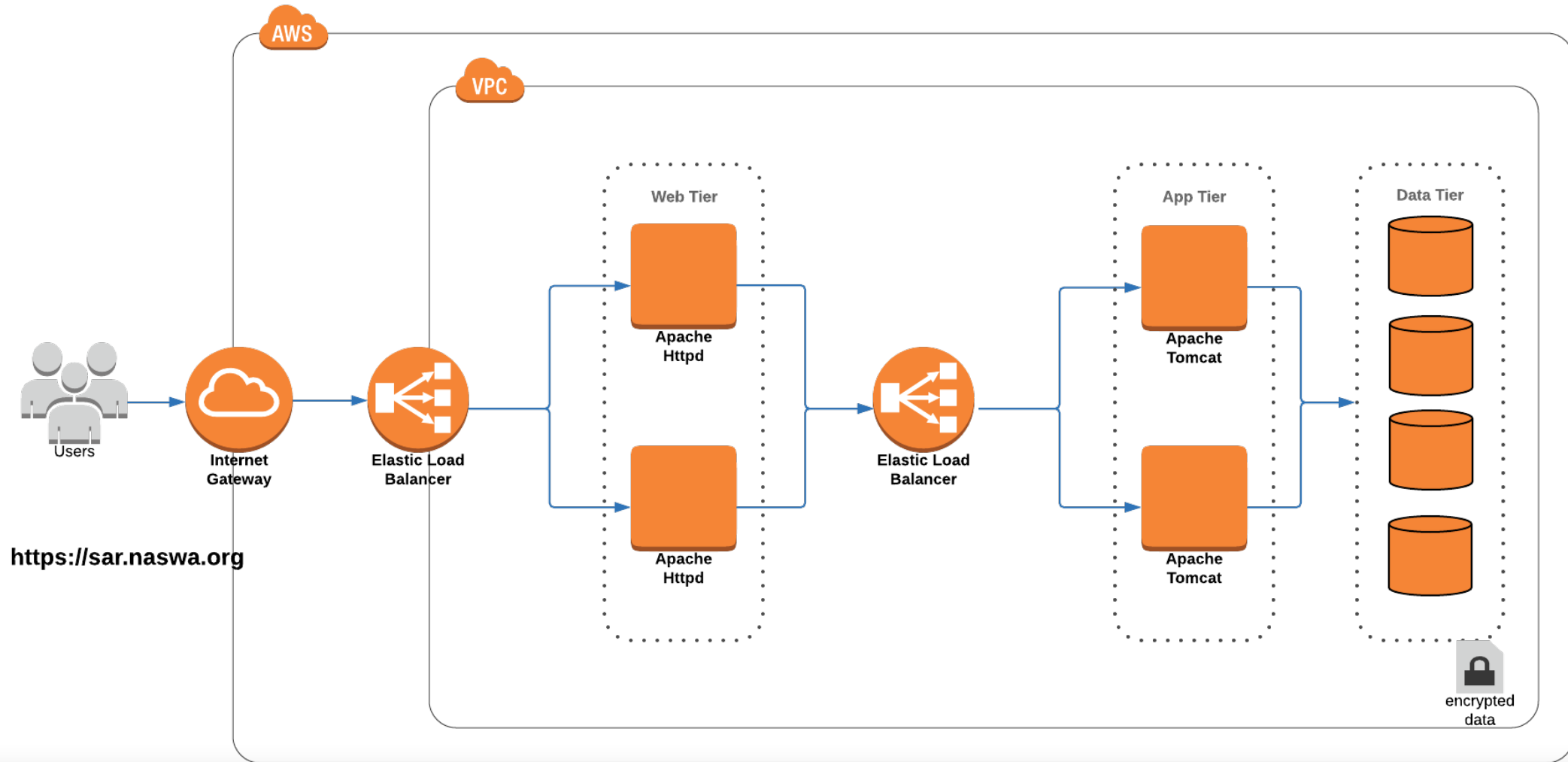


**OPENAM**



**OPENDJ**

# IDH AWS Cloud Environment



# Apache Httpd – Web Server

- Handles Web Requests from the Internet
- SAR Portal Application
  - <https://sar.naswa.org>
    - Requests on urls protected by OpenAM policy agent
    - Validates if the OpenAM security token exists with OpenAM and is valid
- SAR SSO Application
  - <https://sarsso.naswa.org>
- Forward To Internal Load Balancers for Application Servers

# Apache Tomcat – Application

- Custom applications written in Java perform application level processing
  1. SAR Portal
  2. SAR SSO Login
  3. SAR Web Service
  4. SAR FTP Service
- Open AM - SSO software runs on Tomcat

# Apache Cassandra – Database

- NOSQL Database
- DataStax Community Edition 3.0
- Distributed Key-Value Store, Low Latency, Lookup By Key Use Cases
- 4-Node Cluster (m4.2xlarge 8 vCPU 32GB RAM)
  - High Availability - Cluster can still function if one node is not performing
  - 2TB disk space per node (SSD, gp2)
  - Currently 184 MB of data allocation per node
- Predictable Linear Scalability with Added Nodes

# Open AM - Single Sign On



- Open Source Configurable Application By Forge Rock
- Administration Console for Single Sign On (Tomcat)
- Configuration Store is Configured to Use Open DJ
- APIs Validate Users and Manage Passwords
- Policy Agent API for Request Validation



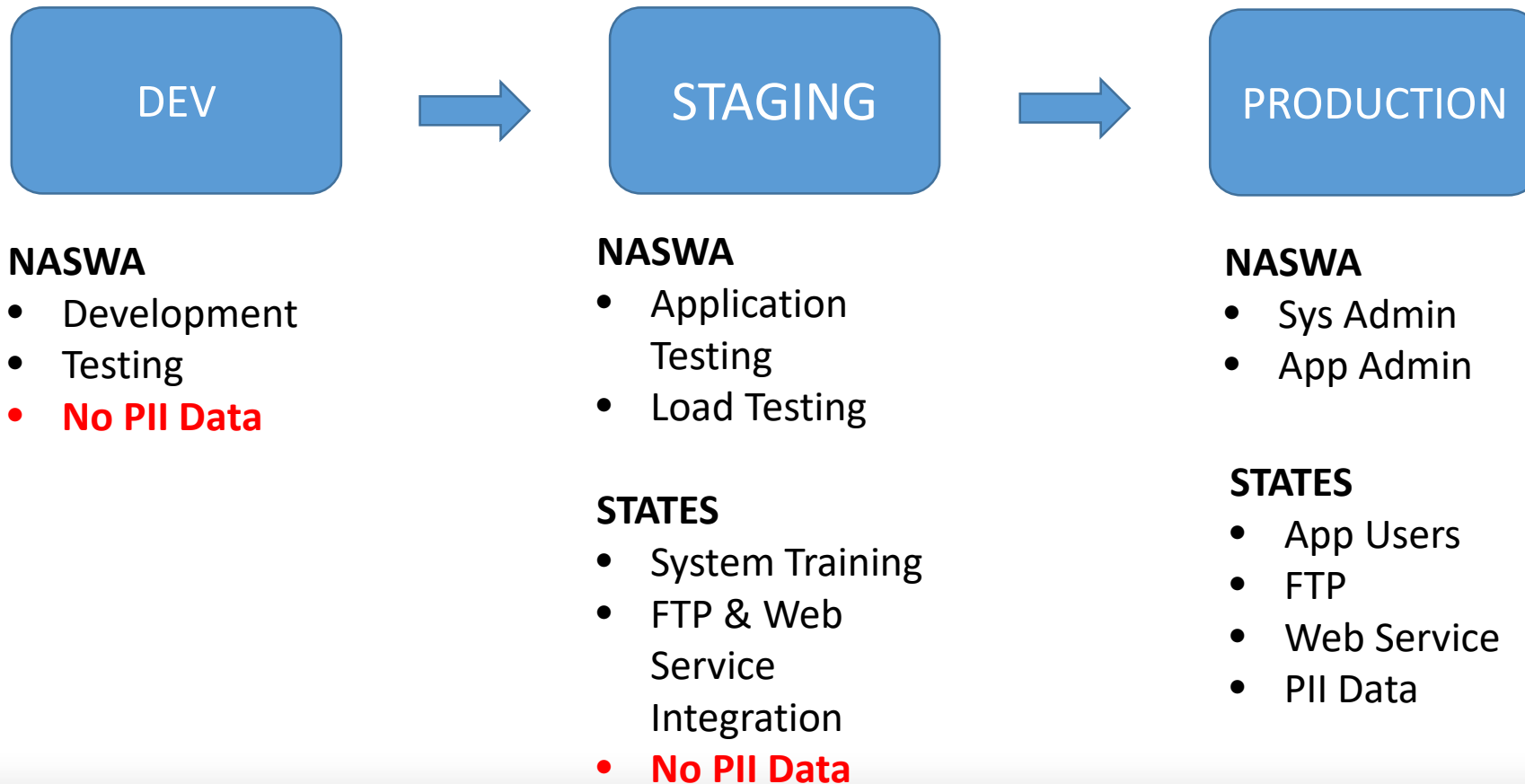
- LDAP Server for Open AM to store user configuration  
dedicated replicated Open DJ LDAP Servers
- User Store Replicated

# IDH AWS Cloud Environment

- Production
  - 16 Servers (Web, App, DB, LDAP)
- Staging
  - Same Topology as Production
  - Facilitates Load Testing
  - Available during core hours 7AM-7PM EST\*
- Development
  - SAR server and SAR SSO server

# IDH AWS Cloud Environment

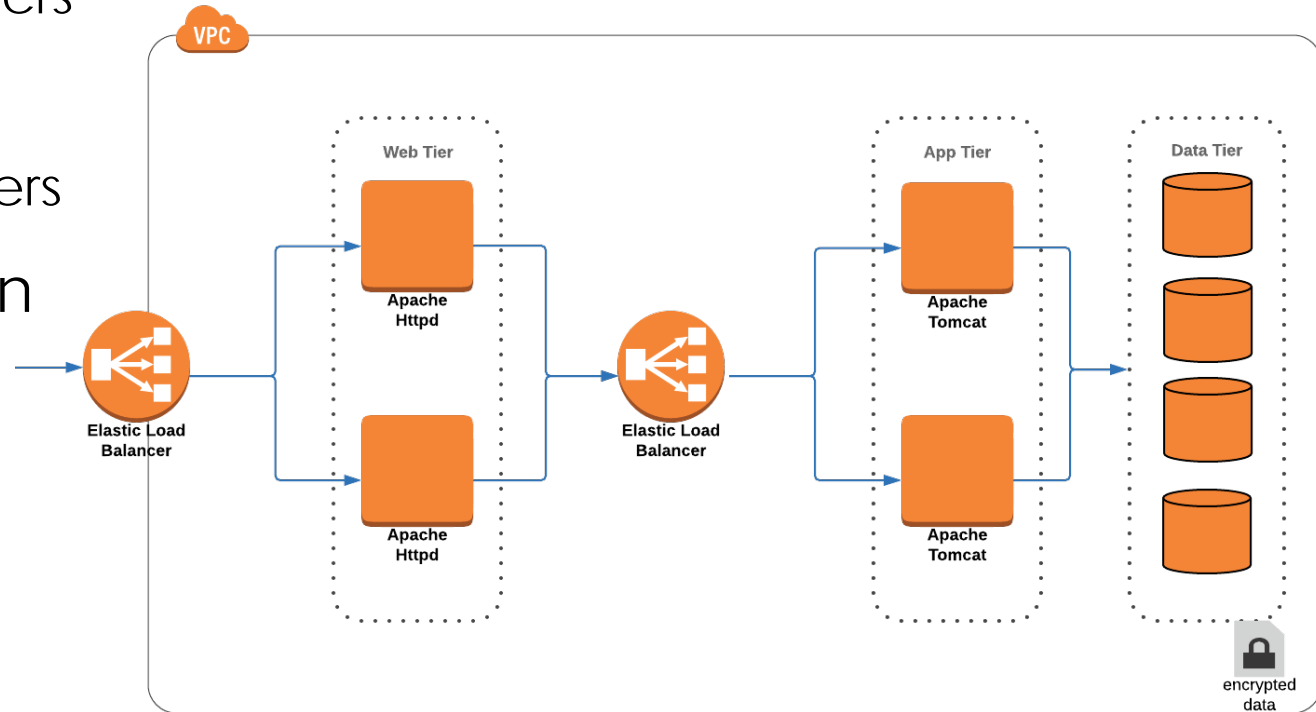
- Activities By Environment





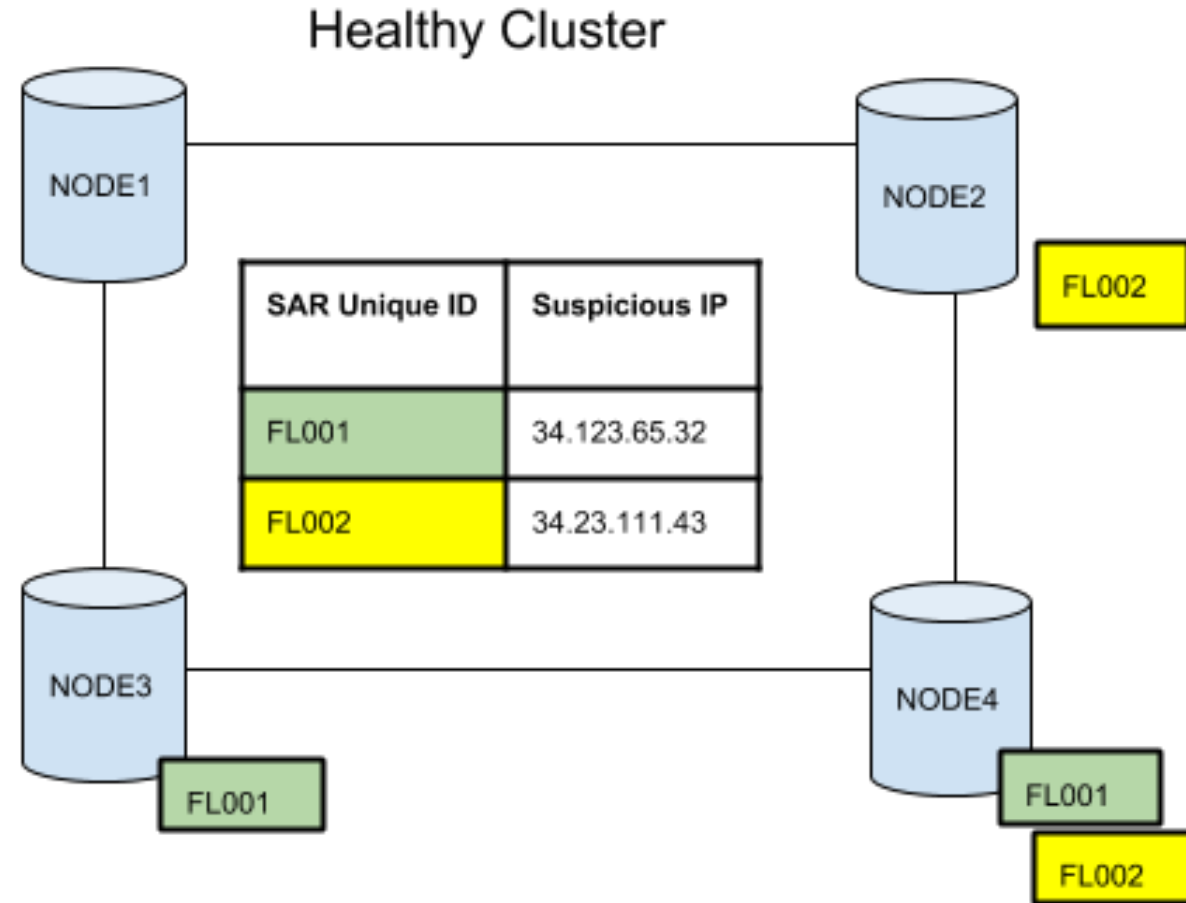
# IDH Redundancy – Web Tier

- Load Balancing
  - SSO Web Servers
  - SSO Login Application Servers
  - SAR Web Servers
  - SAR Core Application Servers
- LDAP Servers - Replication



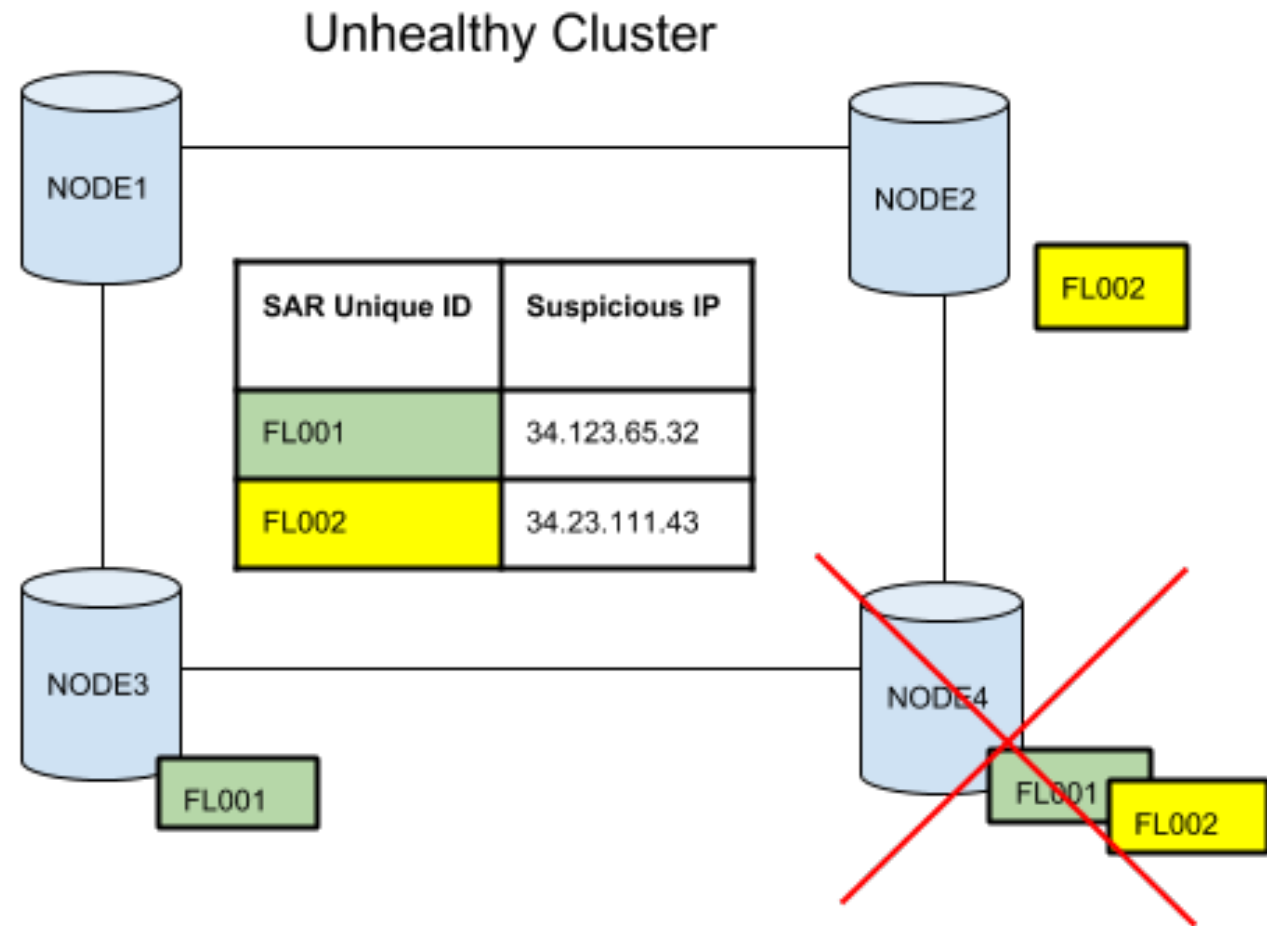
# IDH Redundancy – Data Tier

- Cassandra cluster
- Replication factor of 2
- Each piece of data is automatically replicated



# IDH Redundancy – Data Tier

- One Node Is Down
- Data Still Available
- Tradeoff for Disk Space
- Replication Factor of 3 Common to Increase probability of Availability
- Currently in one Region but across multiple availability zones

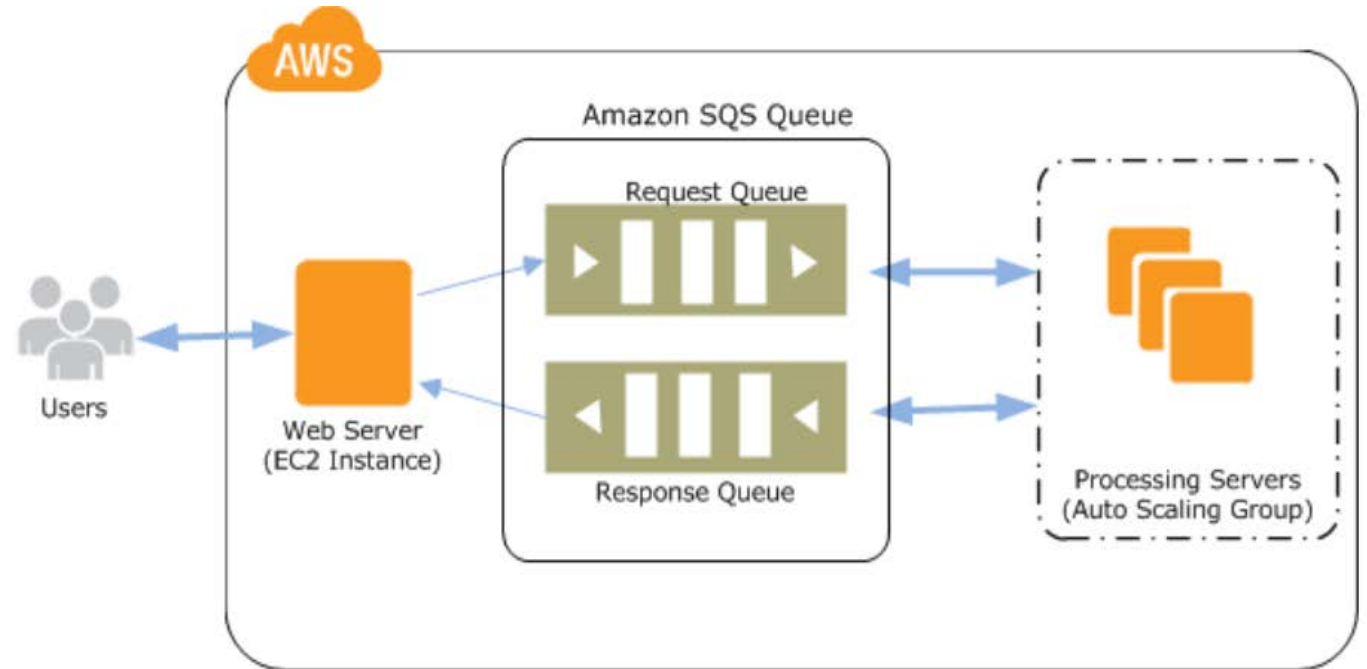


# IDH Scalability - Application

- Horizontal Scaling
  - Architectural foundation for expandability with load balancing
    - Web servers
    - Application servers
    - Web services
  - FTP processors can also be added
- Vertical Scaling
  - Increase processing power and capability of servers

# IDH Scalability - Auto Scaling

- UI demand changes with changes in economics
- Nationwide peak times will vary
- Monitor server usage and automatically deploy more servers to handle the load within minutes
- Servers can then automatically be shutdown when demand is lower saving costs which saves capacity
- Scheduled or On Demand



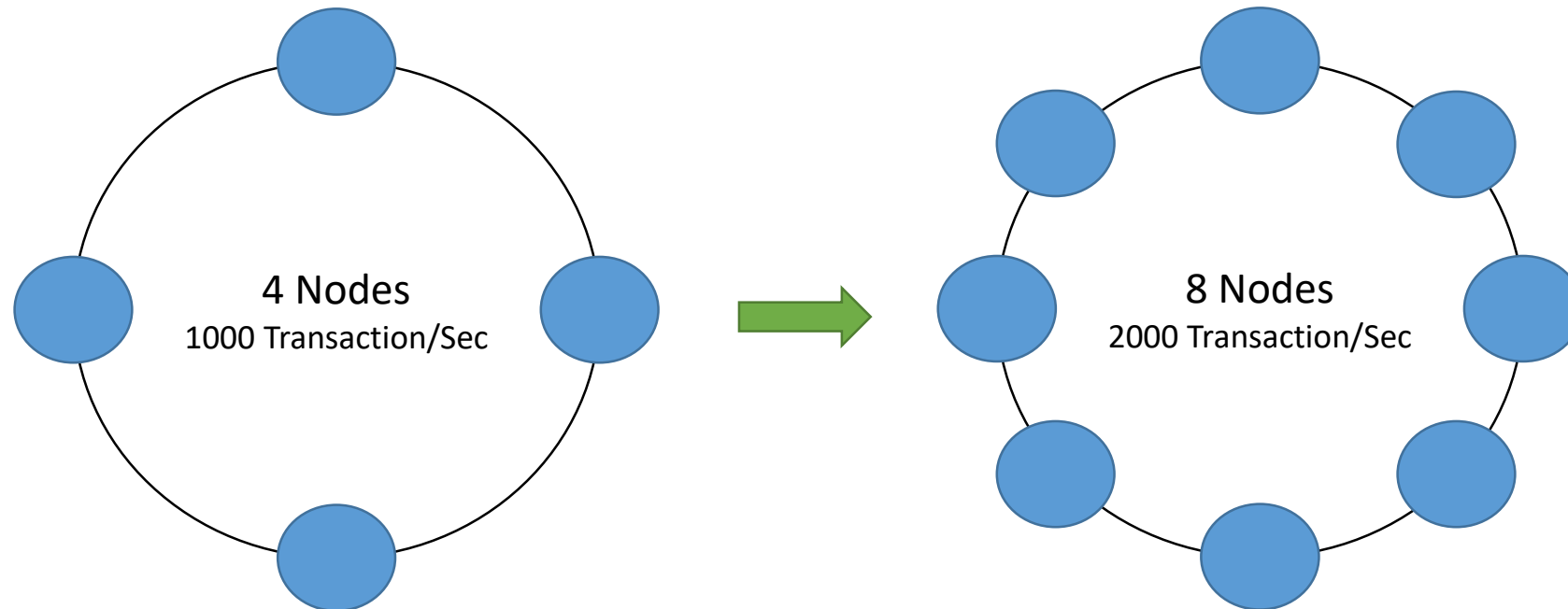
Reference:

<https://docs.aws.amazon.com/autoscaling/ec2/userguide/as-using-sqs-queue.html>

# IDH Scalability - Data

Cassandra

- Ability to add nodes as throughput or storage needs increase
- Distributed key value storage allows linear scalability



# Security

- Data is encrypted in motion and at rest
- Extensive use of https internally and public facing traffic
- Encrypted AWS drives for the database cluster
- Captcha two factor SAR portal authentication
- Secure Data Channels
  - SFTP using asymmetric encryption keys
  - Soap 1.1 Secure Web Service using asymmetric encryption
- Security IV&V on going
- Continuous activity

# IDH Architecture Summary

- AWS Cloud Based Scalable Architecture
- Secured through NIST best practices
- Extensive use of open source software
- Flexible state integration channels to accommodate state level of technical resources





<http://integrity.naswa.org/>