InSight2

An Interactive Web Based Platform for Modeling and Analysis of Large Scale Argus Network Flow Data

THE UNIVERSITY OF TENNESSEE KNOXVILLE

Angel Kodituwakku
J.T. Liso
Dr. Jens Gregor

Jan 10, 2018

This material is based upon work supported by the National Science Foundation under Grant No. IRNC-1450959

InSight2 Foundational Work

- GLORIAD: World wide network for research & education

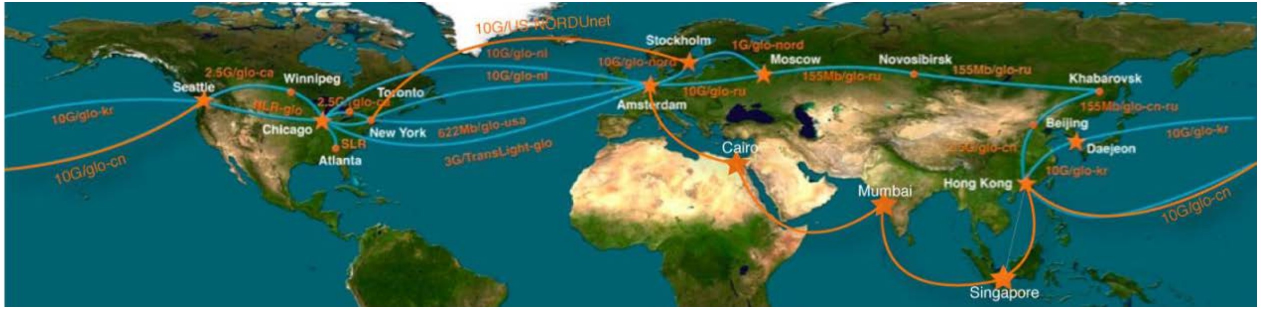
  Global **Ri**ng Network for **A**dvanced Applications **D**evelopment

  NSF sponsored project 2006-2015, Greg Cole (PI)

- InSight: Visualization of GLORIAD Argus flow-data

  Development ended with GLORIAD

- InSight2: Newly developed, completely redesigned tool



InSight2 Motivation

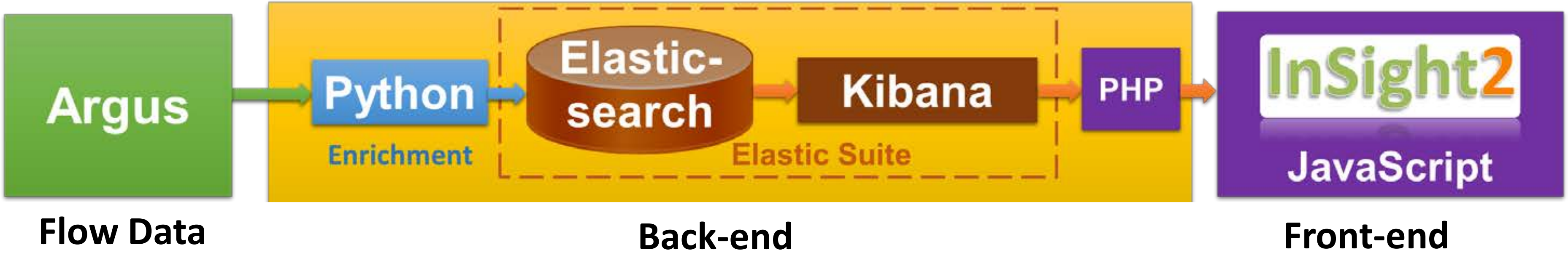
- Open-source Argus flow data analytics platform that provides:
  - **Performance metrics**
  - **Threat detection**
  - **Advanced analytics**
  - **Web based visualization**
- **Modular** architecture that supports **large scale data**, **real-time processing**, and **site-specific requirements**

InSight2 Features

- Core functionality: Performance metrics
- Plug-in extensions: Advanced analytics
  - Markov chain: Behavior prediction
  - Tensor analysis: Anomaly detection
  - Community plugins: TBD
- Data enrichment: Value-added knowledge
  - Geo-IP, Global Science Registry (IP-org mapping)
  - Threat lists, Blacklists (botnets, ransomware etc.)

InSight2 Implementation

- Enrichment: Python
- Database: Elasticsearch
- Visualization: Kibana
- Front-end: HTML/JS



InSight2 Capabilities 1/2

- Measurements
  - **Network statistics** (load, packets dropped, retransmitted)
  - **Usage statistics** (countries, organizations, ISPs)
  - **Diagnostics** (jitter, packet size, hops, delay)

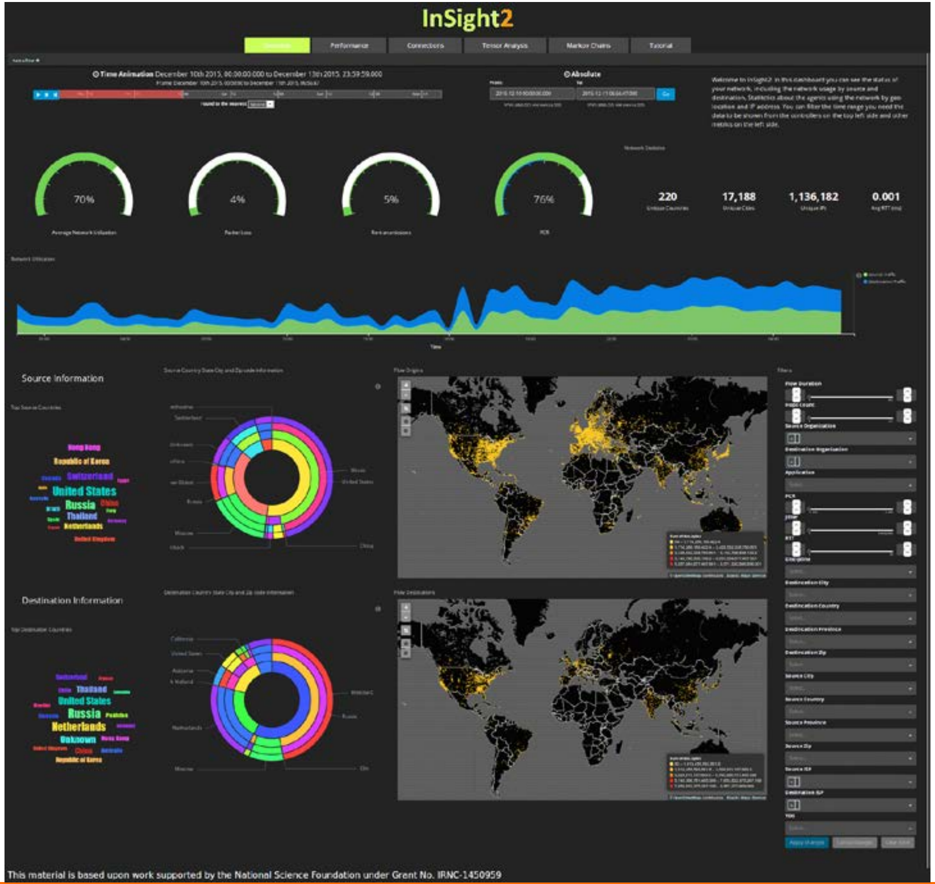
- Visualizations
  - Critical **activity gauges**
  - Overlaid **advanced metrics**
  - **Connections graphs** of top users

InSight2 Capabilities 2/2

- Intuitive filtering by UI interactions
  - Click UI elements to add/remove filters by country, ISP etc
  - Click and drag to filter time range in timeline
  - Click and drag define visual geo-location bounds in geo-maps
- Geo-location mapping: MaxMind Geo-IP database
- Threat detection: Miscl. on-line databases
- Utilization prediction: Markov chain modeling
- Anomaly detection: Tensor based data analysis

InSight2 Traffic Overview

- Main Dashboard
- Activity Gauges
- Country Tag Cloud
- Geo Map
- Intuitive filters

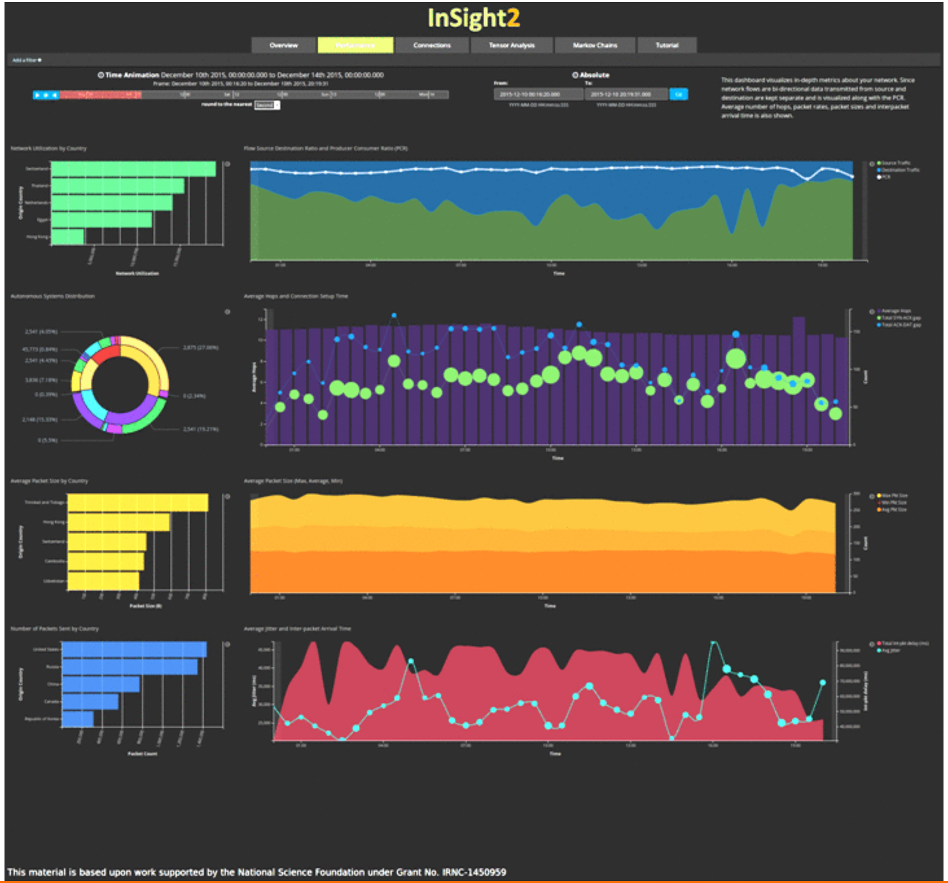


Welcome to InSight2. In this dashboard you can see the status of your network, including the network usage by source and destination, Statistics about the agents using the network by geo location and IP address. You can filter the time range you need the data to be shown from the controllers on the top left side and other metrics on the left side.

This material is based upon work supported by the National Science Foundation under Grant No. IRNC-1450959

InSight2 Performance Metrics

- Traffic ratio and PCR
- Setup time and hops
- Packet size
- Jitter and inter-packet arrival time

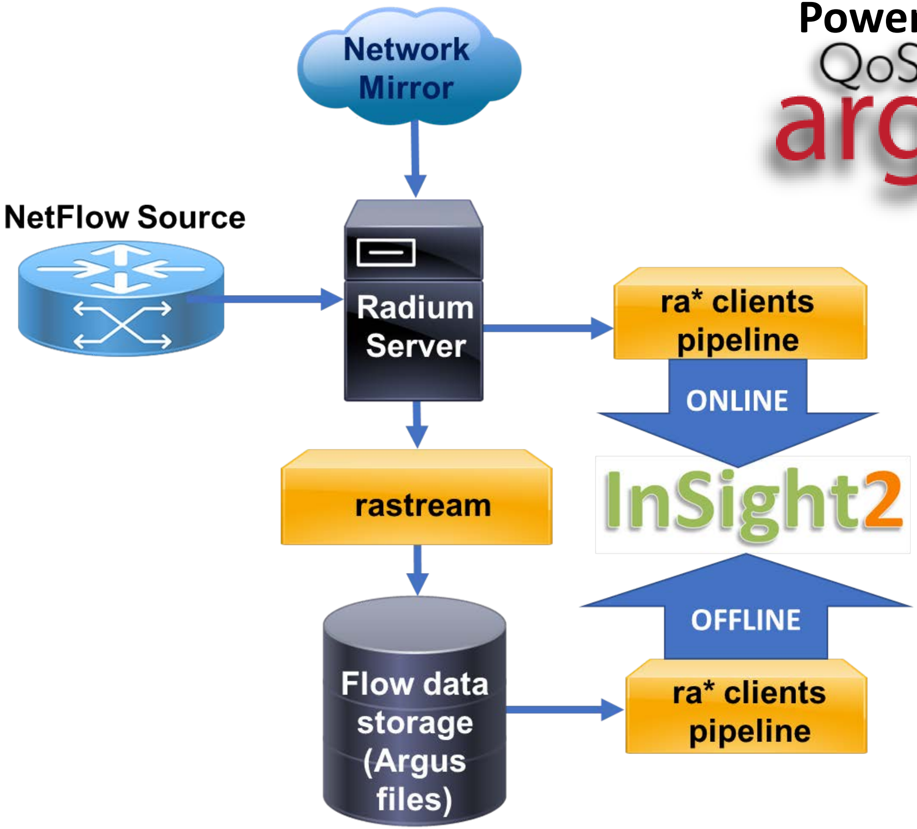

InSight2
Overview
Performance
Connections
Tensor Analysis
Markov Chains
Tutorial
Add a filter +
Absolute
From:
To:
2015-12-10 00:16:20.000
2015-12-10 20:19:31.000
Go
round to the nearest
Second
Network Utilization by Country
Flow Source Destination Ratio and Producer Consumer Ratio (PCR)
Source Traffic
Destination Traffic
PCR
Autonomous Systems Distribution
Average Hops and Connection Setup Time
Average Packet Size by Country
Average Packet Size (Max, Average, Min)
Number of Packets Sent by Country
Average Jitter and Inter-packet Arrival Time
Time

Time Animation December 10th 2015, 00:00:00.000 to December 14th 2015, 00:00:00.000
Frame: December 10th 2015, 00:16:20 to December 10th 2015, 20:19:31

This dashboard visualizes in-depth metrics about your network. Since network flows are bi-directional data transmitted from source and destination are kept separate and is visualized along with the PCR. Average number of hops, packet rates, packet sizes and interpacket arrival time is also shown.

This material is based upon work supported by the National Science Foundation under Grant No. IRNC-1450959

InSight2 Argus Flow Data

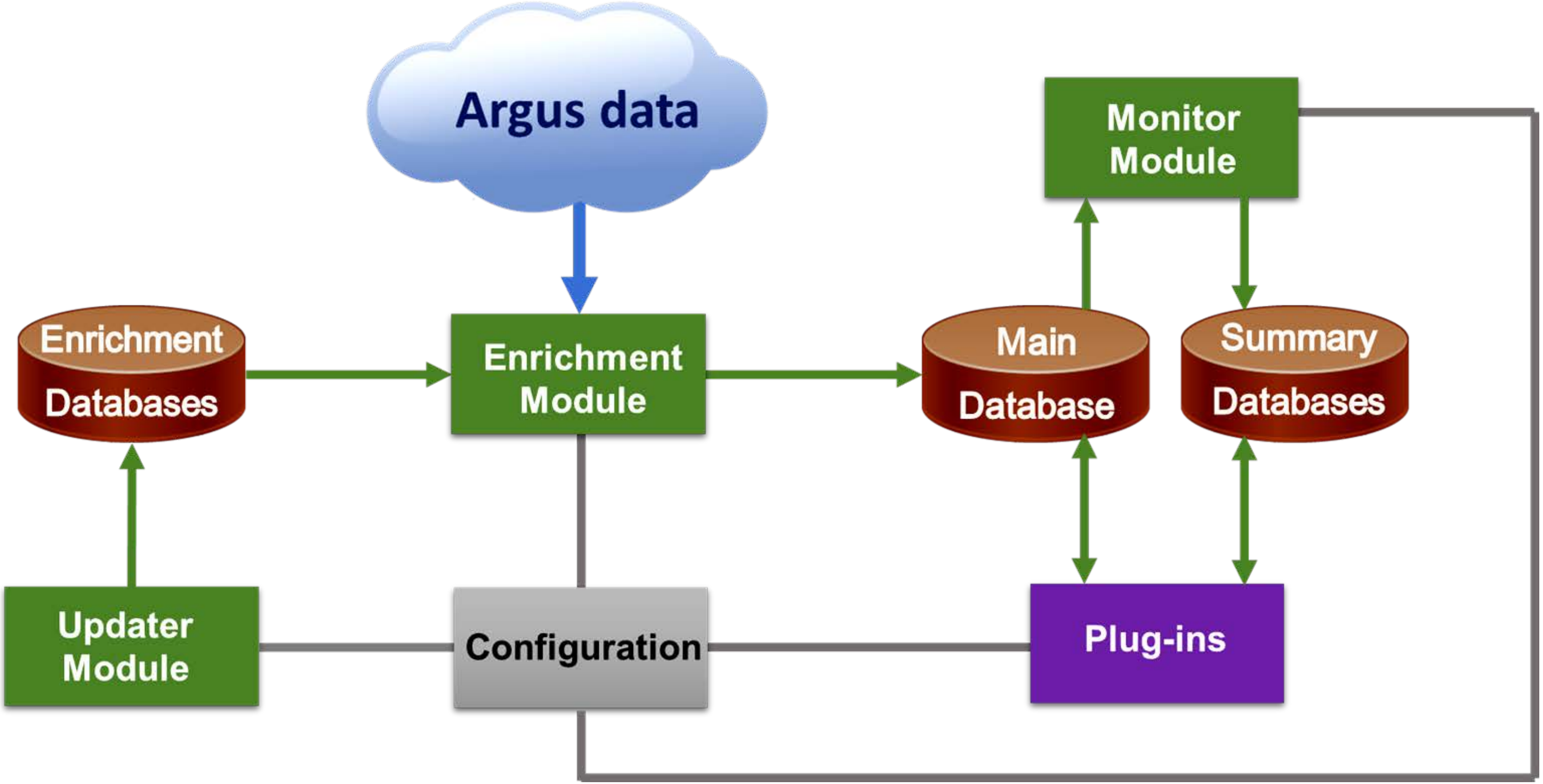


Powered by QoSient argus

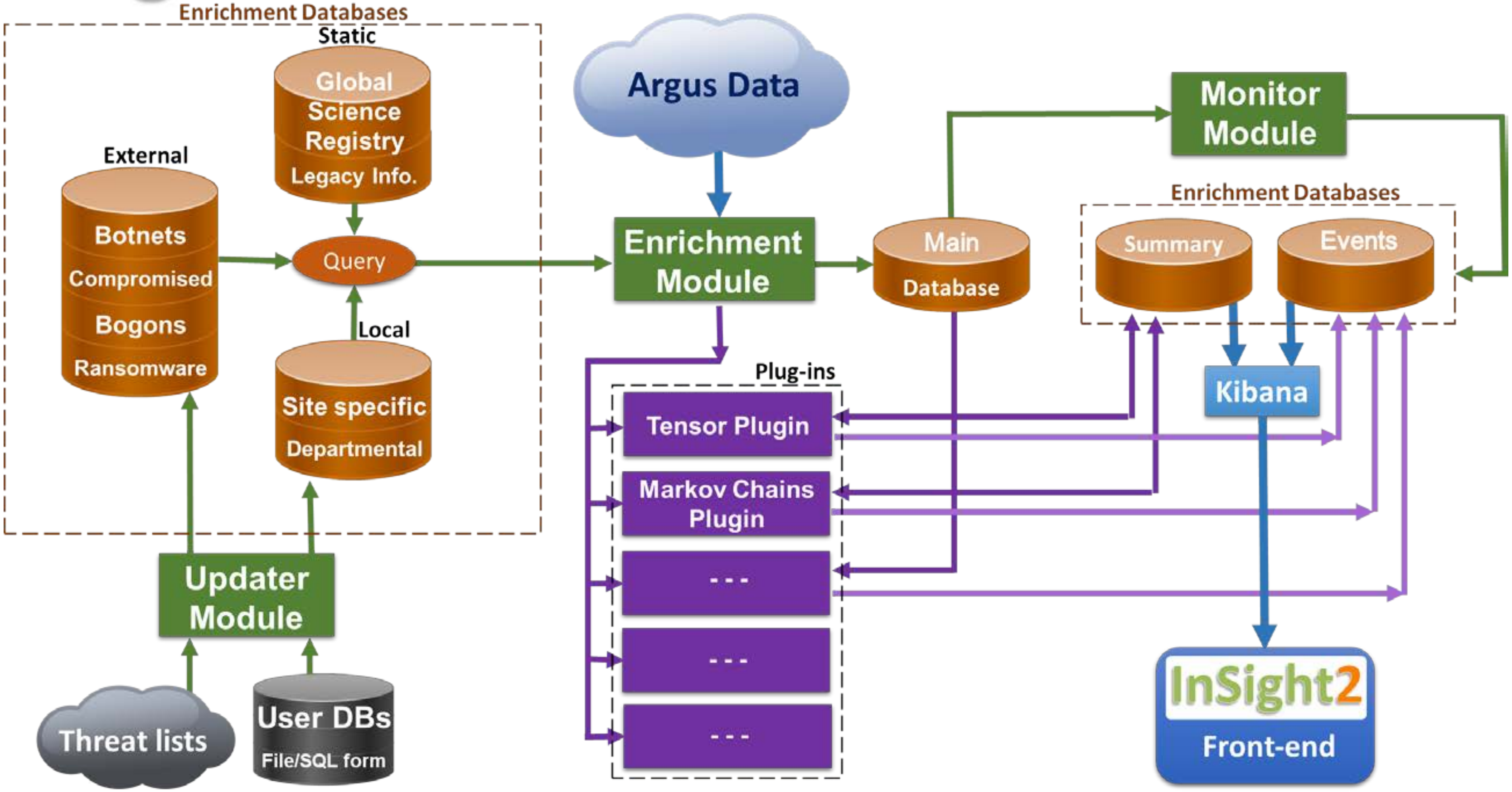
Argus is used by

Oak Ridge National Laboratory
Carnegie Mellon University Software Engineering Institute
Stanford University
U.S. Department of Homeland Security
SONY
NSF
MaxMind®
Defense Information Systems Agency
CERT
Department of Defense United States of America
ArcSight
Pluribus Networks
MySQL

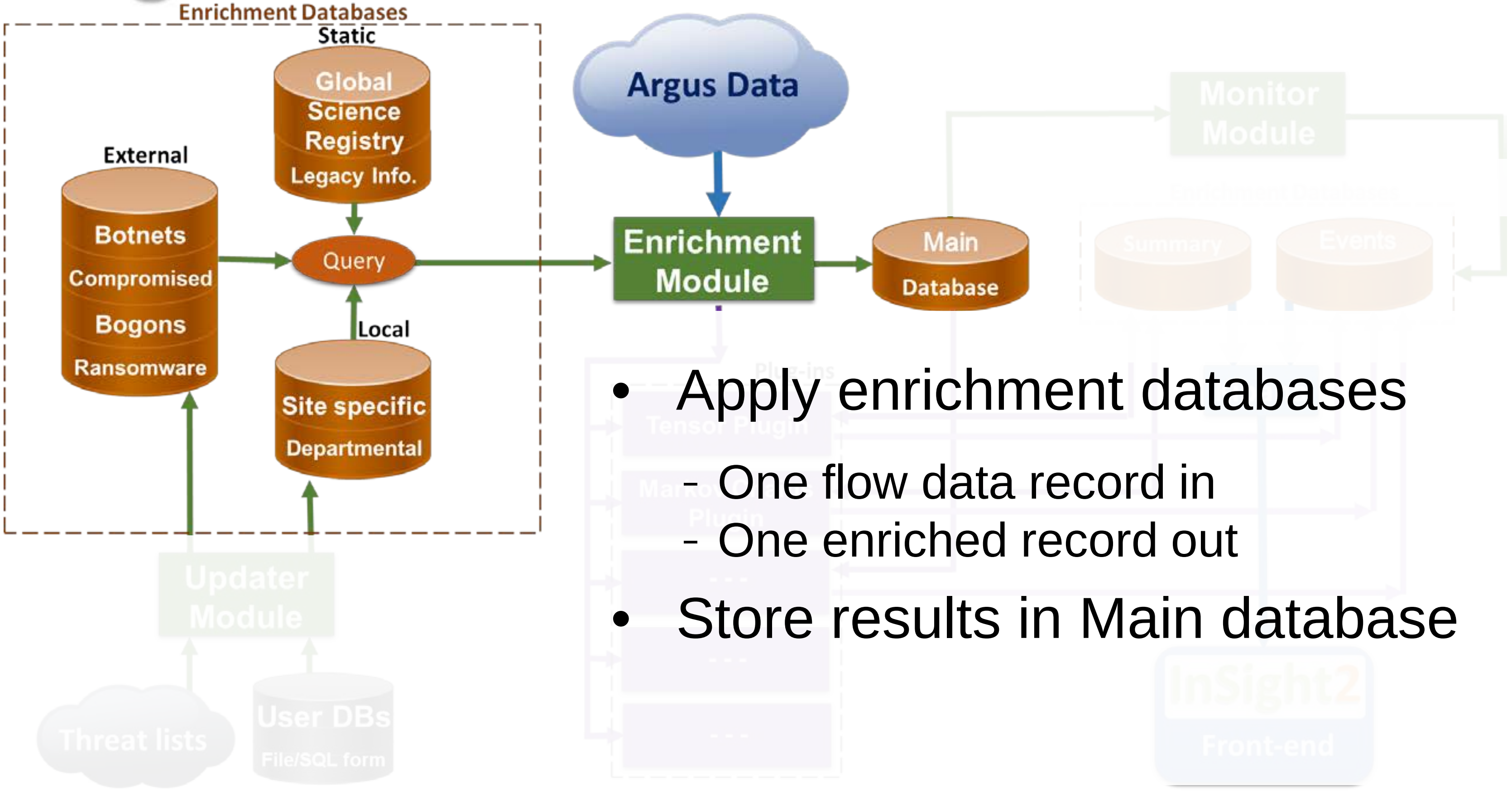
InSight2 Software Architecture 1/6



InSight2 Software Architecture 2/6



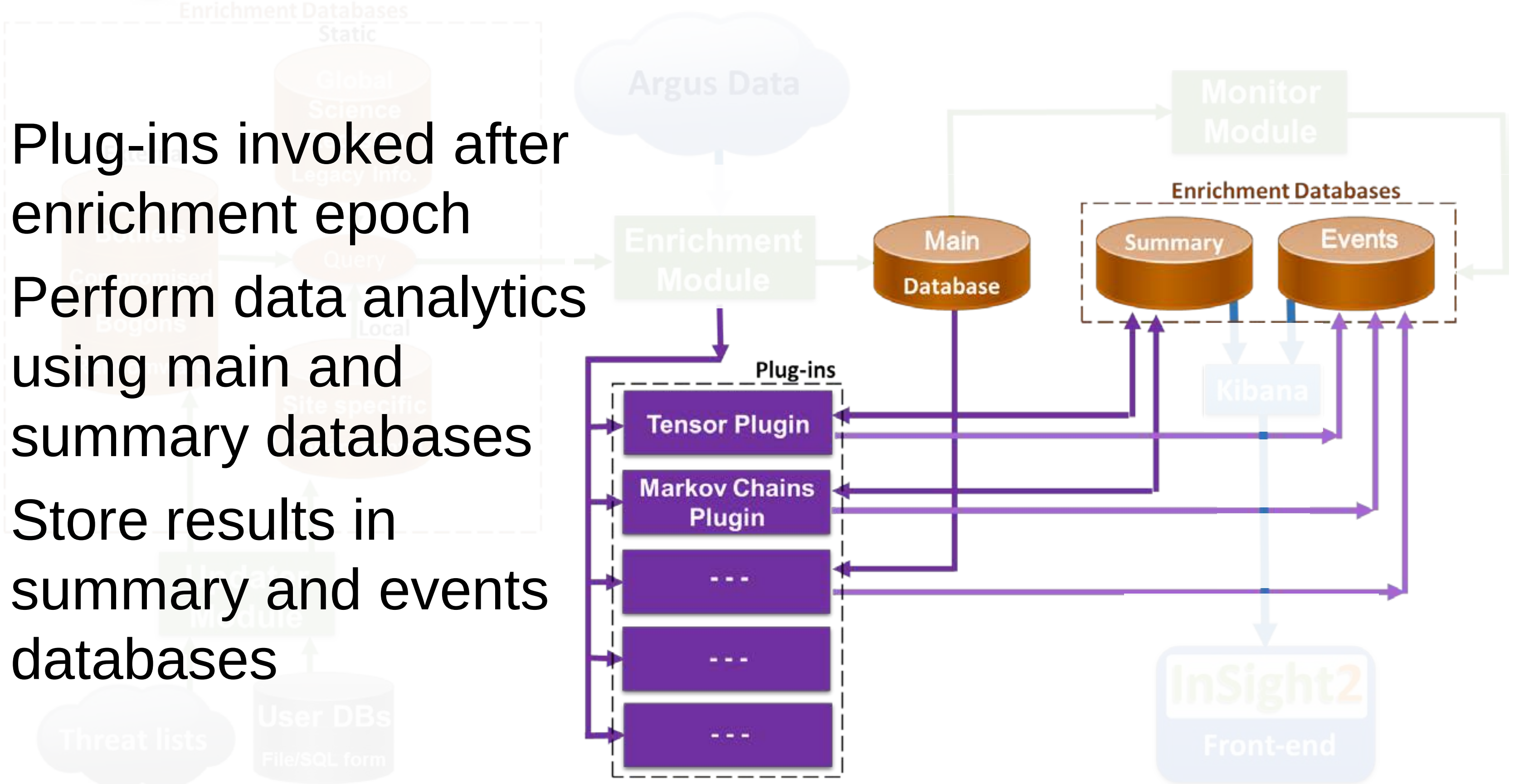
InSight2 Software Architecture 3/6



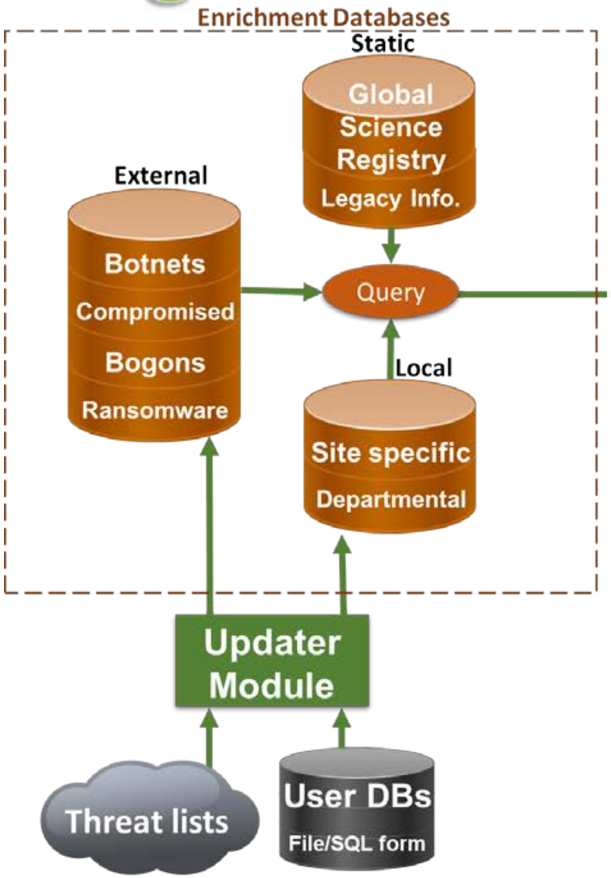
- Apply enrichment databases
  - One flow data record in
  - One enriched record out
- Store results in Main database

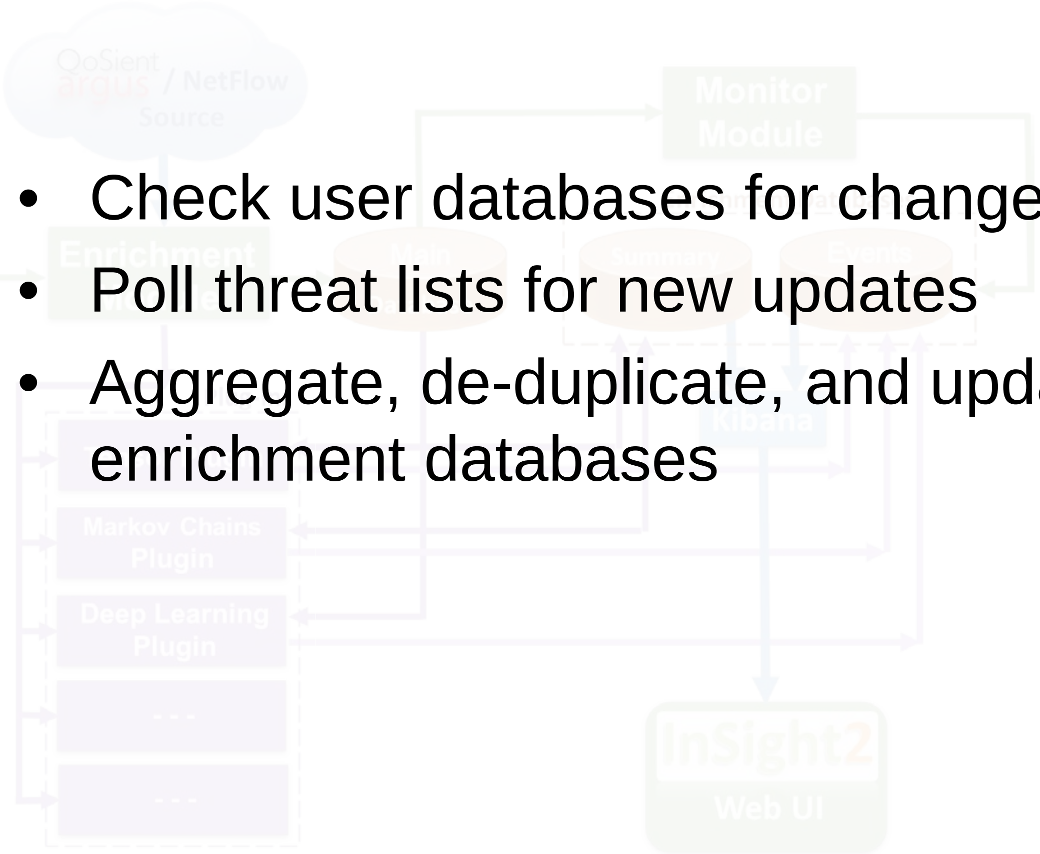
InSight2 Software Architecture 4/6

- Plug-ins invoked after enrichment epoch
- Perform data analytics using main and summary databases
- Store results in summary and events databases



InSight2 Software Architecture 5/6

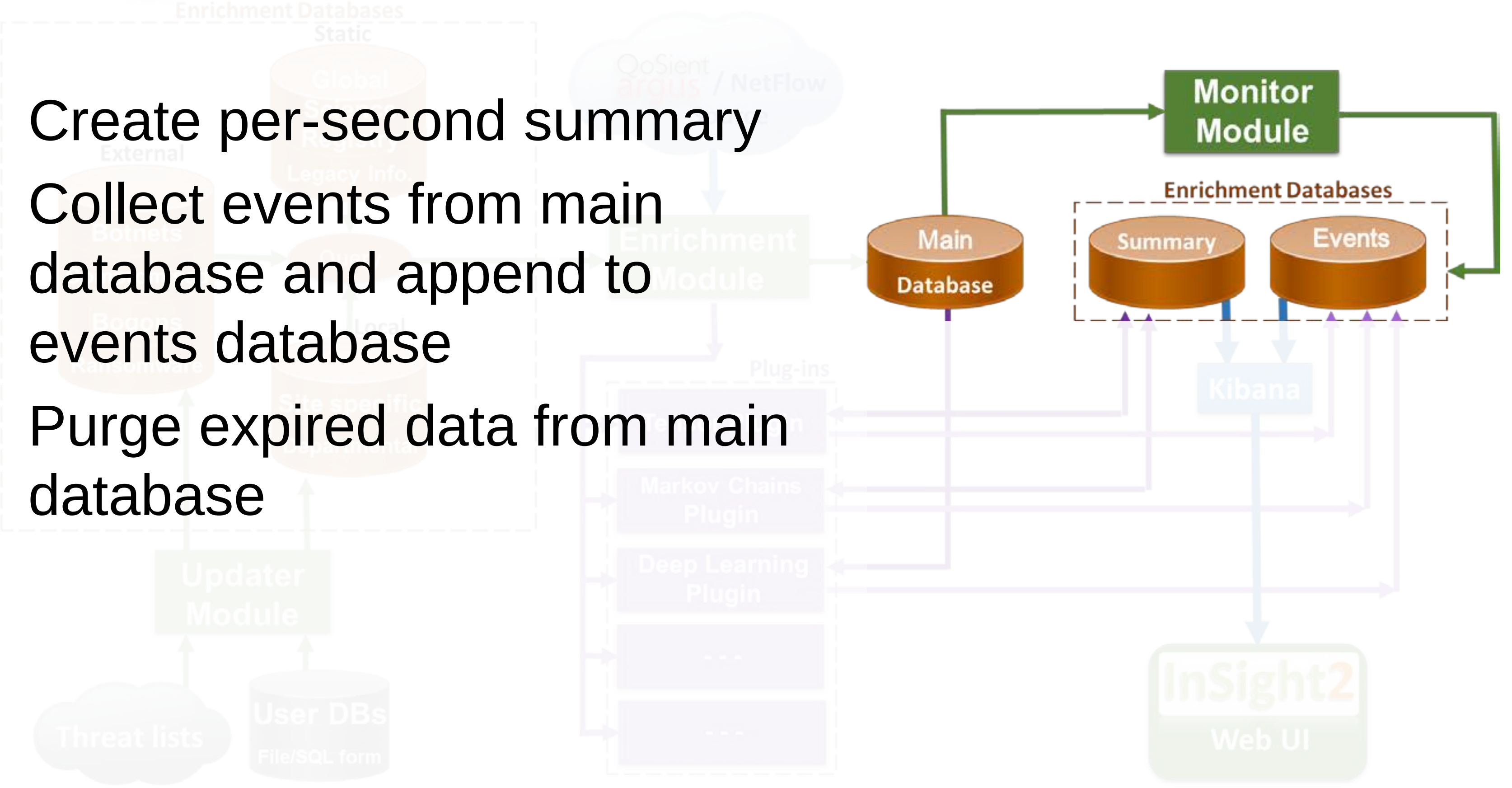




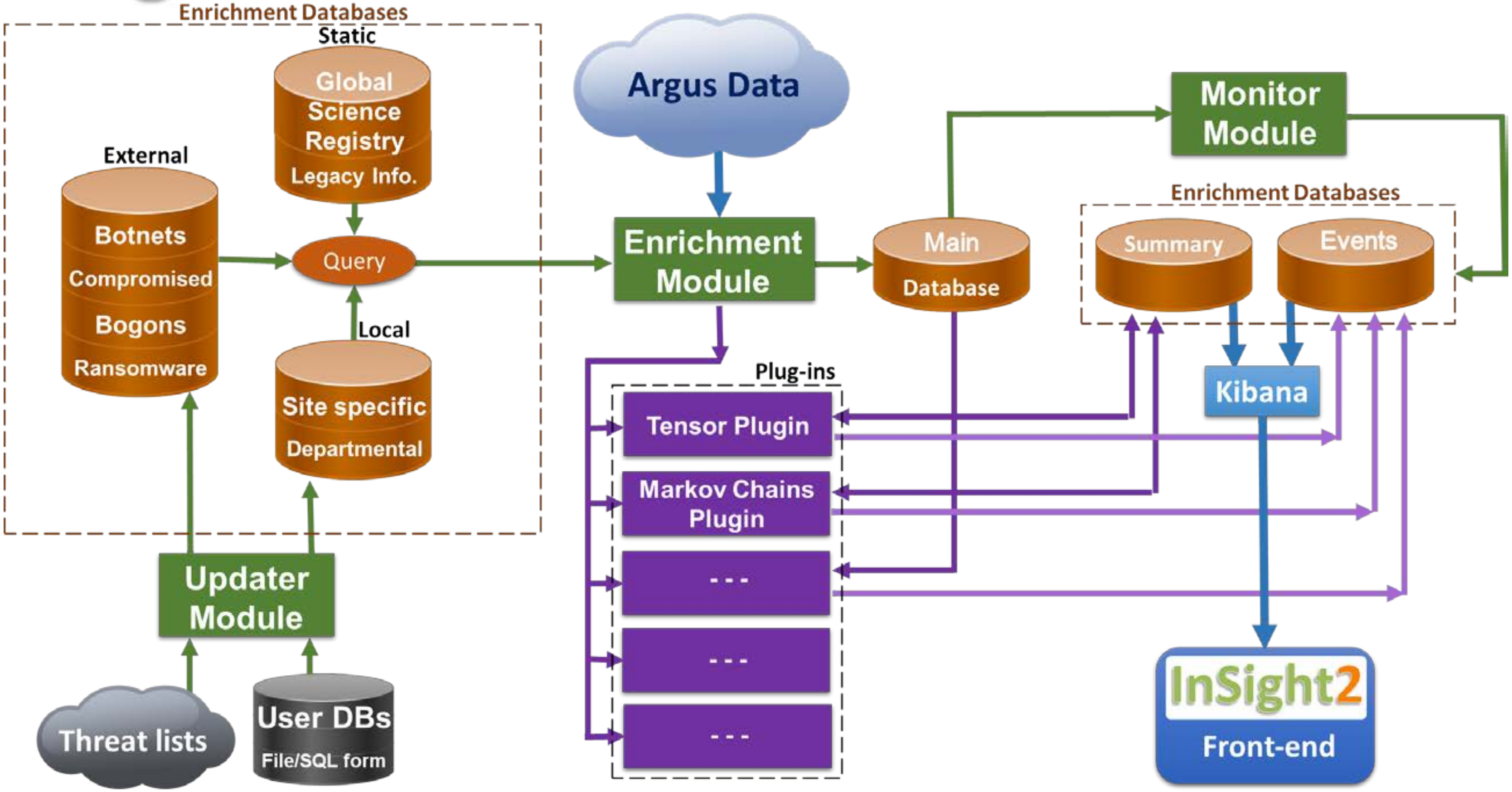
- Check user databases for changes
- Poll threat lists for new updates
- Aggregate, de-duplicate, and update enrichment databases

InSight2 Software Architecture 6/6

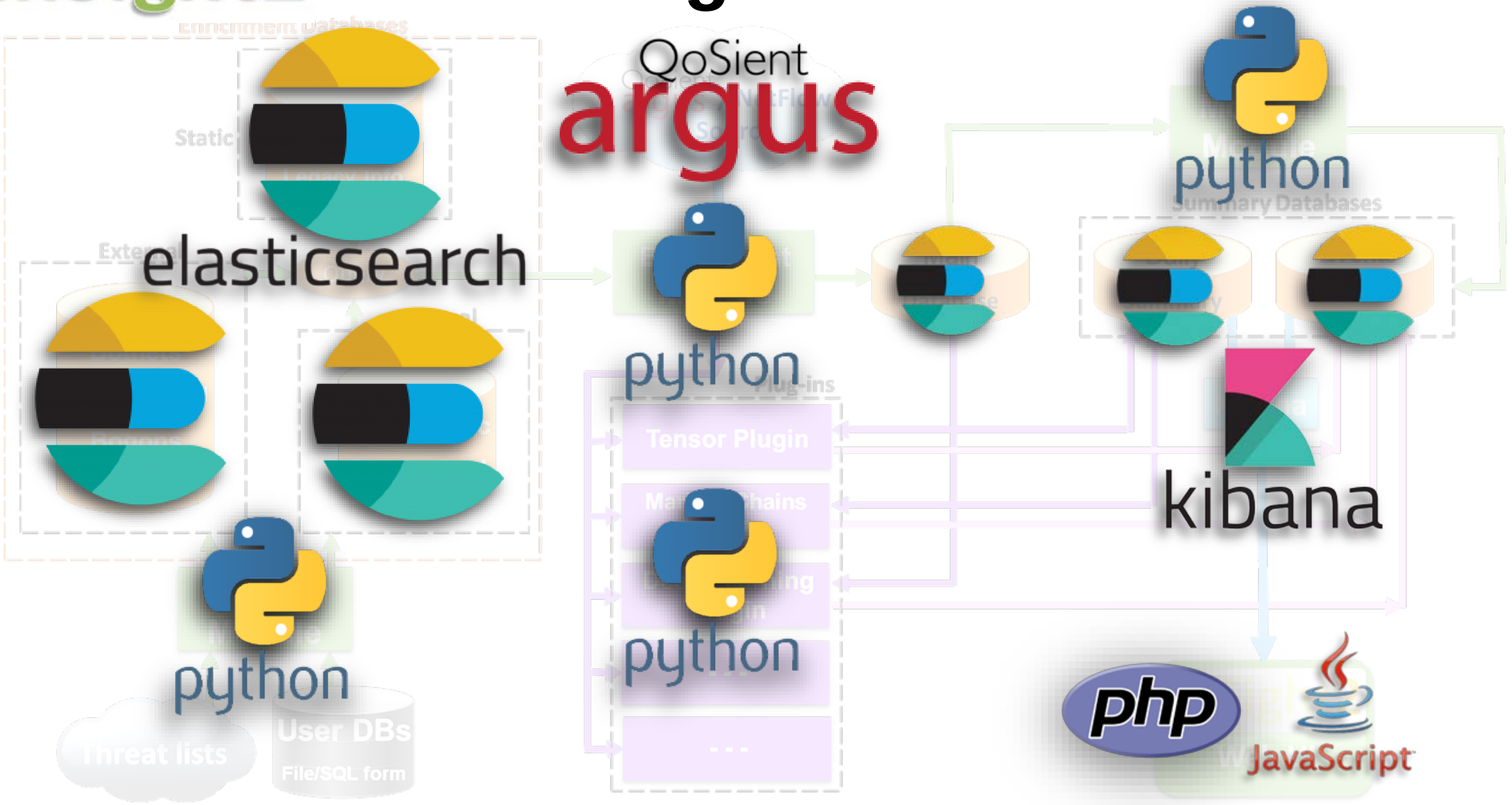
- Create per-second summary
- Collect events from main database and append to events database
- Purge expired data from main database



InSight2 Technologies Used



InSight2 Technologies Used



InSight2 uses elastic

elasticsearch

- Highly scalable
- NoSQL database
- Full-text search engine
- Distributed

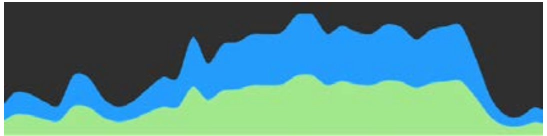
kibana

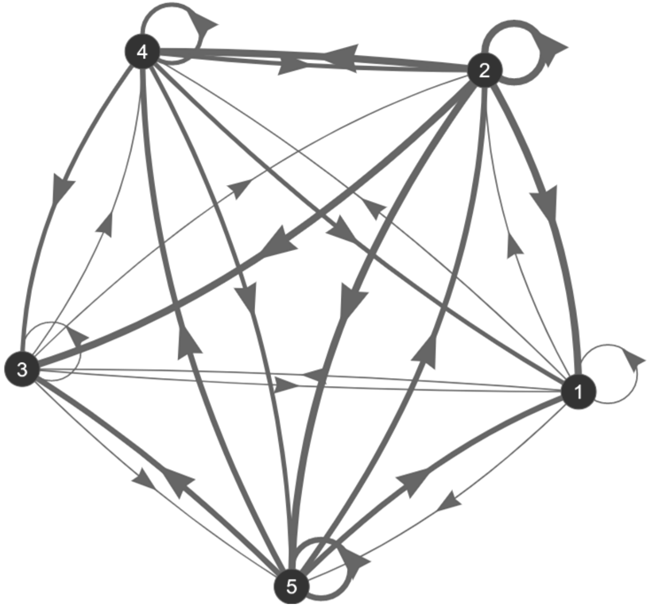
- Visualization platform
- Intuitive dashboards
- Native integration with ES
- Geo-map tile service

InSight2 Plug-in: Markov Chain 1/2

- State transition model
- Stochastic: Prob($s_{i+1}|s_i$)

- Inferred from training data
- Model analysis
  - Steady-state
  - First-transitions
- Live data processing





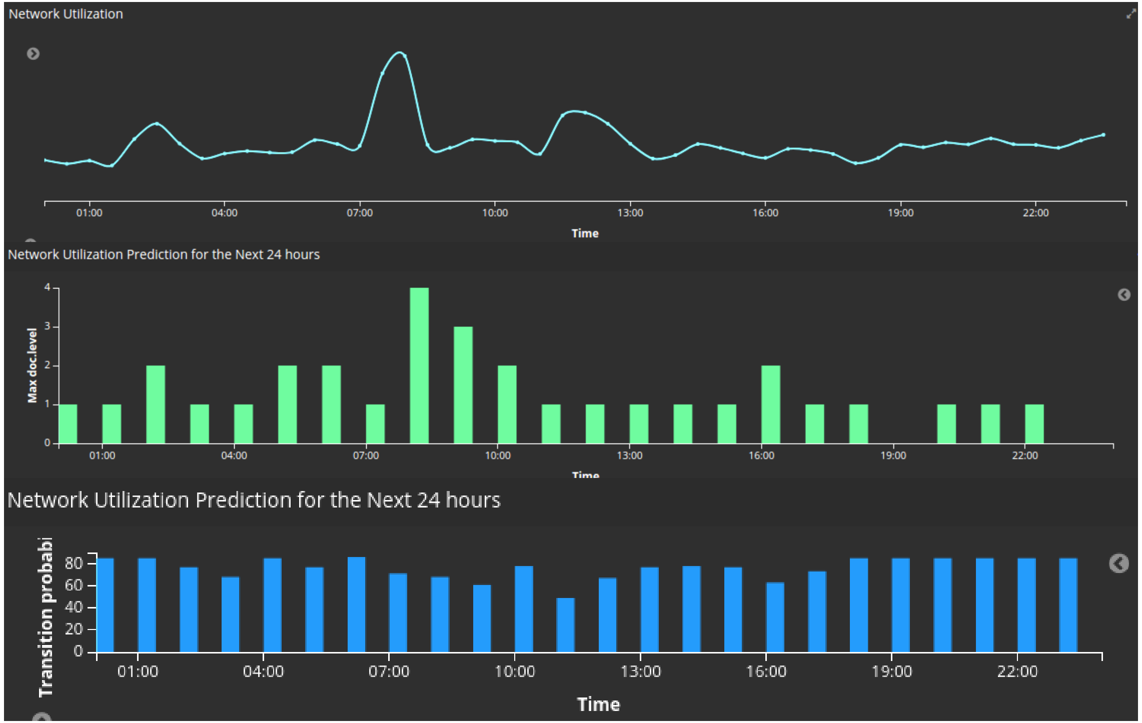
InSight2 Plug-in: Markov Chain 2/2

- Usage: Network utilization prediction

Actual Usage

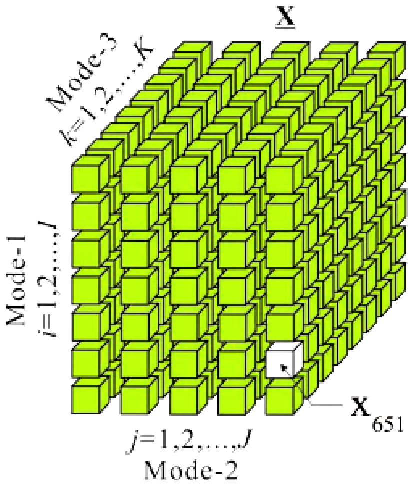
Predicted Usage

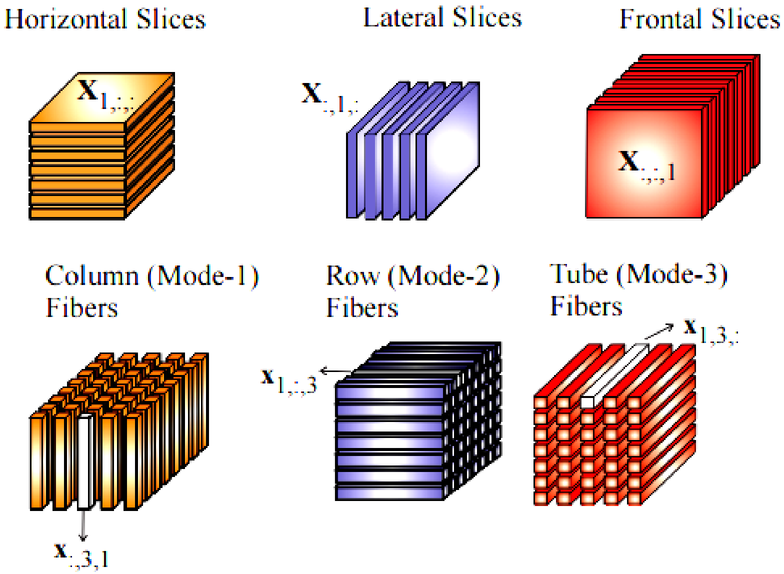
State Transition Probabilities



InSight2 Plug-in: Tensor Analysis 1/3

- Tensor: multidimensional matrix of real numbers
- Each mode is *n*-dimensional matrix (called slice)





InSight2 Plug-in: Tensor Analysis 2/3

- Tensor energy
  - Average sum of squares per slice given mode

- Data sparsification
  - Low energy change data discarded during update

- Event detection
  - High energy change data indicates new trend that may warrant investigation (anomalous behavior?)

S. Papadimitriou et al, Streaming Pattern Discovery in Multiple Time-Series, Proc. VLDB, Trondheim, Norway, 2005

InSight2 Plug-in: Tensor Analysis 3/3

Observed source traffic

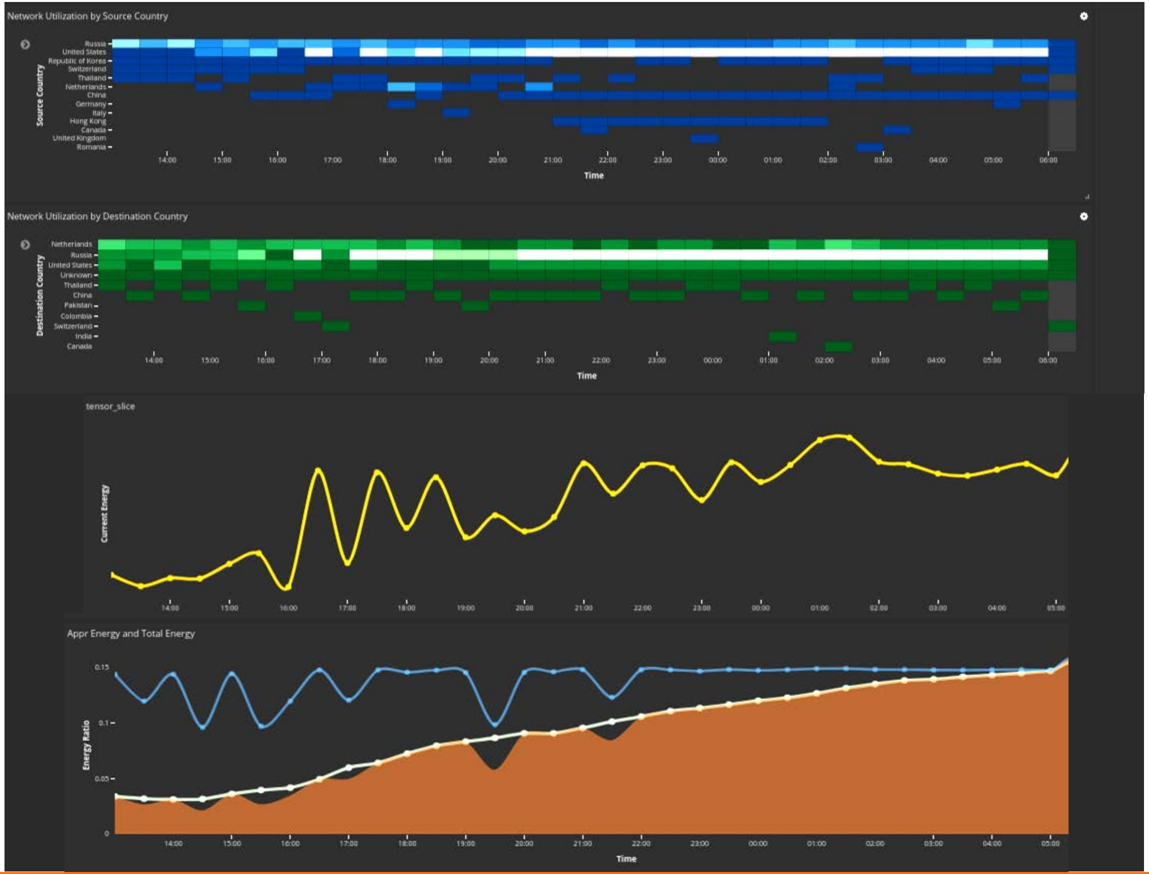
Observed destination traffic

Slice Energy

Anticipated Energy

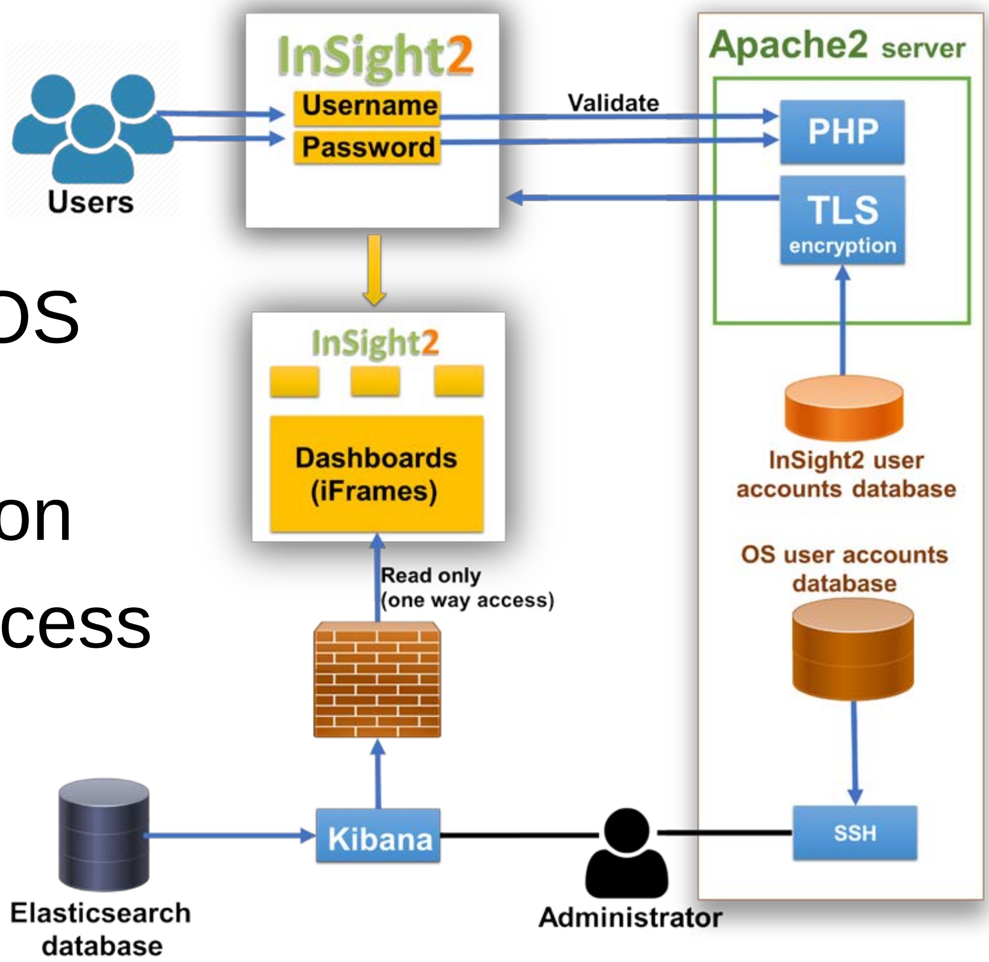
Actual Energy

Energy Ratio



InSight2 Frontend

- TLS 1.2 transport
- Separate InSight2 and OS user authentication
- Server side authentication
- Secure administrator access
- Read only / one way dashboards



InSight2 Summary

- Argus flowdata modeling and analysis
- Interactive web based platform
- Open-source modular software (release TBD)

- Partners
  - QoSient, Cisco ASIG
  - Stanford University, KISTI (South Korea)

- Work supported by NSF: IRNC-1450959