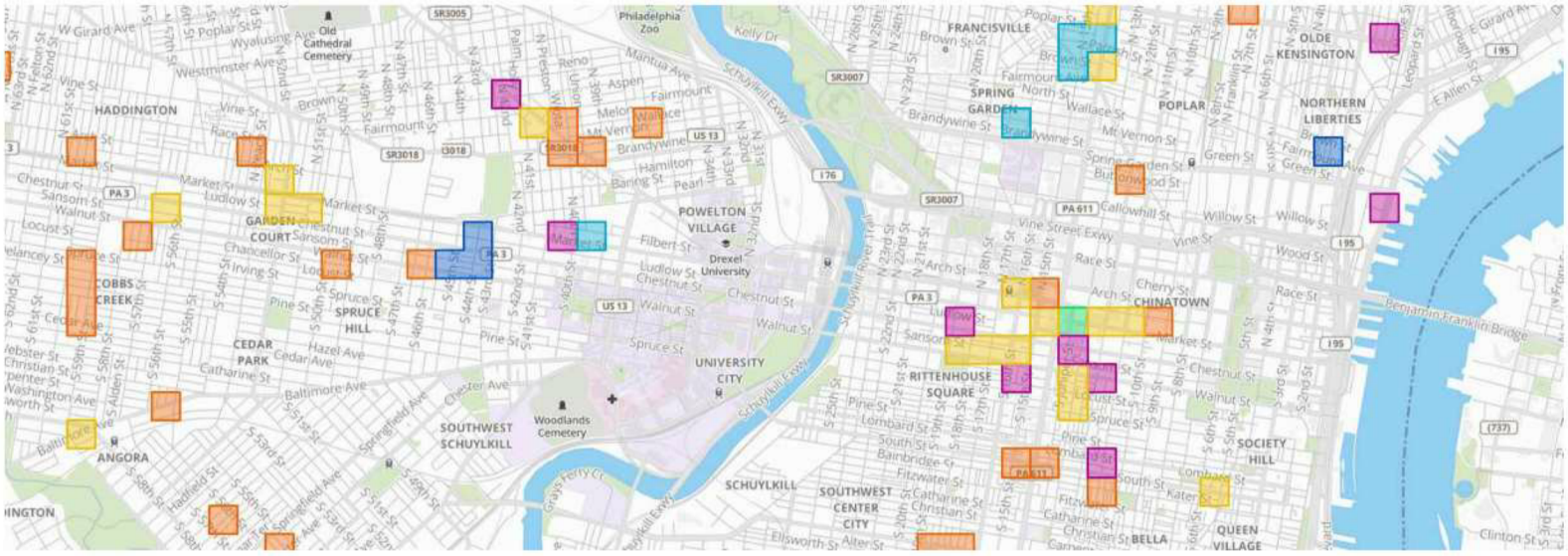
HunchLab Pilot Program for the New York City Police Department



HunchLab

Prepared by

azavea
340 N 12th St, Suite 402
Philadelphia, PA 19107
(215) 925-2600
http://www.azavea.com

Executive Summary

Azavea's HunchLab system provides next generation predictive policing technologies in an easy-to-use web-based solution. Built to utilize a broad set of crime theories and datasets, HunchLab automatically generates robust forecasts for crime and distills these forecasts into target mission areas for easy deployment in the field. We are offering the New York City Police Department the opportunity to pilot HunchLab in an operational setting for six months, which will enable the Department to determine HunchLab's efficacy as a daily tool used by the NYPD.

HunchLab is delivered as a secure cloud subscription, which reduces operational cost and complexity. Azavea will manage the hosting infrastructure, security updates, and 2nd tier support. This cloud-based approach enables the HunchLab to leverage significant amounts of computing power in an elastic manner, a critical requirement for providing the advanced statistical algorithms the system employs. To provide a secure application, we have consulted the FBI's CJIS guidelines within our system architecture, including mitigation techniques such as encrypting data in transit and providing optional 2-factor authentication. More information about HunchLab security is provided in the Appendix at the end of this proposal.

Current trends reflect an interest in predictive policing, increasing adoption of cloud services, pervasive location information, and the proliferation of mobile devices. These trends are occurring in an extremely constrained budget environment in most law enforcement agencies. Azavea believes these factors will combine to cause the law enforcement community to undergo a transformational change in the use of technology, similar in scope to the advent of GIS-based crime mapping in the 1990's. The new HunchLab anticipates these advances in an unprecedented manner; we believe that HunchLab 2.0 truly represents the future of policing. We look forward to working with the New York City Police Department to implement our current technology and spearhead new, innovative approaches to improve public safety.

Pilot Process

Azavea believes that, where possible, vendors should provide low-cost opportunities for governments to validate solutions before committing to long-term acquisition. To support this belief, we have designed a pilot program that provides a test period for clients to gain experience with HunchLab and validate adoption within their agency. The pilot program begins with one month of configuration time in which to tweak settings and align the system with the department's priorities. Following this period, the pilot program provides six months of subscription service at a reduced cost. During this period, the New York City Police Department can use HunchLab operationally within a single police precinct.

Phase 1: Configuration and Vetting (One month)

The initial configuration will commence once we receive historic crime data. Over the course of the configuration timeframe, Azavea will work with NYPD to make sure the mission areas make sense, to tweak settings as necessary so the system reflects the priorities of the department, and to measure the initial accuracy of the system.

Phase 2: Operational Use (Six months)

The six-month operational pilot will begin immediately after the configuration and vetting phase. During this phase, HunchLab will be deployed in a single police precinct of the Department's choosing. Command staff will task officers to focus on mission areas. Mission areas will be disseminated as deemed most appropriate by NYPD (through the web application, electronic PDFs, or physical print outs).

Optional Activities

There are several additional activities NYPD may want to take conduct during the pilot program. The HunchLab API will be fully accessible, allowing your IT department to pull data out of HunchLab and into other mapping or analysis applications such as the Domain Awareness System. You might also want to collaborate with an academic during this period to examine the adoption, impact, and accuracy of the tool.

The team at Azavea that works on HunchLab also looks forward to the feedback that NYPD can provide about the application. This feedback may take the form of user surveys, conference calls, visits to the precinct, and officer "ride-alongs" as deemed appropriate by NYPD. Our past pilot implementations of HunchLab have been a rich source of enhancement ideas and we have often delivered such enhancements before the pilots have concluded.

Pilot Deliverables

The pilot program will deliver a view into the operational impact of predictive analysis at your department. It will also validate the technical requirements of such systems in regards to data formats, data cleanliness, and timely dissemination of information to the field. The HunchLab pilot program provides insight into the expected level of accuracy for predictive modeling of your data through automated accuracy measurements from the HunchLab modeling process. Finally, summary information about the factors represented in the HunchLab models for each crime type will be provided.

Pilot Features

- Automated advanced crime forecasts using crime data and other related data sets
- Interactive modern desktop browser interface to configure and view mission areas
- On-demand generation of PDF reports for mission areas for print and email distribution
- Adaptive mobile browser interface to HunchLab (as available based upon NYPD roll out of devices and tweaks by Azavea)
- User single sign-on via the SAML standard

Pilot Assumptions

- HunchLab will be hosted in Azavea's secure Amazon Web Services cloud infrastructure with data contained fully within the United States.
- NYPD will provide access to key crime data fields in an open format for a five-year period – either in a specified CSV or in a similar, accessible format that Azavea can transform into such CSV such as an ODBC connection.
- Daily (or more frequent) crime data uploads to HunchLab will be posted out of NYPD's internal network automatically via a scheduled task or similar process. Azavea can support the development of this process or NYPD's IT department can handle the process.
- NYPD users will access the application from modern web browsers that support secure connections as specified in the application requirements provided below.
- Azavea will not utilize NYPD data for any purpose other than to provide this pilot.
- Azavea will purge NYPD data from HunchLab upon NYPD request at any time.
- HunchLab will be provided for a period not to exceed seven months (configuration and operational pilot) unless extended by the mutual agreement of both parties.

Pilot Cost

Item	Total
Custom Pilot Program (One precinct for 6 operational months)	$50,000
Total	**$50,000**

HunchLab Application Features

HunchLab 2.0 incorporates concepts such as: temporal patterns (time of day, day of week, day of month, seasonality); weather; risk terrain modeling (locations of bars, bus stops, etc.); socioeconomic indicators (such as collective efficacy indicators); historic crime levels; and near-repeat patterns to help police departments understand and respond more effectively to crime using the resources available to them. After predicting crime expectations across the entire jurisdiction for a shift, HunchLab calculates the relative crime level per unit of patrol effort for each area, sorts these areas from highest to lowest relative risk, and selects the mission areas that can be patrolled by the available resources. This process maximizes the impact of patrols – placing them in areas where crime risk is greatest – and ensures that the right quantity of mission areas is crafted.

At the core of HunchLab 2.0 is a new crime-forecasting engine. These forecasts power the predictive missions feature, enabling departments to proactively generate the appropriate quantity of mission areas based upon the organizational and societal importance of various types of crime. The forecasting engine uses ensemble machine learning approaches that can incorporate the following crime patterns into a single prediction of criminal risk:

- Baseline crime levels
  - Similar to traditional hotspot maps
- Near repeat patterns
  - Event recency (contagion)
- Risk Terrain Modeling
  - Proximity and density of geographic features (points, lines, and polygons)
- Routine activity theory
  - Offender: proximity and concentration of known offenders
  - Guardianship: police presence (historic AVL / GPS data) [Requires additional integration work not available in a pilot.]
  - Targets: measures of exposure such as population, parcels, or automobiles
- Collective Efficacy
  - Socioeconomic indicators, neighborhood heterogeneity, etc.
- Temporal cycles
  - Seasonality, time of month, day of week, time of day, etc.
- Recurring temporal events
  - Holidays, sporting events, etc.
- Weather
  - Temperature, precipitation, etc.

Azavea believes that the use of non-crime data sets as variables within an operational crime prediction system is important because variables based solely upon crime data become skewed as predictions are used operationally. As crimes are prevented in mission areas due to police response, the only variables identifying areas as high risk are skewed in other systems. By including other data sets, our system is more robust against this issue.

While available GIS tools have enabled law enforcement agencies to advance their understanding of crime through more effective geographic visualization for many years, these tools have traditionally required trained analysts, are used by a tiny subset of agency personnel, and are largely reactive in nature. There is a compelling need for new crime analysis tools that perform automated discovery of crime patterns and make that information available in a format that can be easily understood and acted upon at a variety of agency levels, including officers in the field. HunchLab provides these capabilities. It is a groundbreaking system that "learns" which crime theories matter for a given location and automatically calibrates the influence of these theories – both individually and as arbitrarily complex interactions – within a designated geography to identify and address a wide range of crime and other public safety issues.

1. Predictive Missions

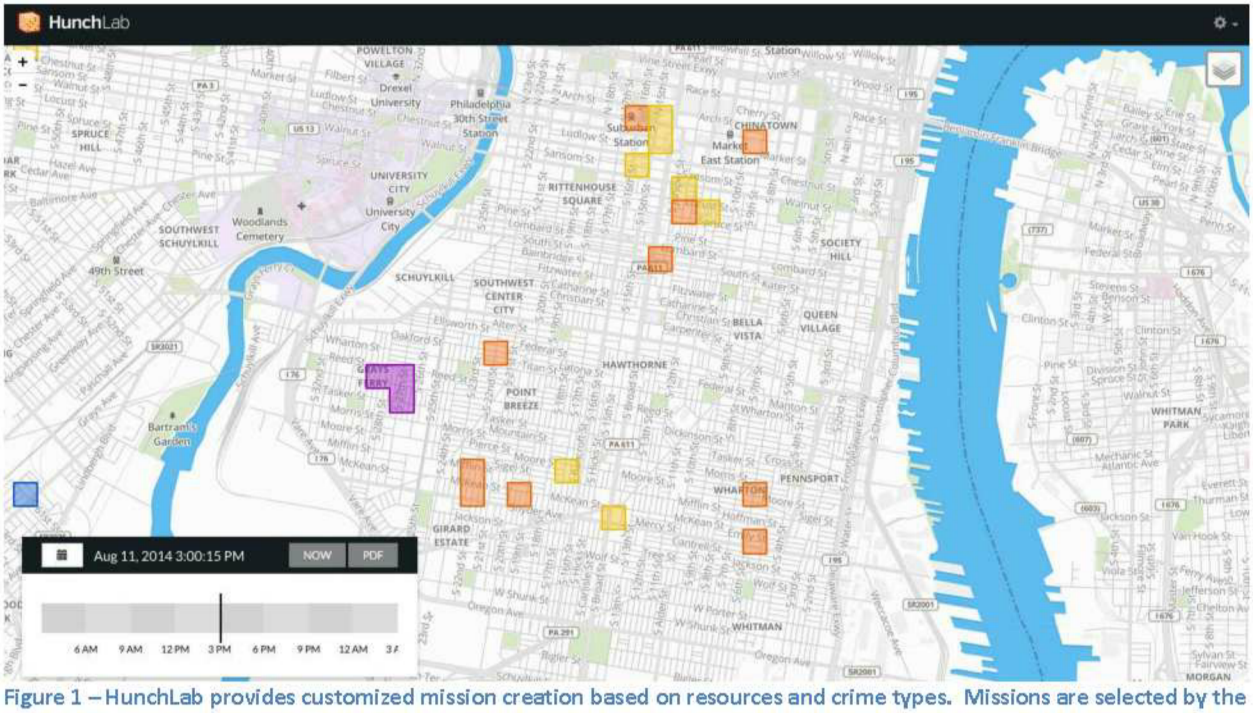


Figure 1 – HunchLab provides customized mission creation based on resources and crime types. Missions are selected by the combined, weighted risk of all configured crime models. Color represents the dominant risk for a mission area.

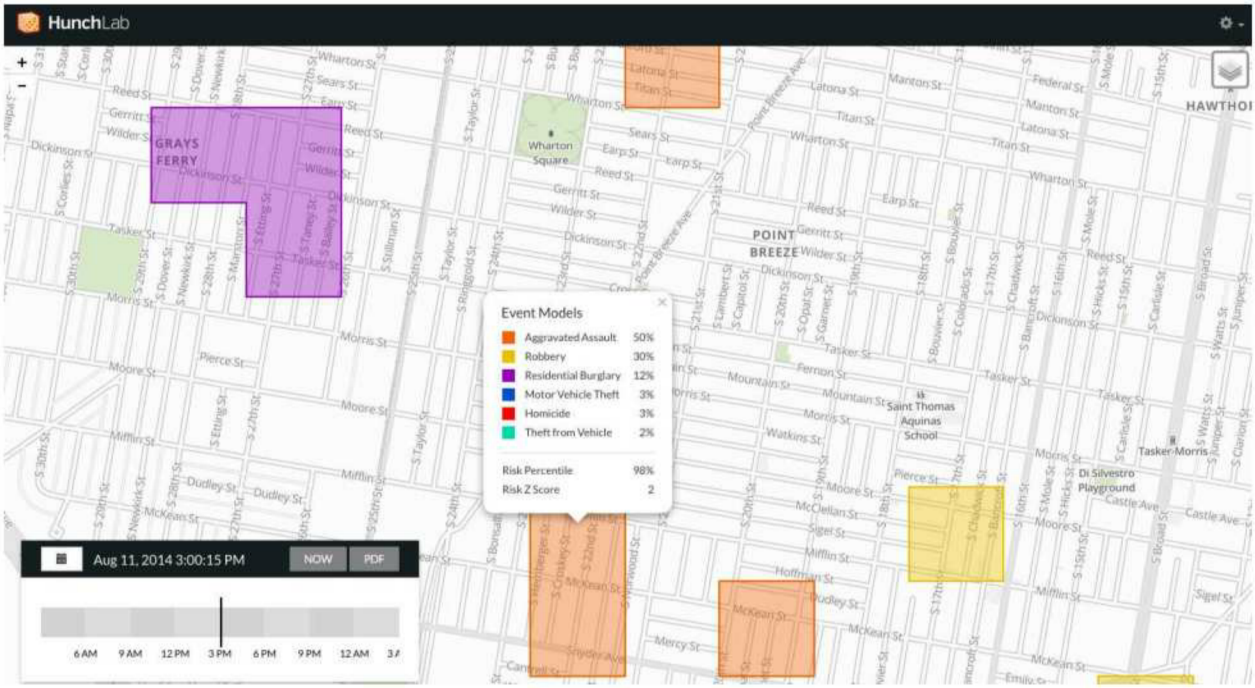


Figure 2- Each mission area displays a risk profile of the crime types that went into selecting this location.

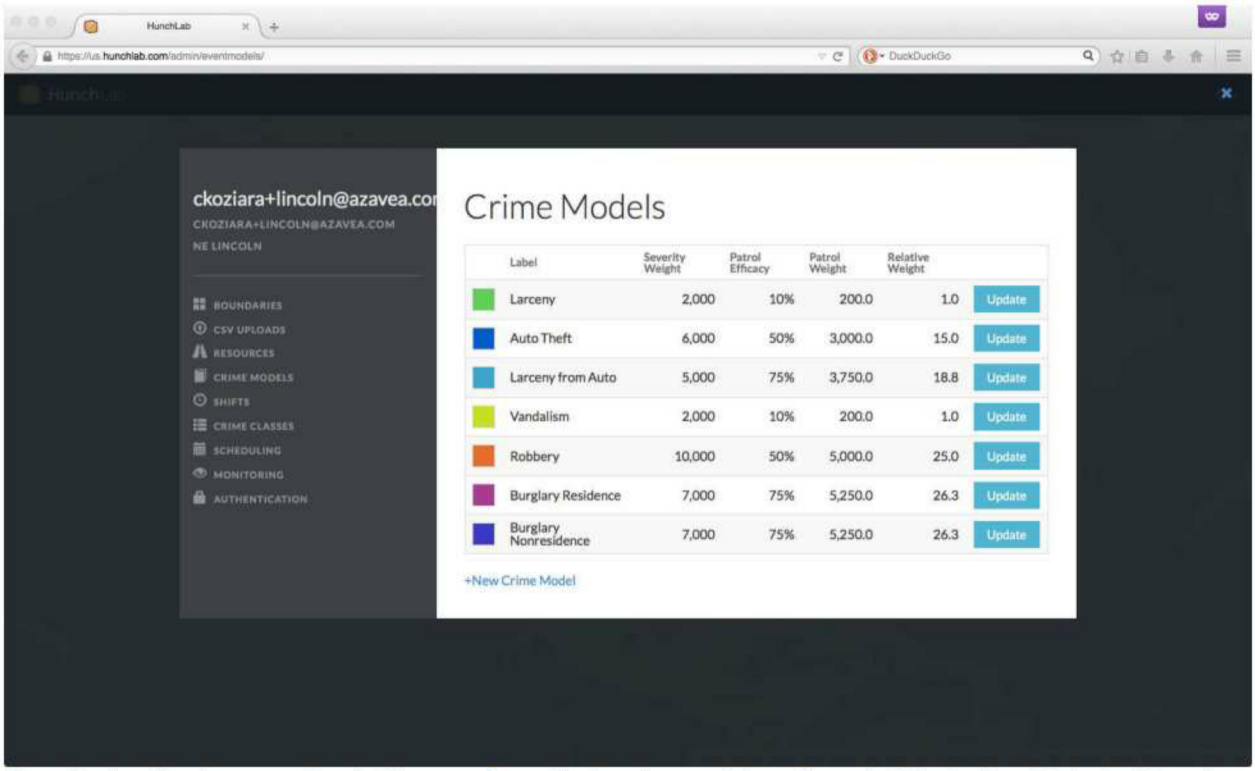


Figure 3 – A police department's priorities are reflected in the crime models configured within the HunchLab administrative user interface. Severity weights enable the department to tell HunchLab how important it is to prevent each type of crime. In this example the cost of crime numbers from the RAND Corporation are utilized to align policing priorities to the societal impact of crime. Patrol efficacy values enable the department to specify how much impact they believe patrols will make on each type of crime. The result is that missions show up where the most important, preventable crimes are likely to occur.

2. Sample Mission PDFs

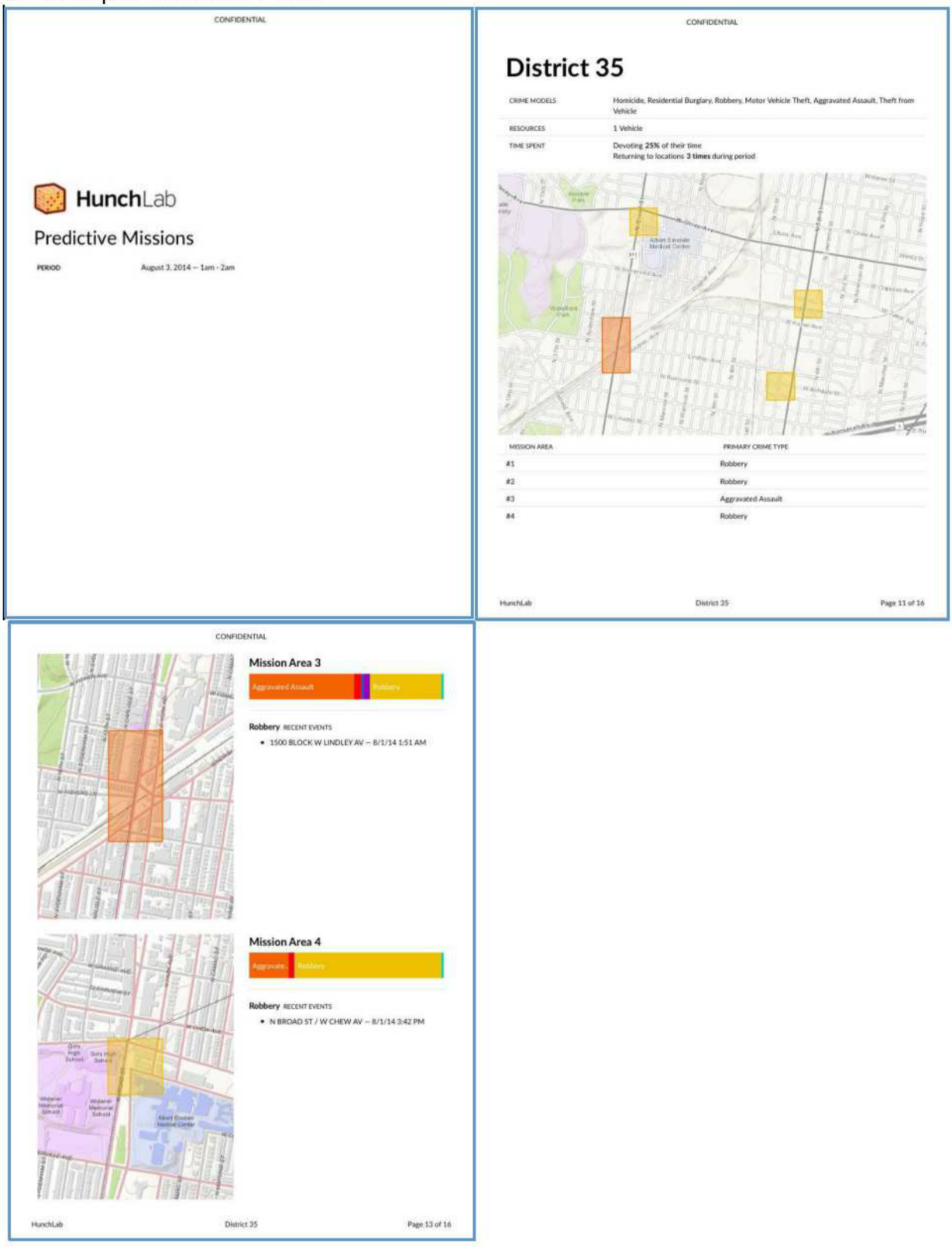


Figure 4 – Mission data can be printed or distributed as PDFs and taken on patrol if network connectivity is not available.

3. Example Model Composition

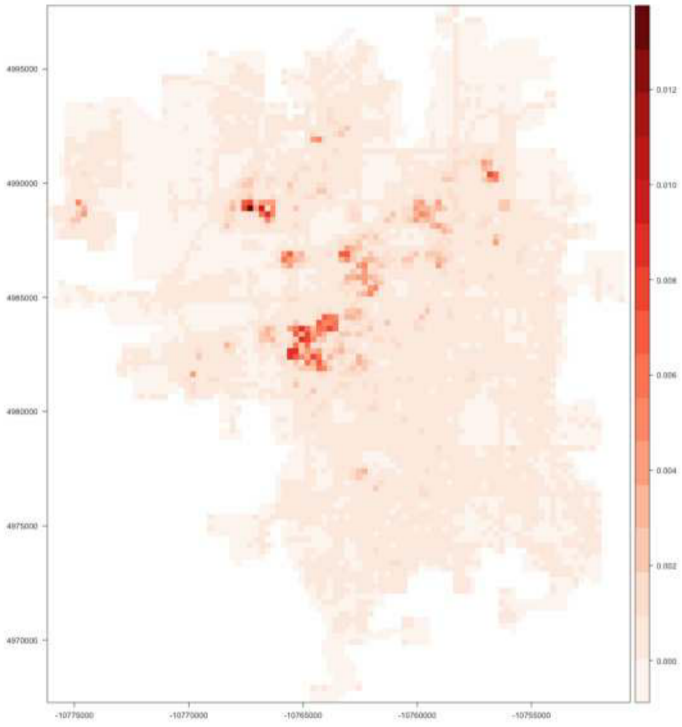


Figure 5 – HunchLab forecasts the expected count for each configured crime type within a shift for each small location within the jurisdiction. This figure shows an example of the map of one such set of forecasts.

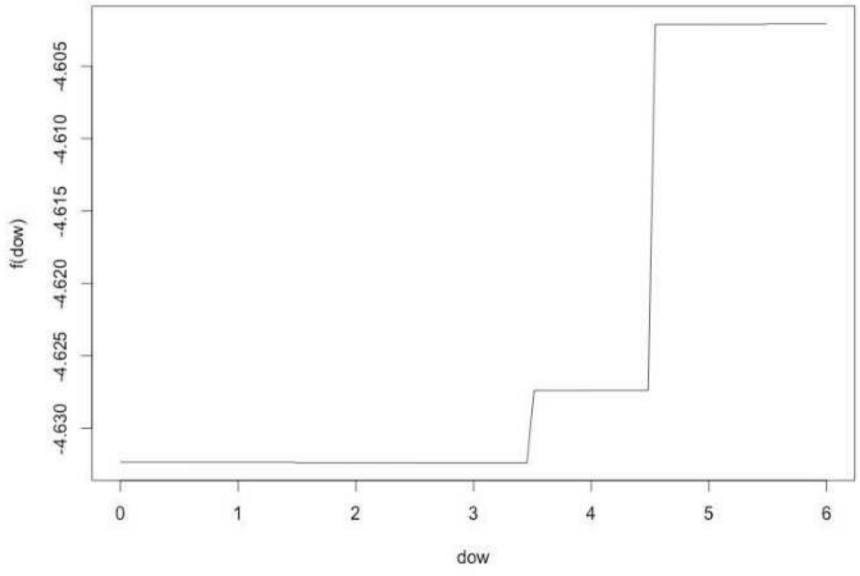


Figure 6 – The forecasting models can be examined to visualize what the system has determined effects the risk levels. In this case, the system learned how Friday, Saturday, and Sunday (4, 5, 6) have higher levels of assaults in Philadelphia.

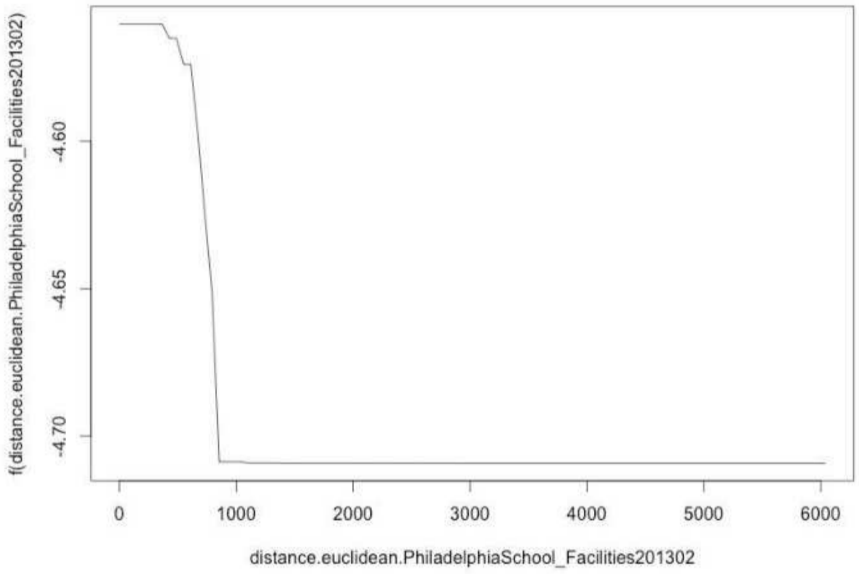


Figure 7 – In this example, the proximity to schools is shown to increase assault risks extending to about 900 meters.

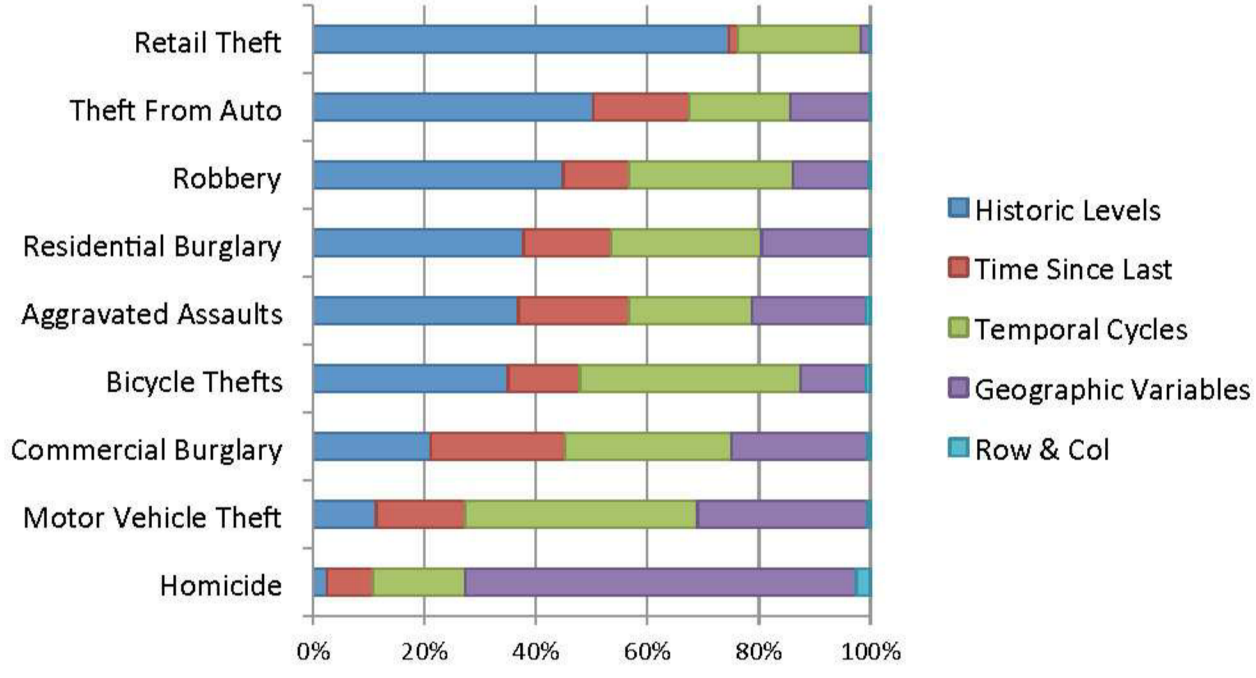


Figure 8 – In this example, the influence on predictions for each set of variables that represent differing crime theories is graphed. This information may inform the short-term and long-term tactics selected to address each type of crime. It also demonstrates the importance of representing risk via more than simply past crime locations.

Related Activities: HunchLab Advisor

HunchLab Advisor is a software service that enables departments to engage in evidence-based policing through the Field Test, Experiment, and Adaptive Tactic initiatives of the module. This module is currently under development by Azavea with statistical design completed and user interfaces being developed. As the module becomes available, Azavea will invite interested pilot clients to test this functionality without incurring additional costs.

Field Test

The Field Test component shows departments the impact of new initiatives within a specified geographic area. For example, if a department adopts a new DDACTS deployment model in a particular part of the jurisdiction, the Field Test can tell the department the likely impact of the new program. This output includes information like how many crimes the new strategy likely prevented.

Experiment

The Experiment component gives departments an easy way to set up a Randomly Controlled Trial (RCT) to test for a causal impact in a rigorous way. New initiatives that have been validated by an initial Field Test are prime candidates for a full Experiment.

Adaptive Tactics

The Adaptive Tactic component uses advanced statistics to determine the most effective tactic in response to individual crime events. Feature administrators (analysts or command staff as a department deems appropriate) can input tactics into the system and the Adaptive Tactics component will match these tactics against outcomes to determine which tactic is likely the best for a given scenario. For example, if thefts from motor vehicle are a problem in your department, you can use Adaptive Tactics to determine the best course of action to reduce future thefts. You might enter three tactics: directed patrol for three days after each event, placing flyers on cars warning of the thefts, or even no action. The system will automatically test these tactics and begin recommending the best tactic to you.

Technical Requirements

Hardware Requirements

The HunchLab application is hosted as a multi-tenant Software-as-a-Service (SaaS) application within the AWS infrastructure. Azavea will manage the hosting infrastructure, security updates, and 2nd tier support (Azavea assumes that police departments will prefer to manage direct end-user support). This cloud-based approach enables HunchLab to leverage significant amounts of computing power in an elastic manner, a critical requirement for providing the advanced statistical algorithms the system employs. Replicating a similar environment on-premise would entail a substantial outlay of capital to provide servers that are utilized only in bursts. In order to provide a secure application, we have consulted the FBI's CJIS guidelines to apply as many guidelines as possible in designing our system architecture, including risk mitigation techniques such as encrypting data connections and optional 2-factor authentication.

The application requires network connectivity from the user to the HunchLab service. Bandwidth requirements are modest, as most application assets are cached locally in the browser.

HunchLab 2.0 is hosted within the Amazon Web Services (AWS) infrastructure. AWS provides best-of-breed security and flexibility for building robust and secure SaaS applications.

Software-Related Information (Including Support and Upgrades)

HunchLab's subscription design includes application hosting, updates (fixes and new functionality within the place-based module of HunchLab 2.0), 2nd tier support, and ongoing training resources. This pricing model allows unlimited users and devices to access the application. All support services are coordinated and provided from Azavea's Philadelphia office and will include incident-based and troubleshooting support services by experienced Azavea staff through e-mail or phone during business hours, Monday to Friday, 9am – 5pm, EST (exclusive of designated US federal holidays). Additional support options outside of business hours are available for discussion as needed. Azavea will provide the same level of support during a pilot as would be provided to a long-term client.

Azavea develops HunchLab through an agile Scrum methodology whereby work is planned in 2-week increments. This structure enables us to quickly develop iterative improvements to the application. New functionality and any necessary operating system updates or patches are deployed on a schedule designed to minimize downtime. For instance, most software updates result in about 0 – 15 minutes of downtime. System updates require no work from the client, as Azavea staff manages the deployment process. The application is hosted as a multi-tenant application, so an update by Azavea for the US hosting environment will update all clients hosted within that environment simultaneously.

In order to comply with security requirements, clients are expected to continue to maintain modern, up-to-date web browsers on the devices that will be accessing the system.

Database Requirements

HunchLab does not require a client to have any particular database available. HunchLab does require event data (crimes or calls for service records) to already be geocoded and include basic attribute data such as the date and time of the event (or time range), event classification, unique identifier, etc. More information about the data interface requirements is in the section below.

Data Interface Requirements

HunchLab will consume canonical data sources, such as CAD and RMS systems. The manner in which this data is transferred to HunchLab varies from client to client. A typical process will consist of transferring records to HunchLab as an extension of existing crime mapping and analysis ETL processes. Alternatively, Azavea can configure an upload process that draws data from an ODBC connection to a read-only database view to fetch data that has changed since the last import was conducted. Most agencies schedule this import process on a daily or hourly basis, but HunchLab can also be configured to import changes on a more frequent basis, such as every few minutes. Ideally, historic data is provided for a 5-year period to allow robust predictive modeling.

Event data (crimes, calls for service, etc.) are transferred to HunchLab in a simple CSV format via a secure RESTful API endpoint. Clients can directly push data into this API endpoint or Azavea can support its use. CSV uploads contain column headers and basic attribute values, such as the location and time of an event. Formatted CSV files can also be uploaded directly via the HunchLab administrative UI.

Alternately, Azavea can fetch data from other RESTful APIs, such as those provided by ArcGIS Server. If the endpoint is available via the Internet (with proper authentication), then no on-premise utility needs to be configured as HunchLab can directly fetch updates. If the endpoint is behind a firewall, then the extraction process would be set up within the client environment to push updates to the HunchLab server.

Desktop & Mobile Requirements

Agency staff access HunchLab through a web-browser. As an advanced web-application, HunchLab supports the following major desktop web-browsers and contemporary operating systems:

- Chrome (last two versions)
  - Windows XP or newer
  - Mac OS
  - Linux
- Firefox (last two rapid release versions and supported extended release versions)
  - Windows XP or newer
  - Mac OS
  - Linux

- Internet Explorer (last two versions; IE 10 and 11 as of February 2015)
  - Windows 7 or newer (*Window Vista and older Windows operating systems do not support secure versions of the Transport Layer Security (TLS) protocol 1.2+ within Internet Explorer. To support these older versions of Windows, we recommend the use of Chrome or Firefox (both free for installation) on these machines since they do support these newer versions of TLS.)*

We also support the following major mobile browsers:

- Safari
  - iOS 7 or newer
- Chrome (current version)
  - Android
- Internet Explorer
  - Windows Phone 8.1 or newer

In all cases, the default HunchLab configuration requires operating systems and browsers that support Transport Layer Security (TLS) version 1.2. This requirement is to prevent known attacks against SSL traffic that impact TLS v1.0 and older protocols. Our supported browsers provide the correct version of TLS either automatically or with minor configuration changes (such as checking a box within the settings panel). If an agency is unable to support TLS 1.2 connections, it may necessitate the creation of a separate access point to the HunchLab system, which can be discussed on a case-by-case basis.

User Management

HunchLab can either operate in a stand-alone authentication mode or integrate with existing directory services. In either mode, users are given an application role that provides only the needed functionality to the user. For instance, officers are provided with a viewer role to consume missions. Analysts and IT administrators can be giving access to administrative settings with the application.

To delegate user management to external systems, HunchLab supports the SAML single sign-on (SSO) standard. For instance, a police department's Active Directory system can be published through Active Directory Federation Services (ADFS) to provide a SAML compliant authentication endpoint. The department then simply creates user groups within Active Directory that represent the distinct application roles within HunchLab. Users login to the ADFS service and select "HunchLab" to be redirected into the application with the proper permissions. In this manner, HunchLab maintains no user credentials improving the overall security of the department. As the department deploys advanced authentication techniques, HunchLab benefits automatically by delegating to a centralized authentication service.

Azavea Capabilities: Law Enforcement Projects

Azavea is an award-winning geospatial software design and development company based in Philadelphia. The firm was organized in 2000 to create technologically advanced solutions for web and mobile geospatial data visualization and analysis. Azavea is a certified B Corporation, a for-profit corporation with a social mission. Our mission is to apply geospatial data and software to create more sustainable, vital and livable communities while advancing the state-of-the-art through research. Azavea provides a range of services that include:

- Web and mobile software development
- User interface and experience design
- Mapping and spatial analysis
- High performance computing
- Spatial data mining and modeling
- Research and development

The firm has designed and implemented geographic data applications for a variety of domains including: crime analysis, public safety, economic development, elections, urban forestry, humanities and land conservation.

Technology and Partners

Azavea's developers work with a broad range of tools and have particularly strong backgrounds with the .Net, Java, Python, Django and Scala frameworks. The firm has also established a number of strategic partnerships to enhance our capabilities.

Azavea is a member of the Amazon Partner Network. HunchLab is hosted in the Amazon Web Services environment and was a 2014 Amazon Web Services Partners in Innovation award winner.

Azavea is an Esri Business Partner and has several years of experience with development and deployment on the ArcGIS platform with dozens of applications implemented on the ArcGIS Server, ArcGIS.com and ArcIMS products. Azavea was named ESRI Business Partner-of-the-Year or Foundation Partner-of-the-Year in 2006, 2007 and 2010. In addition, Azavea is a Microsoft business partner with substantial experience developing the .Net Framework, SQL Server and Windows Server platforms.

In addition to commercial toolkits, Azavea staff is experienced creating web software solutions that use online API's such as GoogleMaps, Bing Maps, ArcGIS Online and OpenStreetMap. The firm also works with a range of open source tools that accelerate and lower the cost of our software development work. In particular, Azavea has a great deal of experience with creating solutions that bring together the

strengths of both commercial and open source toolkits to create high quality and visually attractive applications. The firm not only has experience with open source solutions, but also contributes to them, including significant contributions to OpenLayers and PostGIS.

User Experience Design

Azavea takes great pride in the development of user interfaces that are simple, easy-to-use and are crafted for the specific purpose at hand. Our talented developers and designers work with each client to develop applications that aren't simply functional, they are simple and beautiful.

Commitment to Community

Azavea is committed to working on projects with a strong social value component. Each of Azavea's projects, products and pro bono engagements showcases this commitment. We seek out projects that enhance communities, foster economic development and improve decision-making. Further, we perform research to advance the state of the art.

Azavea R&D

Azavea has an active research and development program through which the firm invests substantial resources toward the development of new solutions and techniques. Each employee is encouraged to develop a personal research project that will both engage the employee and extend the capabilities of the organization. Current research projects include: [redacted] While not all of these research projects results in measurable commercial success, they are an important part of a culture at Azavea that encourages and takes pride in innovative applications of geospatial technology.

Representative Projects and Services

Azavea has several years of experience conducting business with a variety of law enforcement agencies at the local, national and international levels. We have already executed a number of projects that are related to the proposed effort including:

Products and Solutions

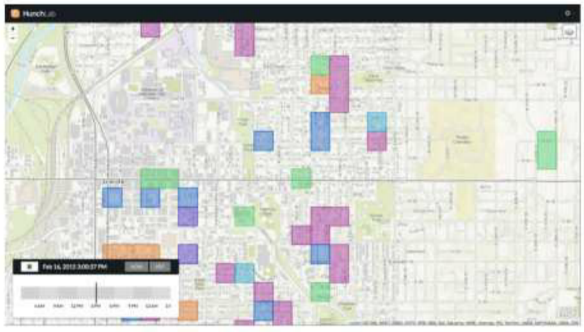
In addition to Azavea's professional services work, the firm has developed several web-based solutions. These solutions have enabled us to develop a broad range of reusable components and software tools that can be applied to many different scenarios. For law enforcement, these include:

HunchLab ***Next generation predictive policing***

Developed with support from the National Science Foundation, HunchLab's advanced software combines many data sources into a unified forecast of crime risk to support effective resource allocation.

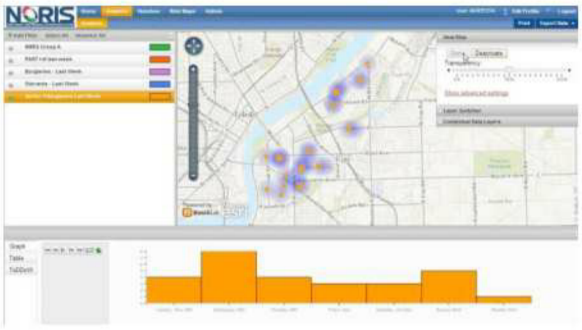
Municipal and County Applications

Lincoln, Nebraska HunchLab 2.0:



The Lincoln Police Department routinely uses a range of community policing tactics other than responding to individual incidents, such as: targeted saturation patrol, bicycle and foot patrol, undercover/plainclothes/surveillance operations, educational presentations, and coordination of efforts with other government or human service agencies. The Lincoln Police Department has shared crime data with Azavea to develop a HunchLab 2.0 instance that underscores HunchLab's ability to meet community policing initiatives.

NORIS HunchLab 1.0



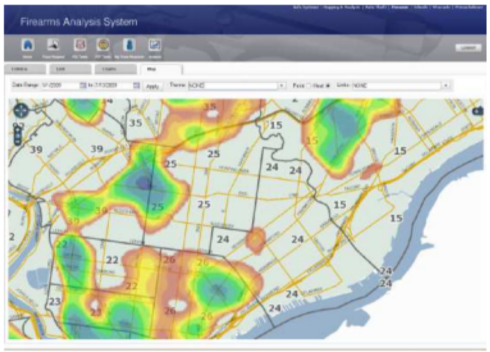
The Northwest Ohio Regional Information System (NORIS), a unit of the Criminal Justice Coordinating Council (CJCC) located in Toledo, Ohio, contacted Azavea about a crime analysis and risk forecasting solution that would enhance law enforcement officials' ability to understand the structure and patterns of crime and other public safety events.

Azavea implemented HunchLab 1.0, the firm's web-based geographic crime analysis, early warning and risk forecasting software solution. HunchLab imports incident level data from multiple sources, such as a police department's existing RMS, CAD and other non-proprietary applications. Users can view incidents from the different data sources either separately or in comparison to one another using a map, table, graph, or other analysis output

Federal Applications

Firearms Analysis System (FAS)

The Firearms Analysis System (FAS) was developed to reduce the amount of time required to trace a firearm seized in a crime and to provide greater analytical ability to look at all such crime guns seized in Philadelphia. The FAS serves both as a workflow / data entry system and as a mapping, charting, search, and reporting system. The Philadelphia Police Department worked closely with the Philadelphia Division of the Bureau of Alcohol, Tobacco, Firearms and Explosives (ATF) and Azavea to develop the application.



Current Research

Predictive Modeling of Long and Short Term Crime Risk

Azavea is working with Dr. Jerry Ratcliffe and Dr. Ralph Taylor at Temple University to develop a methodology to combine long-term risk prediction from underlying socio-demographics with event-created near-repeat risk. Funded by the National Institute of Justice, this ongoing project will create a free software tool that will enable police departments to use their geocoded crime data in combination with freely-available census data to create micro-spatial estimates of future criminal activity at the local block level. As part of this project, Azavea developed the ACS Alchemist, an open source tool (https://github.com/azavea/acs-alchemist) that can help extract specific portions of the American Community Survey (ACS) for use in law enforcement and other GIS research projects.

Other Applications

- *Philadelphia Crime Analysis and Mapping System (PhiCAMS)* – An extensive suite of web-based tools for mapping, visualizing, searching and reporting on crime incidents including integration with arrest and court records, warrants, prison releases and government administrative data.
- *PA Office of Public Safety Radio* – Created a web-based application for spatial models for displaying the status of public safety radio across the state.
- *University of Pennsylvania Office of Public Safety* – Developed applications for geocoding, managing and synchronizing data feeds with the Philadelphia Police Department.
- *International SOS Security Alert Services* – Developed a set of web services that enable global security and medical alerts to be displayed and compared on a map.
- *CrimeBase* – Web-based mapping and analysis tools the enable the public to display and analyze more than 10 years of geographically aggregated crime data.

Appendix 1: Architecture & Security of HunchLab 2.0

Introduction

HunchLab is a web-based application provided under a software-as-a-service (SaaS) model. While the SaaS model of software deployment abstracts architectural decisions behind a simple client-facing web application, we realize that transparency is necessary within the law enforcement community. This document outlines the architecture of the HunchLab application with an emphasis on application security and availability.

Infrastructure

HunchLab 2.0 is hosted within the Amazon Web Services (AWS) infrastructure. AWS provides best-of-breed security and flexibility for building robust and secure SaaS applications.

Security

AWS data centers maintain strict physical access controls including 24x7, trained security. Authorized staff must pass two-factor authentication a minimum of two times to access data center floors. AWS staff members pass criminal background checks prior to employment.

Further, the AWS platform regularly passes third-party evaluations. AWS has achieved ISO 27001 certification and has been validated as a Level 1 service provider under the Payment Card Industry (PCI) Data Security Standard (DSS). AWS annually publishes SOC 1, 2 & 3 audits. AWS is also a FedRAMP Compliant Cloud Service Provider (CSP) with validation at the Moderate level. This validation covers both the regular US regions and the GovCloud region. AWS has been successfully evaluated at the FISMA Moderate level for US federal government systems as well as DIACAP Level 2 for US DoD systems.

AWS Compliance information: *http://aws.amazon.com/compliance/*

Availability

The AWS platform provides robust services to maintain application availability even in the face of infrastructure failure. Within each AWS region, multiple availability zones allow an application to remain available even with the complete failure of an individual data center. Power and network connectivity systems are designed for redundancy with onsite backup power generation.

The HunchLab application is designed to use multiple availability zones within a region to provide availability even in the face of the loss of a complete zone. For instance, if a client's HunchLab application is hosted within the US East region, client data is replicated between multiple availability zones within the region. Availability zones are independent data centers within the region. The application is designed to survive the complete failure of an availability zone (a complete data center) without manual intervention by Azavea.

For US clients of HunchLab, client data can be additionally backed up to a separate US data center. This allows Azavea to redeploy the application within this backup region in the event of a sustained loss of an entire region of the AWS infrastructure. Since there is only one AWS region within the EU, Azavea cannot provide this capability for EU clients while maintaining data residency completely within the EU.

Data Residency

Distinct AWS geographic regions allow applications to be deployed to different parts of the world. This allows HunchLab clients to select a region based upon applicable privacy laws. Data placed within a region is not automatically replicated to other regions by AWS.

Clients can select from residency in:

- United States
- European Union (Dublin, Ireland)
- Japan
- Australia

Access

Logical access to the HunchLab AWS hosting account is limited to Azavea personnel working on the application. Access to the infrastructure is granted via 2-factor authentication using individual credentials for each employee. System development and testing occurs in a separate hosting account so that contact with client data is minimized. Client data is not copied outside of the AWS infrastructure without the explicit consent of the client. Statistical models and other diagnostic data that does not include disaggregated criminal justice information (CJI) may be accessed and examined outside of AWS by Azavea personnel for troubleshooting and support purposes.

More details of the AWS platform can be found in the current version of the *Amazon Web Services: Overview of Security Processes* document available for download at http://aws.amazon.com/security/.

Application Security

Azavea has a long history of handling sensitive law enforcement data sets. The new version of HunchLab is delivered as a secure cloud-based subscription service. As we designed this new version, we focused on incorporating security best practices into our development process. While most deployments of HunchLab contain local department data sets that do not technically require compliance with the FBI's Criminal Justice Information Systems (CJIS) guidelines, we are using the CJIS requirements and recommendations to guide our decision-making process and system architecture. Here are some of the security features and policies available within the new HunchLab.

Data Use & Security Agreement

By default, Azavea agrees to solely use the law enforcement data to provide the agreed upon HunchLab service to the department including using the data for system testing, troubleshooting, and lives operations. Separately, Azavea may seek permission to use the data for research purposes that further the product and crime analysis in general. At no time will Azavea hold any claims to the data nor will Azavea use the data for other commercial purposes. Upon written request, Azavea will purge a customer's law enforcement data from its operational systems. Deletion from operational systems will occur within 7 days. The application maintains automated backup files for the last several weeks. Client data will expire out of these automated backups within 28 days from the request. If requested, Azavea will certify that client data has occurred.

Azavea will gladly sign a CJIS Security Addendum as specified in CJIS v5.3 section 5.1.1.5.

Security Awareness Training

Azavea hires technical staff with an eye toward building reliable and secure web applications. Part of the Azavea onboarding process is acknowledgement of company security practices as well as signing a separate agreement regarding confidentiality of client data. Additionally, staff members with access to the HunchLab system undergo biennial training on best practices when dealing with criminal justice information as outlined in CJIS v5.3 section 5.2.

Reliability & Security Incident Management

The HunchLab service is designed to be resilient to failure with redundancy built into the system architecture. Additionally, Azavea has implemented automatic monitoring of system uptime and incident alerts to provide timely resolution of system issues. In the event of a suspected or confirmed security breach, Azavea will proactively notify the law enforcement agency of the breach in a timely manner as specified in CJIS v5.3 section 5.3.2.

System Auditing

The HunchLab system keeps a running system log of activity by users including log-on attempts and information retrieval. These records are retained for at least 365 days. The auditing system is designed to comply with CJIS v5.3 section 5.4. Additionally, Azavea employs AWS services that log the logical access and control of the AWS environment.

Role-based Security

Access to system functionality is restricted based upon security roles. For instance, only a few users need administrative access to the system. This approach reflects the guidelines in CJIS v5.3 section 5.5.2.

Authentication Credentials

HunchLab can delegate credential management to 3rd party directory services such as Active Directory through the SAML standard. In that case, HunchLab assumes that the 3rd party directory service provides a CJIS compliant security model. Additionally, HunchLab can provide a stand-alone authentication system that complies with both the standard authentication and advanced authentication specifications in CJIS v5.3 sections 5.6.2.1 and 5.6.2.2. Our advanced authentication option provides 2-factor authentication using time-based tokens generated locally by mobile applications for mobile devices. Additional costs may apply if Azavea is managing 2-factor authentication on behalf of the client.

Password Management & Login Failures

If operating in stand-alone authentication mode, HunchLab stores user passwords in a salted cryptographic hash format which increases the computing power necessary to reverse engineer a user's password even if our database is comprised. Additionally, to prevent external attacks on user credentials, the system keeps track of unsuccessful login attempts and locks the account for progressively longer periods of time. This policy is recommended in CJIS v5.3 section 5.5.3.

Session Lock

When a user logs into HunchLab, a temporary security token is kept within their local browser memory. HunchLab assumes that devices logging into the system will employ session locks or screensavers that meet the guidelines in CJIS v5.3 section 5.5.5.

Data Protection

The HunchLab service is hosted within Amazon Web Services (AWS) data centers. These data centers implement state-of-the-art security practices that protect the physical access to data within HunchLab as recommended in CJIS v5.3 section 5.9. Additionally, AWS continuously monitors their infrastructure against denial of service attacks and penetration vulnerabilities.

Within the HunchLab architecture, Azavea has utilized several security features of the AWS platform to harden the system. For instance, all inbound traffic to HunchLab is encrypted via SSL and terminates at a set of load balancers. These load balancers only allow secure HTTPS traffic with specific versions of the TLS protocol (TLS 1.2+) and specific encryption algorithms (AES) and proxy all traffic to the application. Each component of the application is isolated from all others with only the minimum required network traffic for each server instance granted. This security is enforced as inbound and outbound firewall rules on each server as well as redundantly at the network level.

While the physically secure AWS infrastructure constitutes a physically secure location and therefore encryption is not required, Azavea has decided to encrypt data in transit and at rest as much as feasible. All data in transit within the application is encrypted. Data stored on the elastic block storage devices attached to HunchLab servers and within the AWS S3 service is encrypted at rest. Additionally, data stored in the relational database provided by Amazon RDS is encrypted at rest.

These design approaches seek to conform to CJIS v5.3 section 5.10.

[Note: As of February 2015, there are three pending security features referenced above. We have not yet enabled encryption of temporary files stored on EBS volumes. These files exist for only as long as necessary to process the data and are then deleted. When these transient volumes are released by the application, any remaining data is proactively destroyed by AWS. Second, we have added a caching layer within the application. This presently transmits data within the secure environment without encryption. We are working to encrypt the data being stored within the cache. Finally, AWS recently released support for encryption at rest within Amazon RDS. Our deployment of this feature is imminent.]

Personnel

Upon request, Azavea will cooperate with the screening of Azavea personnel with access to the HunchLab system in line with CJIS v5.3 section 5.12.

CJIS Policy v5.3 Review

The following review of CJIS Security Policy version 5.2 outlines how HunchLab aligns with these guidelines.

Section	Requirement	Alignment
4.1	Defines Criminal Justice Information (CJI)	The required data set within HunchLab consists solely of crime event data. This data set does not include personally identifiable information. The most sensitive component of the data set is the location of incidence, but this section of the CJIS guidelines exempts property data when it is not accompanied by PII. As such, CJIS does not technically apply.
5.1.1.5	Private contractors are subject to the CJIS Security Addendum when handling CJI.	Azavea will gladly execute agreements in regards to the handling of CJI.
5.2	Security awareness training shall be required within six months of assignment and biennially thereafter for all personnel with access to CJI.	Azavea already conducts new employee briefs on guidelines and responsibilities in handling client data. Specifically to the team responsible for HunchLab, we are implementing focused training to comply with the minimum topics outlined in the CJIS guidelines.
5.2.2	Records of security training	Azavea shall keep records of security training for staff involved in HunchLab projects.
5.3.1	Security events shall be promptly reported.	Azavea shall promptly report security related events to the relevant clients.

5.4.1	Information systems shall generate audit records for specified events.	The HunchLab API logs user interactions that include the event types specified within the CJIS guidelines. Additionally, the AWS environment generates audit logs of management interactions with the hosting environment through the use of the AWS CloudTrail service.
5.4.3	Audit monitoring shall be conducted at minimum once a week by designated personnel.	The HunchLab environment generates system alerts upon suspicious activity with a view toward maintaining continuous monitoring of suspicious activity. For instance, increased levels of API requests that fail authentication generate alerts to the HunchLab team.
5.4.5	Protection of audit information from modification, deletion, and unauthorized access.	AWS level audit logs are kept in a secure S3 bucket with modification and deletion access limited to a subset of the HunchLab team. HunchLab API audit logs are kept securely within the hosting environment and end users are prevented from modifying or deleting these records.
5.4.6	Audit records shall be retains for at least one year.	Azavea will retain audit logs for at least one year.
5.5.1	Account management shall be in place to validate system accounts and permissions.	HunchLab client agencies manage user access to the system. Administrative access to the hosting environment by Azavea staff is reviewed regularly with only members of the team granted access.
5.5.2	Access enforcement shall be enforced to limit access to privileged functions.	HunchLab application functions are accessible via role-based system that limits access to administrative features within an organization's account. Additionally, components of the HunchLab application are only granted permissions within the AWS environment for systems that they need access to.
5.5.3	Unsuccessful login attempts shall be limited to no more than 5 consecutive invalid attempt per user followed by an automatic lock on the account for 10	This login restriction is in place within the HunchLab application.

	minutes.	
5.5.5	Session locks shall be in place to prevent access to the system after inactivity.	HunchLab assumes that client managed devices will implement screen locks or appropriate measures to meet this requirement.
5.5.6	Remote access shall be monitored and controlled.	By its nature a cloud service provides access over an untrusted network. Access to the application is controlled through login requirements. Access to the hosting environment itself is severely limited and requires multi factor authentication and cryptographic keys.
5.6.1	Identification policies should uniquely identify each user or administrator of the system.	All HunchLab users login with a unique identifier. The AWS environment is also managed through unique credentials assigned to each Azavea team member.
5.6.2.1.1	Passwords shall comply with stated attributes.	The AWS environment is managed through unique credentials assigned to each Azavea team member. These credentials include a password (that meets the stated requirement). HunchLab users can have password restrictions assigned to their accounts. Alternatively, if HunchLab is delegating authentication to another system, then that system would enforce such requirements.
5.6.2.2	Advanced authentication is required for publicly accessible services where the authenticity or security of the requesting device can not be established	HunchLab can either provide 2-factor authentication to end-users directly or can delegate authentication to a client agency to provide a compliant authentication methodology. The AWS environment requires both a password and token to be entered for Azavea staff to access the hosting system.
5.8.1	Electronic and physical media shall be stored within physically secure locations. If not, then the data shall be encrypted.	The AWS hosting environment is a physically secure environment therefore data encryption is not required.
5.8.3	Electronic media shall be sanitized prior to reuse or disposal.	Media within the AWS hosting environment is sanitized before allocation to new customers. Additionally, AWS destroys all media that leaves its data centers for disposal.
5.9	Physically secure locations shall meet stated guidelines.	AWS provides details of its security policies for its data centers. Even Azavea as customers of the

		service are not permitted physical access to the environment.
5.10.1	The network infrastructure shall control the flow of information between connected systems.	The HunchLab application is comprised of distinct functional units. Each unit can only speak to the other units of the application that are necessary for it to complete its functions. Each server has a firewall that only allows inbound and outbound communication as needed. Additionally, the network enforces traffic controls to specific allowed ports. External access to the environment is limited to a single bastion server accessible only by Azavea.
5.10.1.2	Encryption shall protect data outside of the boundary of a physically secure location when being transmitted or encrypted. Cryptographic modules shall be certified to meet FIPS 140-2 standards	External access to HunchLab is over HTTPS. The application only permits TLS 1.2 due to flaws in earlier TLS versions. The server is configured to use either 128/256bit GCM or 256bit CBC AES encryption and prefers ephemeral key exchange that provides forward secrecy (ECDHE). The application uses Amazon's Elastic Load Balancers to terminate inbound SSL connections. These load balancers do not use certified cryptographic modules unless the application is hosted within the GovCloud environment.
5.10.1.3	Intrusion detection shall be implemented.	AWS manages intrusion detection and abuse of their environment. Additionally, HunchLab logs inbound requests to monitoring servers that provide Azavea staff with a real time view of activity.
5.10.1.5	The metadata derived from CJI shall not be used by any cloud provider for any purposes.	Azavea will not use CJI for any purposes other than to provide this service. AWS also agrees to not use client data for any purposes.
5.10.3.1	Partitioning shall separate user functionality from information management functionality.	The HunchLab application is broken up into discrete segments that separate functionality. For instance, an inbound request for data first arrives at a load balancer, which terminates the inbound SSL connection, parses the request, and then wraps the request in a new SSL connection to pass to the web servers. The web servers then receive the request, validate the user's credentials and query the database for needed

		data. The database also resides on a separate virtual machine.
5.10.3.2	Virtualization shall be implemented to isolate machines.	Firewalls are in place that restrict access to each machine within the environment. For instance, the load balancers may not directly communicate with the database server. Requests from the load balancers to the web servers and from the web servers to the database server are encrypted. Requests from all servers to the S3 object store are all enforced as encrypted. Log files on each machine are centrally aggregated for monitoring.
5.10.4.1	Patches shall be maintained.	Azavea maintains a staging environment to validate updates to software. The application utilizes OS releases that are currently supported with security patches. OS security patches are applied upon each deployment of the software via golden images for machines. Additionally, Azavea updates other software packages on a regular basis based upon the severity of the patch.
5.10.4.2	Malicious Code Protection	HunchLab utilizes only Linux based software. It is atypical to run antivirus software on such systems due to the security design of the systems. Additionally, the hosting environment is designed for the rapid replacement of server instances based upon golden images. For instance, no persistent data is stored on web application servers. Upon every deploy of an update, the existing servers are destroyed and replaced with new servers running from a clean image. This approach eliminates the likelihood of an infection of maintaining itself.
5.12.1	Personnel will have fingerprint-based record checks.	Azavea is happy to have relevant staff cleared through these processes.
5.12.2	Upon termination, access to CJI shall be terminated immediately	Azavea maintains a checklist of termination practices, which includes removal of access to the HunchLab environment.

Application Architecture

Overview

The HunchLab application is hosted as a multi-tenant SaaS application within the AWS infrastructure. The application leverages a broad array of open source projects including operating systems, application frameworks, and statistical packages. Further, the application leverages AWS-specific technologies to provide scalability, redundancy, and security. Finally, the application is architected into discrete tiers allowing the logical separation of components.

Open Source Technologies

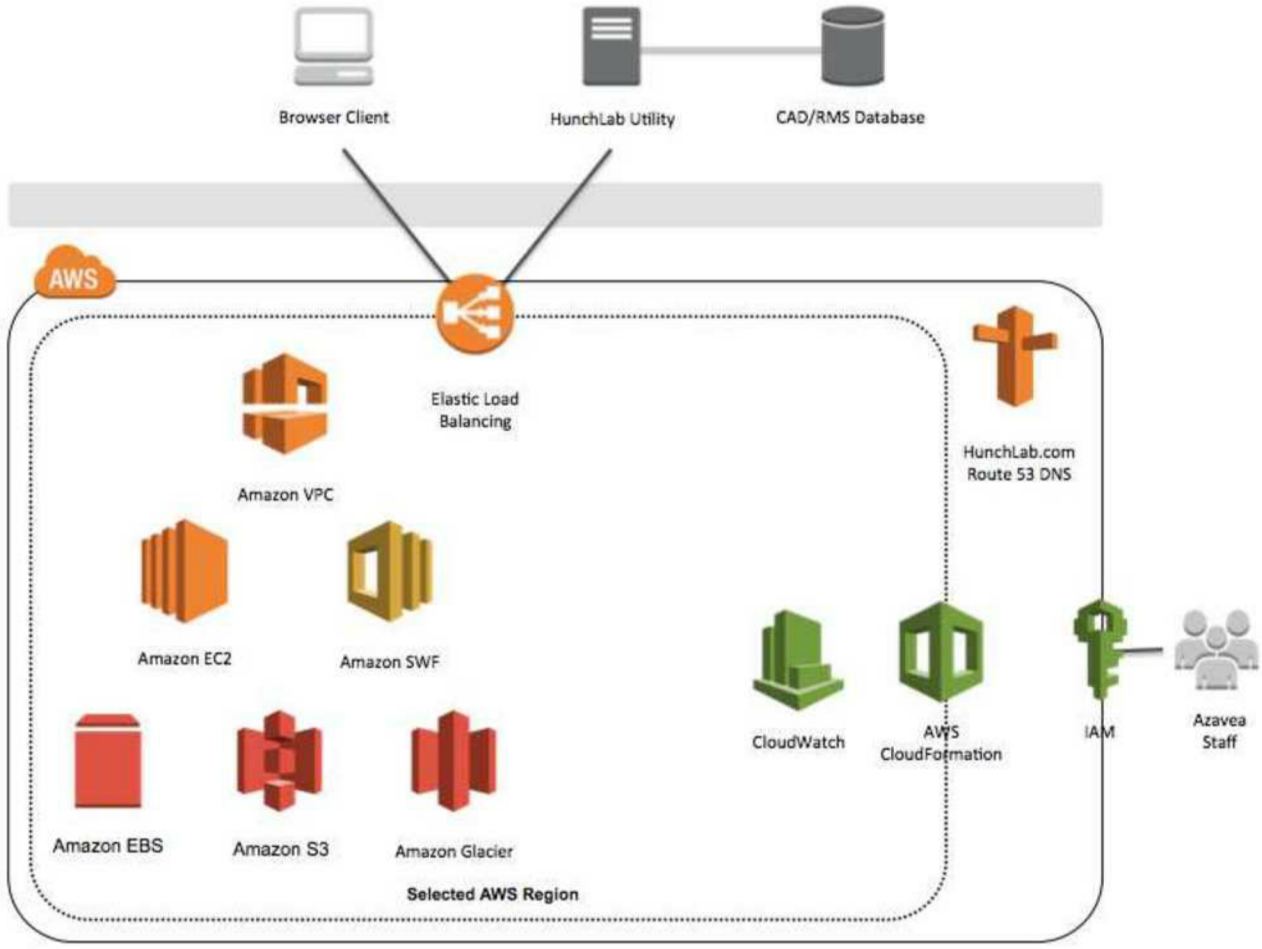
The application consists of a client-side, standards-based web GUI application implemented in JavaScript using the Angular JS framework. This GUI application speaks to a set of RESTful APIs implemented in the Django web application framework with data persistence provided by an AWS-managed PostgreSQL database with geographic queries supported by the PostGIS extension. Additionally, the system uses Azavea's open source GeoTrellis framework for high performance geographic processing and the R framework for state-of-the-art machine learning algorithms.

Amazon Web Services Technologies

The HunchLab SaaS application was designed to take advantage of the breadth of AWS services to provide a secure and scalable application. The application uses the following AWS technologies:

- Route 53
  - DNS for the hunchlab.com domain is managed through the distributed and redundant Route 53 service.
- Virtual Private Cloud (VPC)
  - VPC allows the isolation of application components on individual subnets, enforces network-level traffic rules, and provides both inbound and outbound firewalls.
- Elastic Load Balancing (ELB)
  - The SSL encrypted web traffic for the application is terminated by elastic load balancers which provide secure management of the signed HunchLab SSL certificate and performance under increased application loads.
- Elastic Compute Cloud (EC2)
  - Servers provided by EC2 are used for the web application, database, and machine learning tiers of the application. Many AWS services utilize EC2.
- Elastic Block Storage (EBS)
  - EBS volumes back the root partitions of EC2 instances and are used to store client-specific data.
- Relational Data Store (RDS)
  - RDS provides a managed relationship PostgreSQL database to HunchLab. The RDS instance is configured for real-time replication and automatic failover between availability zones.

- Simple Storage Service (S3)
  - Additional application artifacts are stored in the S3 service using the server-side encryption option. These artifacts include data sets undergoing processing, analytic models and results, and backup files.
- Glacier
  - Long-term backup archives are hosted in the Glacier service.
- ElastiCache
  - An in-memory application cache is provided by the Redis functionality of ElastiCache.
- Simple Workflow Service (SWF)
  - Machine learning and prediction processes are managed via the SWF service allowing distribution of tasks among a cluster of compute instances that scales to meet client needs.
- CloudWatch
  - CloudWatch metrics and alarms are used to scale application resources to meet demand and to notify Azavea staff of failures.
- CloudFormation
  - CloudFormation is used to securely manage and update the application stack with discrete application components isolated from one another and designed to automatically scale to meet user load.
- Identity and Access Management (IAM)
  - IAM is used to provide individual credentials to Azavea staff tasked with supporting the application. IAM security policies require the use of 2 factor authentication tokens when interacting with the AWS infrastructure. Additionally, IAM security roles are used within the application stack to provide credentials to application components.
- CloudTrail
  - CloudTrail provides audit logs of interactions with AWS management commands. These logs are stored securely within S3.



Application Components

The HunchLab SaaS application is designed as a set of loosely coupled components that work together to service the user. The main components of the application include:

- Client-side
  - Browser-based application
    - JavaScript application that provides the graphical user interface
  - HunchLab data upload (varying formats based upon client needs)
    - Data integration utility for crime data
- Server-side
  - Web application tier
    - Serves static files and provides RESTful APIs consumed by the browser application and integration utility
  - Geographic processing tier
    - Conducts geographic processing to support requests from the web tier
  - Machine learning tier
    - Batch processing for creating statistical models and generating crime predictions

- Persistence tier
  - Relationship persistence for the web tier and file persistence for objects shared among tiers