

ASU CIC - Albuquerque PD - Data Lake Press Release

Headline:

APD TACKLES CRIME WITH DATA

APD launches all in one data system designed to help officers fight crime.

Disclaimer: For visioning purposes only, document may not reflect current state of project

Dateline and Overview:

Albuquerque (News Wire) July 1, 2020 The City of Albuquerque Police Department announced today the launch of new public safety data capabilities designed to improve collaboration among law enforcement organizations and help reduce crime in Albuquerque. The new Duke City Heartbeat Data Lake brings key police data together in one place for the first time making it faster and easier for police officials to collect, access, share, analyze and report.

Leader Quote:

“City Hall is committed to the safety of Albuquerque residents and data is a powerful tool to help prevent and respond to crimes,” said City Councilor Tara Bite. “With our new data lake, Albuquerque has harnessed the power of data to give our police force new super powers to fight crime.”

Problem + Opportunity:

Before the data lake, the police department had to track down and find the data they needed, figure out how to access and import it, standardize it so it could be used all while navigating multiple databases and platforms both inside and external to the department. With the data lake the universe of data sets the police use all live in the cloud where the data is secured and easy to access and use.

Customer Quote

“Being a Crime Analyst is a difficult, yet rewarding job. I used to spend the majority of my time acquiring data, running multiple queries from multiple sources, but now I get to do what I love- analyze!” said Anna Littic, ABQ PD Analyst. “Since I switched to the Duke City Heartbeat, I have become more efficient in my work! Instead of fighting data, I use data to fight crime. If you need to have all of your data easily accessible, DCHB is for you! Highly recommend it!”

Customer Experience:

Police officers and staff are now able to log into the data lake and access the data they need all in one place. Because the data is in one place, analytic and reporting applications can access and run more accurate queries in real time allowing officers to access mission critical information when they need it. Because the data lives in the cloud, the department’s applications are more reliable and no longer crash or freeze during complex queries. All of this saves time and money and improves the capabilities of the police to reduce crimes.

Call to Action

For more information on the Duke City Data Lake, please visit www.albuquerquepd.com/datalake.

###

Albuquerque Police Department - Duke City Heartbeat Data Lake

Frequently Asked Questions (FAQ)

1. How can you search for information in different formats?

Free text searching enables the user to comb through multiple fields of structured and unstructured data from multiple datasets. Users can use boolean search criteria and fuzzy matching to find information that does not match exactly but is similar. Users can also select which fields and which datasets to search (i.e. names, aliases, date of birth etc) instead of searching through the entire dataset.

2. How to include databases APD does not control?

Information sharing agreements give APD federated access to the databases that they do not control. They are able to search through the data without altering it in any way. Information that is pertinent can be pulled into the APD data set for future use. To start it will be all internal data.

3. How do we partner with outside agencies and who reached out?

Using their identity and access management model, APD will be able to grant access to partner agencies and organizations. Access controls dictate which datasets and information can be accessed by individual users or organizations.

4. How is security handled with the databases?

Data will be encrypted in transit and at rest with CJIS-compliant encryption. Encryption will be managed by APD via KMS and rotated regularly. Full auditing capabilities will be available to see who is accessing data as well as changes and updates made to the data. Versioning will enable previous iterations of data to be restored if data is corrupted or lost.

5. How does access management work?

AWS provides an Identity and Access Management tool (IAM) which allows data owners to control who has access to the data. Also controls at the application level will allow more granular access controls.

6. Who maintains the master database?

The Duke City Heartbeat Solution will consist of a data lake and not a master database. The data lake will be controlled by the ABQ PD Information Technology Department. The individual data sets will be managed and controlled by the data stewards. The data is ultimately owned by the citizens.

7. How secure is the system?

The system will be very secure. Using the AWS Cloud offers several layers of security under a shared responsibility model. AWS is responsible for the security off the cloud and ABQ PD will be responsible for the security of what's put in the cloud. ABQ will be able to encrypt data if desired and control who has access. AWS will not be able to see or access the data.

8. What data sources are included in the system?

The Duke City Heartbeat Data Lake will include data from the Police Department, other city agencies, other governmental jurisdictions like other municipal, county, state and federal and NGO sources.

The data can be sourced and organized around police department use cases. For example the traffic safety data needed to manage congestion, crashes and parking could all be aligned so users have the data they need in one place.

9. Who is responsible for data uploads and maintenance?

Data will be automatically uploaded into the data lake once the connection strings are set up. The ABQ PD IT Department will assume responsibility for maintenance of the overall data lake.

10. How often is data updated?

How often the data is updated will be set by the requirements of each group that will be using the system. Data will be updated in near live time ideally (as needed). Data will be updated when the original data source is updated by the users of that system. If data transfer costs become prohibitively expensive the frequency of the data transfer can be scaled down.

11. Who is responsible for quality assurance and quality control?

Data stewards of the individual datasets will be responsible for quality control of the data going in. ETL (extract transform load) tools will help normalize the data and address areas where data is messy or flawed and help present where duplicate data exists.

12. Who has access to the data?

Data access will be controlled through Identity and Access Management (IAM) controls set by data managers depending on the sensitivity of the data. Presumably some data and reporting will be open to the public as open data later on in the development of the system. More granular controls can be set at the application level.

13. How do you pull reports and query from the system?

Free-text or "Google-like" searches will be available to search for and display data. Users will also be able to create more complex search queries using drop down menus to choose what fields and datasets to search. Users will also be able to create persistent queries and alerts to be made immediately aware of a subject of interest. Custom reports and BI visualizations will be available that can be produced regularly or on demand.

14. What is the cost to build and maintain and who pays?

The ASU CIC and AWS will aide the ABQ PD in estimating the costs to build and maintain the solution. The cost will depend on the final project scope.

15. Is the system compatible with other datasets and software?

The system will be designed to be compatible with other systems to support ABQ PD cross-jurisdictional needs. The use of API's and other tools will connect data sets and applications as needed.

16. How does this help to reduce crime?

Accurate and timely data and data analytics are proven tools to aid public safety leaders in making informed decisions about resource allocation, strategy and operations decisions and acting in real time.

17. What functionality is available?

The primary functionality is a data lake that brings PD data sets together to increase usability.

Additional features will include integration with PD software like reporting and other tools to improve their performance, accuracy and usefulness.

Other functionality like reporting and analytics can be added as desired.

18. Who will it help with data standardization and fixing data?

The Duke City Heartbeat will be governed by an Oversight Committee and a Data Integrity Plan to ensure the quality and standardization of all data. Data managers will be responsible for complying with data integrity requirements and the system will have tools and review procedures for identifying data issues and resolving them.

19. How do you know the data is accurate?

Having all of the required data sets in one place will make it easier to run reports and control data integrity. The time saved from having the data lake will allow PD staff more time to review and increase the accuracy of data and reporting. The garbage in, garbage out rule applies to the data lake. If we put bad data in it won't create good data.

20. How do I set user permissions?

The solution will include an Identity and Access Management console that allows data owners to set access permissions. It will integrate with APDs directory service to automatically create and assign user roles based on active directory.

21. How do we support interoperability?

Having a data lake supports both internal and external interoperability. Interfaces with applications will allow PD software to use data from the solutions and also connect external data sources to the data lake.

###