Research Notes:

Building an Efficient Data Vault for a Corrections Environment Using SCRUM and AGILE Techniques for Operational Business Intelligence

Gary Harmon

Chapter 1: Data Vault for Big Human Data

The Data Vault methodology allows for best practices from a variety of different components that allow for unparalleled flexibility. The problem in Big Data, as explained by the Corrections Technology Association in its presentation "Big Data and Corrections: What's the Big Issue?", is that Corrections Data is growing at an astounding rate due to the fact that nearly 1 in 33 adults in the United States is under correctional control and that "Human" data in a correctional system is different. "Human Information is made up of ideas, is diverse, and has context. Ideas do not exactly match like data do; they have distance. Human Information is not static; it is dynamic and lives everywhere. Legacy techniques have all fallen short" (Corrections Technology Association 2013). They also cite that with the mixture of data coming in from so many different sources, it creates an increased risk. This increased risk can be uniquely put to a halt by using a Data Vault. The Data Vault will allow for multiple types of data to be accounted for while providing data integrity to key business facts.

Quantity and Quality in the State Correctional System

The Criminal Justice Organizations are beginning to create more advanced data management systems to collect, store, and process large amounts of data. This enables correctional organizations the ability to lower the overall risk associated within a corrections environment by using data to make smart and analytical decisions. It further facilitates advanced management of correctional facilities by giving organizations a data-centered strategic policy. However, the sheer volume of data inside many correctional databases often poses complexity and quality concerns. The success of any major data system depends on the quality of the data that is stored in it. This quality control can be managed by the Data Vault in a specific way that allows for key business facts to remain intact while allowing for scalability based on the environment.

When looking at State Correctional Systems, 3 states have more than 100,000 inmates in custody, 15 states have 20,000-–50,000 inmates in custody, and all other State Correctional Systems have less than 20,000 inmates per state with a combined total of only 289,475 inmates (Carson 2015). This would suggest that California, Florida, and Texas, whose inmates in custody combine to an amount higher than that of 34 states combined, would have far more complex data quality issues than other

states based only on the amount of information per inmate. However, when using a Data Vault system, the complexity of the data may increase as the size increases, but the integrity and quality of the most important data will not change. The unique connection between the Inmate ID in corrections management and the architecture of the Data Vault for business intelligence allows the integrity of the Inmate ID to pass over to the Data Vault and create a unique identification tool for quality control.

Chapter 2: Correctional Environments and Their Operational Connections to Data Vaults

In order to explain the importance of the Data Vault system for a correctional environment, it is imperative that one first understands the correctional environment. The correctional environment is filled with ranging skill sets, from those with little to no technical knowledge (who often serve on operational fronts), to those with novice to medium technical skill levels who serve as support staff to operational staff. There is often a small number of people who have a technical skill set above medium. This range of skill sets is important to this application because it will not only show the value of the Data Vault in this particular industry, but will show the Data Vaults' qualities because of the way this particular industry operates.

The correctional environment revolves around two key business facts:. (1) The inmate resides in housing provided by the Government, whether it be privately monitored or not, the housing is still provided by the Government. (2) The inmate has certain physical characteristics that will not change outside of the normal scope of human progression. An inmate will age just like any other person. An inmate will have specific physiological characteristics such as eye color, hair color, weight, height, race, gender, and so on.

Now, with those two key business facts an inmate is also given a unique identifier known as a unique identification number or Inmate ID that is specific to only that inmate and no one else. This Inmate ID works as a unique identifier. The primary data source in a Corrections Data Management System or (CMS) is entered by the end user when the inmate arrives at the diagnostic intake center. At this point, the inmate's data becomes active and the inmate is given an Inmate ID. This does not happen if the inmate has been in prison before. All inmates entering diagnostic are checked by an Iris Biometric Scanner to ensure connection to correct Inmate ID. The secondary data source in a CMS is entered by users at the local facility, regional headquarters, or at the main headquarters' locations. This unique identifier serves an important purpose when presenting the idea of using a Data Vault in a correctional environment. It allows the vault to be built around one centralized unchanging ID that will hold the key to data integrity. This ID is specific to one inmate and one inmate only, which allows the Data Vault to be built around that fact. It also allows the Data Vault to serve a Small Enterprise System with only 100–200 total inmates, while also being able to serve a Large Enterprise System with as many as 150,000 inmates. The reason for this is that the Data Integrity is based upon the industry standard of one

unique identifier for one inmate. This one-to-one connection means that a system with 150,000 inmates will have 150,000 unique identifiers active at any given time while a system with 250 inmates will have 250 unique identifiers active at any given time. In addition to the data integrity in current time that is forced into the Data Vault by the operational aspects of the correctional industry's Inmate ID as a unique identifier the data integrity of the historical data is also protected. The correctional industry's one inmate ID to one inmate operational security in turn protects historical data because Inmate 1 can never be Inmate 6 Million. These inmates could have the same name, same tattoos, same birthdates, and have every other physiological and psychological connection, but would still be able to be uniquely identified based on the fact that Inmate 1's unique identifier and Inmate 6 Millions' can and never will be the same.

According to Dan Linstedt the "Data Vault represents a system of business intelligence" (Linstedt and Olschimke 2015, p. 11). The correctional industry feeds off of business intelligence making it a prime location for a Data Vault system. The Data Vault system allows for a correctional environment to house the business intelligence it needs and be able to quantify that information based on the industry's unique identification standards of the Inmate ID. The Data Vault system also presents another a unique feature that is in perfect correlation with a correctional environment's ecosystem and its connection to politics.

In R.A. Mcgee's *Prisons and Politics*, published by the National Criminal Justice Reference Service, the author makes a clear connection between state politics and correctional policy roles and legislation. "The State executive branch is presented as the initiator of broad correctional policy. The roles of the Governor are described as political, symbolic and ceremonial, policymaking, administrative, coordinating, and legislative. The performances of the various Governors under which the author has served as a correctional administrator are portrayed. The discussion of the legislative branch's influence on corrections gives particular attention to the situation in California (NCJRS n.d.).". Even though Mr. Mcgee's information looked at the pre-1980 politics, the information still holds true based on the organizational charts of many state governments in reference to corrections where the head of the department is directly appointed by the governor of the state. "The Michigan Department of Corrections is one of the principal state departments. Final responsibility for operation of the department rests with the Governor who appoints the Director, with the advice and consent of the state Senate. The Director serves at the pleasure of the Governor (Corrections Organizational Structure n.d.)". In order to fully understand a correctional environment, it is imperative that the changing political structure be acknowledged not as a mere side note in the ecosystem but as one of its most pivotal and structural measures.

The Need for Data Vaults in Correctional Systems

For business processes the Data Vault, as a data integration architecture, has robust standards and definitional methods which unite information in order to

make sense of it. However, one of the most robust systems ever created for the State of California did not use Data Vault Architecture, according to Bill Inmon's "The Data Vault is the Optimal Choice for Modeling the EDW in the DW 2.0 framework" (Inmon, Strauss, and Neushloss 2008). The Data Vault is a modeling technique that is able to accommodate three important aspects of Big Data in corrections: integrity of key inmate identification data; addition and tracking of new data; and the ability to separate historical data from current data. Now just because it attacks these three main points does not mean that it does not also increase other important business facts in corrections such as speed, efficiency, and the ability to make sense of data. The Data Vault approach for a correctionals environment will allow for simplification in the data entry process by allowing for segregation of data aspects. End users will be able to use one satellite to identify the data to be entered preventing it from being entered incorrectly based on pre-set conventions. Using the Data Vault system architecture a correctional system will be able to allow non-technical users to focus their efforts on efficient and agile operational problem solving using actionable business intelligence instead of waiting on a technical user to produce, decipher, and deliver the business intelligence needed. This happens by first allowing new data sources to be added as they become available without having to disrupt the original data schema. Second, the unique structure of the operational system of corrections and the Inmate ID allows for business intelligence software to be comfortably placed on top of the Data Vault with a unique identifier that allows for multiple applications. The ability to change what business intelligence the operational staff see and are able to decipher without the aid of anything else is an integral part of the equation. As noted before, criminal justice reform may change based on the political structure of the state or nation and the need for additional data points becomes imperative to tracking progress of new reforms and objectives that are put into place. Those new reforms and objectives are able to then be transferred by using Data Vault 2.0 technology to catapult the business intelligence software into the arms of the operational staff in a visually acceptable way that will allow them to be operationally significant.

Chapter 3: SCRUM and Agile Best Practices for Corrections

One of the biggest aspects of SCRUM is helping the team to eliminate anything that is slowing them down. In order to account for the waste of data in corrections, a look into the context of Big Data and the 3V structure is warranted first.

The context of Big Data in corrections is described best by looking at the 3V structure that is present in the data: volume, velocity, and variety. Volume can be explained by the sheer volume of inmates in the United States, therefore creating the sheer volume of data inside a correctional database. The variety of information needed about these individuals to be able to house them safely and securely can range from gang ties to vegan diets. Last, the velocity or speed needed to keep end users equipped with the ever-changing amount of data that exists from the various complexities that exist inside of a correctional ecosystem.

The CMMI Maturity level 5 focuses on continually improving process performance through both incremental and innovative technological improvements (Team 2006). The key words in this definition in relation to correctional environments are incremental and innovative. Correctional environments use the logic of changing a process in a slow and methodical way as not to disturb the current environment too much at one time. They also have to look at innovative ways to deal with problems that other organizations would not have to deal with in relation to technology improvements. A quick example will explain how this works: In a normal business environment a process can be implemented using negative or positive reinforcement. In a correctional environment the lack of a positive or negative repercussion means that efforts that do not depend on positive or negative repercussion must sometimes be developed. In a normal business environment an employee wants to keep his or her job, wants to avoid being incarcerated, and wants to go home after work. In a correctional environment these three outlooks are null and void. This means that the physical security of data and the non-physical security of data have to be changed to fit the environment. Data Vault 2.0 technology will be used to allow for optimization by evaluating quantitative process-improvement objectives, monitoring continually changing business objectives, and keeping the data secure.

SCRUM allows the Data Vault structure to take the volume and variety of the data down to allow for better velocity. This is done in the same way operational variety and volume in a correctional environment is separated. In a typical correctional environment there is a warden who oversees a prison. That warden is only responsible for inmates under his jurisdiction at his/her prison. The warden is not concerned with an operational sense with inmates at another prison except in rare cases where situations from the enterprise system may bleed from one prison to another. The variety of inmates and the volume of inmates have been cut down so that a warden is able to safely and effectively manage a prison at an acceptable velocity. If that is the operational standard, why would the warden need data from the other prisons to operate his prison at an optimal velocity? The Data Vault structure allows the data to be separated in a way that allows for the variety of the data to be reduced, the volume of the data to be reduced without affecting the operational aspects but by actually allowing for scalability which in turn will improve them by using SCRUM-based priority structures. Instead of taking the traditional approach to all data, it is important that the warden is able to look at what data is most important to his/her facility and focus on that information. This approach allows for the different missions of different prisons to be scaled into the actual Data Vault structure. In the current correctional climate, there exist three main prison missions in most correctional systems that effect the operational structure of the prison: re-entry into society, medium to low security, and high security. The Data Vault structure allows for these three missions to operate off of separate hubs and satellites based on the SCRUM-based main priority of the prison. A re-entry prison does not need information about death row and a high-security prison does not need information about work release assignments because those items do not exist operationally at those facilities. If the operational aspect is not

there it only makes sense that data for the operational aspect that is not needed also be absent and not part of any prioritization.

Data Entry and Staging

Data entry will occur at two levels in the Correctional Database Systems Management Solution. The first will occur upon first admittance to the prison system via iris biometric admittance verification at diagnostic intake. This biometric classification to one inmate ID insures the data integrity of his Inmate ID, and will be the first point of record. The information entered at this point will be classified into hub H_Admission noting that the inmate has been admitted and whether or not it is the inmate's first trip to prison. Then information will be entered into the hub H_Inmate_Main. The second level of data entry will occur when a facility enters information about an inmate. Information obtained at this point will be information about an inmate's whereabouts, actions that an inmate commits, programming, counseling, meals, medical, and so on. The third level of data entry will occur at the location of the main headquarters. Information entered by the headquarters' location will be specific to incarceration and will be entered into the hub H_Incarceration.

The staging area will be used to hold raw data and then move it into the data warehouse. Information that will go into the operational Data Vault includes information concerning operational aspects of the inmate while under control of a correctional system. Information will then be declassified and passed into a non-operational Data Vault. This will allow information deemed operational to be used for research purposes without interrupting the operational aspects of the correctional system's main Business Vault.

Operational Business Intelligence

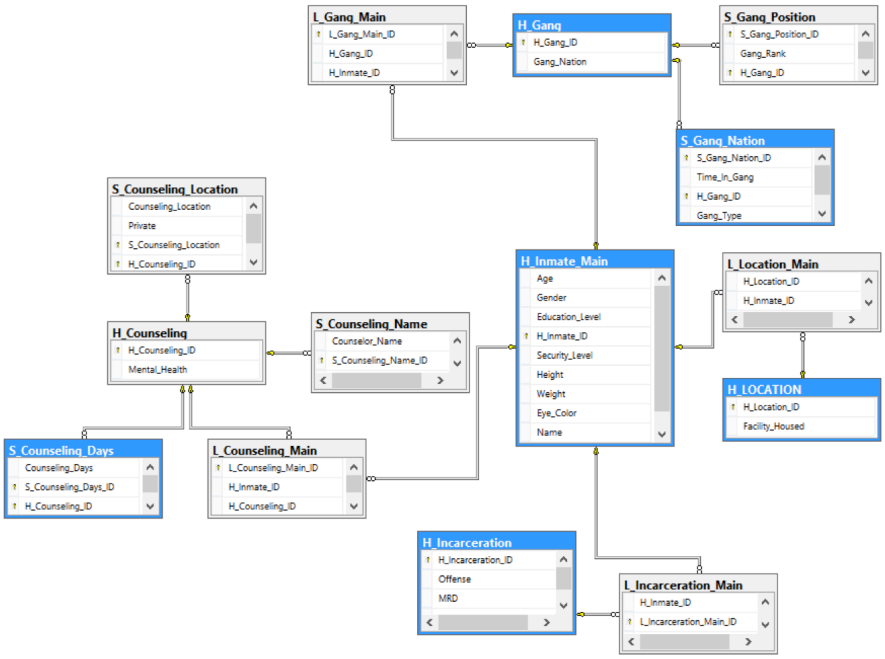
One of the secondary objectives of creating a prototype for a correctional environment that was based on a Data Vault architecture was to have operationally functional business intelligence. The reason the business intelligence needed to be visually optimized was to properly serve the operational use case. The operational use case requires a user to be able to point and click their way to the desired information. In order for this to happen the business intelligence has to look and feel a way that the user is able to easily identify with for operational use. Now, while this might not be technically important to the data integrity or other technical aspects of the Data Vault and business intelligence, it plays a significant role in the operational use case that the business intelligence will play.

The business intelligence software is positioned on top of the Data Vault structure in a way that allows the user to be able to access information in an operational environment for an operational purpose. The end user does not have to figure out which way to count the information that is provided to them. The end user does not have to figure out which sets of data are operationally important that information

is provided to them in a dashboard structure that is pre-approved during the design phase. The business intelligence instead of creating a roadblock for a non-technical end user creates a platform for actual business intelligence. The next example will explain how this process would work.

Due to the recent examinations of policy all members of street gangs who have been incarcerated for violent offenses must attend counseling at least 3 days a week.

Figure 3.1. Database diagram of Information used for business intelligence



The dashboard for the end user has already been set up to only identify the Users Facility G4, identify known gang members, identify counseling days as either TH for those needing 2 days a week, or MWF for those needing 3 days a week.

The end user is able to use the pre-designed dashboard to identify that they have four members of a street gang.

Figure 3.2.

Figure 3.3. They are then able to identify that one has a violent history.

Now from the business intelligence software the end user with only two mouse clicks the has identified that the operational decision needs to be made to have Inmate Marcos's schedule changed from TH to MWF.

Figure 3.4.

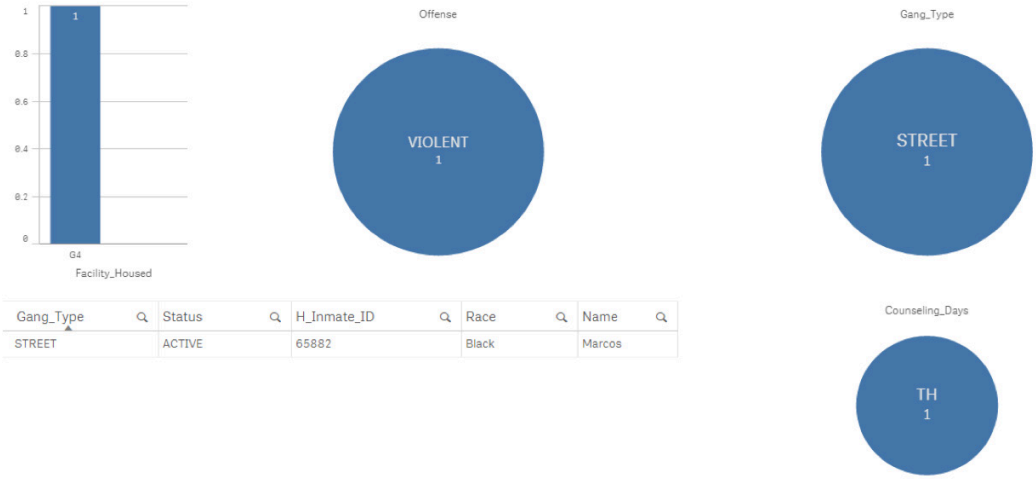


Figure 3.5. SQL of business intelligence for Figure 3.4.

```
select distinct H_Inmate_Main.H_inmate_ID, Gang_Type, Status, H_Inmate_Main.H_inmate_ID, Counseling_days, Facility_Housed, Race, Name
from H_Inmate_Main
 join L_Gang_Main on L_Gang_Main.H_Inmate_ID = H_Inmate_Main.H_Inmate_ID
 join H_Gang on L_Gang_Main.H_Gang_ID =H_Gang.H_Gang_ID
 Join S_Gang_Nation on H_Gang.H_Gang_ID=S_Gang_Nation.H_Gang_ID
 join L_Location_Main on L_Location_Main.H_Inmate_ID=H_Inmate_Main.H_Inmate_ID
 join H_Location on L_Location_Main.H_Location_ID=H_Location.H_Location_ID
 join L_Incarceration_Main on L_Incarceration_Main.H_Inmate_ID=H_Inmate_Main.H_Inmate_ID
 join H_Incarceration on L_Incarceration_Main.H_Incarceration_ID=H_Incarceration.H_Incarceration_ID
 join L_Counseling_Main on L_Counseling_Main.H_Inmate_ID = H_Inmate_Main.H_Inmate_ID
 join H_Counseling on L_Counseling_Main.H_Counseling_ID = H_Counseling.H_Counseling_ID
 Join S_Counseling_Days on H_Counseling.H_Counseling_ID=S_Counseling_Days.H_Counseling_ID
 where Facility_Housed = 'G4' and gang_type = 'street' and Offense = 'Violent'
```

Results | Messages

H_inmate_ID	Gang_Type	Status	H_inmate_ID	Counseling_days	Facility_Housed	Race	Name
65882	STREET	ACTIVE	65882	TH	G4	Black	Marcos

The previous example showed how the information could be used operationally at the local level but the same business model could be used for an enterprise-wide application. In addition, if the information is needed to be changed the same business intelligence dashboard can serve a dual purpose as long as the information resides in one of the predefined tables. Instead of looking at local violent offenders the business question has been changed to those who have committed sex-related offenses.

Due to the recent examinations of policy all members of street gangs who have been incarcerated for sex offenses must attend counseling at least 3 days a week.

Figure 3.6. Fourteen street gang members across all facilities.

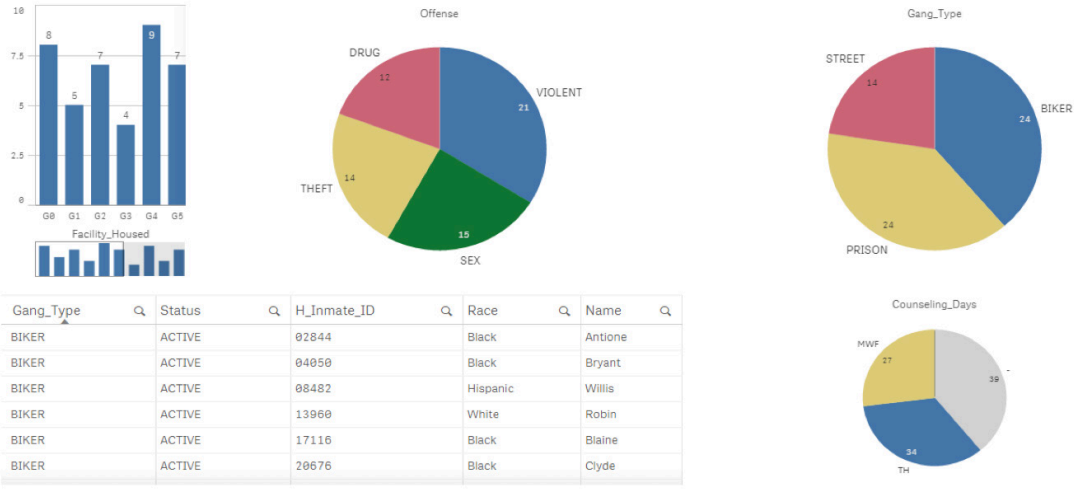
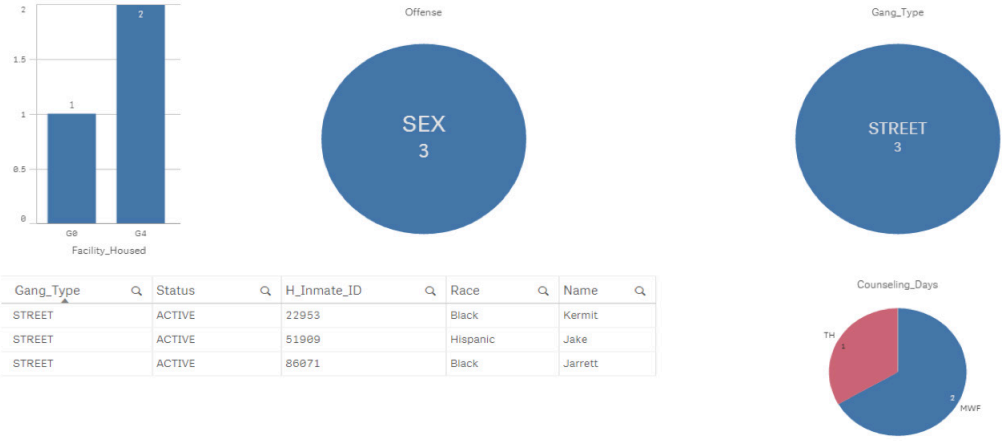


Figure 3.7. Three are incarcerated for sex offenses.

Figure 3.8. One does not have a 3-day counseling schedule.



Now from the business intelligence gained the operational decision that a call needs to be made to Facility G0 to inform them that Inmate Jake needs to have a counseling schedule changed can occur.

Figure 3.9.

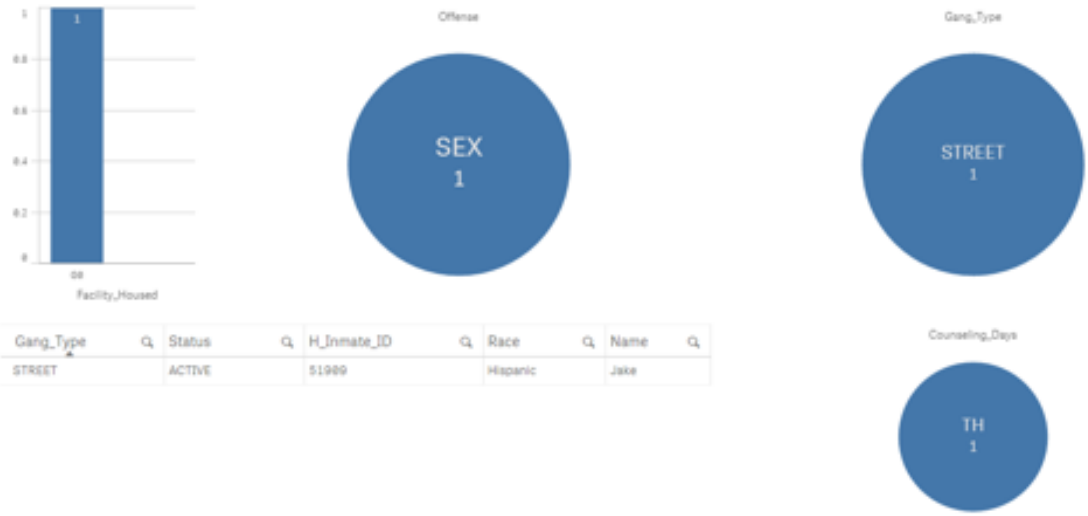


Figure 3.10. SQL of business intelligence for Figure 3.9.

```
select distinct H_Inmate_Main.H_inmate_ID, Gang_Type, Status, H_Inmate_Main.H_inmate_ID, Counseling_days, Facility_Housed, Race, Name
from H_Inmate_Main
 join L_Gang_Main on L_Gang_Main.H_Inmate_ID = H_Inmate_Main.H_Inmate_ID
 join H_Gang on L_Gang_Main.H_Gang_ID =H_Gang.H_Gang_ID
 Join S_Gang_Nation on H_Gang.H_Gang_ID=S_Gang_Nation.H_Gang_ID
 join L_Location_Main on L_Location_Main.H_Inmate_ID=H_Inmate_Main.H_Inmate_ID
 join H_Location on L_Location_Main.H_Location_ID=H_Location.H_Location_ID
 join L_Incarceration_Main on L_Incarceration_Main.H_Inmate_ID=H_Inmate_Main.H_Inmate_ID
 join H_Incarceration on L_Incarceration_Main.H_Incarceration_ID=H_Incarceration.H_Incarceration_ID
 join L_Counseling_Main on L_Counseling_Main.H_Inmate_ID = H_Inmate_Main.H_Inmate_ID
 join H_Counseling on L_Counseling_Main.H_Counseling_ID = H_Counseling.H_Counseling_ID
 Join S_Counseling_Days on H_Counseling.H_Counseling_ID=S_Counseling_Days.H_Counseling_ID
 where gang_type = 'STREET' and Offense = 'Sex' and Counseling_days =' TH'
```

Results | Messages

H_inmate_ID	Gang_Type	Status	H_inmate_ID	Counseling_days	Facility_Housed	Race	Name
51909	STREET	ACTIVE	51909	TH	G0	Hispanic	Jake

SCRUM + Data Vault for Efficiency

Using the Data Vault strategy for Data Warehouse Design allows the development teams to work more efficiently and effectively. Using a SCRUM strategy context switching is held to minimum to reduce the amount of wasted time in development. Instead of pulling teams off of current projects to aid with the creation of new modules it allows database teams to work independently. Let us take a look at this scenario: The introduction of a new referendum on criminal justice reform has required that all inmates with 2 years or more of military service who served in

Vietnam be allowed special considerations, at the current state military service is not tracked at all. In a traditional database the Data Base Administrators would have to be pulled off of their current task to aid in the creation of the new tables and how those tables fit correctly into the current structure of the database. However, using the Data Vault system one is able to evoke total quality management within the organization to prepare the table needed outside of the database structure instead of within it. This allows the DBA to stay on current task while a development team designs the new table. Since the Data Vault system is built for scalability the new table is able to be completed before removing the DBA from their current project if the need arises at all. The new table will be independent of other moving parts in the database so it can be added and will not affect the quality of data that is already in the database. Therefore, a DBA is only needed to overview and assess the situation after the the new table is linked. The DBA is not required during the planning, building, implementation, or preparation phases. This will create a more efficient environment by removing how much context switching occurs from higher level technical staff. Instead of calling off current projects, the staff is able to finish current projects and then assess new projects after completion of current projects thereby reducing wasted time.

Now using a Data Vault system that already had some of the current information available makes the process even more efficient. The data that already exist in the Data Vault can be moved to a View that allows it to be searched from the business intelligence software or through SQL Query without having to create another table. In this example, the information from the Satellite for Army was joined with information from the military hub. This will allow the end user to search only the information needed or the information can be exported to a research team.

```
Create View Army_More_Than_2 as
SELECT S_Mil_Army.Army_Location, H_Military.H_Military_ID, H_Military.Length
FROM S_MIL_Army
Inner JOIN H_Military
ON S_Mil_Army.H_Military_ID=H_Military.H_Military_ID
where length >=2
```

Figure 3.11.

Now the view can be searched for the specifications needed and saved for another time when other information from other locations might be needed.

```
Select * from Army_More_than_2
```

100 %

Results | Messages

	Army_Location	H_Military_ID	Length
1	Mayotte	52026	2
2	Vanuatu	80571	6
3	Japan	18666	7
4	El Salvador	21750	3
5	Tuvalu	95787	9
6	Thailand	59173	4
7	Sierra Leone	45382	3
8	Malaysia	01788	4
9	Azerbaijan	87442	2
10	Azerbaijan	54912	6
11	Zimbabwe	43810	6
12	Sri Lanka	35872	2
13	Peru	33386	3
14	Yemen	44535	6
15	Fiji	94534	4
16	Belgium	35564	9
17	New Zealand	07301	2
18	Nigeria	90074	4
19	Ghana	64794	6
20	Samoa	34490	2
21	Barbados	41110	9
22	Micronesia	90630	5
23	Saint Helena	32721	2
24	Uruguay	37430	3

Figure 3.12.

```
Select * from Army_More_than_2
where Army_Location = 'Vietnam'
```

100 %

Results | Messages

	Army_Location	H_Military_ID	Length
1	Vietnam	00239	4

Chapter 4: Conclusion

The primary reason for using a Data Vault structure in a correctional ecosystem is to allow for agile integration of new data points based on the ever-changing missions of corrections and their inherent connection with the political structure and societal outlooks at any given point in time. The secondary reason for using a Data Vault structure in a correctional ecosystem is to increase the business intelligence to the end user. In an industry where data integrity is of utmost importance, using a Data Vault system makes perfect sense. The Data Vault structure allows for the addition of

new data points without corruption or manipulation of the already established data structures inside the data warehouse. Therefore, as goals change and the need for new data is needed, data integrity of primary data is left untouched. The Data Vault structure would allow for addition of this data without corruption or manipulation of already established data structures inside the data warehouse.

The goal of this paper was to show that the Data Vault system was not only a valid method of Data Base design for the correctional industry, but a best case scenario because of the unique business correlations to business intelligence based on the one-to-one unique identification that is used in the correctional industry. The prototype that was created ran information for 100 inmates and can easily be adopted for many more. The information of only 100 inmates connected to 16 different hubs and 39 different satellites was shown to be able to create a vast collection of Big Data. The prototype created for this paper was able to be run on MS SQL Server 2012, AMD A6 Processor, and 250 GB SSD Samsung Hard Drive running 8 GB of standard RAM in a Windows 8 Environment. The business intelligence suite that produced the information from the Data Vault was the free desktop business intelligence software. Thus, the total cost of Data Vault creation and business intelligence for this prototype was no more than the cost of MS SQL Server 2012, a standard to premium laptop, and labor. This is important because in a correctional environment, like many other environments, being able to maintain efficiency cuts down on cost. As of December 2014 the cost of one bed in a correctional industry in Florida per day is $49.49 a day. This does not include special use beds which can nearly double in cost per day. The Florida DOC houses 100,873 inmates. If the cost of each bed is $49.49, it amounts to a cost of $4,992,204.77 per day (Florida Department of Corrections n.d.). What this means for a Data Vault system is allocation of bed space and tracking of the said bed space is a very important topic that should be able to be looked at by businesseintelligence models to allow for operational intelligence to successfully satisfy the operational need of a nearly 5 million dollar a day industry.

The Data Vault system allows for better business intelligence about bed management, contraband introduction, programming, and a myriad of other scalable ideas and ideologies based on the business structure and political climates that mandates organizational priority. It creates an agile system that allows for the introduction of new data without disturbing the needed data integrity of a correctional system.

In order to show this agility in an architectural sense, an SQL Query file for the entire Data Vault was created and then loaded as another separate Data Vault into MS SQL Server. This process took <2 min from start to finish. This means that structurally the Data Vault is small enough to be able to be transferred from agency to agency. This increases the efficiency as new hubs and satellites can be created specific to a certain agency and then scaled to fit other agencies that may have the same or a similar mission. This thought process would hold true in relation to small jails with only a few hundred inmates who may fall indirectly under the organizational structure of a larger state correctional agency with 20,000+ inmates or more.

In reference to the business intelligence software, depending on the amount

of data requested from the MS SQL SERVER requested took between 15 min and 1 h. However, once the data was loaded into the business intelligence software the information rate had no lag whatsoever, but the initial load was fairly time-consuming. This time would however still be sufficient for small to medium operations that may not have the budgets to acquire the necessary hardware to allow the business intelligence to run in a real-time environment.

The Data Vault allows for optimal business intelligence structures based on the one- to- one identification method of the correctional industry; there is one Inmate_ID for one inmate. This creates a barrier of protection and data integrity when adding historical data or new data.

In order to protect the data integrity of the Data Vault when entering new or historical data to the Data Vault the hub The H_Inmate_Main Column H_Inmate ID has a constraint put on it that only allows unique values. This allows historical data to be entered into the Data Vault and the addition of new data into the Vault without any overlap in the unique identifier that has been set as the main business key as H_Inmate_ID.

```
CREATE UNIQUE NONCLUSTERED INDEX [UniqueID]
ON [H_Inmate_Main] (H_Inmate_ID)
```

To illustrate this process, the data was generated in the replicated Data Vault to create a mock set of historical data as inmates 00-–99 as shown below in Figure 4.1.

Figure 4.1.

```
select * from H_Inmate_Main;
select Count(H_inmate_ID) as #_INmate_ID from H_Inmate_MAin
```

100 %

Results | Messages

	Age	Gender	Education_Level	H_Inmate_ID	Security_Level	Height	Weight	Eye_Color	Name	Race
1	42	F	15	00	Med	77	164	Aquamarine	Hendrix	Native-Am
2	39	M	4	01	Med	53	321	LightSlateGray	Stanley	Black
3	42	M	9	02	Max	42	260	Aquamarine	Page	White
4	57	F	11	03	Med	49	328	DarkViolet	Tyler	Hispanic
5	65	M	14	04	Med	45	258	Thistle	Moore	Black
6	65	M	11	05	Min	52	383	PeachPuff	Reed	Asian
7	48	M	9	06	Med	51	322	LightSkyBlue	Hogan	Asian
8	64	M	14	07	Min	58	214	LightGray	Andrade	White
9	49	M	7	08	Max	43	371	Peru	Scott	Black
10	25	M	7	09	Med	51	362	LightGray	French	Native-Am
11	53	M	8	10	Med	43	318	GoldenRod	Goodman	White
12	64	F	8	11	Max	77	143	DimGrey	Levine	Asian
13	44	M	2	12	Med	75	302	Beige	Mc Cormick	Black
14	19	F	9	13	Min	49	338	Chocolate	Dominguez	Hispanic
15	47	M	11	14	Med	54	117	SlateBlue	Hobbs	Native-Am
16	38	M	7	15	Min	46	312	DarkViolet	Kemp	Hispanic
17	23	F	2	16	Max	63	356	LightSlateGray	Bell	Native-Am
18	61	F	10	17	Min	43	242	Teal	Reeves	Native-Am
19	64	M	7	18	Max	44	293	Thistle	Mckenzie	Asian
20	43	M	16	19	Med	48	289	LightSlateGrey	Costa	Native-Am
21	60	M	9	20	Med	68	231	MidnightBlue	Duran	Asian

	#_INmate_ID
1	100

When trying to alter an Inmate ID, the following exception is thrown and the statement is rejected as shown in Figure 4.2.

Figure 4.2.

```
UPDATE H_Inmate_Main
SET H_Inmate_ID='00'
WHERE H_Inmate_ID='55';
```

100 %

Messages

```
Msg 2627, Level 14, State 1, Line 1
Violation of PRIMARY KEY constraint 'XPKH_Inmate_Main'. Cannot insert duplicate key in object 'dbo.H_Inmate_Main'. The duplicate key value is (00
The statement has been terminated.
```

However, when adding new unique Inmate ID's, the system accepts the new data as shown in Figure 4.3.

Figure 4.3.

```
select * from H_Inmate_Main;
select Count(H_inmate_ID) as #_INmate_ID from H_Inmate_MAin;
```

100 %

Results | Messages

	Age	Gender	Education_Level	H_Inmate_ID	Security_Level	Height	Weight	Eye_Color	Name	Race
36	60	M	12	122	Max	NULL	239	Crimson	Eugene	Native-Am
37	42	M	16	123	Med	64	341	AntiqueWhite	Jarrett	Hispanic
38	25	M	13	124	Med	63	165	BlanchedAl...	Alejandro	Native-Am
39	54	M	12	125	Med	71	130	Ivory	Doug	White
40	31	M	11	126	Med	70	225	CadetBlue	Jered	Hispanic
41	45	M	8	127	Med	73	183	Orange	Jonah	Hispanic
42	52	F	2	128	Max	51	310	Salmon	Jess	Asian
43	64	M	4	129	Max	70	372	BlueViolet	Jody	Asian
44	19	F	9	13	Min	49	338	Chocolate	Doming...	Hispanic
45	26	M	15	130	Med	52	224	LightSlateGray	Cameron	Black
46	46	F	9	131	Med	70	106	ComflowerBl...	Armand	Asian
47	56	M	10	132	Max	70	141	LightPink	Sheldon	White
48	46	M	14	133	Med	70	206	Wheat	Jesse	Asian
49	55	M	11	134	Med	56	382	AntiqueWhite	Carlo	Hispanic
50	21	M	4	135	Med	80	255	Fuchsia	Ray	White
51	62	F	10	136	Min	69	328	DimGray	Alvin	Native-Am
52	30	M	12	137	Max	63	348	Chartreuse	Raul	Black
53	62	F	4	138	Med	76	236	DarkSlateGray	Tad	Asian
54	32	M	11	139	Med	51	125	Indigo	Taylor	White
55	47	M	11	14	Med	54	117	SlateBlue	Hobbs	Native-Am
56	49	M	12	140	Min	52	336	DimGrey	Dave	Asian

	#_INmate_ID
1	200

The simple example speaks of the ability of the Data Vault as an agile, fast, efficient, and accurate means for a correctional environment to be able to enter information with the changing socio-political climate that resides in such an industry, apart from the already proven business intelligence the Data Vault provides for a correctional system. In closing, the purpose of this paper has been appropriately met and exceeded by creating a developmental model suited for a correctional environment that is efficient and effective based on SCRUM and AGILE standards and provides efficient operational business intelligence to the end user.

References

1. Agan, A. (2011). "Sex Offender Registries: Fear Without Function?" The Journal of Law & Economics 54 (1): 207-239.
2. Casters, M., R. Bouman, and J.V. Dongen. (2010). *Pentaho Kettle Solutions Building Open Source ETL Solutions with Pentaho Data Integration*. Indianapolis, IN: Wiley.
3. Carson, A. (September 17, 2015). Prisoners in 2014. US Department of Justice, Office of Justice Programs, Bureau of Justice Statistics. http://www.bjs.gov/index.cfm?ty=pbdetail&iid=5387
4. CDCR's Strategic Offender Management System project earns national award. (n.d.) http://www.insidercdcr.ca.gov/2014/08/cdcrs-strategic-offender-management-system-project-earns-national-award/ (accessed October 15, 2016).
5. Corrections Organizational Structure. (n.d.). Michigan DOC. http://www.michigan.gov/corrections/0,4551,7-119-62761_24147--,00.html
6. Corrections Technology Association. (n.d.) Big Data and Corrections: What's the Big Issue? http://www.correctionstech.org/meeting/2013/Presentations/RR7.pdf
7. Cunningham, M.D., J.R. Sorenson, and T.J. Reidy. (2005). "An Actuarial Model for Assessment of Prison Violence Risk Among Maximum Security Inmates." Assessment, 12 (1): 40-49.
8. Database Data Warehousing Guide. (November 4, 2005). https://docs.oracle.com/cd/B19306_01/server.102/b14223/ettover.htm (accessed January 14, 2016).
9. Florida Department of Corrections. (n.d.) http://www.dc.state.fl.us/oth/Quickfacts.html (accessed February 21, 2017).
10. Foss, E. (1914). "The Ideal Prison System." *Journal of the American Institute of Criminal Law and Criminology*, 4 (5): 674-686. http://www.jstor.org/stable/1132646
11. Gnatovich, R. 2007. "Making a Case for Business Analytics." *Strategic Finance* 88. (8):46.
12. Inmon, W.H. and D. Linstedt. (2014). *Data Architecture: A Primer for the Data Scientist: Big Data, Data Warehouse and Data Vault*. San Francisco, CA: Morgan Kaufmann.
13. Inmon, W.H., D. Strauss, and G. Neushloss. (2008). *DW 2.0: The Architecture for the Next Generation of Data Warehousing*. San Francisco, CA: Morgan Kaufmann.
14. Ivan, M.L. 2014. "Characteristics of In-Memory Business Intelligence." *Informatica Economica* 18 (3): 17-25.
15. Jovanovic, V., I. Bojicic, C. Knowles, and M. Pavlic. 2012. "Persistent Staging Area Models for Data Warehouses." *Issues in Information Systems* 13 (1): 121-132.
16. Jovanovic, V., and I. Bojicic. 2012. "Conceptual Data Vault Model." http://works.bepress.com/vladan-jovanovic/9/
17. King, R.S., M. Mauer, and T. Huling. (2003). *Big Prisons, Small Towns: Prison Economics in Rural America*. Sentencing Project.

18. Linstedt, D. (2001). U.S. Patent Application No. 09/965,343.
19. Linstedt, D. (February 4, 2010). About the Data Vault. http://danlinstedt.com/allposts/datavaultcat/about-the-data-vault/
20. Linstedt, D., and K. Graziano. (2011). *Super Charge Your Data Warehouse: Invaluable Data Modeling Rules to Implement Your Data Vault*. CreateSpace.
21. Linstedt, D. and M. Olschimke. (2015). *Building a Scalable Data Warehouse with Data Vault 2.0*. San Francisco, CA: Morgan Kaufmann.
22. Iliev, A., and S. Smith. (April 2002). "Prototyping an Armored Data Vault." In International Workshop on Privacy Enhancing Technologies. Berlin Heidelberg: Springer, 144-159.
23. Marquis Electronic Offender Management Information System. (2017). http://www.marquisware.com/
24. Nichols, R. (December 2014). 2015 NASCIO Recognition Award Submission Nomination Category: Improving State Operations.
25. NIC State Statistics - California. (n.d.). http://nicic.gov/statestats/default.aspx?st=CA (accessed February 29, 2016).
26. Prisons and Politics Abstract R.A. McGee 1981.
27. Strategic Offender Management System (SOMS) CDCR. (n.d.) http://cdcr.ca.gov/SOMS/index.html (accessed February 8, 2016).
28. Sutherland, J. and J. Sutherland. (2016). *Scrum: The Art of Doing Twice the Work in Half the Time*. New York, NY: Crown Publishing.
29. Team, C.P. (2006). CMMI for Development, version 1.2.
30. What is business intelligence (BI)? (n.d.) http://searchdatamanagement.techtarget.com/definition/business-intelligence (accessed March 4, 2016).