CECIDS

Chicago Early Childhood Integrated Data System

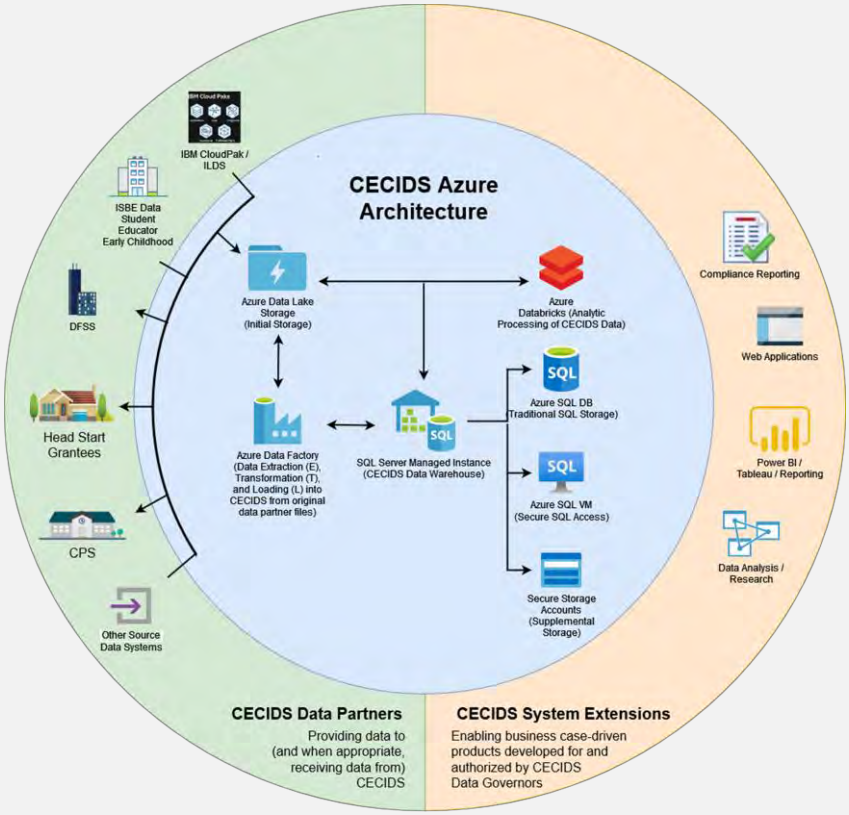
Technical Systems and Security Program Overview

May 2022

1. Architecture

The technical infrastructure of the Chicago Early Childhood Integrated Data System (CECIDS) will integrate disparate data sets to create better, more actionable data for stakeholders. Its system architecture will be built and grown on Microsoft Azure Government, which encompasses a robust set of components that offers performance, flexibility, and scalability. Figure 1 depicts these components—common to data lake solutions—and their relations to CECIDS data sources and the general categories of user products to be populated by the system's data. Component-specific descriptions are found below.

FIGURE 1: CECIDS AZURE ARCHITECTURE DIAGRAM



Azure Data Lake Storage Gen2 provides cost-effective and flexible storage of traditional text and numerical database content as well as images and other semi-structured content. This component is capable of accommodating volumes of data exceeding CECIDS' anticipated needs, making it scalable for any future system

expansion. CECIDS will use it as the main repository for ingested data and as the repository for any post-processed data that needs to be recalled long term for use in reports, dashboards, etc.


Azure Data Factory (ETL)

Azure Data Factory supports ETL or ELT processing of raw data like those to be ingested by CECIDS. This component will serve as the data traffic controller of CECIDS, implementing ETL to enable the movement, transformation, and placement of data across the system. It can be configured to perform automatic data transformations based on situational trigger events, e.g., the end of a day or week or the introduction of a certain data point, and initiate data updates, special reports, or other events.

SQL Server Managed Instance (CECIDS Data Warehouse)

Azure SQL Server Managed Instances are storage systems based on Microsoft SQL Server whose maintenance and security updates are provided for as a part of the CECIDS regular Azure operating costs. While slightly more expensive in daily operating costs, the total cost of ownership is lower than self-managed SQL installations with a consistently higher degree of security for CECIDS' data. SQL Managed Instances will be used in situations where application software being used within the system calls for SQL structures, or in other situations where the incorporation of additional data sets is best facilitated by a more traditional SQL data structure.


Azure Databricks

Azure Databricks is a data analytics system native to the Microsoft Azure platform. It will allow data to be pulled strategically from the CECIDS storage systems (Azure Data Lake Storage or Azure SQL Server Managed Instances) and analyzed, with results of the analyses available in the form of reports, dashboards, or other reporting outputs. The most visible use of Databricks in CECIDS will be the analysis of data across all sources for producing provider, neighborhood, or citywide dashboards.

Azure SQL DB

Azure SQL Databases are Microsoft SQL instances optimized for operation in the cloud but with the scalability and flexibility of traditional server-farm SQL server deployments. CECIDS will use Azure SQL Databases primarily for interactions with external systems preferring or requiring connection to traditional SQL database structures. IBM Cloud Pak for Data, the platform used by the Illinois Longitudinal Data System, is projected as one such use case.

Azure SQL VM

Azure SQL Virtual Machines (VMs) extend the flexibility provided by Azure SQL Databases. They will allow CECIDS to store data in SQL instances that have been specifically tailored—in terms of SQL Server version, edition, and capacity—to the needs of software systems connecting to its data. The automated management and upgrade of cloud-based systems can outpace that of connecting systems. Azure VMs maintain connections with applications whose code continues to look for traditional SQL database structures while preserving the highest level of current security standards across those connections.


Secure Storage Accounts

Azure Secure Storage Accounts are a generic and secure form of storage within which CECIDS data can reside and be accessed as needed. These accounts vary in their structures and performance levels and can provide cost-effective options for storing data rarely accessed by the system. While not projected for immediate use

within CECIDS, they may be deployed as the system's data set grows in breadth and longitudinal depth, and use patterns within the data are identified.

These Azure components are subject to change in branding or function per Microsoft. As necessary and where appropriate, CECIDS will incorporate new or rebranded components that perform similar functions. The CECIDS technical team will communicate to the CECIDS Data Governors any changes to the set of Azure components.

2. Data Flows

Data in CECIDS will generally flow to and through four general layers: ingestion; extraction, transformation, and loading; processing, and presentation. Each layer is described below at a high level.

Ingestion

CECIDS will ingest unprocessed data from multiple sources using secure methodologies—including Application Programming Interface (API), Secure File Transfer Protocol (SFTP), or other methods—matched to the capacities of the specific data provider. These data may vary in file formatting, have different file structures based on their organization of origin, and have varying degrees of error included within them.

Extraction, Transformation, Loading (ETL) / Extraction, Loading, Transformation (ELT)

Following ingestion, unprocessed data will enter a primarily automated ETL or ELT process. The slated environment ETL process will extract the unprocessed data from their original files, transform the extracted data to match the structures used for storage within CECIDS, then load the transformed data into the system's data platform for further processing. An ELT process—which instead extracts, loads, then transforms the data—is being considered as an alternative, with an ultimate determination to be made by the CECIDS technical team pending capacity and need. Regardless of process, Azure's firewall protects data in this layer against public exposure, and automation minimizes exposure of any sensitive Personally Identifiable Information (PII) to the technical team.

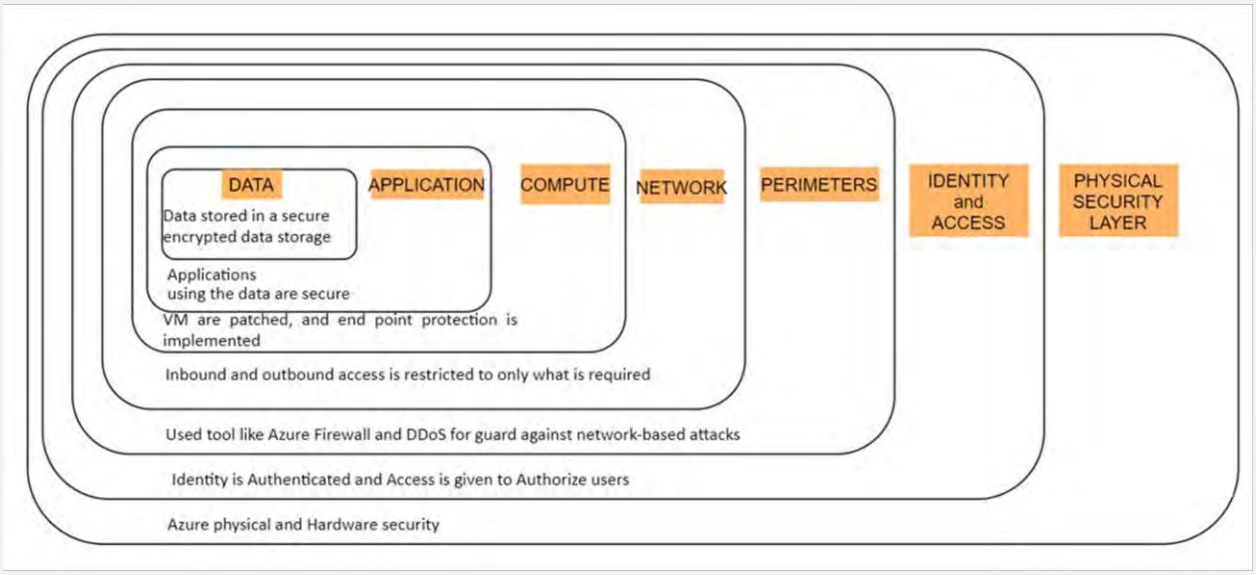
Processing

Processing will make transformed and loaded data available for the advanced manipulation, aggregation, and refinement necessary to populate the CECIDS user products. Only the smallest necessary portion of the CECIDS technical team will have access to PII within this layer, which will host much of the system's development work. The technical team will use primarily pseudonymized data—with PII, e.g., name or SSN, removed or otherwise substituted with pseudonyms—for system construction and refinement. Access restrictions will extend to sensitive aggregate data.

Presentation

The presentation layer will direct processed data to applications supporting CECIDS use cases, including analytics, operations support, compliance reporting, research purposes, or others identified by system stakeholders. Planned applications include interactive dashboards and visualizations, custom web applications and reports, and statistical software to perform analytics and research. Applications will be access-controlled depending upon the sensitivity of their data contents, and all users of access-controlled applications will be authenticated to ensure data security.

FIGURE 2: LAYERS OF SECURITY IN THE AZURE ARCHITECTURE



3. Security Program

Microsoft Azure Government

CECIDS is built on Microsoft Azure Government, which is geared towards public agencies and meets security and compliances standards exceeding those of the commercial Azure product. Azure Government locates all its data centers and networks in the U.S., and it meets various, additional U.S. government requirements around storing and access data. Please visit the Microsoft website for more information describing Azure Government.

Security Layers

CECIDS prioritizes data security in both the design of its technical architecture and operational protocols and the provision of access rights granted to the data professionals responsible for system operations and maintenance. Its Azure architecture provides structural layers of protection for CECIDS and other resident systems.

Physical Security Layer

CECIDS data security starts with the security of Microsoft's physical hardware. The CECIDS Azure spaces will be configured within Microsoft's Platform-As-A-Service (PaaS) model, within which Microsoft manages the physical layer, ensuring operational stability and the most up-to-date maintenance of security patches across the hardware layer.

Identity and Access

Access to data within CECIDS will be restricted to authorized users, who will only be allowed access to portions of the system relevant to their position and purpose.

Perimeter Defense

Azure Firewall and Distributed Denial of Service (DDoS) services will provide vital layers of protection against external attacks on CECIDS and the data therein.

Network

Active management of inbound and outbound channels of access to the CECIDS data repository will further reduce risk by limiting how data can enter and exit the system.

Compute

Azure Virtual Machines (VMs) will provide secure access to the computing resources necessary for CECIDS. Maintenance and update of those VMs is a key feature of the PaaS model.

Application

CECIDS technical staff have years of experience designing and implementing applications that facilitate the secure access to and management and reporting of sensitive data.

Data

Data will be encrypted within the CECIDS data repository.

Security Protocols

FIGURE 3: CECIDS SECURITY PROTOCOLS DIAGRAM

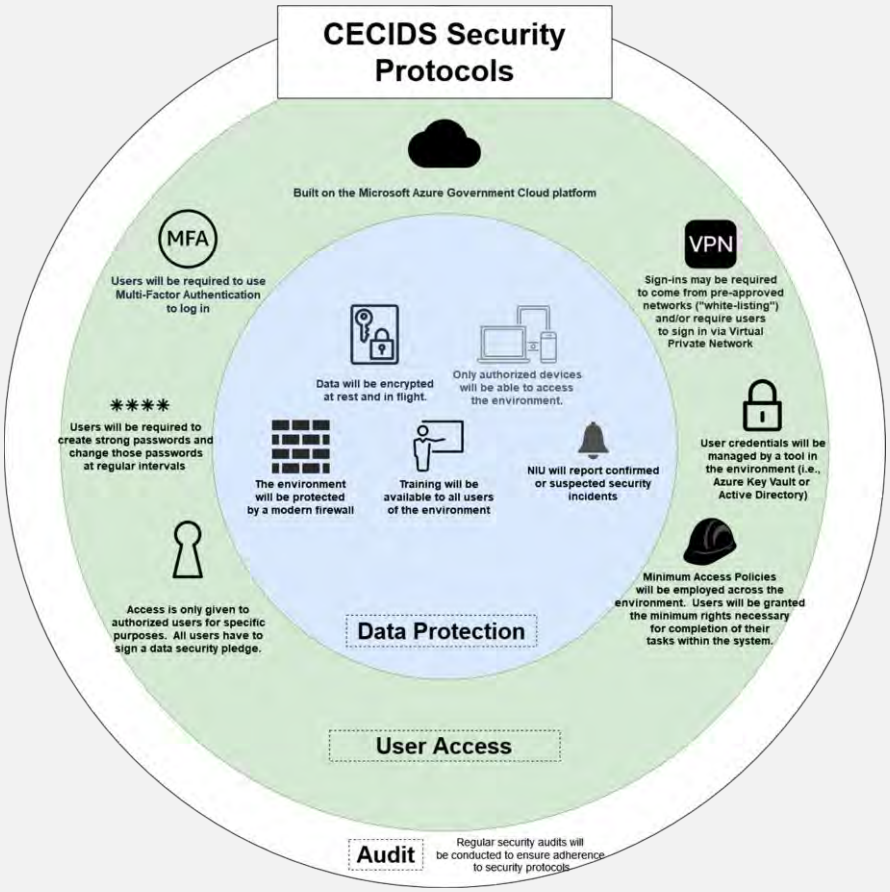


Figure 3 and Table 1 describe the security protocols built into CECIDS and used and/or deployed by the CECIDS technical team. The team and its Technical Affiliates will, at minimum, comply in its treatment of CECIDS Data Governor data with the United States National Institute of Standards and Technology (NIST) Special Publication 800-53 Revision 4 Moderate Level Control.

TABLE 1: CECIDS SECURITY PROTOCOLS

Security Protocol	Notes
Multi-Factor Authentication	Azure Active Directory Multi-Factor Authentication requires multi-factor authentication for every user sign-in.
IP Whitelisting and/or VPN	CECIDS user sign-ins may be subject to authentication via whitelisting of certain IP network(s). CECIDS access may thus require users to sign into Azure Virtual Private Network.
Password Creation and Rotation	Azure Key Vault will require all users to create strong passwords of specific length and complexity per vendor policy. All users will be required to change passwords at regular intervals.
Minimal Access Policy	Access to CECIDS will be granted solely to authorized users for specific data and services. All users must sign a security pledge acknowledging their agreement to abide by the terms of the CECIDS Data Contributor and Participation Agreement
Data Storage and Role-Based Access Control	Sensitive Data will be stored within Azure Data Lake Storage. Access to the data lake will be controlled by a cataloguing and access control system that defines read/write permissions for specific CECIDS user roles. Azure Blob Storage uses a two-key access authentication structure.

Data Encryption	Data is always encrypted at rest within Azure Storage Accounts and Azure Databases.
Firewall	Azure Firewall provides state of the art protection against unwanted traffic within the system.
Security Training	Security training and security support materials will be made available to all CECIDS users.
Incident Response	The CECIDS technical team will immediately report to the relevant Party(s) any confirmed or suspected incidents involving the security of Sensitive Data within CECIDS and fully cooperate with the Party(s) to investigate and resolve the incident. Azure Security Center provides 24/7 monitoring of security conditions across CECIDS.
Security Audit	The CECIDS technical team will establish a process for auditing the CECIDS and its data security. This process, along with the auditing entity, will be determined by the CECIDS technical team with the approval of the Parties.

CECIDS will store and process Sensitive Data, including personally identifiable information, in accordance with industry best practices compliant with the standards set forth in the CECIDS Data Contributor and Participation Agreement executed between Northern Illinois University (NIU)—technical administrator of CECIDS—and each respective CECIDS Data Governor. These standards include appropriate administrative, physical, and technical safeguards to secure Sensitive Information and/or Student Data from unauthorized access, disclosure, and use. CECIDS ("The System") shall comply with the following requirements:

1. All data must be secured in transit using secure FTP services or https/TLS 1.0+. Industry certifications, such as International Organization for Standardization (ISO), SysTrust, Cloud Security Alliance (CSA) STAR Certification, or WebTrust security for SaaS environments are recommended.
2. Such safeguards shall be no less rigorous than accepted industry practices, including specifically the NIST 800-53r4 moderate level, International Organization for Standardization's standards ISO/IEC 27001:2005 (Information Security Management Systems – Requirements), and ISO-IEC 27002:2005 (Code of Practice for International Security Management).

3. Technical Affiliates will conduct periodic risk assessments and remediate any identified security vulnerabilities in a timely manner. Technical Affiliates will also have a written incident response plan, to include prompt notification of the Data Governor in the event of a security or privacy incident, as well as best practices for responding to a breach of Sensitive Information and/or Student Data security practices.
4. All transmission of Data shall be accomplished through encryption of no less rigor than NIST-validated DES standards.
5. The System shall include component and system level fault tolerance and redundancy in system design.
6. The System shall encrypt user passwords in any data storage location and obfuscate password entry fields in any entry interface controlled by the discloser.
7. The System shall encrypt Student Data and Sensitive Information at-rest and in-transit.
8. The System shall support authentication of users at login with a 128-bit or higher encryption algorithm.
9. The System shall secure transmission of login credentials.
10. The System shall provide automatic password change routine.
11. The System shall trace user system access via a combination of system logs and Google Analytics.
12. The System shall secure (encrypt) the audit trails and system generated logs and ensure that they are stored in locations that are inaccessible to automated content discovery software.
13. The System shall conduct or undergo system level testing whenever new functionalities are added to the Solution to reconfirm system security measures are retained and functional, and that interaction with the Board systems is not degraded or compromised.
14. The System shall employ an in-line Intrusion Protection System that inspects incoming data transmissions.
15. The System shall ensure that Student Data is stored in privately addressed network devices that have no direct interaction with public networks.
16. The System shall ensure prevention of hostile or unauthorized intrusion.
17. NIU and Technical Affiliates shall ensure screening of employees and agents with access to Student Data to assure that any agent with access to the Student Data has passed an industry-standard criminal background check. The Parties shall identify the security measures taken to ensure that said employees and agents do not have access to Student Data.
18. The System shall backup of all Data at least once every twenty-four (24) hours.
19. The System shall perform content snapshots at least daily and retain for at least ninety (90) days.
20. The System shall provide a documented disaster recovery plan that includes the following elements:
21. The System shall identify available recovery times.
22. NIU and Technical Affiliates shall conduct 24x7 system monitoring that is capable of detecting Potential outages.
23. NIU and Technical Affiliates shall identify plans for File-level, Database and server recovery after a component/system failure, damage or compromise.
24. NIU and Technical Affiliates shall ensure substantial geographical separation between data centers hosting production, backup and redundant system elements.
25. NIU and Technical Affiliates shall identify recovery/mitigation procedures for all managed sites, including subcontractors, agents, and other recipients.

26. NIU and Technical Affiliates shall ensure no less than annual testing of the disaster recovery plan (at least parts that affect Student Data) with results of the test made available to the Board, as well as information about, and schedule for, the correction of deficiencies identified in the test.
27. The System shall include provisions for at least the following events:
    a. Fire
    b. Natural disaster
    c. Sabotage
    d. Accidental human error
    e. Flooding
    f. Equipment failure
    g. Application/database failure
    h. Other unlikely and/or disastrous events.