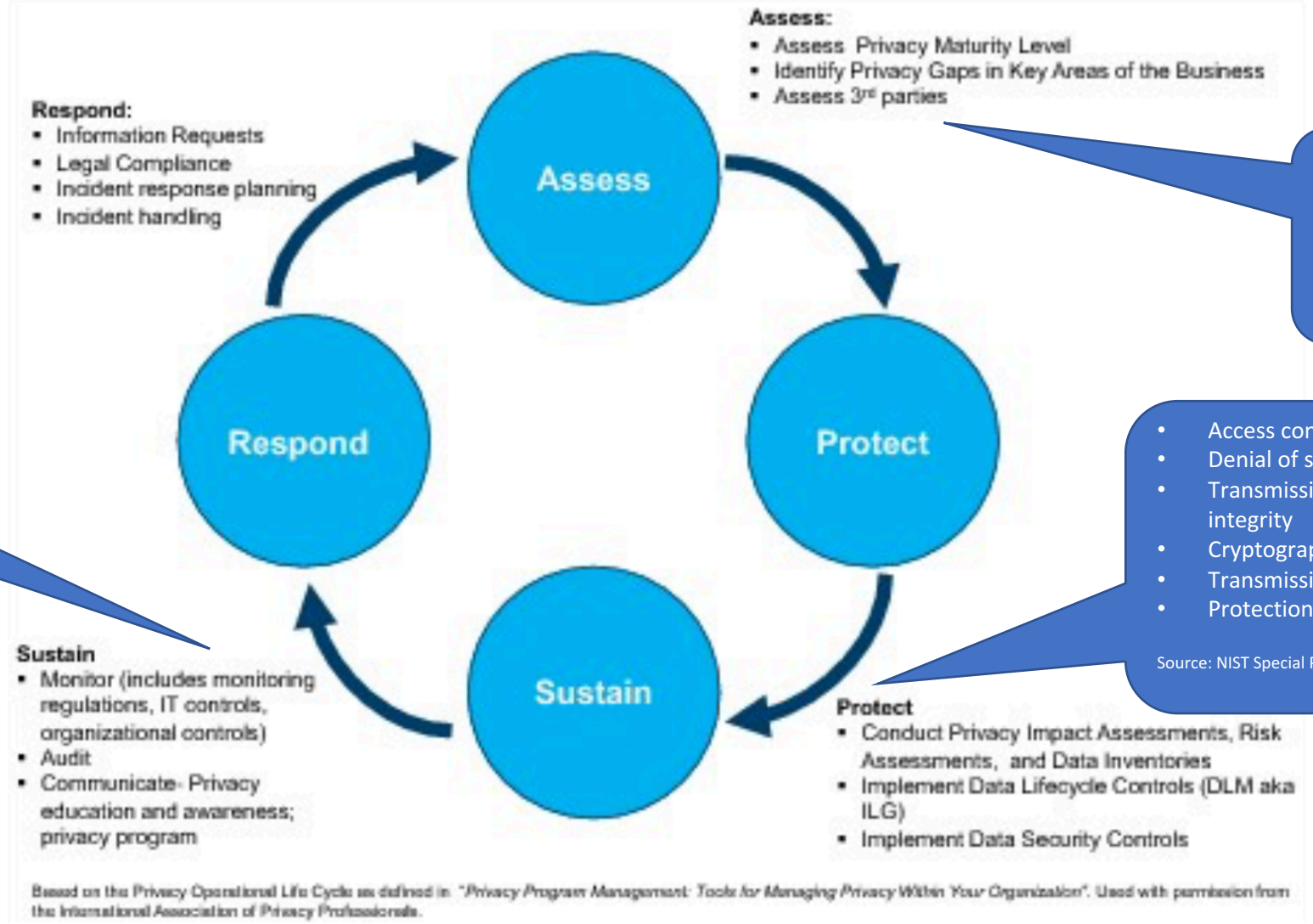
Data Access Governance

Cognitive Enterprise Data Lake

Data Security Operational Lifecycle



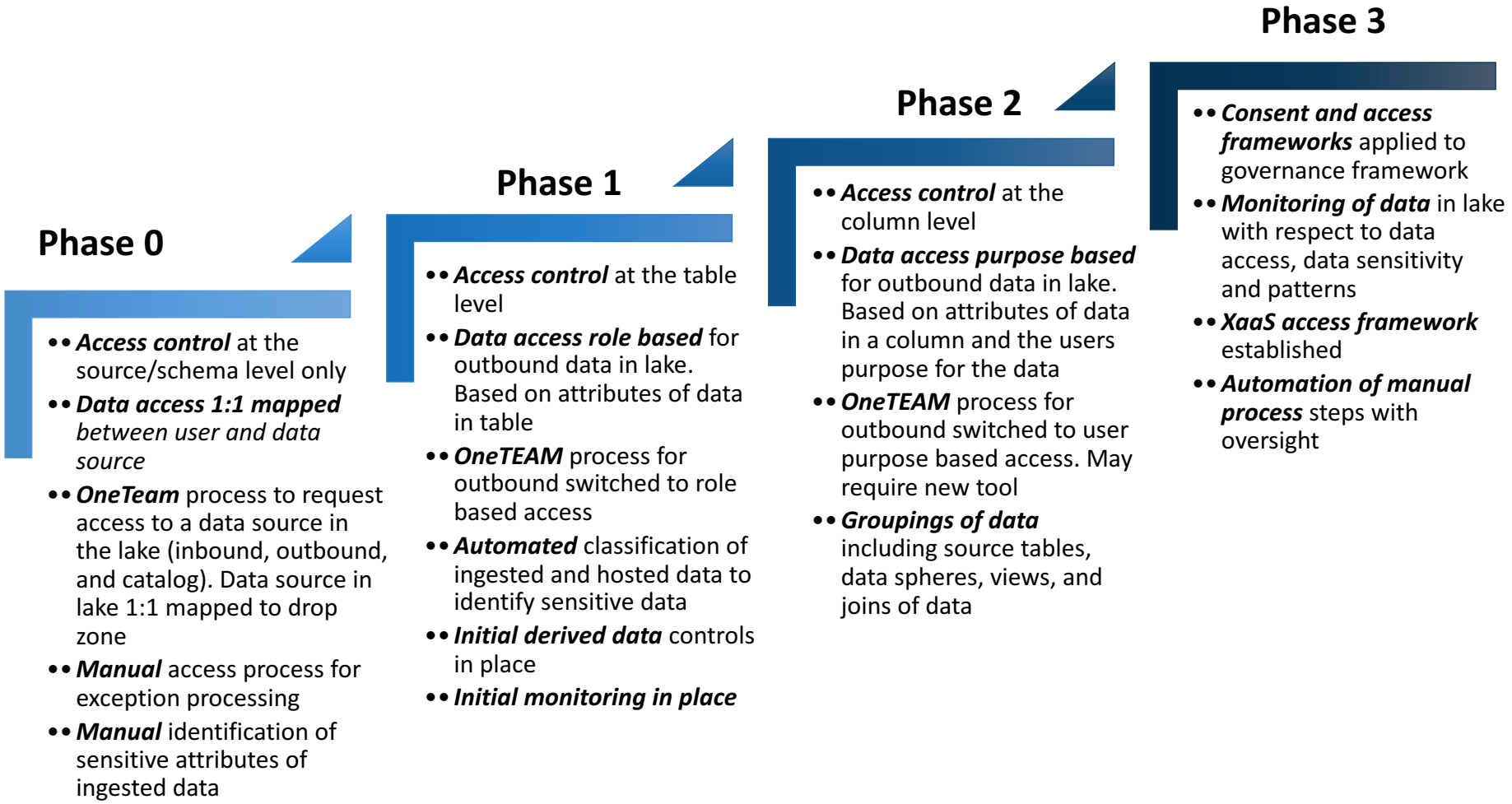
- GPA
- Acquisition process
- Governance review

- Access control
- Denial of service protection
- Transmission confidentiality and integrity
- Cryptographic protection
- Transmission of security attributes
- Protection of information at rest

Source: NIST Special Publication 800-53

- Monitoring
- Breach policies
- Cognitive audit

Access Control Epic Roadmap



0.5 Access Control (Phase 0)

Access Controls and Policies

0.5 Access Governance Policy

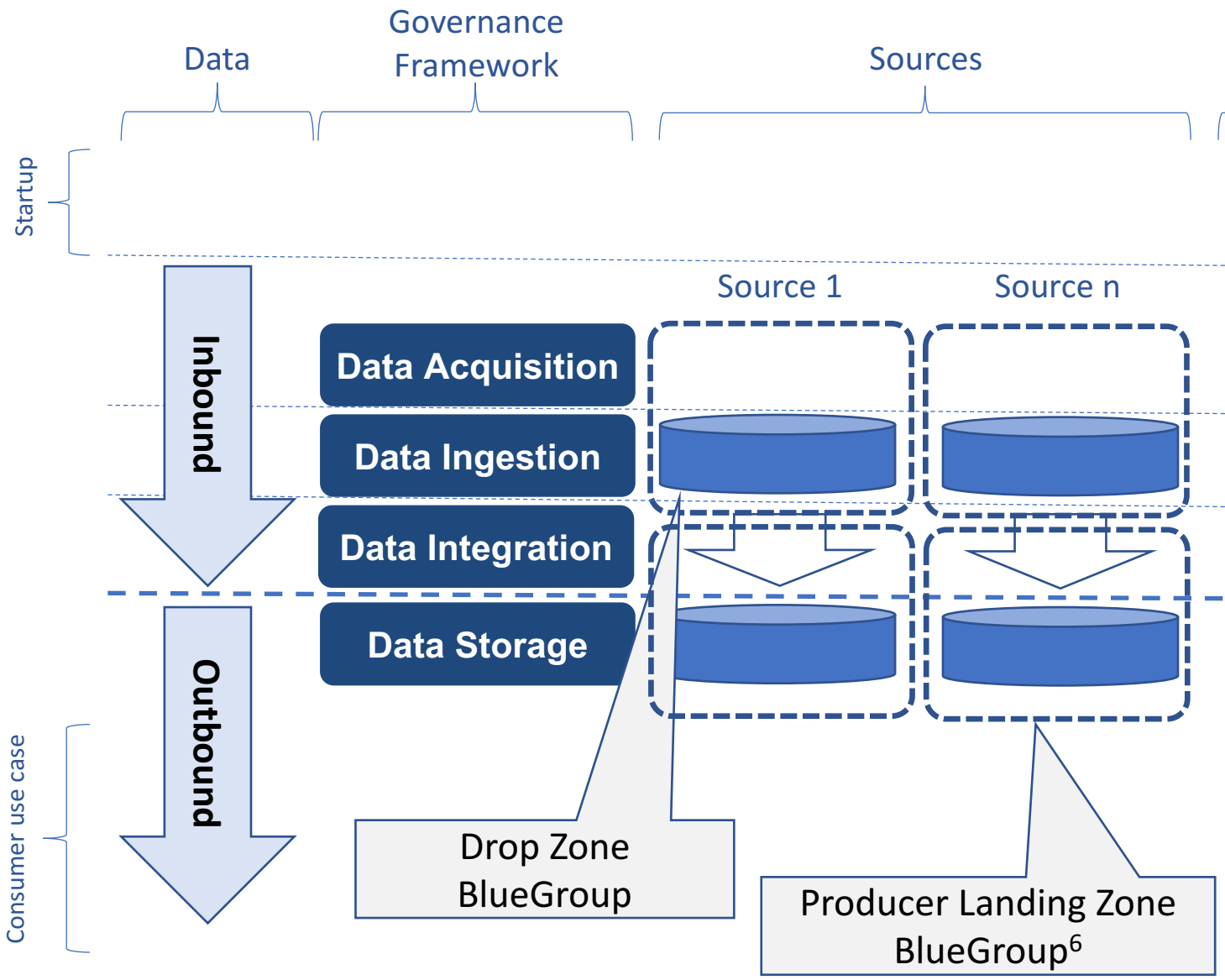
- GPA assessment must be completed on data sources that are to be accessed in the data lake
  - Data tagged if containing sensitive data
- Users must be authorized before accessing data
  - Users must request access to each individual source outside of the CoEDL
  - Business justification for consumer to original source cannot state they need access to the CoEDL
  - Users cannot access sensitive data unless they have the credentials mentioned in previous bullet
  - First line managers must approve user access to CoEDL
  - CDO approve user access to CoEDL
  - Separation of Duties statement must exist for approvers
  - User must be part of a BlueGroup associated with a block of data.
- No access allowed to restricted data during blackout dates for any user
- Control systems must be used for DB access
  - SAML provides user authentication
  - BlueGroup form the basis of authorization policy in the data lake
  - Bluepages provide user authorization to BigSQL and HDFS.
  - DB policy driven by LDAP for user access
  - Geography and divisional access integrity can be maintained via dedicated schemas or by means of views.
- Audit controls must be in place
  - OneTEAM logs for requests and approvals
  - Notifications of request activities sent to a dedicated channel for monitoring and audit
  - CEDP logs maintained for access
- Data must be encrypted in flight
- Data producers govern the access to their data in the lake
  - Revalidation and revocation of access to data managed by original source producers

0.5 Additional Statements

- The GCDO Information Governance organization governs the CEDP infrastructure and [provide the means for data asset owners to regulate access to their respective data sets
- This documents supports 0.5 with a minimal set of access requirements. Assumes:
  - Restricted and small set of users accessing the system inbound and outbound
  - Access is not granted all data in the data lake. Data must be approved for each source in the lake by the source.[1]
  - If access is granted for a user to a data source within the lake, then that user gets access to ALL the data in that source
- OneTEAM is the authoritative source for users and source access
- 1:1 map between inbound and outbound data sources at schema level
  - 1 BlueGroup for each (inbound and outbound)
  - CEDP--- BlueGroup gives access to data lake only
  - An internal LDAP defines authorization for data within the lake. This is synchronized with source BlueGroups
- Architecture and process review to be completed on entire system before release to assess adherence to policies

Notes: 1 – Exception: co-creators and dev have access to all the data in the CoEDL

Drop Zone (Inbound) Data Access Process



- OneTEAM[3] setup

- Source identified[1]
- BlueGroups provided by the … created[2]
- Producers identified[4] and associated with BlueGroup

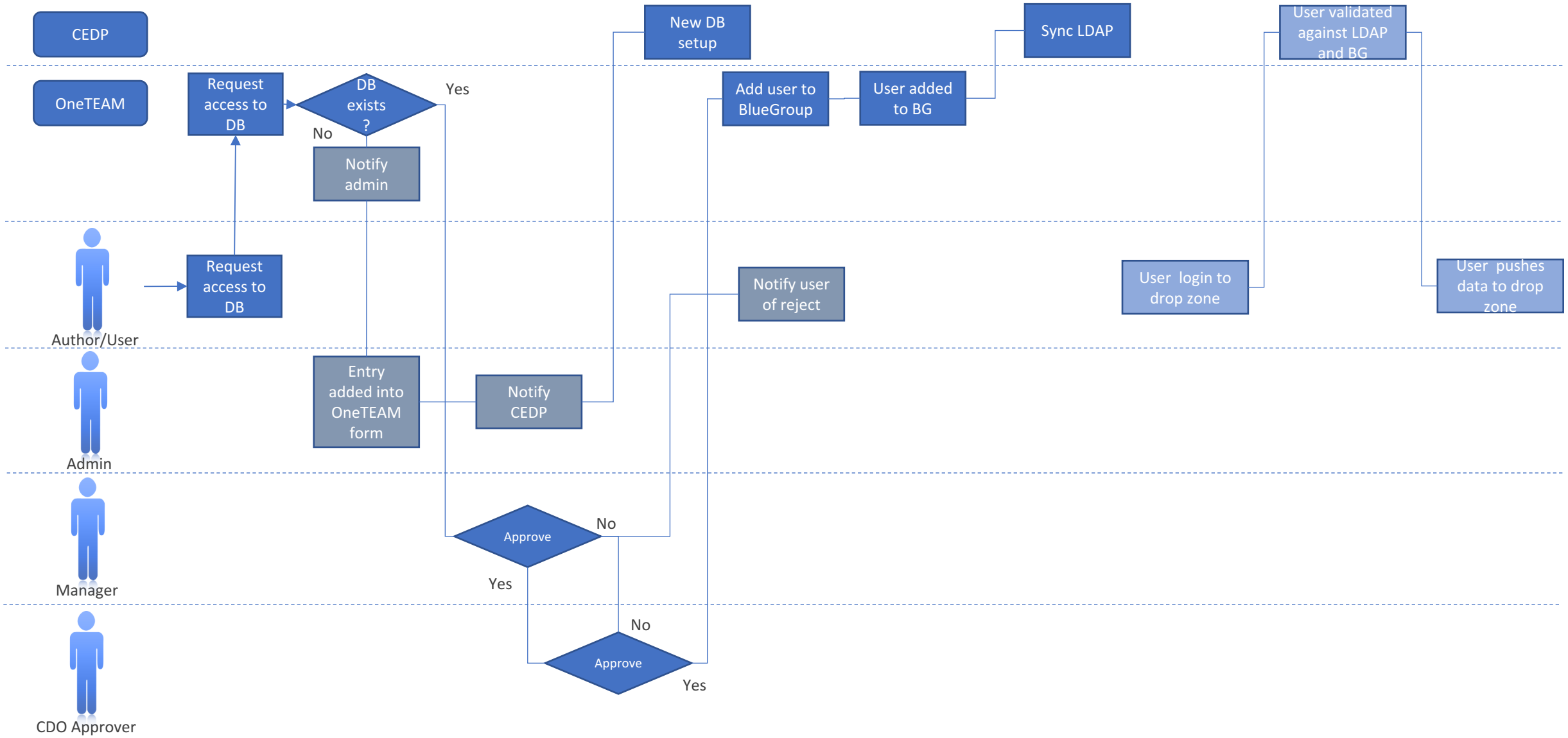
- Source DB created; LDAP synced with BlueGroup
- Data pushed (encrypted) to drop zone (read only)

- Data moved to production; drop zone deleted
- Notifications sent via email & Slack {?}

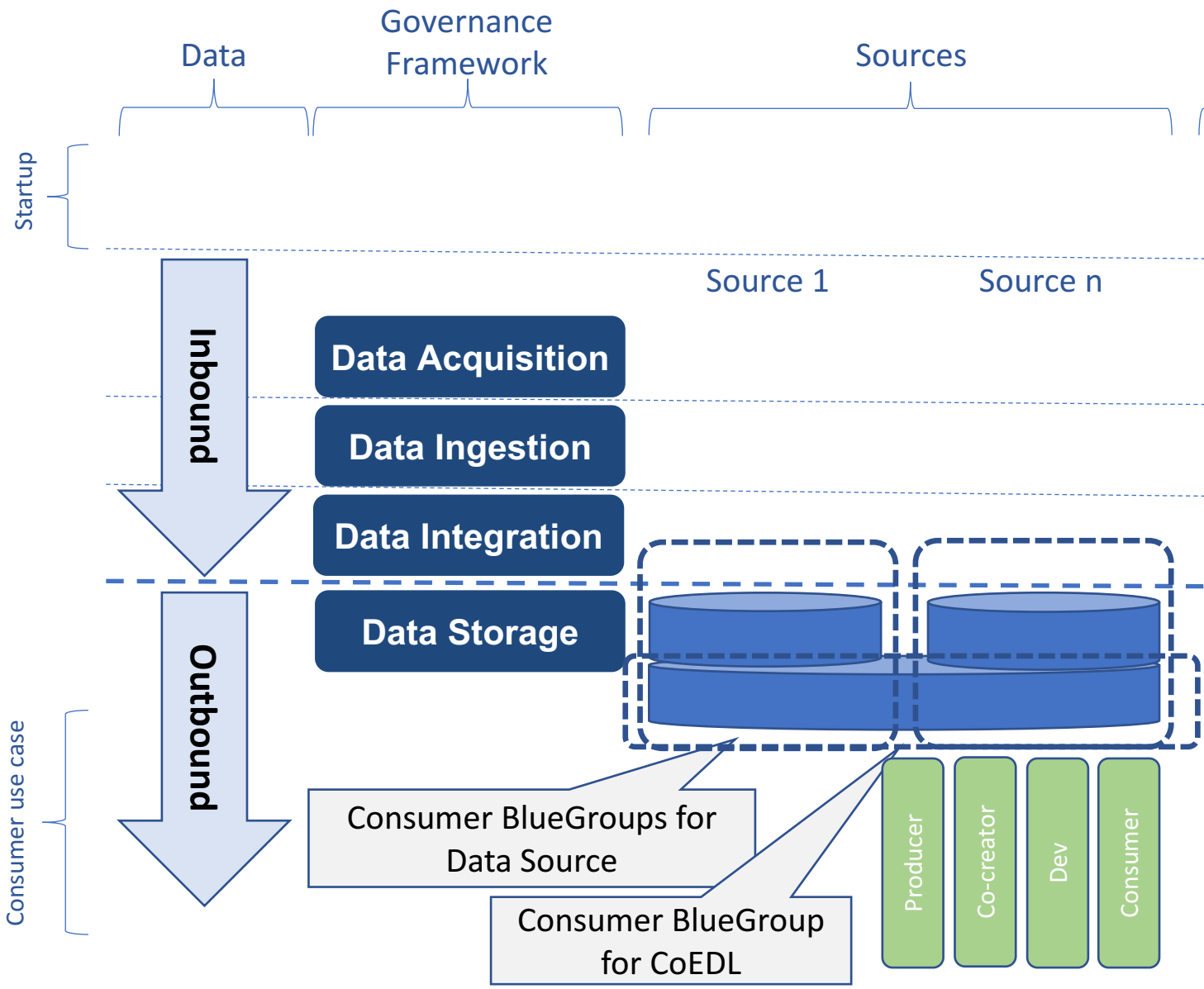
- User access[5] to table(s) requested via OneTeam[3]
- Approvals obtained
- Producer mapped against BlueGroup in landing zone

Notes: 1 – Does not include PI information from end user
2 – Single BlueGroup for each Data Sourceall inbound data
3 – Manual process in place as well where OneTEAM used
4 – First line management and CDO approval required
5 – Access granted at the source level
6 – Producer group to access their schemas. Different from consumer BlueGroups

Drop Zone (Inbound) Access Request Flow



Landing Zone (Outbound) Data Access Process



- User must already have access to original source DB, or must request it before applying for access
- CEDP BlueGroup is created

- At acquisition time, producer identifies bluegroup(s) associated with data source
- Existing users from source are added to CEDP BlueGroup

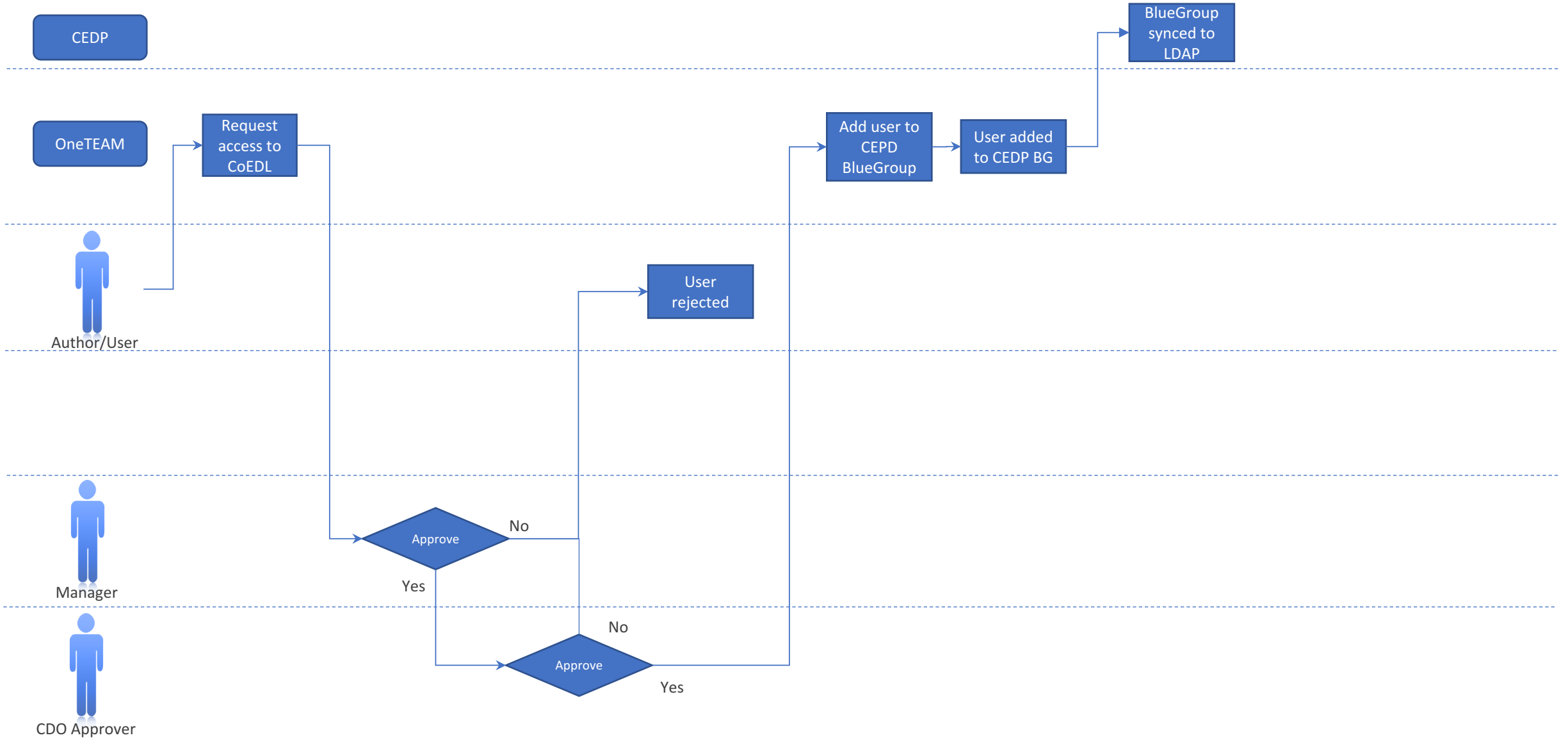
- User access[1] to CoEDL requested via OneTeam[2]
- Approvals obtained[3]
- User added to CEDP BlueGroup[4]
- Periodic sync of users in source BlueGroup(s) to schema LDAPs in CEDP

Notes: 1 – Access granted at the source level. Must show they already have access to original data source
2 – Manual process in place as well where OneTEAM used
3 – First line mgmt and CDO approval required
4 – Single BlueGroup for all outbound data to allow into CoEDL, individual BlueGroup(s) for each data source in CEDP

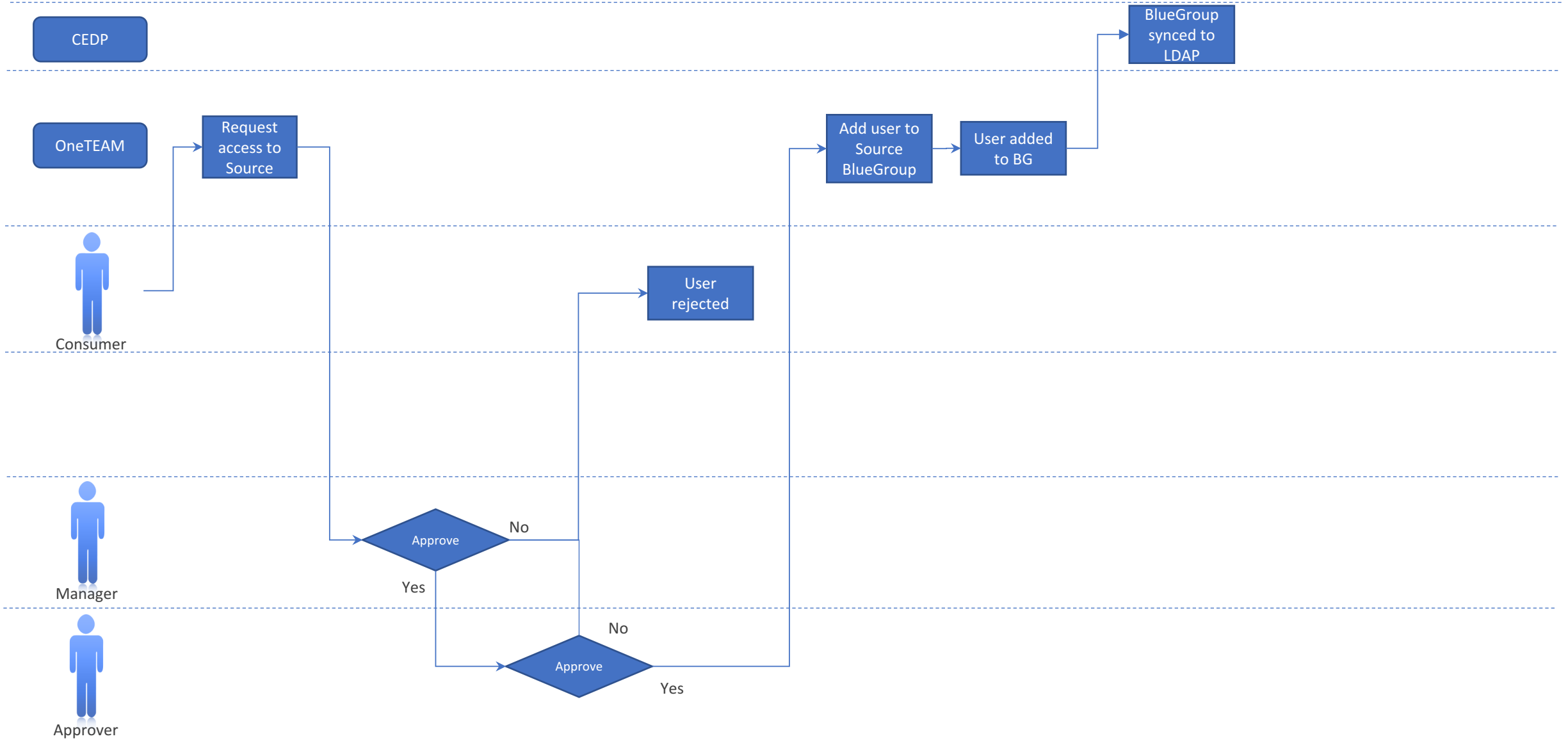
Outbound Access Control Scenarios

New Consumer	Existing consumer, new source	Co-creator	CDO Developer	Derived Data (Producer)
• Consumer requests and is granted access to the CEDP • Consumer gets individual schema and HDFS folder • Consumer requests and is granted access to the IGC (catalog)	• Access to CEDP has already been granted to consumer based on previous scenario • Consumer browses catalog and identifies new data source • Consumer requests and is granted access to the data source outside of CoEDL • Data producer adds consumer to appropriate Bluegroup • Bluegroup is synced to associated schema in CoEDL • Access to the new data is available to consumer	• Co-creator signs DOU for access to the CoEDL • Email request is sent to CDO • Co-creator is added to the co-creator Bluegroup • Co-creator schema and HDFS folder are created • Access to all the data in the CoEDL is available to co-creator	• Email request is sent to CDO • Developer is added to the developer Bluegroup • Developer schema and HDFS folder are created • Access to all the data in the CoEDL is available to developer	• Consumer has already requested and received access to CoEDL and data sources • Consumer creates new data set by joining their data sources • New data set exists in integration zone • Access to new data in IZ driven by BCG for consumer now producer • Consumer/producer can submit data back to acquisition process to create new data source • Consumer/producer owns creating and managing Bluegroup for access to new source

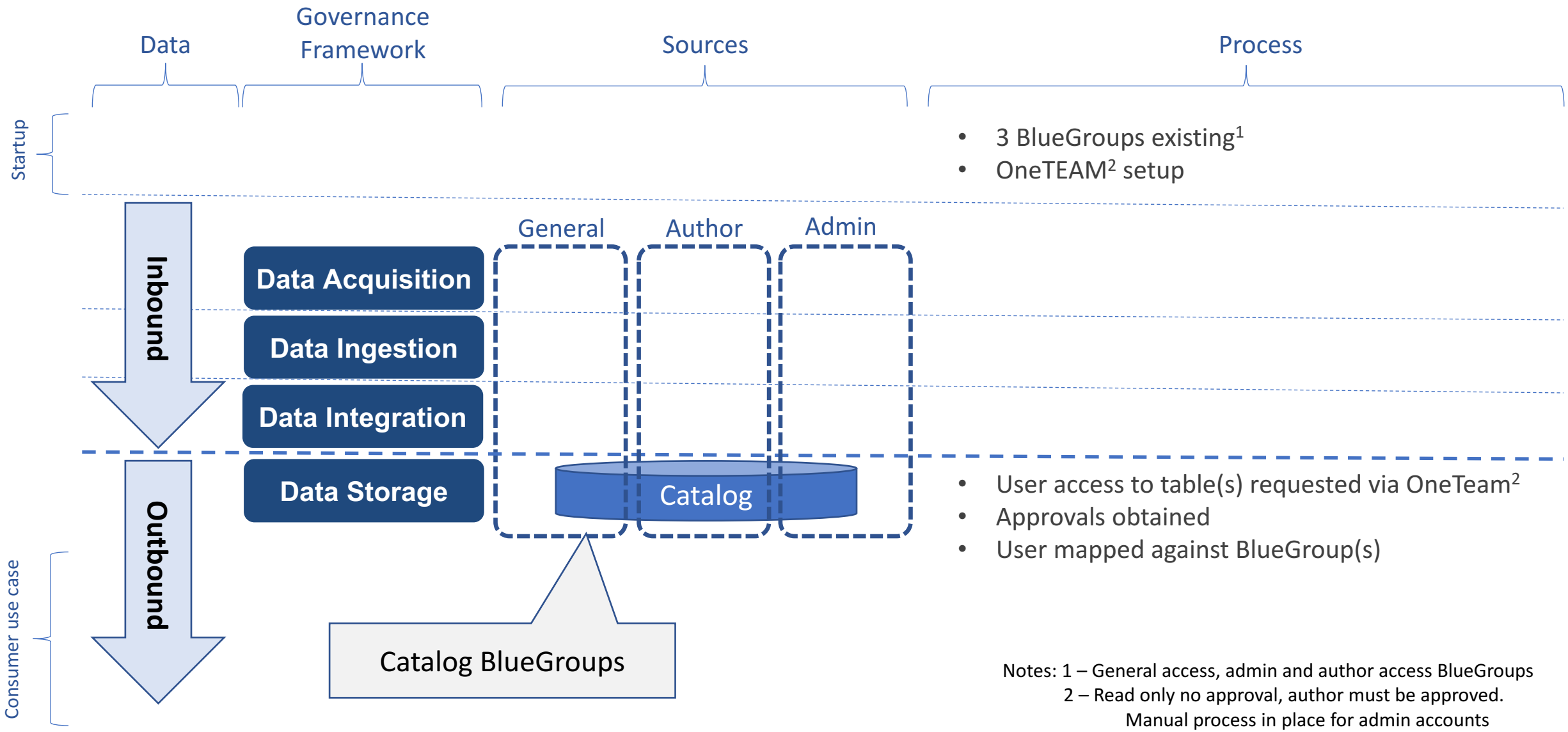
Outbound Access Request Flow - CEDP



Outbound Access Request Flow – Data Source

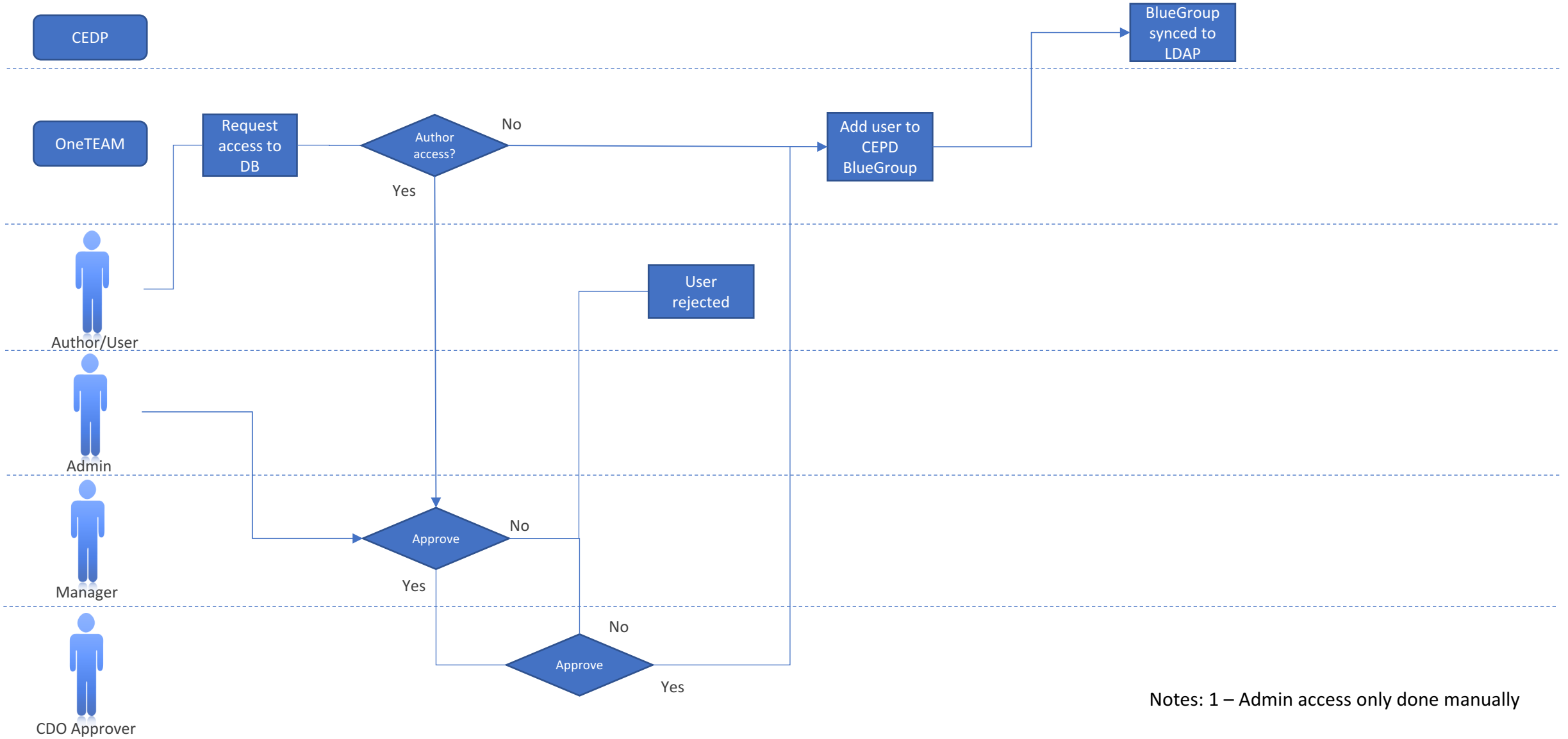


Catalog Data Access Process



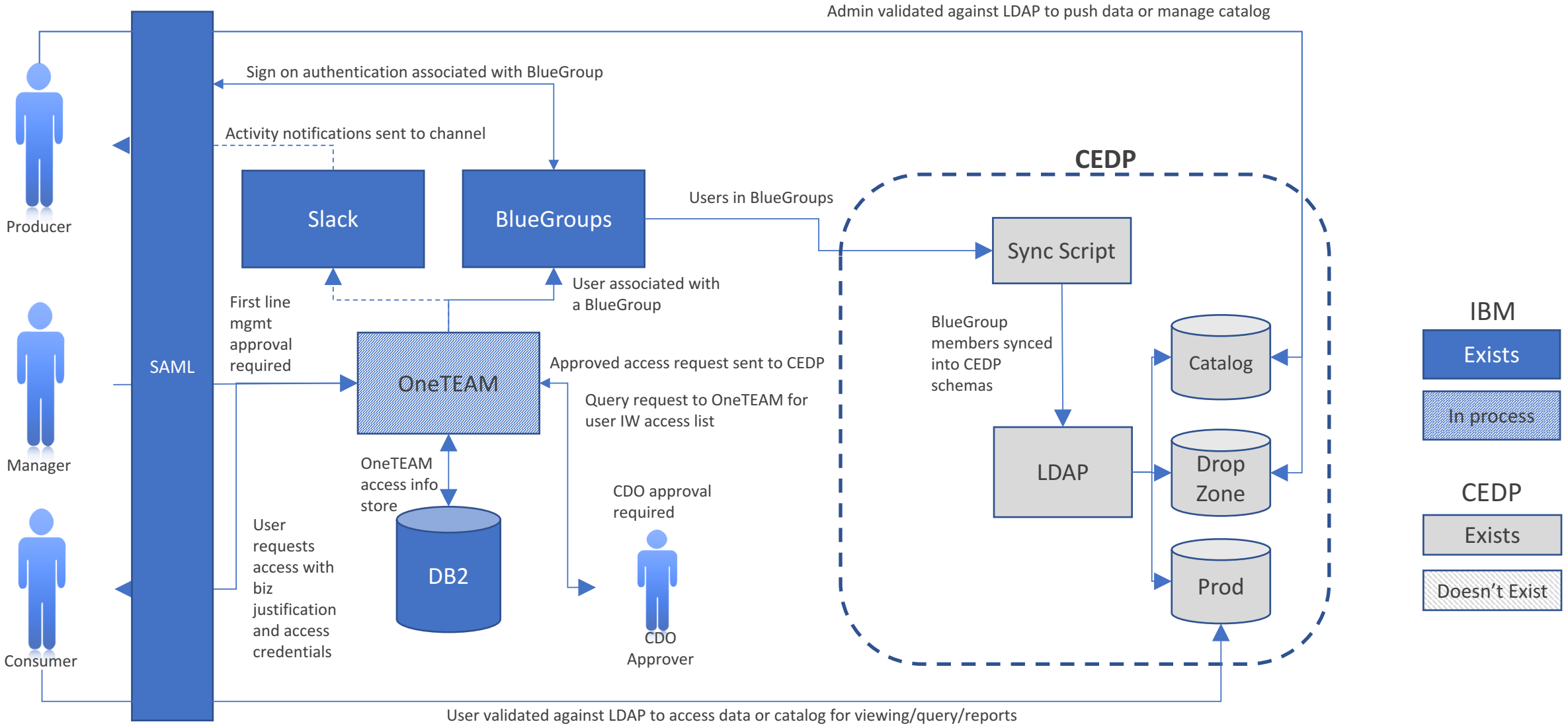
Notes: 1 – General access, admin and author access BlueGroups
2 – Read only no approval, author must be approved.
Manual process in place for admin accounts

Catalog Access Request Flow[1]

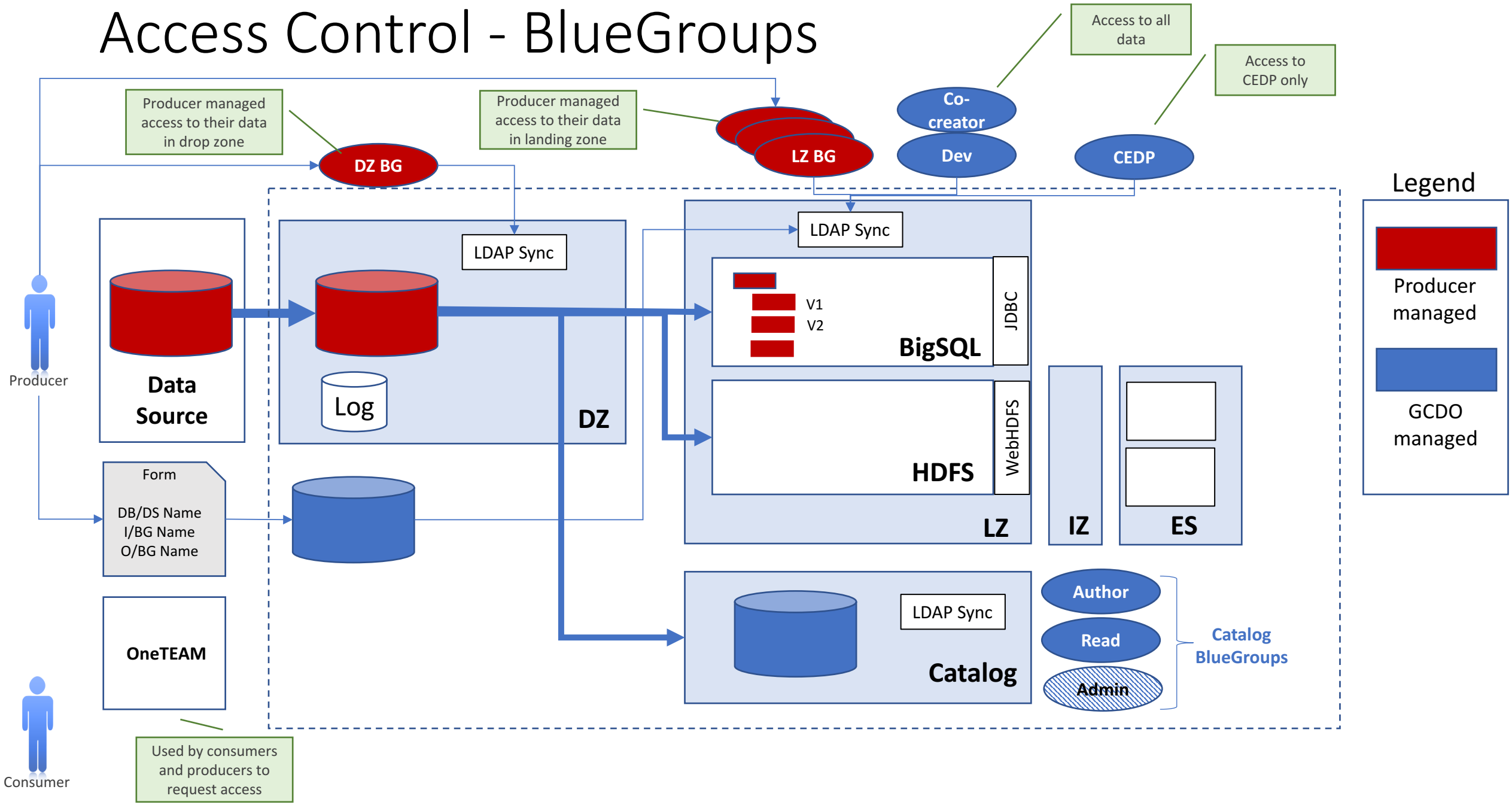


Notes: 1 – Admin access only done manually

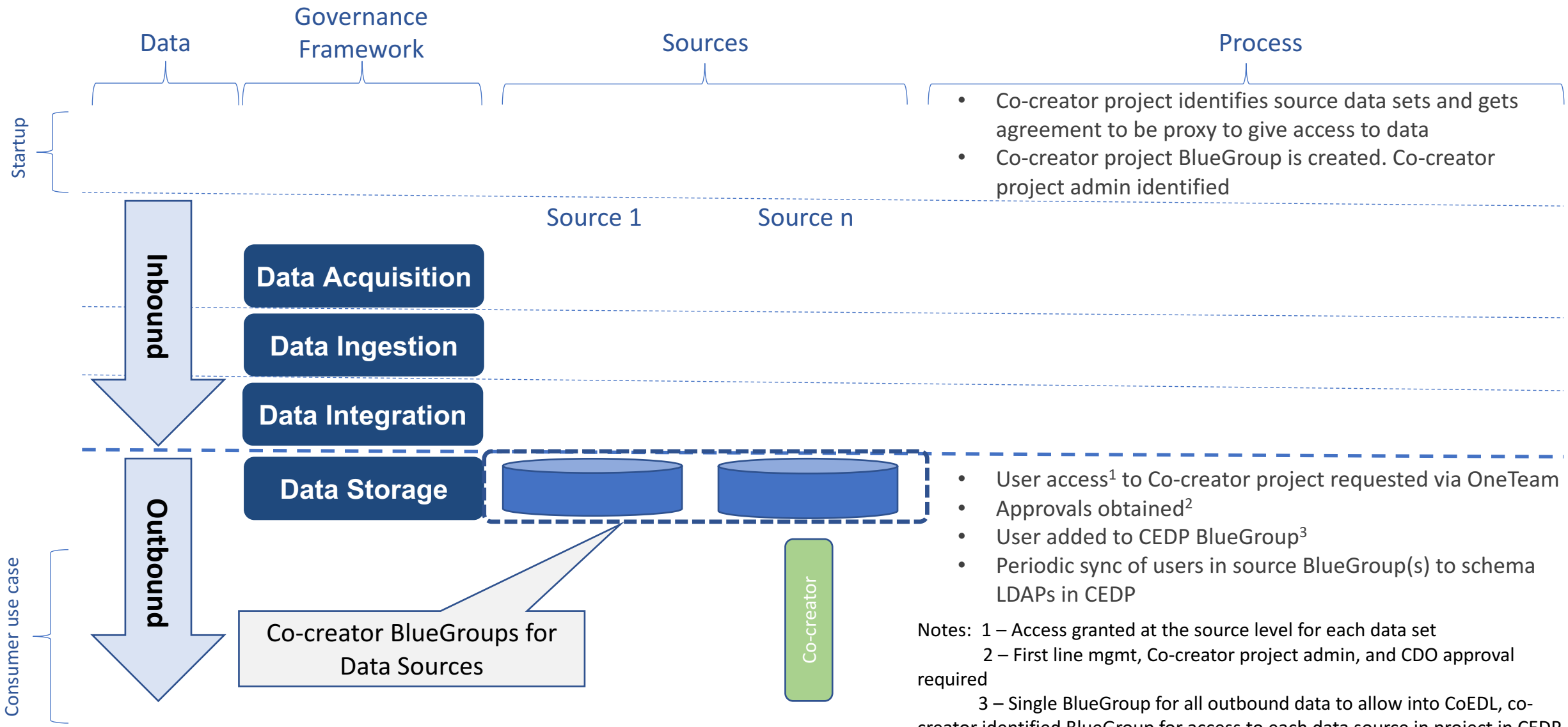
Data Access Control System View



Access Control - BlueGroups



Co-Creator (Outbound) Data Access Process

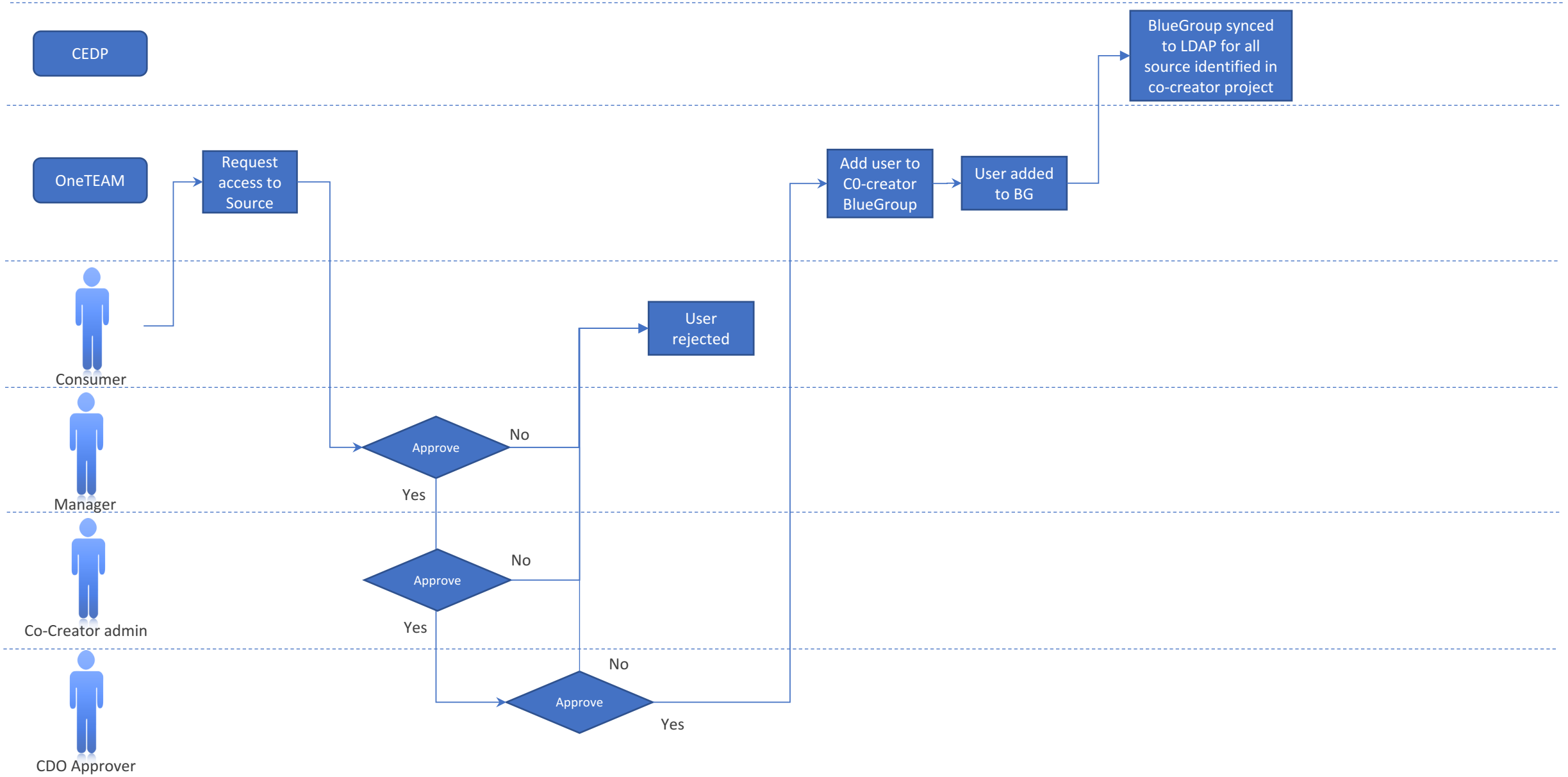


- Co-creator project identifies source data sets and gets agreement to be proxy to give access to data
- Co-creator project BlueGroup is created. Co-creator project admin identified

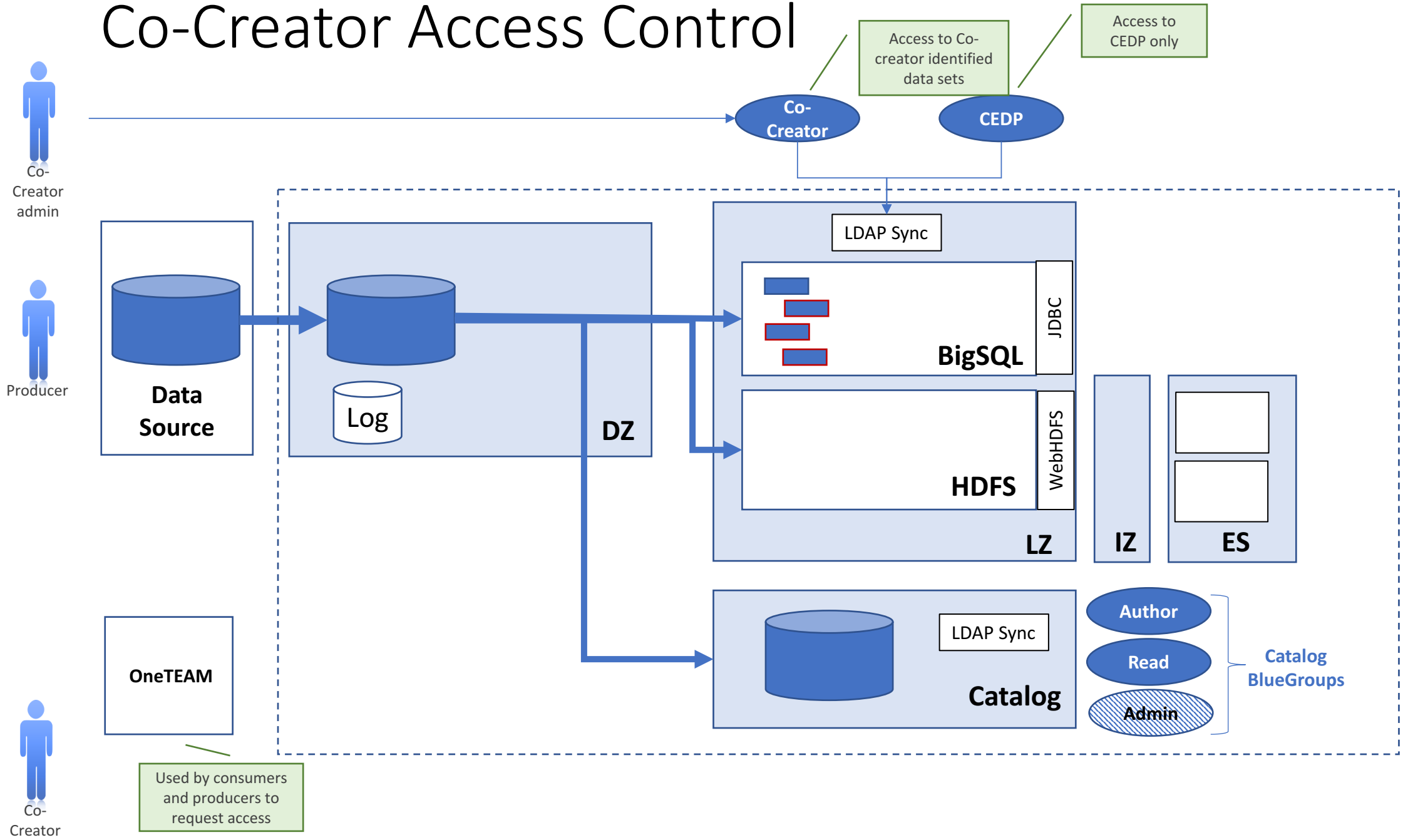
- User access[1] to Co-creator project requested via OneTeam
- Approvals obtained[2]
- User added to CEDP BlueGroup[3]
- Periodic sync of users in source BlueGroup(s) to schema LDAPs in CEDP

Notes: 1 – Access granted at the source level for each data set
2 – First line mgmt, Co-creator project admin, and CDO approval required
3 – Single BlueGroup for all outbound data to allow into CoEDL, co-creator identified BlueGroup for access to each data source in project in CEDP

Co-creator Access Request Flow



Co-Creator Access Control



Backup

Access Controls and Policies

Security Compliance Requirements

4.2 User Access Management

Access to systems, applications or information, whether granting or revoking, is a two-step process:

- An identity token (user ID) is assigned to a user, enabled, and ultimately disabled and/or revoked.
- Access rights are then assigned to, or revoked from, that user ID.

4.2.1 User Identities

a) Leverage centralized identity services, such as w3id or IBMid for applications.
b) Assign unique IDs to a single individual to ensure individual accountability.
c) Disable or remove User IDs upon separation of an employee, or at the end of a contractor engagement. Block or disable access to non-public IT systems within 24 hours.
d) Discard User IDs that are no longer in use. Do not reissue previously assigned User IDs to other users.
e) Provide a means to allow operating units to disable or remove User IDs immediately in case of an emergency.

4.2.2 Application or System Identities

a) Assign Application or System IDs, by default, to the owner of the application or the system, or otherwise assign to an authorized individual. The owner of the Application or System ID is accountable for its use. Application or System IDs may be transferred to a new owner when the original owner's employment or business need ends.
b) Disable or remove Application or System IDs that are no longer needed.
c) Change default vendor passwords immediately following installation of systems or software.

Security Compliance Requirements, Cont'd

4.2.3 Access Provisioning

a) Assign access rights on a business need or need-to-know basis.
b) Implement an authorization process for applications and systems if and as required by the BPO, or as required by the data or asset classification (section 2.0).
c) Complete authorization procedures prior to enabling access rights.

4.2.4 Managing Privileged Access Rights

Assign privileged access carefully and with the least amount of privilege required. Mitigate inappropriate use of privileged access rights, to prevent data breaches or failures.

a) Assign privileged access rights to users on a need-to-know basis, and based on the minimum requirements for the role.
b) Maintain an authorization process and record for assigning privileged access. Complete authorization procedures prior to enabling access rights.
c) Develop and implement procedures to avoid unauthorized use of generic administrative user IDs.
d) Limit privileged access to scripts or other executable utility programs to authorized users. Log and monitor the use of these tools.
e) Identify and track privileged access rights associated with each system or device that the privileged user oversees.
f) Maintain the confidentiality of secret authentication information for generic administrative user IDs (e.g. change passwords frequently and as soon as possible when a privileged user leaves or changes job, communicate them among privileged users with appropriate mechanisms).

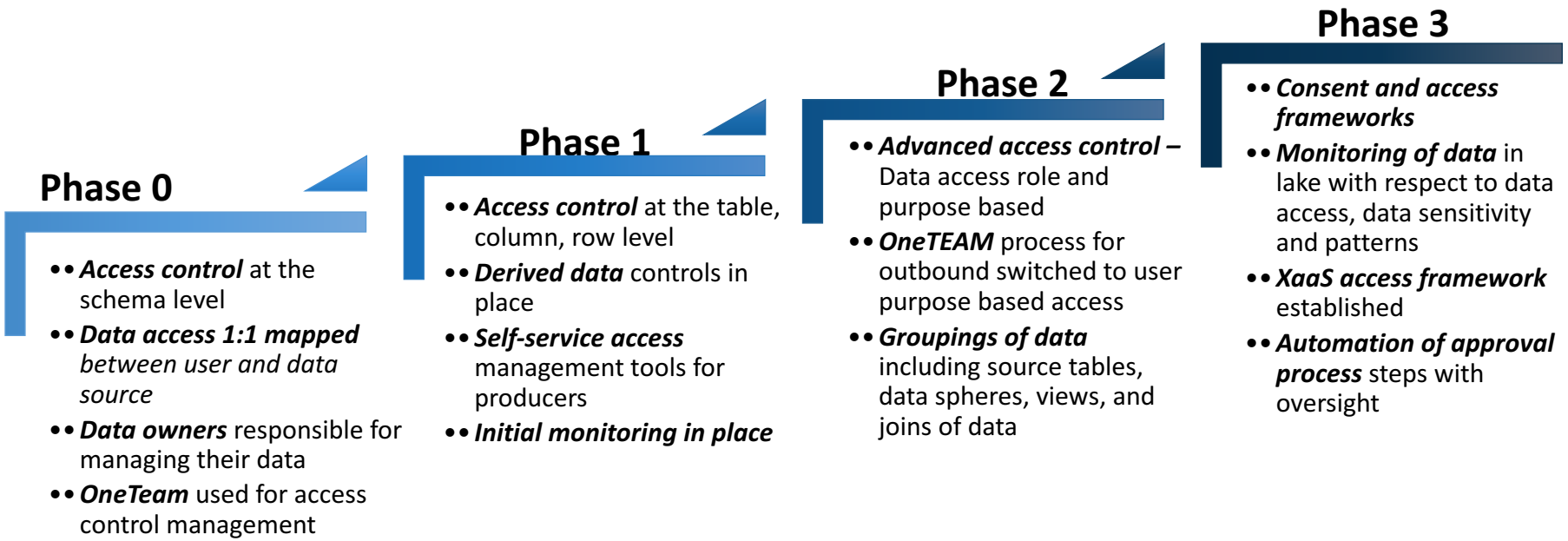
Security Compliance Requirements, Cont'd

4.2.5 Review of Access Rights

a) Implement a re-authorization process for applications and systems if and as required by the BPO, or as required by the data or asset classification (section 2.0).
b) Base re-authorization processes on defined milestones or defined time intervals, depending on the authorities or roles associated with the access or on regulatory or legal requirements. Milestones may include changes to the job role of an individual with access, or changes to the application or system that warrant a review of existing access rights.
c) Limit time-interval based re-authorization to a maximum of 12 months.
d) Review privileged access rights for continued business needs at least every 12 months.

4.2.6 Revocation of Access Rights

a) Remove access rights when there is no longer a business need for the employee or contractor to have the access.
b) Block or remove access within 24 hours of termination of employment for an IBM employee or contractor. Where IBM does not control access, notify the entity controlling access within 24 hours.
c) Provide a means to allow operating units to block or remove access immediately in case of an emergency.


Phase 0
•• Access control at the schema level
•• Data access 1:1 mapped between user and data source
•• Data owners responsible for managing their data
•• OneTeam used for access control management
Phase 1
•• Access control at the table, column, row level
•• Derived data controls in place
•• Self-service access management tools for producers
•• Initial monitoring in place
Phase 2
•• Advanced access control – Data access role and purpose based
•• OneTEAM process for outbound switched to user purpose based access
•• Groupings of data including source tables, data spheres, views, and joins of data
Phase 3
•• Consent and access frameworks
•• Monitoring of data in lake with respect to data access, data sensitivity and patterns
•• XaaS access framework established
•• Automation of approval process steps with oversight