BROAD INSTITUTE

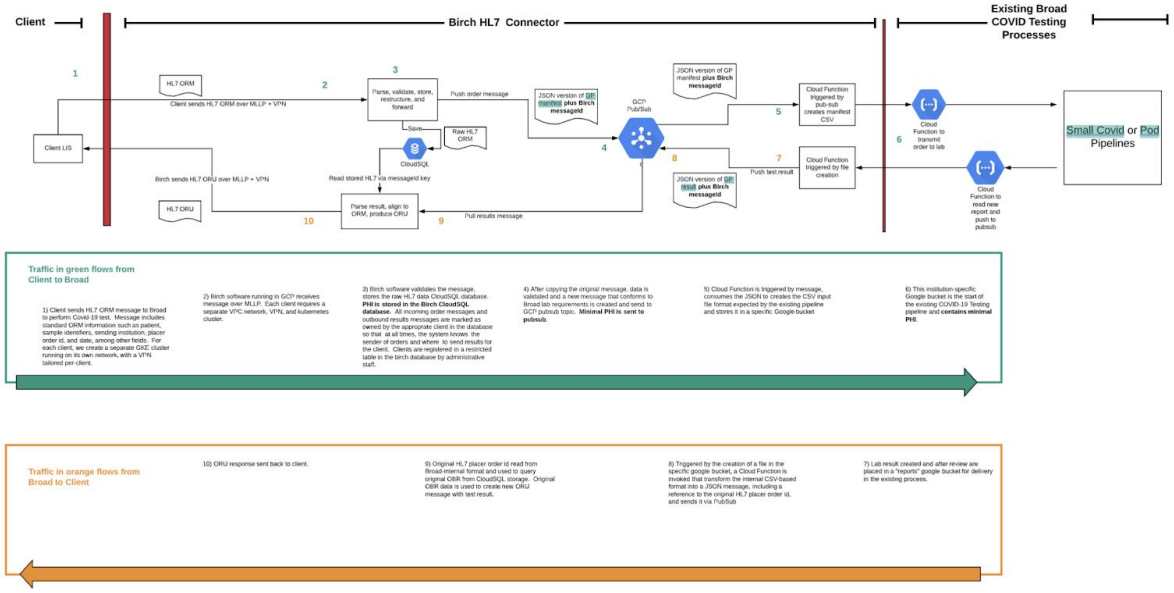
415 Main Street
Cambridge, MA 02142
T 617-714-7000 F 617-714-8972
www.broadinstitute.org

Broad COVID-19 Testing Informatics Infrastructure

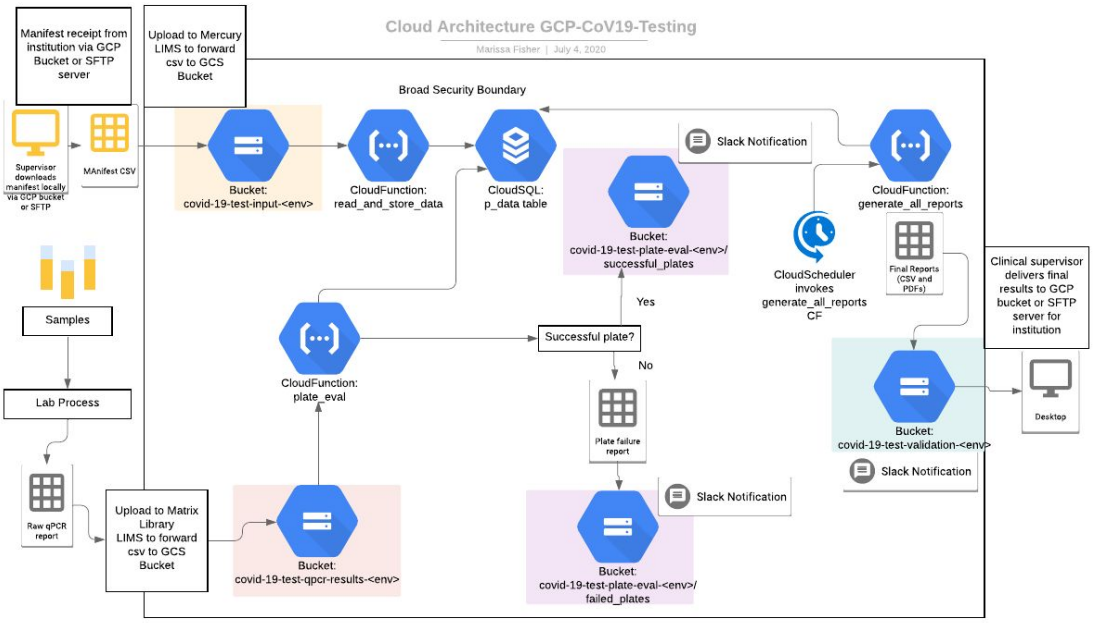
The architecture that Broad is using for this project is one that carries a minimum of risk and is HIPAA Security-rule compliant. The majority of it uses serverless computing and Google's robust Authentication, Authorization and assessment/auditing/protection controls.

The HL7-handling part of the pipeline is built to NIST-800-53 standards and the onprem part of the pipeline is built to NIST-800-171 standards.

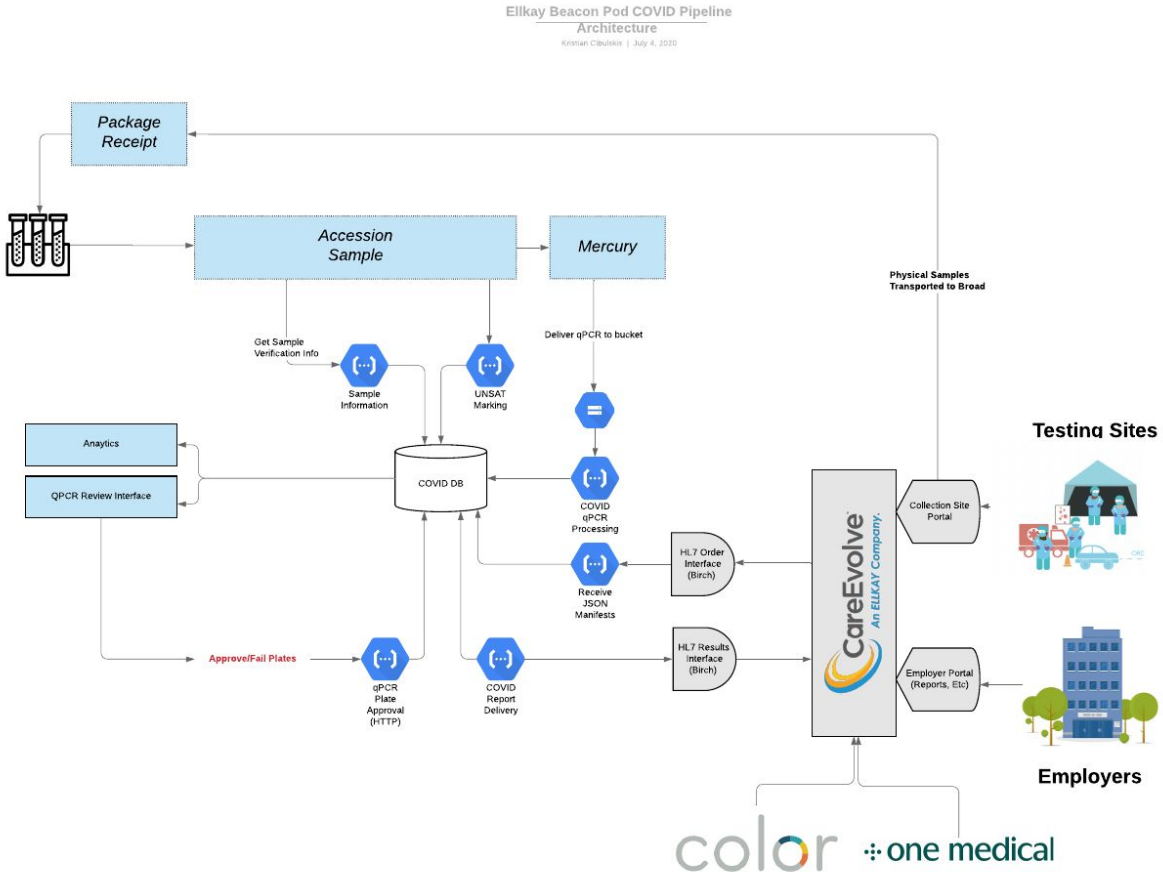
HL7 messages are sent by our HL7 Provider Partners (such as labs where swabs are taken) over client-specific VPNs into the Birch MLLP Kubernetes Cluster. From there, the data flows through event-driven serverless automation in Google's Cloud Platform. It eventually goes to our Onprem lab for actual biological processing, before being returned to Provider Partners via MLLP over VPN.



Birch Diagram - Click here to zoom in. The Birch HL7 handler connects external HL7 Clients (swabs) to Broad Labs. This system connects to our "Small Covid System -- deprecated" or the Pod System.



[Small Covid Diagram] - Click here to zoom in -- this installation is being deprecated for the newer “POD” below.



[POD Covid Diagram] - Click here to zoom in.

Google's Storage (GCS), Cloud Functions (GCF), Kubernetes Engine (GKE), and CloudSQL are all serverless and have no attack surfaces that aren't managed by Google. Other than the Birch MLLP services running on Google Kubernetes, Broad is not spinning up any VMs, networks, firewalls or machines for this infrastructure. The VMs in GKE are managed by GCP and are buttressed by their new Kubernetes Threat Detection. There is a minimum of security responsibility/configuration to Broad engineering.

After being ingested through our Birch pipeline, data goes into our on-prem facility for the actual biological testing. This on-prem facility is a CLIA Certified lab and is HIPAA compliant, though not required to be. The onprem infrastructure is monitored and managed aligned to NIST-800-171 as its guide. This infrastructure is currently used by the US Government for the AllOfUs program and has been authorized to operate under FISMA for that purpose.

Broad uses its own 24/7 Security Operations Center to watch logs and instrumentation coming from Google components to watch that all accesses to GCP are coming from authorized users. Broad's Application Security Staff and Offensive Security Staff ("red team") also review/threat model this infrastructure.

Features of this architecture:

- Only using Google managed components that are covered by Google's extensive security and compliance (including HIPAA and FedRAMP high)
- Exposed custom infrastructure is limited to authenticated VPN connections
- Logging and instrumentation goes to Broad's SOC for analysis and reporting. Same SOC and security staff that's used in our FedRAMP environment (Terra). All data access creates logging and instrumentation.
- End-to-end encryption on every single connection -- even internally
- Every human connection authenticated (with 2FA) and authorized (least privilege)
- All custom code goes through our SDLC and secure coding tools/policy. This features modern DevSecOps tools and techniques

Security appendix:

Logging and Incidents/Events

Broad Institute leverages a variety of logging and monitoring tools for its environment and is always adding new ones. Though logs will be aggregated to a SIEM in the future, for now these tools issue separate alerts but to a central location.

Tool	Log type
Splunk	VM and GCP/Buckets
Sentry.io	Exception handling/alerting

SentinelOne	VM/Endpoint protection
Google Cloud Security Console	GCP Misconfiguration alerts, GCP Security Health Analytics, GCP anomaly detection, GCP event threat detection/Kubernetes threat detection

All connections and transactions are logged and continuously monitored with a variety of software. Alerts and Actions are generated from this auditing practice.

All logs from GCP are sent to Splunk for additional analysis and alerting.

Infrastructure

GCP maintains audit trails of all of their functions. These audit trails are available in Google Stackdriver, where they are kept for 3 years searchable in Google BigQuery tables. Certain GCP logs for relevant services are also forwarded to Splunk for analysis and alerting.
All privileged functions used by Administrators are logged. These privileged functions require using a separate “sa-” account.

All creation/delete/modifications of Privileged users are logged and alerts are sent for each one to be audited via GSuite’s logging mechanisms.

All accesses to the custom app are handled via HTTPS (REST) calls to services, all of which are logged to GCP Stackdriver and go to Splunk:

- date/time
- type of event
- host that generated the message
- service that generated the message
- content of the message
- identity of user that was running the process

In addition, unhandled code exceptions and errors are sent to Sentry (sentry.io) for alerting to developers. This additional layer of monitoring and logging enables a faster and more in-depth response to code problems (and possible attacks) than simple logging does.

We manually review logs based on the above events. Each connection log captures **Who**, **What Action**, and **What Object**.

In addition to the real-time alerts, the following platform events are captured with some events triggering alerts - this list is not exhaustive and we continually add more alerts based on the MITRE ATT&CK framework.

- **Account Management** (privileged accounts only)
    - Account Created - Alerts
    - Account Deleted - Alerts
    - Account Disabled - Alerts
    - Account Expired - Auto Disables
    - Password Changed - Alerts
- **Google Cloud Platform**: All Stackdriver traceable events
    - **Google Cloud Storage**
        - File Created
        - File Deleted
        - File Read
        - File Write
        - File Permissions Changed
        - Object Access
    - **Network Events**: ACL Changed - alerts
- **Policy Change**: All - Alerts
- **Various outlier alerts based on continually changing criteria (threat hunting done by Blue Team)**

Our rationale for these alerts comes from several sources:

- Regulatory requirements
- Best practices as defined by NIST SP 800-92

Security Alert Handling

Audited events are automatically sent to the Systems Administrators in real time via email and Slack. Admins decide if there is an event that needs further investigation that warrants a formal Incident Response. If an Incident is determined, the Incident Response Policy will be engaged.

As per SDLC, developers push code to Github. As a client-side protection all developers are required to install tools such as git-secrets in order to prevent the commit of secrets in source code. As a server-side protection we have tools that detect if a secret is accidentally committed in source code and alert the security team and the developer approximately 2-4 seconds after the incident. This type of real-time detection allows the security team and the developer to take the appropriate next steps and revoke secrets, review logs, remove the incident from git history, and make sure the incident does not happen again.

Assessment

All code is scanned weekly with SourceClear for downstream dependency vulnerabilities.

Broad performs vulnerability scanning of OS/Infrastructure assets and web applications at least monthly. OS/Infrastructure vulnerability scans are conducted by the Broad Information Technology Services (BITS) team. Web application scans are conducted by the Application

Security team, who leverages Zap for automated vulnerability scanning and Burp for manual testing. Raw results from each of these scans are triaged by the team responsible for conducting the scan. Findings are then tracked to remediation.

Any findings classified as high risk/impact and confirmed as legitimate findings trigger a review of the system's historic audit logs to see if the vulnerability had been previously exploited.

Details about security assessment, flaw remediation, and how security fits into our design process is in the SDLC. Finding severities assigned by automated scanning tools may be reassigned by the Security Auditor. We triage findings from multiple sources, including scanners, and assess the severity. In general, High severity vulnerabilities are defined as having a CVSS score of 7 and above and/or being remotely executable. Most other findings fall into our Low or Moderate classifications.

A list of tools used for Vulnerability detection and assessments is here:
https://dsp-security.broadinstitute.org/appsec-team-internal/appsec-team-internal/security-tools

Scanning Table

Type of Scan	What kind of scan is it?	What does it imply?	How is it updated (for new signatures)?	What action is taken?
Google Cloud Security Console	Infrastructure as a service scan	Continually assesses the GCP Projects for alignment to CIS Benchmarks. Alerts go to Slack #dsp-gcp-security-alerts.	Managed by GCP	Broad Infosec gets alert in Slack and takes action after triage.
Github Security Tools	Github	Github security tools is a suite of tools that DSP AppSec Team has built in order to scan for: • Secrets/API keys that are committed in source code • If some user does not have 2FA installed • Repository-level vulnerability alerts automation	DSP AppSec team maintains the project	DSP AppSec team receives Slack alerts #github-security-alerts and follows up on next steps in order to remediate the issue.
Zap Scans	Application scan/DAST	Continuous Dynamic Application Security Testing scans	OWASP ZAP community maintains the open-source ZAP project and DSP AppSec maintains the DSP-specific	Results from scans go to our Vulnerability Management and Correlation Platform, CodeDx where they are then triaged by the

			automation parts.	DSP AppSec Team.
Burp Scans	Application scan/DAST	Continuous DAST scans by leveraging and intercepting traffic generated by tests that are developed either by teams or QA at DSP.	Burp Updates are managed by PortSwigger. DSP AppSec Team manages the containerization of Burp, its automation, and pipeline configuration.	DSP AppSec Team triages issues and follows up with next steps.
Codacy Scan	Static code analysis	Tests code without actually executing it — this is called a non-run-time environment.	Managed by Codacy.	DSP AppSec Team triages issues and follows up with next steps.
SourceClear Scan	Dependency Analysis	Sourceclear is a tool that the DSP AppSec Team has plugged into developers' existing workflows to examine security risks of open-source and third-party code in real time.	Veracode manages SourceClear.	DSP AppSec Team triages issues and follows up with the next steps, working closely with developers to make sure that 3rd party dependencies are properly maintained and updated.
Certificate Transparency Monitoring	Custom monitoring and alerting	Detect potential certificate issuance by leveraging certificate transparency logs, a project that initially was developed by Google to ensure that Certificate Authorities *transparently* issue new certificates.	Managed by DSP AppSec Team.	DSP AppSec Team triages for newly issued certificates to make sure there is no certificate missisuance for domains such as broadinstitute.org, firecloud.org, and terra.bio.