|  |  |  |  |
| --- | --- | --- | --- |
| **Feature/Usage** | **Cloud** | **Appliance** | **Notes** |
| Privacy and Data Protection |  | Easier | IT objections, HIPAA implication and other non-technical barriers. |
| Sales Cycle |  | Longer | Upfront human & capital investment by client |
| Cost Effectiveness | \*Better |  | \*If we have economy of scale (multi-tenancy) |
| Secondary Analysis |  | Easier and Faster | Data upload on client site |
| Development Cost | Higher | High | PAAS has many more parts (except for HW) |
| Maintenance Cost | Low | High | Appliance SW and HW upgrade cycles. May leverage IAAS/PAAS vendors for Cloud. |
| Customization | Easier |  |  |
| Tertiary Analysis | Easier |  | Access to fresh content/model. |
| Latency |  | Higher | Higher grade HW and SW acceleration. |
| Throughput | Higher |  | Scalability and elasticity of computing resources. |
| Scale | High | Low |  |
| Sharing and distribution | Yes | Limited | Limited appliance access |
| Content Integration | Yes | Difficult | Central repository |
| Aggregated Analysis | Yes | No | Per client appliance, limited appliance storage |
| Data Archival | Yes | No | Virtually unlimited cloud storage (and multi-site) |
| High Availability | Yes | No |  |
| Eco-System | Larger |  | May include 3rd Party app/tool developers |
| Turnaround Time | High | Low | Data upload, scheduling |
| \*Phenotype | Possible | Less Likely | EMR integration possible? |

Core Value Proposition for the Bina Cloud

* For researchers:
  + Access to large sample pool for aggregated analysis
  + Access to large resource pool for large/repeat studies
  + \*Experiment with new tools
* For clinicians:
  + Access to content
  + Access to tools (esp. Bina tools)
  + Customize analysis
* For doctors:
  + Access to medical content
  + Sharing and interacting with patients
* For patients:
  + Hosting of medical history (individual and family)
  + Sharing and interacting with doctors
  + Access to education content
* \*For ISV:
  + Access to data and compute resources for the development of new tools

Scale

* 1000 genomes now (low coverage), 10,000 genomes soon (high coverage)
* Millions of genomes eventually
* \*Archival of secondary analysis data (e.g., reads/FASTQ/BAM)
* Online access to tertiary analysis data (e.g., variants/VCF)
* Aggregation of data from known databases (hundreds)

Phase 0: “Virtual Bina Box”

Primary goals:

Business Development for prospects

Secondary goals:

Assess Bina scalability, performance and cost in the cloud setting

Increase integration test coverage (test on all data all the time)

Experiment with the bleeding edge cloud stack for Bina

Nice-to-haves:

Help collection data for Bina box tuning

Scope:

No multi-tenancy (1 cluster = 1 Bina box)

Distributed storage (HDFS) and configuration

Phase 1: SSAS (hosting of data and analysis)

Phase 1B: Content (genomes and interpretive data)

Phase 2: PAAS (open platform)