**Big Data Security Meeting 9/9 10AM – 12PM Mountain**

**Attendees:**

|  |  |
| --- | --- |
| Richard Henning | Paul Gollapudi |
| Hari Ganesan | Adam Taylor |
| Sunil Kumar | Sunil Kumar |
| Debra Flippo | Sampath Anne |
| Joe Gance | Bob Marshall |
| Nagarajan Subbiah | Dan Ferretti |

**Meeting Purpose:**

* Review Data Security Plan for Big Data Program
* Review Technical Implementation Approach

**Challenges**

* Enabling offshore developer access to the data
* Impetus activities planned for Sprint 2 starting 9/14 include:

| **Sprint** | **Duration** | **Task/Milestones** | **Projected Supplier Resources** | **Key Objectives / Deliverables** |
| --- | --- | --- | --- | --- |
| 2 | 2 week | • Data Ingest, Exploratory Data Analysis & Knowledge Transfer (“KT”)  • Requirements Finalization & Sign Off | • 2 Impetus Big Data Architect  • 3 Impetus Sr. Big Data Engineer  • 2 Impetus Data Scientist  • 0.5 Project Manager | • Conduct EDA and provide recommendations and/or requirements on additional data needed and/or required transformation of or aggregations of data.  • Architect and design the solution.  • Review approach and architecture with CenturyLink team and obtain signoff.  **Deliverables –**  • Detailed sprint wise project plan.  • Acceptance on data sources to be ingested.  • Detailed Architecture/Approach documentation.  • Acceptance criteria definition and alignment for DS deliverables. |

**Constraints:**

* Access to the Big Data Lab Environment is restricted to on shore resources (H1B holders are OK)
* Number of resources having access to the unsecured data is being limited / tightly controlled at this time.
* Data cannot be exposed to offshore resources or additional onshore resources until security constraints are clarified

**High Level Approach:**

* Focus immediate efforts on a temporary risk mitigation path utilizing the Big Data Lab Environment temporarily until planned Big Data environment setup is completed
  + Temporary landing zone for source data
  + Limited to on-shore access only
  + Limited number of personnel with access
* Data will be scrubbed, tokenized, masked and filtered in the LAB Environment to create secured data files
* Secured data files would then be exported and loaded into a Cloud-Based Big Data Environment created for actual development
* B2B Cloud Based Cloudera Env to be used by Impetus for B2B Use Case development
  + Per Joe G., the B2B Cloud Based Cloudera Env has been built and is ready
  + Chris Gardner and Keith Young are familiar with this effort.  Chris is on vacation this week.
* Andrew Clyne will be talking with Aamir regarding a request for security waiver allowing offshore access to the secured / filtered data in the B2B Cloud Env.
* Engage Security personnel to participant in the process and confirm data has been secured prior to loading into cloud
  + Need to identify appropriate empowered resources
  + Confirm need to go thru IBRR process prior to export
* Unless a need for offshore access is identified to data related to the Network use cases, that Network related will remain in the LAB environment

**Overview of Planned Activities**

* Focus on data needed most / 1st by Impetus
  + Vijay Joshi will provide document
    - List of prioritized data sources ranked 1-N
    - List of fields in each source
    - Initial assessment of review of fields –
      * White list– No security concern expected
      * Black list – Security concern expected
      * Gray list – Uncertain – further clarification / security ruling required
  + Identify “Whitelist” data sources and get those moving first
  + Identify highest value “Black” / “Gray” list data source and begin development work to “secure” the data via masking, filtering, tokenizing, etc…
    - Leverage “out-of-box” Cloudera functionality
    - <http://www.cloudera.com/documentation/enterprise/latest/topics/sg_redaction.html>
  + Data masking document from Vijay Joshi is expected to explain in detail what is needed to secure the data.
    - Linda Soon has this doc and will share.  She said she would share.
* Form Team
  + Joe Gance / Richard H.
  + PM Support -TBD
  + Nagarajan Subbiah / Sunil Kumar – CLC Developers to be used to secure the data
  + CLC Developers will be paired with Cognilytics developers for assistance / direction
    - Santi Dash; Leela Katari
  + Impetus – Vijay Joshi
    - Impetus on-shore resources to provide direction –Names TBD
  + Cloudera resources – Hari G. and Bob Marshall will support the effort
  + Keith Young and Chris Gardner or equivalents
  + Lori Shields - Architecture
* Schedule GlobalMeeting Sessions
  + Early Next Week to begin work to ingest and secure the data
* Lab Access requested from Dan Ferretti and Russ Jesser for
  + Richard Henning AB68892 - Hue Access
  + Nagarajan Subbiah NXS918221 – Command line access and Hue Access
  + Sunil Kumar AB44252 – Command line access and Hue Access
* Goal – Get as much data as soon as possible to Impetus to allow work to begin
* Confirm whether IRRC approval is required or whether Keith / Chris or equivalent can be added to approve data going into B2B Cloud Env.

**Notes:**

* Robert Marshall = Hadoop Administration
* Hari Ganesam = Cloudera and Hadoop Technical (provide expertise to use the tools to ingest data, and lab environment)

**Risks:**

* Potential to impact the Cloudera deliverables / timeline due to unknown amount of time needed to support the data moves. Actual impacts will depend upon team resources’ level of knowledge and ramp-up time required.