# Customer Project Startup Checklist Revision 2.3

## Note: Text in blue is example customer input. Please replace with customer own input.

## Customer:Customer Name.

## Purpose of Project:A proof-of-concept (POC) for Virtual Master Patient Index.

## Use Case – Describe:

A health worker logs on to a browser screen using and enters search criteria for a patient. If more than one patient is found, the screen should ask for additional search criteria. Once a single patient is found and confirmed to be correct, the health worker should be able to see all data for that patient organized by FHIR HL7-based categories from any and all systems. The data should be standardized according to FHIR HL7. The data seen by the health worker should be determined by a centralized Active Directory role (RBAC) and row-level security (RLS), which will, in-turn, determine what and how that data can be seen (CLS) and data masking and tokenization, and user credential will determine which patients’ data can be seen (RLS). The data should remain in systems and only read as needed in a result-set. Reports should be generate using Tableau

## Timeframe:

Estimated timeframe is 90 days, depending on the final Statement of Work (SOW).

## References:

Ref. 1: <https://www.whamtech.com/documents>

## Requirements:

How many systems are involved?
Unknown at this time.

How many data sources/databases are in each system?
Unknown at this time, but for the POC, would prefer to be limited to three.

What business functions are these data sources used for and an estimate of number – please list?
Operational EHR - 10
Patient management - 3
Lab results - 25

Who are the gatekeepers to each system and what are the access procedures/requirements/restrictions for ‘read’ access to each system? Name 1 (IT Director)/Name 2 (Server Administrator)

### For Each Data Source:

1. Name and type? Works.
2. Hardware? Virtual Machine.
3. Operating system? Windows Server 2019 R1.
4. Drivers available? ODBC.
5. Batch, incremental or real-time update? Real-time.
6. Message queues? Needs clarification.
7. Replication and/or backup in place now? Yes.
8. Percent structured and percent unstructured data? 19% structured/81% unstructured (ballpark).
9. Number of records? 265,000,000.
10. Storage used for entire data source system? Unknown at this time.
11. Storage used for data only? Unknown at this time.
12. % amount of data required for indexing? Unknown at this time.
13. Peak rate of records added? 1,000 per hour.
14. Peak time of day for records added? During normal business hours (8 AM-5 PM PST).
15. For relational data, how many tables? 1,512 (based on a query ran).
16. For relational data, how many records in largest table? Vendor\_Item\_Extension (largest table name) 115,498,862 (Row\_count).
17. Is a copy of the schema and number of records per table – basic stats – available? Requested from vendor.
18. Can we get a sample of the data source? Yes.
19. Access control in place? Yes.
20. HIPAA or other regulations govern? HIPAA.

Is there a preferred standard data model, e.g., HL7? FHIR HL7.

What are the types of reporting applications anticipated, e.g., Cognos, Power BI, Tableau, HTML, Excel, etc.?

Excel, Power BI and possibly XML (for our CQS Data).

What are the types of queries expected, plus some examples of each? Unknown at this time.

Is there a user-level security system in place? Yes.

Are there requirements for additional security? Company internal security standards.

Are there additional items required, such as microservices, entity extraction or Business Process Modeling (BPM)? TBD.

Any additional requirements not anticipated above? TBD.

## Assumptions for POC – production would be different:

* EIQ Adapter software running on Windows 64-bit Server OS
* Standard data sources (Oracle, DB2, SQL Server, etc.) with no significant connector-related work
* If data sources hosted at customer, VPN access and sufficient privileges from EIQ Adapter machines
* Data sources accessible for testing and experimentation
* Standard data types - no unusual data types – images, etc.
* A real-time update mechanism from data sources available or readily obtainable
* EIQ Adapter native security - no OS/Kerberos security

## Deliverables:

What are the expected deliverables?
A demonstration of the POC to managemnet.

What is the anticipated timeline for final implementation and delivery of the solution?
First week in February 2022.

## Challenges:

Obtaining data sources that:

1. Allow for a patient to be matched across more than one data source.
2. Conforms to HIPAA.
3. Can be used for demos outside of Customer Name and affiliated organizations.

## Project Schedule:

| Task Name |
| --- |
| **Customer Name POC** |
|  **Requirements analysis and definition (10% of the project time)** |
|  Demo scenarios |
|  Architecture requirements |
|  Deployment/runtime environment |
|  HIPAA/de-identification/re-identification/transform requirements |
|  Reporting Tool/Data Consumer Client requirements |
|  Security/access control requirements\* |
|  Real-time update requirements\* |
|  **Solution design (10% of the project time)** |
|  **Data Sources and data analysis - for each data source** |
|  Data source type - relational, structured, unstructured, etc. |
|  Data access/connector |
|  Data schema and application/purpose |
|  Data types |
|  **EIQ Adapter Design - for each data source (15% total of the project time)** |
|  EIQ Indexes (tables, columns, masking, VK/NVK, AQP, etc.) |
|  Data model mapping |
|  Data transforms |
|  Real-time update mechanism/simulation\* |
|  Entity extraction data and entities\* |
|  **MDM Design (10% of the project time)** |
|  MDM entity, attributes, matching criteria design |
|  **Implementation (45% of the project time)** |
|  Setup development/test environment |
|  Data source access |
|  EIQ Adapters |
|  Entity extraction configuration\* |
|  EIQ Federation Server configuration |
|  MDM |
|  Index update (real-time/batch, etc.) \* |
|  Reporting Tool/data client integration |
|  **Testing, demo preparation and presentation (10% of the project time)** |

\* - typically, optional, based on requirements

Any questions on the above can be directed to Gavin Robertson, CTO, gavin.robertson@whamtech.com, +1 972-991-5700 x706 (o) or +1 817-995-4912 (m).

## Appendix 1: Solution Features Checklist

| Capability Required | Response | Comment |
| --- | --- | --- |
| Total no. of data sources | 10 (3 for POC) | Maybe more |
| No. of non-SQL data sources | 2 | Hadoop and document store |
| Data discovery with automatic id of personal data | Yes |  |
| Metadata | Yes |  |
| Standard info model, data model and data dictionary | None available | Interested in adopting industry standards |
| Data cleansing | Yes, for personal data | Know there are issues |
| Data transformation (type, sub value lookup split, merge, etc.) | Yes | Depending on standards adopted |
| Data standardization | Yes | Depending on standards adopted |
| Data masking, tokenization and/or encryption | Yes, for sure personal data | Need to coordinate with GDPR group |
| Unstructured data to be processed | Not sure what is available |  |
| Capture entity and attribute links/relationships within and across data sources | Yes, for knowledge graphs | Need to coordinate with analytics group |
| Data catalog | Yes | Would need to learn more and coordinated with analytics group |
| API catalog | No | Maybe in the future |
| Data access control | Yes, AD-based | Need to coordinate with security group |
| Data governance | Yes | Would like to see part of data catalog |
| Master data management (MDM) | Yes, see what we can use of existing customer MDM system |  |
| Data integration | Yes | Patient, ICD-10 code, treatment, doctor, medication and more |
| Data interoperability | No | Maybe in the future |
| Data update propagation and write back, e.g., email, phone numbers and addresses | No | Maybe in the future |
| If indexes, update frequency | Near real-time and overnight | Need to review on data source basis |
| Distributed, and if so, which locations/platforms | Yes, Cloud, data center, multiple network providers and PCs | Need to review |
| Centralized, and if so, which location/platform, e.g., virtual cloud | No | Assuming can centrally access |
| Data monitoring, event processing and/or BPM workflows/process automation | No | Maybe in the future |
| Reporting, BI and/or analytics | Tableau, Power BI, self-developed Web pages and interactive link visualization and analysis |  |
| CCPA/CCPR and/or GDPR | CCPA for sure | CCPR in the future? |

## Appendix 2: Solution Load Factors for Individual EIQ Adapters

1. Size of data in source that needs indexing to pre-process data, execute queries, run MDM and support virtual graph database.
2. Number of fields/columns that need indexing.
3. Amount of structured data versus unstructured (or semi-structured) data.
4. Number of concurrent users.
5. Query rate.
6. Query complexity.
7. Desired response time.
8. Speed of network connection between adapter and data source.
9. Size of adapter RAM cache available.
10. Data cardinality.
11. Data types.
12. Index update frequency.

## Appendix 3: SmartData Fabric® Terminology/Concepts

**EIQ Indexed adapter** – uses indexes, aka EIQ Index, needed for almost all data sources to provide a virtual data management layer to overcome issues associated with data, data source, including query processing, performance and load, access control and compute deployment – in most case, results data is fetched from the source

**EIQ Index** – specific to a data source and contains indexes for selected data source columns and source schema metadata, and can include indexed views, derived tables/columns – usually, cleansing, transformation and standardization algorithms are applied to indexed data and results data – indexes provide many benefits

**Standard Data Model** – usually an imported industry-based model, e.g., ACORD, HL7 and LIMRA’s LDEx

Semantic mapping – mapping of data source-specific columns to Standard Data Model-based schema columns

**EIQ Federation Server** – federates multiple adapter VDSs or other federation servers with associated VDSs that use the same standard data model - a single high-level federation server provides a single virtual database interface to multiple data sources/adapter and Logical Business Objects in the federation available for query as if they are one table - automatically executes union of adapter and other federation server results

**EIQ Conventional Adapter** – in cases a data source can execute and sustain external queries on clean data (also, see EIQ Hybrid Adapter) or there is no option to index data, converts a query from the Standard Data Model to the native data source schema and submits queries directly to the data source – results data can be cleansed, transformed and standardized to the Standard Data Model and combined with the results from other adapters

**EIQ Hybrid Adapter** - for data sources capable of executing and sustaining external queries AND some/most of the queryable data is clean and usable, a combination of an indexed adapter for data that needs indexing and an EIQ Conventional Adapter for the data that does not – from a federation point-of-view, the single data source is treated as two separate data sources using an indexed adapter for some data and an EIQ Conventional Adapter for other data, and the results combined through a joined view

**Logical Business Views Objects** – similar to non-materialized, aka virtual, views, but more flexible and defined as a combination of table JOINs and their mapped columns

**Schema interface** – two connection types: Native schema or Business Data schema – Native schema exposes the source schema tables/columns and Standard Data Model interface provides access only to mapped columns and logical business objects

**Virtual Data Source (VDS)** – similar to a database and is usually a combination of the actual data source and a set of indexes, whereby client applications connect to a VDS as if it is a database through an adapter or federation server – hides source or federation details behind a single database interface

## Appendix 4: Internal SmartData Fabric® Glossary

**ROWID** – a unique value column(s) designated for fetching matching results records from a data source table, such as a primary key in a source table or can be multiple columns that combined uniquely identify a record in a table, e.g., in an association table that does not have a primary key

**VK (Virtual Key)** – EIQ Index has an index for this column, but no data – data needs to be fetched from the source through a connection using ROWIDs

**NVK (Non-Virtual Key)** – EIQ Index has an index AND data – no need to fetch data from the source

**VC (Virtual Column)** – EIQ Index has only the metadata for this column – no index or data – cannot be used for filtering, but can be used for fetching data from columns as part of results data, and is often used for columns containing large objects such as files and image objects – this is different from normal columns that are read as part of results data

Note: WhamTech encourages using any of the above only for data that needs indexing or metadata (in the case of VC). Most data in transaction systems is non-human-generated and does not need indexes for data issues – there may be data source, access control, compute deployment or other issues where indexes are needed, but if the data source can execute and sustain external queries on a lot of the data that does not need indexing, then WhamTech may recommend the use of an EIQ Hybrid Adapter (see previous slide on SmartData Fabric® terminology/concepts). In many cases, an EIQ SuperAdapter is all that is needed, as queries may be limited to indexed data only.