

Software Engineering Methodology for the Data Warehouse Project

Arman Kanooni

101001010100111101000010010111010010 110101010101110100001000101001001
0010000101001010010010100001011010010101000011110100101010011110100001001011010010
110101010101110100001000010100101001010000101101001010100001111010010101

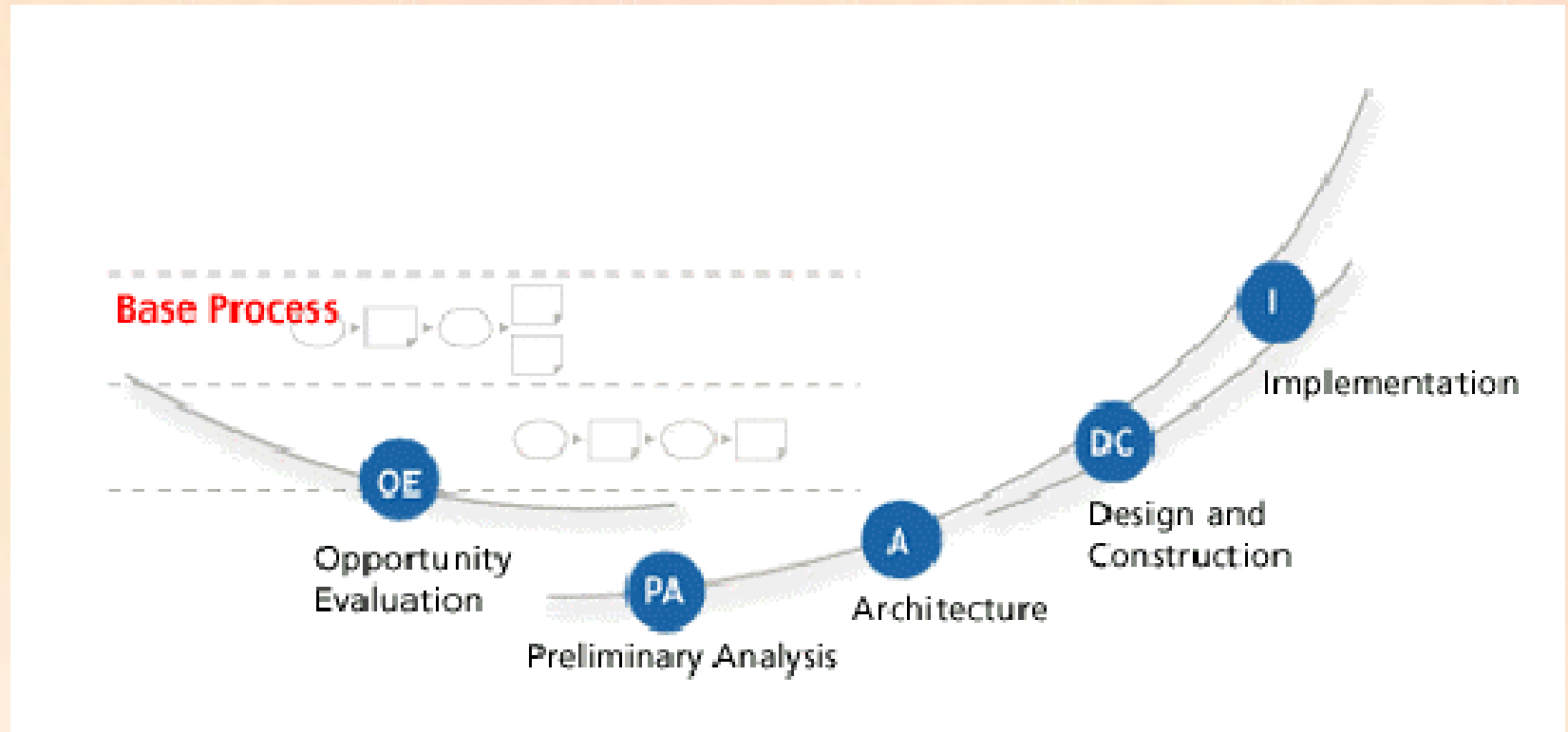
Contents

- Difference between Data Warehouse and Stand Alone Systems
- ProductivityCenter Version 4.0
- Phase Question and Results
- Release Strategy
- Iterative Approach
- What is a deliverable?
- Concurrent Engineering
- Roadmap

Difference between Data Warehouse and Stand Alone Systems

- *Data Warehouse are mostly driven by Business Opportunity rather than Business Need*
- *Data Warehouses implement a cross-organizational decision-support strategy rather than departmental decision-support silos*
- *Business Intelligence decision-support requirements are mostly strategic information requirements rather than operational functional requirements*
- *Analysis of Business Intelligence projects emphasis business analysis rather than system analysis*
- *Ongoing Data Warehouse release evaluation promote iterative development and software release concept rather than big-bang development*

IT Life Cycle



Phase Question and Results

- **Opportunity Evaluation:** Does the project make business sense?
 - Business value is defined
- **Preliminary Analysis:** Is the proposed solution financially, organizationally and technically feasible?
 - Feasibility is established
- **Architecture:** Have we defined the overall solution well enough to minimize rework in building the upcoming release and to meet the customer's needs for maintainability?
 - Overall design and implementation plan is confirmed

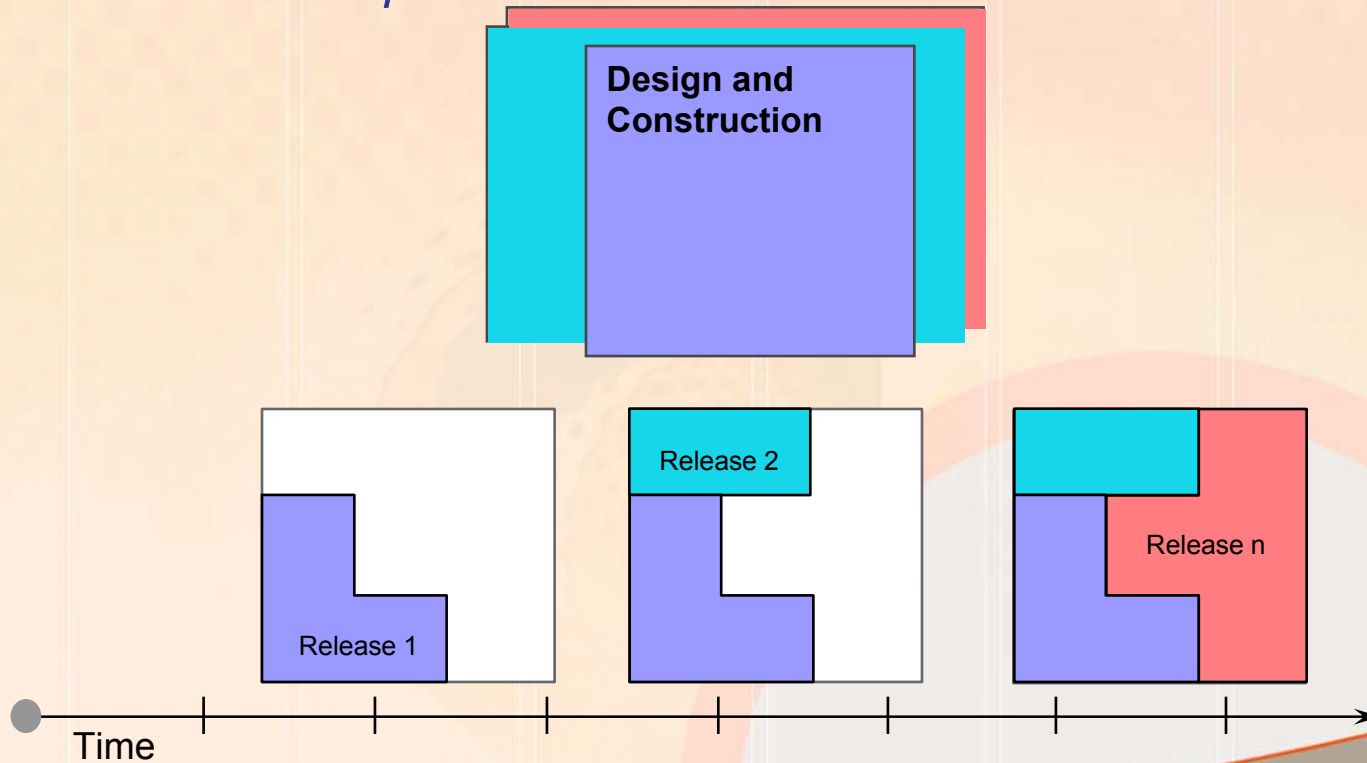
Phase Question and Results

- **Design & Construction:** *How well do the product and implementation plan meet the specifications and business needs?*
 - *Detailed plans and product are created and tested*
- **Implementation:** *How can the solution be successfully installed and integrated into the business?*
 - *Product is accepted and deployed*



Release Strategy: Incremental Approach

- *System components can be built and delivered progressively*
 - *Progressive construction of releases*
 - *Progressive implementation, each release building on the prior releases*



Releases

A release is a part of the system that delivers value to the business to be installed as a separate and functioning unit.



Release
1

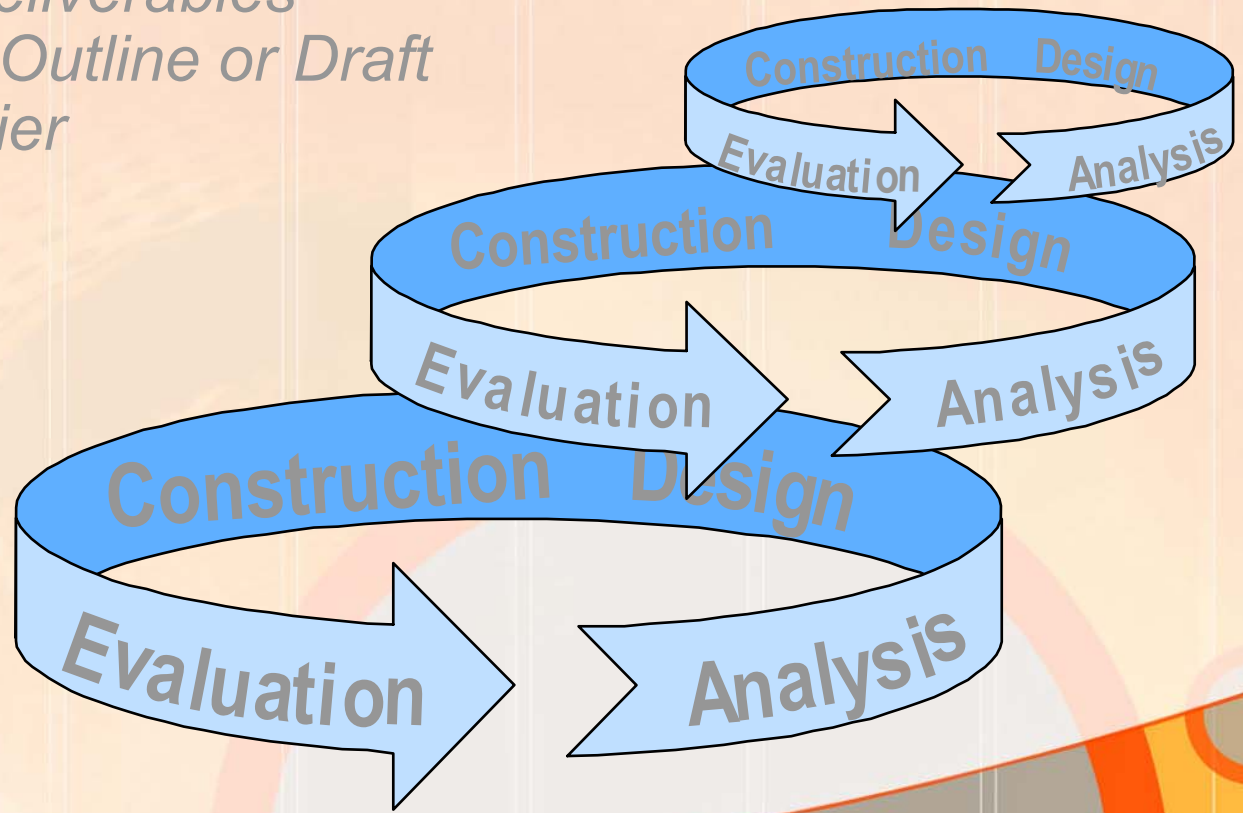


Release 2

Iterative Approach

The Iterative approach is also used

- In various Prototyping Techniques described in Macroscopic*
- In revisiting deliverables worked on to Outline or Draft stages in earlier Activities or Phases*



What is a Deliverable?

A Deliverable:

- *Is a tangible product resulting from project activity*
- *Always has a specific purpose, often a clearly identified audience*
- *Is used to communicate, capture decisions, & manage a project*
- *Is gradually completed through concurrent engineering*

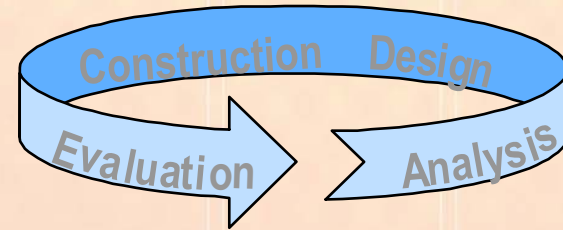
It could be

- *a document*
- *a plan*
- *a spreadsheet*
- *specifications*
- *software*

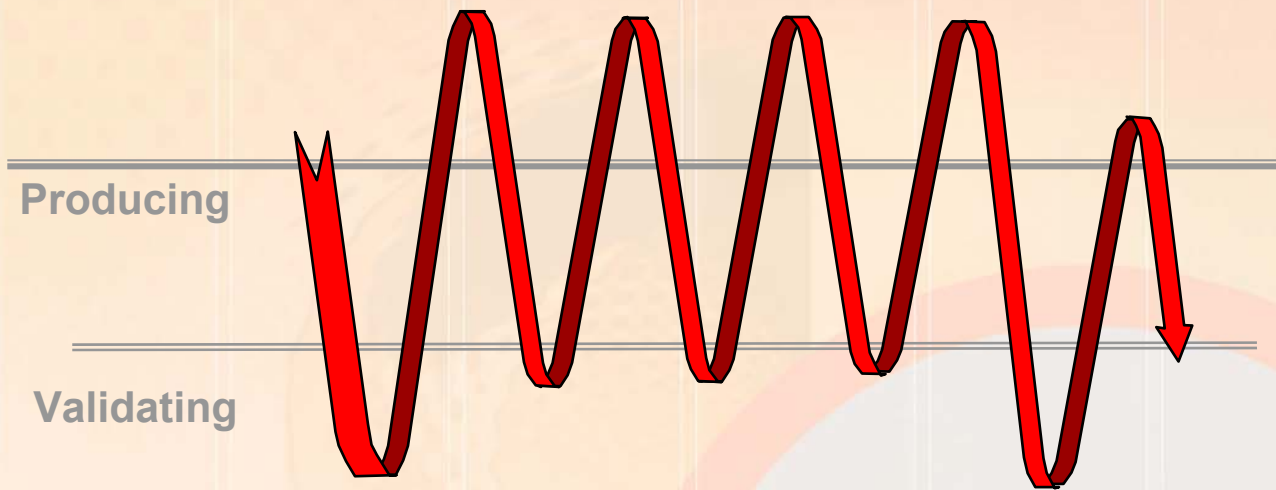


Concurrent Engineering: Deliverables

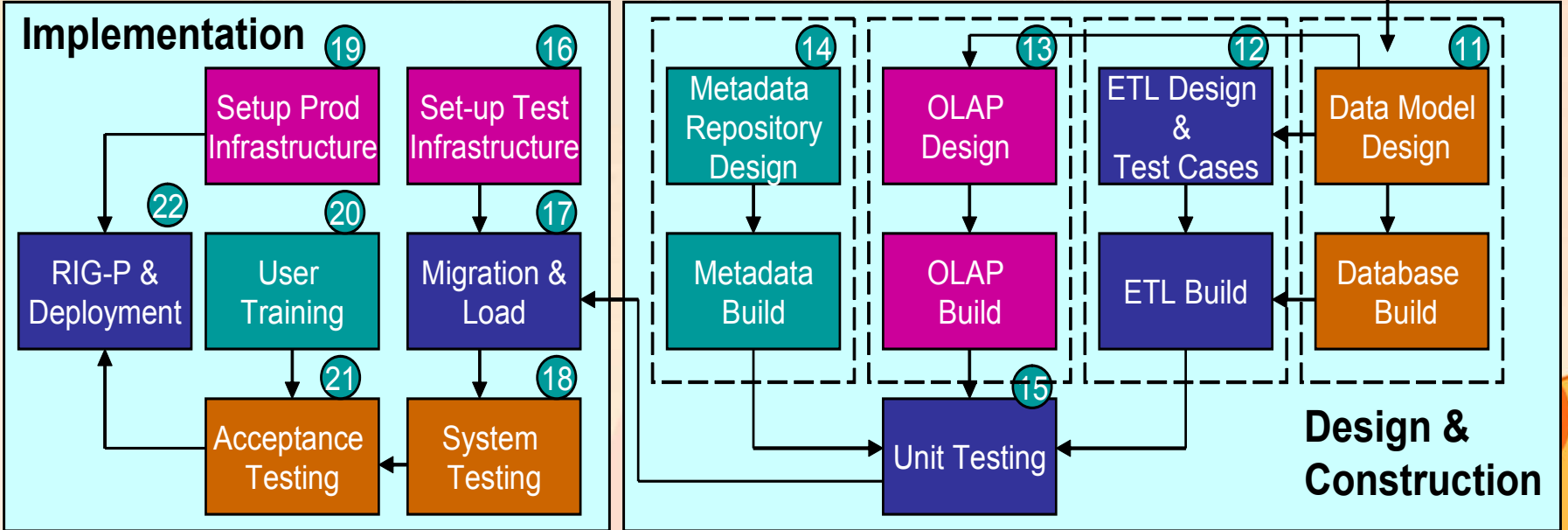
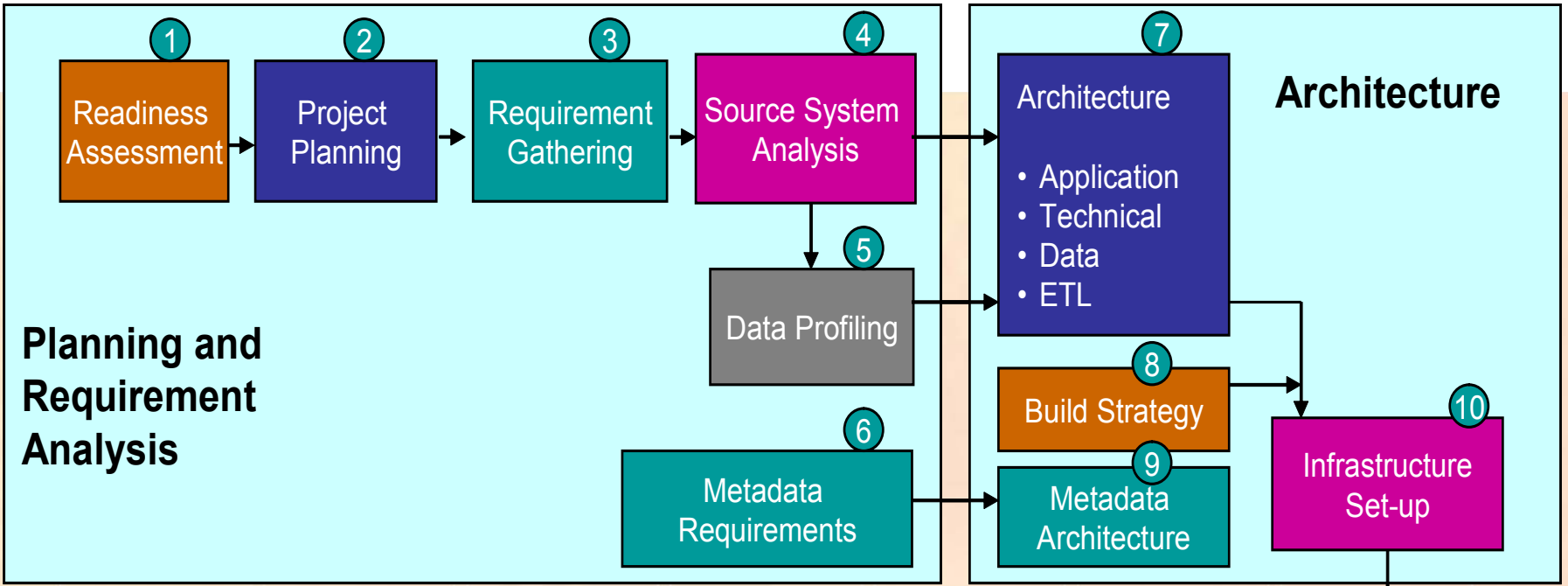
- *Get feedback*
- *Sort and prioritize*
- *Discuss*
- *Achieve consensus*



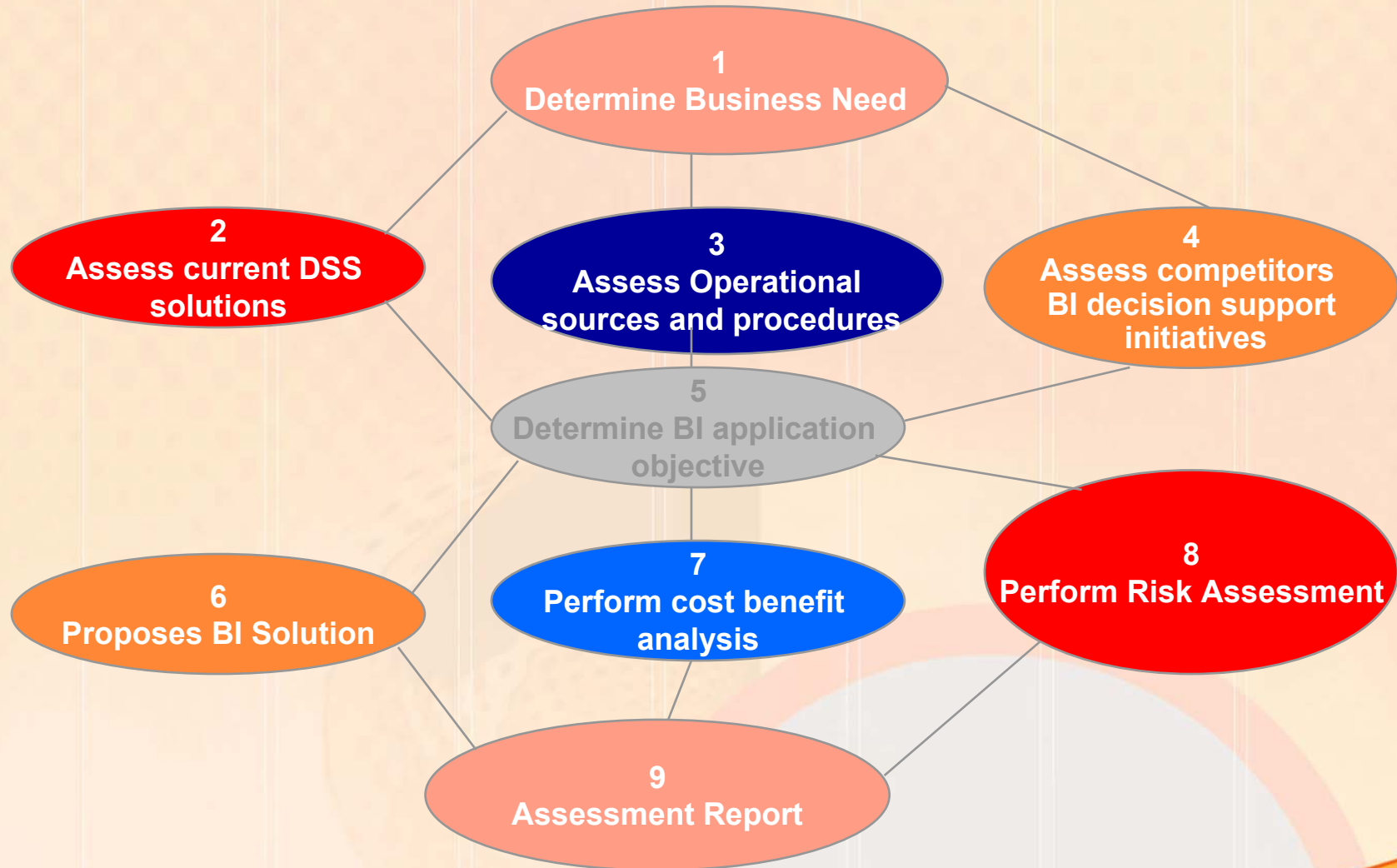
Validation Cycle



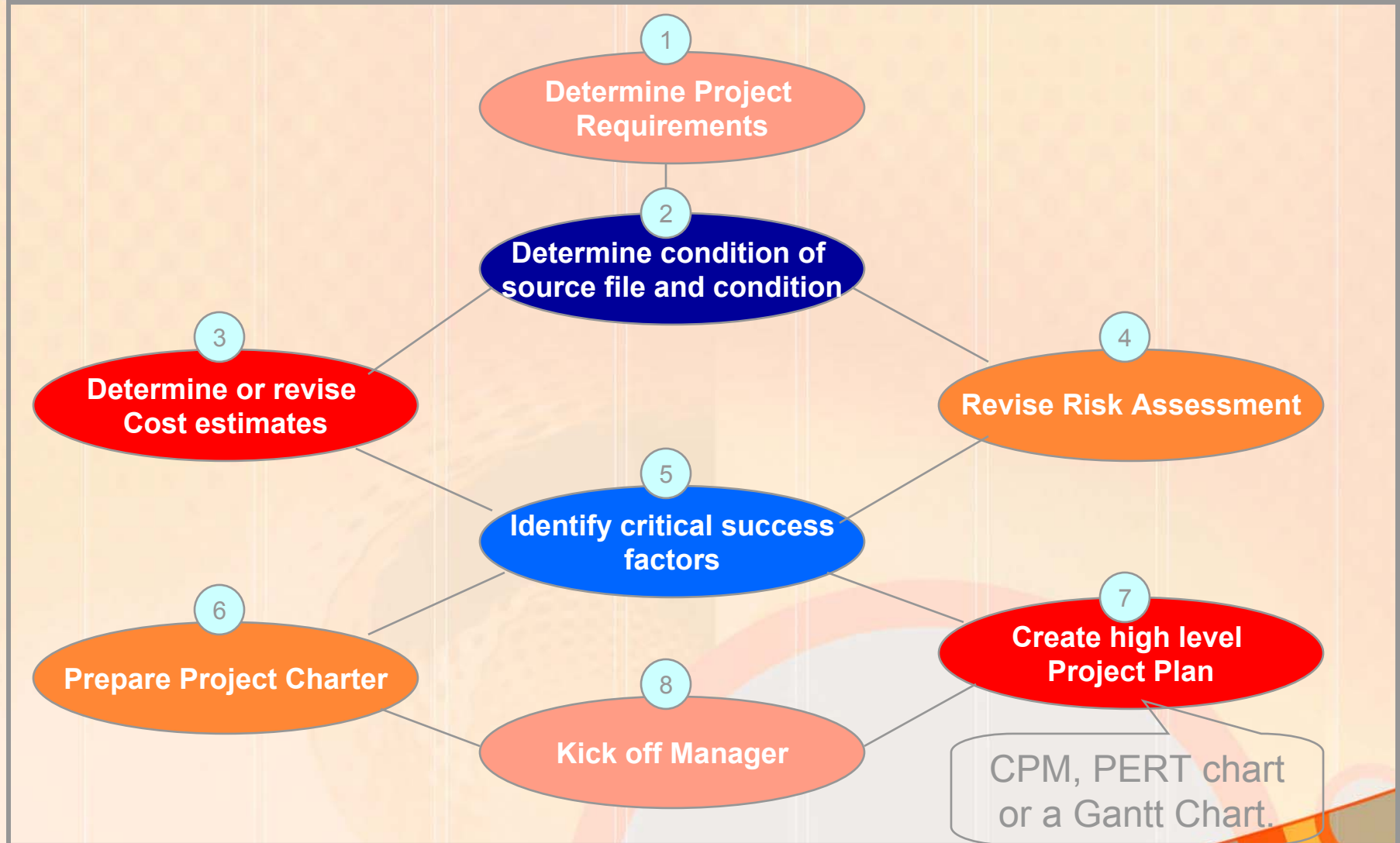
Continuous Validation



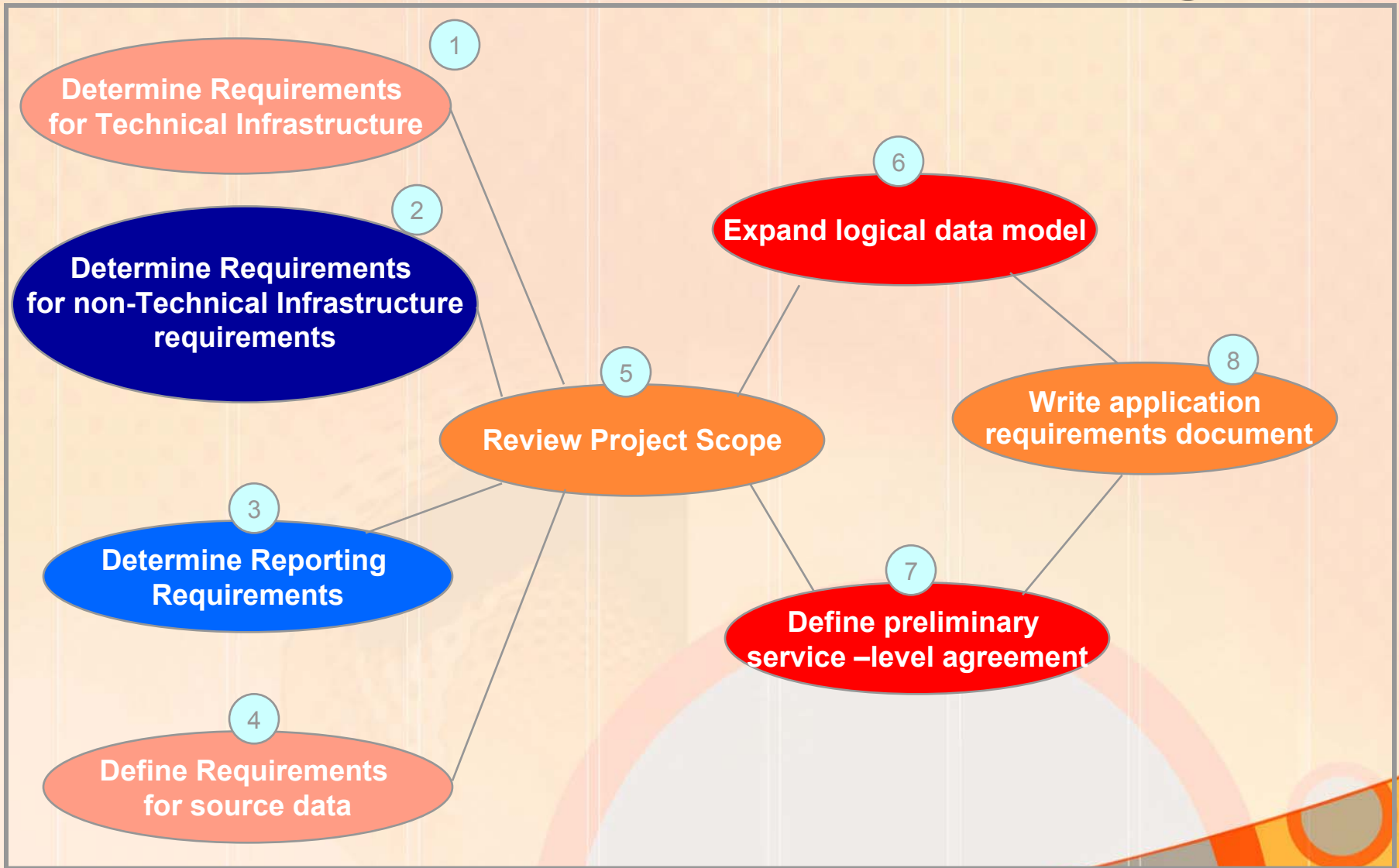
1. Readiness Assessment



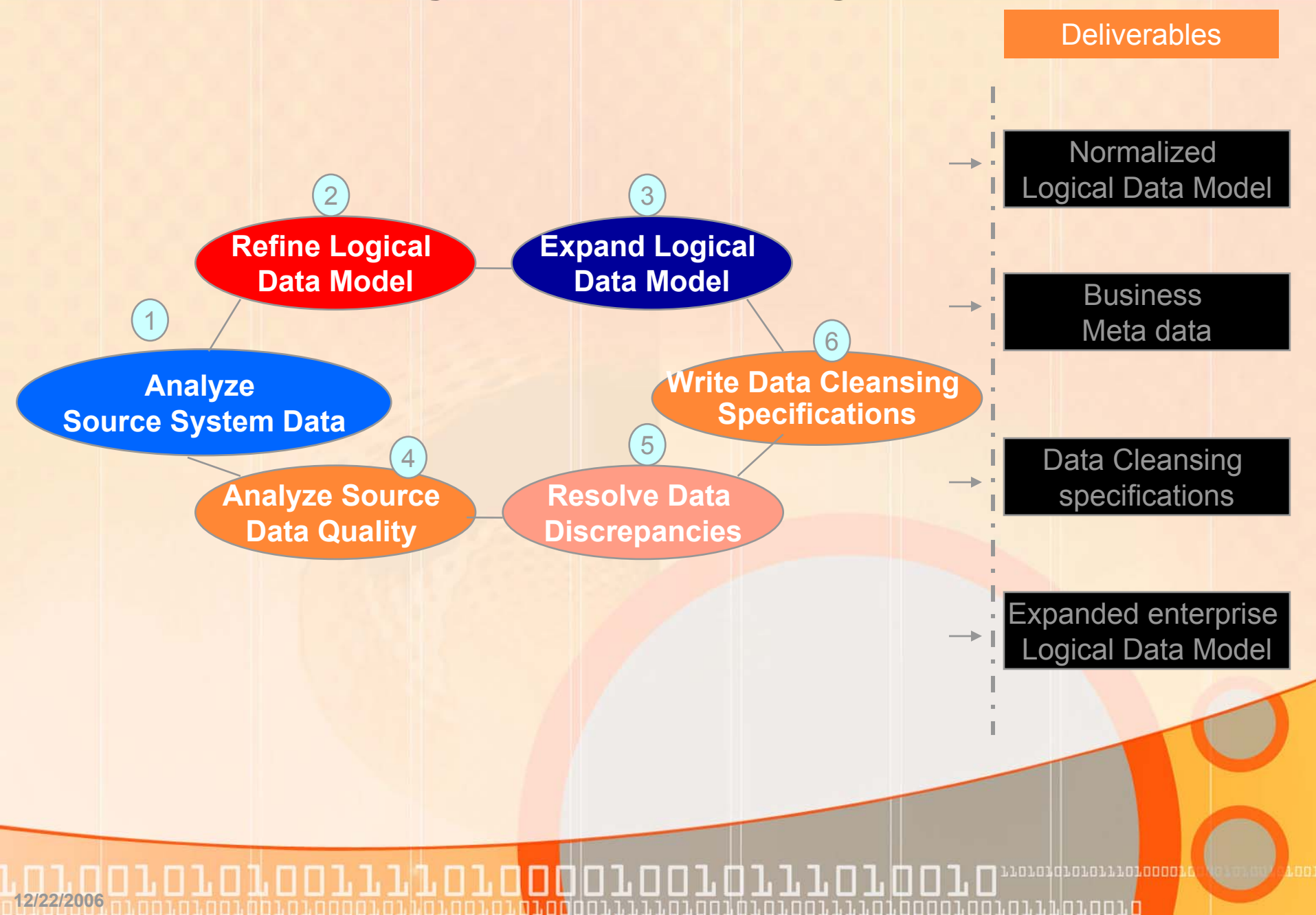
2. Project Planning



3. Requirements Gathering



4. Source System Analysis

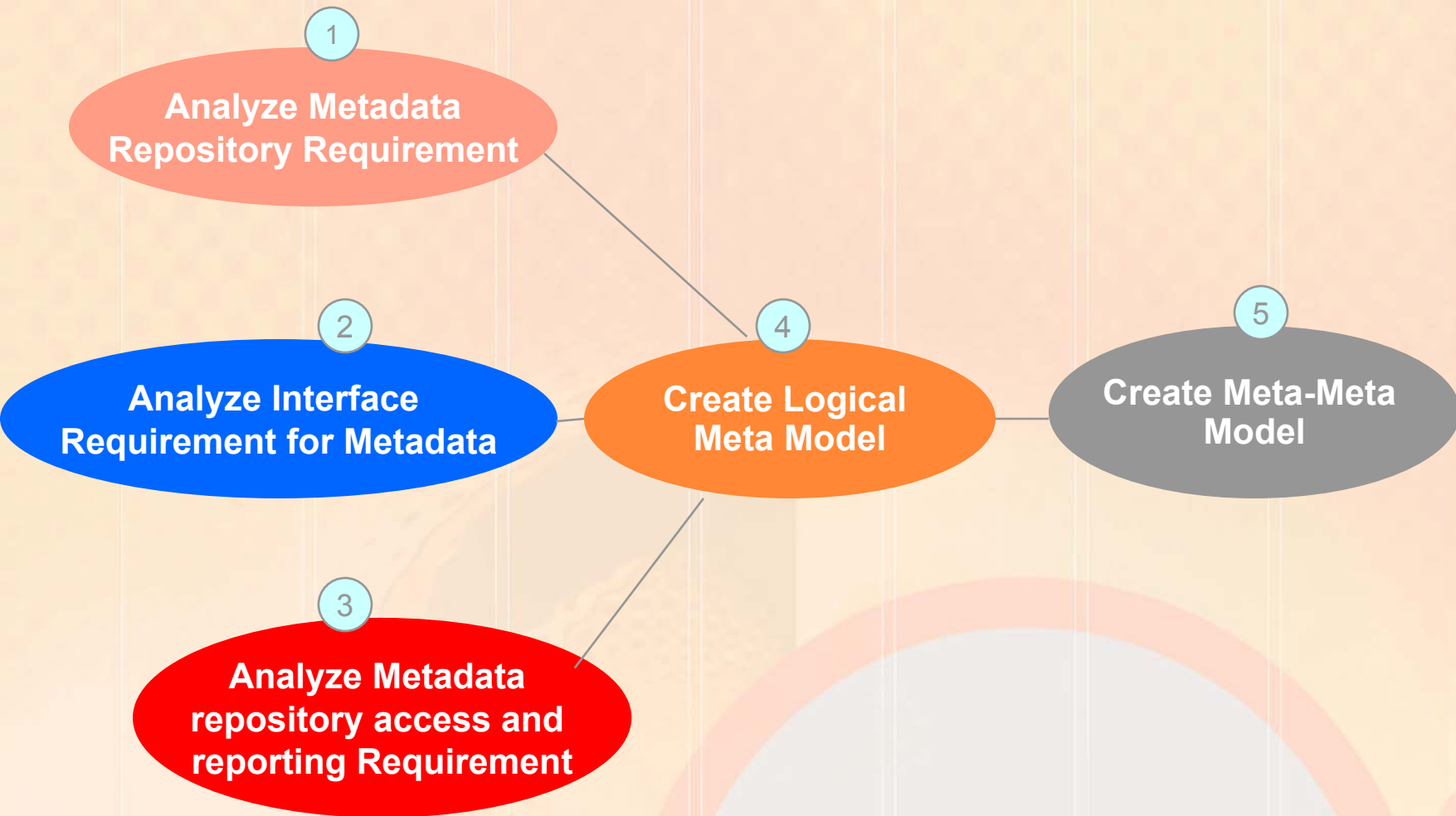


5. Data Profiling

- *The information analysis also involves profiling the data to analyzing source system data quality.*
- *Deliverable is Data Quality Findings Report*



6. Metadata Requirements



7. Architecture

- *Application Architecture: The purpose of this phase is to define a structure for the entire application. This architecture shows the various subsystems / layers / components that together make the application and how these are linked. This architecture also includes End user Applications.*



7. Architecture (Continued)

- *Technical Architecture: The purpose of this phase is to design the technical platform to base the project on. Typically, this service identifies details of hardware and software specifics to be used in the project.*



7. Architecture (continued)

- *Data Architecture: This data architecture provides high level data model for Data Warehouse and identify the key dimensions and fact based on the requirements.*



7. Architecture (continued)

- *ETL Architecture: The ETL Architecture Phase involves the identification of a high level approach to the process of populating the data warehouse and/or data marts. This process will encompass three steps namely Extraction, Transformation and Loading.*



8. Build Strategy

- *The Build strategy is used to refine the increment definition, plan for their execution, develop the testing plan.*



9. Metadata Architecture

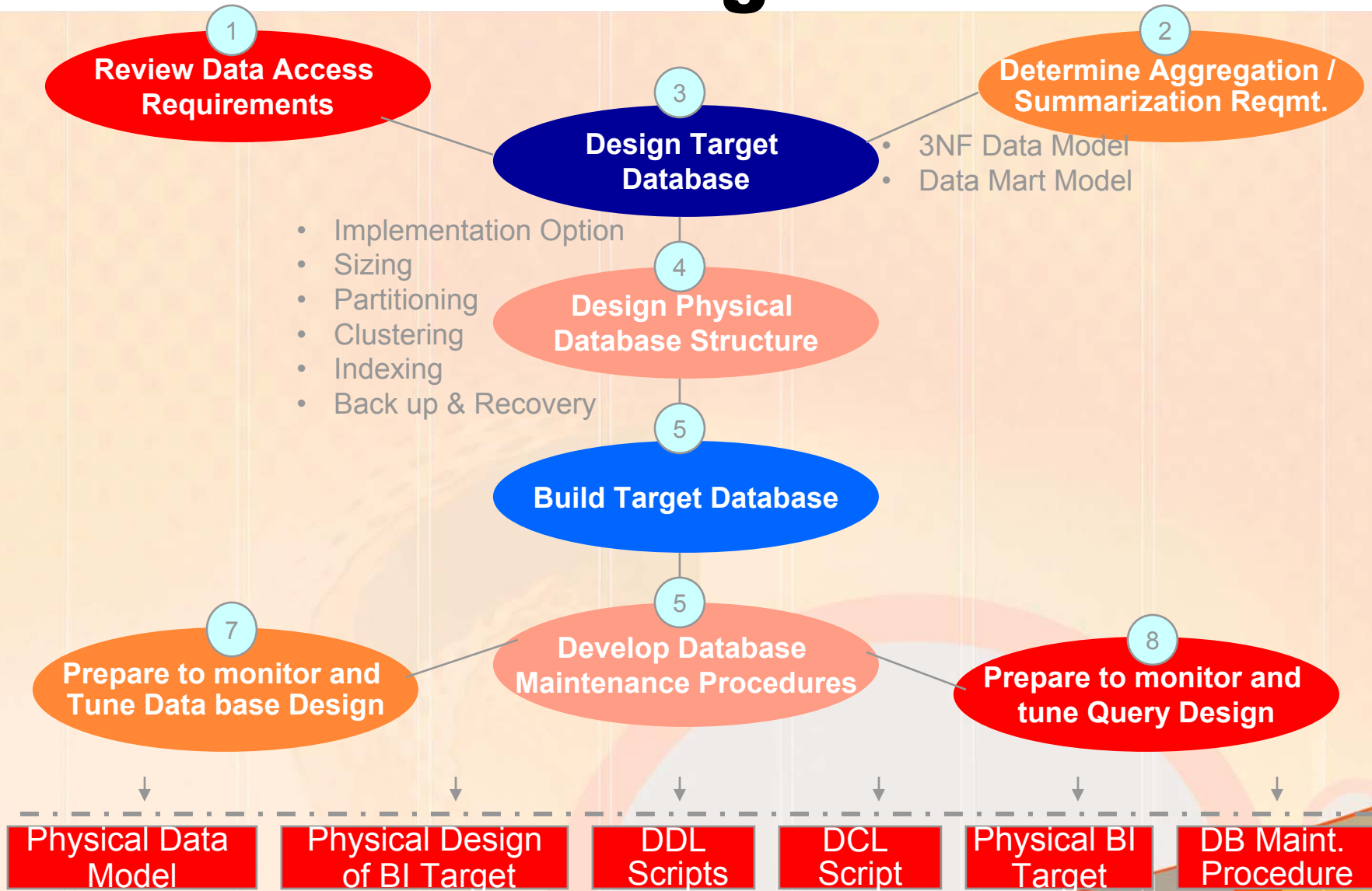
- *The metadata architecture details designing the architecture, standard to be used, nature of interfaces to be used / build mechanism for data collection and porting.*



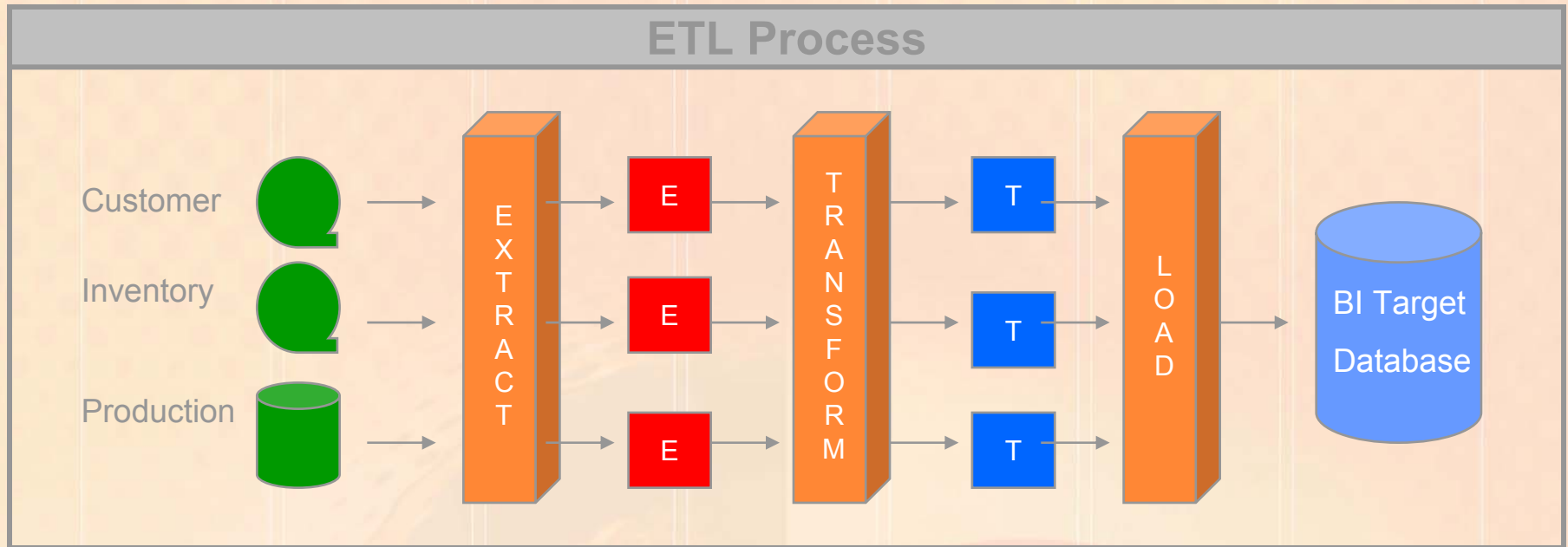
10. Development Infrastructure Setup

- *Set-up the development environment in line with the approved technical and application architecture. Ensure the required hardware, software and tool licenses are procured and installed.*
- *Data requirement for development environment also need to be identified as the project progresses. Although, volume of data is not a consideration in development environment, however it would be beneficial if the test data sample covers adequate business scenario.*
- *For outsourcing, the data may either need to be sanitized if it falls under ITAR / EAR guidelines.*

11. Database Design and Build



12. ETL Design and Build



- Heterogeneous Data Source System
- Inconsistent Primary Keys
- Inconsistent Data Values
- Different Formats
- Inaccurate Data Values
- Embedded Process Logic
- Historical Load
- Incremental Load

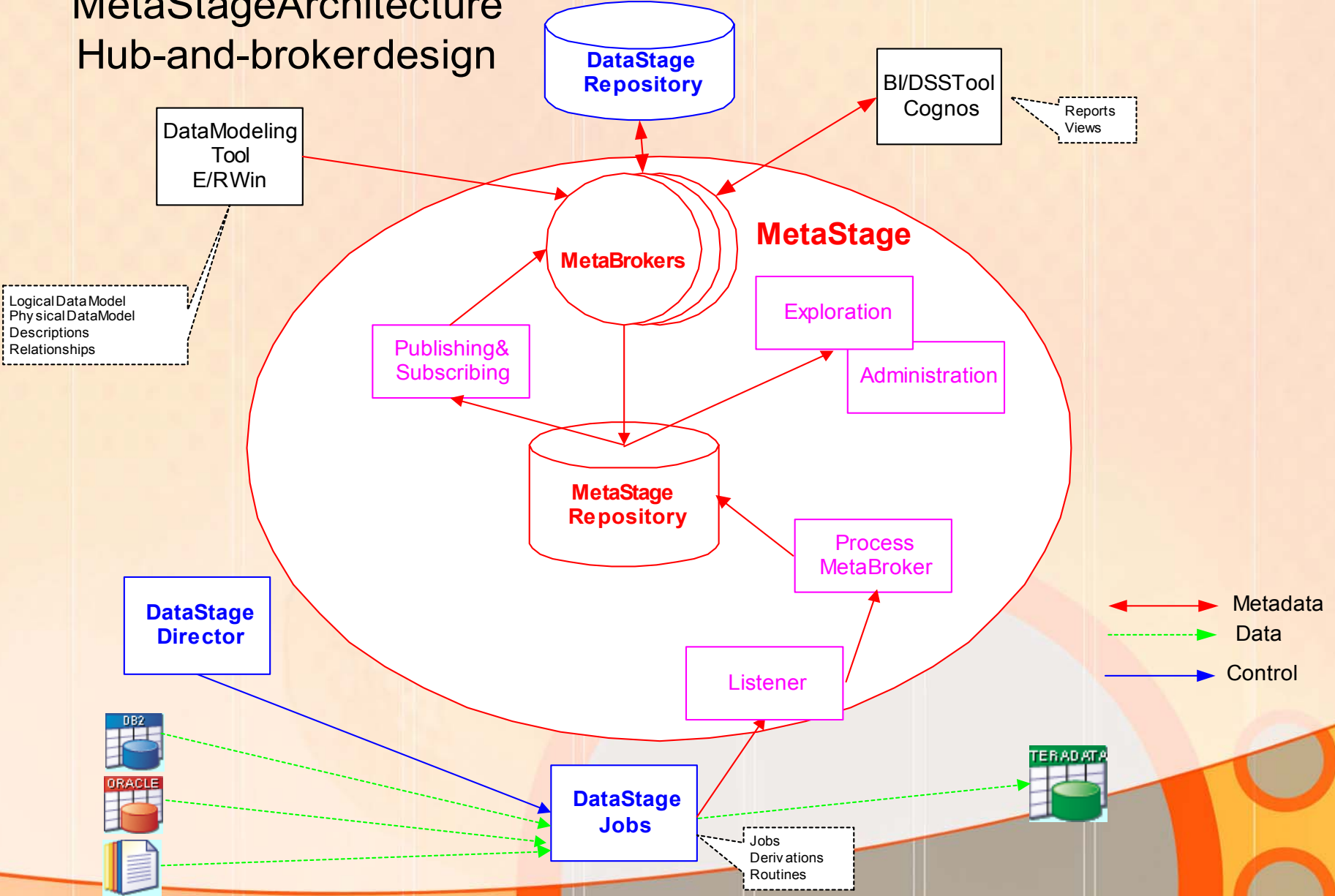
13. OLAP Design and Build

- *The OLAP build involves detailed design of the end user application (Cognos catalogs/cube) and report development.*
- *Design Document for Cubes*
- *Design Document for Reports*
- *Unit Test Cases*
- *Cubes and Reports*



14. Metadata Design and Build

MetaStageArchitecture
Hub-and-brokerdesign



15. Unit Testing

- *Each of the above code (ETL, OLAP, Metadata interfaces) will undergo unit testing.*



16. Setup Test Infrastructure

- *Design and learning during the development, set-up the test environment in line with the approved architecture guidelines and principles,*
- *Ideally, test environment should be a replica of production environment in terms of data volume and source system interfaces. Test data requirements should be identified well in advance during the design as this could be a time consuming activity. The application should be tested on acceptance before moving to the production.*



17. Migration and Load

- *Test Application is migrated and data populated in respective source system.*



18. System Testing

- *The complete system is tested during this phase to check the quality attributes as per the test plan.*
- *The phase also involves preparing Release Installation Guide (RIG) for moving to testing environment.*



19. Production Infrastructure

- *Provide data sizing details to the support vendors for recommending necessary production infrastructure upgrades in terms of disk, memory, processing requirement.*
- *Vendor may require substantial lead time to deliver components. Provide vendors with adequate lead time to avoid any project delay. Carry out the setup/upgrade as appropriate as a parallel activity during the testing phase.*



20. Training

- *The task involves engaging test users and training them to carry out the acceptance testing.*
- *Test users also prepare test cases for acceptance testing. Test Manager usually engages the business users and trains them on writing test cases and system functionality.*



21. User Acceptance Testing

- *The user acceptance phase involves the functional acceptance testing of the application.*



22. Deployment

- *This involves Installation plan & prepare Release Installation Guide (RIG-P) for installation.*
- *Prepare help desk scripts to support user queries on deployment.*

