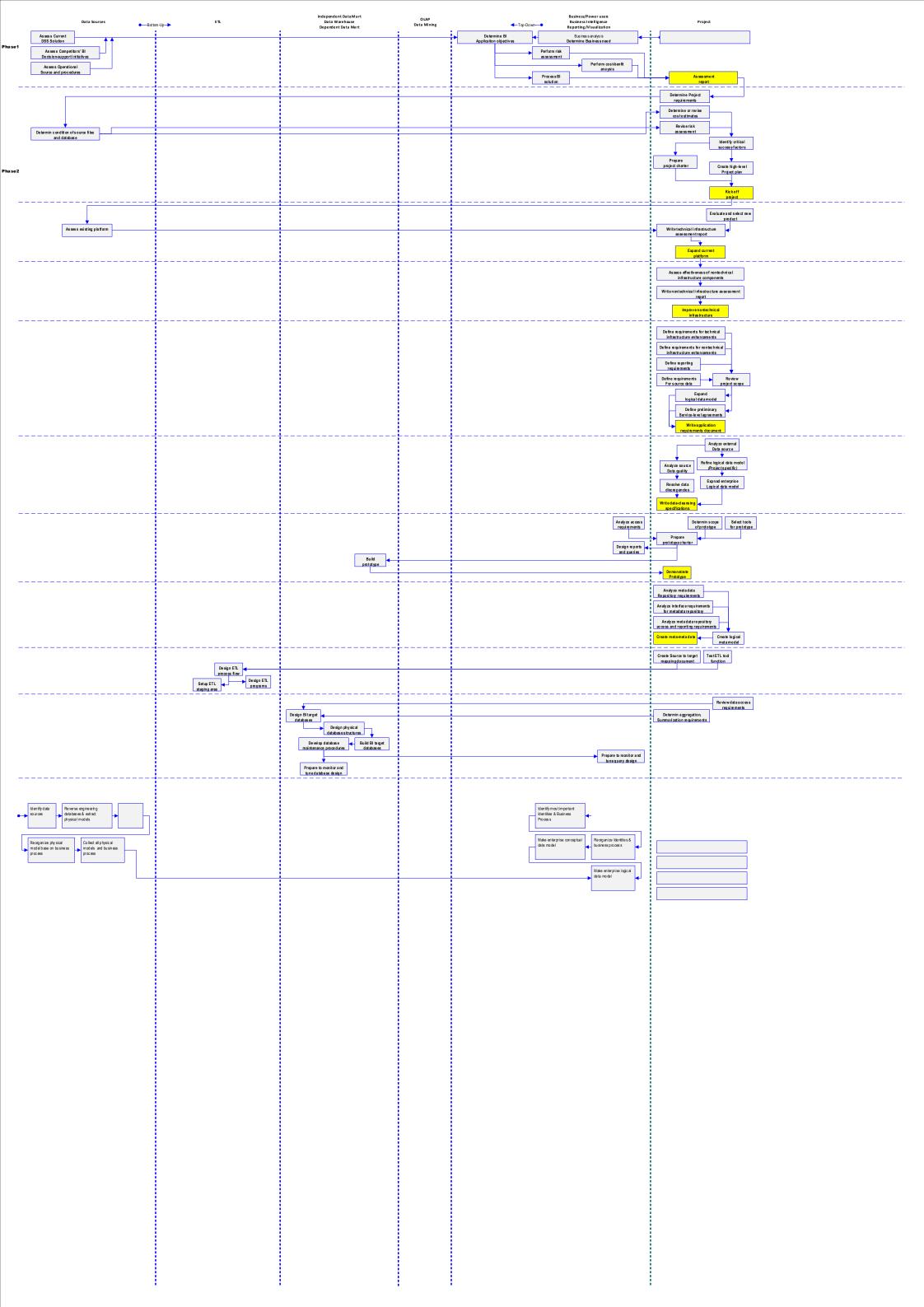
Data velocity through the Data Warehouse



Corporate/Government Information Factory, CIF/GIF

Prepared by: Khashayar jamsahar www.jamsahar.com last update Apr, 2011

Government Level of Information

<u>Data Globalization</u>(Make a Universal View) Sanitization

Business Performance Management

Strategic Planning, Balanced Score cards & KPI & Dashboards

Power Users Decision Making (Visualization)

Monitoring & Control + Feedback

Job schedulers, Backup/Recovery system, Recovery and Restart process

Version control/migration system, Workflow monitor, Security

Parallelize/Pipeline, Compliance manager, Maintain/Update Metadata repository manager

Data Mining(Advanced & Predictive Analysis, Alert & Proactive Notifications)

OLAP Analysis(Pivoting, Drill Down/Up, Slice & Dice)

Ad hoc/Static/Dynamic Reports

Data Mart (ROLAP/MOLAP/HOLAP Cube)

Data Warehouse (Relational Data Model or Star/Snowflake/Galaxy/Vault)

Post Load

Optimizing, Active Constraints, Rebuild indexes, Recalculate Summarization/Aggregation

Load Subsystems

Initial/Incremental Refresh, Partitioning, Deploying/Synchronization, Log/Track changes, Generating Load metadata

Deliver

Compression/Condensation, Generating Surrogate key, Summarize and aggregate data, Manage Type I/II/III Dimensions Manage slowly change and late arriving Fact/Dimensions, Manage Time Dimension, Hierarchy Manager, Referential integrity Fact/Dimension table builder, Propagation, Generating Deliver metadata

<u>**Data Integration**</u>(Make a Unified/Global/Enterprise View)

Source/Target Mapping, Projection, Transposing Single Version of Truth, Make Subject-Oriented Fashion
Restructuring and mutating data, Enforcing business rules, Semantic Integration
Conforming/Resolve conflict between data, Conforming business labels/metrics
Derive/Calculate data, Manage Date/Timestamps, Revision/Check + Soft/Hard Reject system
United Value/Metrics/Currency, Data Quality Assurance/Firewall, Feedback to source operational systems
Generating integration activities metadata, Staging

Data Migration/Mediation (Make a Local/Department View)

Source/Target Mapping, Platform(Hardware/OS/RDBMS/Engine) Conversion, Set Validations/Conformations
Set Default/Missing/Null/Dummy Value, Replacing text string and Labels, Denormalization
Set standard metrics and common data format, Dedupticating data, Remove out of range data
Restructuring and mutating data, Set new data types, Data type Conversion, Conforming/Resolve conflict between data
Data Quality Assurance/Firewall, Feedback to source operational systems
Generating migration activities metadata, Staging

Transformation Subsystems

Source/Target Mapping, Refine/Reformatting(Select/Analyze/Edit/Clean/Match/Split/Join/Merge/Combine/Reset/Balance-Control/Filtering/Screening/Resequencing/Realigning/Deduplication/Reduction/Sort/Scrubbing/Classify/Categorize/Enrichment and Modeling data)

Adding Date/Timestamp, Conforming system, Error handling + Soft/Hard Reject system

Generating transformation activities metadata, Staging

Data Conversion (Make a Division View)

Source/Target Mapping, Decoding, Character Set Conversion
Data Quality Assurance/Firewall, Feedback to source operational systems
Generating data conversion activities metadata, Staging

Extraction (Population)

Source/Target Mapping, CDC/Delta-Files/Compare-snapshot/Timestamps/Partitioning/Triggers/Log-Audit/Replication/Sniffing/Materialized-View/Application-Code Scheduling, Generating extraction activities metadata, Staging

Who? How(action)? What? Where? When? Why?

Define Conceptual/Logical/Multi-Dimensional(Star/Snowflake/Galaxy/Vault)/Physical models around business subject areas

- 1- Enterprise reference modeling is done subject by subject.
- 2- Enterprise knowledge coordination is done artifact by artifact.
 - 3- Information factory development is done **topic by topic**.
 - 4- Data profiling and mapping are done source by source
- 5- Data correction is done attribute by attribute.
 6- Infrastructure management is done component by component.
- 7- Total information quality management is done process by process to improve and error-proof processes.

Business Requrements Analysis (Top-Down Procedures)

Discover Business Entities and Business Process and Business Rules, Describe Direct/Indirect requirements

Source System Analysis (Bottom-Up Procedures)

Data Profiling/Stores/Elements/Entities/Model/Architecture, Reverse engineering data models(Data Flow/State Diagram)

Set data quality assurance keys and SLA, Find business rules

Data Access Level Protocols(ODBC, JDBC, OLEDB)

Knowledge Management (OLTP/Archived/Internal/External/Unstructured Data) + Master Data Management