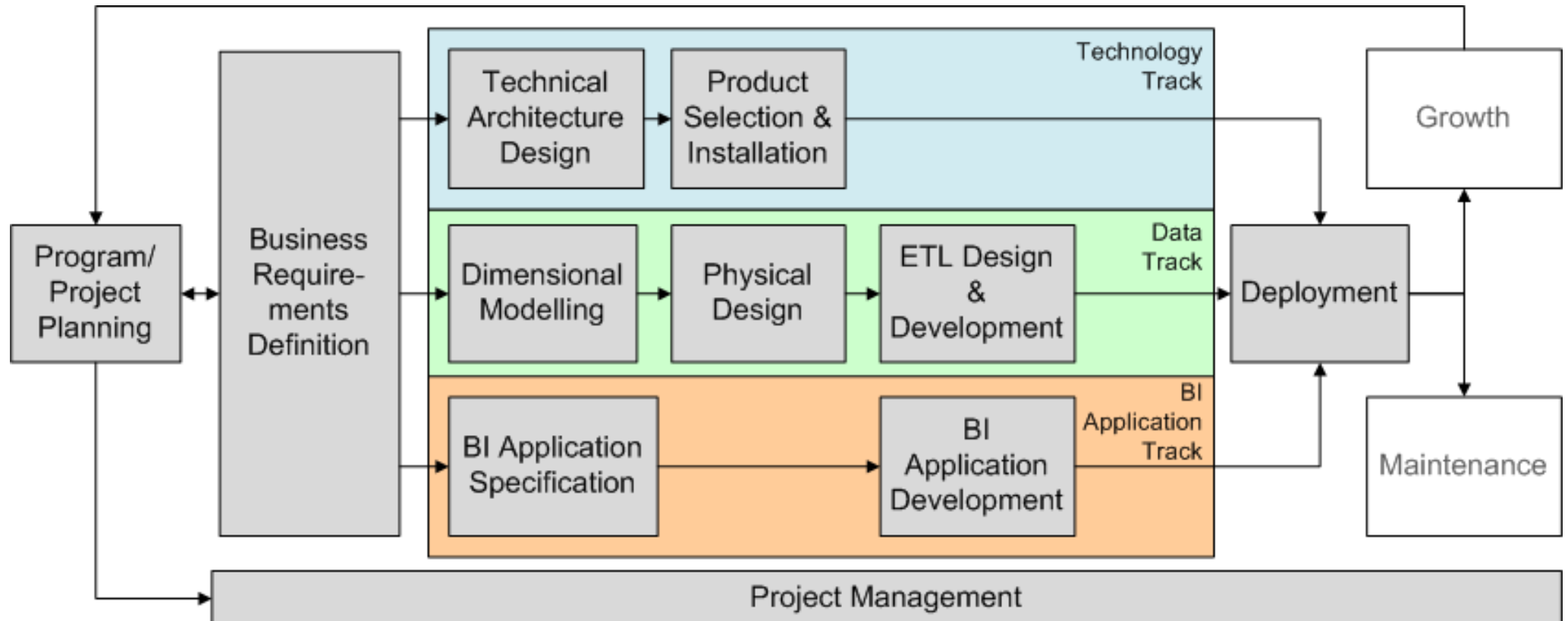


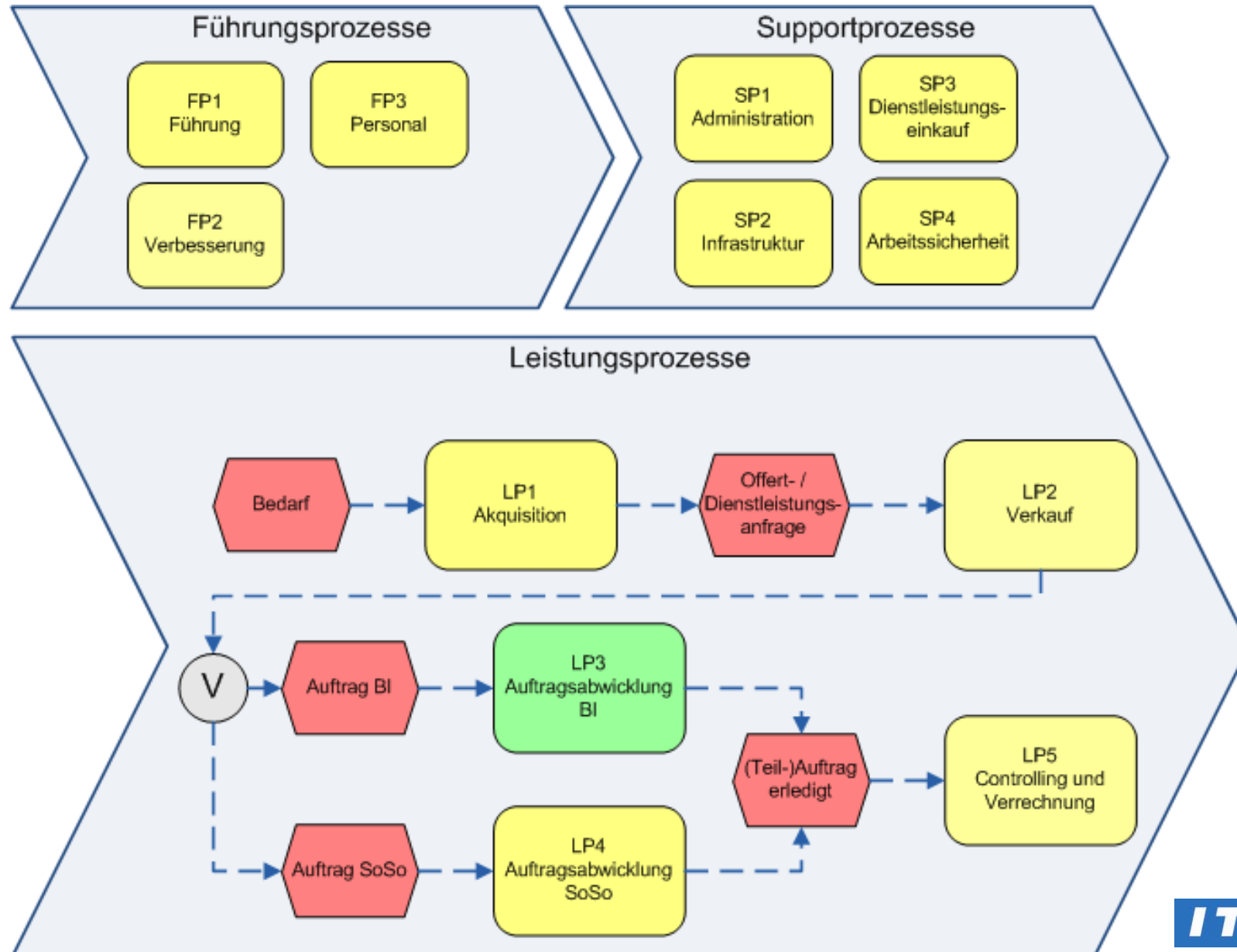
Kimball – The Data Warehouse Lifecycle Toolkit

Strukturiertes Projektvorgehen im
BI / DWH Umfeld

Urs Grunder, Senior Consultant

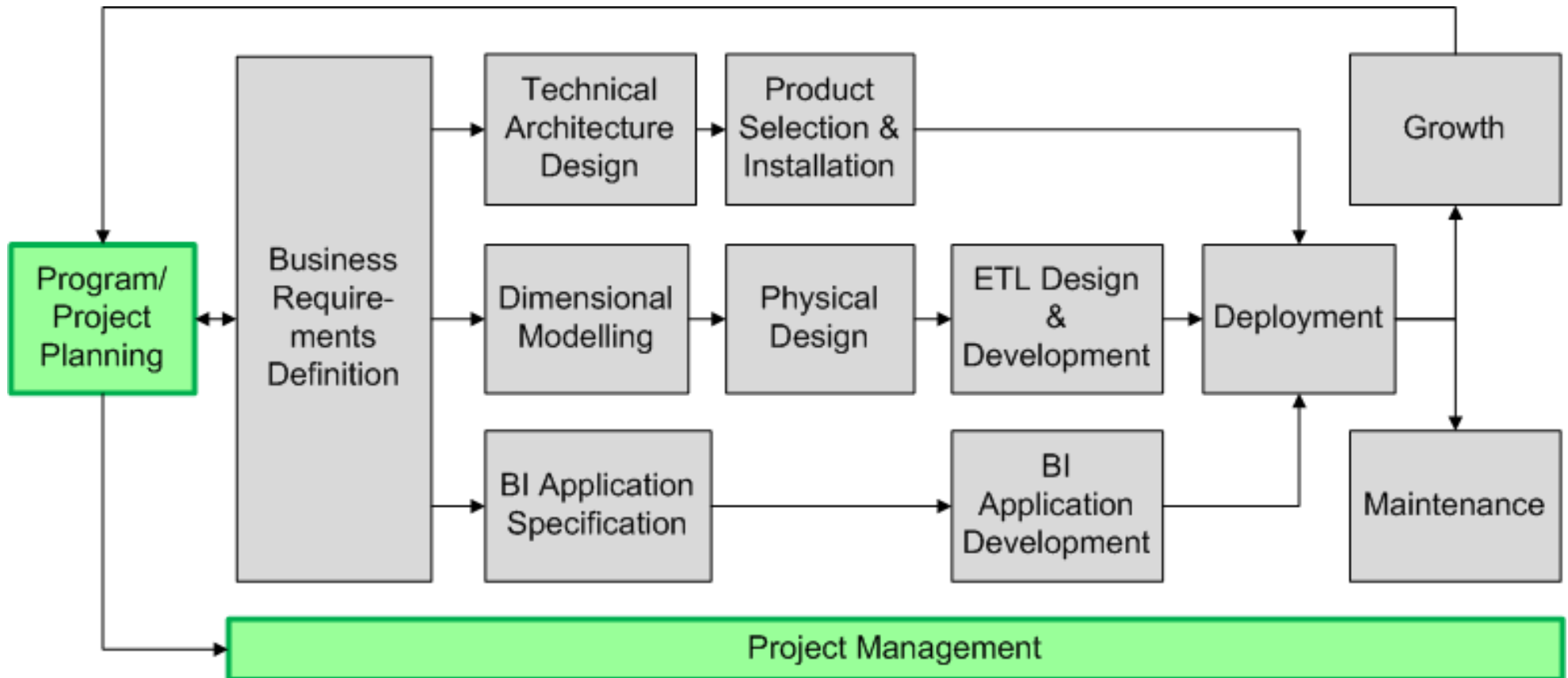
Kimball – Data Warehouse Lifecycle Toolkit -> Aus Best Practices entstandenes Vorgehensmodell für DWH/BI Projekte





- Praxisorientiert mit vielen Tips
- Umfänglich – nichts geht verloren
- Tailoring auf Projektgrösse
- Tailoring auf Teilprojekte - z.B. nur ETL Design und Implementation

Program / Project Planning & Management



- *Project* -> eine Iteration des Lifecycles mit definiertem Anfang und Ende (auch innerhalb des Projektes kann es auf verschiedenen Stufen zu Iterationen kommen!)
 - *Program* -> Beinhaltet mehrere *Projects* und stellt somit den Gesamtrahmen dar wie in einem Unternehmen die BI-Strategie umgesetzt wird
- > Im Folgenden Project Scope

- Projekt definieren
- Projektplanung und –controlling aufsetzen
- Issue-Tracking und Change-Control aufsetzen

Prozesshauptverantwortliche sind ...

- Project Managers
- Business Project Lead

Weitere Rollen arbeiten mit oder liefern Input

Vorgehen – Project Definition



- Readiness prüfen – BI Audit
- Projektumfang definieren
- Projekt finanziell „rechtfertigen“

- Projekt staffen (Rollen besetzen gemäss Taskliste)
- Projektplan ausarbeiten (Tailoring!)
- Projektkommunikationsplan ausarbeiten
- Kickoff planen und durchführen

- Laufendes Projektmanagement
- Status überwachen mit Statusberichten und -meetings
- Projektplan anpassen
- Change Control und Issue Tracking

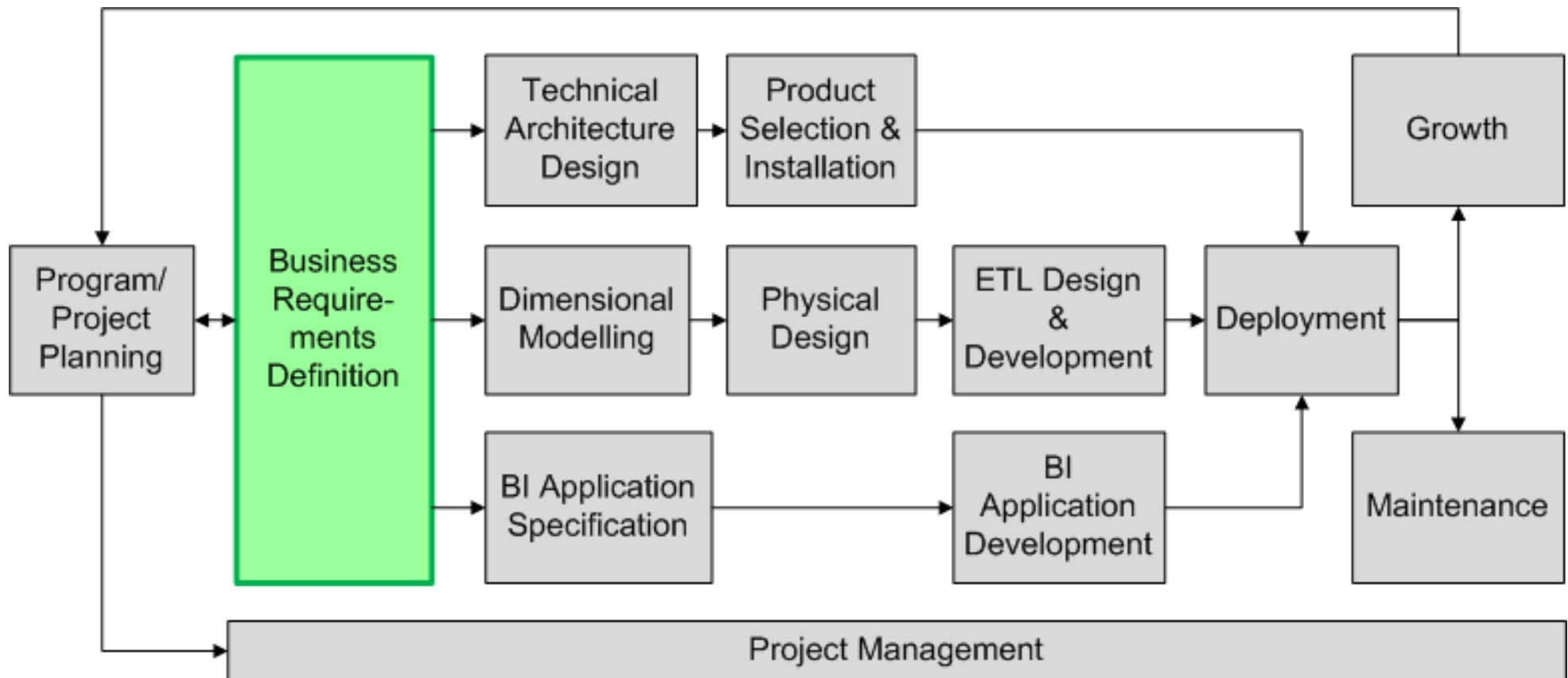
- Project Scope
- Project Business Justification (ROI)
- Project Communication Plan
- Project Plan and Task List
- Project Status Meeting Agenda
- Project Issue and Change Control Logs

Taskliste / Projektplan Demo



- Taskliste Kimball
- MS Projects Plan Template
- Project Communication Plan

Business Requirements Definition



- Verständnis entwickeln für die Vorgehensweise und Entscheidungsfindung der Business User
- Sammeln und dokumentieren der Business Requirements (Project)
- Prüfen ob die Verfügbarkeit und Verlässlichkeit der Sourcedaten die Business Requirements abdecken

!! Kritischer Prozess für Projekterfolg !!

Prozesshauptverantwortliche sind ...

- Project Managers
- Business Analysts
- Dimensional Modelers
- Project Sponsors

Weitere Rollen arbeiten mit oder liefern Input

Business Requirements:

- Interviews – Viele Details, einfach zu koordinieren. Interviewer und Protokollierer, optional 1-2 Beobachter
- Workshops – Mehrere Sichtweisen, breit abgesicherter Output, schwierig zu koordinieren
- Fragenkatalog, Attributlisten – Gut zur Vorbereitung oder Ergänzung
- Zielpublikum: Business Users, Management

Data Audit:

- Interviews
- Zielpublikum: IT Verantwortliche für Sourcedaten

- Enterprise DWH Bus Architecture Matrix
- Project Requirements Findings Document

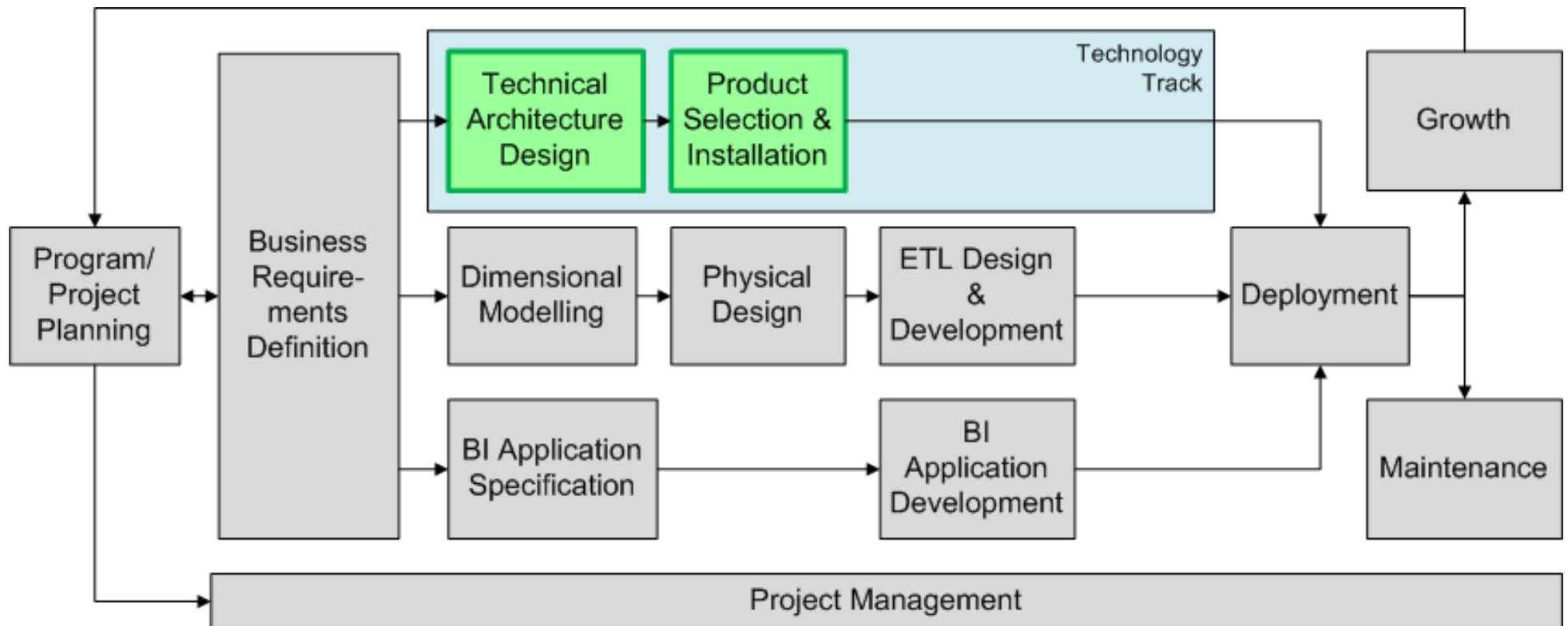
Beispiel High Level Enterprise DW Bus Matrix



Enterprise Data Warehouse Bus Matrix

| Business Process / Event | Common Dimensions | | | | | | | | |
|------------------------------|-------------------|----------|-----------|-------------------|---------|-----------|-----------------|-------|----------------|
| | Datum | Material | Lieferant | Produktionsanlage | Produkt | Spediteur | Wiederverkäufer | Kunde | Promotion Deal |
| Materialeinkauf | X | X | X | X | | X | | | |
| Materiallieferung | X | X | X | X | | X | | | |
| Materialinventar | X | X | X | X | | | | | |
| Verrechnung Material | X | X | | X | X | | | | |
| Produktfertigung | X | X | X | X | X | | | | |
| Auslieferung Wiederverkäufer | X | | | X | X | X | | | |
| Produktinventar | X | | | X | | | | | |
| Bestellung | X | | | | X | X | X | X | X |
| Kundenlieferung | X | | | | X | X | X | X | X |
| Fakturierung | X | | | | X | | X | X | X |
| Zahlung | X | | | | X | | X | X | X |
| Retouren | X | | | | X | X | X | X | X |

Technology Track



- Umsetzen der Business Requirements in einen *Architecture Plan*
- Produktauswahl
- Aufbau der Infrastruktur mit den benötigten DW/BI Produkten

High Level Application Architecture Model



Verantwortliche Rollen



Hauptverantwortlich:

- DW/BI Technical Architect (Lead)

Mitarbeit:

- ETL Architect / Developer
- BI Architect / Developer

Weitere Rollen arbeiten mit oder liefern Input

- *Architecture Task Force* bilden. Jemand mit Backroom-Erfahrung und jemand mit Front-Room Erfahrung einbinden. Ideal 2-4 Personen
- Extrahieren der relevanten Anforderungen für die Architektur (Antwortzeiten, Verfügbarkeit, Anzahl User, analytische Funktionalität, etc.)
- IT-Fachleute interviewen um vorgegebene *IT-Architektur-Standards* einzubinden

- Erstellen eines *Draft Architectural Implications Document*. Liste mit allen architekturelevanten Anforderungen, deren Auswirkung auf die Architektur und deren Priorität
- Erstellen des *Architecture Models*. Zuordnen aller architektonischen und funktionalen Auswirkungen

- Priorisierung der Komponenten für die Implementation (Iterationen) – Direkt auf *Architecture Model*
- *Architecture Plan* erarbeiten – Umfassendes Dokument beinhaltend alle Informationen aus den vorherigen Schritten mit Detailbeschreibungen

- Vier Hauptevaluationen:
- Hardware Plattform
- DBMS für DWH und OLAP
- ETL Tools
- BI Tools

- *Evaluation Matrix* erstellen – Direkt übernehmen der Anforderungen aus dem *Architecture Plan*, gewichten
- Geeignete Anbieter suchen (Internet, Network, Publikationen, Beraterfirma) – Shortlist
- Präsentationen und Hands-On Evaluation (Testversionen)
- Möglicherweise Prototyp

- Produktempfehlung
- Trial – Wenn möglich das Produkt inhouse für 90 Tage testen
- Vertrag – Zu beachten sind Lizenzart (Concurrent Users, Named Users, per CPU, etc.), Development Lizenzen, Consulting Services, Trainings, Dokumentation, Support levels

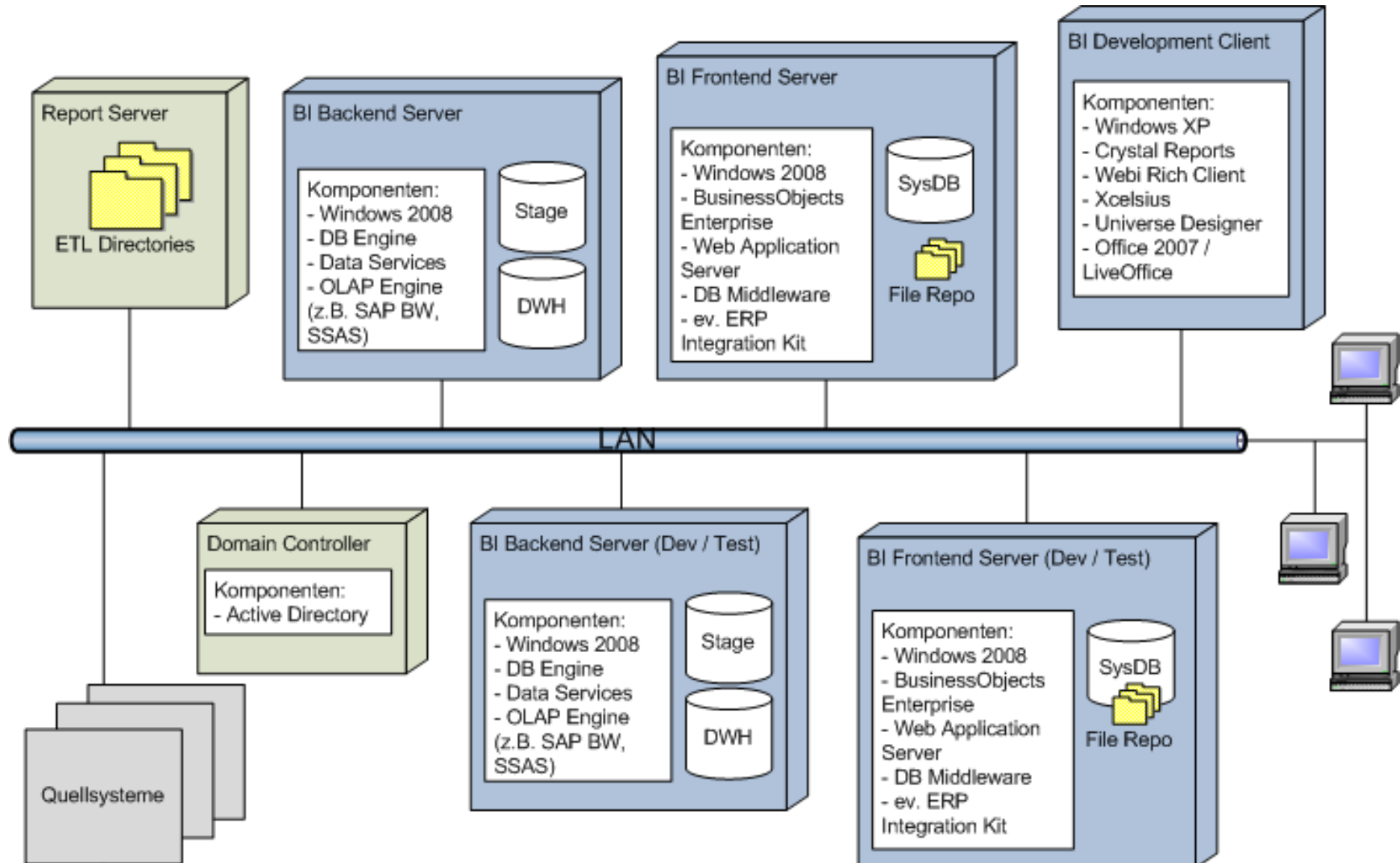
- Erstellen einer Infrastructure Map:
- Servers: Welche HW und SW
- Disk Systems: Welche Massenspeicher (z.B. SAN)
- Network: Welche Netzwerkanforderungen (vor allem für ETL Prozess)
- Security: Wie sollen Zugriffe geregelt werden
- Desktop: Betriebssystem und BI-Applikationen

- Architecture Plan
- Product Evaluation Matrix und Empfehlung
- Infrastructure Map
- Installierte DW/BI Infrastruktur

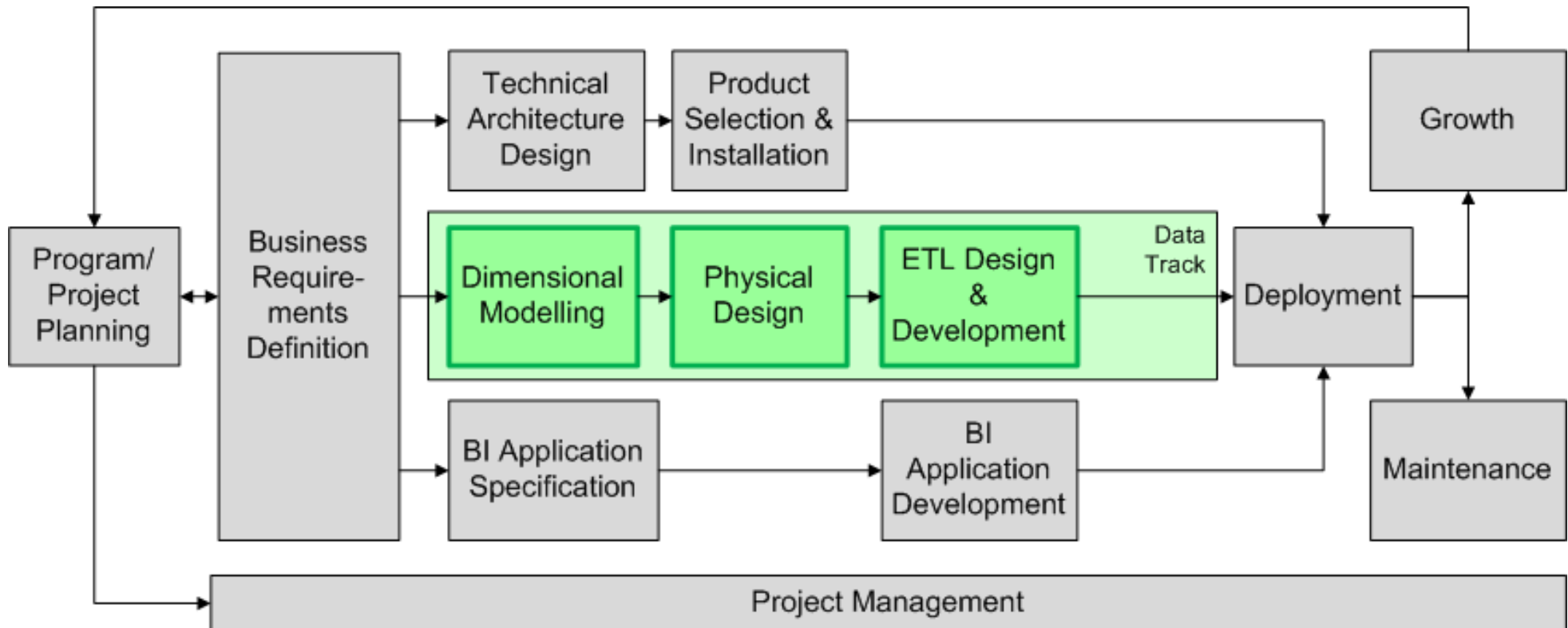
Weitere, hier nicht besprochene Outputs:

- Metadata Strategy, Metadata Inventory, Security Policies, Security Monitoring

Beispiel Infrastructure Map



Data Track



- Strukturieren von Daten optimiert auf Verständlichkeit und Query Performance

Dimensional Modelling: Verantwortliche Rollen

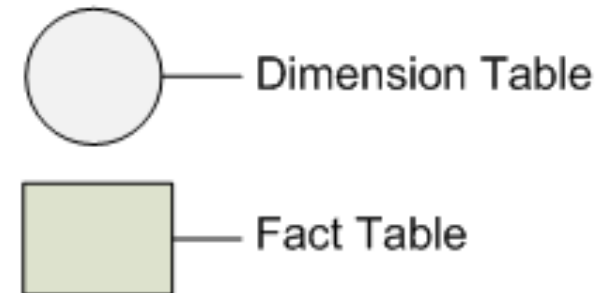
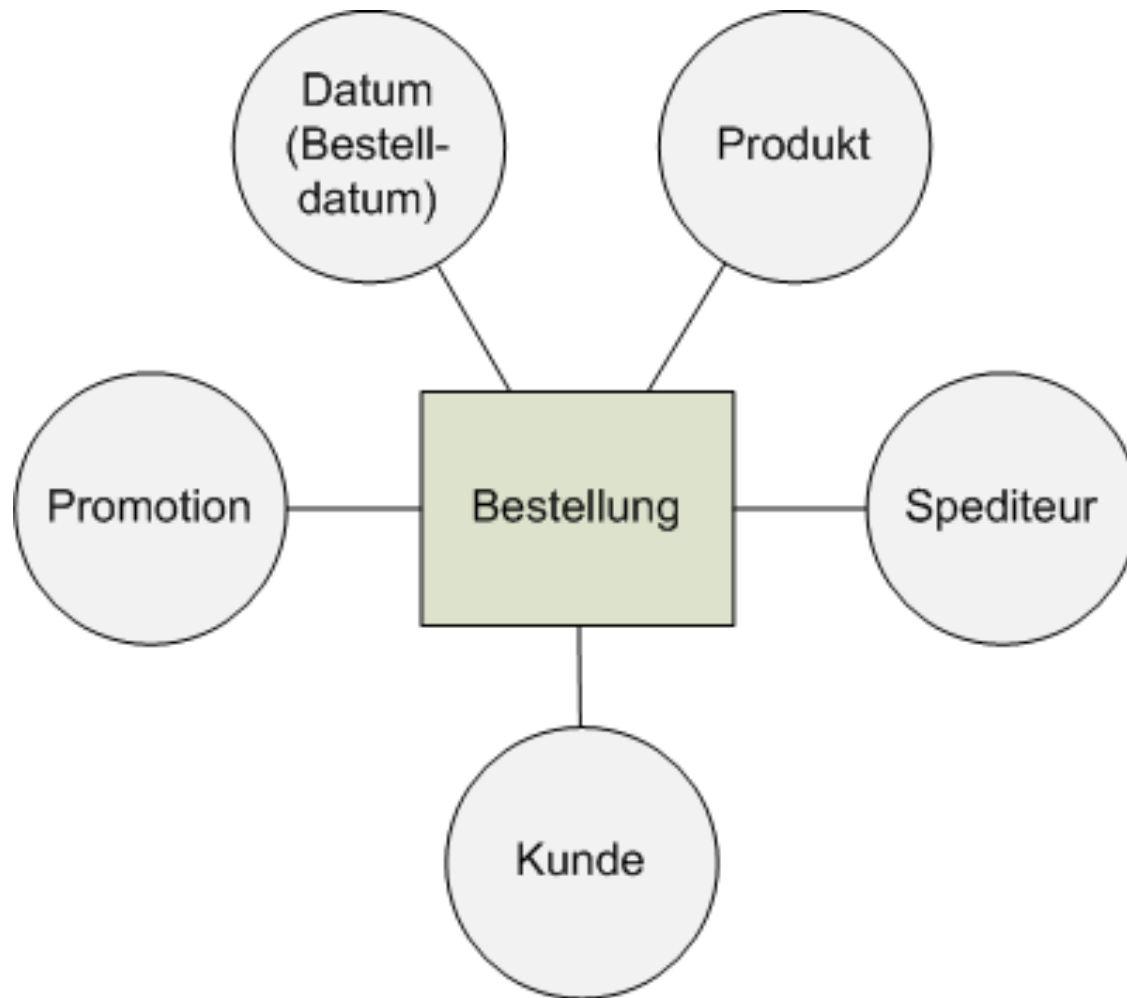


Hauptverantwortlich:

- Data Architect / Data Modeler (Lead)
- Business Analyst

Weitere Rollen arbeiten mit oder liefern Input

Dimensional Modelling: High Level Model Diagram



- Enthält die *Measures* die aus den Business Aktivitäten und Events generiert werden
- Enthält nur numerische Werte die additiv sind und Fremdschlüssel auf Dimensionen (Nicht additive numerische Werte wie Produktpreis gehören in die Dimensionen!)
- Grosse Datenmengen
- Höchst möglicher Detailgrad um Flexibilität für Auswertungen und Anpassungen zu gewährleisten

Dimensional Modelling: Dimension Table



- Enthält Daten für Einschränkungen und Filter in Querys sowie Namensgebung im Result Set
- „Nach“ ist das Wort für Dimensionen.
Daten...
- ...nach Jahr
- ...nach Produkt
- ...nach Region
- etc.

- Business Prozess auswählen der abzubilden ist (aus der Enterprise DW Bus Matrix)
- Granularität (Detailgrad) der Fact Table bestimmen
- Dimensionen mit Attributen identifizieren
- Facts (Measures, Metrics) mit Attributen identifizieren

Dimensional Modelling: Deliverables



- High Level Model Diagram
- Attribute and Metrics List
- Dimensional Design Worksheet
- Issues List (Offene Punkte Liste)

Dimensional Modelling: Attribute and Metrics List



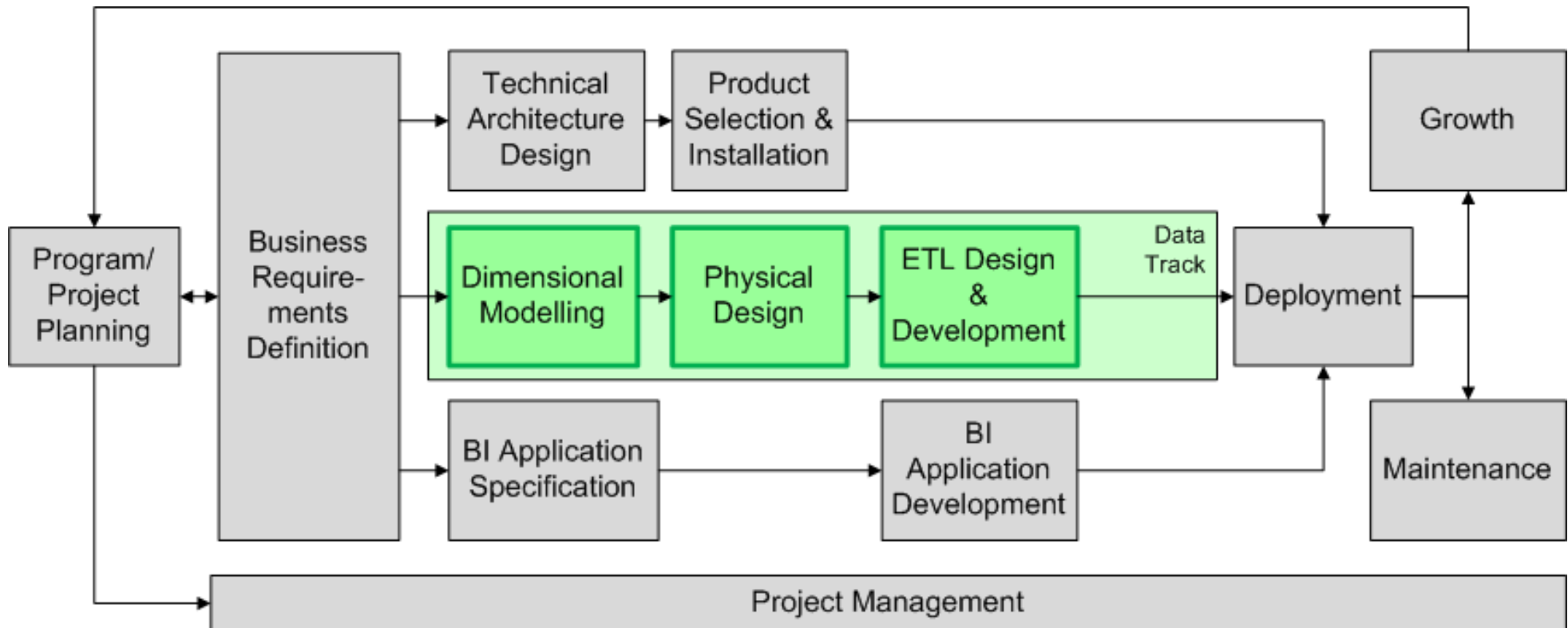
Attribute and Metrics List

Table: Promotion
Tabletype: Dimension

| Attribute Name | Description | Alternate Names | Sample Value |
|----------------------------|--|---------------------|----------------------------------|
| Spezialangebot ID | Source System Key | | |
| Spezialangebot Bezeichnung | Name / Beschreibung des Spezialangebots | Promotionsname | Volumenrabatt 11 bis 14 |
| Rabatt Prozent | Prozent Rabatt auf dem produkt | | |
| Spezialangebot Typ | Beschreibung des Rabatttyps | Promotionstyp | Volumenrabatt; Auslaufprodukt |
| Spezialangebot Kategorie | Verkaufskanal für den die Spezialangebot gültig ist | Promotionskategorie | Wiederverkäufer; Kunde |
| Startdatum | Erster Tag an dem das Spezialangebot gilt | | 01.01.2008 |
| Enddatum | letzter Tag an dem das spezialangebot gilt | | 30.06.2008 |
| Minimum Bestellmenge | Minimale Bestellmenge für das Spezialangebot | | 0 |
| Maximum Bestellmenge | Maximal erlaubte Bestellmenge für das Spezialangebot | | |

- *The Data Warehouse Toolkit, Second Edition* geht detailliert und mit branchenspezifischen Beispielen auf das Thema Dimensional Modelling ein

Data Track: Physical Design



- Implementation auf Development- und Test-Umgebung:
- DWH
- Staging Tables
- Security
- Aggregation
- Indexierung

Physical Design: Verantwortliche Rollen



Hauptverantwortlich:

- DBA (Lead)
- Data Architect / Data Modeler
- Security Manager

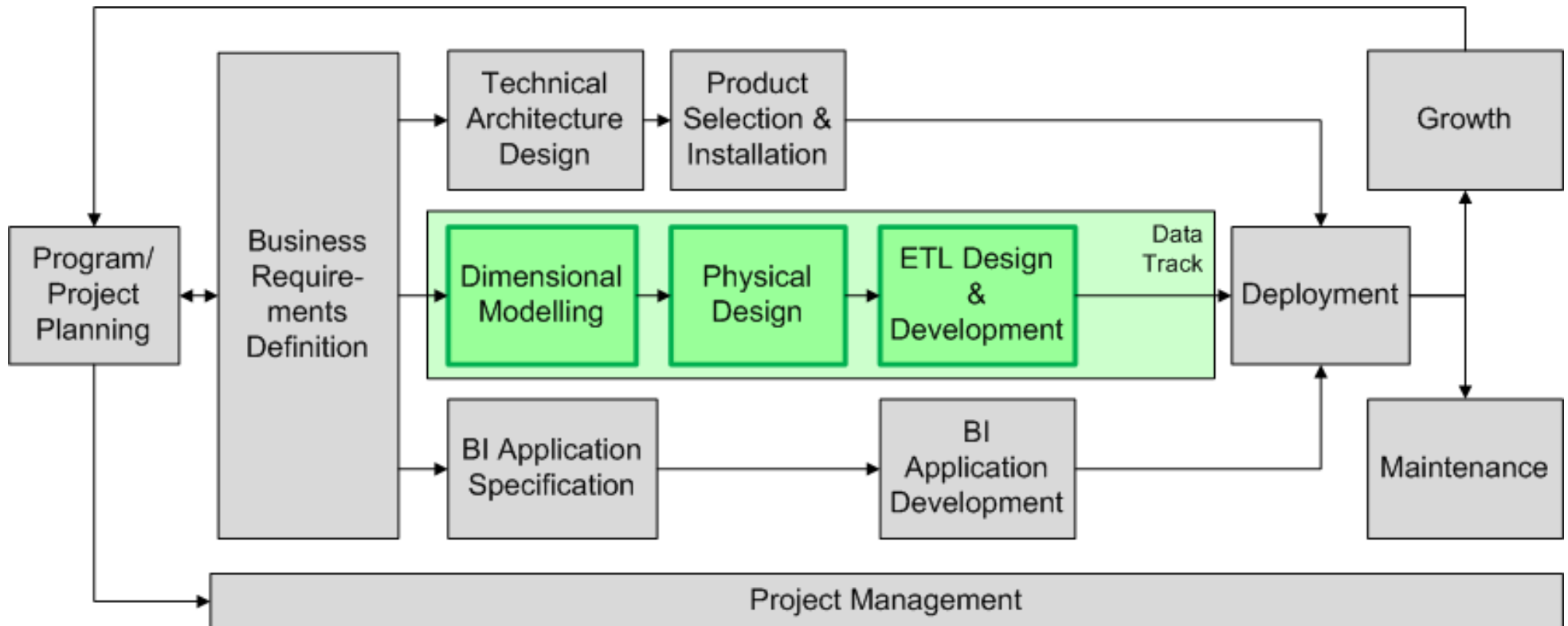
Weitere Rollen arbeiten mit oder liefern Input

- *Standards* für Namenskonventionen, Null-Handling, Sourcecodeablage (Source Control System), Stage-Tables, Primary Keys (Surrogate keys), Foreign Keys, Synonyme erstellen
- ERD *Modell* erstellen
- *Instanz* auf Developmentserver erstellen
- Design der *Processing Datastores* – Staging Tables, Auditing Tables, Monitoring Tables (User Access), Security Tables

- Initialen Indexierungsplan erstellen
- OLAP Design – Granularität und Drillthrough, Kalkulationen
- Instanz auf Testserver erstellen
- Aggregationen designen und Indexierung optimieren

- Physical Data Model on Development Server
- Endgültige Source-to-Target Map
- Initial Index Plan
- OLAP DB Design
- Aggregation Plan
- Partition Plan

Data Track: ETL Design & Development



ETL Design & Development: Ziele



- Einbringen der Sourcedaten in das DWH in der geforderten Qualität und Granularität

- Aufwendigster Prozess im ganzen Lifecycle. Kann bis 70% des ganzen Design- und Implementierungsaufwands ausmachen
- 4 Hauptbereiche mit 34 Subsystemen:
- Extracting
- Cleaning and Conforming
- Delivering
- Managing

ETL Design & Development: Vorgehen



ETL-Subsystems

| Prozessschritt | ETL Subsystem | | | |
|--|-----------------|-------------------------|------------------------------|--------------------|
| | Extracting Data | Cleaning and Conforming | Delivering for Presentation | Manage Environment |
| Plan | | | | |
| High level Plan Source-to-Target | 1 | | | |
| Testen, auswählen und implementieren ETL-Tool (siehe Technology Track) | | | | |
| Default Strategien für Dimensionen, Error Handling | 3 | 4, 5, 6 | 10 | |
| Drill Down pro Zieltabelle - grafische Darstellung Daterestrukturierung und Transformationen sowie Jobablaufplan | | 4, 5, 6 | 11 | 22 |
| Implementierung einmaliger Historic Load | | | | |
| Implementieren und testen des initialen Dimensionsladeprozesses | 3 | 4, 7, 8 | 9, 10, 11, 12, 15 | |
| Implementieren und testen des initialen Factsladeprozesses | 3 | 4, 5, 8 | 13, 14 | |
| Implementierung inkrementeller Load | | | | |
| Implementieren und testen des inkrementellen Dimensionsladeprozesses | 2, 3 | 4, 7, 8 | 9, 10, 11, 12, 15, 16, 17 | |
| Implementieren und testen des inkrementellen Factsladeprozesses | 2, 3 | 4, 5, 8 | 13, 14, 16, 18 | |
| Implementieren und testen des Aggregationstabellen und/oder OLAP Ladeprozesses | | | 19, 20 | |
| Implementieren der ETL-System-Automation | | 6 | 17, 18, 21 | 22, 23, 24, 30 |

- *ETL Plan* entwerfen:
- High Level Plan – Grafische Darstellung Source-to-Target mit den wichtigsten Herausforderungen
- ETL Tool Auswahl (Technology Track)
- Default Strategien für Dimension Management, Error Handling
- Drill Down pro Target Table – grafische Darstellung aller Transformationen, Jobablaufplan

ETL Design & Development: Vorgehen



- *ETL Plan* entwerfen:
- Zusammenstellen aller Informationen plus zusätzlichen Detailinfos im *ETL Specification Document*

ETL Design & Development: Vorgehen



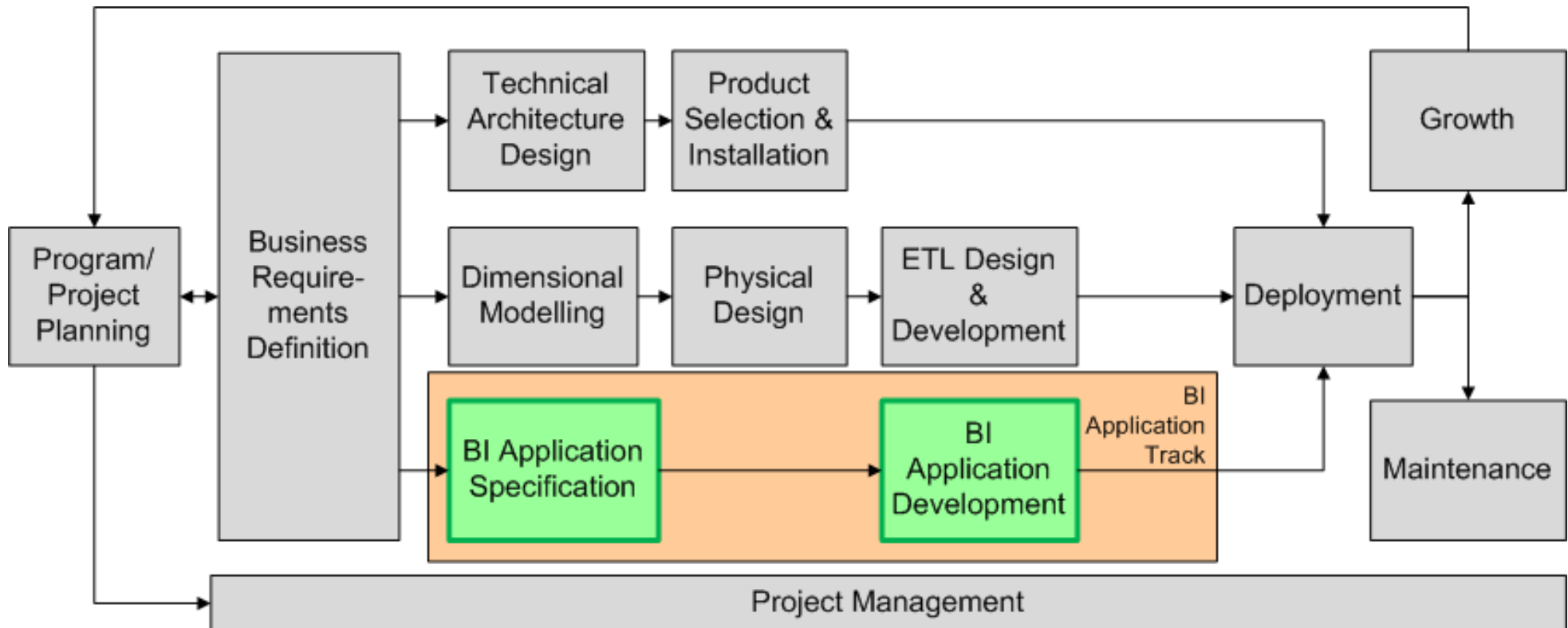
- Einmaligen *Historic Load* implementieren:
- Implementieren und testen des initialen Dimensionenladeprozesses
- Implementieren und testen des initialen Factladeprozesses

- Inkrementellen Ladeprozess implementieren:
- Implementieren und testen des inkrementellen Dimensionenladeprozesses
- Implementieren und testen des inkrementellen Factladeprozesses
- Implementieren und testen des Aggregationstabellen- und/oder OLAP-Ladeprozesses
- Implementieren der ETL-System-Automation

- *The Data Warehouse ETL Toolkit* geht detailliert auf das Thema ETL ein

- *ETL System Specification* mit High Level Plan, Default Strategies und ETL Plan für jede Tabelle
- *ETL System* mit einmaligem historischem Load, inkrementellem Load, Aggregationen/OLAP und Automation

BI Application Track



BI Application Specification : Ziele



- Korrekte, performante, einfach zu benutzende und „gutausssehende“ BI-Applikationen nach Benutzerbedürfnissen
- Einfacher, einheitlicher Zugang zu den BI-Applikationen (Portal)

BI Application Specification: Ziele

BI Application Type

Consumer Mode

Analysis Type

Nbr of Users

Direct Access
Query,
Reporting, Data
Mining

Standard
Reports

Analytic
Applications

Dashboards,
Scorecards

Operational BI

BI Portal



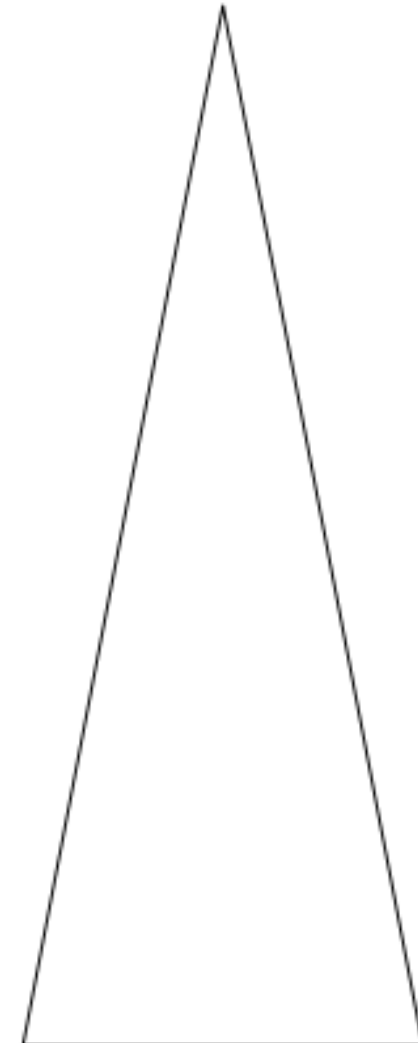
Ad hoc Access

Strategic

Push-button
Access

Operational
Reporting

Tactical



BI Application Specification : Verantwortliche Rollen



Hauptverantwortlich:

- BI Application Architect / Developer (Lead)
- Business Analyst

Mitarbeit:

- Key Business Users (Report Design)

Weitere Rollen arbeiten mit oder liefern Input

- Kann und soll direkt nach Aufnahme der Business Requirements beginnen (Alles noch frisch in den Köpfen!)
- *Standards* und *Templates erstellen* – Namens- und Formattierungskonventionen, Applikationskonventionen (Inhalt, Struktur), Reporttemplate(s), Dashboardtemplate(s), Logo, etc.

- Bestimmen des initialen *Application Set* –
Identifizieren der Report Kandidaten (Liste),
Konsolidieren der Liste (Redundante
Reports), Priorisieren (grosser Business
Value, kleiner Aufwand -> hohe Prio)
- Mit einem „kleinen“ Set anfangen
- Detaillierte *Application Specifications*
erstellen – Pro Report/Dashboard Section:
Prototyp, Liste der User Interaktionen,
Dataset, Kalkulationen, Business Rules,
Interaktionen mit anderen Reports

- Top Level Design des *Navigation Frameworks* und des *Portals* – Organisation nach Business Prozessen hat sich bewährt
- Review und Validierung der Applications Specs gegen das Dimensional Model – Alles da?
- Review der Specs mit den Business Users

- Vorbereitung der Implementierung – Installation und Parametrisierung der BI-Tools (sehr unterschiedlich nach Anbieter), Test, Security aufsetzen
- *Reports* implementieren – User Interaktionen definieren, Queries erstellen, Formattieren des Resultats
- *Dashboards/Scorecards* implementieren

- *Operational BI* implementieren – Kann Mix aus operativen und DWH Systemen sein, möglichst auf operativen Systemen (grosse Zahl User, aktuelle Informationen)
- *Data Mining Models* implementieren
- Validieren der Daten und des Datenmodells anhand von Tests, Rework
- *Portal* implementieren
- Scheduling und automatische Verteilung

BI Application Development: Vorgehen



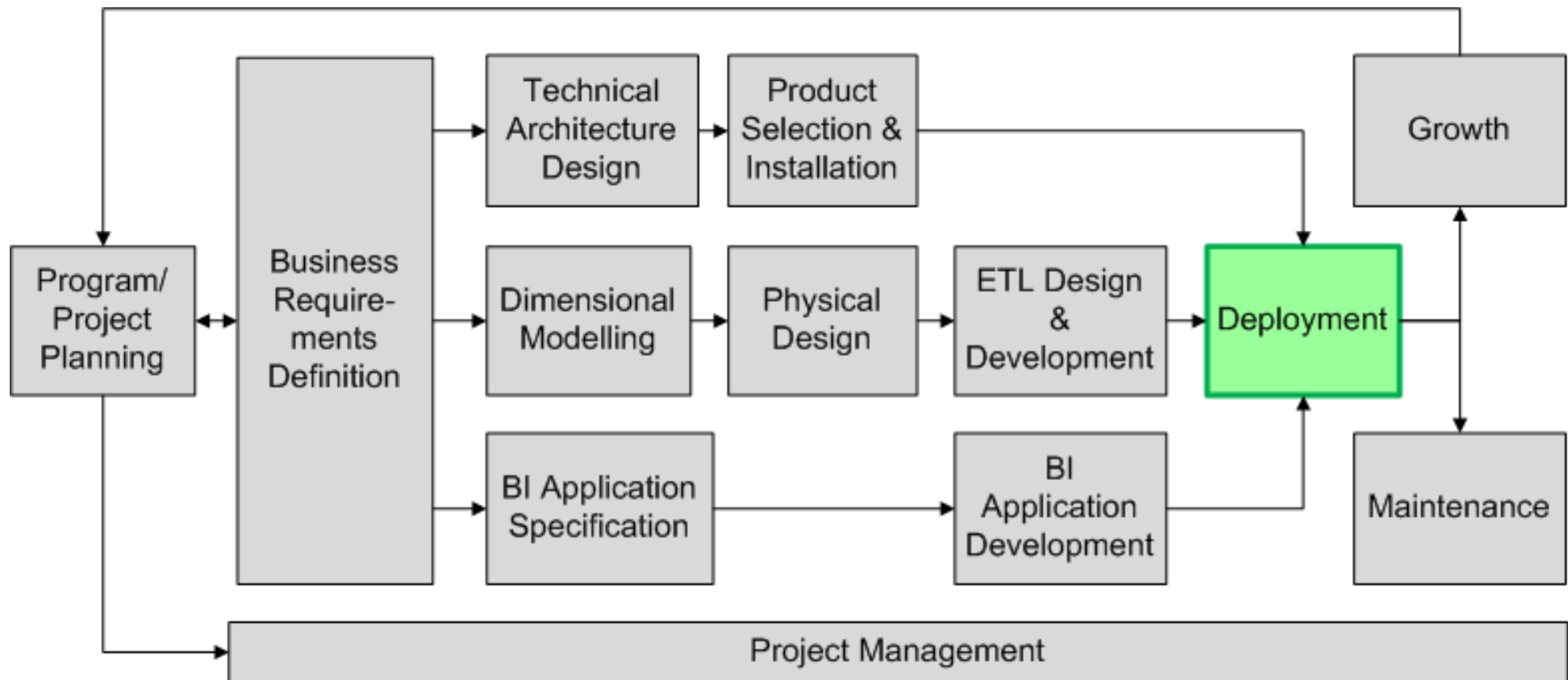
- Test, Rework, Performance Tuning
- Dokumentation
- Deployment planen

BI Application Development: Deliverables



- Application Development and Naming Standards Document
- Templates
- Candidate Application List
- Application Specs
- Navigation Framework and BI Portal Design
- Initial Applications and BI Portal

Deployment



- *Pre-Deployment Testing* – Datenqualität, End-to-End Prozess, Performance, Usability, Deployment, User Desktop
- *Deployment* – Relationale DB (via Scripts), ETL, OLAP DB, Reports/BI-Applications
- Dokumentation und Training

- Test Plan – Tools, Scripts, Daten
- System Deployment Playbook – genaue Vorgehensbeschreibung
- Business User Documentation
- Training Materials

Besten Dank!!



Vielen Dank für Ihr Interesse
und
Auf Wiedersehen