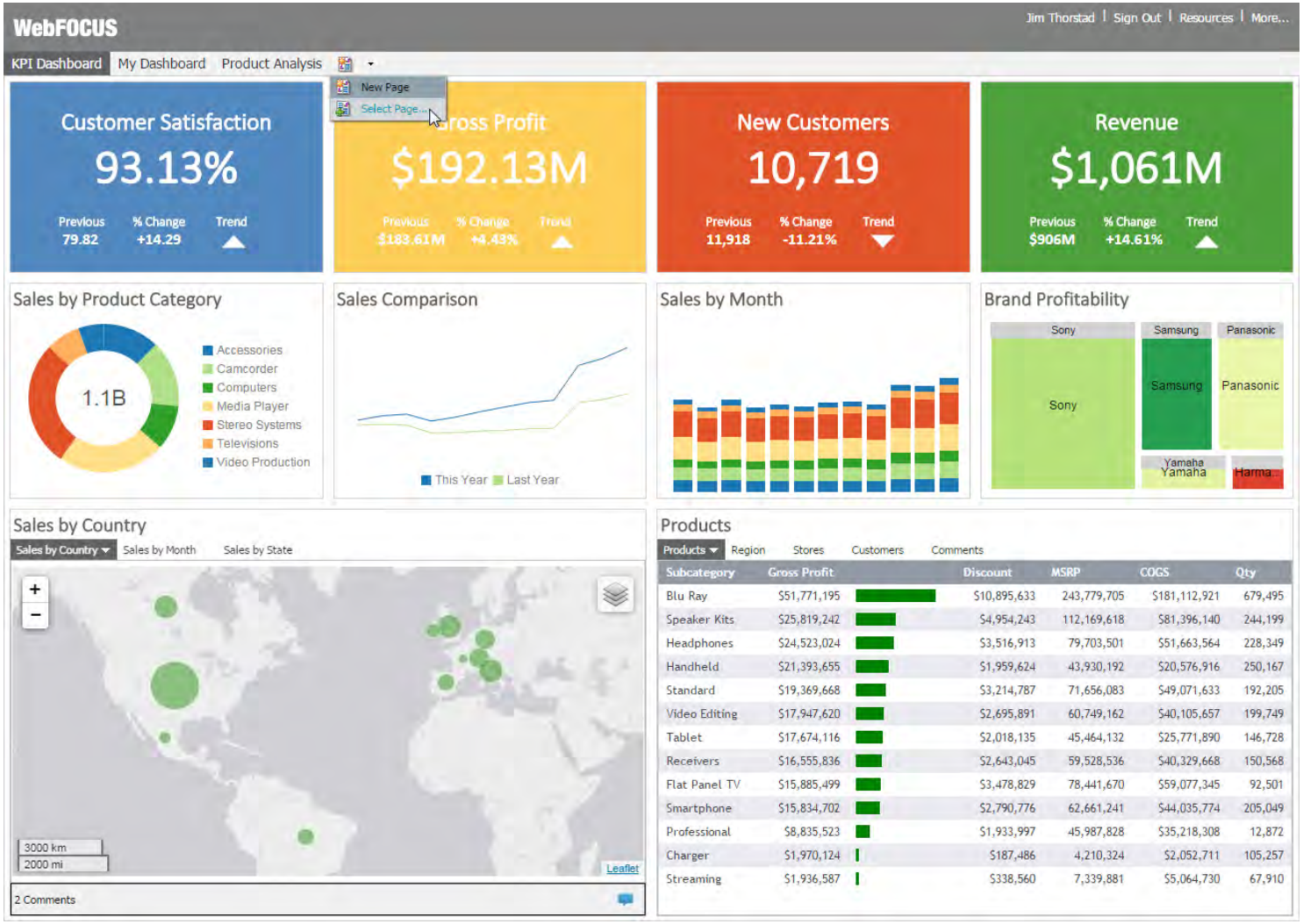

BUSINESS
INTELLIGENCE
DATA
BI
IMPLEMENTATION
INFORMATION
SYSTEM
SYSTEMS
ANALYSIS
INTERVIEW
REQUIREMENTS
ORDER
PROCESS
ORGANIZATION
SPONSOR
DECISION
USED
SEMI-STRUCTURED
INTERVIEWEES
SUPPORT
MANAGEMENT
UNSTRUCTURED
METADATA
ANALYTICS
NEED
PROJECT
GATHERING
NEEDS
USERS
MUST
DATA
CONTENT

Who this presentation is for

- Future ETL developers (back end)
- Future development DBAs
- Future Data Modelers/Data Architects
- Future BI Developers aka Report Developers (front end)

What do you need to know to proceed?

- Relational database theory
- Data normalization
- Any reporting/presentation tool that makes use of charts and graphs



Subcategory	Gross Profit	Discount	MSRP	COGS	Qty
Blu Ray	$51,771,195	$10,895,633	243,779,705	$181,112,921	679,495
Speaker Kits	$25,819,242	$4,954,243	112,169,618	$81,396,140	244,199
Headphones	$24,523,024	$3,516,913	79,703,501	$51,663,564	228,349
Handheld	$21,393,655	$1,959,624	43,930,192	$20,576,916	250,167
Standard	$19,369,668	$3,214,787	71,656,083	$49,071,633	192,205
Video Editing	$17,947,620	$2,695,891	60,749,162	$40,105,657	199,749
Tablet	$17,674,116	$2,018,135	45,464,132	$25,771,890	146,728
Receivers	$16,555,836	$2,643,045	59,528,536	$40,329,668	150,568
Flat Panel TV	$15,885,499	$3,478,829	78,441,670	$59,077,345	92,501
Smartphone	$15,834,702	$2,790,776	62,661,241	$44,035,774	205,049
Professional	$8,835,523	$1,933,997	45,987,828	$35,218,308	12,872
Charger	$1,970,124	$187,486	4,210,324	$2,052,711	105,257
Streaming	$1,936,587	$338,560	7,339,881	$5,064,730	67,910

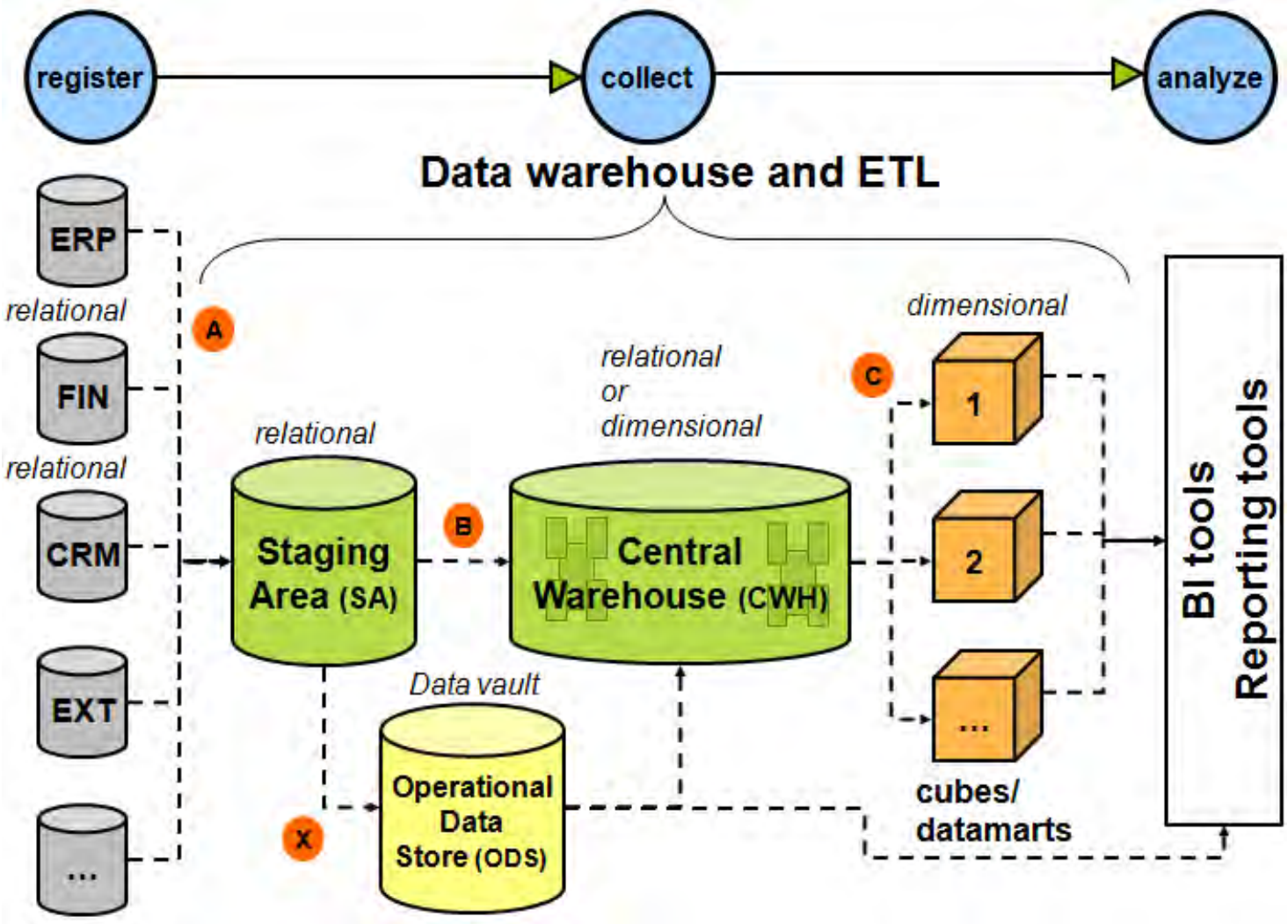
Software Components of BI

- BI = Integrated Data Store + ETL + Reports
  ETL = Extract Transform and Load
  Integrated Data Store = Data Mart or Data Warehouse


BI TOOLS
COGNOS
BO
SAP BW
ACTUATE BIRT
SSRS
JASPERSOFT
TABLEAU
QLIKVIEW
DATABASES
ORACLE
DB2
TERADATA
SQL SERVER
ANALYTICS SERVERS
NETEZZA
COGNOS
BO
SAS
SSAS
ETL TOOLS
INFORMATICA
BODI
PLSQL
ASCENTIAL
SWISSSQL
SSIS
PENTAHO/KETTLE

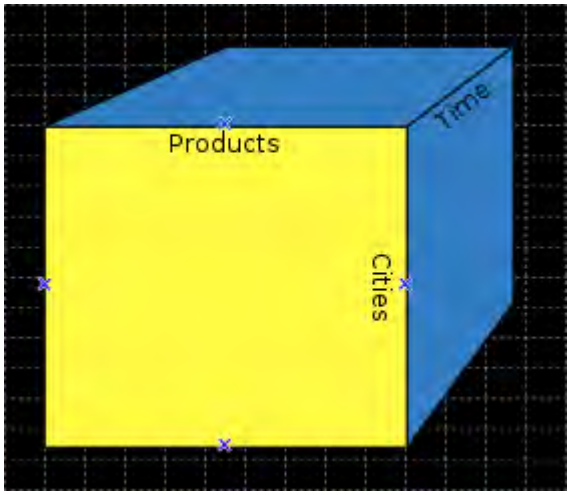
Why Integrate?

- One stop shop for the reporting
- Everything is part of the system
- Single version of the truth
- To improve performance and availability
- To enable history
- The right granularity level
- Reporting data models

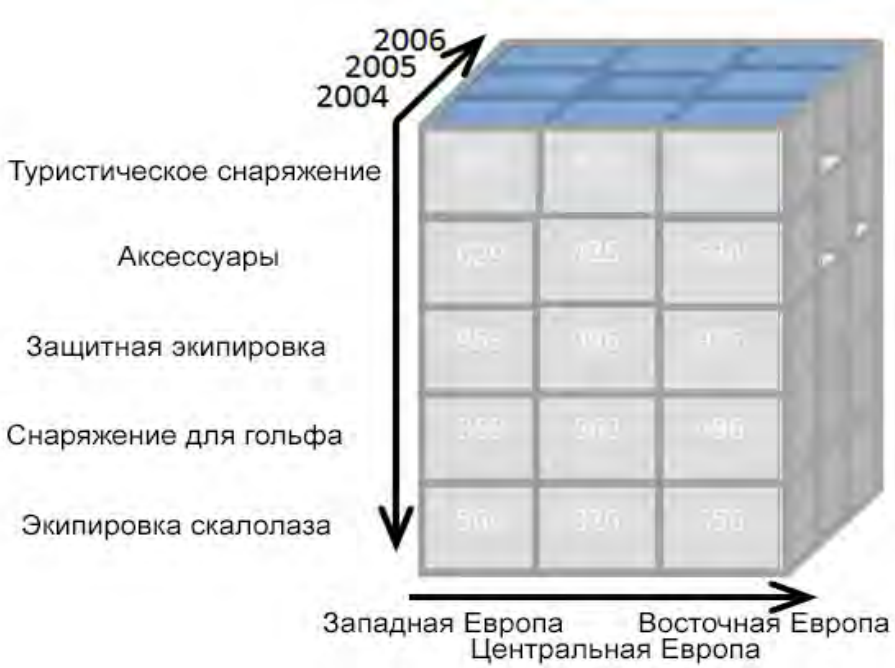

register
collect
analyze
Data warehouse and ETL
ERP
relational
FIN
relational
CRM
EXT
...
A
relational
Staging Area (SA)
B
relational or dimensional
Central Warehouse (CWH)
C
dimensional
1
2
...
cubes/ datamarts
BI tools
Reporting tools
Data vault
X
Operational Data Store (ODS)

OLAP CUBE

Online Analytical Processing



SLICE

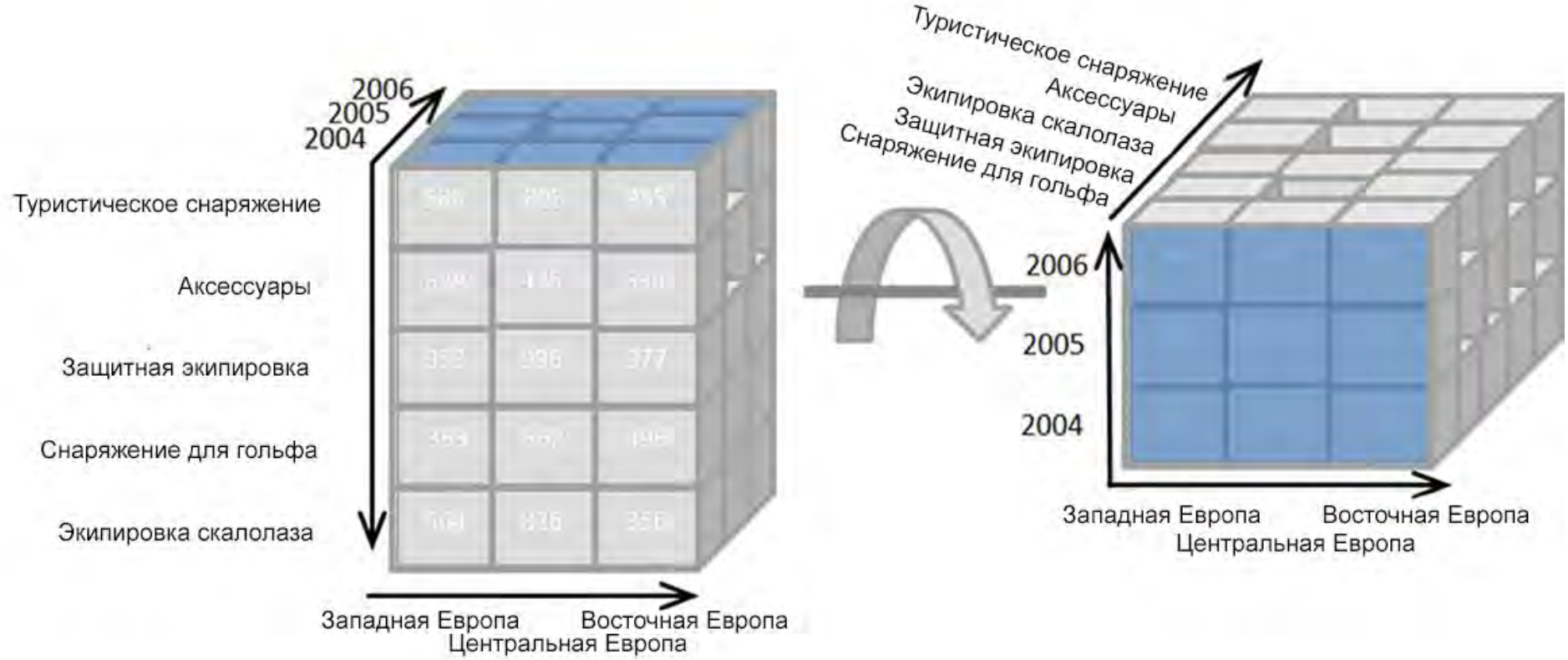


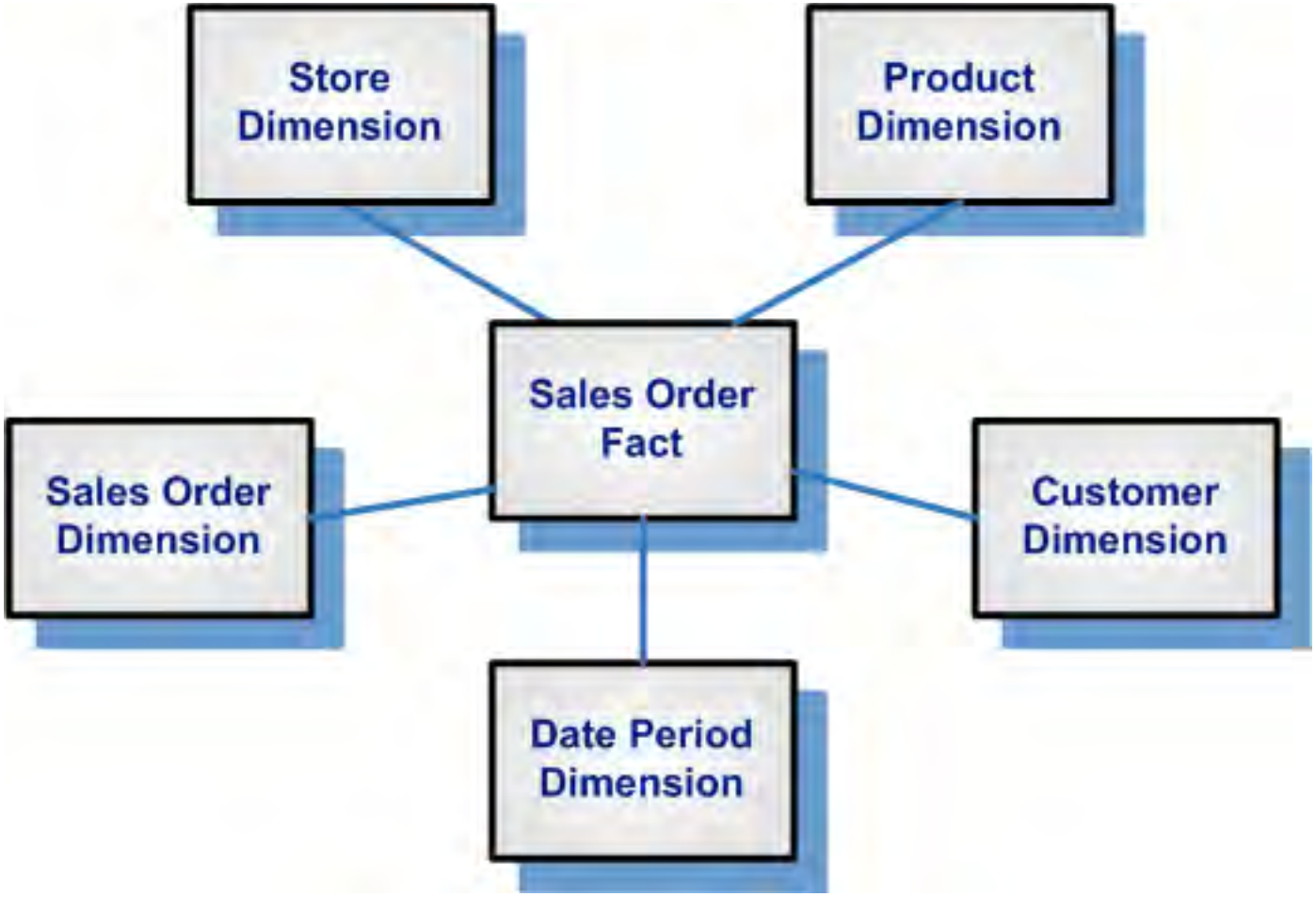
DICE

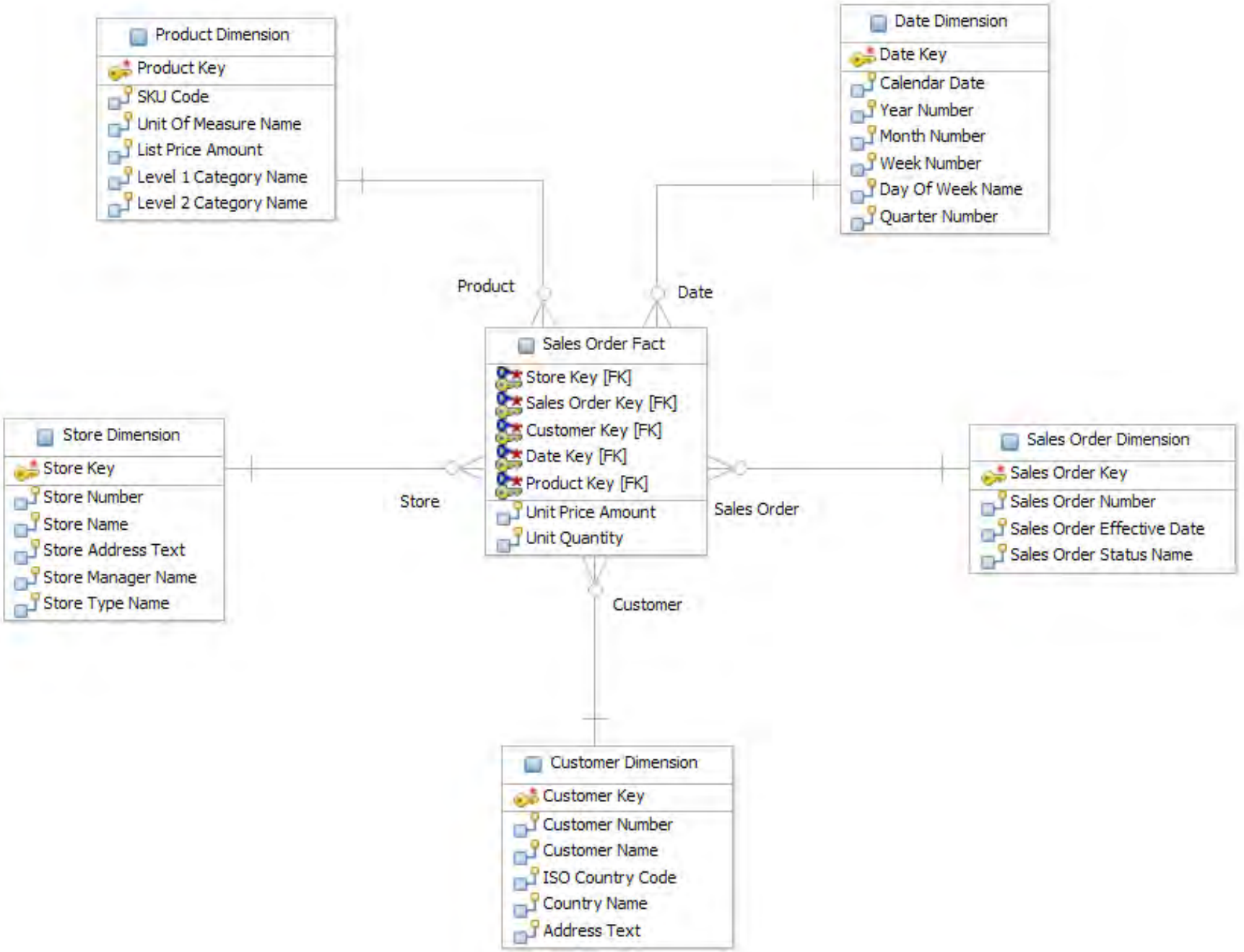

2006
2005
2004
Туристическое снаряжение
Аксессуары
Защитная экипировка
Снаряжение для гольфа
Экипировка скалолаза
Западная Европа
Восточная Европа
Центральная Европа
Туристическое снаряжение
Защитная экипировка
Снаряжение для гольфа
2006
2005
2004
Западная Европа
Восточная Европа
Центральная Европа

Drill Up/Down

PIVOT


2006
2005
2004
Туристическое снаряжение
Аксессуары
Защитная экипировка
Снаряжение для гольфа
Экипировка скалолаза
Западная Европа
Восточная Европа
Центральная Европа
Туристическое снаряжение
Аксессуары
Экипировка скалолаза
Защитная экипировка
Снаряжение для гольфа
2006
2005
2004
Западная Европа
Восточная Европа
Центральная Европа


Store Dimension
Product Dimension
Sales Order Fact
Sales Order Dimension
Customer Dimension
Date Period Dimension


Product Dimension
Product Key
SKU Code
Unit Of Measure Name
List Price Amount
Level 1 Category Name
Level 2 Category Name
Date Dimension
Date Key
Calendar Date
Year Number
Month Number
Week Number
Day Of Week Name
Quarter Number
Product
Date
Sales Order Fact
Store Key [FK]
Sales Order Key [FK]
Customer Key [FK]
Date Key [FK]
Product Key [FK]
Unit Price Amount
Unit Quantity
Store Dimension
Store Key
Store Number
Store Name
Store Address Text
Store Manager Name
Store Type Name
Store
Sales Order
Sales Order Dimension
Sales Order Key
Sales Order Number
Sales Order Effective Date
Sales Order Status Name
Customer
Customer Dimension
Customer Key
Customer Number
Customer Name
ISO Country Code
Country Name
Address Text

```
SELECT
   'sof.unit quantity'*'sof.unit price amount‘  as ‘total amount’
FROM
   'sales order fact' sof
INNER JOIN
   'sales order dimension' sod
ON
   'sod.sales order key'='sof.sales order key'
INNER JOIN
   'customer dimension' cd
ON
   'cd.customer key' = 'sof.customer key'
INNER JOIN
   'store dimension' sd
ON
   'sd.store key' = 'sof.store key'
INNER JOIN
   'date dimension' dd
ON
   'dd.date key' = 'sof.date key'
INNER JOIN
   'product dimension' pd
ON
   'pd.product key' = 'sof.product key'
```

Sales by customer - SLICING

Where ‘customer dimension.customer name’=‘IBM’

- Sales by product category and date - DICING

Where ‘product dimension.level2 category name’ in (‘Firewall’,’Antivirus’) and ‘date dimension.year number’=2018

- Sales by product category, DRILL DOWN from level 1 to level 2

Where ‘product dimension.level1 category name’=‘Security Software’

- DRILL DOWN from level 2 to the product level

Where ‘product dimension.level1 category name’=‘Security Software’ and ‘product dimension.level2 category name = ‘Firewall’

Пример взаимодействия в BI проекте

- Заказчик принимает решение о создании нового отчета
- Бизнес аналитик готовит требования, работая со специалистами на стороне заказчика
- Архитектор данных данных принимает требования от аналитика и согласует их или шлет на доработку

- Архитектор данных делает изменения в модели данных и согласует их с разработчиками отчетов (если необходимы изменения в модели данных)
- Архитектор данных формулирует техническое задание для ETL разработчиков
- Backend и Front-end разработчики выполняют свою работу
- Готовые Front End и Back End части устанавливаются на тест среду для тестирования бизнес аналитиком и/или SME заказчика

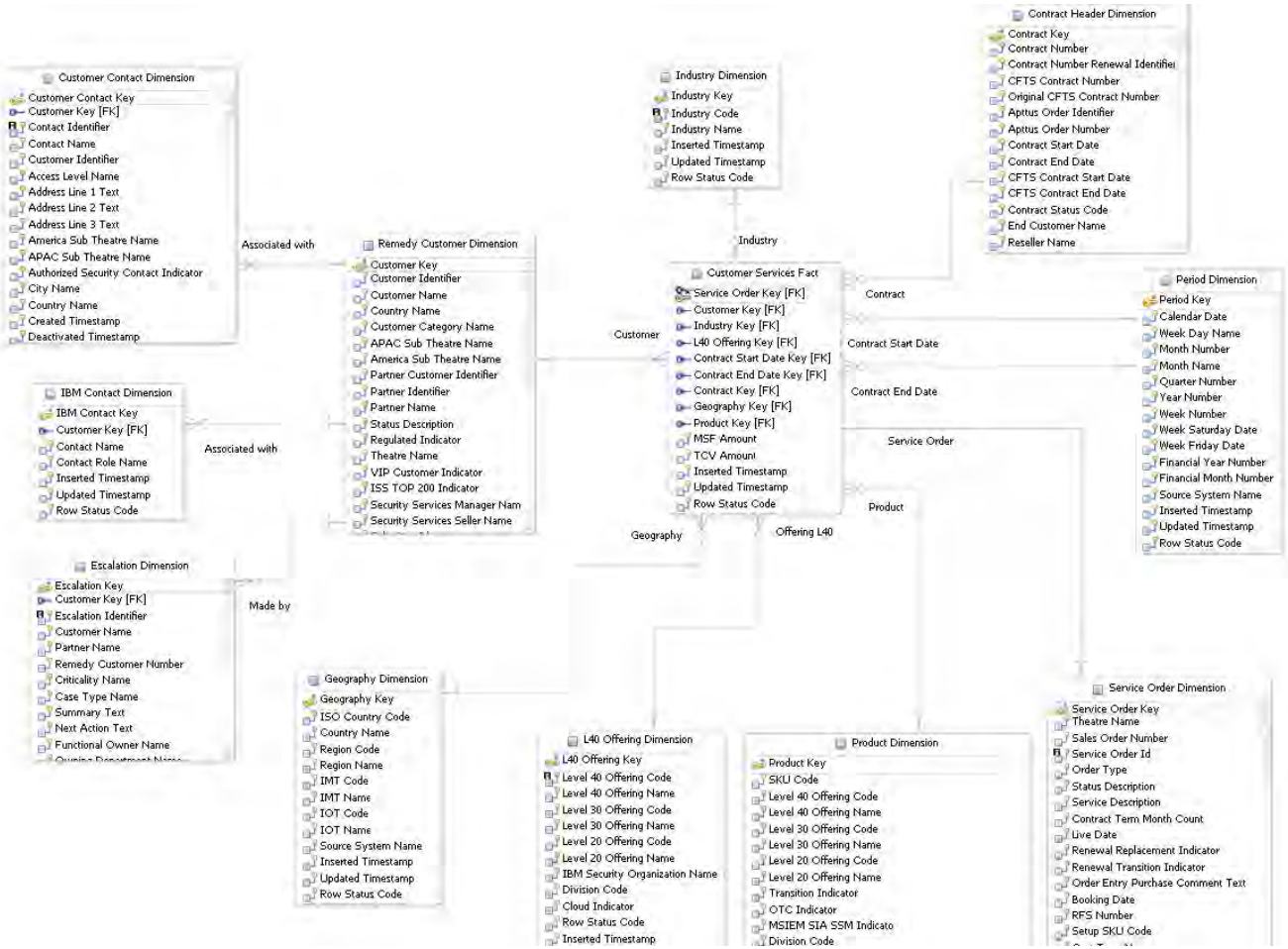
Роли и качества участников BI проекта

- Знание бухучета
- Работа с заказчиком на его «языке», умение «извлечь» нужную информацию
- Постоянное изучение новых для себя областей производств и систем автоматизации
- Документирование требований
- Работа с базами данных и системами построения отчетов на уровне пользователя
- Хорошее понимание принципов построения пользовательского интерфейса
- Углубленное понимание возможностей, особенностей и ограничений используемых в проекте программных средств

Data Architect

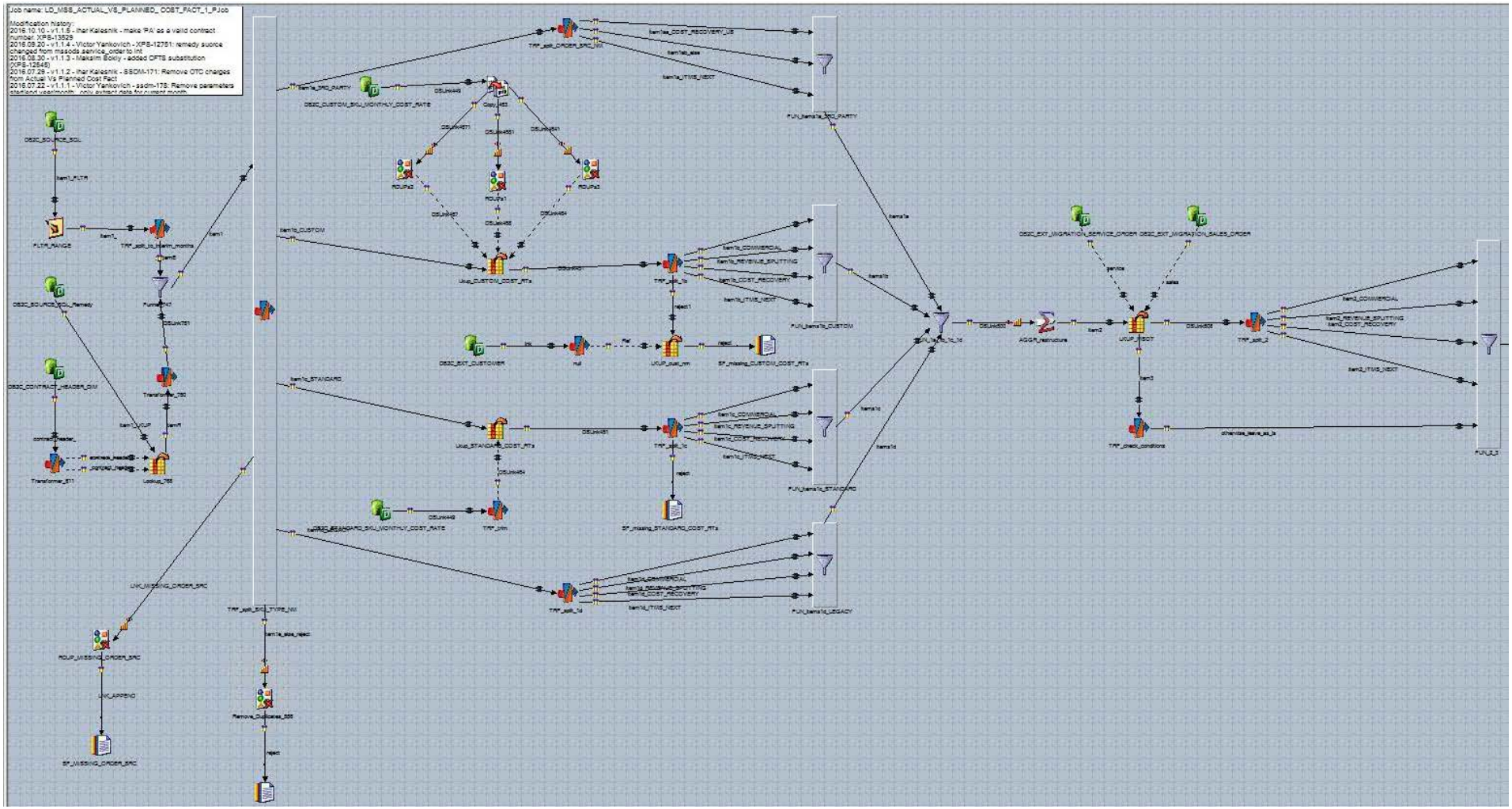
- Работа со средствами моделирования данных
- Углубленное понимание основных концепций моделирования и интегрирования данных
- Анализ данных
- Углубленное понимание принципов построения ETL систем, а также конкретных программных продуктов, используемых в проекте
- Хорошее понимание принципов построения автоматизированных отчетов
- Руководство ETL командой
- Продвинутый уровень пользователя СУБД, знания из области администрирования используемых операционных систем и баз данных
- Работа над улучшением производительности запросов

Star Schema Data Model

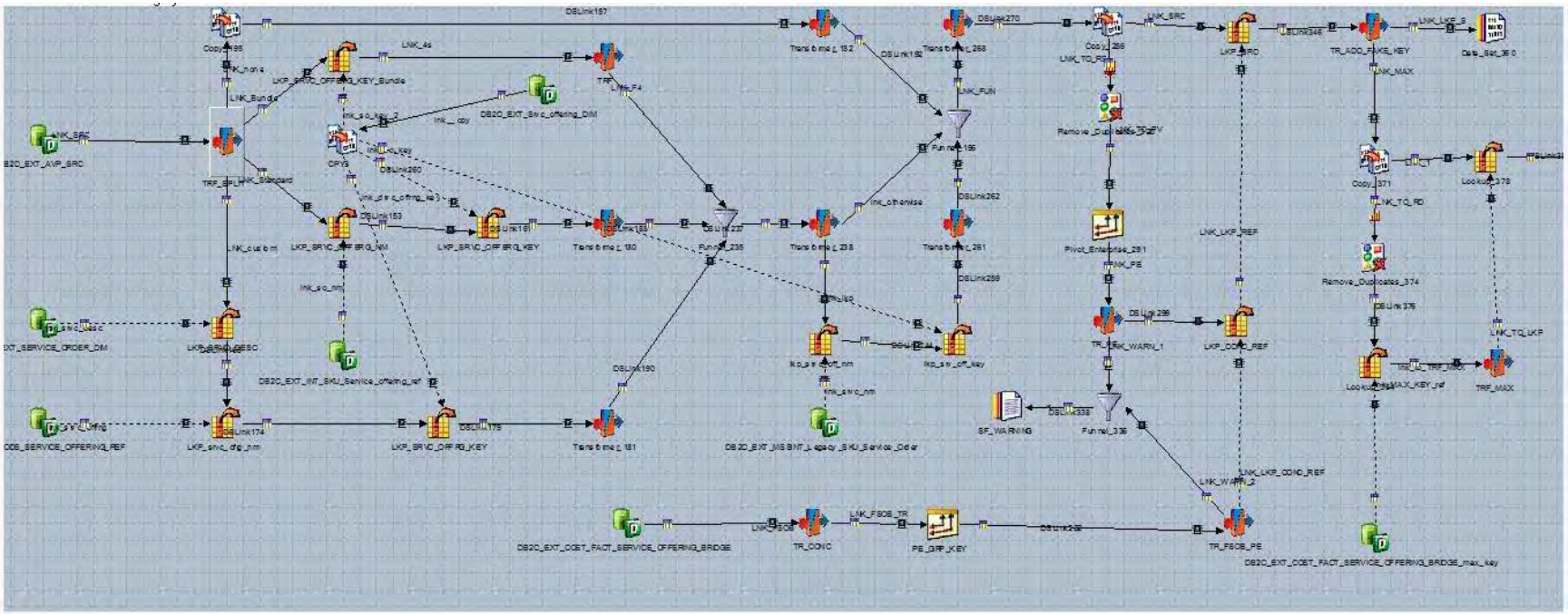


- Работа с используемыми в проекте средствами ETL
- Отличное владение SQL
- Хорошее понимание принципов оптимизации производительности запросов
- Уверенная работа на уровне пользователя с используемыми операционными системами
- Знание теории моделирования баз данных и умение читать модели данных
- Ведение документации программного продукта

ETL Job Layout 1



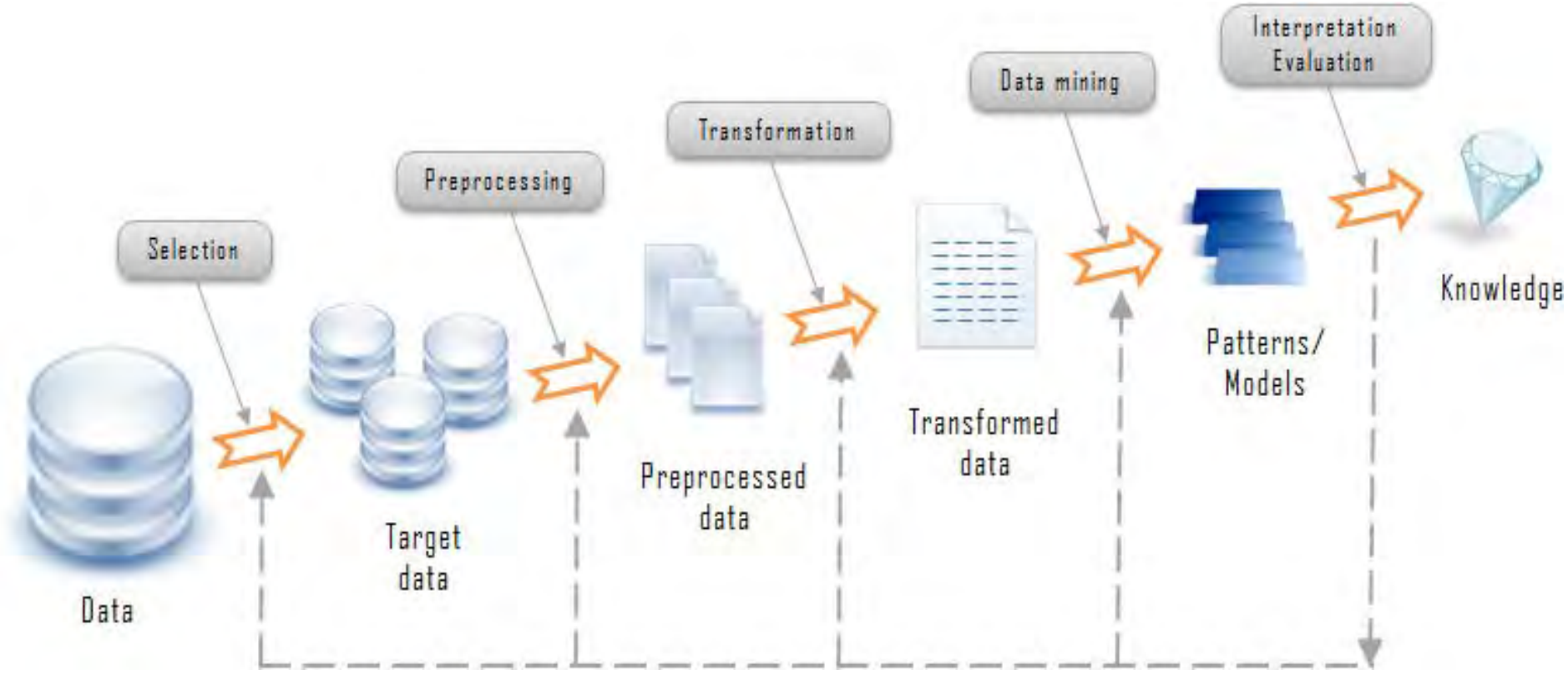
ETL Job Layout 2



- Установка и конфигурирование используемых в проекте СУБД
- Развертывание базы данных и постоянное обновление согласно требований архитектора данных
- Знание в области администрирования используемых операционных систем
- Навыки в написании шелл-скриптов
- Умение работать со средствами моделирования данных
- Ведение глоссария данных
- Поддержание в рабочем состоянии политик контроля доступа
- Мониторинг баз данных и оперативное реагирование на проблемы с доступностью
- Глубокое понимание оптимизации производительности

Сложности BI проектов

- Низкое качество входных данных и нежелание заказчика менять ситуацию
- Недостаток информации на начальных этапах разработки о дальнейшем развитии проекта, не позволяющий архитектору планировать дизайн наперед.
- Балансирование между простотой отчетов и универсальностью моделей
- Сложности, вызванные несовершенством Star Schema дизайна в целом (уход от нормализации)
- Большое количество источников данных, каждый из которых время от времени имеет проблемы с доступностью, производительностью, качеством
- Определение аттрибутов, чью именно историю хранить.
- Физические ограничения по возможности хранения исторических данных
- Сложность организации инкрементальной загрузки из источников, не имеющих аудит полей
- Разнообразие интерфейсов с базами источниками


Selection
Preprocessing
Transformation
Data mining
Interpretation
Evaluation
Data
Target
data
Preprocessed
data
Transformed
data
Patterns/
Models
Knowledge


BIG
DATA
hadoop

Data Science vs. Business Intelligence

	Business Intelligence (BI)	Data Science
Data analysis	Yes	Yes
Statistics	Yes	Yes
Visualization	Yes	Yes
Data Sources	Usually SQL, often Data Warehouse	Less structured (logs, cloud data, SQL, noSQL, text)
Tools	Statistics, Visualization	Statistics, Machine Learning, Graph Analysis, NLP
Focus	Present and past	Future
Method	Analytic	Scientific
Goal	Better strategic decisions	Advanced functionality

The two fields are closely related. In some ways Data Science is an evolution of BI.

Рекоммендации

- Ralph Kimball - The Data Warehouse Toolkit
- Ralph Kimball – The Data Warehouse ETL Toolkit