

ORACLE®

Интеграция данных в реальном времени

Александр Рындин
Старший консультант



Agenda

- Тенденции в развитии BI
- Решения Oracle GoldenGate для Real-Time BI с примерами клиентов
- Решение Oracle Complex Event Processing для Real-Time аналитики
- Вопросы&Ответы

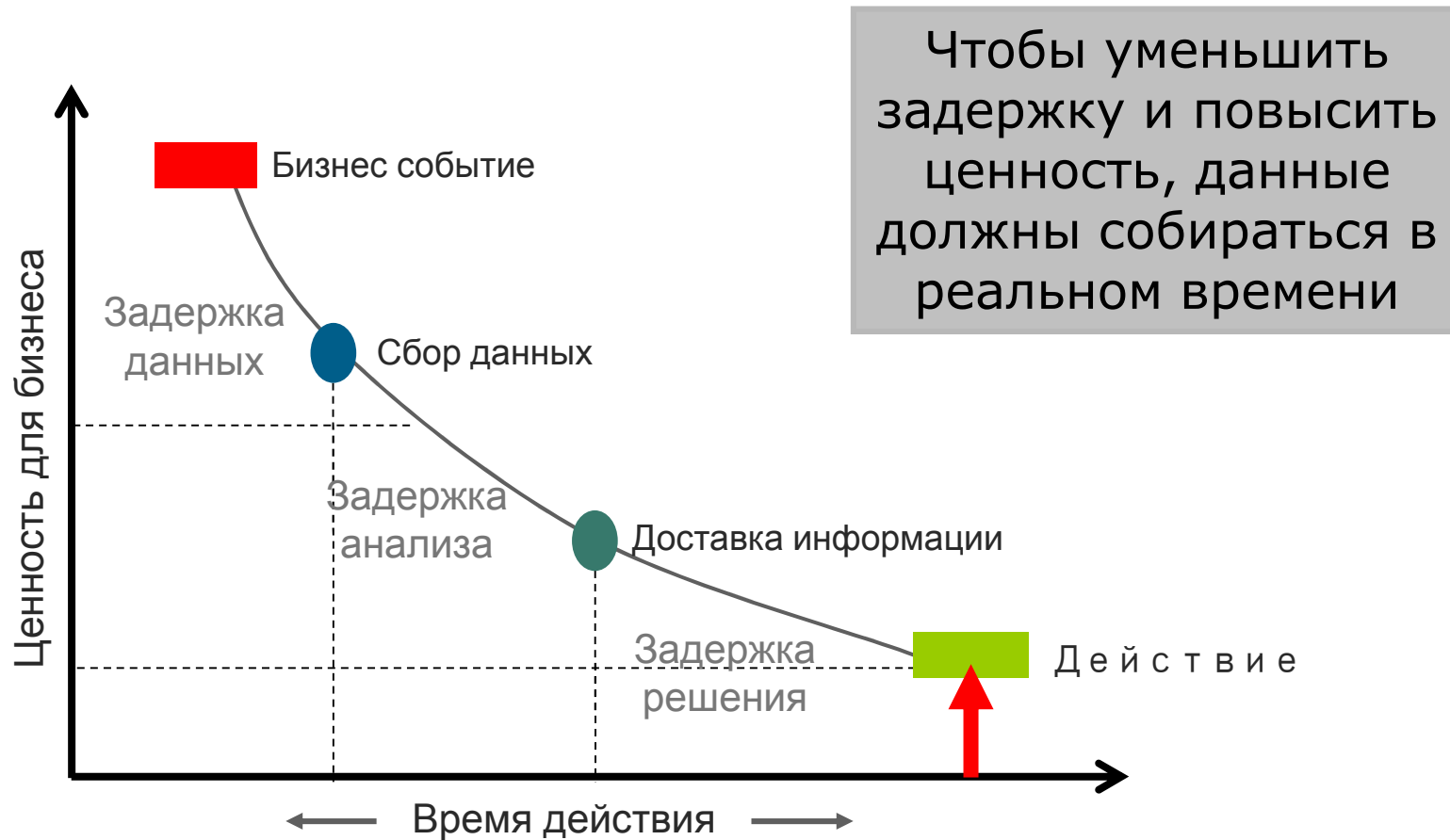


Real-Time – один из основных трендов BI

- Начиная с 2011 линия между BI и **Complex Event Processing (CEP)** будет стираться, а BI-вендоры начнут предлагать эти решения совместно¹
- 59% опрошенных респондентов сказали, что уже используют или планируют использовать **CEP** в ближайшие 12 месяцев²
- Постоянный рост интереса к аналитике в реальном времени. Это на CEP, но более mainstreams направление (Real-Time DWH)¹
- Сокращение разрыва между OLTP и хранилищами данных. Традиционные BI начинают использоваться в интеллектуальном принятии решений²

Почему реальное время?

Увеличиваем ценность корпоративной информации





Операционные BI реального времени в действии

- Обнаружение мошенничества (Fraud detection)
- Удержание клиентов (Customer churn)
- Оптимизации рекламы (Online promotion optimization)
- Оптимизация работы с клиентом в реальном времени
- Улучшение цепочек поставки



Agenda

- Тенденции в развитии BI
- Решения Oracle GoldenGate для Real-Time BI с примерами клиентов
- Решение Oracle Complex Event Processing для Real-Time аналитики
- Вопросы&Ответы

GoldenGate – стратегическое решение для Oracle

Лучшее в своем классе решение



- Лидирующее решение в области интеграции и репликации данных в реальном времени
- Более 500 клиентов с 4,000+ внедрениями в финансовом секторе, связи, медицине, госсекторе, ритейле и коммуникациях:
 - Top 3 из 5 крупнейших коммерческих банков
 - Top 3 из 3 самых нагруженных АТМ сетей
 - Top 7 из 10 компаний финансового сектора
 - Top 4 из 5 телекоммуникационных провайдеров
 - Top 3 из 5 крупнейших продовольственных и аптечных сетей

ORACLE

Что такое Oracle GoldenGate?

Oracle GoldenGate обеспечивает сбор, маршрутизацию, преобразование и доставку транзакционных данных между **гетерогенными** средами в **реальном времени** с **минимальной дополнительной нагрузкой**

Ключевые отличия:

Производительность

Не нарушаем работу, минимальное влияние, задержки меньше секунды

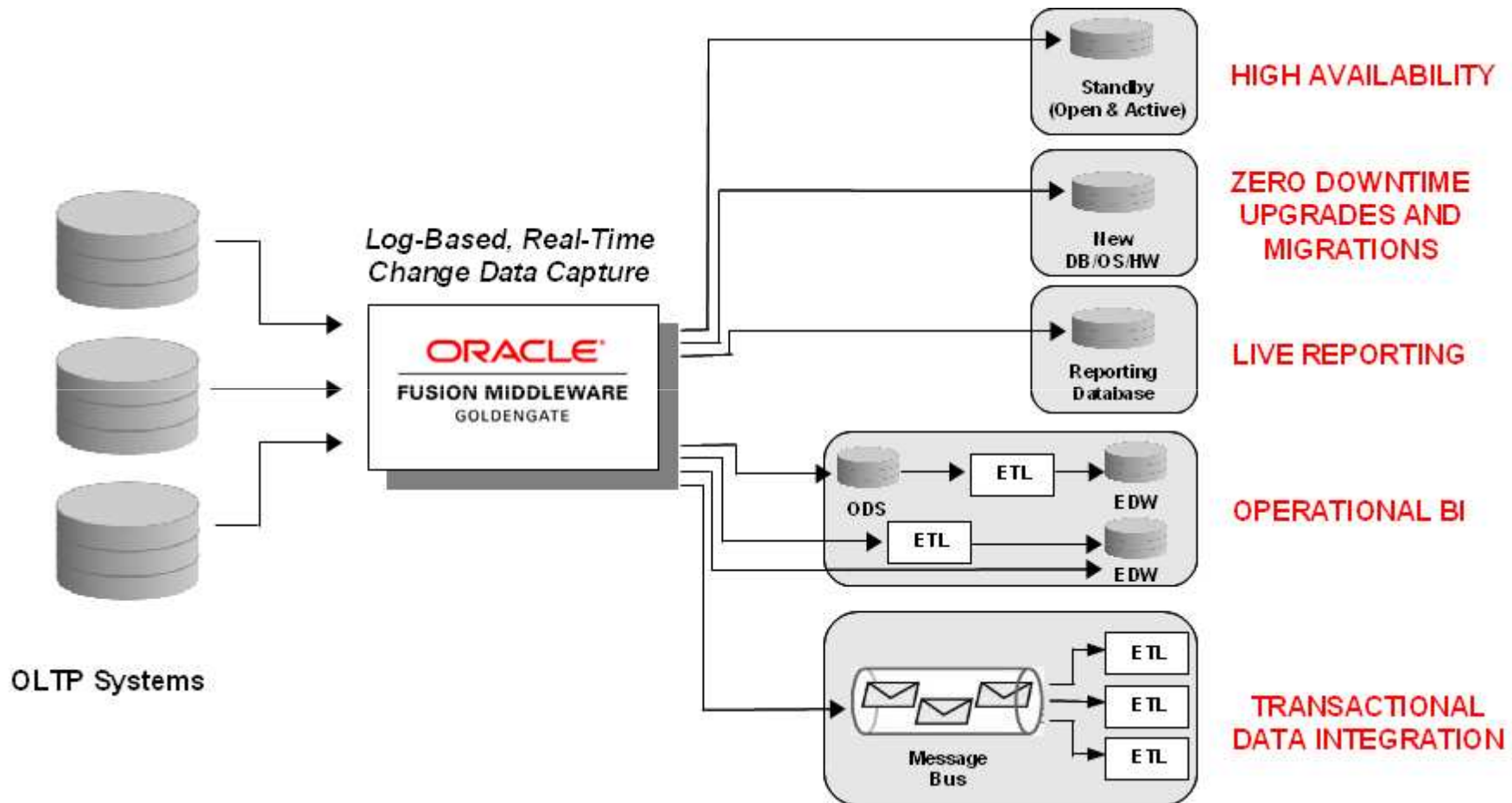
Гибкость и расширяемость

Открытая, модульная архитектура – поддержка гетерогенных источников и целей

Надежность

Транзакционная целостность – устойчиво к прерываниям работы и сбоям

Oracle GoldenGate – единая платформа



Oracle GoldenGate гетерогенная поддержка

Базы данных	O/S и платформы
<p>Сбор:</p> <ul style="list-style-type: none">▪ Oracle▪ DB2▪ Microsoft SQL Server (+2008)▪ Sybase ASE▪ Ingres▪ Teradata▪ Enscribe▪ SQL/MP▪ SQL/MX▪ MySQL▪ JMS очереди <p>Применение:</p> <ul style="list-style-type: none">▪ Вышеперечисленные плюс:▪ TimesTen, Netezza, Greenplum, HP▪ ETL продукты▪ JMS очереди	<p>Windows 2000, 2003, XP</p> <p>Linux</p> <p>Sun Solaris</p> <p>HP NonStop</p> <p>HP-UX</p> <p>HP TRU64</p> <p>HP OpenVMS</p> <p>IBM AIX</p> <p>IBM z/OS</p>

Как работает Oracle GoldenGate

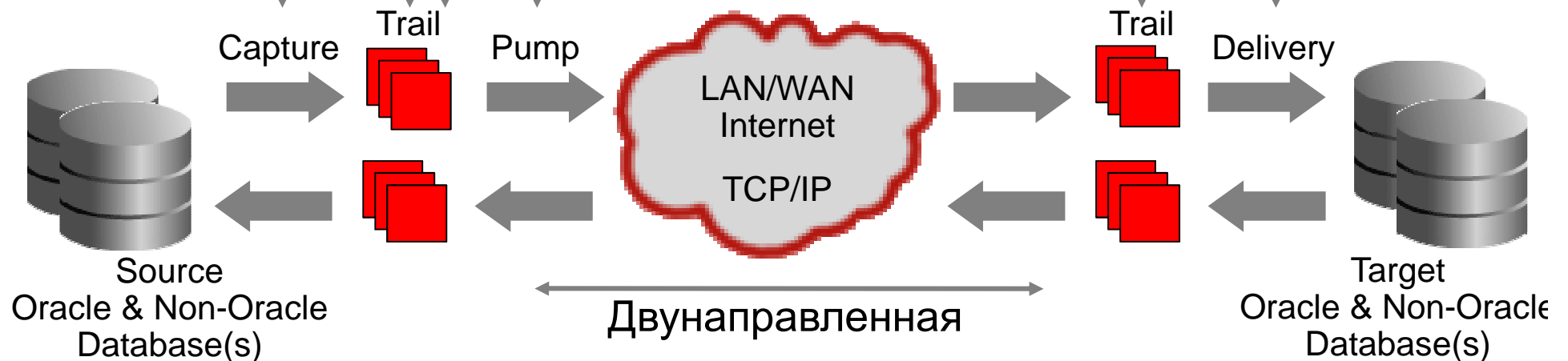
Capture: зафиксированные транзакции захватываются (фильтруются) по мере их появления в журналах (online, arc).

Trail: записываем данные для маршрутизации.

Pump: передает данные на целевую систему

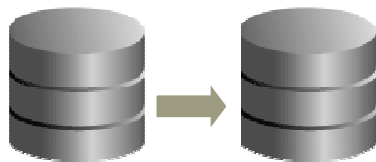
Маршрутизация: данные могут быть сжаты и зашифрованы при передаче

Delivery: применяет данные с транзакционной целостностью

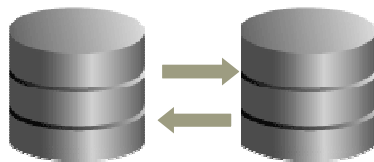


Топологии Oracle GoldenGate

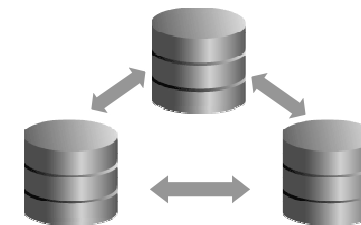
Однонаправленная
Перенос нагрузки
на чтение



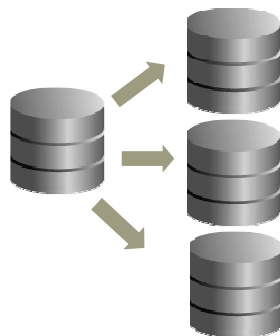
Двунаправленная
Live Standby или
Active-Active for HA



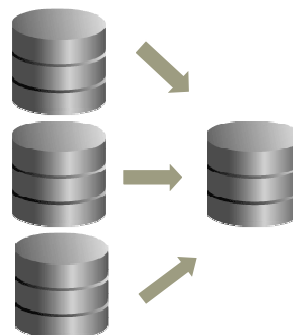
Peer-to-Peer
Балансировка нагрузки с
несколькими главными



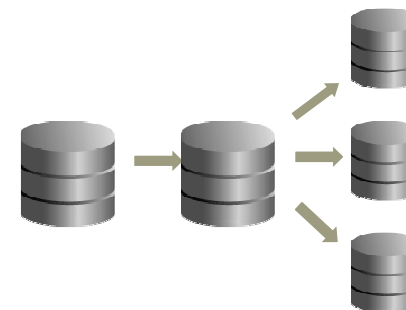
Трансляция
Распространение
данных



Интеграция/консолидация
Хранилище данных



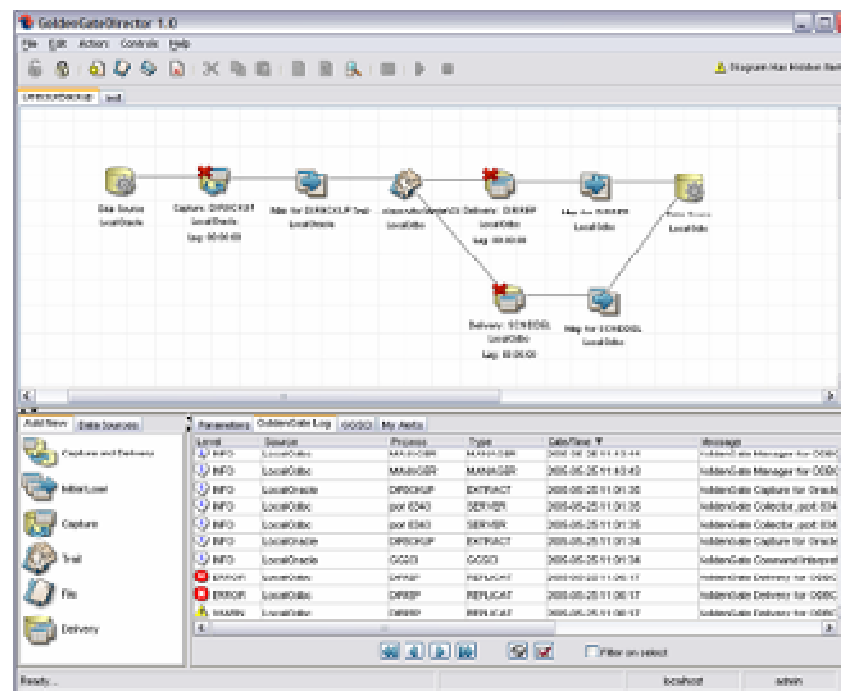
Каскадирование
Data Marts



Oracle Management Pack for Oracle GoldenGate

Простое управление решением Oracle GoldenGate

- Управляет, формирует, конфигурирует и отслеживает компоненты Oracle GoldenGate
- Ключевые особенности:
 - Централизация управления модулями GoldenGate
 - Rich-client и Web-based интерфейс
 - Отслеживает работу в реальном времени
 - Не влияет на работу

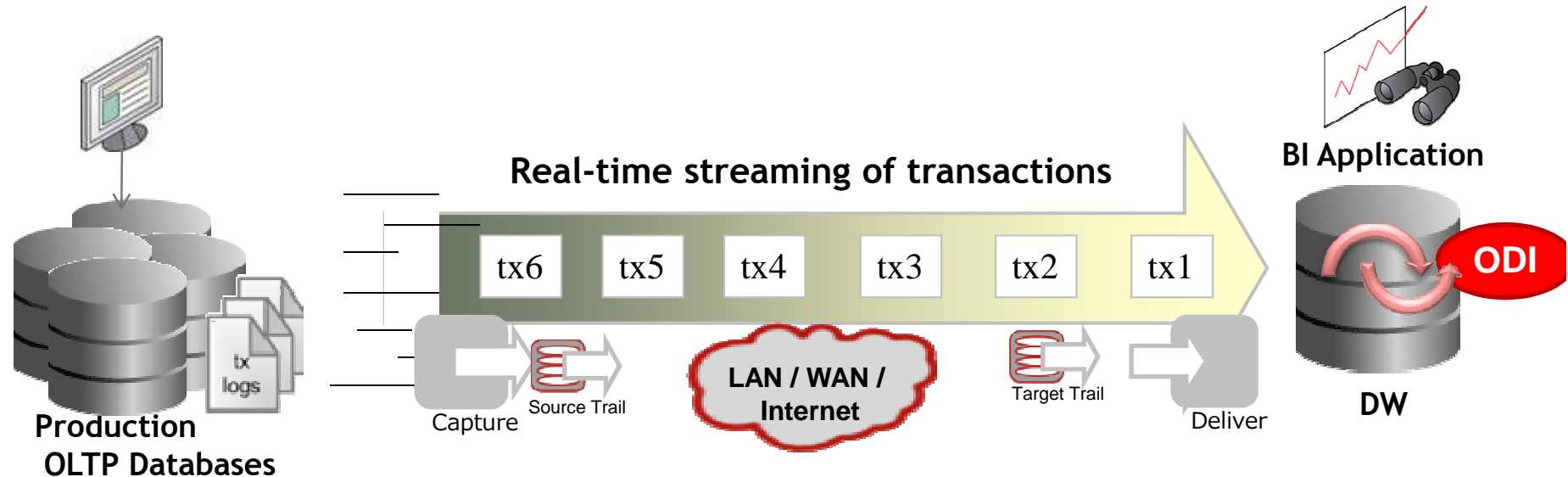


Comparison of Data Acquisition Methods

Attribute	Batch Scripts	SQL Query	Database Triggers	Oracle GoldenGate
Latency	Hours	Hours-Minutes	Minutes	Sub-seconds
Data Volume	Medium	Low	Low	High
Processing Overhead	Very High 20-50+%	High 15-20%	Medium-High 10-15%	Low 1-3%
Transaction Integrity	No	No	No	Guaranteed
Transformations	Intermediate	Basic	Basic	Basic - supplemented with ODI EE
Implementation Complexity & Risks	Medium	Medium to High	High	Low
Data Recoverability	Medium	Poor	Poor	Full
Management Overhead	Medium	Medium	High	Low

Oracle для хранилищ данных в реальном времени

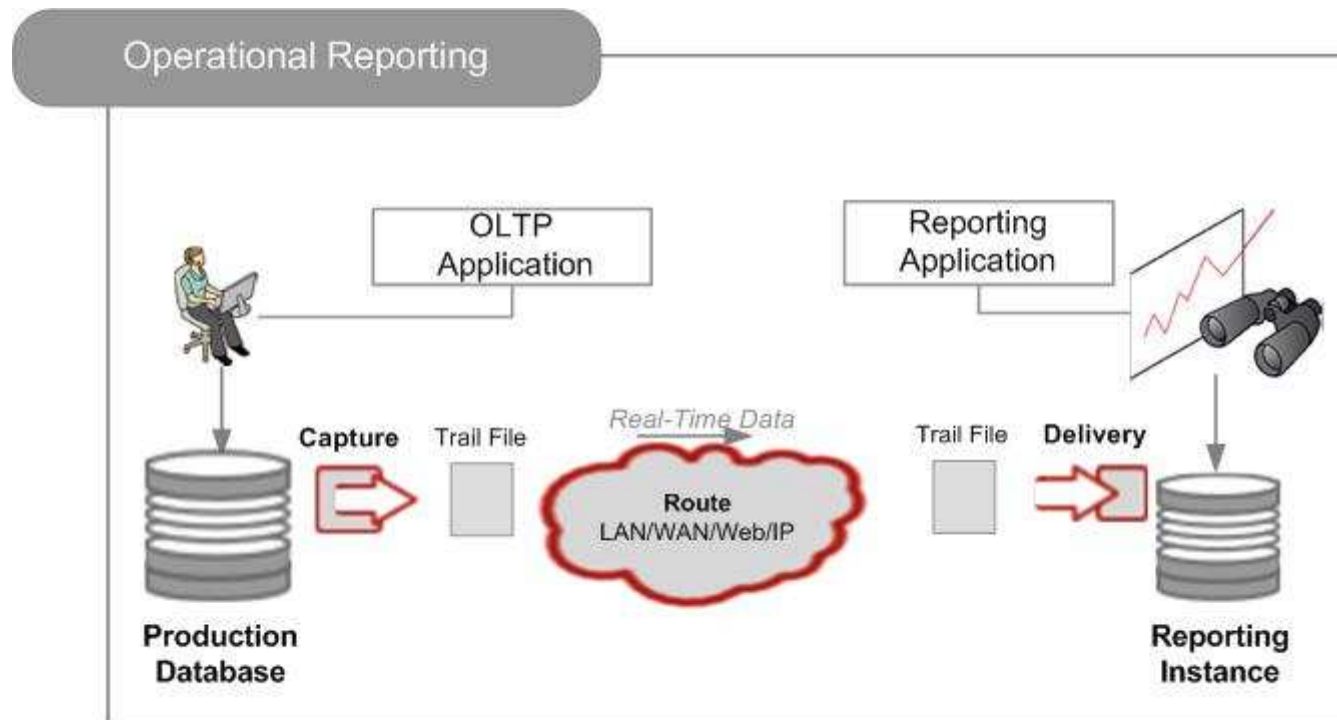
Высочайшая скорость интеграции данных в реальном времени,
Высочайшая скорость пакетной трансформации



- Задержка данных менее секунды
- Минимальная доп. нагрузка, нет пакетного окна
- Высокая скорость, преобразования в базе данных
- Согласованные на чтения данные, ссылочная целостность
- Полная восстановимость данных с помощью Trail файлов

Oracle GoldenGate для операционной отчетности

Перенесите отчетность на более дешевую отчетную базу



- **Задержка менее секунды**
- **Не снижает производительность исходной системы**
- **Консистентные данные со ссылочной целостностью**
- **Полная восстановимость данных через Trail файлы**

Пример клиента: Research in Motion (RIM)

Глобальная интеграция данных и высокая доступность



Oracle GoldenGate

Oracle RAC

Oracle Exadata

- Каждая фабрика должна приспособливаться к производству новых устройств и планировать загрузку в зависимости от производства на других фабриках
- Фабричные системы должны быть доступны в режиме 24x7, чтобы работать на глобальном рынке

Решение

- Перенос данных о производстве во многих направлениях менее чем за секунду
- Позволяет транзакциям приложения быть распределенным по нескольким базам данных
- Поддержка 24*7*365 глобального процесса с точными и надежными данными
- Перенос операционной отчетности на Oracle Exadata для повышения производительности OLTP

Oracle GoldenGate 11g – JMS как транспорт

Capture: зафиксированные транзакции захватываются (фильтруются) по мере их появления в журналах (online, arc).

Trail: записываем данные для маршрутизации.

Pump: передает данные на целевую систему

Маршрутизация: данные могут быть сжаты и зашифрованы при передаче

Delivery: применяет данные с транзакционной целостностью





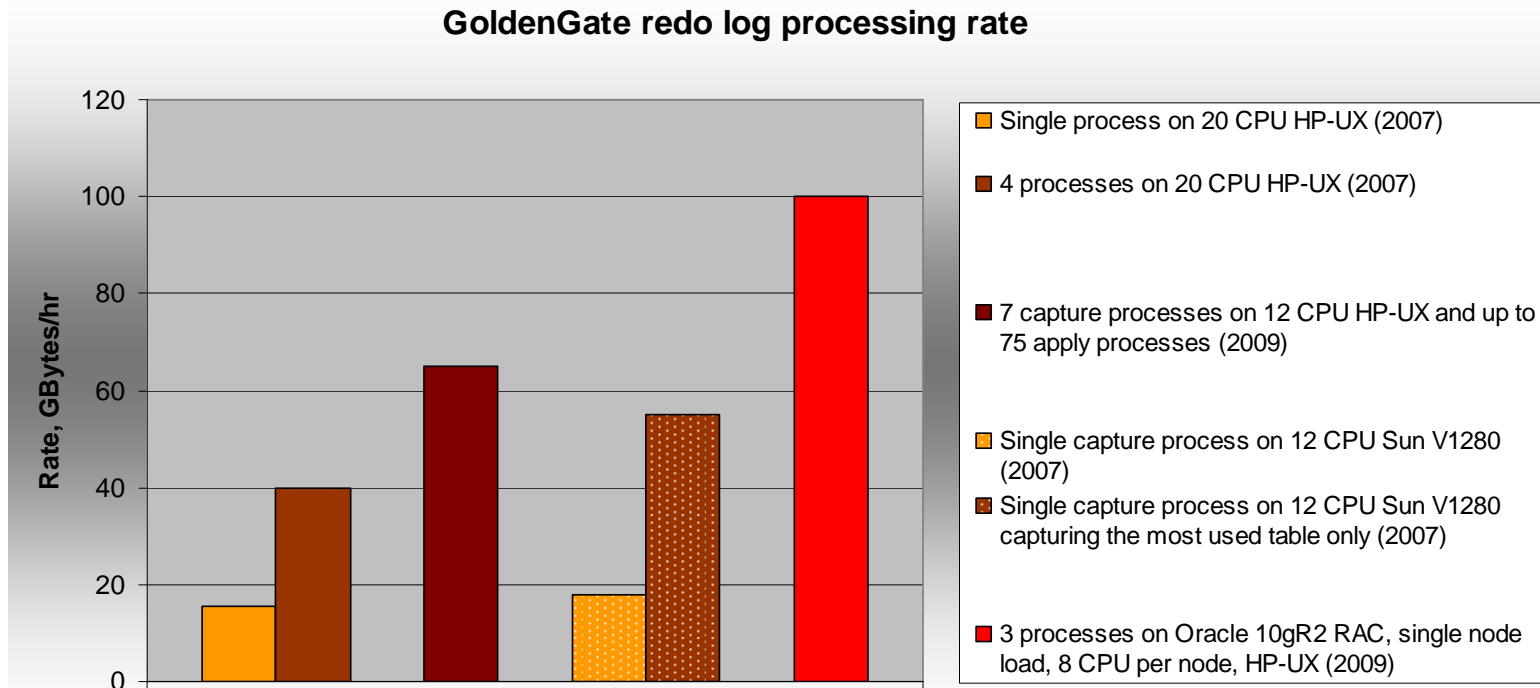
Опыт проектов

- Использовать можно синтетическую нагрузку, но лучше реальную.
- Тестируйте все – insert, delete, update, ddl
- Скачивайте последнюю версию – иногда производительность выше 1,5-2 раза
- Нагрузка 50 Гб/час для Extract на источнике без настройки. Не распараллеливайте Extract по Range
- Используйте сжатие при передаче по сети
- Используйте Client Result Cache
- Отключайте во время внедрения PURGE*TRAIL
- Во время ПОС тестируйте не только производительность, но и первоначальную инициацию

End to End Low Latency High Throughput

Year 2009:

- Major telco: ~65 GB/hr peak load end to end
- Major trading system: ~100 Gb/hr redo processing end to end





Agenda

- Тенденции в развитии BI
- Решения Oracle GoldenGate для Real-Time BI с примерами клиентов
- Решение Oracle Complex Event Processing для Real-Time аналитики
- Вопросы&Ответы

Streaming Event-Driven Architecture (EDA) Solutions

Легкий, с низкой задержкой, высочайшая пропускная способность, Java-платформа

<http://www.oracle.com/technology/products/event-driven-architecture/complex-event-processing.html>


Oracle Complex Event Processing Производительность & масштабируемость

► Обогащение потоков

- Из любой источника: data streams, web services, Java, Database

► Адаптеры

- Преобразуют внешние события/данные в java-объекты для обработки



1 Million Events Processed per Second

- <1 Millisecond Latency for Events – Avg 69 μ s
- 100s of Rules & Event Queries
- Events from Java, Queues, DBMS, SOA
- Oracle Event Server 10gR3
- Oracle Enterprise Linux 5.0
- 4X2.93GHz Quad Core Processors
- 6 GB RAM Per Server

► Процессоры

- Набор запросов, применяемых к потоку

► Прослушиватели

- Срабатывают по результатам работы процессоров

► События

- Реализованы как JavaBean или Map

Агрегация, корреляция, фильтрация данных
Может обрабатывать «бесконечные» запросы

Входящие потоки данных
1,000,000+ Messages/Sec

Задержка
Микросекундная задержка (среднее.)

Результирующий поток данных, реализуемый с помощью Java

ORACLE

Обработчик Событий

Oracle Complex Event Processing (CEP)

BA	BOEING	D	77.075	41.575	20080305 10:03:02.78
DD	DUPONT	D	41.575	3000	20080305 10:03:04.12
BA	BOEING	D	77.075	800	20080305 10:03:05.05
C	CITIGROUP	D	34.125	2000	20080305 10:03:05.05
BA	BOEING	D	77.075	800	20080305 10:03:05.78
L	UNITGROUP	M	34.125	2000	20080305 10:03:05.99
CAT	CATERPILLAR	D	22.5	600	20080305 10:03:05.46
DD	DUPONT	D	41.575	3000	20080305 10:03:04.12
AA	ALCOA INC	D	20.125	1000	20080305 10:03:01.05
ASP	AMER EXPRESS CO	D	45.875	500	20080305 10:03:02.10
BA	BOEING	D	77.075	800	20080305 10:03:02.78
C	CITIGROUP	D	34.125	2000	20080305 10:03:05.05
CAT	CATERPILLAR	D	22.5	600	20080305 10:03:05.46
DD	DUPONT	D	41.575	3000	20080305 10:03:04.12
AA	ALCOA INC	D	20.125	1000	20080305 10:03:01.05
ASP	AMER EXPRESS CO	D	45.875	500	20080305 10:03:02.10
BA	BOEING	D	77.075	800	20080305 10:03:02.78
C	CITIGROUP	D	34.125	2000	20080305 10:03:05.05
CAT	CATERPILLAR	D	22.5	600	20080305 10:03:05.46
DD	DUPONT	D	41.575	3000	20080305 10:03:04.12
AA	ALCOA INC	D	20.125	1000	20080305 10:03:01.05
ASP	AMER EXPRESS CO	D	45.875	500	20080305 10:03:02.10
BA	BOEING	D	77.075	800	20080305 10:03:02.78

➤ Поток

- Постоянные источники сообщений, часто с высоким объемом
- Упорядочены по времени
- Бесконечные
- *Невозможно обрабатывать/анализировать в реальном времени с помощью традиционных реляционных СУБД*

Пример: Необработанный поток событий с сенсоров, GPS, Биржевые Поток данных

Обработка событий обеспечивает новую инфраструктуру для поддержки и анализа Поток в реальном времени

In-Memory, Continuous Queries

Oracle Complex Event Processing (CEP)



➤ Результат Обработки Событий

➤ Фильтрация

- Новый поток, отфильтрованный по заданному критерию, например, стоимость акций > \$22

➤ Корреляция & Агрегирование

- Прокручиваемые метрики, построенные по временному окну, среднее # операций с акциями за последний час

➤ Сопоставление с Шаблоном

- Уведомление об обнаруженном событийном шаблоне W, например, стоимость изменялась X, Y и Z в течение 15-минутного окна

⋮

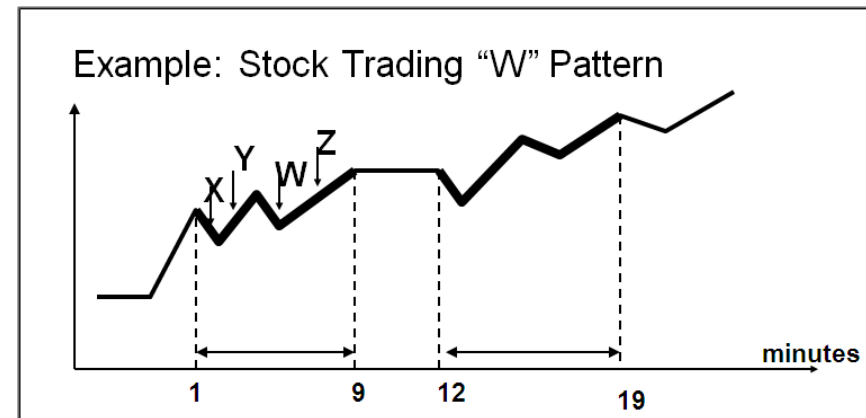
DD	DUPONT	D	41.575	3000	20080305 10:03:04.12
AA	ALCOA INC	U	20.125	1000	20080305 10:03:01.00
APP	AMER EXPRESS CO	D	45.875	500	20080305 10:03:02.10
DD	DUPONT	D	41.575	3000	20080305 10:03:04.12
AA	ALCOA INC	D	20.125	1000	20080305 10:03:01.00
APP	AMER EXPRESS CO	D	45.875	500	20080305 10:03:02.10
BA	BACARD	D	77.575	800	20080305 10:03:02.78

COMPLEX QUERIES

⋮

```

<view id="SubstationSummaryView">
  <![CDATA[
    ISTREAM {
      SELECT
        houseEvents.substation,
        MAX(date) AS maxDate,
        COUNT(*) AS eventCount,
        COUNT(DISTINCT name) AS distinctSourceCount,
        COUNT(isSmartRate) AS smartRateCount,
        COUNT(isNotSmartRate) AS nonSmartRateCount,
        AVG(houseEvents.percentnormkwh) AS averagepercentnorm,
        SUM(houseEvents.kwh) AS totalkwh
      FROM
        ExpandSmartMeterView [RANGE 1 SECONDS SLIDE 500 MILLISECONDS] AS houseEvents
      GROUP BY houseEvents.substation
    }
  ]>
</view>
  
```

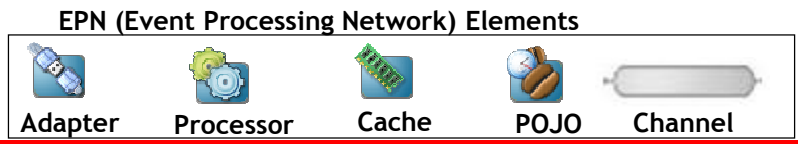
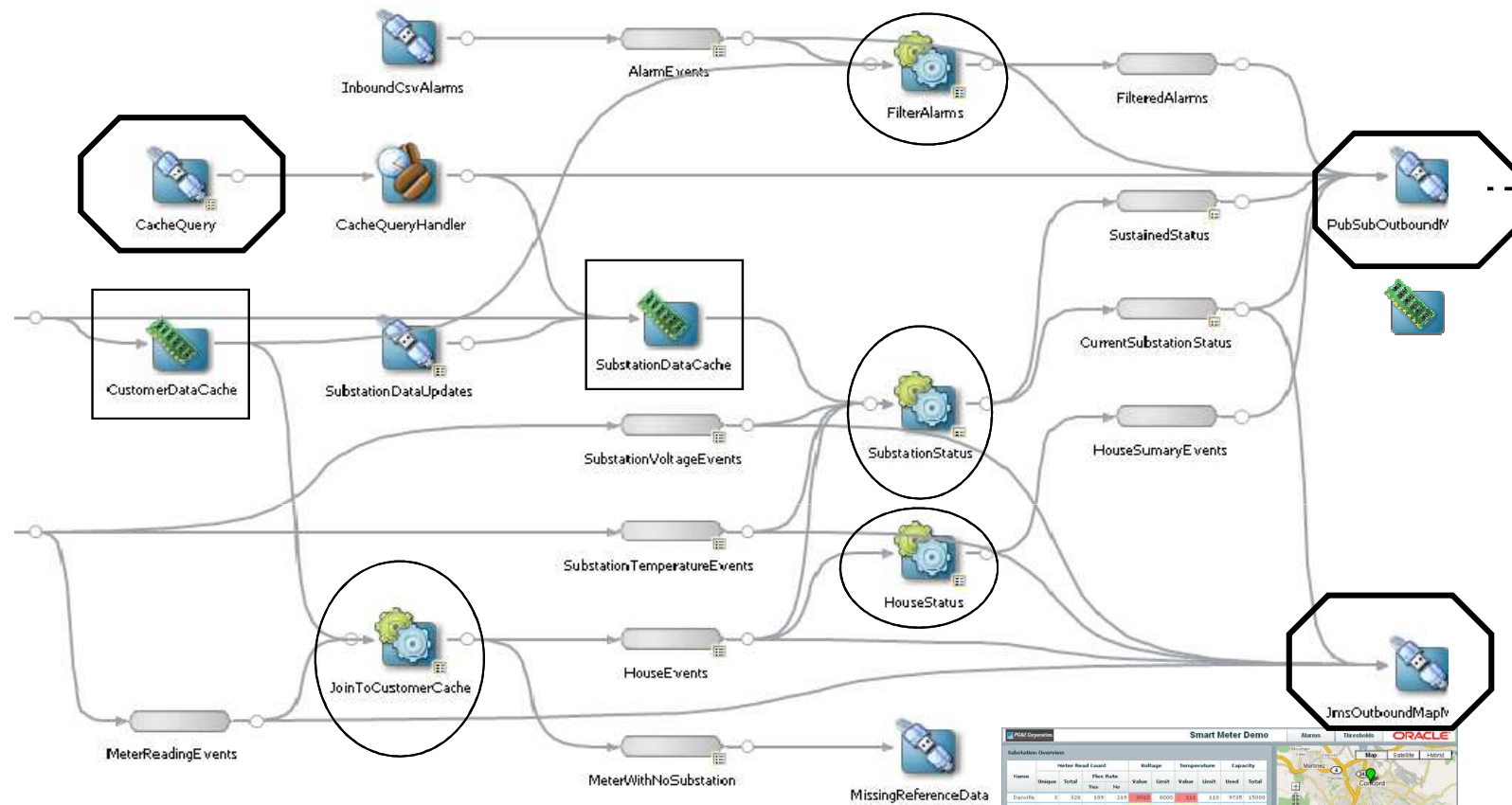


- Работает В Памяти (не в базе данных)
- Логика определяется с помощью Continuous Queries к данным

EDA Application Model Architecture

Эффективная обработка потока данных

- Поток данных -



Names	Usage	Total	Flow Rate	Value	Unit	Temperature	Capacity
Demofile	0	329	159	217	3000	111	110
Control	2	220	226	9	7500	7000	93

JSON

HTTP Pub/Sub

CQL Example – Join To Cache

Join the incoming meter read stream to the cache containing customer/house data. This adds contextual data (e.g. substation and baseline values) to the event.

We use an outer join to be able to route events with no matching meter to a resolution process.

```
-->
<view id="JoinToCustomerCacheView">
  <![CDATA[
    RSTREAM(
      SELECT
        meter.date,
        meter.house AS name,
        meter.kwh,
        customer.address,
        customer.normkwh,
        customer.smartrate,
        customer.substation
      FROM
        MeterReadingEvents [NOW] AS meter,
        CustomerDataCache AS customer
      WHERE
        meter.house = customer.name (+)
    )
  ]]>
</view>
```

[NOW] is a Stream to Relation Operator



CQL Example - Aggregation

```
<!--  
We output a stream here because we need to join the resulting summary values to cache  
to get contextual information about the substation.  
-->  
<view id="SubstationSummaryView">  
  <![CDATA[  
    ISTREAM (  
      SELECT  
        houseEvents.substation,  
        MAX(date) AS maxDate,  
        COUNT(*) AS eventCount,  
        COUNT(DISTINCT name) AS distinctSourceCount,  
        COUNT(isSmartRate) AS smartRateCount,  
        COUNT(isNotSmartRate) AS nonSmartRateCount,  
        AVG(houseEvents.percentnormkwh) AS averagepercentnorm,  
        SUM(houseEvents.kwh) AS totalkwh  
      FROM  
        ExpandSmartMeterView [RANGE 1 SECONDS SLIDE 500 MILLISECONDS] AS houseEvents  
      GROUP BY houseEvents.substation  
    )  
  ]]  
</view>
```

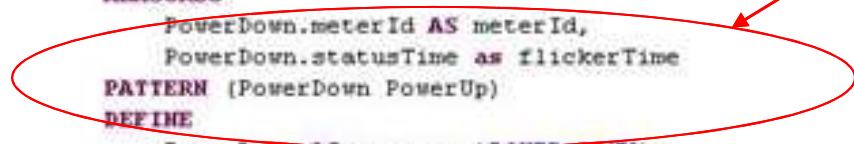
**Calculates aggregate values
over a *sliding* 1 second
window with output every half second**



CQL Example – Pattern Matching

```
<?xml version="1.0" encoding="UTF-8"?>
<wlevs:config xmlns:wlevs="http://www.bea.com/ns/wlevs/config/application"
  xmlns:jdbc="http://www.oracle.com/ns/ocep/config/jdbc">
  <processor>
    <name>MeterFlickerProcessor</name>
    <rules>
      <query id="FlickerQuery">
        <![CDATA[
          SELECT
            flicker.meterId as meterId,
            flicker.flickerTime as flickerTime
          FROM MeterStateInputChannel MATCH_RECOGNIZE
          (
            PARTITION BY
              meterId
            MEASURES
              PowerDown.meterId AS meterId,
              PowerDown.statusTime as flickerTime
            PATTERN (PowerDown PowerUp)
            DEFINE
              PowerDown AS state = 'POWER_DOWN',
              PowerUp AS state = 'POWER_UP'
          ) AS flicker
        ]]>
      </query>
    </rules>
  </processor>
  <channel>
    <name>MeterFlickerOutputChannel</name>
    <selector>FlickerQuery</selector>
  </channel>
</wlevs:config>
```

Identifies Meter “Flicker” Event Patterns – Power Up following Power Down



Oracle Business Activity Monitoring

Meaningful, Event-driven Visibility for Business Users

➤ **Monitor** business processes & services in real-time

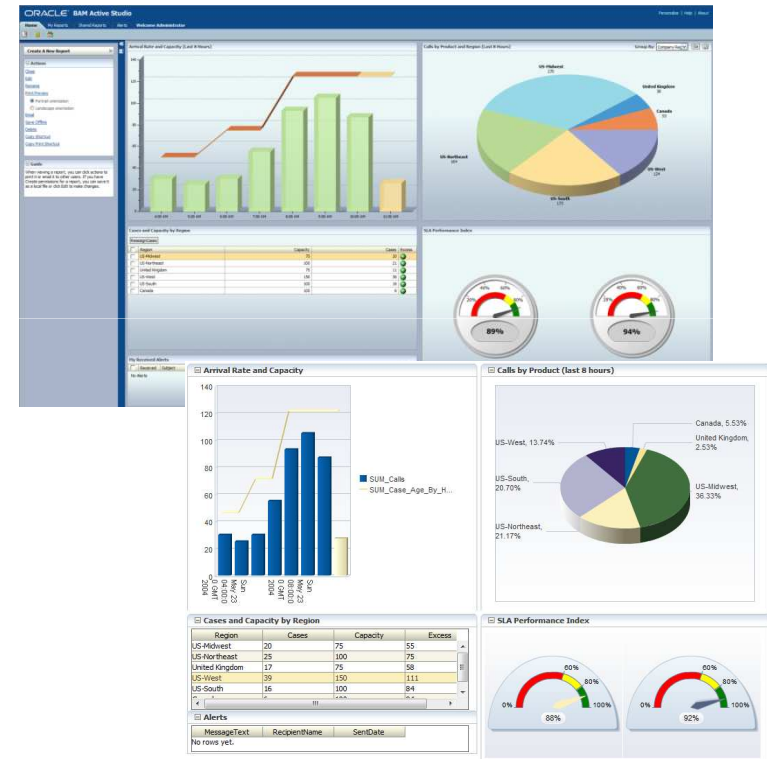
- Key Performance Indicators (KPIs)
- Service-Level Agreements (SLAs)

➤ **Analyze** events as they occur

- Correlate events & KPIs
- Identify trends as they emerge
- Alert users to bottlenecks & solutions

➤ **Act** on current conditions

- Event-driven alerts
- Real-time dashboards
- BPEL processes & web services integration



Select Report

Print Preview

Personalize

Reprint

Save Offline

Email

Substation Status: Percent of Normal Usage



Danville



Concord

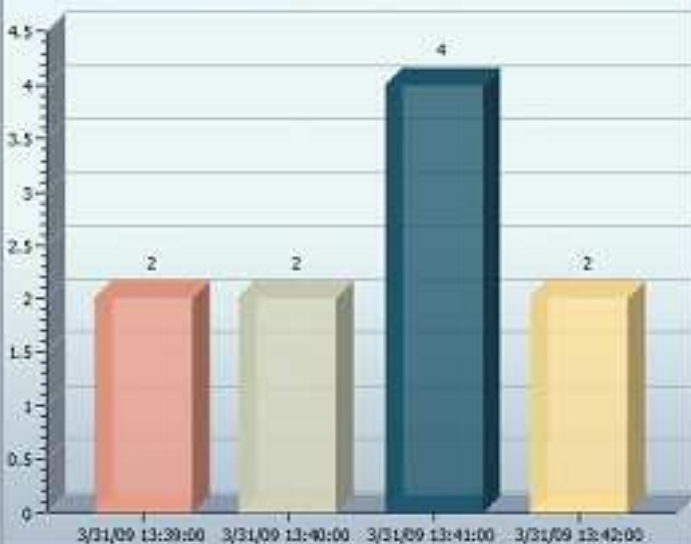
Substation Temperature Updates

2176 new transactions

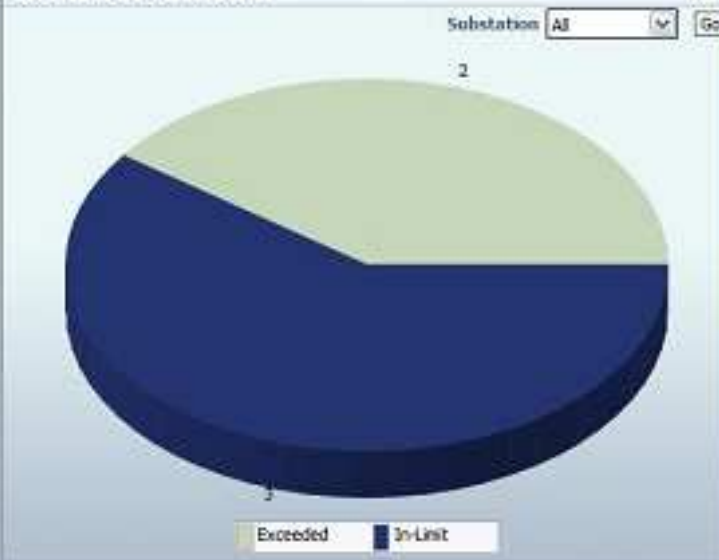
1-50 51-100 101-150

Status	Temperature	Substation	Timestamp
Fine	95	Concord	3/31/2009 1:42:09 PM
High	111	Danville	3/31/2009 1:42:09 PM
Fine	95	Concord	3/31/2009 1:42:07 PM
High	111	Danville	3/31/2009 1:42:07 PM
Fine	95	Concord	3/31/2009 1:42:05 PM
High	111	Danville	3/31/2009 1:42:05 PM
Fine	95	Concord	3/31/2009 1:42:03 PM
High	111	Danville	3/31/2009 1:42:03 PM
Fine	93	Concord	3/31/2009 1:42:02 PM
High	111	Danville	3/31/2009 1:42:01 PM
Fine	93	Concord	3/31/2009 1:42:00 PM
High	111	Danville	3/31/2009 1:42:00 PM
Fine	95	Concord	3/31/2009 1:41:58 PM
High	111	Danville	3/31/2009 1:41:58 PM
Fine	93	Concord	3/31/2009 1:41:56 PM
High	111	Danville	3/31/2009 1:41:56 PM
Fine	93	Concord	3/31/2009 1:41:54 PM

Households Consuming Over 300KWH



Power Consumption by House



Alert

<input type="checkbox"/>	3/31/09 1:38 PM	SubstationOverload : En all Substation Danville is experiencing unusually high usage. Please re-route power if necessary
<input type="checkbox"/>	3/31/09 12:03 PM	SubstationOverload : En all Substation Danville is experiencing unusually high usage. Please re-route power if necessary
<input type="checkbox"/>	3/30/09 8:58 PM	SubstationOverload : En all Substation Danville is experiencing unusually high usage. Please re-route power if necessary
<input type="checkbox"/>	3/31/09 11:51 AM	SubstationOverload : En all Substation Danville is experiencing unusually high usage. Please re-route power if necessary
<input type="checkbox"/>	3/30/09 6:21 PM	SubstationOverload : En all Substation Danville is experiencing unusually high usage. Please re-route power if necessary



Примеры использования

- Online магазины
- Охранные системы
- Энергетика (Enernoc)
- Фармацевтика
- Автомобильная телеметрия (Exadata+Exalogic)
- Автоматические торговые системы
- Телекоммуникации (Turkcell)



- Real Time обнаружение шаблонов
 - На данном этапе:
 - Кто из клиентов находится на стадионе Şükrü Saraçoğlu?
 - Кто из клиентов послал более чем 5 sms из определенного торгового центра за последние 10 минут?
 - С каких вышек за последние полчаса не было сделано ни одного звонка?
 - В будущем:
 - Кто из клиентов посещал место А, затем В и потом С в последние 15 минут?

What the GSM Network data looks like?

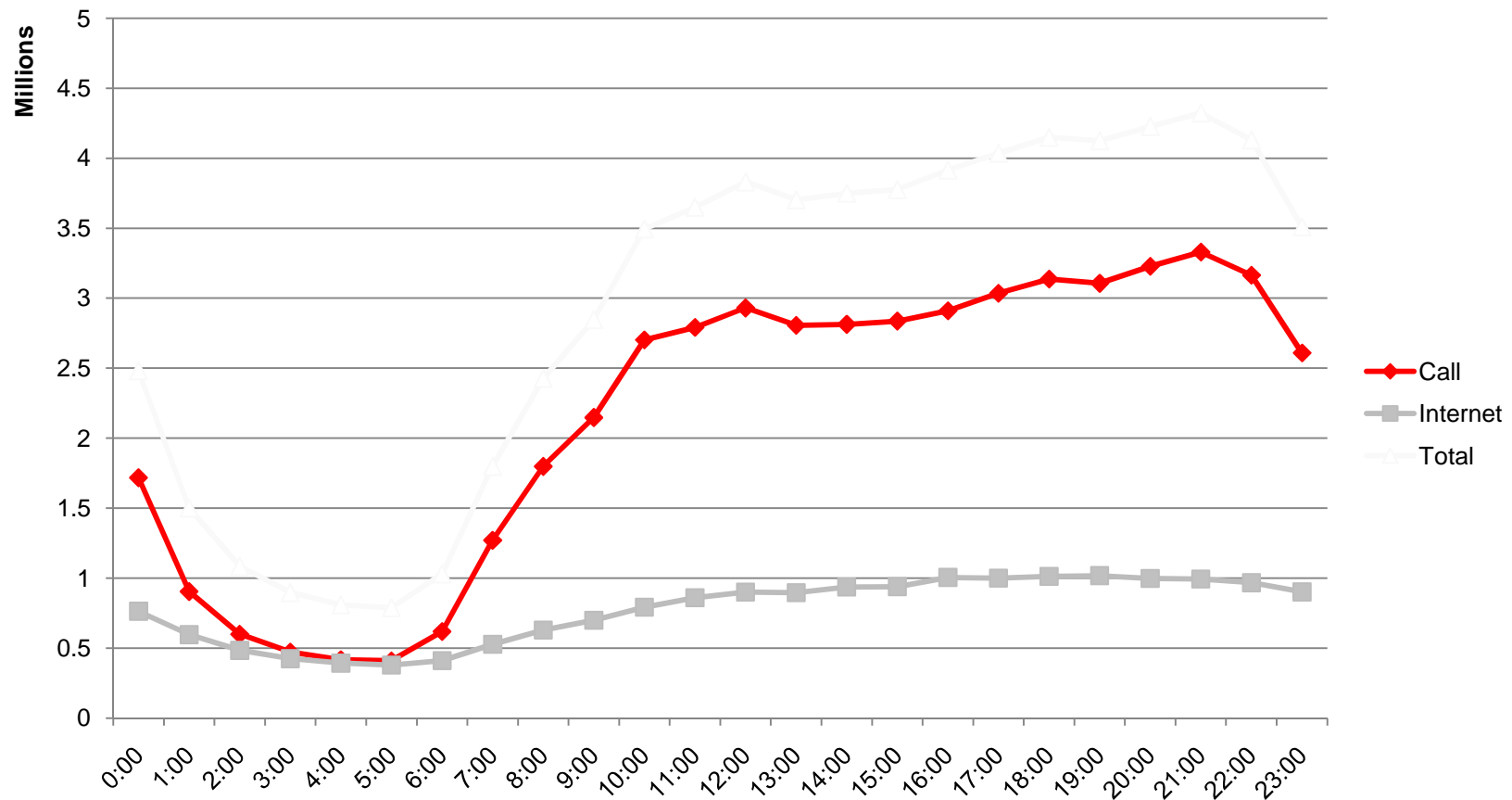


04,012c06a75479,012c06a75b60,286015248522797,3556770075057064,f69e,78da

04	→	Activation Type
012c06a75479	→	When activation started
012c06a75b60	→	When activation finished
286015248522797	→	IMSI number of who has started the activation
3556770075057064	→	IMEI number of who has started the activation
f69e,78da	→	Location information Where the activation occurs



AVG Processed MB per Second for Event Types

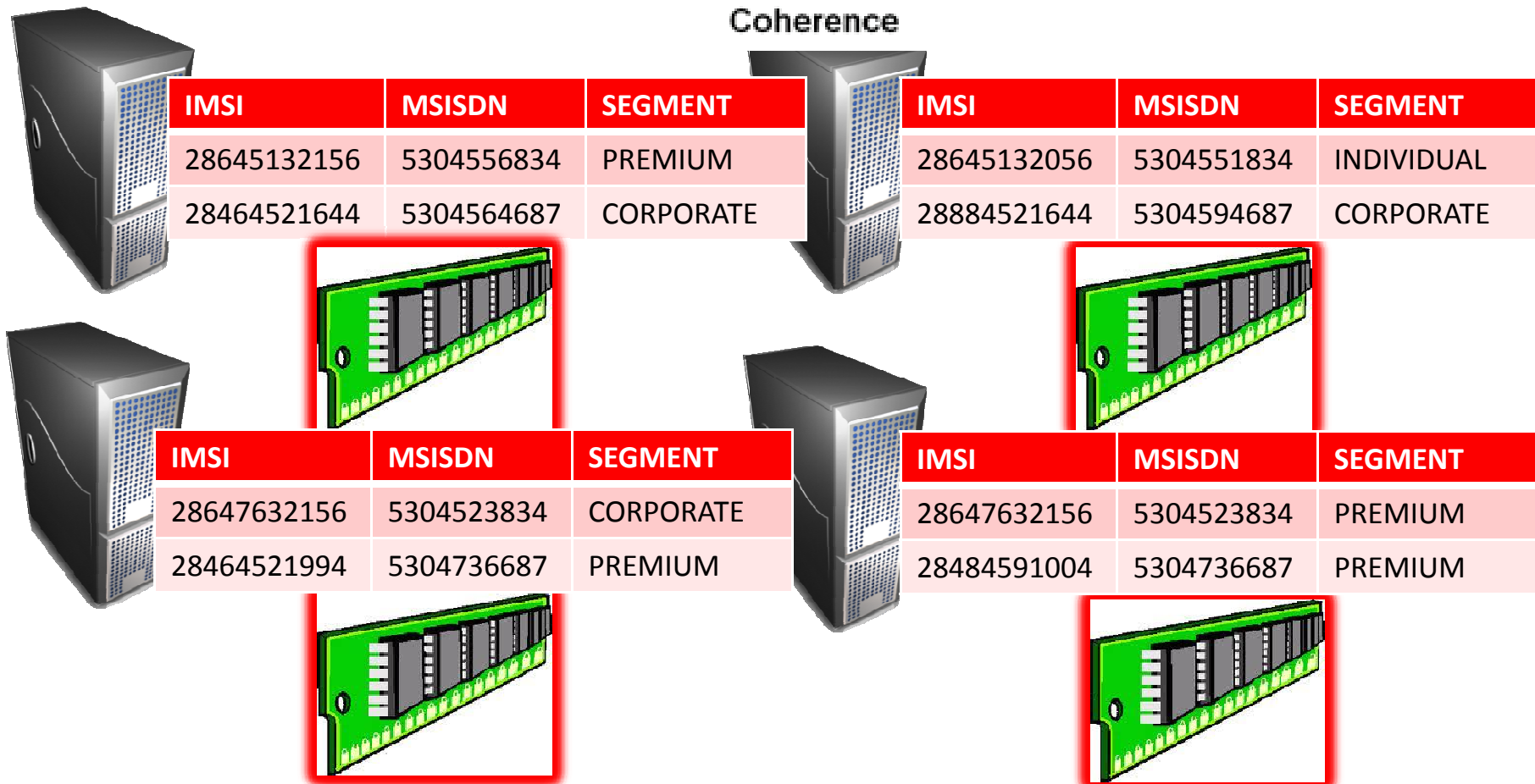


Oracle Coherence



- In the processed file there are only one information about customer, IMSI. We should access the other information

ORACLE[®] Coherence





- **Oracle CEP**
 - **Cons**
 - **Lacking of wide community**
 - **Documentation has not deep information**



- **Oracle CEP**
 - **Pros**
 - **Totally in J2EE**
 - **Supports transparently Oracle Coherence**
 - **Supports scalability and high availability**
 - **Advanced Continuous Query Language**
 - **Supports Jrockit JVM**



ALEXANDER RYNDIN

*Старший консультант по
пространственным технологиям и
технологиям хранилищ данных*

Oracle CIS

**Moscow,
Krasnopresnenskaya
nab., 18, Block C**

phone +7(495) 641-1400
fax +7(495) 641-1414
alexander.ryndin@oracle.com
<http://www.oracle.com/>
<http://www.oraclegis.com/>