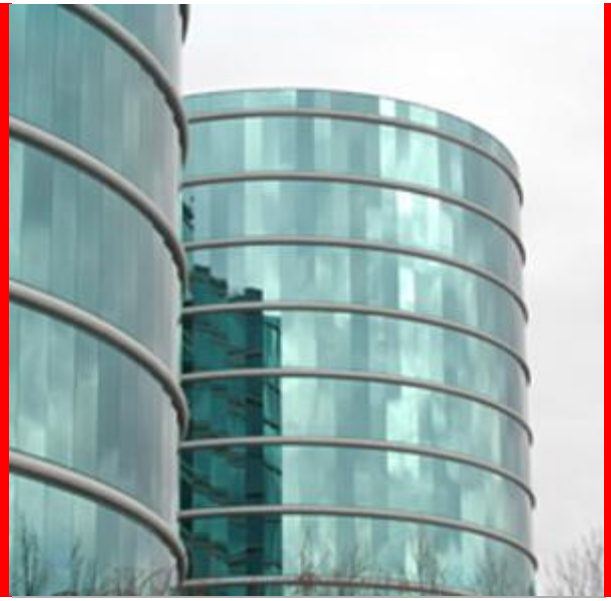


ORACLE®

Использование Oracle (Complex) Event Processing для обработки Big и Fast Data

Александр Рындин
Старший консультант

Agenda

- Что такое потоковые данные?
- Почему CEP актуален сегодня?
- Решение Oracle по обработке событий
- Примеры внедрений
- Вопросы&Ответы

Потоки событий

- События: измерение температуры, написан twit, сделан звонок
- Источники событий: датчики, бизнес-системы, социальные сети

- Непрерывный поток событий, часто большого объема
- Отсутствует конец потока
- Упорядочен по времени
- Нужно на лету уметь обнаруживать «шаблоны»
- Невозможно или не эффективно обрабатывать/анализировать в реальном времени с применением баз данных

Потоковые данные (1)

- Задача: “посчитайте все красные машины на парковке”.
- Ответ с помощью реляционной базы:
  - Обойти всю парковку.
  - Посчитать все средства передвижения, которые
    - Красные
    - Машины

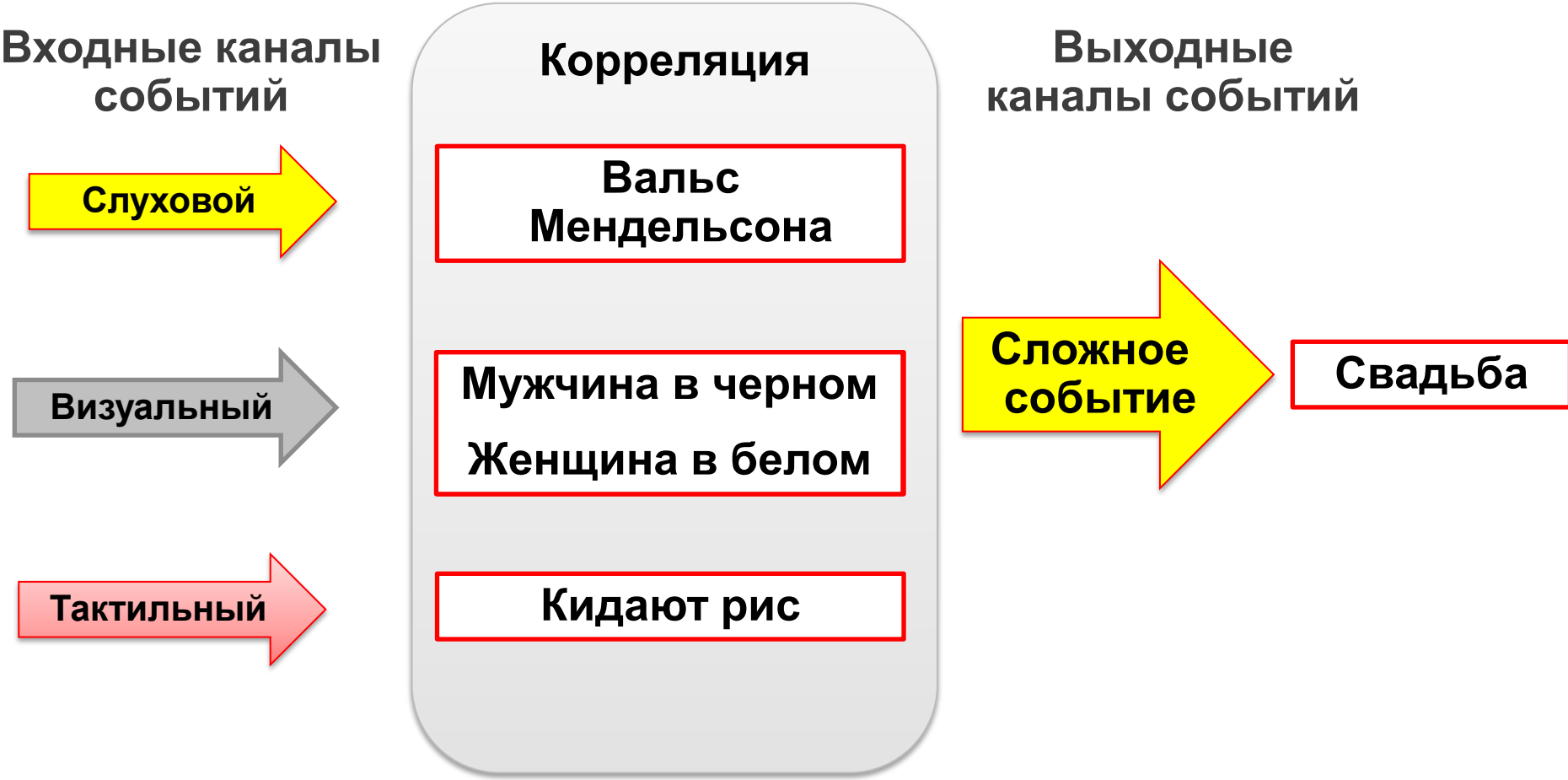
```
SELECT COUNT(*) FROM ParkingLot
WHERE type = ‘AUTO’
AND color = ‘RED’
```

Потоковые данные (2)

- Задача: “Посчитать все красные машины, свернувшие с Тверской на Большую Садовую за последний час”
- Ответ с помощью реляционной базы
  - Собрать все машины за час на парковке и посчитать
- Ответ с помощью Oracle Event Processing
  - Встать на повороте и считать проезжающие красные машины

Основная идея:
обработка на лету

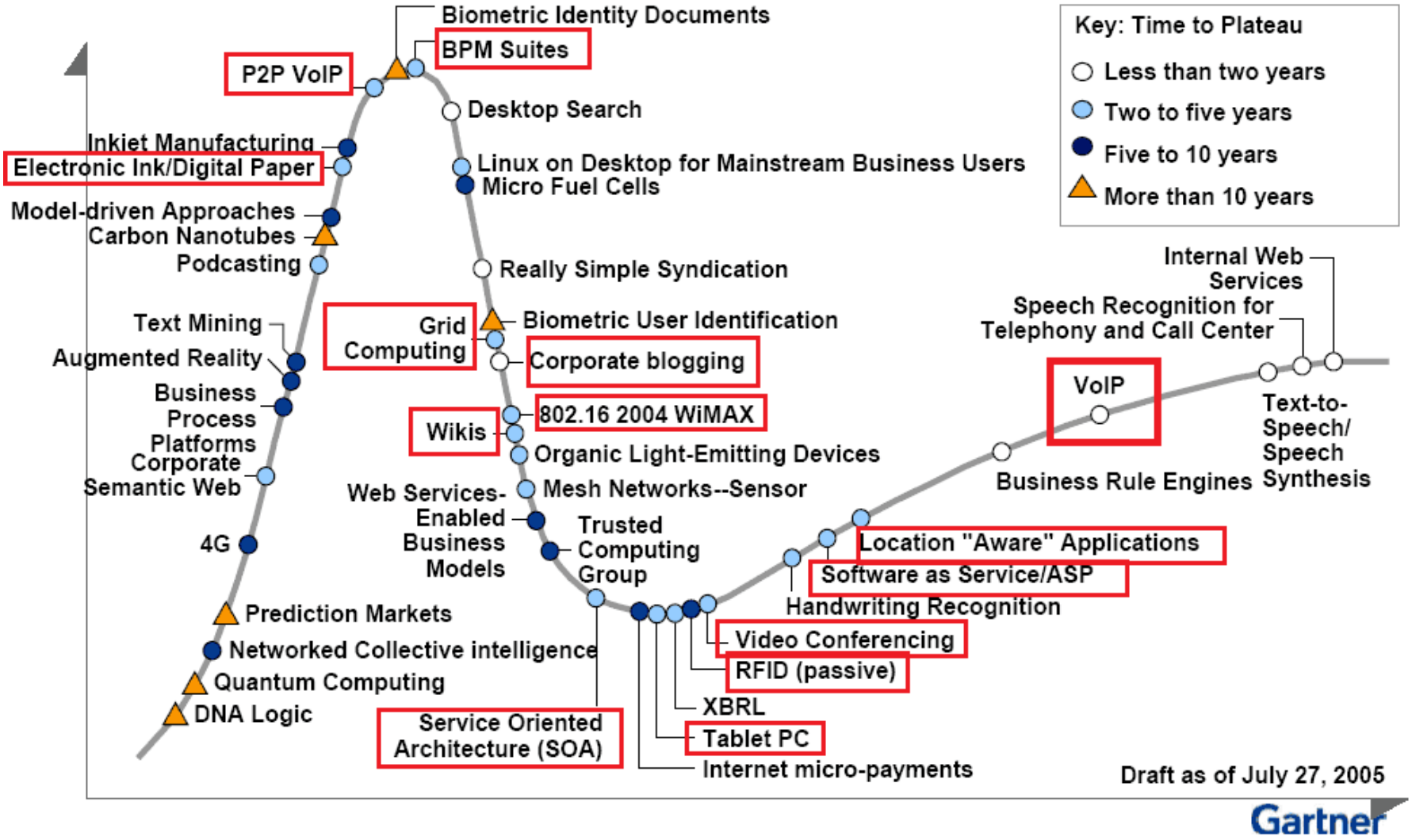
Почему Complex Event?



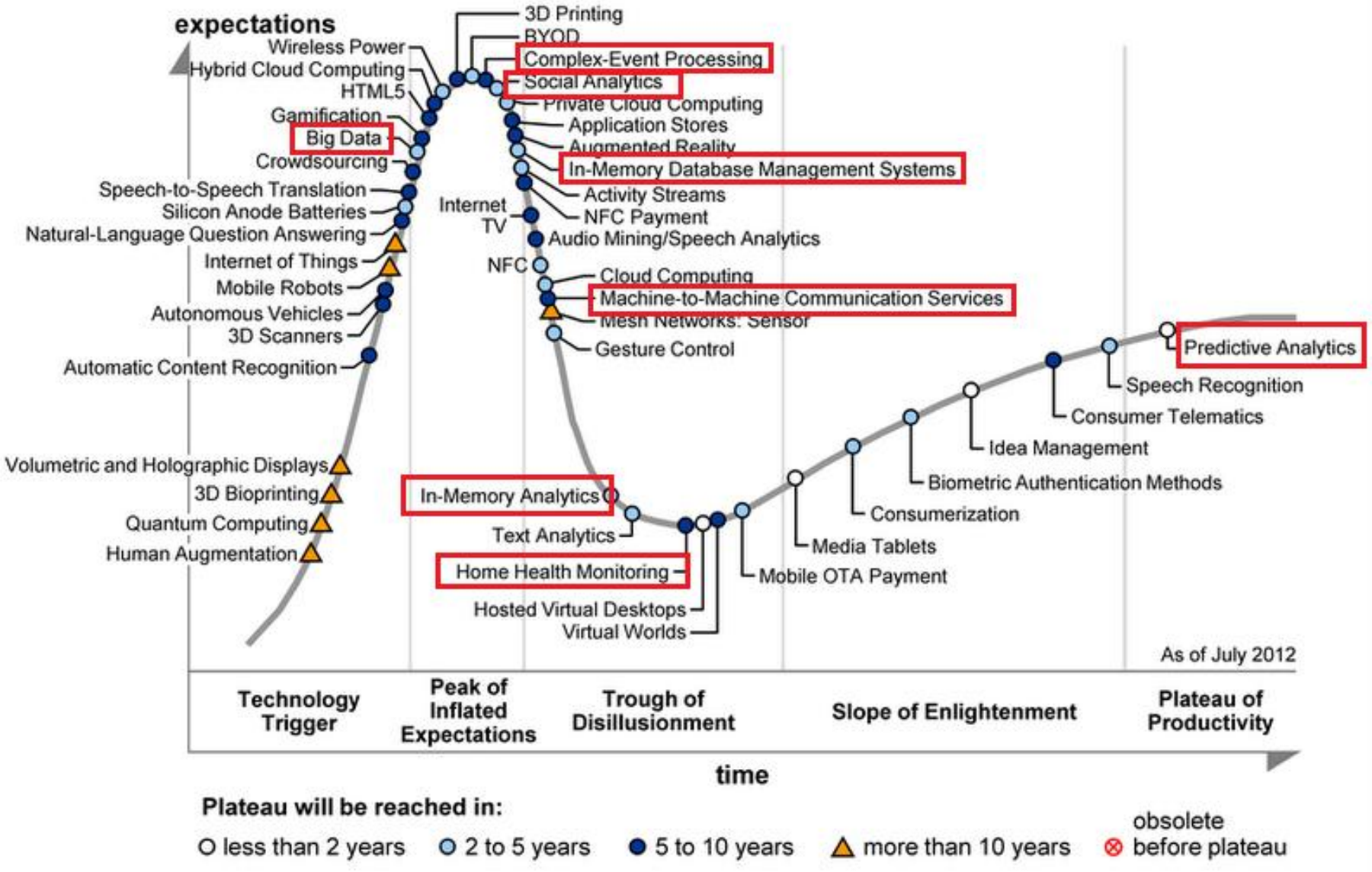
Agenda

- Что такое потоковые данные?
- Почему CEP актуален сегодня?
- Решение Oracle по обработке событий
- Примеры внедрений
- Вопросы&Ответы

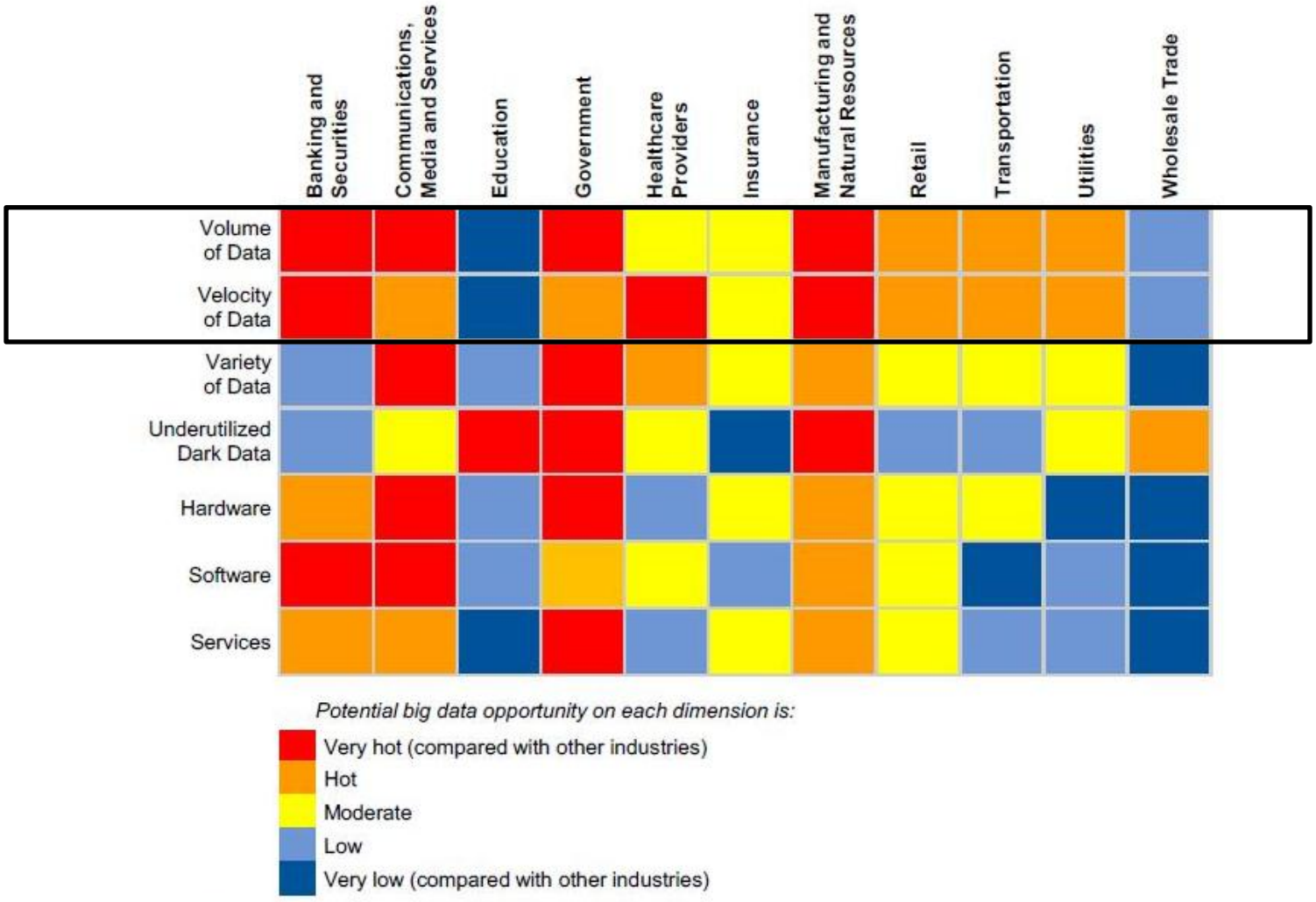
Hype Cycle 2005



Hype Cycle 2012

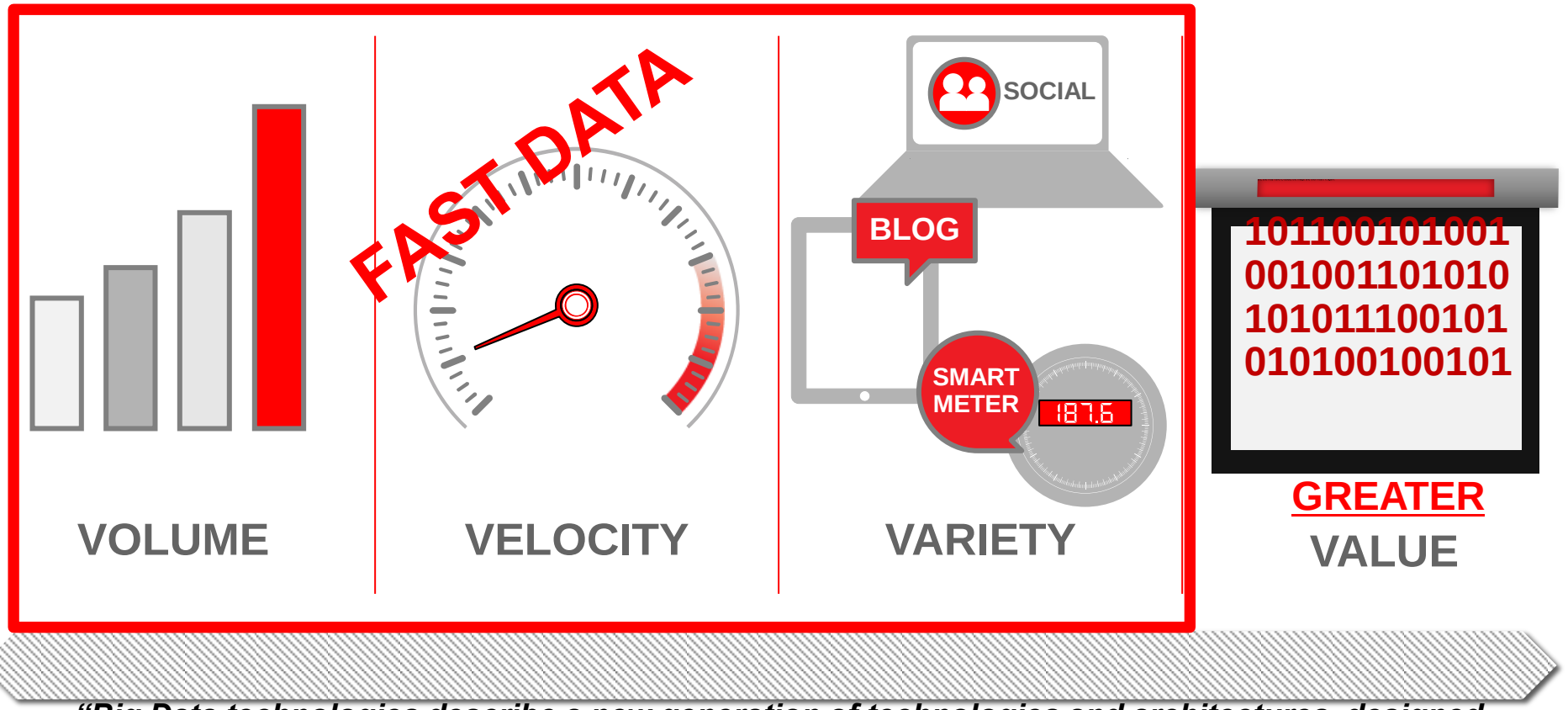


BigData by Industry



Source: Gartner (July 2012)

What Makes it Big Fast Data: The 4 Vs Principle



"Big Data technologies describe a new generation of technologies and architectures, designed to economically extract value from very large volumes of a wide variety of data, by enabling high velocity capture, discovery and/or analysis."

Source: Matt Eastwood, IDC

Agenda

- Что такое потоковые данные?
- Почему CEP актуален сегодня?
- Решение Oracle по обработке событий
- Примеры внедрений
- Вопросы&Ответы

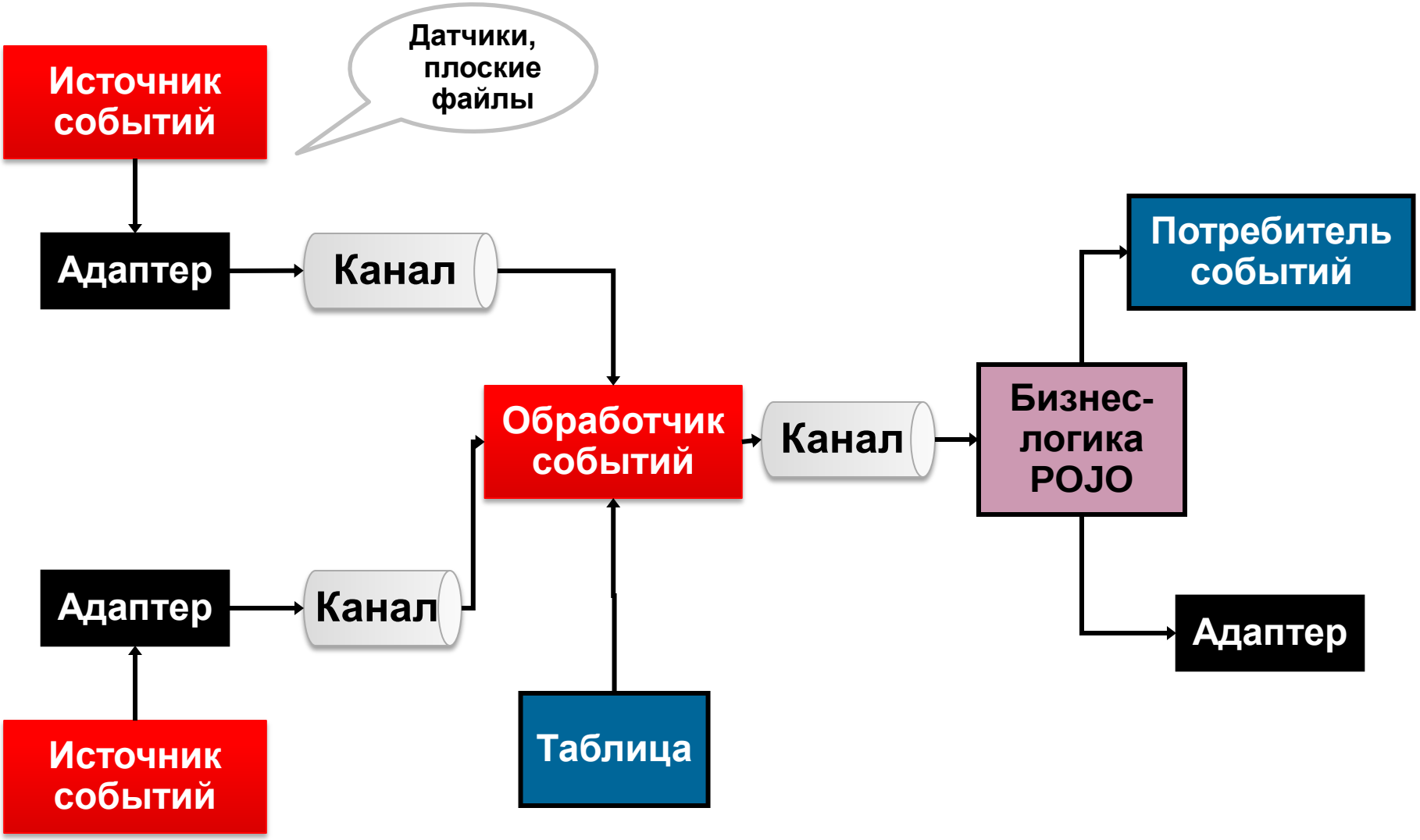
Evolution of Oracle Event-Driven Architecture

- 1992 Database: Oracle DML Triggers
- 1997 Database: Oracle Advanced Queuing
- 2002 Database: Oracle Data Streams
- 2004 Middleware: Oracle Business Activity Monitoring
- 2006 Middleware: Event-Driven Architecture Suite
- **2008 Middleware: BEA Web Logic Event Server with enhancements as Oracle Complex Event Processing 10.3 in Oracle Event-Driven Architecture Suite 10g**
- Now Available: Oracle Event-Driven Architecture Suite 11gR1
  - Oracle Business Activity Monitoring (BAM), java edition
  - Oracle Complex Event Processing with 100% SQL 99 compliance and industry standard, CQL, for pattern matching

Oracle Event Processing

- OEP это полное решение для создания приложений, фильтрующих, коррелирующих и обрабатывающих события в реальном времени
  - Позволяет создавать основанную на событиях архитектуру для аналитики в реальном времени
  - Построено на индустриальных стандартах:
    Java, Spring, OSGi, SQL
  - Высокая пропускная способность (сотни тысяч событий в секунду) и низкая задержка
  - Java container для развертывания приложений, написанных на Java и CQL (Continuous Query Language -- SQL extension)
  - Не требует полноценного сервера приложений JEE

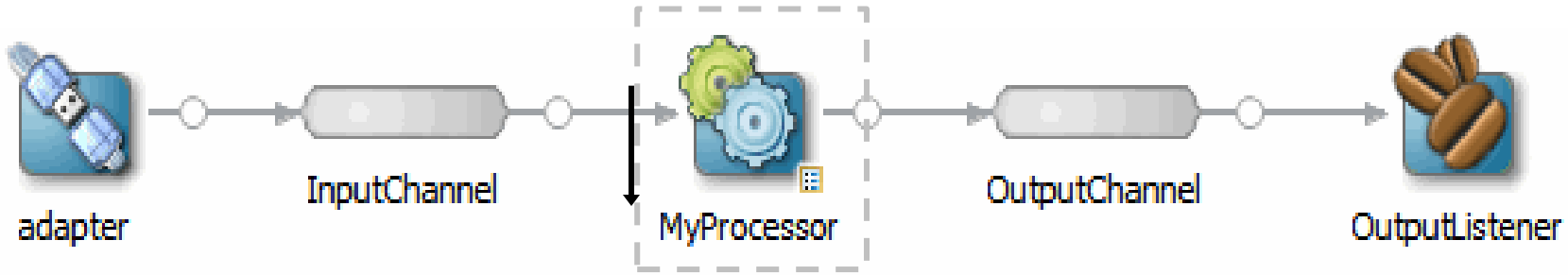
Архитектурные блоки



Адаптеры

- **csvgen** (Входящий) адаптер, используется для тестирования
- **File** (Входящий) адаптер читает локальный файл и создает событие для каждой строки в файле.
- **JMS** (Входящий и Исходящий)
  - Поддерживает автоматическую обработку Map сообщений
  - Предоставляет специальное API для преобразования различных типов сообщений
- **HTTP Pub/Sub** (Входящий и Исходящий)
  - Поддержка автоматической обработки JSON
  - Предоставляет специальное API для сложных преобразований в/из JSON строк
- **СУБД** Oracle, MSSQL, DB2, MySQL с помощью GoldenGate
- **Собственные адаптеры** на Java

Процессоры




```
<processor>
    <name>MyProcessor</name>
    <rules>
        <query id="ExampleQuery">
        <![CDATA[ select * from InputChannel[now] ]]>
        </query>
    </rules>
</processor>
```

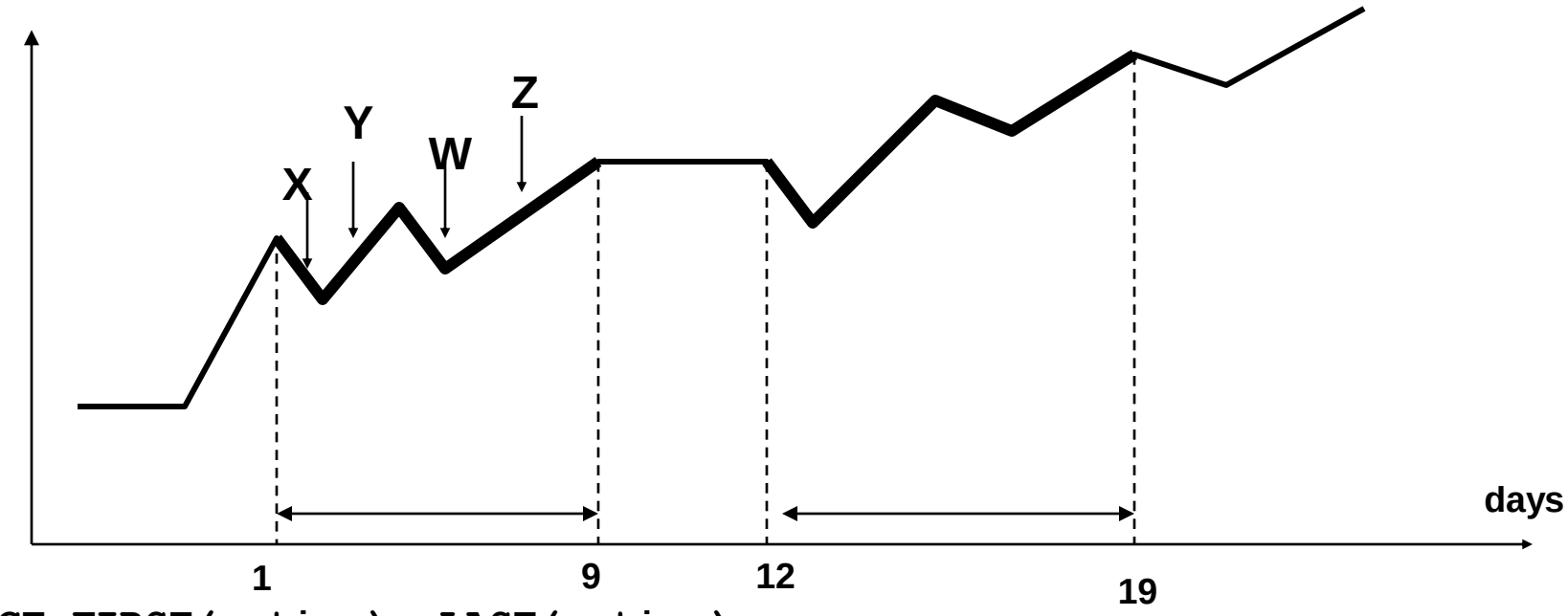
OEP: Пример запроса

В памяти, непрерывные запросы

- **Фильтрация (следим за звонками со стоимостью больше 2200)**
  - `SELECT * FROM S WHERE S.A > 2200;`
- **Корреляция и агрегация (fraud)**
  -
```
SELECT ID
(SELECT S1.ID, SUM(S1.COST) C1, SUM(S2.COST) C2,
      FROM S1 [RANGE 1 hour]
      JOIN S2 [RANGE 1 week]
      WHERE S1.ID = S2.ID)
WHERE C1*24*7>C2*10
```

Выявление шаблонов

Торговля на бирже – шаблон “W”




```
SELECT FIRST(x.time), LAST(z.time)
FROM ticker MATCH_RECOGNIZE (ONE ROW PER MATCH PARTITION BY name
       PATTERN (X+ Y+ W+ Z+)
       DEFINE X AS (price < PREV(price))
              Y AS (price > PREV(price))
              W AS (price < PREV(price))
              Z AS (price > PREV(price)))
```

Обогащение потоков

Поток с датчиков

Time	Input (STOCKSTREAM)	Output (QUERY)
1	{“AAA”, 10.0, 10.5}	{“AAA”, “Сургут”, “Участок 10 км”}
2		
3		

Идентификатор датчика “AAA” используется для в реляционном источнике.

Датчики

Symbol	Fullname	Location
AAA	Сургут	Участок 10 км
BBB	Сковородино	Нефтеналивная станция

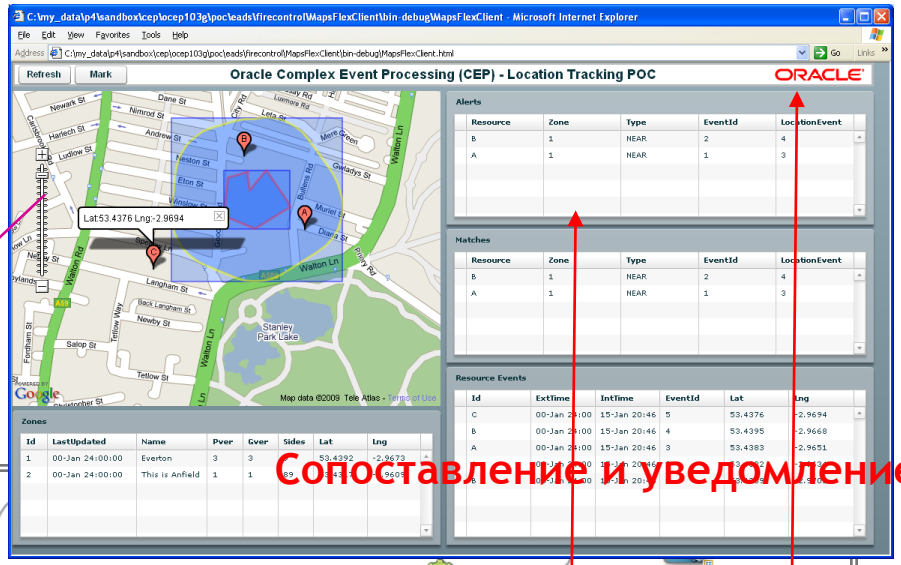
Местоположение извлекается из таблицы и добавляется к результирующему событию

Отслеживания перемещения

Готовая интеграция с Oracle Spatial

Инфопанель

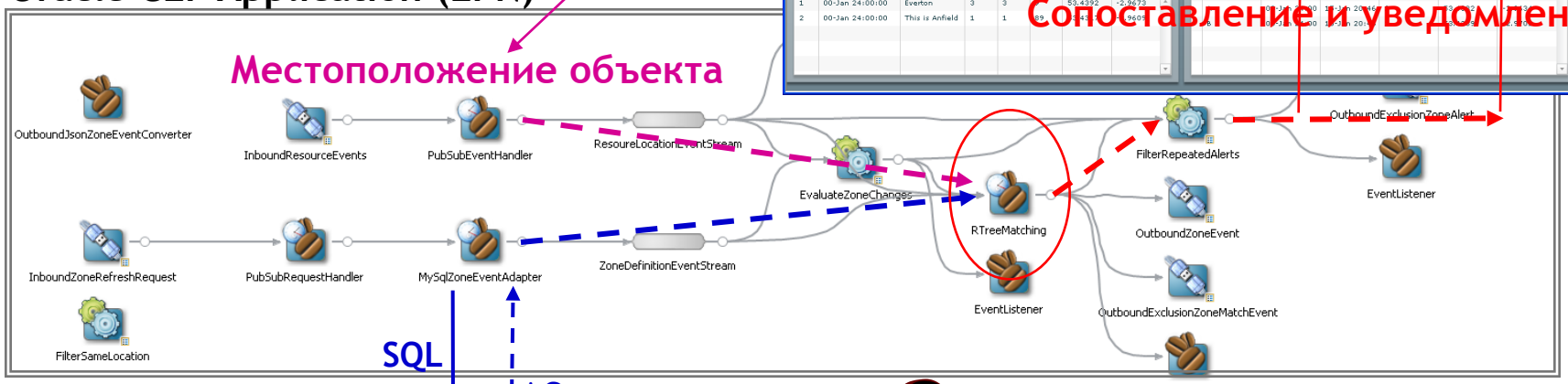
- Менеджер полигонов
- Oracle CEP Application
- Инфопанель



Oracle CEP Application (EPN)

Сопоставление и уведомление

Местоположение объекта



SQL

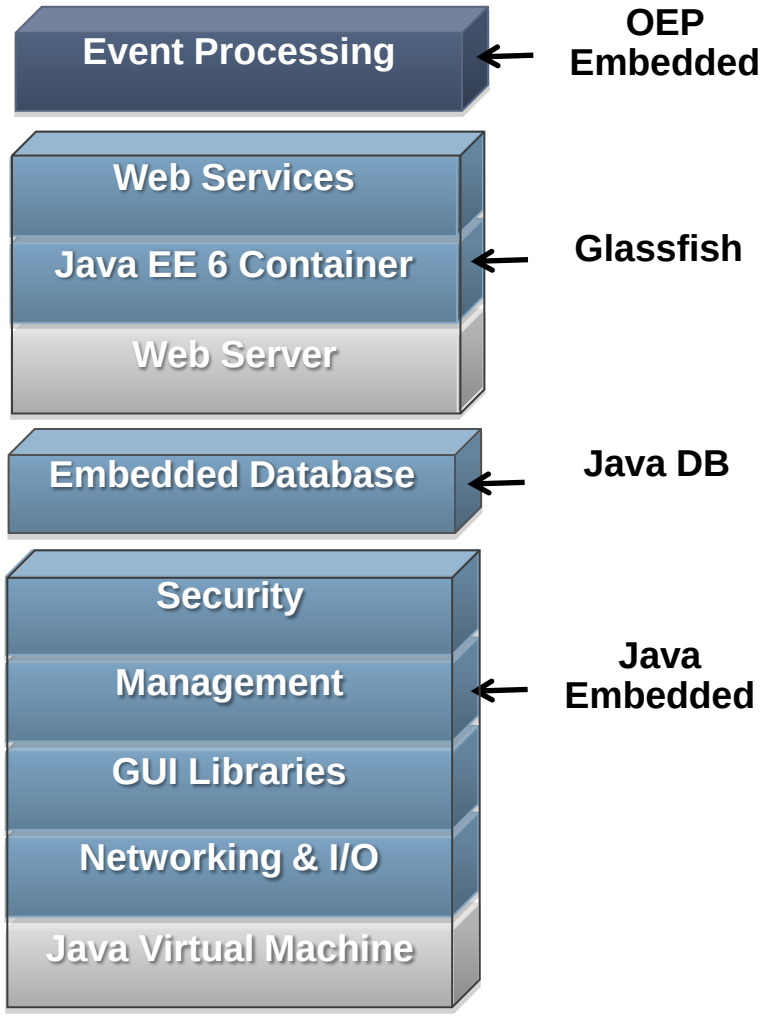
AQ

Polygon Manager

Определение зон

OEP Embedded on Smart Devices

OEP Embedded работает на базе стека Java Embedded и позволяет приложениям, работающим в устройствах **отслеживать и анализировать данные в реальном времени**



Case Study: Smart Grid Deployments

Распределение управления событиями между сервером и устройствами

Smart Meter

Smart Meter

Concentrator

Gateway

- Локальный анализ данных
- Фильтрация в реальном времени, корреляция и простая обработка
- Бизнес правила загружаются с сервера

Java Embedded Suite

- Glassfish
- Java DB
- Java SE-E
- OEP Embedded

J2EE Application Server (Weblogic, iAS)

Coherence

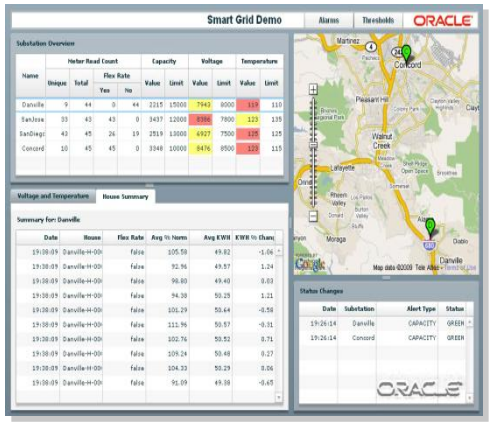
Java SE

- Oracle BAM
- Oracle Event Processing
- Oracle WebCenter
- Oracle DB

Exadata/Exalogic

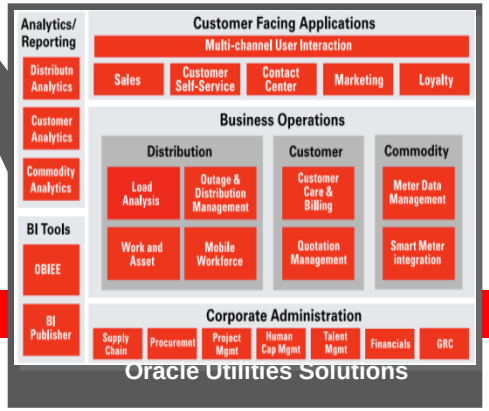
Utilities Business Applications

- Complex event processing для аналитики в зависимых приложениях, SOA and event-driven architectures
- В памяти, непрерывные запросы по множеству потоков из различных источников данных



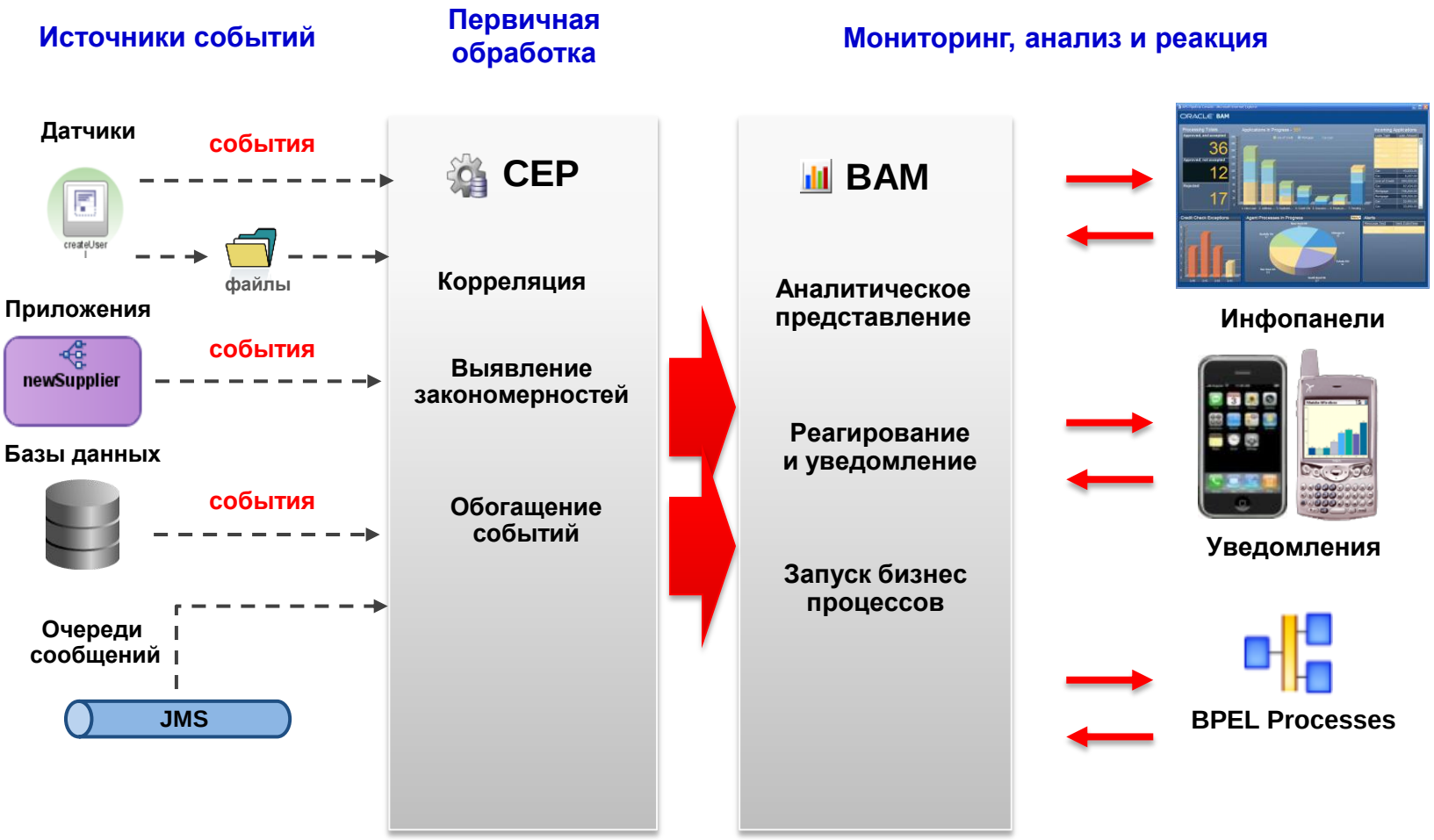
Operations Dashboard

Management Console



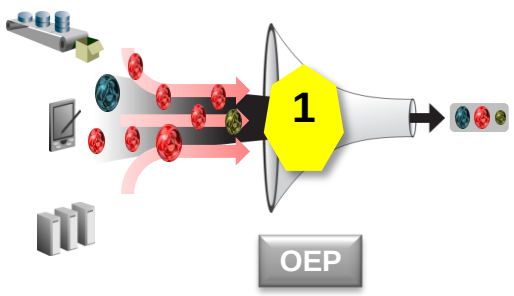
Oracle Utilities Solutions

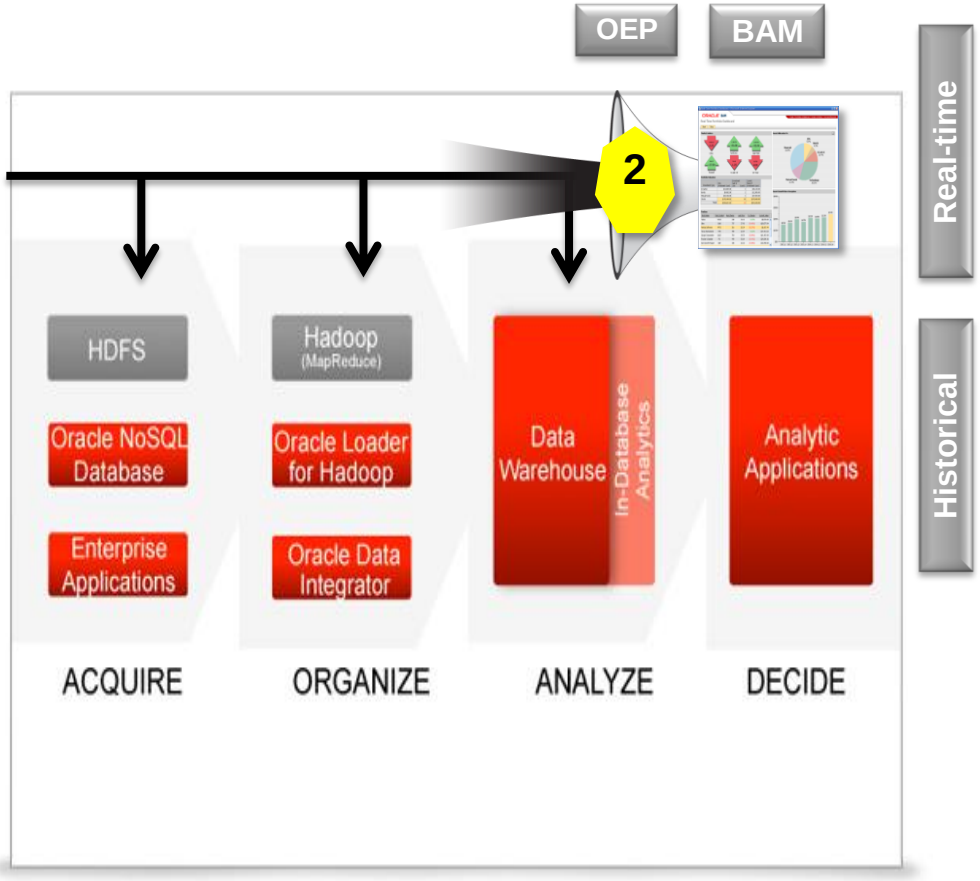
Архитектура решения по обработке событий



Big Data & Fast Data

Adding real-time capabilities with OEP & BAM





1. **Filter out useless, redundant data, correlate *; find and focus on events that matter***
2. **Analyze *and act ; execute critical decisions* in real-time, complementing historical approaches**

Производительность

2 сервера, в каждом по 2 Xeon с 6 ядрами

Conn. per Instance	Injection Rate per Instance (events/ sec)	Total Injection Rate (events/ sec)	Output Rate (events/ sec)	Average Latency (microsecs)		99.99% Latency (millisecs)		Overall Max Latency (ms)
				CEP 1	CEP 2	CEP 1	CEP 2	
1	50,000	100,000	5275	20.1	20.1	0.2	0.2	7.4
2	100,000	200,000	10,540	20.7	20.2	0.1	0.1	8.6
3	150,000	300,000	15,733	21.1	20.5	0.2	0.1	8.8
4	200,000	400,000	20,949	21.5	20.7	0.2	0.1	9.3
5	250,000	500,000	26,201	21.9	20.8	0.2	0.2	16.8
6	300,000	600,000	31,410	22.5	21.2	0.2	0.2	23.2
7	350,000	700,000	36,652	25.5	24.5	0.3	0.5	17.9
8	400,000	800,000	41,889	28.1	26.8	0.6	0.7	22.7
9	450,000	900,000	47,100	30.3	29.1	1.2	1.3	25.9
10	500,000	1,000,000	52,173	32.5	31.0	1.8	1.8	34.5

Agenda

- Что такое потоковые данные?
- Почему CEP актуален сегодня?
- Решение Oracle по обработке событий
- Примеры внедрений
- Вопросы&Ответы

Fast Data

Volume

Turkcell

The Location-based Temporal Mobile Billboard

TURKCELL TECHNOLOGY

TURKCELL TECHNOLOGY RESEARCH & DEVELOPMENT INC.

COMPANY OVERVIEW

- Turkcell Technology has more than 15 years of development experience with its solutions applied and proven at leading operators in more than 10 countries
- Specialized for over 10 years in BI in the Telecommunications industry, designing, building and running one of the largest data warehouses in telecom industry with a proven record of success in BI operations
- Early and leading Industry adopter of new real time BI trends, taking advantage of Event Driven Architectures, Event Pattern Analysis in a Spatial and contextual data context

CHALLENGES/OPPORTUNITIES

- Requirement to improveTurkcell's ability to capture real-time events in its GSM network
- Demands for Turkcell to give its subscribers a new service that provides a dynamic, wide and targeted range of offers
- Increase service clients revenue generation by adding new marketing capabilities to their portfolio

PROJECT OBJECTIVES

- Provide a Mobile Billboard temporal location based marketing platform for its subscribers, to make offers and advertisements, and send alerts at the right place and the right time

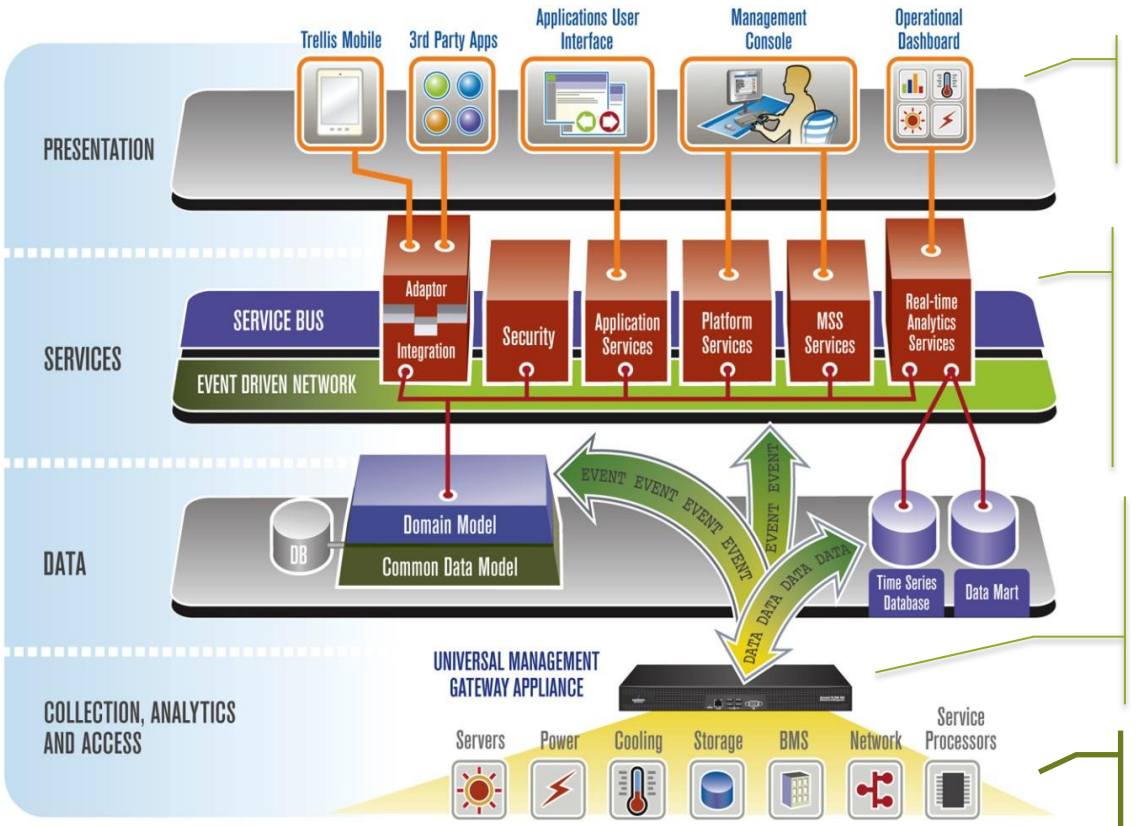
RESULTS

- *New solution infrastructure exceeded expectations by processing over 800,000 subscriber related events per second (with 1.5 Billion Events Daily)*
- *Provide and execute over 50 simultaneous campaigns*
- *Ensured customer Responsiveness with less than 1 second times with a scalable architecture, ready to expand on demand*

Trellis Platform Architecture

Multi-Tier Architecture for Event Management

Push as much element-specific correlation as possible to the lowest tiers of the architecture



Top Tier – General purpose Apps Console
Single Integrated View

OEP, Coherence , Modeling
Middle tier – Platform
Multi-domain event management, correlation & analysis

Embedded OEP & Coherence
UMG – aggregation tier across devices and support for non intelligent devices

Embedded OEP
Device Specific, specialized
Service processors, Unity Card

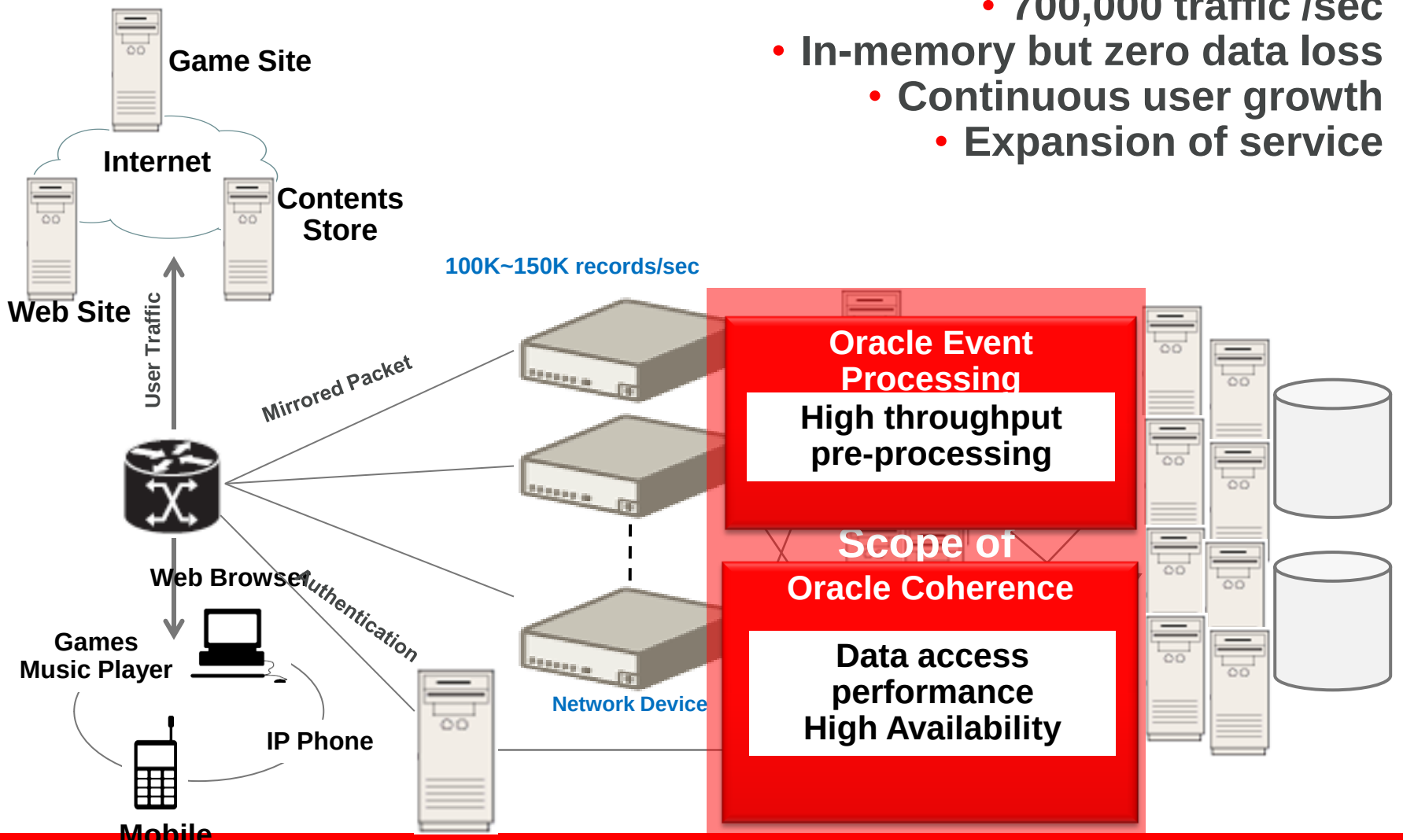
Data and Event Enrichment

Avocent

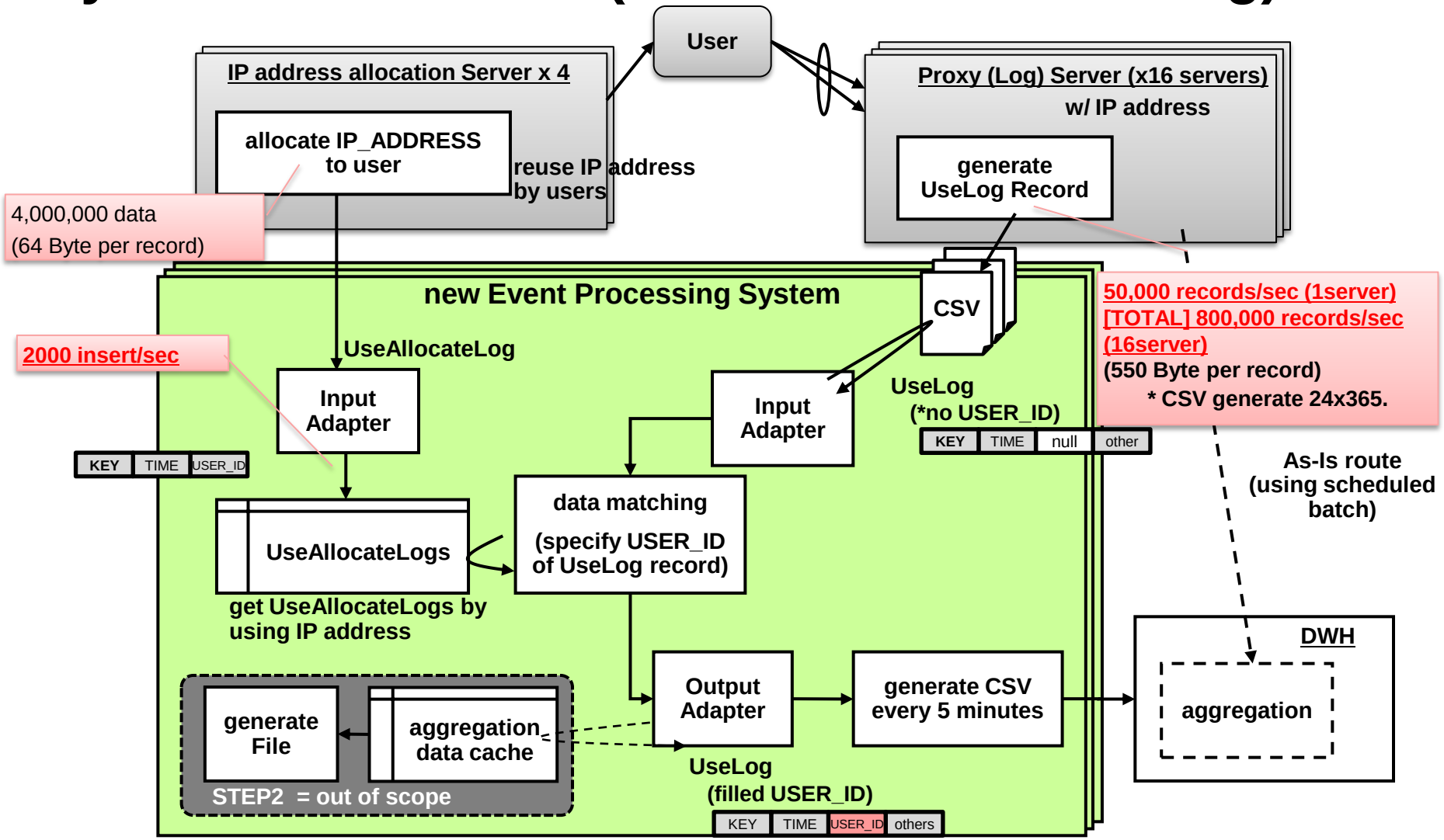
EMERSON Network Power

Role of Oracle Event Processing and Coherence

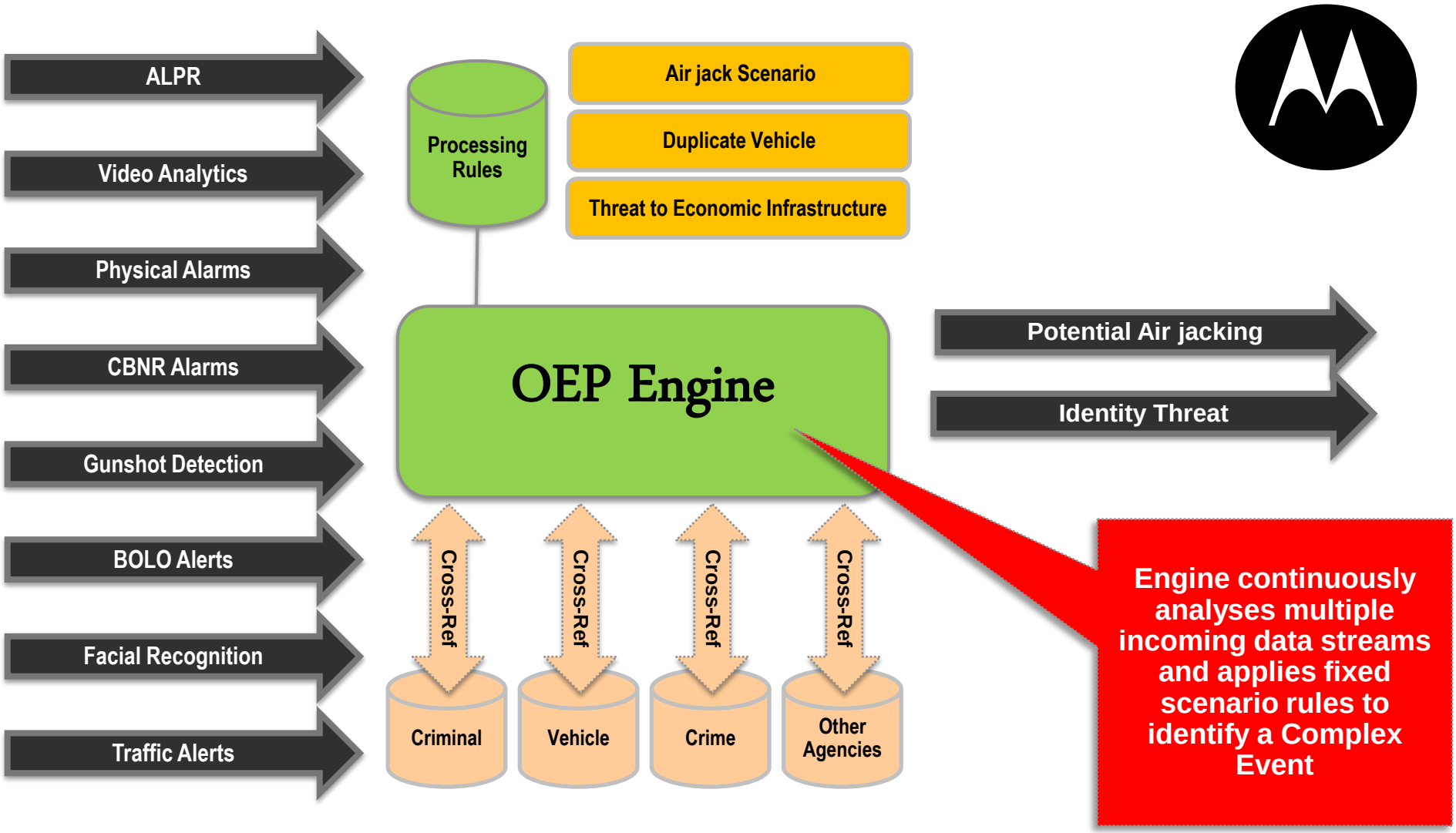
- 700,000 traffic /sec
- In-memory but zero data loss
- Continuous user growth
- Expansion of service



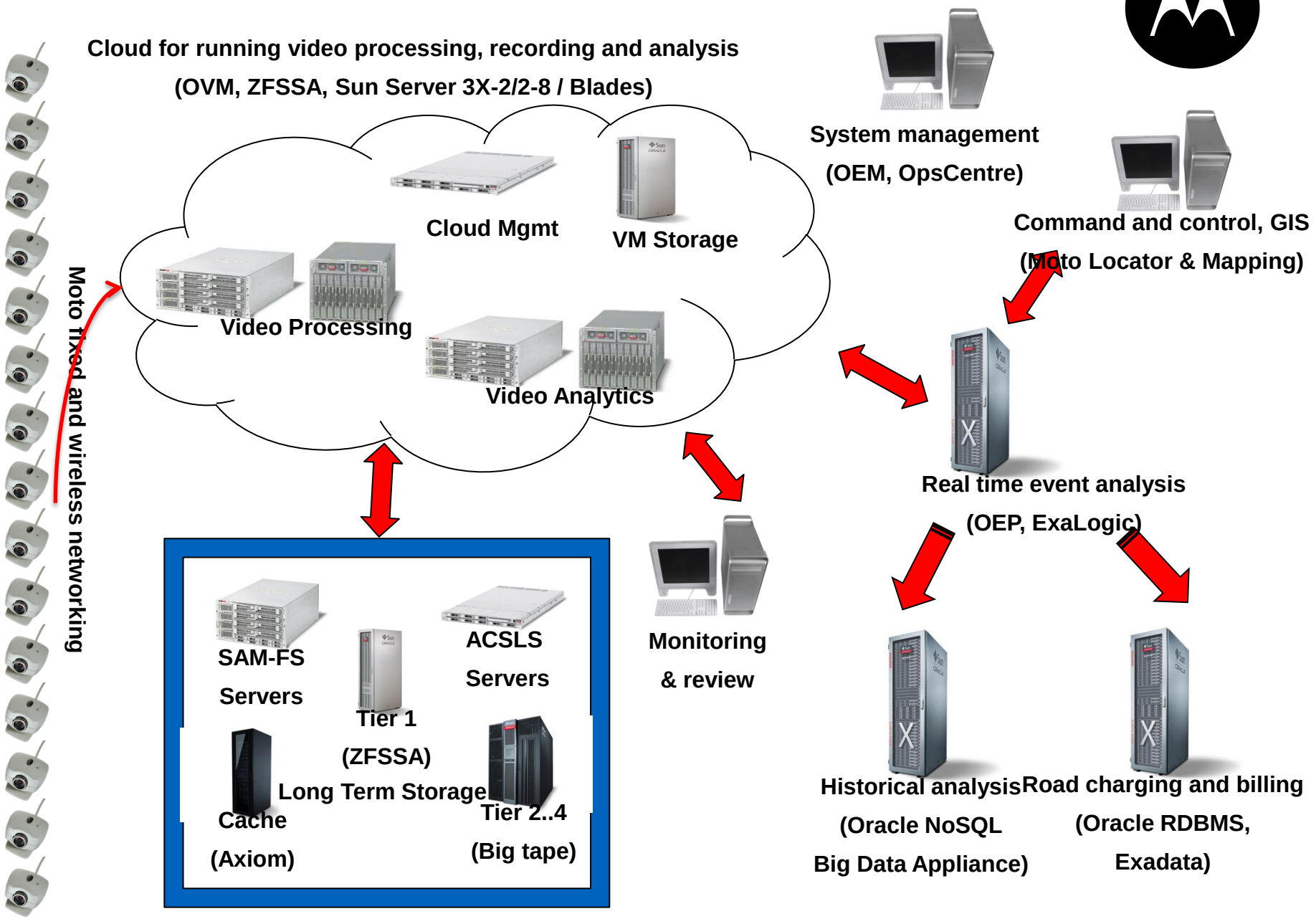
NTT DoCoMo
System Architecture(new: Event Processing)



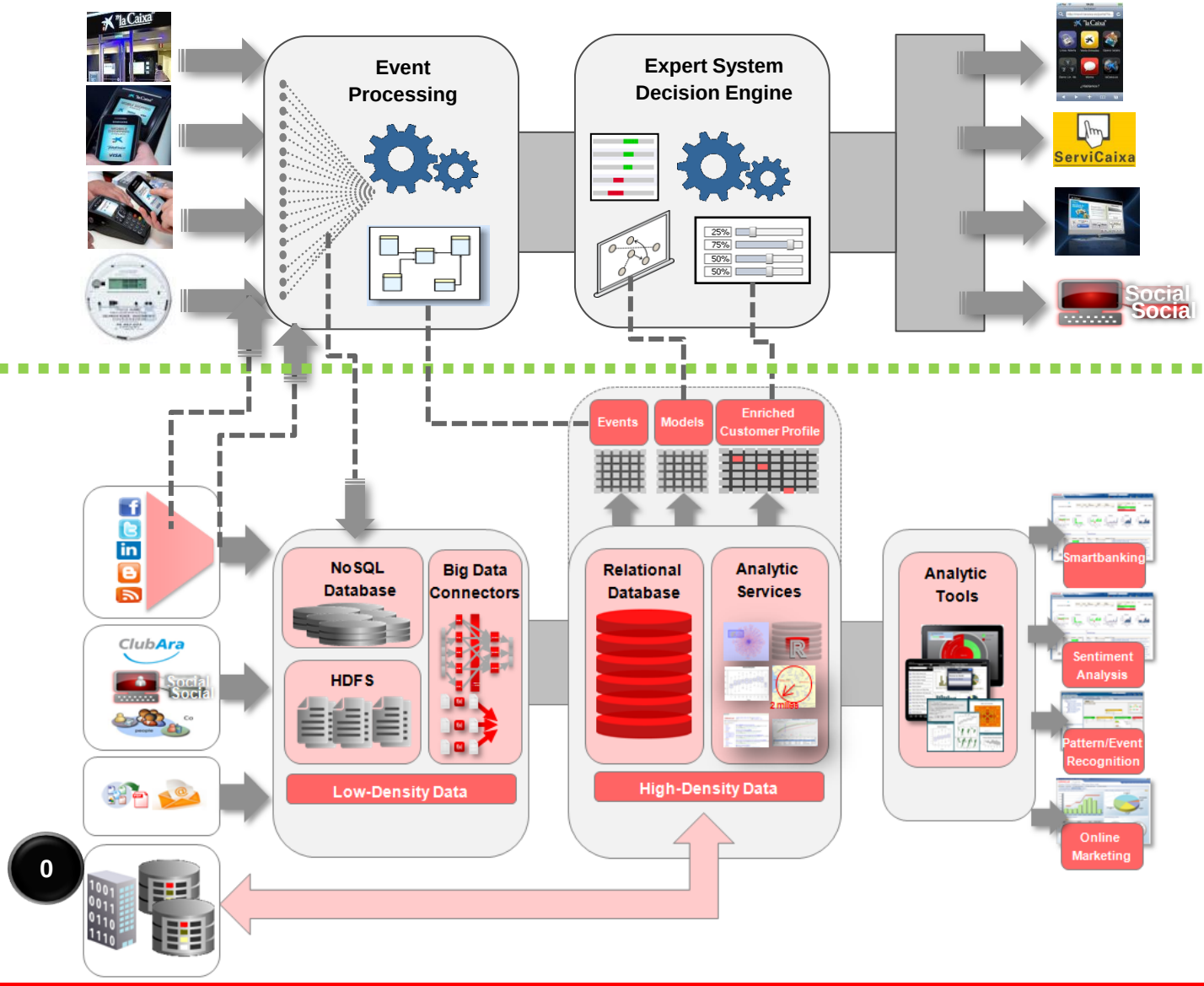
RULES ARE FIXED, DATA IS CONTINOUS



ARCHITECTURE



Fast Data Architectures

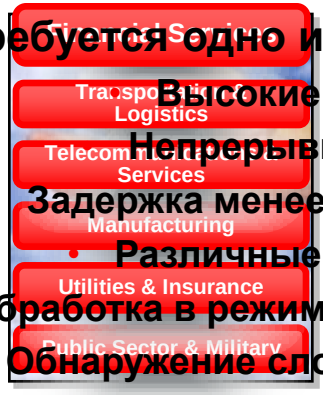


Event-Driven Applications –

- Распределенное управление обработкой заказов
- Потоковый анализ СМС

- Geo-Fencing
- Управление активами & логистика
- Управление трафиком

- Алгоритмическая торговля
- Умная маршрутизация заказов
- Online обнаружение мошенничества



Требуется одно из следующего:

- **Высокие объемы**
- **Непрерывный поток**
- **Задержка менее милисекундыу**
- **Различные источники**
- **Обработка в режиме временного окна**
- **Обнаружение сложных шаблонов**
- **Визуализация бизнес-событий**

- Online Retail Predictive
- Обработка жалоб в фармацевтике

- Отслеживание близости/положения
- Обнаружение проникновения
- Управление военными системами

- Управление инфраструктурой
- Реагирование на аварии – землетрясения, наводнения

ALEXANDER RYNDIN

Старший консультант по пространственным технологиям и технологиям хранилищ данных

Oracle CIS

Moscow,
Presnenskaya nab., 10,
Block C

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