Agile Enterprise Analytics

OLIVER RATZESBERGER
Sr. Director Architecture & Operations
eBay Inc.
February 2010
eBay Inc Overview

Pierre Omidyar founded eBay on a simple idea: **People are basically good.** This fundamental belief created a completely new kind of marketplace, forever transforming e-commerce. Today our values continue to guide how we do business and connect buyers and sellers around the world.

Our Values

- We create opportunities for people.
- We care. Because we know people depend on us.
- We make a difference in the world.
eBay Inc Overview

With more than 88 million active users globally, eBay is the world's largest online marketplace, where practically anyone can buy and sell practically anything. With a diverse and passionate community of individuals and small businesses, eBay offers an online platform where millions of items are traded each day.

With more than 75 million active registered accounts worldwide, PayPal is the faster, safer way to pay and get paid online. Available in 190 markets and 19 currencies around the world, PayPal enables global e-commerce by making payments possible across different locations, currencies, and languages.
eBay Inc Overview

Newsweek Ranks eBay Inc. One of the Greenest Companies in America - Newsweek magazine ranks eBay Inc. as one of the top green companies in America – coming in at No. 76 on the list of the 500 largest corporations in the United States.

eBay Inc.: The Most Trusted Company for Privacy - eBay tops the list of the "Most Trusted Companies for Privacy" in the 2009 consumer study commissioned by TRUSTe and Ponemon Institute, a privacy and information security research company. In 2008, eBay Inc. ranked in the top five.

eBay Inc. Recognized by the President of the United States - eBay Inc. was awarded the 2007 National Medal of Technology and Innovation, becoming the first Internet company to receive the prestigious award. The medal is presented annually to companies and individuals by the President of the United States during a ceremony at The White House.
Every Second – in over 50,000 Categories
eBay Analytics Technology Highlights

>50 TB/day of new, incremental data
>50 PB/day Processed
>150^10 new records/day
>100k data elements
>50k chains of logic
>5000 business users & analysts

Active/Active
turning over a TB every 5 seconds

24x7x365 Always online

99.98+% Availability

Millions of queries/day

Near-Real-time
eBay Analytics Core

- **MicroStrategy**
- **Unica**
- **Crystal**
- **Hadoop**
- **SAS**
- **SQL**
- **MAX**
- **SOA/DAL**

**Primary**
- **Relational Data**
  - **Teradata** 2.5PB
    - MPP
    - Local Interconnect
  - 'Singularity' 8PB
    - Deep Data, Semi Structured
    - Linux

**Secondary**
- **Relational Data**
  - **Teradata** 2.2PB
    - MPP
    - Local Interconnect
    - Wide Area Interconnect 1000 miles
  - 8PB
    - MPP/HPC/Grid
    - Linux

- **Ab Initio**
- **Informatica**
- **Golden Gate**
- **UC4**
- **BES**
- **MAX**
- **SOA**
Design for the Unknown

>85% of eBay analytical workload is NEW & Unknown

**Exploration** is the core of an analytical company

The metrics you know are ‘cheap’

The metrics you don’t know are expensive but also high in potential ROI

Design **can’t be static** or dependent on specific questions or dimensions
Analytics DNA

Embedded in our daily life

Bottoms-up & Tops-down

Think and Live Analytics

Always

But know when to avoid Analysis Paralysis!
Types of Analytics at eBay

Basically measure anything possible - A few examples:

- Marketing
- Buyer Experience
- Finance
- Trust & Safety
- Technology Operations
- Customer Service
- Loyalty
- Information Security
- Infrastructure
- Finding
- User Behavior
- Seller Experience
Key Performance Indicators

Align individual and departmental performance objectives with corporate goals
KPI Example: Technology Operations

Parallel Efficiency – (simplified) The effectiveness of distributing large amounts of workload over pools and grids of servers.

100% is GOOD  Less than 70% is BAD

10,000 Server running at an average PE of 50%

Established through Analytics of Operations Data – Minute by minute utilization metrics of entire infrastructure

Raising PE from 50% to 80% equals Millions in OpEx savings
KPI Example: Technology Operations (cont)

- Individual process rollup
- Grid level Parallel Efficiency at 99.9%

Note: Based on EMQD data. Queries run less than 5 seconds are not included.
Proliferation of Analytics

Decentralized Analytics
Shortened Time to Market Requirements
Adhoc Exploration

Departmental Data

Prototyping - Can’t wait for EDW

"We Need Data Marts!"
Storage Utilization

- Capacity Planning
- Facilities
- Search
- Operations
- Marketing
- Finance
- Corporate
- Fraud
- Customer Support
- Strategy
- Legal
- IT
- Shipping
- Supply Chain
- HR
Processing Capacity
Processing Capacity
Processing Capacity

- Capacity Planning
- Facilities
- Search
- Procurement
- Customer Support
- Finance
- Fraud
- Corporate
- Marketing
- Shipping
- Supply Chain
- Strategy
- Legal
- IT
- HR
- Operations
- Marketing
- Corporate
- Shipping
- Supply Chain
- Strategy
- Legal
- IT
- HR
- Operations
- Marketing
- Corporate
- Shipping
- Supply Chain
- Strategy
- Legal
- IT
- HR
- Operations
Processing Capacity
Processing Capacity

- Capacity Planning
- Facilities
- Search
- Operations
- HR
- Marketing
- Finance
- Corporate
- Shipping
- Supply Chain
- Customer Support
- Fraud
- Strategy
- Legal
- IT

eBay Inc. confidential
Processing Capacity
Processing Capacity

- Capacity Planning
- Facilities
- Procurement
- Customer Support
- Search
- Finance
- Fraud
- Corporate
- Marketing
- Supply Chain
- Shipping
- IT
- Legal
- Strategy
- Operations
- HR
- eBay Inc. confidential
Data Mart Implications

- Locked in
- Fragmented
- Redundant
- Slow Time to Market

10-20x (and more) Storage Waste

10-20% Overall Resource Utilization

FTEs to operate: 40+
Data Mart Dilemma

Total Cost of Ownership (TCO)
Fully loaded cost staggering $500k-M++

Biggest drivers are

- Maintaining separate databases
  weekly/daily/hourly data transfers
- Data inconsistencies
- Data redundancy
- Increased complexity
- Loss of lineage over time

A Data Mart cannot be ‘cheap’ enough to justify its existence
Agile Analytics needs Analytics as a Service

Massive scale **Analytical Utility Computing**

Bring your data - Perform your Analytics

From Simple Web based data upload ...to fully private Utility access

Combine custom data and code with ALL existing data
Virtual Data Marts in the Enterprise
Virtual Data Marts
Virtual Data Marts

Capacity Planning

Facilities

Search

Procurement

Customer Support

Procurement

Finance

Corporate

Fraud

Supply Chain

Shipping

IT

Legal

Strategy

Operations

HR

Marketing
Virtual Data Marts

- Capacity Planning
- Facilities
- Search
- Operations
- Marketing
- HR
- Finance
- Corporate
- Shipping
- Supply Chain
- Procurement
- Customer Support
- Strategy
- Legal
- IT
Virtual Data Marts
Virtual Data Marts

- Capacity Planning
- Facilities
- Search
- Procurement
- Customer Support
- Finance
- Fraud
- Corporate
- Supply Chain
- Marketing
- Shipping
- Operations
- Strategy
- Legal
- IT

eBay Inc. confidential
Virtual Data Marts
Virtual Data Marts in the Enterprise

Single MPP Infrastructure
One Physical Grid
Virtualized Analytics

Extremely Flexible and Nimble
Can adjust to workload changes in seconds

Share a **SINGLE** copy of data

1/10th of the storage for the Enterprise

5-10 x Capacity per $ invested

FTEs to operate: 2-4

Unbeatable
Virtual Data Marts

Various Types:
- Prototype/Sandbox
- Production Data Mart
- Development
- QA
- Testing and Experimentation

Logical Definitions NOT physical instances
Analytics as a Service 2.0

Virtual Data Mart Self Provisioning
   Your personal Data Mart in 5 min

Templatized Data Marts
   Pick from a variety of predefined templates
   
   Click Stream
   Financial
   Site Analytics
   ...

eBay Inc. confidential
Self Service Example
Key to success for AaaS

**Unlimited** Scalability

- Data Volume
- Concurrent Users
- # Applications
- # Queries
- Complexity

Large Scale MPP

**Sophisticated Workload Management**

- Resource Partitions
- Workload Queuing

Application Level Virtualization Support
Technology Stack Decision is Multi Dimensional

- IO
- CPU
- Concurrency
- Flexibility
- Storage
- Governance

A
B
C
Analytics as a Service - Benefits

Improved Time To Market - **Days/Weeks** vs Months

Enable the business to do **agile** prototyping

Enable the users to **“Fail Fast”** - Make it easy to try out new ideas

Eliminate stray Data Marts
Questions