“BIG DATA” Appliances

R Sathyanarayana

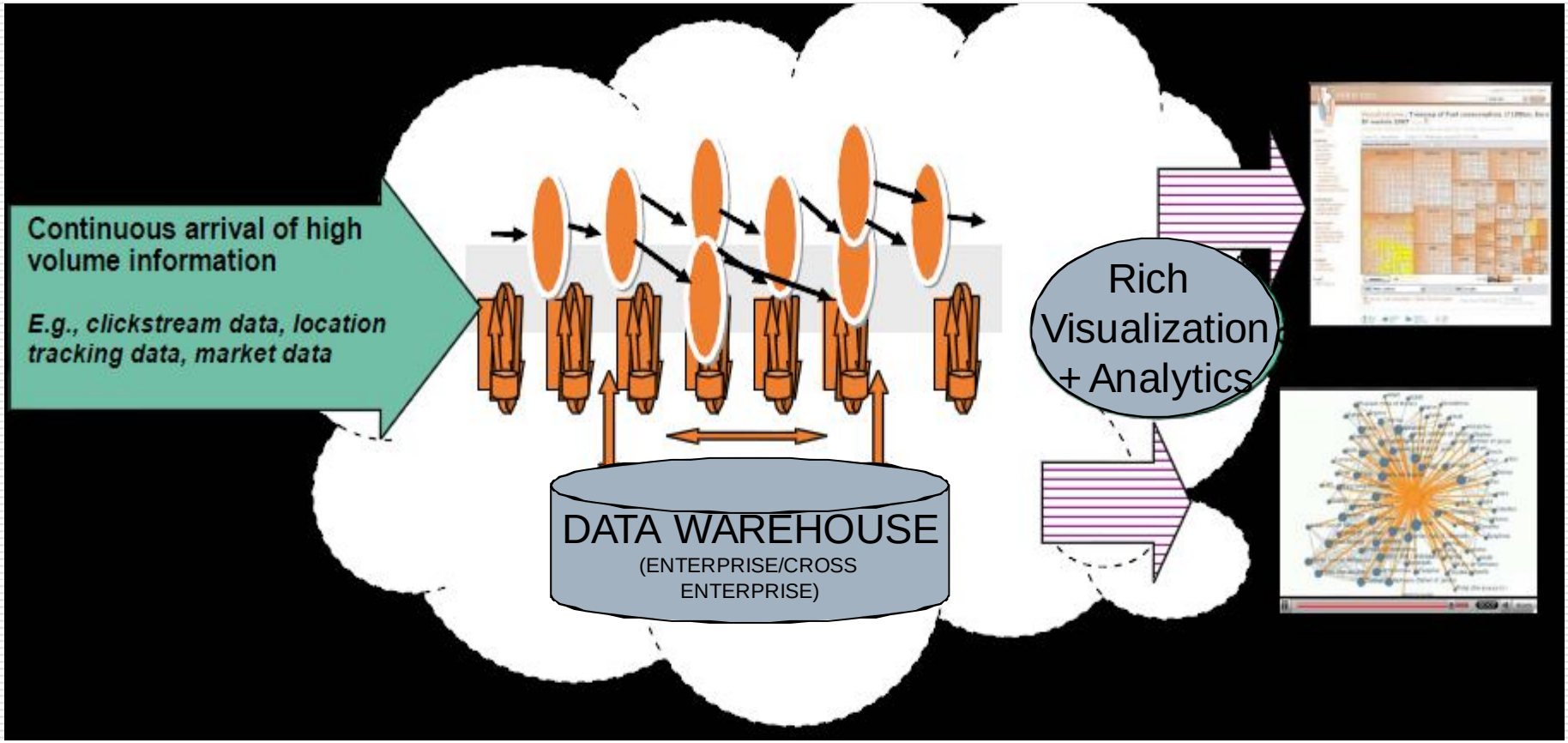
TDWI Bangalore
5 Feb 2010

Enterprise Information Management & Analytics practice, EMC Consulting

Starter Kit : Why Analytics on the Cloud?

- Internet scale analytics capability
  - Analyze enormous volumes of data with *cost efficiency and response time* unimagined a few years ago
  - Integrate cross platform data to derive meaningful insights : Ingest not gigas but teras in a day
- Transform fundamentally what is currently possible
  - Can we promote offerings/products to individuals?
  - Can we respond quickly to emergencies, frauds?
- Problem statement
  - What capabilities are required and how enterprises would use such capability?

Picture This….



“Big Data”

- Big data are datasets that grow so large that they become awkward to work with using on-hand database management tools. Difficulties include capture, storage, search, sharing, analytics, and visualizing. This trend continues because of the benefits of working with larger and larger datasets allowing analysts to "spot business trends, prevent diseases, combat crime.“
- One current feature of Big data is the difficulty working with it using relational databases and desktop statistics/visualization packages, requiring instead "massively parallel software running on tens, hundreds, or even thousands of servers."
- Big data sizes are a constantly moving target currently ranging from a few dozen terabytes to many petabytes of data in a single data set.
- Sample This : web logs, RFID, sensor networks, social networks, Internet text and documents, Internet search indexing, call detail records, genomics, astronomy, biological research, military surveillance, medical records, photography archives, video archives, and large scale eCommerce

Data Warehouse Appliances : Setting the context

- As IT organizations build up massive numbers of databases to deal with the explosion of data, the ability to make real-time decisions on new questions (BI) that involve enormous amounts of information (DW) will need to be a core competency for many organizations.

  Due to this shift, DW/BI customers need a solution that can provide extreme predictable performance, scale-out architecture for **'Big Data'** analytics and an enterprise-proven feature set all at the lowest TCO.

UNDER WORKS: Definition of DataWarehouse Appliances

- Original Definition : Hardware + Software – built and supported by a single vendor

- Partial Technology Stack : May or may not bundle with other vendors' hardware and / or operating system

Architectural considerations are evolving

CONFIDENTIAL - REMOVED

Trends: Consolidation in the industry : Small Vendors Vs Infrastructure Providers

- Focus is shifting in multiple areas:
  1. *From whole technology stack to pieces of it*
  2. *From hardware to software*
  3. *From proprietary to commodity hardware*
  4. *From new vendors to infrastructure providers*
  5. *From single to mixed workloads*
  6. *From data marts to enterprise data warehouses*

- These trends affect the content & capabilities of DWAs, where to get them, how to define them, how to use them.

Source: TDWI Research 2010.

Major differences between the DW appliances

- Column vs. Row Storage
- Polymorphic Storage (Both Column and Row)
- Proprietary and Commodity Hardware
- In-Memory Processing
- Relationship with Existing Architecture
- Shared Nothing Architecture

The Challenges in Today's Data Warehousing Environments

- Sources of data and the amount of data to analyze is growing exponentially
- Stale data exists because DW solutions cannot ingest the vast amounts of data fast enough
- Lack of performance for advanced analytics and complex queries
- The number of users and the concurrency of users is increasing rapidly

Considerations of DW Solution for the Big Data

These are the characteristics you want in your DW solution:

- Easily scales to analyze the growing amounts of data
- Rapidly ingests large amounts of data from sources
- Provides high performance in database analytics
- Supports high user concurrency securely, reliably
- Handle multiple workloads

THANK YOU

- Sathyanarayana.Ranganatha@EMC.com