Evolving Solutions Disruptive Technology Series
Modern Data Warehouse

Presenter – Kumar Kannankutty
Big Data Platform Technical Sales Leader Host - Michael Downs, Solution Architect, Evolving Solutions
Modernizing the Data Warehouse

Pure Data for Analytics powered by Netezza Technology

Kumar Kannankutty
Big Data Platform Technical Sales Leader
kkannankutty@us.ibm.com
A different approach

A different technology

A different set of offerings

Uncommon results
Would You Still Use Google If It Took 3 Days And 7 People To Get A Search Result?
Ring any bells?

How can I deliver faster queries for my end users cost effectively?

How can I cut costs?

How can I be more responsive?

I'm drowning in large data sets

Feeling pressure to lower your overall data warehouse TCO?

Interested in using data for traditional and new types of analytics?

How can I improve price performance?

How can I be more innovative?

What's the most cost effective way to add new workloads and capabilities without affecting performance?
Traditional Data Warehouse Architecture

Source Systems

Data Marts

Data Integration

Enterprise Data Warehouse

- Billing
- HR
- ERP
- CRM

Data Sources
The Future Data Warehouse will be Logical

- NoSQL Doc Store
- Data Warehousing
- Deep Analytics, Modeling
- Transactional Systems
- Landing, Exploration, Archive
- Reporting, Analytics

Logical Data Warehouse

Traditional Data Warehouse
How the traditional DW was built

Observations:
1. Infrastructure and architecture designed for OLTP…30 years ago…and works well.
2. Complex – multiple interfaces, requires teams of people to manage
3. Bringing data to query processing layer is not efficient for DW and analytics!
Traditional Data Warehouses

are just too complex

They do NOT to meet the demands of advanced analytics on big data.

- Too complex an infrastructure
- Too complicated to deploy
- Too much tuning required
- Too inefficient at analytics
- Too many people needed to maintain
- Too costly to operate

Too long to get answers
We asked ourselves: Why can’t a data warehouse be a true appliance? Isn’t it about the experience?

- Dedicated device
- Optimized for purpose
- High OOTB Performance
- Complete solution
- Fast installation
- Very easy operation
- Standard interfaces
- Low TCO
A different approach

A different technology

A different set of offerings

Uncommon results
Inside the IBM PureData System for Analytics

**True Appliance**
- MPP
- Zero Configuration
- Call Home
- Vendor Updates SW
- Vendor Fixes HW
- Low Administration

**Compute Nodes**
- Hardware-based query acceleration with FPGAs
- Blistering fast results
- Complex analytics executed as the data streams from disk

**Storage**
- User data, mirror, swap partitions
- High speed data streaming
We predicted (15 years ago) that I/O would still be the bottleneck…. We were right.

CPU Performance has continued to outpace I/O performance.

How to do we accelerate I/O since it’s losing the war?

Enter the FPGA.
FPGA “I/O Turbo Charger” - No One Else Has This

- Like a CPU, but only operates on streaming data

- Customizable aspect of this technology lets us adapt it to SQL processing

- You have FPGA’s in your house
  - DVD/BluRay players
  - HDTV

- The military uses them for leading-edge data performance at low power

- Real-time video and audio processing
SQL at the speed of physics - The “Secret Sauce”

```sql
SELECT State, Age, Gender, count(*)
FROM MultiBillionRowCustomerTable
WHERE BirthDate < '01/01/1960'
AND State IN ('FL', 'GA', 'SC', 'NC')
GROUP BY State, Age, Gender
ORDER BY State, Age, Gender
```
Snippet-Blade™ (S-Blade) Components

CPU + RAM

IBM BladeCenter Server

FPGA’s

Netezza DB Accelerator Card
Netezza’s revolutionary approach

The Appliance
Simpler, faster, more accessible analytics

“This is what Netezza has done in the data warehousing market: It has totally changed the way we think about data warehousing.

- Philip Howard, Bloor Research
A different approach

A different technology

A different set of offerings

Uncommon results
IBM PureData System for Analytics
The Simple Data Warehouse Appliance for Serious Analytics

What makes it different?

**Speed** - 10-100x faster than traditional custom systems

**Simplicity** - minimal administration and tuning

**Scalability** - petabyte+ scale user data capacity

**Smart** - high performance, advanced analytics

**Secure** – hardware encryption

---

1 Based on IBM customers’ reported results. “Traditional custom systems” refers to systems that are not professionally pre-built, pre-tested and optimized. Individual results may vary.
The PureData System for Analytics N3001 Family

<table>
<thead>
<tr>
<th>Specification</th>
<th>N3001-001</th>
<th>N3001-002</th>
<th>N3001-005</th>
<th>N3001-010</th>
<th>N3001-020</th>
<th>N3001-040</th>
<th>N3001-080</th>
</tr>
</thead>
<tbody>
<tr>
<td>Racks</td>
<td>4U</td>
<td>¼</td>
<td>½</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>Active S-Blades</td>
<td>0 SW FPGA</td>
<td>2</td>
<td>4</td>
<td>7</td>
<td>14</td>
<td>28</td>
<td>56</td>
</tr>
<tr>
<td>CPU cores</td>
<td>40</td>
<td>40</td>
<td>80</td>
<td>140</td>
<td>280</td>
<td>560</td>
<td>1,120</td>
</tr>
<tr>
<td>User data (TB) *</td>
<td>4</td>
<td>8</td>
<td>24</td>
<td>48</td>
<td>96</td>
<td>192</td>
<td>384</td>
</tr>
</tbody>
</table>

* Before Compression

Linear Scalability!
Introducing PureData System for Analytics N3001-001

Bringing speed and simplicity to midsize organizations for big outcomes

- Rack Mountable 4U Solution
- Same SW
- Same simplicity
- Same features
- No FPGA’s
- Ideal for SMB
- Lower TCO than even SQL*Server

Solution Highlights
- Production ready
- Full function appliance
- User data capacity 16 TB*
- 2x10 core IvyBridge CPU per server
- High availability - All redundant hardware, 4 disk spares, hot swap power supply
- Self encrypting drives, Kerberos support, LDAP/Active directory

*Assumes 4x compression

Powered by Netezza Technology

© 2015 IBM Corporation
Spend Less Time Managing and More Time Innovating

- Easy Administration Portal
- No software installation
- No indexes and tuning
- No storage administration

- No dbspace/tables size and configuration
- No redo/physical/Logical log size and configuration
- No page/block size and configuration for tables
- No extent size and configuration for tables
- No Temp space allocation and monitoring
- No RAID level decisions for dbspace
- No logical volume creations of files
- No integration of OS kernel recommendations
- No maintenance of OS recommended patch levels
- No JAD sessions to configure host/network/storage

Data Experts, not Database Experts
The Premier Analytics Platform

*IBM Netezza In-Database Analytics*

- Transformations
- Mathematical
- Geospatial
- Predictive
- Statistics
- Time Series
- Data Mining

- No data movement
- Analyze deep and wide data
- High performance, parallel computation
Big Data and Business Intelligence Ready
Unlocking Data’s True Potential

Entitlements Included with the PureData System for Analytics N3001

Data Warehouse Appliance

Included
$0

Built-in, In-Database analytic capability and integration with a variety of 3rd party tools

Exceptional value provided

Business Intelligence
Cognos software, 5 Analytics User licenses, plus 1 Analytics Administrator license

Data Integration & Transformation
InfoSphere DataStage 280 PVUs, 2 concurrent Designer Client licenses and InfoSphere Data Click

Hadoop Data Services
InfoSphere BigInsights Software licenses to manage ~100 TB of Hadoop data

Real-time Analytics
InfoSphere Streams Developer Edition 2 users, non-production licenses

Optional

Industry Process & Data Models
Models for Banking, Financial Markets, Healthcare, Insurance, Retail, Telco

IBM InfoSphere Data Privacy and Security for Data Warehousing
A different approach

A different technology

A different set of offerings

Uncommon results
The DW Appliance Leader 1200+ Clients

**Marketing Services Providers**
- ACXIOM
- datalogix
- epsilon
- Kelley Blue Book
- MERKLE
- TRIDENT MARKETING
- appnexus
- edmunds.com
- exelate
- MediaMath
- nielsen

**Financial Services**
- American Express
- Bank of America
- Nationwide
- Capital One
- Wells Fargo
- State Farm
- MasterCard
- Fidelity
- CitiBank
- NYSE Euronext

**Health & Life Sciences**
- PREMIER
- Harvard Medical School
- Capital BlueCross
- Shire
- University at Buffalo
- Humana
- MassMutual
- UW Health

**Retail / Consumer Products**
- KOHL'S
- Bass Pro Shops
- LANDS'END
- carter's
- Best Buy
- Barnes & Noble Booksellers
- Neiman Marcus
- Kraft
- Walmart
- Yum!

**Telecom**
- Sprint
- AT&T
- iBasis
- QUALCOMM
- T-Mobile
- The Carphone Warehouse
- MACH
- TEOCO
- Virgin Media
- verizon

**Other**
- AIRWAVE
- Con-way Freight
- Olympus
- Foxwoods Resort & Casino
- bp
- CenterPoint Energy
- DAY Birger et Mikkelsen
- John Deere
- Fuso
- CONTRA COSTA COUNTY
- US Department of Defense

© 2015 IBM Corporation
They shipped us a box, we put it into our data center and plugged into our network. Within 24 hours we were up and running. I'm not exaggerating, it was that easy.

- Joseph Essas, Vice President of Technology, eHarmony
...when something took 24 hours I could only do so much with it, but when something takes 10 seconds, I may be able to completely rethink the business…

- SVP Application Development, Nielsen
Scalability

Credit scores the entire US population in **20 minutes**

“We score the entire population of the US in 20 minutes. It doesn’t get any easier than that.”

Edgar Denny, Sr. Software Engineer
With Netezza’s in-database technology, we can now individualize offers to millions of customers, resulting in coupon redemption rates that are unheard of in the industry.

- Eric Williams, CIO and Executive VP, Catalina Marketing
With the incredible performance benefits and improved time to market we are now realizing, we can easily keep pace as data volumes spike. Netezza has simplified our entire data warehouse environment.”

Brian Clark, Chief Architect and Managing Director
NYSE Euronext
Test Drive Netezza
Your Data. Your Site. Our Appliance.

www.ibm.com/software/data/puredata/analytics/
Presentation Download and registration for the next in the series: Optimizing Structured Data Archive July 28
http://www.evolvingsol.com/events-and-education

Webinar Replay available on the Evolving Solutions YouTube Channel

Evolving Solutions
Michael Downs, Solution Architect
michael.d@evolvingsol.com
612-805-5579
@emeldi

IBM Pure Data
Kumar Kannankutty
Big Data Platform Technical Sales Leader
kkannankutty@us.ibm.com