Microsoft Developer Day
Big Data Architecture

Pradeep Menon
Data Solutions Architect
Agenda

- Introduction
- Big Data Architecture Pattern
- Azure Data Service Offering
- Demos
- Q & A
**Key Notes:**

- Philosophy: Understand data. Transform data. Load data. Analyze data.
- More time spent on ETL. Less time for spend on analysis.
- Data discovery is a challenge.
- Not all data is analyzed.
Data in the Era of Big Data

- The definition of analyzable **data** has changed.
- Traditional sources of data is **disrupted**.
Processing capability has increased 10,000 times since 2000

Cost of storage has reduced by 1000 times since 2000

1 TB blob storage = $24/month
Data Lake Concept

Data Lake

Data Warehouse/Mart
Conceptual Data Lake Architecture

Structured Data Sources

Unstructured Data Sources

Analytical Sandboxes
- Data Discovery
- Exploratory Data Analysis
- Predictive Modeling

Raw Data Store

Batch-Processing Engine

Processed Data Stores

Real-Time Processing Engine

Presentation Layer
- Reports

Data Cataloging and Curation

Data Security and Governance
Key Component Groups - Lambda

**Process:**
- Batch Layer stores all the data in the rawest form possible, this set is the master data set.
- The master data set is immutable.
- Data follows a fact-based model, where each piece is atomic and timestamped.
- Streaming data follows both path.
- The speed layer allows random writes and the result are transient.

Caters to both data on rest and data in motion
Key Component Groups - Analytical Sandboxes

**Analytical Sandboxes**
- Data Discovery
- Exploratory Data Analysis
- Predictive Modeling

**Process:**
- Develop and test hypothesis.
- Explore data.
- Rapid prototyping.
- Discover data.
- Generate use-cases.

Discover data. Extract value. Transform business.
How much does this painting cost?

Catalog Information:
- The Old Guitarist
- Pablo Picasso
- 1903

Est: Over $100 million

Data cataloging and curation is the key to extract value from data.
## Difference between Data Lake and EDW

<table>
<thead>
<tr>
<th>Data Lake</th>
<th>Data Warehouse</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retains all data i.e. stores data in its raw form.</td>
<td>Refined and transformed data structure for specific business requirement</td>
</tr>
<tr>
<td>Supports all kinds of data</td>
<td>Optimized only for structured data stores</td>
</tr>
<tr>
<td>Supports all kinds of users</td>
<td>Targeted to support operational users</td>
</tr>
<tr>
<td>Adapts to change easily as requirements evolve</td>
<td>Relatively difficult to change as the data is highly structurized</td>
</tr>
<tr>
<td>Provides active cataloging of raw and transformed data</td>
<td>Provides limited metadata capture features</td>
</tr>
</tbody>
</table>
Azure Services for Data Lake Architecture

Analytical Sandboxes
- Data Discovery
- Exploratory Data Analysis
- Predictive Modeling

Real-Time Processing Engine
- Stream Analytics
- Event Hubs

Batch-Processing Engine
- HDInsight
- Data Factory
- SSIS

Processed Data Stores
- SQL Data Warehouse
- DocumentDB
- Azure SQL database
- DataLake
- Storage blob

Raw Data Store
- EL
- SSIS
- Data Factory

Data Cataloging and Curation
- Azure Active Directory
- Access Control

Data Security and Governance
- Azure Active Directory
1. Data Lakes is a new paradigm shift for Big Data Architecture.

2. Data Lakes caters to all kinds of data, stores data in the raw form, caters to spectrum of users and enables faster insights.

3. Meticulous data cataloging and governance is key for successful data lake implementation.

4. Azure offers end-end solution for implementation of data lake architecture in an economical and scalable way.
Demo Architecture - Lambda

Twitter → Event Hubs → Stream Analytics → Storage blob → Diagram
Demo Architecture – Data Catalog

Wine Quality Data

Storage blob

Adventure Works

On-Prem SQL DB

Data Cataloging and Curation
References

Related references for you to expand your knowledge on the subject


---

Azure Portal
http://azure.microsoft.com

Azure Updates
http://azure.microsoft.com/blog/

Microsoft Virtual Academy
aka.ms/mva

Developer Network
msdn.microsoft.com/
Thank you

Follow us online

Twitter: @rpradeepmenon
Email: prmen@microsoft.com
Tell us what you think

Help us shape future events by sharing your valuable feedback.
Link - https://aka.ms/msdevdayfeedback

#MSDevDay