Next Phase of Evolution in Storage Industry: Impact of Machine Learning

Udayan Singh, Head SPE-Storage, Compute & Manageability

30 May 2017
Agenda

1. Digital 5 Forces | Impact on Storage Industry Companies
2. Storage Industry Value Chain & Technology Trends.
3. AI/ML and the area’s of Impact on Storage Industry
4. Use Cases
5. Example Implementations
Digital 5 Forces | Impact on Storage Industry

Digital 5 forces impacting multiple dimensions of the Storage Industry Companies

Business Models
- Pay as you grow cloud storage based models

Products and Services
- Prescriptive analytics-based issue identification and auto-resolution
- Next Gen Products: HyperScalers, Hyper Convergence, AFA, Cloud Storage

Customer Segments
- Movement from only B2B to B2B and B2C

Partner Network
- Cloud Storage through partners
- Mobility connect

Analytics & Big Data

Social Networking

Cloud

Application

AI/Robotics

Mobility

Business Models

Products and Services

Customer Segments

Partner Network

Copyright © 2017 | TATA CONSULTANCY SERVICES
**Storage Industry Value Chain and Storage Technology Trends**

1. **Increased Collaboration** - ISVs, Switch and Storage OEMs
2. **Increased competition** - Cloud Storage Companies
3. **Consolidation** - SSD/HDD Companies
4. **Emergence** Niche Technology

**Hyper Converged Infrastructure**
- Delivering through simplicity

**Cloud Storage**
- Operational costs of maintaining high-growth internal storage infrastructures

**Flash Storage (Hybrid Flash Array/All Flash Array)**
- New interfaces - NVMe

**Software Defined Data Center**
- Simpler, intuitive and intelligent Data Center Management

**Artificial Intelligence / Machine Learning**
- Self-service infrastructure
- active cognition and analytics-based automation
AI/ML and the area’s of Impact on Storage Industry

**Artificial Intelligence**
- Robotics
- Sensory Perception
- Natural Language Processing
- Machine Learning
- Speech Recognition
- Image Analysis
- Natural Language Generation
- Deep Learning
- Knowledge Engineering
- Cognition

**Impact on Storage**
- PDLC: Requirements Gathering; Development; Quality Assurance; Sustenance
  - Time To Market
  - Cost Reduction
- Customer Support Services
  - User Experience
  - Improved Quality
- Storage Admin Software: Provisioning; Management; Troubleshooting
  - User Experience
  - Cost Savings
- End User Storage Software Products: File Storage;
  - User Experience
  - Improved Quality
- Storage Devices: SSD, HDD
  - Improved Quality (endurance, performance)

Copyright © 2017 | TATA CONSULTANCY SERVICES
Use Cases: Data Center Storage Software

**Machine Learning**

- Cloud Storage Management
- Data Protection
- Storage Monitoring & Management
- Storage Analytics
- Customer Support Software
- Performance Management
- System Health Monitoring
- Storage Software Platform

**Executing Task**

- Provision
- Monitoring
- Manage
- Issue Identification
- Issue Resolution
- Optimization
- Demand Prediction

**Learning by experience**

- Automated Update
- Manual Analysis
- Manual Performance Tuning
- Automated Data Storage Optimization

**Improving Performance**

- Proactive Management
- 1 Admin per 4 Pb/500 Servers

**Current Stage**

- Automated Root Cause Analysis & Corrective Action
- Intelligent Performance Tuning
- Intelligent Data Storage Optimization

**Future Stage**

- Intelligent Proactive Error detection and Update
- Self Managed Data Center
  - Customer Support
  - Storage Management Software
- 1 Admin per 20 Pb/10,000 Servers

Copyright © 2017 | TATA CONSULTANCY SERVICES
Capture systems requirements through spoken communication and converting it into PRD (Chatbots + NLP).
Example 1: Storage Performance

Machine Learning Method

Selection of Right Data Set

Data

Quality Check of Data

Pattern Matching

Analysis of the Data

Features:
- Total IOPS
- Cache Hit Ratio (%)
- Disk Response Time
- Disk Utilization
- [%age Write Miss
- Data Transfer Rate (MB/s)
- No. of Active Network Ports
- Spindle Count
- RAID Configuration
- Disk Type]

Models:
- MultiVariate Regression Analysis (Causal)
- Time Series Analysis
- Final: Multivariate Time Series (VAR Model)

Prediction:
- Average Response Time (Latency)

Similarly for:
- Capacity
- CPU Failure
- Disk Failure

Insight Generation
Example 2: Prediction of Database Response Time

**Business Problem**

To predict the database response time and facing challenges as below
- To create predictive models which predict DB response time.
- To predict 1-hour-later SLA violation of DB response time with 90% of recall and 70% of precision.
- The predictive models predict the Probabilistic distribution of DB response time at the time of 1 hour later.

**Solution**

- Creation of Lab environment to load on DB and measurement of various parameters for model building
- Predictive models to:
  - Predict number of queries on the DB after 1 hour
  - DB response time after 1 hour
- To identify load on the system after 1 hour and corrective action planning
- Techniques - Unobserved component models and Poisson Regression

**Benefits**

- All models are having Mean Absolute Percentage Error (MAPE) less than 5%
- Precision & recall rate for 1 hour later prediction was 60% and 45% respectively.
Thank You