Who am I?
... and what am I doing here?

Son Nguyen

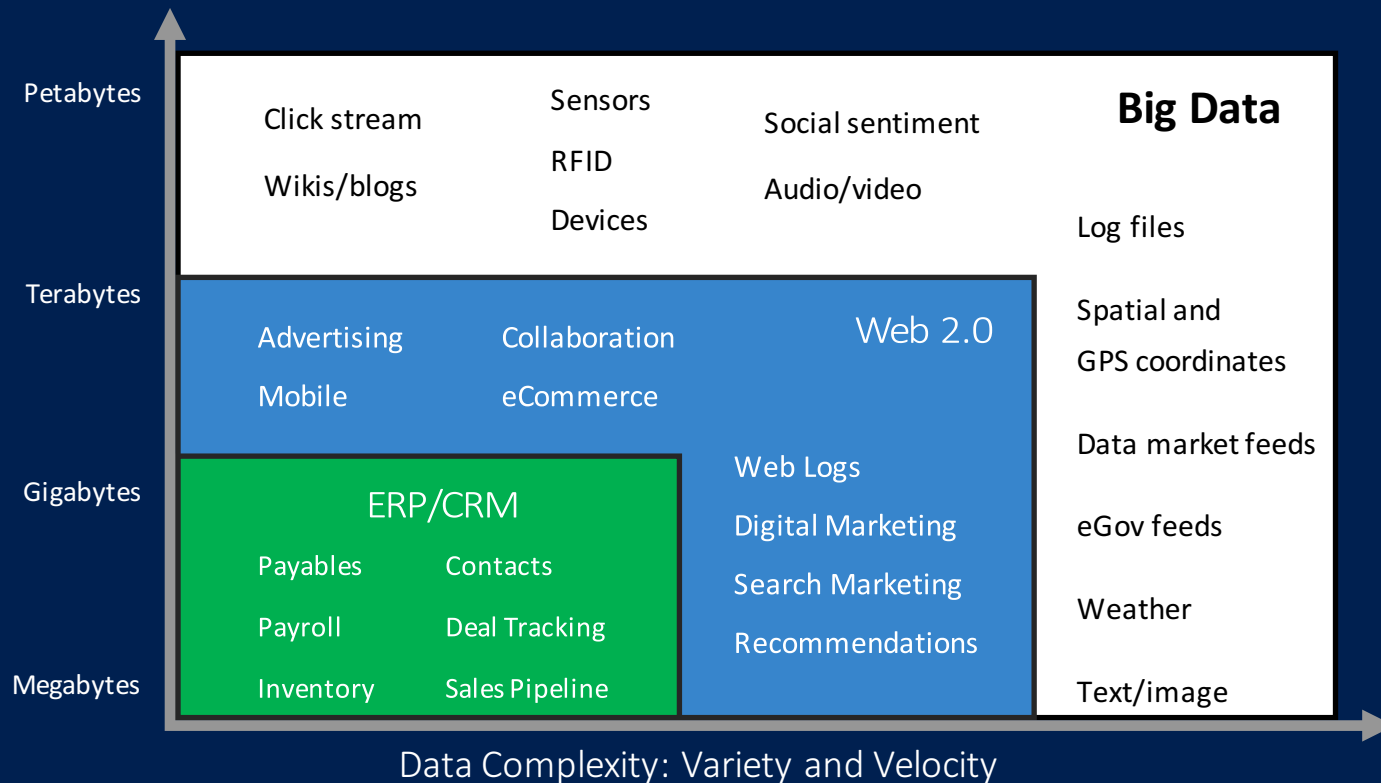- Cloud Consultant at FPT Software
- 5 years' experience in Microsoft Azure
- Provide Cloud & IoT solution to JP, US and EU customers
- Blog: http://sonn2.com
- Email: nguyens.developer@outlook.com
- LinkedIn: https://www.linkedin.com/in/nnson4nguyen

I'm here to share with you my experience in how to use Azure to simplify big data solution.
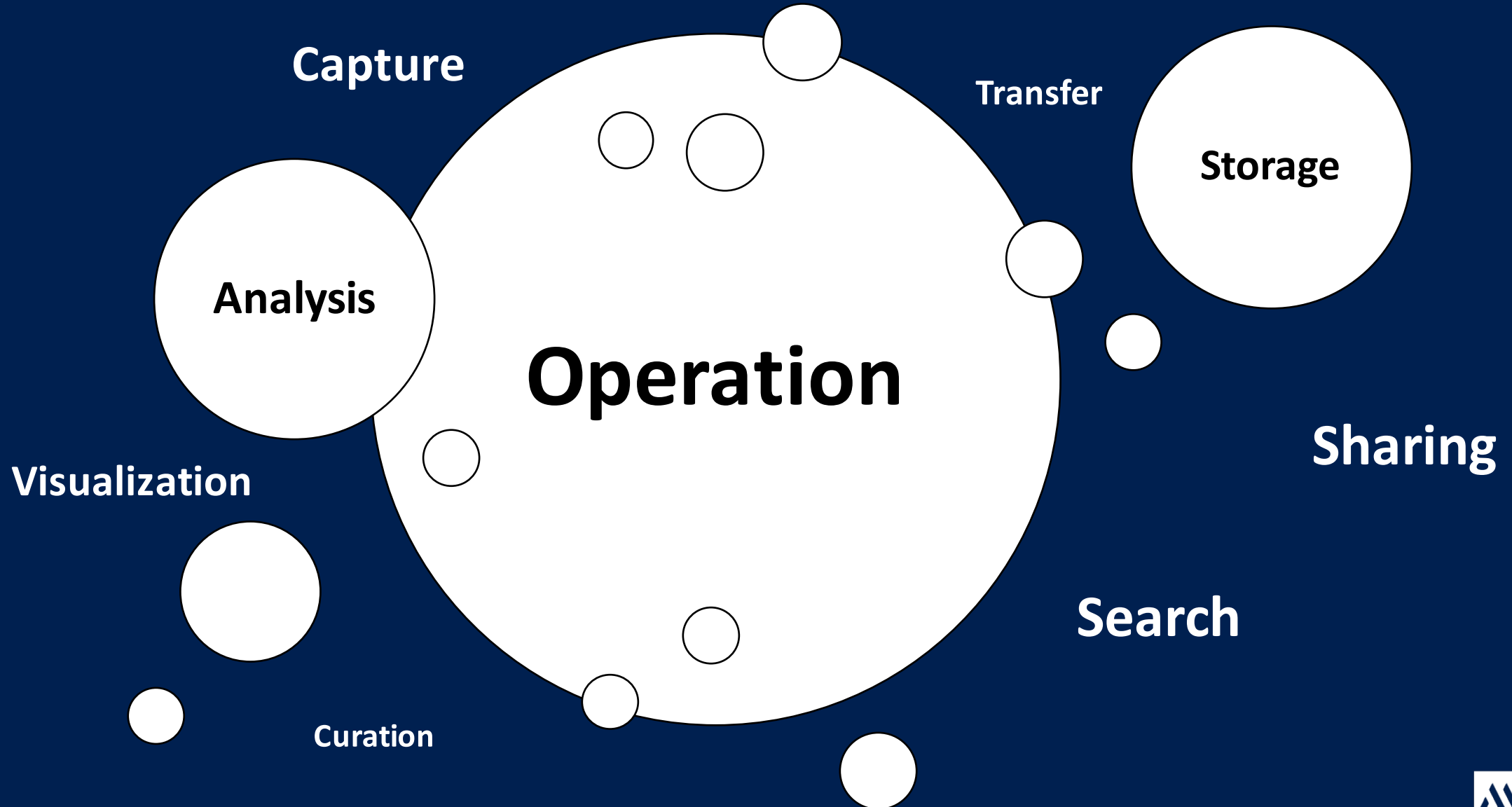
VIETNAM WEB SUMMIT

What is big data?

*"Big data is the term for collection of data sets so **large** and **complex** that it becomes **difficult to process** using on-hand database management tools or traditional data processing applications."*

Wikipedia.com

Petabytes | Click stream | Sensors | Social sentiment | **Big Data**
Wikis/blogs | RFID | Audio/video | Log files
Devices

Terabytes | Advertising | Collaboration | Web 2.0 | Spatial and GPS coordinates
Mobile | eCommerce | | Data market feeds

Gigabytes | ERP/CRM | Web Logs | eGov feeds
Payables | Contacts | Digital Marketing | Weather
Payroll | Deal Tracking | Search Marketing
Megabytes | Inventory | Sales Pipeline | Recommendations | Text/image

Data Complexity: Variety and Velocity

VIETNAM WEB SUMMIT

Big data challenges

Capture

Transfer

Storage

Analysis

Operation

Sharing
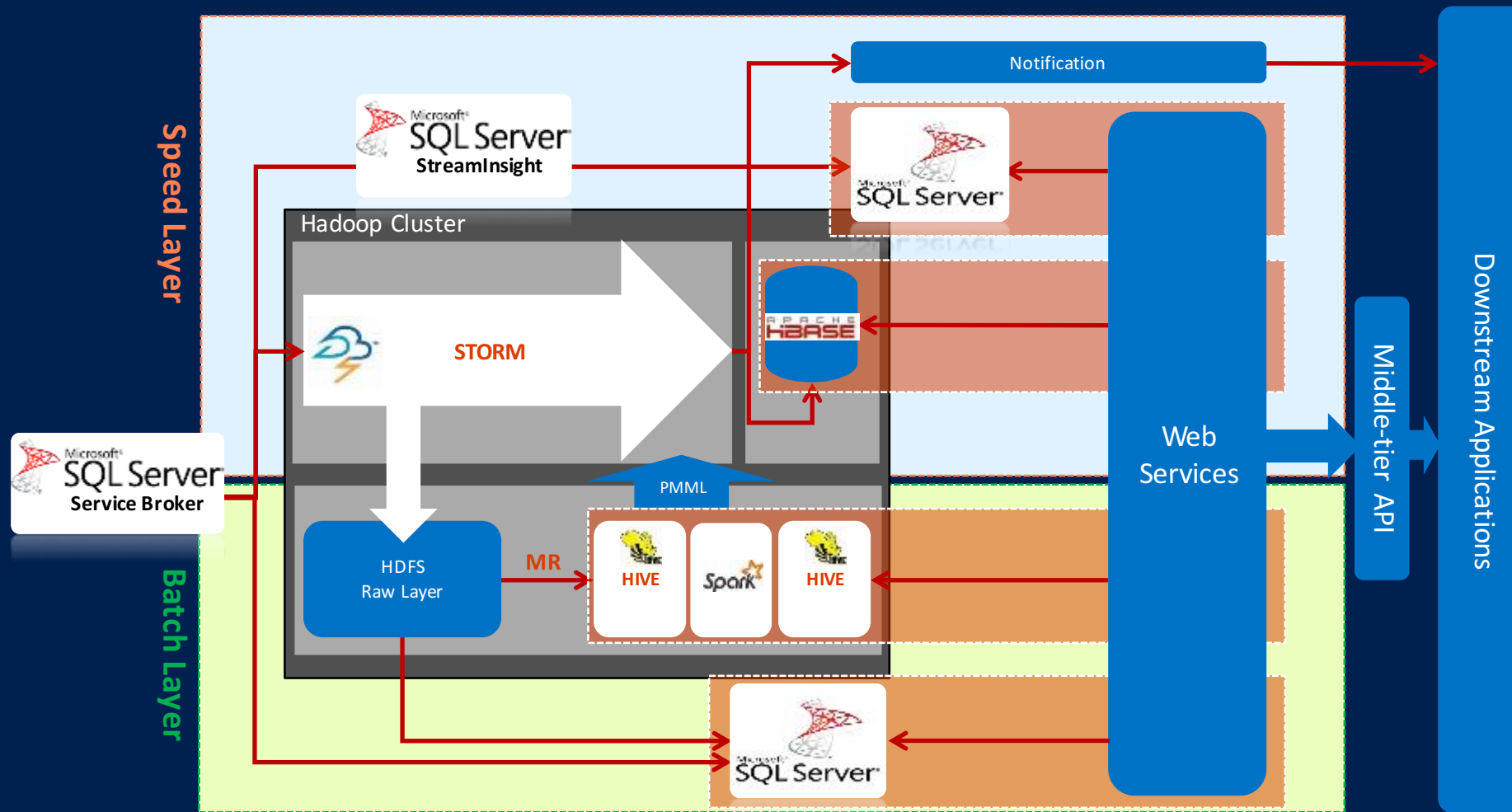
Visualization

Search

Curation

VIETNAM WEB SUMMIT

Business case
... fleet management system

- Vehicle tracking
- Health and safety tracking
- Fuel and speed management
- Sales order transmission
- Route planning
- Driver management
- Vehicle diagnostics
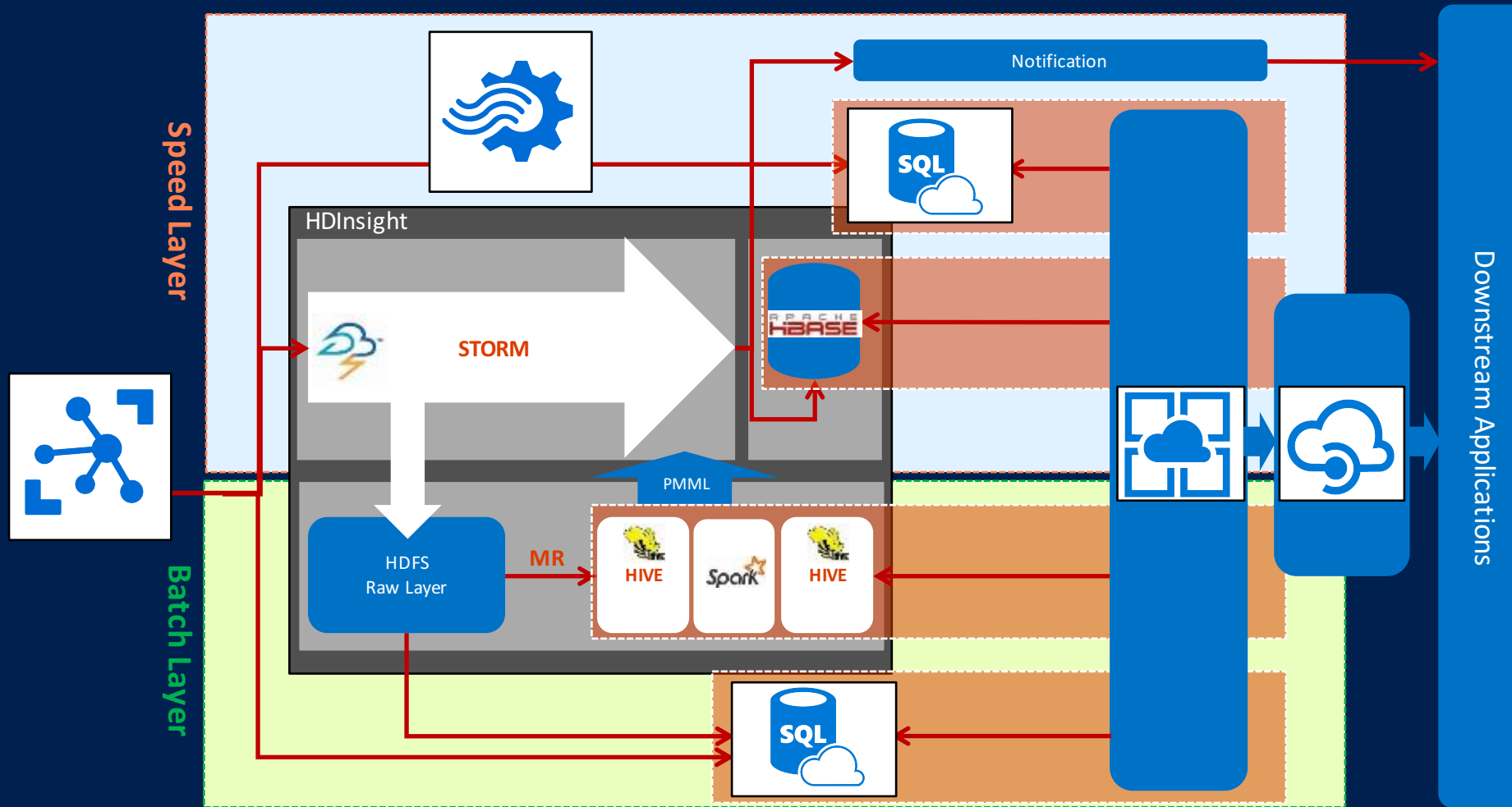
On-premises solution

Buildable but have many technical debts need to be resolved.

Top 5 challenges
- Scalability
- Cost
- Operation
- High availability
- Security

Enhancement solution on Azure

Speed Layer

Batch Layer

HDInsight

STORM

HDFS
Raw Layer

MR

HIVE   Spark   HIVE

PMML

APACHE HBASE

Notification

SQL

SQL

Downstream Applications

Enhancements

- System is deployed on Azure.
- Use IoT Hub, Stream Analytics services
- Use Azure SQL Elastic Database

and more ...

Top 5 challenges

- Scalability
- Cost
- Operation
- High availability
- Security

VIETNAM
WEB
SUMMIT

Azure big data eco-system

Information Management	Big Data Store	Machine Learning and Analytics	Intelligent	Dashboard and Visualization
Data Factory	Data Lake Store	Machine Learning	Cognitive Services	PowerBI
Data Catalog	SQL Data Warehouse	Data Lake Analytics	Bot Framework	
Event Hub	DocumentDB	HDInsight (Hadoop, Spark)	Cortana	
IoT Hub	SQL Database	Stream Analytics		

THANK YOU

VIETNAM
WEB
SUMMIT