

# Building Business Data Lake on AWS



Manish Bhatt – Head of Big Data, Capgemini Australia
AWS Big Data Solution Day, HK
31 October 2017

- 1. Big data and Cloud
- 2. AWS Big Data Customer Stories
- 3. Some considerations for building Big Data solutions

- 1. Big data and Cloud
- 2. AWS Big Data Customer Stories
- 3. Some considerations for building Big Data solution

### Big data and Cloud - No Brainer



The Data Explosion is transformational – and can unlock answers to critical business needs and opportunities



"This one's a real no-brainer. That's why I asked for you specifically."

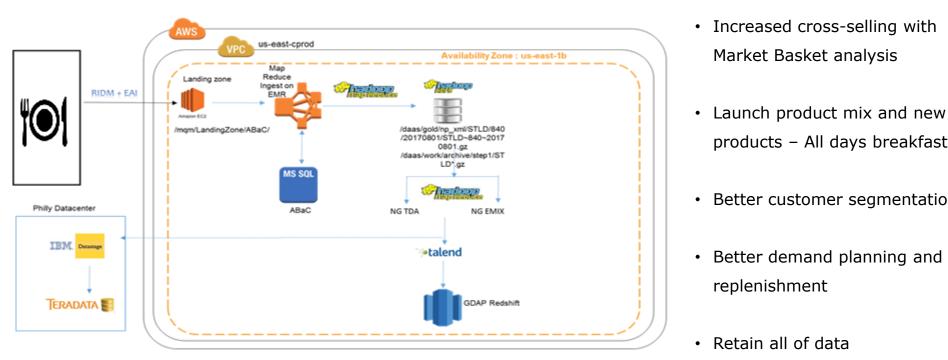
- 1. Big data and Cloud
- 2. AWS Big Data Customer Stories
- 3. Some considerations for building Big Data solution

### Cloud Migration for a leading Quick Service Restaurant



#### **Challenges**

- Support huge jump in data volume due to upgrade in POS systems in stores - which provided ability to send order level product items data.
- Existing infrastructure was inflexible and costly



**Benefits** 

- products All days breakfast
- Better customer segmentation

Key **Project Features** 

\$5M per year incremental cost saving

 $\sim$ 50k+ files daily

1TB everyday 100TB in S3

**Solution** 

130M **Transactions** per day

Comprehensive Metadata Management

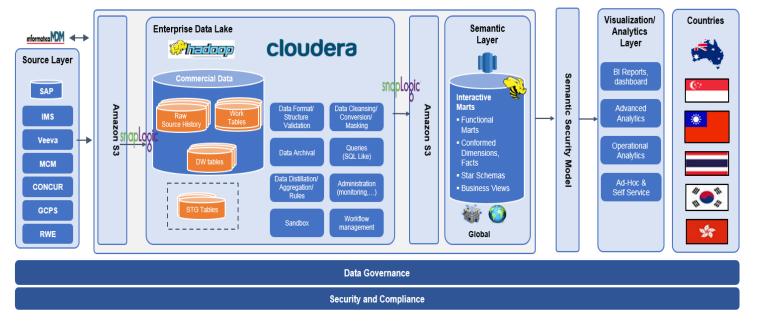
### Global Data Lake for a major life science firm



#### **Challenges**

Build end-to-end solution that facilitates integration of data from various global regions and support self service capabilities, reporting and analytics on multichannel marketing and sales data.

#### Solution



#### **Benefits**

- Build a single ,
   consistent, and logical
   view of commercial data
   across JAPAC region
- Standardization of all KPI across regions
- More targeted marketing
- Lower cost

Key Project Features

Global Deployment 40+ Data sources

Standardization of all KPI across regions

Self-service reporting DWH

### Cloud Migration for a leading telecommunication company



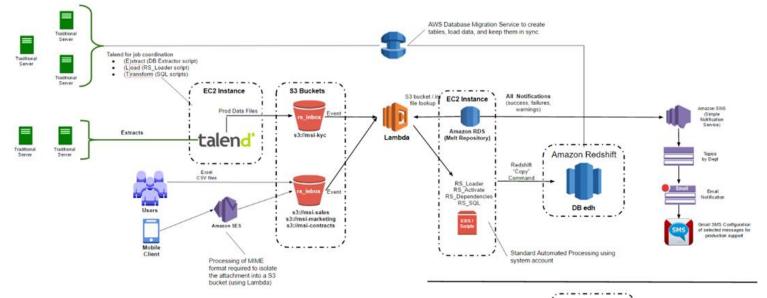
#### **Challenges**

- Business was looking for an agile solution to reduce time for reporting and insight after a very long ERP implementation cycle store to provide one view of historical and new erp data
- Engage business in an agile information delivery system

Key Project Features

70 Oracle schemas migrated

#### **Solution**



- **Benefits**
- Enterprise data store for business continuity which provided historical as well as new data
- · Data as a service function
- Data driven campaigns,
   Install base analytics,
   Field inventory accuracy

Data as a Comprehensive service Metadata Management

+ableau

EC2 Instance

8 TB Data in

Redshift

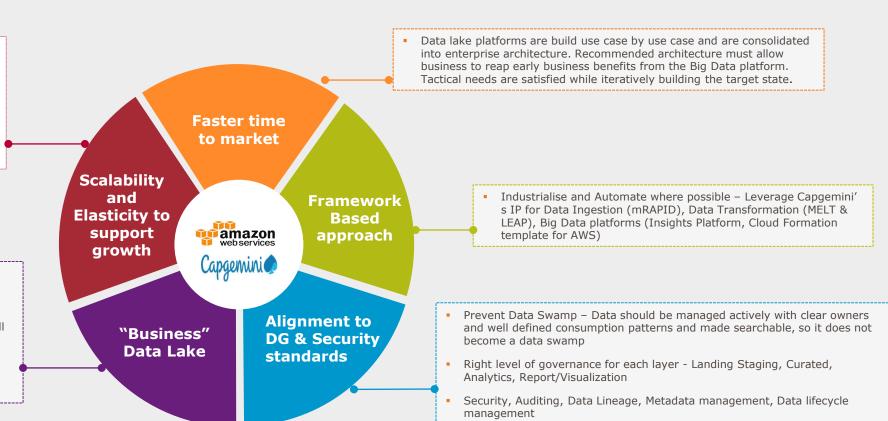
- 1. Big data and Cloud
- 2. AWS Big Data Customer Stories
- 3. Some considerations for building Big Data solutions

### Some considerations for building Big Data solutions



- Platform with established elastic capabilities public vs.
   Hybrid
- Proprietary versus Open source
- Tools and technology can scale to handle volume for organic as well external data volume growth
- Tools to handle future multi-structured data sources in real time, micro-batch and batch mode

- Follow Best Practices for Data Lake Store everything as-is and store it cheaply, Distil on demand, Govern where it matters
- Provide a simple, business-oriented information catalog of all data assets available
- Metadata Management
- Self-service ETL and reporting capabilities





## Innovation requires building and deploying new technologies









manish.a.bhatt@capgemini.com