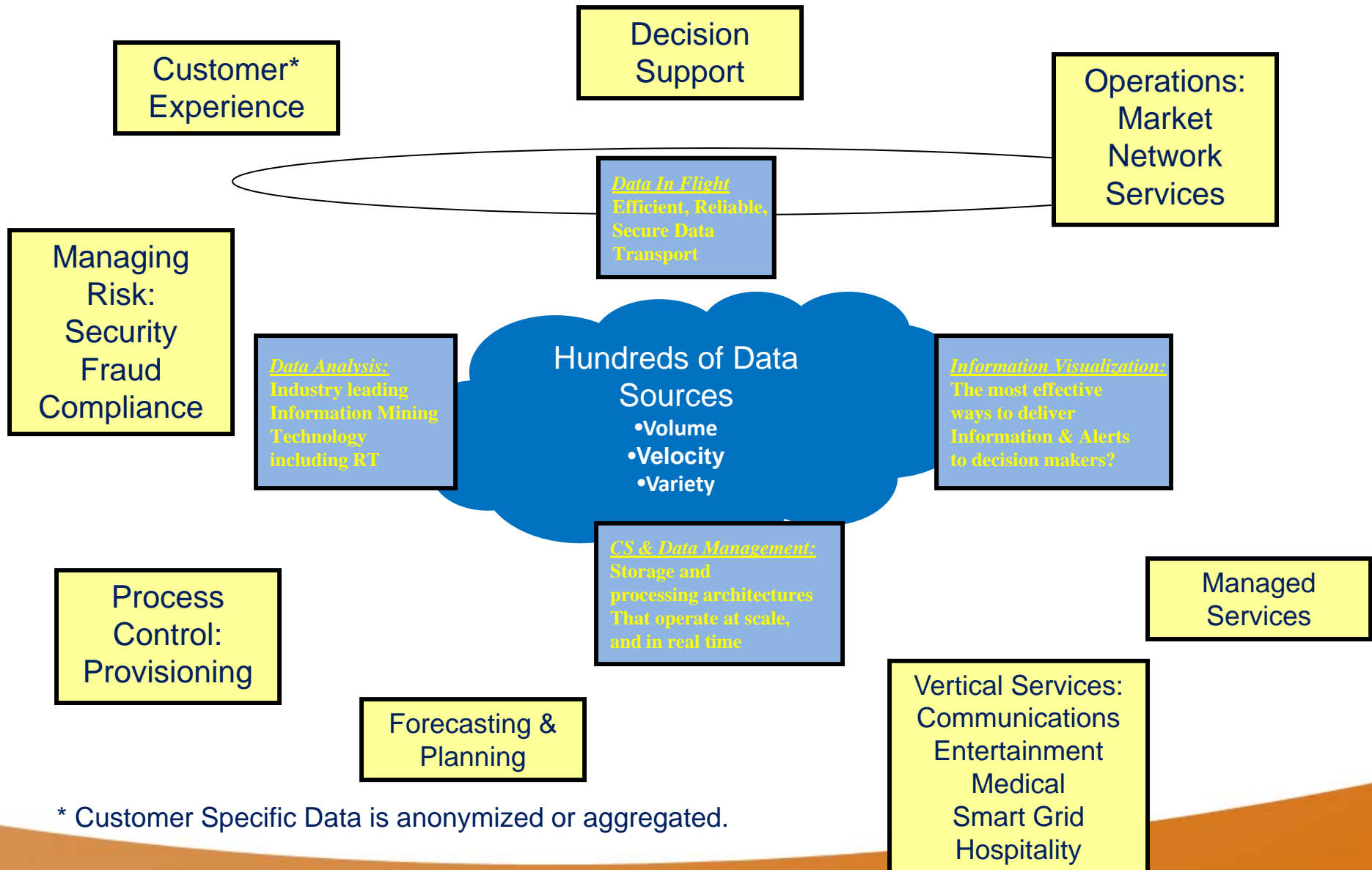


Data Mining and Anomaly Detection

David Belanger, PhD
Senior Research Fellow,
Stevens Institute of Technology
AT&T Labs Chief Scientist (retired)



Getting More Value from Data

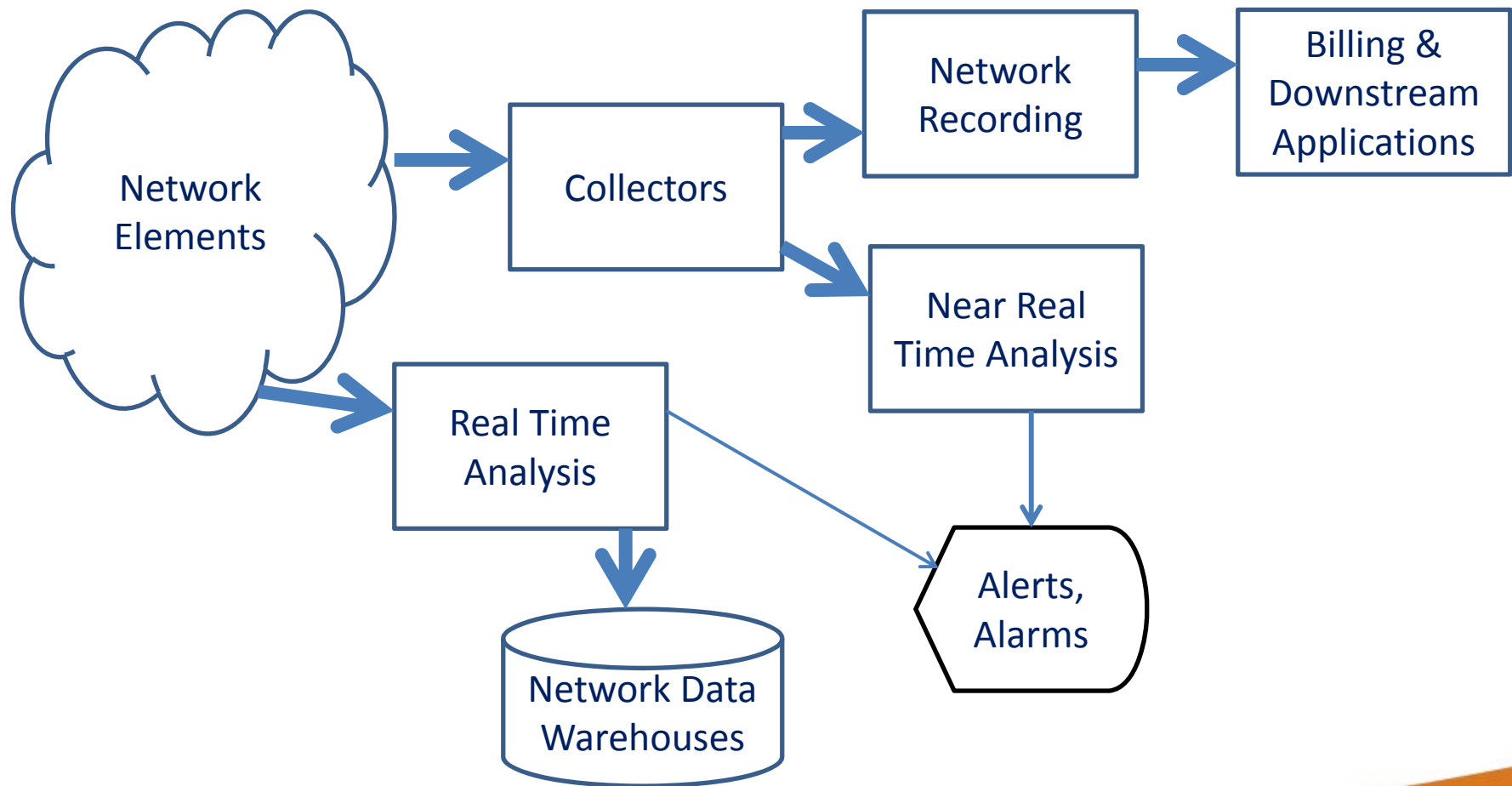


* Customer Specific Data is anonymized or aggregated.

Sources of Data - Communications

- **Network**
 - **Circuit Switch: Signaling**
 - **IP: Netflow, Packets**
 - **Mobility, Wireless: RAN, RNC, Routers**
 - **Services**
 - **Call Detail Records**
 - **Usage**
 - **Customers**
 - **Billing**
 - **Customer Service**
 - **Surveys**
-
- **Challenges – Alert/Alarm Systems**
 - **Scale**
 - **Integrity**
 - **Security, Privacy**
 - **Efficiency**

Sources of Data - Communications



Call Detail Record Data Analysis

Example Fields:

- **Initiating Number**
- **Terminating Number**
- **Start Time and Date**
- **Call Duration**
- **Billing Number**
- **Phone Exchange**
- **Call Results – Busy, Failed, etc.**
- **Call Type – Voice, SMS, etc.**

Example Uses of Network Recording Data:

- **Billing**
- **Fraud**
- **Forecasting**
- **Media Event Voting**
- **Network Operations**
- **Customer Experience**
- **Non-syndromic**
- **Many more**

Call Detail Record Data Analysis

Example Analytic Techniques - Fraud:

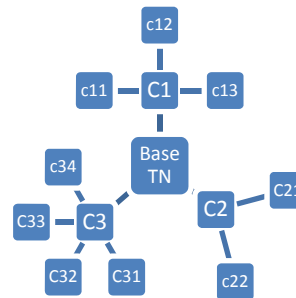
- Thresholds

Call Duration > X

- Signatures

Day of Week	Region of World
Time of Day	% Countries Called
Call Duration	

- Communities of Interest



Application of Data

Metric

	Aggregate	Individual	Relational
Input	Aggregate	Wire Center, Zip, Cell Tower	
	Individual	Thresholds	Signatures
	Relational	Graph Measures	Col
	Crowd	Twitter	Mark the Spot
			Social Network