# Breach Prediction Model

This project will be a collaborative effort between Analytics, Data Science and Product Tech to provide predictive capability related to breach at a shipment, asset, node, hop and lane levels enabling proactive solutions by operations.

**Business Context & Objective**

Today Ops and CT teams rely heavily on data to take actions to curb breach. However, the data and metrics visibility today stands at D-2 or D-3 level. There are currently challenges for metric visibility at D-0 level, which misses the point of proactive solutions.

The alternate approach is to provide visibility in a preemptive fashion which can help ground Ops to be prepared for the upcoming days.

This framework and model for Breach Prediction will try to effectively and predictably solve for possible upcoming breaches.

In the future as in when real time data becomes available for ingestion, this solution will become much more powerful providing vis

**Expected Outcome**

* Visibility of probable breach as a shipment travels through the supply chain
* Recommended actions for the node to take corrective actions
* Reduction of breach leading to improvement in L0 metrics

**Audience**

This model will be consumed by Control Tower, Operations and Asset Heads (multiple levels)

# Problem Statement

* Predict the Customer Promise Date (CPD) **breach probability score** for a given shipment at any **node** of shipment life cycle
* **Interpretation** of Nodal Breach Probability as either on a quantitative or qualitative scale (H-M-L)
* **Recommended Actions** and possible Breach reduction **goodness** achievable

## What is Node?

These are the different events that happen for a shipment/unit at different assets of the supply chain.

**Events:**

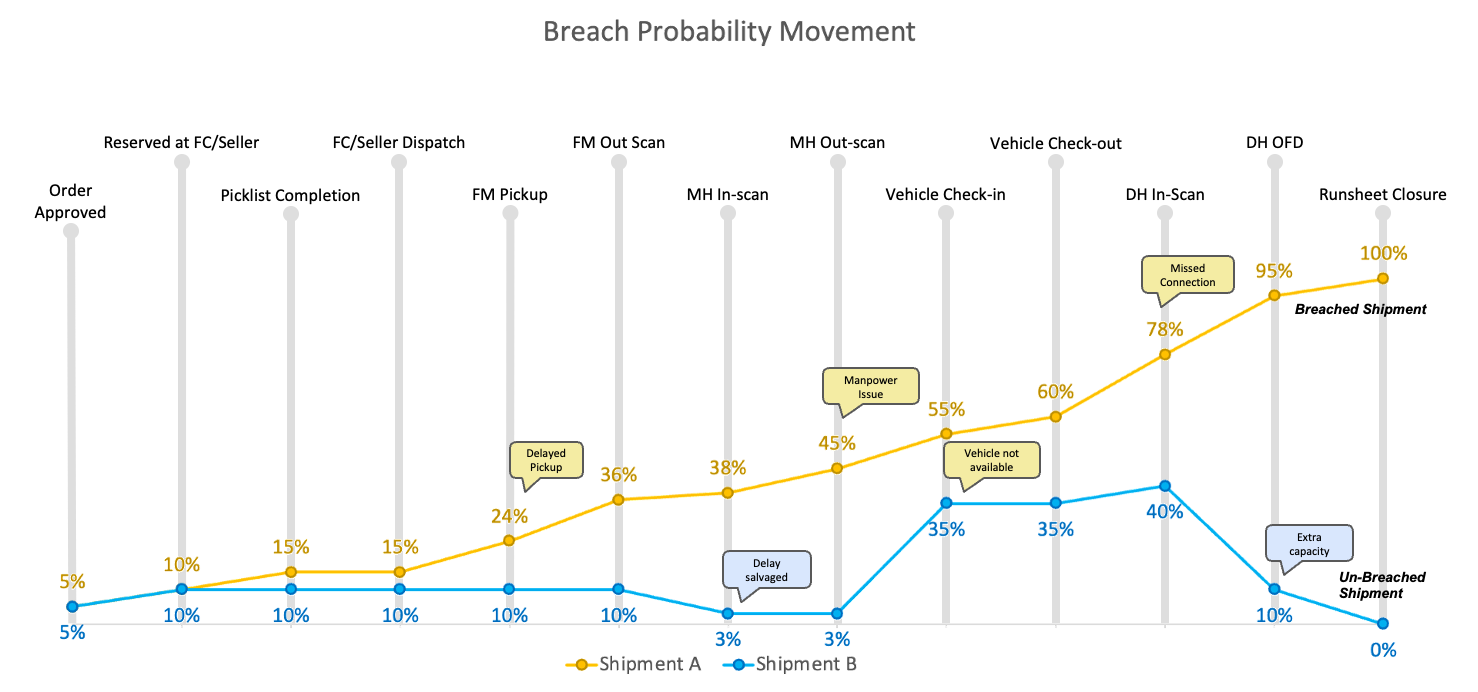
* Reserved at FC/Seller
* Picklist Creation & Completion at FC
* FC/Seller Dispatch
* FM Pickup from Seller - Seller
* FM Out-Scan from Pickuphub - Seller
* MH In-Scan - **Multiple**
* MH Out-Scan - **Multiple**
* Vehicle CheckIn - **Multiple**
* Vehicle Checkout - **Multiple**
* DH-In
* DH- OFD
* Runsheet Closure

We want to predict the **breach probability** for each of these events.

## Input Data Availability at each node

1. **Reservation at FC/Seller**
   1. Order datetime,Reservation datetime (Actual & Planned) - check for delay in reservation of the order.
   2. SLA promised
   3. Entire network plan - assets/hops that has allocated
   4. Planned process and travel time at each leg
   5. Historical Lane Breach (Historic data - Functional Breach/ CPD Breach ) (last 1-2 weeks)
      1. Source Breaches (MotherHub)
      2. Destination Breaches (Delivery Hub)
   6. L/Z/N (Local, Zonal, National) Lane Information
   7. Planned Datetime at each node - so that we can compare the planned time v/s the assets average processing time in past 1-2 weeks
   8. Number of hops (Planning Data) - (correlation ?)
   9. Number of days/hours left for promise
   10. Stress Signals
       1. Historical 1-2 weeks data - HighLandings - DH Stress/FC Stress
   11. Each Node’s performance index - average processing time v/s that of last 7 days
   12. Holidays at FCs and MHs
   13. Consider Functional Breaches( Historic Data)
2. **FC/Seller Dispatch**
   1. Dispatch Datetime (Actual & Planned) - FC/Seller Processing time - to check for delay/deviation from ideal scenario
   2. Reservation datetime - pickup2dispatch time - dispatch by datetime
   3. All previously listed data
3. **FM Pickup DateTime from Seller**
   1. Pickup requested ,Pickup datetime (Actual & Planned)
   2. Total time exhausted
   3. All previously listed data
4. **FM Out-Scan**
   1. Outscan datetime (Actual & Planned), FM processing time - to check for delay
   2. Total time exhausted
   3. All previously listed data
5. **MH In-Scan**
   1. Delays at previous assets (Seller/FM/FC)
   2. Delay in previous MH hops
   3. All previously listed data
6. **MH Out-Scan**
   1. Delay in in-scan at MH
   2. All previously listed data and delays
7. **Vehicle Checkin/Checkout**
   1. In-Transit delays by LHs/Vehicle placement delay/TAT at source (when vehicle did not leave on time)
   2. LH has to reach before MH connection - Mh Tasks(sort,resort,cross-dock)
   3. All previously listed data and delays
8. **DH In**
   1. Vehicle arrival v/s Planned Arrival - whether there was multirun/early landing
   2. DH Stress level at shipment landing date
   3. All previously listed data and delays
   4. Manpower/processing - get attendance
   5. check for holidays and its effect. (External Parameters)
9. **DH - ofd**
   1. Runsheet creation
   2. Runsheet assignment
   3. All previously listed data and delays
10. **Runsheet closure**
    1. Applicable for cases when out for delivery is before the CPD and shipment is undelivered on that date

# Model Approach



1. **Model Development**

Based on historical data, simulate sample observations for shipment journey till every node.

**Develop individual models for every node, with incrementally introducing new features** reflecting the situation in the node reached.

**2) Threshold Probability**

For every node, based on historical probability distribution, arrive at threshold values beyond which it significantly increases the shipment breach probability. For example, a breach probability at MH of 32% leads to a 80% CPD breach probability, hence, the threshold for MH node will be 32%.

**3) Model Validation**

Model needs to be validated based on threshold probabilities at every node

**4) Set Triggers and Alerts at shipment level**

In case threshold probability is crossed, raise alert and recommend actions for subsequent node

**4. a) How the alerts gets consumed**

**Example; for event= ‘reservation’**

|  |  | Breach Probability | | |
| --- | --- | --- | --- | --- |
|  |  | <10% | 10-80% | >80% |
| CPD days left | <=1 day | Action 2 |  |  |
| 2 |  |  |  |
| 3-5 |  |  |  |
| >5 | No Action | No Action | Action 1 |

**5) Set Triggers and Alerts at Node (hops and lane) level**

At every node, (and subsequently at lanes and hops) compute % shipment crossing threshold breach probability. Define control limit for raising alert if x% shipments have likelihood of breaching.

**6) Recommendation Engine**

Triggers at node level lead to invocation of Recommendation Engine, which will recommend possible actions for reducing breach probability. Every action should have a goodness measure for breach probability reduction.

**7) Overall Breach prediction**

Based on shipment status of particular CPD, compute probable FK breach % for the current week. This will be computed for orders which have been placed.

**Model Deployment**

1. Live machine learning model, which will refresh with additional data points added, with capability of ingesting user inputs
2. Recommendation engine on platform with recommended actions at stress points
3. Hourly and daily reports on stressed assets and lanes

**Platform Development**

*(In this section, we are outlining the requirements from Analytics POV. Detailed platform development related approach to be owned and updated by Product/Tech)*

A robust platform is required which can be accessed by operations across locations, Pan India providing predictive visibility for breach probability and recommended actions.

**Desired Feature List**

1. Create a map of the entire supply chain with source, destination, facilities, lanes and hops
2. Zoom in and zoom out feature
3. Highlight said facilities, lanes and hops when significant shipments cross threshold breach probability.
4. Allow users to input certain values and flags which will be used in model refresh, e.g. - embargo hubs, upload hub capacity, etc.
5. Export capability of datasets and reports
6. Call to Action for a) recording user action, b) link to apps (for actionability), e.g. - stressed hub needs to create an optimized picklist, hence a link to an app allowing Automated/Prioritized Pick List creation based on a defined objective

## Project Management

**High level Tasks and RACI Framework**

| **Broad Steps** | **Analytics** | **Data Science** | **Product / Tech** | **CT** |
| --- | --- | --- | --- | --- |
| Data availability | R, C | I | R, A | C, I |
| Rule Book for Actions | R | C | C, I | R, A |
| Model development | R, A | R, C | C, I | C |
| Model validation | R, A | R | I | C |
| Model deployment | R, A | R | C, I | C |
| Platform development | C, I | C, I | R, A | C, I |

*R = Responsible, A = Accountable, C = Consult, I = Informed*

**Owners for data availability**

| **Available in FDP** | **Available in System but not in FDP** | **Not Available** |
| --- | --- | --- |
| Analytics | Product/Tech | Product Tech |

## Data Ingestion Plan and Tracker

This section will track progress on ingestion of new data and will be owned by Engineering

| **Data required** | **Assigned to** | **ETA** | **Comments** |
| --- | --- | --- | --- |
| Planner data | Siddharth Barua | July end |  |
| Capacity data | Gajendra Likhitkar |  |  |
| Live triggers for Node change |  |  |  |
|  |  |  |  |
|  |  |  |  |

## 

## Recommendations and Actions Playbook

This section will outline the recommendations and possible actions taken by CT and Ops at each level of prediction. This section will be created jointly with the CT team.

| **Node** | **Breach Probability** | **Insights** | **Actions** | **Expected Goodness** |
| --- | --- | --- | --- | --- |
| 1. Order reservation | Network breach | Assets likely to breach | Plan capacity preemptively in stressed assets in the network |  |
| 2. Order Shipped | Seller / FC breach | Revised TAT for forward assets |  |  |
| 3. |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

## 

## Phase Wise Plan

|  | **Phase 0** | **Phase 1** | **Phase 2** | **Phase 3** |
| --- | --- | --- | --- | --- |
| *POC Phase to gather learnings from model development and ease of adoption* | *Test first few nodes of prediction and learn to improve model accuracy* | *Implement learnings from previous phases and work towards creation of an Analytics Product to reduce CPD Breach* | *Drive implementation and adoption and measure goodness* |
| **Product Engineering** | **Data:** Network plan at a shipment level | **Data:** 1. Planner data,  2. Asset wise Capacity data | **Data:** Real-time triggers for movement of shipment |  |
| **Analytics & DS** | **Model:** 1. Historical inherent Asset Breach probability model  2. Network breach prediction at a shipment level  **Insights:** Stressed hub based on all orders placed and orders moving in the supply chain | **Model:** 1. Breach probability for all nodes  2. Threshold probability for all nodes  **Insights:** Heuristic triggers | **Model:** 1. Scale up plan for Large and Hyperlocal | **Model:** Feedback based model refresh  **Insights:** Measure on ground goodness |
| **CT & Ops** | **Rulebook:** Possible Actions to be taken for stressed hubs and measure of goodness (analytics to assist)  **Live trial:** North Zone network testing | **Rulebook:** Possible Actions for further stages  **Feedback:** on prediction accuracy and platform usage | **Rulebook:** Possible Actions  Feedback:  **Adoption:** Drive Live Trial with 100% compliance  **Feedback:** For model tweaking and Platform Usage | **Adoption:** Drive org wide adoption |
| **Product Tech** |  | **Platform:** MVP for Breach Prediction Platform with visibility and recommended actions, allowing user inputs and feedback | **Platform:** Visualizations and Map features  Reporting features | **Platform:** Roll-out platform |
| **ETA** | **Aug’22 end** | **Dec’22 end** | **Mar’23 end** | **Jun’23 end** |

1. For **Phase** 1, we are planning to initially create the model for **one SC pipeline** and then expand to others eg: large,11\*11 etc
2. We will only work on **boundary** conditions for each asset for phase 1 and later enrich the model with other detailed conditions and events.
3. We need to consider other events where shipment gets lost on the way eg: **BRSNR** or **Misroute** cases where the breach probability in such cases increases exponentially or 100% breach
4. **External Factors** - Zone/Regional (Later stages) can be considered in further stages
5. External shipment load/11\*11/grocery load/stress at different assets
6. Prioritization happening based on **plus/non plus** customers

**Phase 1**: Visibility

**Phase 2**: Recommendation

**Phase 3**: Actions

**Appendix**

# Definitions and Prerequisites

* **CPD - Customer Promise Date** - The date that is promised to the customer for delivery at the time of placing an order.
  + CPD Breach - When we fail to deliver by the promised date
* **LPD - Logistic Promise Date** - The date that logistics partners promises to deliver the order by
* **Functional Breach -** Assets breaching their latest possible handover time (LPHT)
* **EOB - Extent of Breach** - As % of promised shipments
  + Denominator = All promised shipments.
  + Numerator = All shipments which did not have a single attempt by CPD+1 closing on CD status or delivered.
* **Assets** - Different facilities that the shipment passes through.
  + e.g - Fulfillment center, MotherHub, Delivery Hub etc.
* **FC - Fulfillment Center** -> The different ways that we can have inbounds fresh goods, IWIT, customer returns and seller rejects
* **MH - MotherHub** -> This is the place where a shipment arrives once it has been dispatched from an FC. Most consignments move to one or more intermediate Mother Hubs before reaching the DH (Delivery Hub).
* **TC - Transport Center** -> The intermediate MHs are referred to as TC (Transport Centers).
* **FM- First Mile** -> Responsible for picking up all shipments from sellers and depositing them at the pickup hubs as well as returning the shipments to the sellers from the pickup hubs.
* **LM - Last Mile** -> Responsible for transport of an item from a delivery hub to the end customer or from the customer back to the pickup hub in case of returns.
* **LH - Line Haul** -> The path between the Source MH and one or more Transport Centers / Delivery Hub is called a Line Haul. A shipment could go through a single MH or it could go through multiple MHs or through a combination of an MH and a TC (transport center).



## List of Important Documents

| **Document** | **Link** |
| --- | --- |
| Supply Chain Overview | [eKart Overview.pdf](https://drive.google.com/file/d/1VkWAZEzDgQ9PXVP6mY24ZeCRgP9eZeqM/view) |
| Assets Functioning & Processes | [Ekart Analytics](https://docs.google.com/document/d/1RSVYR8TdxFpnRqslhu-m23GZK7Z0JSEKBFHPQ2LjoUs/edit) |
| Assets Breach Definitions | [SFC-Breach](https://docs.google.com/document/d/1l568ERKUtgwDNqSX-oom3K7HbvtDnbgO4IcEx7wgvtw/edit) |
| CPD Breach Attribution Logic | [CPD Breach: Attribution Logic](https://docs.google.com/document/d/1W-19uTz2Q-i3AnWEOB58sJzsJZRtVJwxoY9bbKrsvEw/edit#) |
| Unified Logic | [Unified Breach - Rollout](https://docs.google.com/presentation/d/1vr__AKweeOld112T7cJOCFdq4WKHJBsO2yYRmhJOlYw/edit#slide=id.g103e4a926d8_0_132) |
| Promise Definitions | [Promise breach definition document](https://docs.google.com/document/d/12Q2IdjqUcfJ8bqMXB5ySGYoRCd4dHe7SPM5I1HHGLAI/edit) |

**Drivers of Breach:**

* Functional Failures
* External Factors
* Spillover from previous functional failures
* Tech outages

## Sub projects:

* **Relationship and conversion metric for Breach - Divya / Aakansha / Aman**
  + Conversion metric for Breach to NPS - Divya
  + Conversion metric for EOB to NPS - Aakansha
  + Conversion metric for Fake attempts to NPS - Aman
* **Conversion metric for functional failures to Breach - Aakansha**
  + Conversion metric for DH Highlanding to Breach
  + Conversion metric for
* **Breach Impact on Customers**
  + Customer level breach mapping to orders
  + Segmentation of customers based on % of breach, breach x ticket size, breach x BU
  + Mapping NPS, RTO, lapsed customer against each segment
  + Recommendations for each segment
* **Functional Breach of Assets leading to CPD Breach – Divya / Aakansha**
  + Asset wise identification of Functional Breaches
  + Asset wise identification of task wise functional failure
  + Asset wise identification of design time buffer
  + Model to predict breach probability of shipment based on functional failure of assets
* **Impact of Tech outages on Spillovers/Highlanding – Aman (intern project)**
  + Identification of tech outages in history with timestamp and duration
  + Multi-class model for outage duration buckets leading to spillovers
* **Impact of Spillover on Breach – Aakansha**
  + Identify points of failure leading to spillover at various assets and sub-assets
  + Identify shipment stages which are affected by spillovers
  + Identify any pattern for shipments moving to subsequent stage (prioritization)
  + Evaluate the factor of impact on breach probability
* **Impact of External Factors leading to Breach – Aman (intern project)**
  + Identify external factors from RCA in the past 1-2 years
  + Identify recurring and seasonal factors
* **Lane level and Hub level breach performance over time**
* **Volumetric measures of consignment**
* **Dynamic E2E Planner Prediction – Aakansha**
* **Dynamic E2E Planner - Darshan**