# Assignment1_TeamA

**ZHIHAN WANG, LEI KANG, YUXIANG PENG**

## 1)    A list of project goals and deliverables

- Determine how to best leverage the Synchronoss analytics stack to fulfill the reporting and analytics use cases for IL customers
- Gather integration data points that will help decide the approach to potentially unify the analytics architecture
- PoC to demonstrate the following:
    - Using Synchronoss interfaces, ingest IL platform data into Synchronoss data lake from S3 or RAP ODS in AWS Redshift
    - Using Synchronoss analytics stack, process data as appropriate to prepare for reporting and analytics consumption
    - Optionally offer visualization of 1 or 2 reports in the self serve portal to demonstrate complete end to end flow
    - Study and report on ease of integrating into Synchronoss stack and the operational characteristics

## 2)    A list of tasks you completed this week and will be working on next week (with name of person(s) who owns the task)

- This week:
    - (Zhihan, Lei, Yuxiang) Had a group meeting with client on Friday and Monday discussing about the whole background and project goal
    - (Zhihan, Lei, Yuxiang) Decided how to  team communicate: Weekly chat on Monday 12:00 PM /email
- Next week:
    - (Yuxiang) Learn how to extract data from company's database
    - (Lei) Data scope established for transfer into Synchronoss data lakes
    - (Zhihan) Data model (star) and associated views for scoped out facts and dimensions

## 3)    List of challenges you are facing and possible risks

- Learning new softwares may require much time and patience.
- As the client mentioned, we will have to handle massive data from the existing database, the data processing needs to be efficient.
- According to the project goal, the timeline is tight and the workload is hugh
- It's difficult to conduct helpful and connective data analysis from different sources

## 4)    Preliminary thoughts on timeline for a few major deliverables  (ok if some dates are TBD)

- IL platform data successfully processed and persisted by the Synchronoss ingestion pipeline (8 weeks). Deliverables to minimally include:
    - Data scope established for transfer into Synchronoss data lakes

- - Data model (star) and associated views for scoped out facts and dimensions
  - Working ETL pipeline moving data into the Synchronoss data lakes – "pull" based will suffice
  - Reference scripts to demonstrate scalable data access for reporting made available
  - Performance characteristics of ingestion and other processing to be noted Stretch Goal
  - Options for data security – row level authorization and data masking Stretch Goal
- 1 or 2 reports made available in the self serve portal to demonstrate data availability and their performance (3 weeks). Deliverables to minimally include:
  - Document use of UI technology for report rendering and/or construction of dashboard
  - Approach to Authentication and Authorization for customers from IL Global User Directory Stretch Goal