MapReduce Assignment

Charlie Mei cm3947

**Inputs**

Doc1: Named Doc1, contains anchor ‘Columbia University’ which points to Columbia University Landing Page, anchor ‘SPS’ which points to the SPS Landing Page, anchor ‘NYU’ which points to NYU Landing Page, and ‘Columbia’ which points to Columbia University Landing Page

Doc2: Named Doc2, contains ‘Ivy League school’ which points to Columbia University Landing Page, ‘Apple’ which points to Apple Landing Page

Doc3: SPS Landing Page (hence, named 'SPS Landing Page') contains the anchor ‘the university’ which points to Columbia University Landing Page, the anchor ‘APAN’ which points to the Applied Analytics Program page

Doc4: Named Doc4, contains ‘the university’ which points to NYU Landing Page

Doc 5: Named Doc5, contains ‘iOS’ which points to Apple Landing Page, and contains ‘windows’ which points to Microsoft Landing page.

**Mapping**

<Columbia University Landing Page<Doc 1: ‘Columbia University’> 1

<SPS Landing Page<Doc 1: ‘SPS’>> 1

<NYU Landing Page<Doc 1: ‘NYU’>> 1

<Columbia University Landing Page<Doc 1: ‘Columbia’>> 1

<Columbia University Landing Page<Doc 2: ‘Ivy League school’>> 1

<Apple Landing Page<Doc 2: ‘Apple’>> 1

<Columbia University Landing Page<Doc 3: ‘the university’>> 1

<Applied Analytics Program page<Doc 3: ‘APAN’>> 1

<NYU Landing Page<Doc 4: ‘the university’>> 1

<Apple Landing Page<Doc 5: ‘iOS’>> 1

<Microsoft Landing Page<Doc 5: ‘windows’>> 1

**Sort and shuffle**

<Applied Analytics Program page<Doc 3: ‘APAN’>> 1

<Apple Landing Page<Doc 2: ‘Apple’>> 1

<Apple Landing Page<Doc 5: ‘iOS’>> 1

<Columbia University Landing Page<Doc 1: ‘Columbia’>> 1

<Columbia University Landing Page<Doc 1: ‘Columbia University’> 1

<Columbia University Landing Page<Doc 2: ‘Ivy League school’>> 1

<Columbia University Landing Page<Doc 3: ‘the university’>> 1

<Microsoft Landing Page<Doc 5: ‘windows’>> 1

<NYU Landing Page<Doc 1: ‘NYU’>> 1

<NYU Landing Page<Doc 4: ‘the university’>> 1

<SPS Landing Page<Doc 1: ‘SPS’>> 1

**Reduce**

<Applied Analytics Program page<Doc 3: ‘APAN’>> 1

<Apple Landing Page<Doc 2: ‘Apple’, Doc 5: ‘iOS>> 2

<Columbia University Landing Page<Doc 1: ‘Columbia’, Doc 1: ‘Columbia University’, Doc 1: ‘Ivy League school’, Doc 3: ‘the university’>> 4

<Microsoft Landing Page<Doc 5: ‘windows’>> 1

<NYU Landing Page<Doc 1: ‘NYU’, Doc 4: ‘the university’>> 2

<SPS Landing Page<Doc 1: ‘SPS’>> 1

**NODE ASSIGNMENTS**

